Recommended citation:
Burket, G. R. (1964). ***A Study of Reduced Rank Models for Multiple Prediction*** **(Psychometric Monograph No. 12). Richmond, VA: Psychometric Corporation. Retrieved from http://www.psychometrika.org/journal/online/MN12.pdf**

PSYCHOMETRIC MONOGRAPH NUMBER 12

A STUDY OF REDUCED RANK MODELS FOR MULTIPLE PREDICTION

This Study Was Supported in Part by
Office of Naval Research Contract Nonr-477(33)
and Public Health Research Grant M-743(C7)
Principal Investigator: Paul Horst

The William Byrd Press, Inc.
Richmond, Virginia

A STUDY OF REDUCED RANK MODELS FOR MULTIPLE PREDICTION

By

GEORGE R. BURKET
AMERICAN INSTITUTE FOR RESEARCH
AND
UNIVERSITY OF PITTSBURGH

A STUDY OF REDUCED RANK MODELS FOR MULTIPLE PREDICTION

By

GEORGE R. BURKET

AMERICAN INSTITUTE FOR RESEARCH

AND

UNIVERSITY OF PITTSBURGH

PREFACE

Prediction problems frequently arise in which the regression weights must be based on a relatively small number of criterion observations. In such cases, current techniques permit the utilization of only a very few predictors, even though many more may be available. Unless one or more of the predictors is closely related to the criterion, accurate predictions cannot be made. The possibility of increasing the accuracy of prediction under such circumstances through the use of reduced-rank methods is investigated in this study.

On the basis of normal regression theory, a general reduced-rank model is formulated in terms of prediction from factor scores. The problems of selecting a method of factoring, of selecting an optimal subset of prespecified size from among a given set of factors, and of selecting an optimal rank are considered. It is shown that in the absence of criterion observations, the optimally chosen reduced-rank solution will be the one that accounts for the greatest proportion of variance in the full-rank predictor matrix. Prediction either from subsets of the original predictors, which are equivalent to triangular factors, or from principal-axes factors is considered. It is concluded that, when degrees of freedom are sufficiently limited, the most accurate predictions obtainable will be those based on the largest principal-axes factors. As a tentative solution to the problem of optimal rank, estimates are derived which are intended to indicate the accuracy of prediction to be expected when regression weights computed on the basis of data in one sample are applied to data in other samples.

An empirical comparison of five reduced-rank methods is carried out, employing a variety of ranks, sample sizes, and criteria. The five methods include prediction from the principal-axes factors, selected in three different ways, and from the original predictors, selected in two different ways. The results indicate that weights computed by the method of largest principal-axes factors on samples with as few as 30 cases can give predictions as accurate as those from weights computed by conventional techniques on samples of several hundred cases.

The present monograph was submitted as a doctoral dissertation at the University of Washington in July 1962. The writer wishes to thank his sponsor, Professor Paul Horst, for the invaluable blend of criticism and encouragement that he provided. The work for the present monograph was largely supported by Office of Naval Research Contract Nonr. 477(33) and Public Health Research Grant M-743(C7) (principal investigator: Paul Horst). Acknowledgment is due Mrs. Judy Goodstein and Mrs. Helen Ranck for their work in typing and proofreading the manuscript.

George R. Burket

Pittsburgh, Pennsylvania
October, 1963

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1

INTRODUCTION

Basic Requirements

Accurate predictions of an individual's degree of success or failure in such socially significant activities as a college course, training for some vocation, or a particular job would be of incalculable utility, both to the individual concerned and to the community. Remarkably accurate predictions of this nature can be obtained with existing statistical techniques, provided that two basic requirements are satisfied. First, there must be measurements available on a number of variables related to performance in the activity of interest. It must be possible to obtain these measurements on any individual before he engages in the activity. Second, such measurements must be obtained for a large number of persons who subsequently engage in the activity.

The first requirement can almost always be met. Indeed, it is usually possible to find many variables having at least some relation to performance in the criterion activity. To obtain measurements on a large number of variables may be expensive, but accurate predictions of many activities are of sufficient value to warrant large expenditures. The second requirement is much less likely to be satisfied, since the number of persons who actually engage in a particular activity is often limited. This is particularly true for activities requiring an unusual degree of ability, where accurate predictions are apt to be most desired. Many socially significant activities are full-time occupations which individuals must pursue for years before their success or failure can be determined. If the number of persons engaging in such an activity is too small to permit application of existing techniques, no feasible expenditure will yield accurate predictions. We need new techniques.

The Statistical Model

A system for obtaining the best possible predictions for a given criterion would be the following. First, determine all variables, termed predictors, not statistically independent of the criterion. Then obtain measurements of predictors and criterion on a sufficiently large validation sample so that every possible configuration of predictor values is represented by a large number of cases. Compute the criterion mean for each of these configurations. To make a prediction for a particular case, determine the configuration of the predictors for that case. The prediction will be the criterion mean for cases in the validation sample having that configuration.

Such a system is unworkable because of practical limitations on sample size and number of predictors. Under certain circumstances, moreover, a much simpler system could give equally accurate predictions. If, for example, the criterion means were known to be functionally related to the predictors, it would only be necessary to determine this function. In practice, such a functional relation is virtually always assumed. It may also happen that a small subset of all variables statistically related to the criterion will give predictions as accurate as the entire set. Even where a very large number of independent predictors is readily available, the number that may actually be used is limited by the available sample size. This is because it is necessary to have many more cases than there are parameters in the assumed functional relation between predictors and criterion mean. Otherwise one could not obtain stable estimates of these parameters.

In least-squares or regression theory and also in correlation theory, the mean of the criterion is assumed to be a linear function of the predictors. In correlation theory, predictors and criterion are assumed to be random variables having a joint multivariate normal distribution. In regression theory, the criterion is assumed to be a normally distributed random variable, while the predictors are thought of as being fixed. Anderson (1958, p. 61) recommends using one model or the other depending on whether or not the predictors may be considered random. Mood (1950, p. 312) states that, in practice, most correlation problems can be more appropriately handled by regression methods. In many cases, the two models have led to equivalent procedures; under the null hypothesis, estimates of regression weights, test criteria, and probability theory are all the same. However, when the null hypothesis (viz., that predictors and criterion are independent) is not true, the probability theory differs.

In prediction problems in psychology, the predictor variables are generally random rather than fixed, and the null hypothesis is rarely true. Thus correlation theory would appear to be more appropriate. However, since correlation theory is considerably more complex and difficult to apply than regression theory, the latter is generally used, with the hope that the practical differences between conclusions drawn from the two models will be negligible. In the present study, prediction problems will for the most part be considered within the context of regression theory.

It may prove useful at this point to make the distinction between actual prediction problems and validation problems. In validation problems, the goal is to demonstrate a systematic relationship between a number of "independent variables" and a "dependent variable." To accomplish this, one formulates the null hypothesis of no relationship and hopes to reject it at some level of confidence. Thus, for validation problems, correlation theory and regression theory are equivalent. In prediction problems, on the other hand, the null hypothesis is assumed to be false. The goal is to obtain a regression equation which, when applied to predictor measures in future samples, will give the most accurate estimate possible of the corresponding criterion values. Having obtained such a regression equation, one would also wish to have estimates or confidence intervals indicating the accuracy to be expected when the regression equation is applied to new samples. In validation problems, the multiple correlation is often used as a measure of relationship between the dependent and independent variables. It is sometimes termed a validity coefficient, or simply a validity. In prediction problems, the correlation between the prediction and the criterion in new samples may be used as a measure of accuracy of prediction. Such a coefficient may be termed a weight-validity to distinguish it from the multiple correlation coefficient between the prediction battery and the criterion in the original sample.

Purpose of the Study

The present study is concerned with prediction problems as opposed to validation problems. Regression theory in its current form is adequate for those applications in which the available number of cases far exceeds the available number of predictors, i.e., in which the number of degrees of freedom is large. In such cases, weight-validity will be very close to battery validity, and the least-squares estimates of the regression weights will provide optimal predictions. But when the number of predictors available is relatively large in relation to sample size, as is perhaps more often than not the case, problems arise that lack satisfactory theoretical answers. One such problem is that of estimating an index, such as weight-validity, that will provide some idea of the accuracy of prediction to be expected in new samples. A more important problem is that of determining the regression weights which will give the most accurate predictions possible in new samples.

These optimal weights will not in general be given by the conventional least-squares solution applied to all available predictors. For example, if the number of predictors is the same as the number of cases in the sample, the least-squares weights for an arbitrary subset of predictors will usually give better weight-validity (though lower validity) than the weights for the entire set. More generally, in such an extreme case, any lower-rank approximation to the matrix of predictor values would give better predictions than the complete matrix. As the situation becomes less and less extreme, there must come a point where some ranks and some methods of rank reduction and not others are preferable to the complete matrix. At a still less extreme point, the entire set of predictors will presumably give better predictions than any reduced-rank approximation. Still, when predictors are discarded, the loss of accuracy of prediction may be so slight as to be more than offset by the practical savings of not having to measure as many predictors.

Thus in any prediction problem where the number of degrees of freedom is limited, the question of rank reduction arises: can the complete predictor matrix be improved upon, and if so, which method of reduction and which rank will give the greatest improvement? When its purpose is to give more accurate prediction by increasing degrees of freedom, the much-studied predictor selection problem is a special case of the rank-reduction problem. Predictor selection methods are more often used, however, in situations where an upper limit on the size of the prediction battery is given by considerations of cost. The emphasis is thus on obtaining an optimal set of predictors of a particular size rather than on obtaining optimal predictions regardless of battery size. Perhaps because of the prevalence of the former emphasis, particularly before the advent of electronic computers, the problem of predictor selection has received a great deal more attention than the general problem of rank reduction.

Most methods of predictor selection are alike in selecting first the variable having the highest single validity, and adding, step by step, the variable which, together with those previously selected, will give the greatest increase in the multiple correlation with the criterion. These so-called accretion methods differ with respect to computational procedure and method of deciding how many predictors to use. Perhaps the computationally simplest such method is the square-root (or triangular-factoring) method described by Summerfield and Lubin (1951). Horst has generalized and extended this method for absolute (1955) and differential (1954) prediction of multiple criteria. Horst and MacEwan (1960) have described a method which is essentially the reverse of the accretion method. Here one eliminates at each step the predictor contributing least to the multiple correlation. The accretion and elimination methods will not in general result in the same battery, nor will either of them necessarily give the battery of given size having the highest obtainable validity.

Horst (1941) has suggested two models for reduced-rank prediction. His rationale is based upon the factor analysis hypothesis that the predictor matrix is basic only because of the presence of error or specific factors. One of these models assumes the presence of specifics. Accordingly, the matrix of predictor intercorrelations is augmented by the vector of criterion correlations and communality estimates are placed in the diagonal prior to factoring. Least-squares weights are then computed for the common factors. This method was tested by Leiman (1951) using 12 predictors and computing weights on samples of 30 cases. A rank-3 solution gave weight-validities which were significantly higher than those obtained with the full-rank solution. This method has the disadvantage of being difficult to treat theoretically, since the nature of communalities and of the factor scores (which are not unique) are not well understood. The other model suggested by Horst accomplishes rank reduction by attempting to remove error factors rather than specific factors. Here the best least-squares approximation to the predictor intercorrelation matrix is used, the principal-axes solution. One advantage of this method is that it is theoretically straightforward. Another advantage is that rank reduction is accomplished independently of the criterion and thus does not capitalize on the errors in the criterion.

Virtually the exact opposite of this model has been implicitly suggested by Guttman (1958). Since the inverse of the predictor correlation matrix is directly involved in computing regression weights, one might well base predictions on the best lower-rank approximation to the inverse rather than on the approximation to the intercorrelation matrix. The best set of factors for approximating the inverse is, as Guttman points out, the worst for approximating the intercorrelation matrix. In view of this paradox, perhaps one should abandon approximation as a criterion for selecting the factors to be retained for prediction and simply use those factors giving the highest multiple correlation, as is attempted in the predictor-selection methods. Certainly the basic assumption of the rationale for approximating the intercorrelation matrix may be questioned: that the reliable variance is concentrated in the larger princpal-axes factors, the smaller factors being composed mainly of error. For example, in a study by Davis (1945) involving nine principal-axes factors, a strict correspondence between variance contribution and reliability was not found; e.g., the split-half reliability for the eighth factor was larger than for the fourth factor.

The present study proceeds along both theoretical and empirical lines. First an attempt is made to work out some of the consequences of regression theory for reduced-rank models. Since, as noted above, there is reason to question the appropriateness of regression theory for psychological prediction problems, an empirical comparison of five reduced-rank procedures is also carried out. The methods used were predictor elimination, predictor selection, the method of approximating the intercorrelation matrix, the method of approximating the inverse, and the method using the principal-axes factors giving the highest multiple correlation. As will be seen, both the theoretical and the empirical evidence favors the method of approximating the intercorrelation matrix.

CHAPTER 2

IMPLICATIONS OF REGRESSION THEORY FOR REDUCED RANK MODELS

The General Linear Hypothesis

Regression theory was first worked out at the beginning of the 19th century by Gauss and Legendre and has since, of course, been presented by innumerable authors from various points of view. Among recent sources, a rigorous presentation with geometrical interpretations has been given by Scheffé (1959). A simpler presentation entirely in terms of matrix algebra is given by Kempthorne (1952). Anderson (1958) provides a generalization to multiple criteria. A presentation in terms of deviation scores may be found in Cramér (1946). Some results from regression theory which are relevant to the rank-reduction problem are summarized below. The derivations, which are for the most part omitted, may be found in the sources mentioned above. Let

y be a column vector of N observations on the criterion;
x be an $N \times M$ matrix of rank $M < N$, each row of which represents an observation on each of M predictors;
e be an Nth-order column vector of uncorrelated errors, each distributed normally with mean zero and variance σ^2;
β be an $M \times 1$ vector of population regression coefficients;
C be a covariance matrix of the variable given in the subscript.

The general linear hypothesis is that

(1) $$y = x\beta + e.$$

The assumptions regarding e, apart from normality, may be stated as

(2) $$E(e) = 0,$$

(3) $$C_e = E(ee') = \sigma^2 I.$$

From these equations it follows that the criterion has the expectation

(4) $$E(y) = x\beta,$$

and the covariance matrix

(5) $$C_y = E[(y - x\beta)(y - x\beta)'] = \sigma^2 I.$$

Let

$\hat{\beta}$ be the $M \times 1$ vector of least-squares estimates of the population regression coefficients;
$\tilde{y}$ be the $N \times 1$ vector of estimates of the criterion based on $\hat{\beta}$.

Then

(6) $$\hat{\beta} = (x'x)^{-1}x'y,$$

and

(7) $$\tilde{y} = x\hat{\beta}.$$

The vector $\hat{\beta}$ has the property of minimizing the sum of squares of the errors in estimating y from $\tilde{y}$. These errors will be orthogonal to the predictors and also to the estimates themselves. The error sum of squares has the expectation

(8) $$E[(y - \tilde{y})'(y - \tilde{y})] = (N - M)\sigma^2.$$

Thus

(9) $$\hat{\sigma}^2 = \frac{(y - \tilde{y})'(y - \tilde{y})}{N - M}$$

provides an unbiased estimate of σ^2. What is generally termed the standard error of estimate is given by $\hat{\sigma}$. The variable $\hat{\sigma}^2$ is distributed independently of $\hat{\beta}$.

The estimates of the regression coefficients have the expectation

(10) $$E(\hat{\beta}) = \beta,$$

and the covariance matrix

(11) $$C_{\hat{\beta}} = E[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'] = \sigma^2(x'x)^{-1}.$$

The estimates of the criterion have the same expectation as the criterion itself,

(12) $$E(\tilde{y}) = E(x\hat{\beta}) = x\,E(\hat{\beta}) = x\beta,$$

but are not independent, since from (7), (11), and (12),

(13) $$C_{\tilde{y}} = E[(x\hat{\beta} - x\beta)(x\hat{\beta} - x\beta)'] = xC_{\hat{\beta}}x' = \sigma^2x(x'x)^{-1}x'.$$

The canonical form of the general linear hypothesis may be obtained as follows. Let x be expressed as

(14) $$x = ub',$$

where u is an $N \times M$ orthonormal matrix of factor scores, and b is an $M \times M$ matrix of factor loadings. Let V be an N by $N - M$ orthonormal matrix such that the $N \times N$ matrix H in

(15) $$H = [u \quad v]$$

is an orthonormal matrix. The matrices u, b, and v are always obtainable, and can be determined solely from the predictors without reference to the criterion. Then the Nth-order vector of transformed criterion values

$$z = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = H'y = \begin{bmatrix} u'y \\ v'y \end{bmatrix} \tag{16}$$

has the expectation

$$E(z) = \begin{bmatrix} E(z_1) \\ E(z_2) \end{bmatrix} = \begin{bmatrix} b'\beta \\ 0 \end{bmatrix}, \tag{17}$$

and the covariance matrix

$$C_z = \sigma^2 I. \tag{18}$$

Thus the best possible predictions for the $N - M$ transformed observations z_2 will always be zero, regardless of the true regression coefficients or of the particular values of the criterion. The least-squares estimates of the regression weights are so chosen as to reproduce exactly the M transformed observations z_1 from

$$z_1 = u'y = b'\hat{\beta}, \tag{19}$$

so that

$$\hat{\beta} = b'^{-1}u'y. \tag{20}$$

Equation (20) may also be obtained by putting (14) in (6). Thus, errors can occur only in estimating z_2, and since the estimated value of z_2 is zero, we have

$$(y - \tilde{y})'(y - \tilde{y}) = z_2'z_2. \tag{21}$$

Metric and the Status of the Multiple Correlation

In regression theory, the multiple correlation coefficient and other functions of the predictors such as means, standard deviations, and covariances do not have the status of population parameters. This is because the predictors are not assumed to be random variables but rather fixed values. Thus, regression theory does not admit of statistical inferences about such functions. However, one can make statistical inferences about such characteristics of future samples as depend on the criterion, provided that the relevant features of the predictor matrix in the future samples are assumed to be known in advance. For example, one can assume that exactly the same predictor matrix will be obtained in future samples or merely that the predictor intercorrelations will be the same. Using the latter assumption and scaling the criterion appropriately, one can define both a sample and a population multiple correlation coefficient.

Except where correlations are concerned, no assumptions about metric are made in the present paper. However, it should be noted that if the equations of the preceding section were to be applied to data in the original units of observation, a correction for origin would be required. This correction will be accomplished if a predictor is added which is defined to be unity for all cases. If this is done, equation (6) of the preceding section may be shown to be identical to the usual formulas for raw-score regression weights, which are typically expressed in terms of means and covariances or correlations and standard deviations.

The question of metric also arises in connection with defining multiple correlation. The assumption made here whenever correlation coefficients are discussed is that all measures are normalized, i.e., expressed as deviations from the sample mean in units of the sample standard deviation multiplied by the square root of the number of cases in the sample. We may now define the square of the multiple correlation in the sample as

$$R^2 = \hat{\beta}'x'x\hat{\beta} = y'x(x'x)^{-1}x'y \tag{22}$$

and in the population as

$$\rho^2 = \beta'x'x\beta. \tag{23}$$

If we let r be the $M \times M$ matrix of predictor intercorrelations, (23) may be written as

$$\rho^2 = \beta'r\beta, \tag{24}$$

since, on the basis of the assumption about the metric,

$$r = x'x. \tag{25}$$

Thus ρ will be a population parameter if it is assumed that the predictor intercorrelations will be the same in all samples.

An unbiased estimate for ρ may be obtained as follows. The expectation of the criterion sum of squares is, from (1),

$$E(y'y) = E[(x\beta + e)'(x\beta + e)] = \beta'x'x\beta + 2\beta'x'E(e) + E(e'e). \tag{26}$$

From (23), the first term on the right is ρ^2 and from (2) the second term is zero. The third term is the trace of (3). Thus

$$E(y'y) = \rho^2 + N\sigma^2. \tag{27}$$

Since the errors of estimate are orthogonal to the estimates, we have

$$y'y = \tilde{y}'\tilde{y} + (y - \tilde{y})'(y - \tilde{y}). \tag{28}$$

From (7) and (22), the first term on the right is R^2. Thus from (8) and (27),

$$\begin{aligned} E(R^2) &= E(y'y) - E[(y - \tilde{y})'(y - \tilde{y})] \\ &= \rho^2 + N\sigma^2 - (N - M)\sigma^2 = \rho^2 + M\sigma^2. \end{aligned} \tag{29}$$

Given the assumed metric, the criterion sum of squares will always be unity, so from (27),

$$\sigma^2 = \frac{1 - \rho^2}{N} \tag{30}$$

and (29) may be written as

$$E(R^2) = \rho^2 + \frac{M(1 - \rho^2)}{N}. \tag{31}$$

From (31) it is clear that the extent to which R^2 overestimates ρ^2 will vary directly with the number of predictors and inversely with the sample size. Solving equation (31) for ρ^2 we obtain the following unbiased estimate for ρ^2:

$$R_C^2 = \frac{NR^2 - M}{N - M}. \tag{32}$$

Equation (32) will be recognized as the familiar "shrinkage" formula for multiple R.

It is perhaps worth noting that R_C, or "shrunken R" is not an estimate of weight-validity or of the shrinkage to be expected in the correlation between the criterion and its estimate if weights computed on one sample are applied in other samples. It does provide an estimate of the correlation that would have been obtained between the criterion and its estimate if the population regression weights had been used instead of their least-squares estimates. Shrunken R may also be thought of as an estimate of the multiple R that could be obtained in a very large sample having the same predictor intercorrelation matrix as the observed sample.

The Accuracy of Prediction in Future Samples

In prediction problems we wish to compute a set of weights from a given sample which will give the most accurate predictions obtainable when applied to other samples. Specifically, we will assume that the sum of squares of the errors of prediction in each other sample is the quantity to be minimized. If we let $\bar{\beta}$ be a set of weights obtained in some fashion from a previous sample, this sum of squares may be written (Kempthorne, 1952) as

$$(y - x\bar{\beta})'(y - x\bar{\beta}) = (y - x\hat{\beta})'(y - x\hat{\beta}) + e'x(x'x)^{-1}x'e + 2(\beta - \bar{\beta})'x'e + (\beta - \bar{\beta})'x'x(\beta - \bar{\beta}). \tag{33}$$

The expected value is

$$E[(y - x\bar{\beta})'(y - x\bar{\beta})] = N\sigma^2 + (\beta - \bar{\beta})'x'x(\beta - \bar{\beta}). \tag{34}$$

Now the second term on the right has an expectation in the sample from which $\bar{\beta}$ was obtained. Assuming that the usual least-squares estimates are employed, we have, using equation (11),

$$(35) \quad E[(\beta - \hat{\beta})'x'x(\beta - \hat{\beta})] = \text{tr}\,[E[x(\beta - \hat{\beta})(\beta - \hat{\beta})'x']] = \text{tr}\,(xC_{\hat{\beta}}x') = \sigma^2\,\text{tr}\,[x(x'x)^{-1}x'].$$

Using (14), we may write the matrix whose trace we require as

$$(36) \quad x(x'x)^{-1}x' = ub'(bb')^{-1}bu' = ub'b'^{-1}b^{-1}bu' = uu'.$$

Putting (36) in (35), we may write

$$(37) \quad E[(\beta - \hat{\beta})'x'x(\beta - \hat{\beta})] = \sigma^2\,\text{tr}\,(uu') = \sigma^2\,\text{tr}\,(u'u) = \sigma^2\,\text{tr}\,(I) = M\sigma^2.$$

Now if we assume that $x'x$, or equivalently the factor-loading-matrix b, is the same in all samples, we would expect the sum of squares of errors of prediction to be $(N + M)\sigma^2$. More generally, if $\bar{\beta}$ is any estimate of β computed from the original sample, we would expect the sum of squares of errors of prediction in future samples, provided that the factor-loading matrix is the same as in the original sample, to be

$$(38) \quad \psi_{\bar{\beta}} = N\sigma^2 + E[(\beta - \bar{\beta})'x'x(\beta - \bar{\beta})].$$

Thus $\psi_{\bar{\beta}}$ will be taken as an inverse index of weight-efficiency: the smaller it is, the more suitable $\bar{\beta}$ will be for a prediction problem. In particular,

$$(39) \quad \psi_{\hat{\beta}} = (N + M)\sigma^2.$$

Since the interpretation of (38) is basic to the following development, we will examine its derivation with some care. Certainly $\psi_{\bar{\beta}}$ is not a mathematical expectation in the usual sense, but rather an expectation of an expectation. Since N, σ^2, β, and (by assumption) $x'x$ are fixed, the expectation in (34) is a function of $\bar{\beta}$, and is thus fixed as soon as the original sample is drawn. Since this quantity is a function of the criterion in the original sample, its expectation in this sample is $\psi_{\bar{\beta}}$. The quantity that we are directly concerned with minimizing is the one in (34). This quantity is itself not determined in advance of drawing the first sample, but its expectation is determined. Rather than minimize the quantity of direct interest, then, we attempt to minimize its expectation.

An estimate of weight-validity may be obtained from (39). Assuming the metric of the previous section, and using (9) and (22),

$$(40) \quad \hat{\sigma}^2 = \frac{y'y - \tilde{y}'\tilde{y}}{N - M} = \frac{1 - R^2}{N - M}.$$

Thus, an unbiased estimate for $\psi_{\hat{\beta}}$ is, from (39)

$$(41) \quad \hat{\psi}_{\hat{\beta}} = \frac{N + M}{N - M}(1 - R^2).$$

For an arbitrary set of weights $\bar{\beta}$, the weight-validity is

$$W = \frac{y'x\bar{\beta}}{\sqrt{\bar{\beta}'x'x\bar{\beta}}}. \tag{42}$$

The sum of squares of errors of prediction is

$$S = (y - x\bar{\beta})'(y - x\bar{\beta}) = 1 - 2y'x\bar{\beta} + \bar{\beta}'x'x\bar{\beta}. \tag{43}$$

If (42) is substituted in (43),

$$S = 1 - 2W\sqrt{\bar{\beta}'x'x\bar{\beta}} + \bar{\beta}'x'x\bar{\beta} . \tag{44}$$

Since $\bar{\beta}$ is the vector of least-squares weights from the original sample, under the assumption that $x'x$ is constant, the radical in the second term on the right of (44), and the third term on the right become, respectively, R and R^2 of the original sample. Solving (44) for W gives

$$W = \frac{1 + R^2 - S}{2R}. \tag{45}$$

Now to obtain an estimate of W, we substitute for S in (45) the estimate of its expectation given by (41). Simplifying, we obtain

$$\hat{W} = \frac{NR^2 - M}{R(N - M)}. \tag{46}$$

To see the relation of the estimated weight-validity to the estimated population multiple correlation as defined in the preceding section, we put (32) in (46), obtaining

$$\hat{W} = \frac{R_C^2}{R} = \frac{R_C}{R} R_C. \tag{47}$$

Since R_C is less than R (unless R is unity), the left-hand factor on the right of (47) will be less than one, so $\hat{W}$ will be less than R_C.

Perhaps a more important application of (38) is its use as a criterion for evaluating reduced-rank models for computing regression weights. An alternate approach is indirectly suggested by Leiman (1951, pp. 3–4). There, the assumption is made that the least-squares weights for the lower-rank system will give better predictions than least-squares weights for the full-rank system to the extent that they provide closer approximations to the population regression weights for the full-rank battery. The reason for rejecting this position is as follows: It is well known that the optimal weights for a subset of predictors may differ greatly from the weights of the same predictors when the full battery is retained. A mathematical statement of this fact is given in (104). Thus one cannot properly measure the suitability of a reduced-rank set of weights in terms of how closely they approximate the full-rank weights. It seems more likely that the least-squares weights for a subset of predictors or of factor scores may, because of the increased number of degrees of freedom, be so much more stable than the weights for the full set as to give more accurate predictions despite the loss of information. In any case, the criterion in (38) involves no assumptions other than those usually made in applications of regression theory to prediction problems and is, moreover, referred directly to the expected errors of prediction.

In evaluating reduced-rank solutions, a question arises as to the number of factors to be included in the general linear hypothesis. If the full-rank hypothesis is retained, then the quantity $N\sigma^2$ in (38) is fixed, so that the only way of improving on $\hat{\beta}$ will be to find a $\bar{\beta}$ for which the second term is less than $M\sigma^2$. If, however, a smaller set of, say, L predictors (either the original ones or factor scores) is hypothesized, both terms change. The variance of the errors, σ^2, will of course increase in proportion to the systematic variance in the criterion accounted for by the discarded predictors. If we denote this larger variance by σ_L^2 and the least-squares weights for the reduced battery by $\bar{\beta}$, then

(48) $$\psi_{\bar{\beta}} = (N + L)\sigma_L^2,$$

as will be seen in the next section. Thus the $\bar{\beta}$ for any subset of L predictors for which $(N + L)\sigma_L^2$ is less than $(N + M)\sigma^2$ will be an improvement over $\hat{\beta}$.

Another possible application of (38) would be in obtaining a criterion for how many predictors to retain in the standard predictor-selection procedures. If we denote by R_L the multiple correlation obtained with a set of L predictors, this criterion is obtained directly from (41):

(49) $$\hat{\psi}_{\bar{\beta}} = \frac{N + L}{N - L}(1 - R_L^2).$$

One would retain those L predictors for which $\hat{\psi}_{\bar{\beta}}$ is the smallest. We use $\hat{\psi}_{\bar{\beta}}$ rather than $\hat{W}$ since weight-validity is an indication not of the actual errors of prediction but of the errors which would have been obtained if the predictions could themselves have been weighted after the criterion had been observed. In other words, a correlation coefficient between two variables is independent of differences in location and scale, whereas actual errors of prediction are in part determined by such differences.

The General Reduced-Rank Model

The reduced-rank solution will first be developed in terms of a general factor model. Predictor selection and prediction from principal-axes factors will then be considered as special cases of this model. Let

(50) $$x'x = bb'$$

be any complete factoring of $x'x$. Then

(51) $$u = x(b')^{-1}$$

will be the orthonormal matrix of factor scores. The matrices x, u, and b are the same as those in (14). Now we partition u and b after the Lth column so that, from (14),

$$x = [u_1 \quad u_2] \begin{bmatrix} b_1' \\ b_2' \end{bmatrix} = u_1 b_1' + u_2 b_2'. \tag{52}$$

We will assume that the columns of u and b have been permuted so that the L factor scores retained for prediction are given by u_1, or (if one prefers to think of prediction from a rank-L approximation to x) by $u_1 b_1'$. We will now show that the two assumptions are equivalent for prediction problems. Note first, however, that in future samples the weights must be applied to the predictors rather than to the factor scores or to the lower-rank approximation. The latter must be obtained as a row transformation of the prediction matrix, since a prediction equation must be applicable to individual cases.

Let the inverse of b be conformably partitioned and denoted by B' so that

$$B'b = \begin{bmatrix} B_1' \\ B_2' \end{bmatrix} [b_1 b_2] = \begin{bmatrix} B_1' b_1 & B_1' b_2 \\ B_2' b_1 & B_2' b_2 \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & I \end{bmatrix}. \tag{53}$$

Then

$$u_1 = x B_1 \tag{54}$$

is a unique solution for u_1 as a transformation on the rows of x. To see this, we let γ be any other such transformation, and let

$$E = \gamma - B_1. \tag{55}$$

Then

$$u_1 = x\gamma = xB_1 + xE = u_1 + xE \tag{56}$$

so that

$$xE = 0, \tag{57}$$

which, since x is basic, implies that E is zero. Now let $\hat{\beta}_u$ be a set of least-squares weights for u_1. Since u_1 is basic, $\hat{\beta}_u$ is unique. Let $\hat{\beta}_b$ be a set of least-squares weights for $u_1 b_1'$. Since $u_1 b_1'$ is nonbasic, $\hat{\beta}_b$ is not unique. If

$$u_1 b_1' \hat{\beta}_b - y = \epsilon_b \tag{58}$$

and

$$u_1 \hat{\beta}_u - y = \epsilon_u, \tag{59}$$

the sums of squares of ϵ_b and of ϵ_u will be minimized by $\hat{\beta}_b$ and $\hat{\beta}_u$, respectively. The former sum of squares can be no less than the latter, for we could always take

$$\hat{\beta}_u = b_1' \hat{\beta}_b. \tag{60}$$

The two sums of squares will be equal if we let

$$\hat{\beta}_b = B_1\hat{\beta}_u. \tag{61}$$

Therefore, a set of least-squares weights for (58) will be given by $\hat{\beta}_b$ in (61) and

$$\epsilon_b'\epsilon_b = \epsilon_u'\epsilon_u. \tag{62}$$

But since $\hat{\beta}_u$ is unique, $b_1'\hat{\beta}_b$ must be unique, and (60) holds for all least-squares solutions $\hat{\beta}_b$ of (58). Thus, (58) and (59) are identical, and because of the uniqueness of B_1 in (54), we have

$$\bar{\beta} = B_1\hat{\beta}_u \tag{63}$$

as a unique set of least-squares weights for x under the assumption of reduced rank.

If it is assumed that the criterion depends solely on the subset of L factors retained for prediction, the general linear hypothesis takes the form

$$y = xB_1\beta_u + e_L, \tag{64}$$

where x, y, and e_L are defined in the first section of this chapter. All of the results of that section may be obtained for the present hypothesis if we replace x by xB, and β by β_u in (1) through (13). In like manner, (48) may be obtained from the derivation of (39). Thus, from (6) and (54) the least-squares estimate of β_u is given by

$$\hat{\beta}_u = (u_1'u_1)^{-1}u_1'y = u_1'y. \tag{65}$$

It has, from (10), the expectation

$$E(\hat{\beta}_u) = \beta_u \tag{66}$$

and, from (11), the covariance matrix

$$C_{\hat{\beta}_u} = \sigma_L^2(u_1'u_1)^{-1} = \sigma_L^2 I. \tag{67}$$

An unbiased estimate of the vector of weights to be applied directly to the predictors is given by $\bar{\beta}$ as defined in (63), since

$$E(\bar{\beta}) = E(B_1\hat{\beta}_u) = B_1E(\hat{\beta}_u) = B_1\beta_u. \tag{68}$$

The covariance matrix for these weights will be

$$C_{\bar{\beta}} = E[(B_1\hat{\beta}_u - B_1\beta_u)(B_1\hat{\beta}_u - B_1\beta_u)'] = B_1C_{\hat{\beta}_u}B_1' = \sigma_L^2B_1B_1'. \tag{69}$$

The estimates of the criterion will now be, from (7),

$$\tilde{y}_L = xB_1\hat{\beta}_u = x\bar{\beta}. \tag{70}$$

The expected sum of squares for the errors of estimate becomes, from (8),

$$E[(y - \tilde{y}_L)'(y - \tilde{y}_L)] = (N - L)\sigma_L^2. \tag{71}$$

The matrix H for transforming the criterion observations to canonical form may take exactly the same form as in (15):

$$H = (u_1\ u_2\ v). \tag{72}$$

The matrix $[u_2 \quad v]$ is now arbitrary to the extent that only v was arbitrary before. It will be convenient, however, to define H as in (72). Partitioning the transformed observations somewhat differently from the way it was done in (16), we let

$$z = \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} = Hy = \begin{bmatrix} u_1'y \\ u_2'y \\ v'y \end{bmatrix}. \tag{73}$$

The elements of z_2 and z_3 will all have expected values of zero, while the expectation of z_1 will be

$$E(z_1) = E(u_1'y) = E(\hat{\beta}_u) = \beta_u. \tag{74}$$

The unbiased estimate for σ_L^2 may be expressed in terms of z_2 and z_3 as

$$\hat{\sigma}_L^2 = \frac{z_2'z_2 + z_3'z_3}{N - L}. \tag{75}$$

The implications of using a reduced-rank solution instead of the conventional solution can perhaps be better understood if the full-rank hypothesis of (1) is retained, rather than the rank-L hypothesis of (64). We first observe that $\bar{\beta}$ is a biased estimate of β, since

$$E(\bar{\beta}) = E(B_1u_1'y) = B_1u_1'x\beta = B_1b_1'\beta. \tag{76}$$

Its covariance matrix, which will now be proportional to σ^2 instead of to σ_L^2, is given by

$$C_{\bar{\beta}} = E[(B_1u_1'y - B_1b_1'\beta)(B_1u_1'y - B_1b_1'\beta)'] = B_1E(u_1'ee'u_1)B_1' \tag{77}$$

since premultiplying (1) by u_1' gives

$$u_1'y = b_1'\beta + u_1'e. \tag{78}$$

Continuing, with (3) in (77),

$$C_{\bar{\beta}} = B_1u_1'E(ee')u_1B_1' = \sigma^2B_1B_1'. \tag{79}$$

The first and second moments about β will be

$$E(\bar{\beta} - \beta) = B_1b_1'\beta - \beta = -(I - B_1b_1')\beta = -B_2b_2'\beta \tag{80}$$

and

$$\begin{aligned} &E[(\bar{\beta} - \beta)(\bar{\beta} - \beta)'] \\ &\quad = C_{\bar{\beta}} + [E(\bar{\beta} - \beta)][E(\bar{\beta} - \beta)]' = \sigma^2B_1B_1' + B_2b_2'\beta\beta'b_2B_2'. \end{aligned} \tag{81}$$

Equation (11) may be written as

$$C_{\hat{\beta}} = \sigma^2(x'x)^{-1} = \sigma^2 BB' = \sigma^2 B_1B_1' + \sigma^2 B_2B_2'. \tag{82}$$

Thus, from the standpoint of relative efficiency (Mood, 1950, p. 149) in estimating β, $\hat{\beta}$ and $\bar{\beta}$ may be compared in terms of the diagonals of the rightmost terms of (81) and (82). If the trace of the former is smaller, on the average the reduced-rank estimates will be more efficient than the full-rank estimates.

The expected value of z as given by (73) will now be

$$E(z) = \begin{bmatrix} u_1'x\beta \\ u_2'x\beta \\ v'x\beta \end{bmatrix} = \begin{bmatrix} b_1'\beta \\ b_2'\beta \\ 0 \end{bmatrix}. \tag{83}$$

We recall from (19) that $\hat{\beta}$ is computed so that

$$\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} b_1'\hat{\beta} \\ b_2'\hat{\beta} \end{bmatrix}. \tag{84}$$

But $\bar{\beta}$ is computed to reproduce only z_1:

$$z_1 = u_1'y = b_1'B_1u_1'y = b_1'\bar{\beta}. \tag{85}$$

We have

$$b_2'\bar{\beta} = b_2'B_1u_1'y = 0. \tag{86}$$

Thus, the reduced-rank solution, in effect, predicts a value of zero for z_2 rather than a value of $b_2'\hat{\beta}$. If the elements of $b_2'\beta$ are smaller than σ^2, then the prediction of zero would have the higher relative efficiency.

The statistic $\hat{\sigma}_L^2$ will be an overestimate of σ^2. To see this, first note that

$$\begin{aligned} E(z_2'z_2 + z_3'z_3) &= \text{tr}\,[E(z_2z_2')] + \text{tr}\,[E(z_3z_3')] \\ &= \text{tr}\,(\sigma^2 I + b_2'\beta\beta'b_2) + \text{tr}\,(\sigma^2 I) \\ &= (M - L)\sigma^2 + \beta'b_2b_2'\beta + (N - M)\sigma^2 \\ &= (N - L)\sigma^2 + \beta'b_2b_2'\beta. \end{aligned} \tag{87}$$

Then from (75),

$$E(\hat{\sigma}_L^2) = \sigma^2 + \frac{\beta'b_2b_2'\beta}{N - L}. \tag{88}$$

Next, we describe the effect of hypothesized rank on our inverse index of weight-efficiency, $\psi_{\bar{\beta}}$. We will denote this index and its estimate by ${}_M\psi_{\bar{\beta}}$ and ${}_M\hat{\psi}_{\bar{\beta}}$, where the full rank M is assumed, and by ${}_L\psi_{\bar{\beta}}$ and ${}_L\hat{\psi}_{\bar{\beta}}$, where the reduced-rank, L, is assumed. Mathematical expectation under the hypothesis of full rank will be denoted by $E_M(\ )$ and under the hypothesis of reduced-rank by $E_L(\ )$.

The reduced-rank index ${}_L\psi_{\bar{\beta}}$ was given by (48). To obtain the full-rank index, we first evaluate the rightmost term in (38). Using (81),

$$(89)\qquad \begin{aligned} E_M[(\beta - \bar{\beta})'x'x(\beta - \bar{\beta})] &= \operatorname{tr}\,[xE(\bar{\beta} - \beta)(\bar{\beta} - \beta)']x'] \\ &= \sigma^2 \operatorname{tr}\,(xB_1B_1'x') + \operatorname{tr}\,(xB_2b_2'\beta\beta'b_2B_2'x') \\ &= \sigma^2 \operatorname{tr}\,(u_1u_1') + \operatorname{tr}\,(u_2b_2'\beta\beta'b_2u_2') \\ &= \sigma^2 \operatorname{tr}\,(u_1'u_1) + \beta'b_2u_2'u_2b_2'\beta \\ &= L\sigma^2 + \beta'b_2b_2'\beta. \end{aligned}$$

Substituting (89) in (38), we obtain

$$(90)\qquad {}_M\psi_{\bar{\beta}} = (N + L)\sigma^2 + \beta'b_2b_2'\beta.$$

An unbiased estimate of ${}_L\psi_{\bar{\beta}}$ is, from (75) and (48),

$$(91)\qquad {}_L\hat{\psi}_{\bar{\beta}} = (N + L)\hat{\sigma}_L^2 = z_2'z_2 + z_3'z_3 + \frac{2L}{N - L}(z_2'z_2 + z_3'z_3).$$

An unbiased estimate of ${}_M\psi_{\bar{\beta}}$ is, from (87),

$$(92)\qquad {}_M\hat{\psi}_{\bar{\beta}} = z_2'z_2 + z_3'z_3 + \frac{2L}{N - M}z_3'z_3.$$

The latter will also be an unbiased estimate of ${}_L\psi_{\bar{\beta}}$, since

$$(93)\qquad E_L\left(\frac{z_3'z_3}{N - M}\right) = \sigma_L^2.$$

It would not, however, be as stable an estimate as ${}_L\hat{\psi}_{\bar{\beta}}$, since the rightmost term of (91) is based on more observations than the rightmost term of (92). If ${}_L\hat{\psi}_{\bar{\beta}}$ were used to estimate ${}_M\psi_{\bar{\beta}}$, it would have a positive bias, since, from (88) and (90),

$$(94)\qquad E_M({}_L\hat{\psi}_{\bar{\beta}}) = (N + L)\left(\sigma^2 + \frac{\beta'b_2b_2'\beta}{N - L}\right) = {}_M\psi_{\bar{\beta}} + \frac{2L}{N - L}\beta'b_2b_2'\beta.$$

In practice, it would often be convenient to express these estimates in terms of the multiple correlation coefficient. If the metric of the third section is assumed, the elements of z_1 and z_2 will be the correlations between the factor scores and the criterion, or factor validities. Since the factor scores are uncorrelated, the squared multiple correlation between the first L factors and the criterion will be

$$(95)\qquad R_L^2 = z_1'z_1 = 1 - z_2'z_2 - z_3'z_3.$$

Hence (91) and (92) are equivalent to

$$ {}_L\hat{\psi}_{\bar{\beta}} = 1 - \underline{R_L^2} + \frac{2L(1 - R_L^2)}{\underline{N} - L}, \tag{96} $$

and

$$ {}_M\hat{\psi}_{\bar{\beta}} = 1 - \underline{R_L^2} + \frac{2L(1 - R_M^2)}{\underline{N} - M}. \tag{97} $$

Equation (96) is, of course, equivalent to (49). Although ${}_L\hat{\psi}_{\bar{\beta}}$ and ${}_M\hat{\psi}_{\bar{\beta}}$ will in general differ only very slightly, the former is to be preferred in applications, since R_L will be less inflated by overfit than will R_M.

In theoretical comparisons of different factor solutions, ${}_M\psi_{\bar{\beta}}$ will be most useful, since it is a function of the loadings of the discarded factors. The optimal factor solution would be that which minimized the rightmost term of equation (90).

Some Particular Reduced Rank Procedures

Of the five particular rank-reduction procedures considered in the present study, three involve prediction from principal-axes factors, and two involve prediction from a subset of the original predictors. Summerfield and Lubin (1951) have shown that a subset of predictors is equivalent to a subset of orthogonal triangular (or square-root) factor scores. The first factor is simply the first predictor. The second factor is that portion of the second predictor which cannot be predicted from the first. The third factor is that portion of the third predictor which cannot be predicted from the first and second. The remaining factors are similarly obtained. Each factor will thus be independent of the earlier factors and of the predictors corresponding to them, and will therefore have zero loadings on those predictors. Accordingly, the factor-loading matrix will be a lower triangular matrix, i.e., its supradiagonal elements will all be zero.

The predictor-selection and predictor-elimination methods may be thought of as procedures for placing the predictors in the approximate order of their contribution to the multiple correlation with the criterion. Since the triangular factors are determined by the ordering of the predictors, the first L factors will tend to give the highest multiple correlation obtainable with a subset of L predictors.

Prediction from the principal-axes factors giving the highest validity is similar to these methods in that the subset of factors to be retained is entirely determined by the characteristics of the sample from which regression weights are to be computed. Under these circumstances, none of the indices of validity or weight-validity is directly applicable, since all are based on the assumption that, for given L, the subset of predictors to be retained is determined in advance of observing the criterion. A detailed analysis of the consequences of choosing factors on the basis of the observed y will not be attempted. Clearly, however, the fewer the degrees of freedom available, the larger will be the variance of the sample validities, and the smaller the probability that the subset of L factors having the largest true validity will give the largest sample validity. Moreover, the true validity for the subset chosen would tend to fall short of the true validity for the optimal subset, and the sample validity for the chosen subset would tend to overestimate its true validity, in inverse proportion to the degrees of freedom. Still, it seems that subsets of predictors selected in this way would usually have higher true validities than would arbitrarily chosen predictors.

Although the foregoing discussion is not concrete enough to lead to precise conclusions, it does suggest the desirability of having a method of factoring that would provide an a priori expectation as to the contributions to validity of the individual factors. The success of using approximation to the intercorrelation matrix or to its inverse as a criterion for selecting predictors will in part be determined by the extent to which contribution to the approximation is related to contribution to validity.

In describing the two particular factor methods in terms of the general model of the preceding section, we will consider first the triangular factors. For the general factor-loading matrix, b, we substitute a lower triangular factor-loading matrix, t. But where b was partitioned only after the Lth column, we will partition t also after the Lth row, so that

$$t = [t_1 \quad t_2] = \begin{bmatrix} t_{11} & 0 \\ t_{12} & t_{22} \end{bmatrix}. \tag{98}$$

We will partition the inverse of t similarly, and denote it by T'. It may be readily verified that

$$T' = \begin{bmatrix} T_1' \\ T_2' \end{bmatrix} = \begin{bmatrix} t_{11}^{-1} & 0 \\ -t_{22}^{-1} t_{21} t_{11}^{-1} & t_{22}^{-1} \end{bmatrix} = t^{-1}. \tag{99}$$

It will also be convenient to partition the predictor matrix x after the Lth column, and to partition the regression vectors β and $\bar{\beta}$ after the Lth element.

We first note, from (52), that

$$x = [x_1 \quad x_2] = u_1 t_1' + u_2 t_2' = [u_1 t_{11}' \quad u_1 t_{12}'] + [0 \quad u_2 t_{22}']. \tag{100}$$

Thus

$$u_1 t_1' = [x_1 \quad u_1 t_{12}'] \tag{101}$$

and

$$x_2 = u_1 t_{12}' + u_2 t_{12}'. \tag{102}$$

The first term on the right of (102) is that portion of x_2 which can be predicted from x_1, while the second term is that portion of x_2 which is independent of x_1. Thus the "reduced-rank approximation" of x on which predictions are based is from (101) composed simply of the retained predictors augmented by the portion of the discarded predictors that is determined by those retained.

From (63) and (65), the estimated regression weights will be

$$\bar{\beta} = T_1 u_1' y = \begin{bmatrix} (t_{11}')^{-1} u_1' y \\ 0 \end{bmatrix} = \begin{bmatrix} \bar{\beta}_1 \\ \bar{\beta}_2 \end{bmatrix}. \tag{103}$$

Their expected values, under the full-rank hypothesis, will be, from (76)

$$E(\bar{\beta}) = T_1 t_1' \beta = \begin{bmatrix} (t_{11}')^{-1} \\ 0 \end{bmatrix} [t_{11}' \quad t_{21}'] \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \begin{bmatrix} \beta_1 + (t_{11}')^{-1} t_{21}' \beta_2 \\ 0 \end{bmatrix} = \begin{bmatrix} E(\bar{\beta}_1) \\ E(\bar{\beta}_2) \end{bmatrix}. \tag{104}$$

The value for $E(\bar{\beta}_1)$ in (104) may be thought of as an expression for the optimal weights for a subset of predictors in terms of the optimal weights for the entire set. The original weights for the retained predictors are altered as a function of the original weights for the discarded predictors. This illustrates the point made in the section on accuracy of predictions, to the effect that weights for a subset of predictors cannot be properly evaluated in terms of how closely they approximate the weights for the entire set. The covariance matrix of the sample regression weights, obtained from (79), is

$$C_{\bar{\beta}} = \sigma^2 T_1 T_1' = \sigma^2 \begin{bmatrix} (t_{11}'^{-1}) t_{11}^{-1} & 0 \\ 0 & 0 \end{bmatrix}. \tag{105}$$

The expected values of the transformed criterion observations will be, from (83),

$$E(z) = \begin{bmatrix} E(z_1) \\ E(z_2) \\ E(z_3) \end{bmatrix} = \begin{bmatrix} t_1' \beta \\ t_2' \beta \\ 0 \end{bmatrix} = \begin{bmatrix} t_{11}' \beta_1 + t_{22}' \beta_2 \\ t_{22}' \beta_2 \\ 0 \end{bmatrix}. \tag{106}$$

From (90), the inverse index of weight efficiency ${}_M\psi_{\bar{\beta}}$ is given by

$${}_M\psi_{\bar{\beta}} = (N + L)\sigma^2 + \beta' t_2 t_2' \beta = (N + L)\sigma^2 + \beta_2' t_{22} t_{22}' \beta_2. \tag{107}$$

To obtain the principal-axes solution, we first express the predictor matrix x in terms of its basic structure (Horst, 1961, ch. 17):

$$x = P \Delta Q'. \tag{108}$$

Now, in place of the general factor-score matrix u we have the principal-axes factor-score matrix P. The principal-axes factor-loading matrix, corresponding to the general b is given by $Q\Delta$, where Q is a square orthonormal and Δ a diagonal matrix. Equation (50) now takes the form

$$x'x = Q \Delta^2 Q'. \tag{109}$$

The eigenvalues and eigenvectors of $x'x$ will be given by the elements of Δ^2 and the columns of Q respectively. We may partition the factors on the right of (108) to obtain

$$\begin{aligned} x &= [P_1 \quad P_2]\begin{bmatrix} \Delta_1 & 0 \\ 0 & \Delta_2 \end{bmatrix}\begin{bmatrix} Q_1' \\ Q_2' \end{bmatrix} \\ &= [P_1 \quad P_2]\begin{bmatrix} \Delta_1 Q_1' \\ \Delta_2 Q_2' \end{bmatrix} \qquad (110) \\ &= P_1\Delta_1 Q_1' + P_2\Delta_2 Q_2'. \end{aligned}$$

As before, both the factor-score and factor-loading matrices are considered to be partitioned after the Lth column. For the inverse of the factor-loading matrix, B', we will now have

$$[Q_1\Delta_1 \quad Q_2\Delta_2]^{-1} = \begin{bmatrix} \Delta_1^{-1}Q_1' \\ \Delta_2^{-1}Q_2' \end{bmatrix}. \qquad (111)$$

The sample regression vector is, from (63) and (65),

$$\bar{\beta} = Q_1\Delta_1^{-1}P_1'y. \qquad (112)$$

Under the full-rank hypothesis, the lower-rank sample regression weights will have the covariance matrix, from (79),

$$C_{\bar{\beta}} = \sigma^2 Q_1\Delta_1^{-2}Q_1'. \qquad (113)$$

From (83), the canonical form of the criterion will have the expectation

$$E(z) = \begin{bmatrix} E(z_1) \\ E(z_2) \\ E(z_3) \end{bmatrix} = \begin{bmatrix} \Delta_1 Q_1'\beta \\ \Delta_2 Q_2'\beta \\ 0 \end{bmatrix}. \qquad (114)$$

Equation (90) will now take the form

$$_M\psi_{\bar{\beta}} = (N + L)\sigma^2 + \beta' Q_2\Delta_2^2 Q_2'\beta. \qquad (115)$$

The specific reduced-rank prediction models may be obtained from the foregoing development by assuming appropriate permutations either of the predictors, in the case of triangular factors, or of the columns of P and Q, and of the elements of Δ, in the case of principal-axes factors. We note from (73) and (83) that each element of z_1 and z_2 is determined by only one factor: the observed value by the factor scores, the expected value by the factor loadings. In predictor selection, each time a predictor is selected, a factor, and hence an element of z_1, is determined. At each step in the procedure, that predictor is selected which will make the next element of z_1 as large (in absolute value) as possible. In predictor elimination, a factor and hence an element of z_2, is determined each time a predictor is eliminated. At each step, that predictor is eliminated which will make the next element of z_2 as small (in absolute value) as possible.

In the method of predicting from the factors giving the best least-squares approximation to the predictor intercorrelation matrix, the elements of Δ are placed in order from largest to smallest, so that the largest are in $\overline{\Delta_1}$ and the smallest in Δ_2. If the inverse is to be approximated, the elements of Δ are placed in the opposite order, i.e., from smallest to largest. (When we speak of ordering the elements of Δ, we assume, of course, that the columns of P and Q are permuted correspondingly.) In the method of predicting from the principal-axes factors giving the highest validity, the factors are permuted so as to place the elements of z_1 and z_2 in order of absolute value from largest to smallest, with the largest values in z_1, the smallest in z_2.

The Problem of Finding an Optimal Reduced-Rank Solution

There are three major problems involved in obtaining an optimal reduced-rank solution. The first concerns the method of rank reduction: whether subsets of the original predictors, of the principal-axes factors, or of factors obtained by some other method will give the most accurate prediction in future samples. The second problem is, having obtained the factors, to specify the subset of a given size that may be expected to provide the greatest accuracy of prediction. The third problem is, having specified the subset which would be used for any given rank, to determine the particular rank that will tend to lead to the most accurate predictions.

The estimate of the inverse index of weight-efficiency given in (91) and (96) provides a solution (or a potential solution) to the third problem. It does not, however, enhance our ability to deal with the second problem, since, as can be seen from (96), it merely indicates the traditional approach; namely, to attempt to select that subset of predictors of given size having the highest multiple correlation with the criterion. The drawbacks of such an approach when degrees of freedom are limited were discussed in the preceding section. Since a reduced-rank solution is indicated only when degrees of freedom are limited, a selection method that is independent of the criterion might well be preferable. Some evidence favoring this view is provided in the empirical portion of the present study. In the present section we assume that view to be correct and accordingly consider only methods of selection which are independent of the criterion.

If the present analysis is correct, an optimal solution will be one which minimizes ${}_M\psi_{\bar{\beta}}$ as given in (90). In the absence of observations on the criterion, nothing can be said about β or σ^2, so our only course is to seek a value for b_2 which will minimize $\beta' b_2 b_2' \beta$ for general β. The quantity to be minimized may also be expressed as the sum of squares of the expected values of the z_2, as given in (83):

$$\underline{\beta' b_2 b_2' \beta} = [E(z_2)]'[E(z_2)]. \tag{116}$$

Minimizing this quantity will be equivalent to making the elements of $E(z_2)$ as small (in absolute value) as possible. We let the ith element of

$$\bar{z} = \begin{bmatrix} E(z_1) \\ E(z_2) \end{bmatrix} \tag{117}$$

be denoted by $\bar{z}_i$. If we knew these values, the second of the problems stated above would be solved by discarding those factors for which $\bar{z}_i$ was smallest. Denoting the column of factor loadings for the ith factor by $b_{.i}$, we have, from (83),

$$\bar{z}_i = b'_{.i}\beta. \tag{118}$$

Let D be a diagonal matrix whose ith element is given by

$$D_i = \sqrt{b'_{.i}b_{.i}}. \tag{119}$$

Let

$$W = bD^{-1}. \tag{120}$$

Denoting the ith column of W by $W_{.i}$, we have

$$W'_{.i}W_{.i} = \frac{b'_{.i}b_{.i}}{b'_{.i}b_{.i}} = 1. \tag{121}$$

The expected values of z_1 and z_2 can now be expressed in terms of D and W as

$$\bar{z} = b'\beta = DW'\beta, \tag{122}$$

or

$$\bar{z}_i = D_i W'_{.i}\beta. \tag{123}$$

Since we have assumed that nothing is known about β, and since (121) holds for all i, we can have no a priori expectation as to the magnitude of $W'_{.i}\beta$. Thus our only basis for predicting the rank order of the $\bar{z}_i$ in the absence of criterion observations will be the magnitudes of the D_i. A tentative solution for the problem of which factors to retain for prediction, then, will be to discard those factors having the smallest values of D_i. From (119), we see that D_i^2 is the sum of squares of the loadings for the ith factor, or the variance accounted for by that factor. Thus, for a rank-L solution, we wish to retain those L factors giving the best least-squares approximation to the predictor matrix.

It is well known that the principal-axes factors will give a better least-squares approximation to the predictor matrix than will factors obtained by any other method. Thus, as a tentative answer to the first of the above problems we obtain the principal-axes solution.

Now, given the restriction that the factors be selected independently of the criterion, we can state that the best prediction possible with a reduced-rank solution will be obtained from the principal-axes factors giving the best least-squares approximation to the correlation matrix. We note that, for a principal-axes solution, D and W become the Δ and Q of the preceding section. Thus we can also state that the method of approximating the inverse will give the worst possible predictions, since with that method one discards the factors corresponding to the largest elements of Δ.

We have shown that, with appropriate assumptions, the principal-axes factors making the largest contribution to the variance of the predictors (or simply, the largest principal-axes factors) are optimal with respect to our index of expected accuracy of prediction. It may be shown that the factors are also optimal with respect to the variance of the sample regression weights. The sum of these variances will be smaller than for any other method of rank reduction. From (69) (or (79)), this sum will be proportional to the trace of B_1B_1'. We let

$$g' = Bu' = B_1u_1' + B_2u_2', \tag{124}$$

so that

$$g' - \underline{B_2u_2'} = B_1u_1'. \tag{125}$$

It is well known that

$$\mathrm{tr}\,(u_1B_1'B_1u_1') = \mathrm{tr}\,(B_1B_1') \tag{126}$$

will be a minimum when B_2 is composed of the largest principal-axes factors of

$$g'g = BB' = (x'x)^{-1} = Q\Delta^{-2}Q'. \tag{127}$$

Equivalently, the above trace will be a maximum when b_1 is composed of the largest principal-axes factors of $x'x$.

The major conclusion of this section is that, in the absence of criterion observations, the best index to use for selection of predictors or factors will be the amount of variance accounted for in the predictor data matrix. In the case where a subset of the original predictors is to be used, one would eliminate those predictors for which the trace of $t_{22}t_{22}'$ in (107) is a minimum. Where a factor solution is feasible, the largest principal-axes factors would be retained. The important question of how many degrees of freedom must be available before the criterion observations can be used to advantage in the selection process has been left open. Thus a sound basis for deciding whether to use the methods above or to use methods which attempt to maximize the sample multiple correlation with the criterion is still lacking.

CHAPTER 3

AN EMPIRICAL COMPARISON OF FIVE REDUCED RANK PROCEDURES

The Data

A typical application of regression methods is to the problem of predicting academic success as measured by college grades. The data for the present comparisons were taken from a recent study of academic prediction by Shanker (1961). Twenty-nine predictor variables and five separate criterion variables are used. Fifteen of the predictors are those composing the University of Washington Entrance Battery. These have been in use for predicting college grades since 1953, and include age, sex, test scores, and high-school grades. The remaining predictors are taken from the Edwards Personal Preference Schedule (EPPS). The 15 variables of the EPPS are ipsative; i.e., any one can be computed exactly from the remaining 14. Accordingly, only 14 are used here, since the 15th would be completely redundant for purposes of prediction. The EPPS variables are described by Edwards (1954). Descriptions of the Entrance Battery variables are given by Shanker (1961). Since the specific nature of the predictors is not of immediate interest in the present study, we simply list them here.

Edwards Personal Preference Schedule Variables

1. Achievement
2. Deference
3. Order
4. Exhibition
5. Autonomy
6. Affiliation
7. Intraception
8. Succorance
9. Dominance
10. Abasement
11. Nurturance
12. Change
13. Endurance
14. Heterosexuality

High-School Grade-Point Averages

15. English
16. Mathematics
17. Foreign Language
18. Social Science
19. Natural Science
20. Electives

Test Scores

21. Vocabulary
22. Mechanical Knowledge
23. English Usage
24. English Spelling
25. Mathematics
26. Social Science
27. Quantitative Reasoning

Other Variables

28. Age
29. Sex (coded 0 for male, 1 for female)

The criterion variables consist of grade-point averages in various college course areas. The five criteria chosen for the present study were those having 500 or more cases available, as listed below.

1. All-University, 973 cases
2. Mathematics, 541 cases
3. English Composition, 804 cases
4. Chemistry, 526 cases
5. Psychology, 507 cases

The cases used were 973 students who entered the University of Washington as freshmen between 1953 and 1958. Only those students were included for whom measurements on all predictors and at least one criterion variable were available. Scores on the criterion variables and on the Entrance Battery (predictors 15–29) were obtained from the files of the University of Washington Division of Counseling and Testing Services. The EPPS data (predictors 1–14) were obtained partly from Edwards, partly from Wright (1957), and largely from the Division of Counseling and Testing Services files.

Method

The five reduced-rank prediction methods chosen for comparison were the following.

1. The predictor-elimination method (Horst and MacEwan, 1960)
2. Predictor selection by the accretion method (Horst, 1955)
3. The method of largest principal-axes factors (Horst, 1941)
4. The method of smallest principal-axes factors (Guttman, 1958)
5. The method using the principal-axes factors giving the highest multiple correlation.

As noted in the introduction, we can be virtually certain that, for sufficiently small samples, one or more of these methods will give more accurate predictions than will the standard full-rank method. And as shown in the last section of Chapter 2, there is reason to believe that method 3 will be superior to the others for samples below some critical size. Similarly, method 4 would be expected to give the poorest predictions. We would expect also that the statistics ${}_L\hat{\psi}_{\beta}$ as given by (91) and $\hat{W}$ as given by (46) would give some indication of the accuracy of prediction in future samples obtainable from a particular set of weights.

The method used for the empirical comparisons consisted essentially of replications of the following procedure. All cases with measurements available on a particular criterion were taken as the statistical population. From this population a random sample was drawn. Regression weights were computed for each reduced-rank method for each rank from 1 to 29. Thus 29 sets of weights for each method were computed. The sets of weights for rank 29 were, of course, the same (aside from rounding error) for all methods. From the cases remaining in the population after the original sample was removed, a new random sample was drawn. Each set of weights computed in the original sample was then applied to the new sample, and measures of accuracy of prediction were computed. For all computations, predictor and criterion variables were normalized as described in the second section of Chapter 2. In effect, then, means and sums of squares were equated for all variables on all samples. Differences in these values, therefore, do not show up in the total squared errors of prediction.

For each of the five criterion variables, this design, using all five reduced-rank methods, was replicated for six different original-sample sizes: 255, 210, 165, 120, 75, and 30 cases. The new samples consisted of 252 cases for all replications. Weight-validities were used as measures of accuracy of prediction.

An additional set of replications was carried out for criterion 1 (All-University) only, and omitting method 4. Here the estimates of weight-validity and of total squared errors of prediction were also computed from the original samples. A wider range of original-sample sizes was used: the six sizes above and also sizes of 435, 390, 345, and 300 cases. A second new sample was randomly drawn for each replication from the cases remaining in the population after the original sample and the first new sample were removed. Both new samples again consisted of 252 cases for all replications. As measures of accuracy of prediction when the original sample weights were applied to each of the two new samples, total squared errors of prediction as well as weight-validities were computed.

All phases of the above procedures were carried out on the IBM 709 computer, using programs written especially for this study. The method of drawing the samples was as follows. The cases in a particular criterion population of, say, NT students were assigned sequential numbers from 1 to NT. A sequence of random numbers was generated using a procedure described in the WDPC *Users Manual* (Western Data Processing Center, 1961, sec. 9.2.4). The original sample of size N_0 consisted of the cases corresponding to the first N_0 distinct numbers modulo NT from the sequence of random numbers. The remaining $NT - N_0$ cases were renumbered sequentially from 1 to $NT - N_0$. The new sample of size N_1 consisted of the first N_1 distinct numbers modulo $NT - N_0$ from a second sequence of random numbers. In a similar way, all other samples were obtained, using a new sequence of random numbers for each sample.

After obtaining the original sample, the matrix of predictor intercorrelations and the vector of the correlations between the predictors and the criterion were computed. Retaining the notation of the preceding chapter and recalling that the variables in x and y were normalized, the predictor intercorrelation matrix was computed by (25) and the vector of predictor-criterion correlations by

$$r_c = x'y. \tag{128}$$

Next the predictor elimination and predictor selection procedures were carried out and the corresponding regression weights computed, using the procedures described by Horst and MacEwan (1960) and by Horst (1955), respectively. The matrix r was then factored as in (109). The regression weights for the three principal-axes methods were computed as follows. We let z_L denote the Lth element of z_1, $Q_{.L}$ denote the Lth column of Q_1 and Δ_L the Lth element of Δ_1.

First the vector of factor validities z_1 was computed from

$$z_1 = \Delta_1^{-1}Q_1'r_c. \tag{129}$$

Equation (129) is equivalent to (73), since, from (108), (110), and (128),

$$\Delta_1^{-1}Q_1'r_c = \Delta_1^{-1}Q_1'x'y = \Delta_1^{-1}Q_1'(Q_1\Delta_1P_1' + Q_2\Delta_2P_2')y = P_1'y. \tag{130}$$

The regression vector for rank L was computed by

$$\bar{\beta}_L = Q_1\Delta_1^{-1}z_1 = \sum_{i=1}^{L} Q_{.i}\Delta_i^{-1}z_i, \tag{131}$$

which, it may be noted, is equivalent to (112). Thus the regression vector for rank $L + 1$ was obtained from the vector for rank L by

$$\bar{\beta}_{L+1} = \bar{\beta}_L + Q_{.L+1}\Delta_{L+1}^{-1}z_{L+1}. \tag{132}$$

The weights for methods 3, 4, and 5 were all computed in the same way, the only difference being in the order of summation.

The new sample was drawn and the various correlations computed as for the original sample. The weight-validity and total squared errors of prediction obtained with a particular vector of weights were computed respectively by

$$W = \frac{r_c'\bar{\beta}_L}{\sqrt{\bar{\beta}_L'r\bar{\beta}_L}} \tag{133}$$

and

$$\psi = 1 - 2r_c'\bar{\beta}_L + \bar{\beta}_L'r\bar{\beta}_L. \tag{134}$$

Equations (133) and (134) are, of course, equivalent to (42) and (43). Note that r and r_c in (133) and (134) are computed on the new sample while $\bar{\beta}_L$ was computed on the original sample.

Results and Discussion

The weight-validities obtained with methods 1, 2, 3, and 5 on all five criteria are given in Table 1. The six pages of Table 1 correspond to the

TABLE 1

Weight-Validities for Four Methods and Five Criteria

(N_0 = 255)

Criteria:	All-Univ				Math				Engl Comp				Chem				Psych			
Methods:	1	2	3	5	1	2	3	5	1	2	3	5	1	2	3	5	1	2	3	5
Ranks 1	455	455	551	551	305	390	414	414	462	547	640	640	404	404	441	441	406	406	416	416
2	484	484	568	576	375	382	416	407	615	676	616	643	372	459	446	496	477	475	468	468
3	536	536	569	591	415	415	416	400	607	645	580	618	418	426	418	473	489	481	469	487
4	529	529	571	595	421	421	416	401	646	646	608	665	448	451	411	460	489	504	482	488
5	521	521	577	555	435	411	421	418	645	653	659	640	418	422	393	450	492	511	487	488
6	522	522	575	530	422	412	421	414	661	637	666	643	409	412	399	451	509	507	485	488
7	498	498	575	529	404	396	414	426	663	634	669	651	389	389	403	437	510	497	485	486
8	494	494	577	531	392	383	417	426	661	627	644	623	389	393	391	426	501	489	482	484
9	494	494	572	529	393	393	405	419	661	622	644	608	385	413	390	431	492	486	481	480
10	488	491	567	530	374	374	416	417	648	624	648	609	392	406	380	417	477	469	486	463
11	496	488	566	524	371	371	416	407	635	629	630	608	399	412	377	406	481	475	500	473
12	490	496	564	532	368	368	412	399	634	626	635	608	397	410	375	410	475	468	500	468
13	490	492	553	527	375	375	414	395	636	627	635	633	411	419	377	411	478	481	498	470
14	486	487	553	524	372	372	406	395	637	626	638	635	406	418	376	411	485	485	499	473
15	489	498	544	508	371	371	400	389	635	637	640	637	405	412	385	415	486	481	499	471
16	485	498	541	511	369	369	406	380	625	637	642	640	414	409	367	412	477	481	505	468
17	482	500	575	515	372	372	404	385	628	638	643	641	413	408	372	410	474	483	508	470
18	483	499	577	514	376	376	404	385	629	635	644	638	408	415	365	406	470	478	490	468
19	483	502	551	511	379	379	403	385	628	631	641	639	412	415	363	408	471	474	484	466
20	479	499	551	505	384	384	402	386	631	632	642	639	410	410	417	410	470	470	483	473
21	490	496	545	501	383	383	408	384	636	632	638	640	407	408	413	415	476	476	482	474
22	493	497	541	499	381	381	405	381	638	632	638	642	403	404	413	414	478	478	481	472
23	494	494	522	502	382	382	395	384	636	633	639	639	407	408	413	413	478	478	482	470
24	497	495	524	500	383	383	399	385	635	634	635	639	411	408	412	407	477	477	483	471
25	498	498	521	500	384	384	393	384	636	634	636	636	408	412	414	409	477	477	485	474
26	498	498	506	502	384	384	393	382	636	636	638	637	409	409	410	412	476	476	485	473
27	499	499	507	501	385	385	392	383	637	639	636	637	410	411	409	411	477	477	482	473
28	501	501	507	502	384	384	384	384	637	637	636	637	410	410	411	410	477	477	481	474
29	500				383				637				410				477			
R_0	659				539				705				623				626			
R_1	667				515				770				557				580			

Decimal point preceding each entry has been omitted.

TABLE 1 (Cont.)

Weight-Validities for Four Methods and Five Criteria

(N_0 = 210)

Criteria:	All-Univ				Math				Engl Comp				Chem				Psych			
Methods	1	2	3	5	1	2	3	5	1	2	3	5	1	2	3	5	1	2	3	5
Ranks 1	407	407	478	478	361	300	464	464	499	499	543	543	432	432	463	463	353	353	409	409
2	462	460	491	479	439	382	465	477	540	540	543	546	454	471	466	450	445	445	480	480
3	459	490	491	484	440	440	469	451	596	596	547	596	488	481	473	482	443	443	483	477
4	439	474	501	491	433	433	448	440	627	627	596	594	528	508	473	484	461	461	488	463
5	479	473	504	490	424	424	429	431	614	614	601	568	523	530	478	457	475	475	492	469
6	451	466	497	504	411	411	418	440	620	620	602	573	527	525	483	477	466	466	489	448
7	445	463	500	492	374	374	415	435	628	628	602	589	518	533	485	486	461	461	487	425
8	456	489	488	505	375	375	394	446	623	635	607	607	513	525	484	494	452	452	491	436
9	456	491	487	513	362	362	383	416	622	630	619	618	516	522	485	503	460	438	491	432
10	461	470	488	503	365	361	380	395	616	628	620	607	526	527	485	506	460	428	499	437
11	448	484	491	504	377	363	382	395	618	625	607	614	519	524	487	500	446	436	503	434
12	465	483	491	498	384	375	393	383	623	618	623	613	527	526	486	502	439	439	511	428
13	472	472	491	497	382	382	412	382	628	621	633	616	533	526	486	505	437	437	488	439
14	473	473	485	494	374	379	401	372	625	625	630	616	533	531	487	511	439	433	490	437
15	478	478	481	489	366	370	396	363	618	618	629	613	533	530	466	515	445	438	490	442
16	486	486	481	484	357	361	392	353	622	622	628	609	532	530	486	518	448	441	492	442
17	485	485	486	485	358	355	395	348	620	620	627	613	528	535	486	517	445	445	488	437
18	481	481	486	481	358	352	395	340	628	628	627	616	526	530	520	516	446	446	488	434
19	479	484	499	481	355	354	395	343	628	628	621	617	525	532	521	518	449	449	486	440
20	478	482	502	482	354	351	393	347	628	628	622	613	525	528	513	521	447	450	488	444
21	476	484	497	484	358	350	377	346	625	625	622	612	522	531	513	518	444	448	485	446
22	477	482	494	483	355	348	357	342	618	618	602	621	519	530	511	517	445	445	478	445
23	479	479	480	484	349	348	357	340	620	620	617	621	516	527	520	517	445	445	481	444
24	479	479	481	484	346	352	359	341	620	620	614	619	518	524	517	517	442	442	473	447
25	479	479	480	483	344	349	368	343	620	620	613	619	517	521	518	517	441	446	473	446
26	480	480	480	483	339	345	369	342	619	619	616	619	516	519	517	516	443	447	461	446
27	480	479	480	483	339	339	369	342	620	620	621	619	517	518	519	517	446	446	448	446
28	481	480	480	483	342	342	358	342	619	619	620	618	517	517	526	517	446	446	449	447
29	480				340				619				516				446			
R_0	718				502				768				577				672			
R_1	574				562				722				616				568			

TABLE 1 (Cont.)
Weight-Validities for Four Methods and Five Criteria
(N_0 = 165)

Criteria:	All-Univ				Math				Engl Comp				Chem				Psych			
Methods:	1	2	3	5	1	2	3	5	1	2	3	5	1	2	3	5	1	2	3	5
Ranks 1	507	507	542	542	266	322	393	393	513	513	539	539	413	413	440	440	463	374	398	398
2	509	509	577	527	346	341	393	351	546	554	531	537	433	440	441	368	533	362	501	501
3	545	545	578	556	356	364	393	331	595	601	538	564	477	450	432	366	496	456	500	430
4	553	553	588	553	337	348	386	296	587	596	540	590	430	476	426	382	510	501	508	395
5	542	542	583	557	318	329	389	302	588	597	573	613	416	430	425	410	473	508	513	388
6	551	551	583	571	352	337	397	293	578	605	596	599	410	433	426	397	467	513	509	401
7	539	556	597	563	353	329	396	295	560	597	594	594	400	422	425	381	456	498	503	417
8	548	564	610	559	345	353	397	301	589	582	599	598	407	427	420	382	471	483	519	412
9	558	556	622	533	323	347	363	276	583	602	595	595	412	419	425	390	446	485	518	426
10	565	565	621	552	297	339	356	270	576	607	594	601	398	418	424	379	443	470	516	422
11	576	576	619	544	297	326	365	280	554	596	598	610	398	417	423	380	467	484	511	422
12	563	563	601	536	291	314	364	284	559	580	591	610	401	403	424	387	468	468	529	418
13	558	558	599	542	287	314	366	272	564	586	590	611	403	404	420	394	468	465	545	412
14	556	556	602	542	292	301	355	284	560	582	612	613	404	394	413	393	467	465	537	412
15	556	556	616	551	294	298	344	293	559	577	603	613	405	400	420	385	462	465	518	413
16	562	562	600	560	287	294	343	285	561	581	594	604	412	401	423	378	452	460	514	421
17	568	568	598	563	282	300	343	282	569	576	600	598	412	404	425	374	452	441	501	419
18	564	562	587	561	284	306	344	281	576	575	602	591	405	405	406	379	445	438	503	425
19	564	556	575	559	266	297	357	287	574	581	602	592	405	405	406	383	443	429	500	415
20	564	555	573	557	273	287	351	286	574	582	598	591	402	400	444	378	421	431	502	404
21	558	559	576	555	281	285	359	283	577	581	598	585	402	398	449	379	420	425	490	403
22	562	556	576	556	285	289	367	277	577	581	598	589	399	399	428	379	415	416	478	401
23	557	557	577	556	280	286	347	277	578	582	597	586	397	397	439	380	409	417	477	399
24	559	559	579	555	278	278	315	278	577	583	597	585	391	391	450	380	407	409	468	398
25	559	559	580	557	278	278	315	278	579	582	597	585	387	387	454	380	410	408	467	400
26	557	557	579	556	280	280	302	278	582	583	592	583	386	386	402	382	407	405	448	403
27	557	557	577	558	277	277	302	279	583	583	591	583	387	387	394	380	409	407	446	401
28	556	556	558	558	278	278	297	280	585	585	584	583	387	387	392	382	408	408	445	400
29	556				277				582				388				402			
R_0	711				670				716				617				641			
R_1	683				508				700				595				631			

TABLE 1 (Cont.)
Weight-Validities for Four Methods and Five Criteria
(N_0 = 120)

Criteria:	All-Univ				Math				Engl Comp				Chem				Psych			
Methods:	1	2	3	5	1	2	3	5	1	2	3	5	1	2	3	5	1	2	3	5
Ranks 1	367	367	557	557	275	275	383	383	467	550	526	526	369	369	425	425	370	370	440	440
2	448	448	564	567	335	365	387	392	576	593	528	522	409	329	423	444	456	454	492	492
3	514	514	555	532	293	340	385	362	576	601	525	568	469	372	428	446	418	420	491	491
4	519	521	565	484	343	332	385	321	589	588	570	583	431	374	411	407	430	430	478	479
5	496	493	575	466	320	343	355	275	621	600	548	564	403	398	413	394	436	442	479	479
6	459	515	576	485	325	317	354	288	611	611	558	558	394	433	422	406	437	446	483	440
7	442	511	564	487	315	321	350	289	594	612	565	564	391	413	428	406	451	440	493	422
8	467	516	563	494	313	313	351	259	591	604	565	577	397	402	429	404	448	446	476	418
9	457	516	568	510	307	308	351	265	591	590	572	556	408	402	421	381	424	461	479	408
10	442	513	566	520	306	292	364	270	592	589	586	558	420	408	417	379	397	412	488	410
11	459	527	555	511	291	271	371	269	589	586	587	567	435	418	370	380	388	399	468	396
12	458	528	571	503	271	271	382	283	579	583	580	563	434	435	372	390	370	391	450	383
13	469	522	573	508	277	269	378	256	573	584	582	571	432	436	375	401	364	380	449	364
14	477	509	572	503	273	275	378	265	577	573	581	573	428	437	387	406	367	373	450	329
15	471	518	573	494	271	272	348	277	575	575	581	562	430	434	386	399	364	369	451	334
16	476	513	574	487	265	270	356	272	570	577	595	553	430	429	387	391	359	370	433	325
17	483	506	582	493	261	265	350	266	570	575	589	556	437	427	394	399	371	378	434	330
18	494	505	534	494	261	261	351	262	566	578	579	557	425	423	430	400	376	376	449	345
19	495	499	522	490	266	261	342	265	567	574	575	560	424	425	438	404	375	375	448	355
20	488	494	527	480	265	265	349	257	569	578	575	566	420	426	439	399	374	374	443	356
21	476	484	517	482	263	265	341	253	572	580	579	561	419	426	446	401	369	369	437	352
22	478	472	496	473	262	266	347	253	563	575	580	564	411	424	446	409	369	367	414	350
23	472	470	496	473	262	266	359	257	567	575	581	566	412	430	442	414	367	367	406	344
24	470	474	496	473	264	267	367	254	569	570	571	568	414	431	442	415	365	364	404	346
25	470	473	486	472	264	264	328	258	570	570	574	570	415	424	445	414	363	364	380	348
26	471	476	478	472	260	260	326	258	566	571	575	571	416	417	447	421	356	363	380	350
27	469	471	477	470	258	258	297	258	567	567	582	570	418	417	439	414	355	362	376	350
28	470	472	473	470	259	259	257	259	567	567	586	569	418	418	421	414	354	356	377	349
29	470				259				567				418				355			
R_0	764				670				788				629				692			
R_1	688				546				697				558				589			

TABLE 1 (Cont.)
Weight-Validities for Four Methods and Five Criteria
(N_0 = 75)

Criteria:	All-Univ				Math				Engl Comp				Chem				Psych			
Methods:	1	2	3	5	1	2	3	5	1	2	3	5	1	2	3	5	1	2	3	5
Ranks																				
1	399	399	503	503	363	363	403	403	458	506	542	542	476	476	492	492	443	443	381	381
2	360	360	492	515	300	375	381	299	581	545	542	580	387	387	470	391	409	527	367	538
3	338	410	493	501	315	304	361	293	588	529	528	567	304	304	479	372	441	469	532	523
4	314	388	492	511	249	298	381	282	596	542	546	543	248	324	480	226	425	441	511	533
5	320	419	507	510	299	291	336	244	582	591	575	551	229	269	486	202	414	450	465	485
6	294	396	500	446	300	261	314	249	588	593	561	579	238	253	499	216	385	445	452	468
7	325	382	503	421	295	248	302	249	582	582	554	535	259	247	493	215	400	443	452	455
8	317	365	497	373	268	273	281	221	565	582	553	537	251	252	513	180	359	390	445	437
9	349	354	492	372	263	270	279	211	561	575	553	529	259	227	514	211	353	371	446	420
10	347	363	492	377	272	264	284	213	557	575	541	528	229	222	436	206	356	363	429	414
11	370	359	486	367	267	257	257	202	545	568	537	500	252	237	429	164	341	360	418	386
12	354	341	495	374	258	250	280	220	527	569	530	483	247	248	450	179	312	383	449	384
13	352	335	493	371	254	243	280	231	519	568	567	482	241	258	426	156	291	376	454	386
14	324	340	489	366	255	239	278	218	492	561	576	480	208	261	430	169	320	337	435	369
15	326	316	479	358	246	240	213	207	500	564	572	471	219	250	300	168	320	333	419	370
16	325	332	468	353	231	239	224	222	507	563	557	463	212	245	305	178	316	340	434	332
17	333	333	444	355	230	219	226	215	482	543	554	459	219	253	307	190	328	326	433	315
18	335	335	445	353	224	225	233	201	470	541	558	469	225	240	316	212	335	330	418	317
19	344	344	441	348	222	213	239	196	462	520	558	475	215	223	265	194	337	328	411	310
20	336	336	427	338	223	212	248	198	458	496	556	466	211	211	283	201	339	342	423	301
21	325	325	378	338	222	209	242	203	471	483	567	465	215	202	272	190	328	332	418	315
22	325	325	371	343	222	211	245	199	475	476	571	464	215	201	292	203	332	335	403	325
23	320	320	371	336	221	211	221	204	477	474	535	468	207	205	252	199	323	333	396	320
24	322	322	360	336	216	212	201	206	467	480	532	467	204	205	233	200	320	326	370	315
25	320	320	362	330	212	211	197	208	471	469	522	472	205	206	219	201	319	321	371	317
26	322	322	358	329	210	212	197	206	471	471	511	476	206	204	219	205	319	320	353	317
27	324	324	331	326	210	205	200	205	470	470	514	478	203	205	225	204	319	320	351	318
28	325	325	324	325	208	208	204	205	470	471	470	476	203	203	204	206	319	319	354	319
29	325				208				475				204				319			
R_0	655				755				748				760				790			
R_1	572				528				716				620				609			

TABLE 1 (Cont.)
Weight-Validities for Four Methods and Five Criteria
(N_0 = 30)

Criteria:	All-Univ				Math				Engl Comp				Chem				Psych			
Methods: (Ranks)	1	2	3	5	1	2	3	5	1	2	3	5	1	2	3	5	1	2	3	5
1	461	428	481	481	347	347	397	397	440	501	577	577	365	292	431	431	298	298	464	080
2	366	400	478	361	361	316	424	319	543	451	534	534	375	267	432	303	285	285	406	406
3	416	333	531	432	398	272	423	156	519	486	502	502	354	288	440	010	347	347	430	385
4	372	390	513	022	340	267	443	174	556	527	504	466	347	349	437	022	343	301	419	403
5	393	381	472	022	270	275	433	212	518	483	563	509	331	364	431	018	390	291	444	303
6	380	398	467	022	168	247	370	−006	520	514	563	520	329	426	421	014	408	281	451	142
7	354	395	472	022	179	251	364	006	477	514	575	−004	317	365	426	096	357	312	426	152
8	356	365	467	022	149	265	367	006	482	470	574	−004	308	350	430	086	313	289	422	143
9	332	353	467	015	130	283	347	076	484	471	590	−004	317	332	421	095	274	248	430	147
10	309	377	449	018	112	273	346	089	489	450	597	−004	298	316	434	−018	209	232	405	142
11	328	365	454	018	106	271	317	092	462	460	598	−003	293	304	434	−018	186	259	389	137
12	324	338	451	018	101	278	318	084	397	468	598	−002	284	271	430	−019	174	248	386	145
13	342	324	414	018	098	256	318	088	389	453	554	−004	289	245	408	−019	174	259	352	138
14	325	317	410	018	064	234	295	−002	374	418	553	−004	285	208	397	−018	181	259	343	130
15	328	322	391	017	056	229	278	000	347	395	530	−005	282	177	397	−018	182	269	342	124
16	301	338	379	019	016	231	288	−006	304	360	521	−004	220	160	361	−016	162	256	349	126
17	289	354	345	019	031	194	287	−008	228	363	497	−003	129	137	334	−015	136	226	341	−029
18	280	370	310	019	021	167	287	−009	195	334	488	−003	055	131	334	−017	083	181	311	−024
19	223	387	249	021	−011	148	177	−006	147	284	483	−003	011	138	295	−015	068	190	333	−024
20	169	336	247	021	−024	086	202	−007	141	267	477	−002	−015	135	298	−016	011	172	266	−025
21	137	315	349	021	−034	088	177	−009	137	261	476	−002	−018	125	301	−014	010	170	235	−024
22	091	312	379	021	−033	086	219	−009	140	257	474	−002	−017	063	301	−013	005	158	237	−025
23	076	334	390	021	−033	079	210	−009	113	181	454	−002	−021	056	268	−015	−002	156	257	−026
24	062	321	373	021	−028	076	149	−011	078	173	423	−001	−029	048	211	−014	−036	155	224	−029
25	057	170	318	021	−022	069	236	−012	050	157	403	−001	−028	049	215	−014	−025	155	166	−028
26	031	075	311	021	−013	044	204	−011	036	142	333	−002	−027	064	212	−014	−013	144	166	−027
27	019	074	302	021	−011	031	180	−011	002	135	218	−002	−014	110	025	−014	−013	139	134	−023
28	021	076	073	021	−011	018	042	−012	000	055	214	−002	−023	029	104	−013	−014	048	143	−024
29	021				−021				−002				−013				−024			
R_0	999				999				999				999				999			
R_1	663				584				693				565				615			

six original-sample sizes used, ranging from 255 down to 30 cases. This size is denoted by N_0. In each instance, the new sample contained 252 cases. An original sample and a new sample were independently drawn for each size and each criterion, for a total of 30 original samples and 30 new samples. Since for rank 29, all methods are equivalent (aside from rounding error), the corresponding weight-validity is listed only under method 1. The full-rank (rank 29) multiple correlations for each sample are also listed under method 1, the subscripts 0 and 1 denoting the original and new samples, respectively.

Although the weight-validities using method 4 were computed on the basis of the data given above, they are not presented. For all ranks, criteria, and sample sizes, these weight-validities were substantially lower than those for any other method or for the full-rank weights. They were frequently negative, rarely greater than .10, and virtually always less than half as large as the weight-validities obtained by any of the other methods. Our expectation that the method of smallest principal-axes factors would give less accurate predictions than the other methods is thus unequivocally confirmed.

To assist in comparing the other four reduced-rank methods, Table 2 was prepared from Table 1. For each original-sample size and each criterion, two comparisons are made. In each of the first five columns, the number of ranks for which each method was superior to the other three methods is given. In making the counts, ties were divided equally among the methods sharing the high value for a particular rank. In each of the second five columns of Table 2, the number of ranks for which a particular method was superior to the full-rank weights is given. When for a particular rank a method had the same weight-validity as the full-rank weights, the count was increased by one half.

Of the four methods, the method of largest principal-axes factors most often gave the highest weight-validities in 26 of the 30 samples. This trend was most marked when the weights were computed on smaller samples, particularly samples of size 30. The only exceptions occurred for samples of 210 and 255 cases. The superiority of method 3 was most pronounced for Psychology and Mathematics and less clear-cut for English Composition and Chemistry. Method 3 was also more often superior to the full-rank weights than were the other methods. Thus it appears that our expectation as to the superiority of method 3 is also confirmed, but with the qualification that, for larger samples and for certain criterion variables, one or more of the other methods may be preferable.

Another possible basis of comparison would be the number of samples for which a particular method gave the highest weight-validity for any rank. Of the 30 samples, method 3 gave the highest validity in 12.5, method 5 in 8.5, method 1 in 5, and method 2 in 4 samples. The comparisons of Table 2 would appear to be more meaningful than this comparison, however, since

TABLE 2

Comparisons Between Four Reduced-Rank Methods With Respect to Weight-Validities for Five Criteria

Sample Size	Methods	Number of ranks for which weight-validity is higher than for other methods					Number of ranks for which weight-validity is higher than full-rank method				
		All-Univ	Math	Engl	Chem	Psych	All-Univ	Math	Engl	Chem	Psych
	1	0.	2.75	5.17	2.	6.	5.	13.	9.5	9.5	16.5
255	2	0.	.75	3.33	5.	2.	6.5	13.5	8.	15.5	17.
	3	24.5	19.75	13.83	6.	19.5	28.	28.	18.	11.5	25.
	5	3.5	4.75	5.67	15.	.5	26.	24.5	17.	21.	7.
	1	.5	0.	6.17	8.	0.	5.	26.	19.	22.5	12.
210	2	.5	0.	9.67	17.	0.	13.	26.	20.	24.	11.
	3	11.5	19.5	8.	2.5	27.	24.5	28.	14.5	8.	27.
	5	15.5	8.5	4.17	.5	1.	26.	27.	4.	12.	9.
	1	0.	0.	0.	1.	3.	18.5	24.5	7.	24.	28.
165	2	0.	0.	7.	4.	1.	17.5	27.5	15.	24.	26.
	3	27.	27.5	13.5	22.5	24.	27.	28.	23.	28.	27.
	5	1.	.5	7.5	.5	0.	14.	24.	25.	6.	18.
	1	0.	.33	5.5	6.	0.	15.5	26.5	22.	15.	26.5
120	2	0.	.33	9.5	5.	0.	25.5	26.5	26.	16.	28.
	3	26.5	25.5	13.	15.5	25.	28.	27.	21.	18.	28.
	5	1.5	1.83	0.	1.5	3.	26.	18.	10.5	4.	14.5
	1	.33	7.	3.5	0.	.5	15.5	27.5	17.5	25.5	23.
75	2	.33	2.5	9.5	0.	.5	20.5	26.5	23.	24.5	27.5
	3	22.5	18.	13.5	26.5	22.	27.	23.	27.	27.5	28.
	5	4.83	.5	1.5	1.5	5.	27.5	15.5	17.5	11.5	18.5
	1	0.	0.	3.	0.	0.	26.5	23.	28.	19.	26.
30	2	5.	0.	0.	2.	1.	28.	28.	28.	28.	28.
	3	22.5	27.5	24.5	25.5	26.5	28.	28.	28.	28.	28.
	5	.5	.5	.5	.5	.5	13.	27.	12.	10.	19.

TABLE 3

Total Squared Errors of Prediction and Weight-Validities for Four Methods and a Single Criterion

	First New Sample								Second New Sample							
	Weight-Validities				Total Errors				Weight-Validities				Total Errors			
Methods / Ranks	1	2	3	5	1	2	3	5	1	2	3	5	1	2	3	5
1	461	504	582	582	788	746	663	663	385	385	488	488	860	865	763	763
2	509	556	596	570	742	691	647	676	407	422	487	459	852	831	764	795
3	579	572	596	583	666	673	647	661	462	452	487	456	802	804	765	801
4	592	595	599	589	650	646	643	653	467	462	491	469	797	800	760	789
5	597	612	603	593	644	625	638	649	480	478	499	477	789	785	753	782
6	610	612	608	600	628	626	633	640	493	485	503	487	773	781	749	774
7	616	617	619	599	620	619	619	642	493	494	513	480	773	770	738	785
8	616	609	620	605	621	630	617	634	490	476	513	493	781	795	738	770
9	616	608	620	604	621	631	617	635	484	471	515	491	790	801	737	775
10	622	608	620	600	613	631	618	640	486	469	516	485	786	807	736	783
11	616	616	613	596	621	621	625	646	471	471	514	477	807	807	738	790
12	607	607	613	594	634	634	625	648	472	472	514	478	806	806	738	789
13	607	607	613	588	634	634	625	655	470	470	514	477	808	808	738	790
14	605	601	614	593	636	642	624	650	470	465	511	485	807	814	741	783
15	606	601	613	588	635	642	625	657	469	466	507	481	810	813	747	789
16	601	601	617	590	642	642	620	654	464	464	518	475	816	816	734	798
17	606	606	617	595	635	635	619	648	464	464	514	474	816	816	739	801
18	608	608	624	595	633	633	611	648	468	468	525	471	812	812	729	806
19	609	609	622	602	632	632	613	640	467	467	526	468	813	813	728	812
20	612	612	629	607	628	628	604	633	470	470	521	471	809	809	734	808
21	611	611	614	611	628	628	623	628	468	468	491	474	812	812	776	805
22	612	612	619	611	627	627	617	629	472	472	500	472	807	807	767	806
23	611	611	617	610	628	628	620	629	472	472	496	472	806	806	774	806
24	611	611	621	610	628	628	616	630	473	473	492	470	805	805	780	808
25	613	613	615	612	626	626	623	628	472	472	488	471	807	807	787	808
26	613	613	615	612	626	626	624	627	472	472	486	472	807	807	789	807
27	613	613	610	612	626	626	630	627	472	472	478	472	806	806	797	808
28	612	612	608	612	627	627	633	627	472	472	472	472	806	806	805	808
29	613				626				472				806			

$N_0 = 435$ $R_0 = 626$ $R_1 = 684$ $R_2 = 582$

Decimal point preceding each entry has been omitted.

Ranks																
1	349	450	481	481	893	799	770	770	340	412	502	502	901	835	749	749
2	410	481	507	469	843	771	743	782	411	462	516	511	842	791	734	740
3	467	497	506	497	789	756	744	753	451	486	514	524	807	768	736	726
4	480	508	518	511	776	744	732	739	473	504	535	542	786	750	714	707
5	491	498	519	510	767	756	731	742	506	508	535	521	749	745	714	732
6	482	519	530	523	781	732	720	728	495	514	548	538	762	739	700	714
7	486	517	519	522	779	735	731	729	485	502	533	537	777	754	716	714
8	509	503	518	521	752	753	733	732	491	491	535	528	770	766	715	725
9	504	526	518	526	756	728	733	726	495	496	516	537	765	765	737	717
10	500	521	516	522	764	735	737	730	492	492	514	534	771	769	740	720
11	493	512	514	500	771	747	740	756	493	500	508	528	770	760	747	727
12	498	507	517	511	765	755	736	745	497	496	510	541	766	766	746	713
13	495	501	518	503	770	761	736	754	499	497	510	529	764	765	746	728
14	495	503	518	506	768	758	736	752	511	497	509	526	750	766	746	731
15	490	500	522	505	774	762	731	754	511	499	518	521	749	764	738	738
16	485	501	516	499	779	761	738	761	512	512	523	527	748	750	732	731
17	484	504	528	497	780	757	724	765	507	522	536	524	755	737	717	735
18	475	499	522	497	791	762	730	764	504	522	542	519	758	737	713	740
19	481	495	525	491	786	767	728	772	507	524	540	518	755	734	715	741
20	485	493	502	494	780	769	755	769	508	519	535	517	753	740	722	743
21	488	485	495	496	776	779	763	767	508	515	533	515	754	745	723	745
22	492	488	497	496	771	776	761	767	510	514	532	514	752	746	725	746
23	492	489	496	493	771	775	762	771	510	511	530	515	752	749	727	745
24	497	489	497	495	764	775	761	768	515	512	525	516	744	748	733	744
25	495	493	496	494	767	770	766	769	514	514	515	517	746	747	745	742
26	495	497	495	494	767	764	767	769	515	516	516	516	746	743	743	744
27	495	495	492	494	767	767	771	768	515	515	517	516	745	745	743	744
28	495	495	491	495	767	767	772	768	515	515	516	516	745	745	743	744
29	495				767				515				745			
	$N_0 = 390$				$R_0 = 619$				$R_1 = 646$				$R_2 = 638$			

TABLE 3 (Cont.)

Total Squared Errors of Prediction and Weight-Validities for Four Methods and a Single Criterion

	First New Sample								Second New Sample							
	Weight-Validities				Total Errors				Weight-Validities				Total Errors			
Methods	1	2	3	5	1	2	3	5	1	2	3	5	1	2	3	5
Ranks 1	383	408	511	511	859	843	747	747	377	406	531	531	866	845	723	723
2	402	446	524	507	872	820	732	759	426	450	535	523	837	813	718	736
3	477	440	518	530	791	839	741	731	480	466	530	532	780	799	724	727
4	489	489	516	526	780	780	753	740	496	496	524	514	764	764	735	754
5	487	487	523	511	788	788	742	759	491	491	530	505	773	773	727	767
6	521	521	527	509	745	745	735	760	515	515	536	508	746	746	720	763
7	511	511	530	504	759	759	731	770	512	512	536	504	750	750	721	768
8	514	514	534	505	754	754	726	770	516	522	535	505	747	741	722	768
9	514	528	535	518	754	737	726	756	513	520	535	510	752	745	722	761
10	505	529	537	524	769	737	724	746	508	512	532	515	762	757	727	753
11	518	528	536	520	754	738	727	754	509	504	531	504	764	766	730	768
12	531	538	534	514	737	726	735	759	506	503	513	505	771	770	756	767
13	534	538	531	517	733	727	738	756	500	507	506	502	780	766	768	772
14	534	542	530	530	734	721	739	740	505	513	505	510	774	759	768	761
15	529	536	549	525	741	729	716	746	506	512	516	509	773	759	759	764
16	530	532	546	524	739	737	722	747	511	512	507	502	766	761	771	776
17	536	531	543	527	732	739	726	742	519	508	510	508	756	771	766	768
18	538	524	541	530	731	747	728	738	514	502	513	507	763	780	762	770
19	533	527	541	535	738	744	728	733	509	504	512	506	770	777	763	773
20	533	527	553	536	737	745	712	733	506	500	519	507	774	783	752	773
21	535	530	547	533	736	739	719	738	506	507	521	501	775	774	750	783
22	533	533	550	532	737	737	716	739	503	503	518	495	779	779	756	792
23	535	535	552	530	735	735	712	740	503	503	519	497	781	781	755	788
24	534	534	548	531	735	735	717	740	502	502	518	501	783	783	757	784
25	534	534	535	534	735	735	732	736	503	503	509	502	781	781	769	782
26	535	535	535	534	733	733	732	736	504	504	509	502	780	780	769	782
27	534	534	535	533	735	735	732	737	502	502	503	502	782	782	778	782
28	533	533	532	532	736	736	737	737	502	502	497	501	782	782	788	783
29	533				736				502				782			
	$N_0 = 345$				$R_0 = 676$				$R_1 = 649$				$R_2 = 630$			

Ranks																
1	412	490	561	561	831	762	691	691	477	466	542	542	776	783	710	710
2	454	493	556	556	794	757	692	692	478	460	545	545	771	789	704	704
3	471	485	557	561	781	766	691	685	505	474	548	539	745	776	701	710
4	502	538	558	557	750	710	689	690	531	524	546	548	718	726	702	701
5	538	550	564	578	711	698	682	667	554	578	540	556	693	668	708	692
6	519	534	560	550	733	716	686	698	550	577	539	544	698	668	710	704
7	526	525	562	550	726	726	685	697	577	569	540	555	667	677	709	693
8	522	520	560	552	732	732	686	695	567	572	554	559	679	672	694	688
9	527	524	560	542	726	728	686	707	557	568	554	573	690	677	694	673
10	519	520	547	543	734	733	701	706	547	559	565	574	701	688	682	671
11	515	525	551	535	740	727	697	715	542	548	566	568	707	700	681	677
12	511	531	548	531	746	720	700	718	547	556	569	564	701	691	677	682
13	507	519	547	529	747	734	701	721	541	556	566	560	708	692	681	687
14	512	518	551	533	742	734	696	717	551	554	563	552	697	694	683	696
15	515	519	544	532	738	732	704	718	540	548	571	555	709	701	675	692
16	514	508	546	525	738	743	702	725	546	544	574	556	702	706	671	691
17	499	521	547	519	755	730	701	732	540	547	580	553	709	701	665	695
18	512	528	568	518	741	723	677	733	544	549	585	557	705	700	659	690
19	515	535	560	518	737	715	686	732	542	554	578	554	706	694	666	693
20	513	536	557	522	740	715	690	728	537	556	575	552	713	691	670	695
21	515	537	554	528	737	713	694	722	536	546	571	553	714	703	674	695
22	522	535	553	528	729	716	694	723	537	546	573	548	712	702	672	700
23	528	530	553	528	722	721	694	722	541	545	570	546	708	704	675	702
24	525	530	548	530	726	721	701	721	546	543	572	549	702	706	673	698
25	528	527	547	527	722	724	702	724	547	545	572	547	701	704	673	701
26	530	524	525	526	721	727	726	725	548	544	559	547	700	705	688	702
27	528	525	529	527	723	726	721	724	549	550	551	548	700	698	697	700
28	528	528	529	527	723	723	722	724	550	548	546	548	698	700	703	700
29	527				723				550				698			
		$N_0 = 300$				$R_0 = 608$				$R_1 = 664$				$R_2 = 664$		

TABLE 3 (Cont.)

Total Squared Errors of Prediction and Wave-Validities for Four Methods and a Single Criterion

	First New Sample								Second New Sample							
	Weight-Validities				Total Errors				Weight-Validities				Total Errors			
Methods	1	2	3	5	1	2	3	5	1	2	3	5	1	2	3	5
Ranks 1	457	457	561	561	794	794	687	687	383	383	473	473	871	871	780	780
2	503	526	592	592	749	723	650	650	422	422	443	443	838	838	820	820
3	575	561	590	587	670	686	652	655	432	420	443	445	838	845	820	822
4	565	586	590	600	684	657	652	640	442	439	443	450	831	833	819	819
5	574	574	597	580	673	672	644	665	454	446	446	442	819	827	817	830
6	574	579	596	580	673	667	645	665	463	456	439	433	812	817	825	841
7	568	574	612	572	682	673	627	676	462	463	446	437	814	811	817	840
8	569	570	611	571	680	679	627	677	463	463	446	442	811	813	818	831
9	561	581	612	584	692	666	626	661	456	459	446	449	822	814	821	822
10	561	579	612	584	692	670	626	662	456	464	447	450	823	812	821	825
11	565	580	614	592	686	669	624	652	456	464	449	448	824	814	818	826
12	566	580	608	582	685	668	631	665	449	463	441	452	832	814	827	823
13	569	574	607	575	682	675	633	674	450	463	436	458	831	815	832	816
14	574	569	616	584	675	683	622	663	449	459	435	462	833	820	834	812
15	580	570	612	581	667	681	625	667	456	458	438	465	825	820	832	810
16	577	571	608	579	672	681	630	669	452	462	441	467	831	815	830	809
17	580	571	608	578	667	680	631	672	447	457	435	465	836	820	836	812
18	582	573	620	574	665	677	616	676	447	456	442	459	836	822	828	817
19	578	575	602	573	671	674	638	676	447	456	434	453	836	823	841	825
20	580	579	600	571	668	669	640	679	445	454	435	453	837	825	840	825
21	582	575	599	575	666	674	642	674	448	449	435	451	833	831	839	827
22	578	578	600	576	671	670	640	673	448	452	443	451	833	828	833	828
23	579	578	589	575	670	670	654	674	448	453	449	452	832	827	828	826
24	578	580	591	577	671	667	652	672	449	447	449	454	831	834	829	824
25	577	577	581	574	672	672	667	675	448	447	452	454	832	834	828	825
26	575	578	573	573	674	670	676	677	448	448	457	453	833	833	821	825
27	575	577	572	573	674	672	678	677	448	448	457	453	833	832	822	825
28	576	575	569	573	673	674	682	677	448	448	456	454	832	833	823	824
29	576				673				448				832			
	$N_0 = 255$				$R_0 = 646$				$R_1 = 717$				$R_2 = 563$			

Ranks																
1	377	377	498	498	869	869	753	753	445	445	571	571	803	803	676	676
2	371	371	506	519	889	889	745	732	489	489	566	575	763	763	680	670
3	456	456	507	513	805	805	744	744	573	573	563	571	672	672	684	674
4	457	457	508	497	805	805	743	772	536	536	563	568	720	720	683	680
5	469	469	498	496	793	793	754	772	525	525	565	548	740	740	681	706
6	458	458	525	492	806	806	727	778	511	511	572	549	757	757	673	705
7	445	445	519	501	832	832	740	772	488	488	570	552	792	792	677	705
8	459	459	522	505	819	819	736	763	488	488	576	562	790	790	670	692
9	470	470	521	489	809	809	737	783	489	489	575	554	793	793	671	704
10	476	476	522	495	800	800	735	777	497	497	573	551	781	781	673	712
11	479	487	521	488	796	784	737	791	499	502	571	552	781	773	677	715
12	489	483	526	485	785	792	732	800	498	507	569	534	782	769	679	742
13	490	491	520	493	785	783	740	785	510	517	552	528	767	757	701	746
14	488	490	519	493	790	788	743	786	510	512	554	519	770	765	699	759
15	494	491	509	483	781	786	762	801	515	514	560	516	761	763	695	764
16	486	486	514	488	794	796	756	794	519	519	561	517	757	758	695	764
17	489	493	500	488	793	789	774	796	522	520	554	516	754	757	705	766
18	488	491	499	492	794	792	773	790	522	522	539	516	753	757	725	768
19	492	492	499	497	789	792	774	784	522	523	538	526	753	755	727	755
20	491	490	499	499	790	794	773	781	519	523	540	524	758	755	725	758
21	487	488	504	495	796	796	766	786	521	526	532	520	758	750	732	763
22	490	494	507	497	792	788	762	782	527	528	542	518	751	748	720	766
23	494	496	506	496	788	785	764	785	529	526	542	517	750	751	722	767
24	492	496	500	496	790	785	775	786	526	529	542	516	753	750	724	770
25	490	494	488	495	793	787	795	788	524	526	539	515	756	754	730	771
26	492	492	490	495	790	790	794	787	524	524	523	517	756	756	754	769
27	493	493	486	493	789	789	798	789	524	524	519	519	756	756	759	767
28	493	493	486	494	789	789	797	788	524	524	520	519	757	757	758	767
29	493				789				524				757			
		$N_0 = 210$				$R_0 = 670$				$R_1 = 605$				$R_2 = 652$		

TABLE 3 (Cont.)

Total Squared Errors of Prediction and Weight-Validities for Four Methods and a Single Criterion

Ranks	First New Sample								Second New Sample							
	Weight-Validities				Total Errors				Weight-Validities				Total Errors			
Methods	1	2	3	5	1	2	3	5	1	2	3	5	1	2	3	5
1	350	457	587	587	889	791	658	658	295	408	540	540	940	839	709	709
2	471	520	588	585	798	730	657	658	445	499	541	521	833	753	708	731
3	512	507	581	569	753	756	665	677	486	475	543	521	789	797	705	734
4	509	529	591	569	761	732	652	678	495	499	537	512	781	772	712	747
5	504	526	579	570	777	742	665	679	492	495	539	502	789	781	710	760
6	520	520	578	561	756	756	666	695	502	502	534	507	776	776	716	757
7	520	512	582	563	757	768	661	696	505	511	547	505	772	763	702	763
8	537	528	584	556	733	745	660	707	508	513	537	501	768	762	716	769
9	540	530	573	566	731	744	676	698	510	514	538	469	762	760	717	811
10	535	533	574	559	739	740	675	710	503	516	541	464	773	753	714	819
11	546	529	563	558	725	746	696	709	507	514	527	473	765	755	736	806
12	541	538	557	561	733	734	707	706	513	518	522	485	757	748	743	792
13	538	537	559	561	737	738	704	706	518	511	523	492	750	758	741	783
14	540	533	555	557	736	744	710	712	514	508	522	493	755	764	744	782
15	545	539	555	559	729	737	712	710	514	508	529	487	755	763	735	793
16	550	541	554	551	723	734	713	724	521	514	529	485	747	757	735	797
17	555	544	550	556	716	731	719	717	506	504	528	488	766	769	736	791
18	559	549	550	559	712	724	719	714	508	508	529	490	763	764	735	792
19	559	552	552	559	712	722	717	715	508	499	534	490	766	776	729	792
20	560	551	558	559	712	723	708	716	510	503	538	495	765	770	722	785
21	557	551	550	557	719	724	721	720	505	504	536	495	771	770	725	786
22	555	554	543	554	722	721	733	724	503	505	532	495	775	769	731	786
23	556	556	543	555	720	718	733	724	503	506	532	497	775	769	731	784
24	556	556	544	556	721	721	732	722	503	503	527	499	774	775	741	782
25	554	556	539	555	724	721	740	723	501	503	528	500	778	774	740	780
26	554	554	539	555	724	724	740	723	501	501	528	501	777	778	740	779
27	554	555	538	554	724	723	741	724	501	502	529	500	777	776	738	779
28	554	555	551	554	724	723	727	724	501	502	499	500	777	776	778	780
29	555				723				502				777			
	$N_0 = 165$				$R_0 = 666$				$R_1 = 679$				$R_2 = 646$			

Ranks																
1	380	380	546	546	865	865	703	703	451	451	514	514	797	797	738	738
2	418	418	536	531	843	843	718	720	481	481	543	519	774	774	706	731
3	474	474	536	526	806	806	718	729	467	467	545	548	811	811	704	700
4	492	492	534	521	792	792	720	741	490	490	545	540	788	788	704	712
5	529	529	534	504	749	749	721	767	504	504	545	514	776	776	704	747
6	517	517	541	516	773	773	713	759	482	482	535	527	809	809	717	734
7	517	514	542	526	769	777	712	748	471	483	533	521	825	807	720	745
8	512	498	521	514	781	804	743	766	483	468	501	499	820	830	759	778
9	496	473	512	492	807	840	754	807	467	456	505	494	845	848	752	804
10	462	467	516	495	858	854	750	808	451	463	504	470	875	850	753	846
11	457	465	509	480	865	852	759	839	435	455	490	453	895	866	770	866
12	439	473	509	483	894	843	759	831	420	451	490	448	927	870	770	869
13	432	452	509	470	910	895	759	852	431	437	490	441	917	905	770	880
14	423	433	520	457	928	929	752	881	435	422	503	445	919	939	758	889
15	427	433	522	446	926	925	749	898	429	415	502	449	931	943	760	886
16	424	430	506	448	935	932	775	898	435	425	480	447	935	932	790	894
17	430	427	496	441	931	940	789	909	437	430	487	454	936	934	783	887
18	436	426	479	439	920	942	816	915	443	428	479	451	918	940	795	897
19	434	426	479	440	916	944	816	916	448	426	473	449	901	947	802	901
20	433	431	484	440	925	934	817	917	449	429	474	447	901	939	810	907
21	438	425	485	440	920	945	816	916	442	429	473	442	915	942	812	915
22	444	427	475	440	910	944	835	918	446	432	480	442	906	938	809	914
23	447	428	478	442	905	941	838	915	443	430	454	441	909	940	852	915
24	449	426	482	440	902	943	833	917	443	433	450	441	908	935	862	915
25	444	428	466	440	912	935	863	918	444	437	432	441	910	922	895	916
26	445	433	462	441	910	926	870	917	441	441	428	440	915	915	905	918
27	443	432	448	442	913	931	904	914	441	444	434	440	916	912	921	916
28	443	442	441	442	914	916	917	915	441	443	440	441	916	912	917	916
29	443				914				441				916			
	$N_0 = 120$				$R_0 = 737$				$R_1 = 638$				$R_2 = 642$			

TABLE 3 (Cont.)

Total Squared Errors of Prediction and Weight-Validities for Four Methods and a Single Criterion

	First New Sample								Second New Sample							
	Weight-Validities				Total Errors				Weight-Validities				Total Errors			
Ranks \ Methods	1	2	3	5	1	2	3	5	1	2	3	5	1	2	3	5
1	253	335	513	513	971	903	737	737	350	480	592	592	885	770	655	655
2	389	376	497	425	914	917	753	868	474	522	591	532	800	738	656	723
3	475	341	485	434	820	1037	768	884	549	504	566	559	727	808	680	706
4	436	428	482	426	935	936	783	921	540	555	586	565	794	755	660	709
5	419	445	484	425	1004	938	782	1002	522	565	588	534	846	757	657	821
6	392	422	491	435	1085	995	774	1005	489	541	591	545	925	809	654	816
7	395	409	456	434	1099	1036	822	1034	521	526	570	537	865	840	684	849
8	390	407	450	412	1140	1050	830	1093	537	542	568	525	870	809	687	877
9	392	383	451	405	1126	1116	830	1107	551	532	568	521	841	835	686	883
10	356	392	458	399	1193	1140	820	1150	516	531	569	518	924	868	686	906
11	363	380	441	393	1181	1199	847	1156	518	545	557	518	918	873	703	902
12	357	394	432	377	1203	1210	863	1199	518	537	566	486	913	913	695	987
13	374	391	426	372	1195	1217	900	1220	516	539	560	503	931	899	716	949
14	398	366	388	375	1145	1281	1034	1221	528	523	532	504	905	929	792	954
15	411	378	388	370	1117	1234	1034	1246	530	524	533	511	902	920	791	955
16	400	393	389	376	1129	1201	1036	1257	524	534	520	501	903	893	816	981
17	391	391	400	477	1170	1170	1027	1259	527	527	537	492	901	901	799	1004
18	393	393	400	382	1186	1186	1027	1249	528	528	536	492	914	914	801	1005
19	398	392	404	384	1169	1192	1023	1242	527	521	540	495	917	932	797	1003
20	395	392	410	389	1174	1184	1013	1228	520	518	543	498	935	939	791	997
21	395	388	411	384	1178	1204	1011	1242	515	507	543	493	945	968	790	1014
22	396	390	390	380	1182	1214	1084	1252	511	510	540	492	958	972	813	1016
23	387	386	390	381	1201	1224	1104	1250	499	505	544	491	978	987	809	1016
24	390	387	381	385	1204	1224	1111	1244	504	506	516	493	978	983	872	1016
25	387	393	376	388	1235	1208	1123	1249	493	508	512	493	1014	980	883	1021
26	386	397	381	388	1249	1216	1228	1247	490	504	504	494	1031	994	984	1020
27	388	391	384	388	1251	1235	1230	1248	490	494	504	493	1031	1013	989	1023
28	389	389	385	389	1248	1248	1226	1249	491	491	505	493	1029	1029	988	1022
29	389				1251				491				1030			
	$N_0 = 75$				$R_0 = 854$				$R_1 = 608$				$R_2 = 684$			

Ranks																
1	−137	344	429	429	1009	1001	817	817	−139	436	411	411	1009	874	838	838
2	083	241	429	503	*	1269	817	754	114	376	411	544	*	1029	837	709
3	235	204	503	267	1330	1316	754	1483	355	365	545	302	1113	1043	709	1440
4	201	187	492	224	1350	1302	766	1597	322	325	537	260	1146	1071	718	1549
5	208	208	467	173	1314	1295	791	1951	345	354	529	254	1096	1039	725	1821
6	176	200	426	150	1432	1336	836	2008	322	373	479	220	1201	1020	779	1920
7	179	222	412	148	1594	1307	850	2061	323	398	473	237	1312	989	786	1856
8	143	232	424	191	1772	1334	846	2004	298	407	477	294	1503	1004	790	1833
9	076	210	432	187	2291	1427	852	1986	231	373	509	293	1960	1114	758	1823
10	065	164	347	197	2308	1630	984	2748	184	333	413	281	2137	1256	909	2120
11	074	159	336	209	2379	1626	1015	2766	205	350	442	301	2168	1257	865	2137
12	044	179	337	289	3088	1602	1012	3817	197	379	440	213	2767	1207	867	3708
13	032	186	317	296	3642	1616	1036	3869	196	391	427	222	3194	1207	878	3739
14	026	205	321	273	3573	1611	1045	4735	192	392	427	259	3149	1222	887	4338
15	029	184	279	267	3773	1935	1117	4798	179	363	403	255	3332	1451	938	4388
16	050	191	270	258	3622	1999	1126	4795	190	330	400	250	3153	1523	943	4323
17	047	205	264	242	3672	2009	1152	6552	183	306	419	159	3257	1588	933	6374
18	047	174	257	223	3908	2365	1181	8157	175	263	387	105	3487	1865	980	7761
19	044	204	305	225	4863	2608	1169	8160	142	275	447	107	4445	2015	957	7803
20	039	232	222	218	6698	3134	1547	8106	108	258	411	105	6227	2568	1169	7823
21	027	249	196	217	*	5221	1648	8195	074	205	391	097	*	4682	1237	7856
22	012	252	198	210	*	6507	1708	8322	058	189	376	091	*	5703	1326	7902
23	−030	244	138	199	*	7984	2427	8718	−007	175	279	086	*	7060	2056	8218
24	−037	240	164	−085	*	8435	3041	*	−015	173	274	−077	*	7470	2310	*
25	−046	242	143	−085	*	9995	4042	*	−048	159	287	−077	*	8797	3290	*
26	−050	237	134	−085	*	*	4934	*	−065	171	215	−076	*	9289	3873	*
27	−067	225	209	−085	*	*	6529	*	−069	161	164	−076	*	8726	5785	*
28	−072	227	196	−085	*	*	9073	*	−069	138	092	−076	*	*	8519	*
29	−085				*				−077				*			
	$N_0 = 30$				$R_0 = 999$				$R_1 = 619$				$R_2 = 672$			

* Value greater than ten.

the outcome of the latter would presumably be much more subject to random variability of weight-validities from rank to rank.

In Table 3 are presented data from ten additional original samples from the criterion-1 (All-University) population, with sizes ranging from 435 down to 30 cases. Here all sets of weights from each original sample were cross-validated on two new samples, where again each new sample consisted of 252 cases. Total squared errors of prediction are presented as well as weight-validities for each of the 20 new samples. Method 4 was omitted from this phase of the computations. At the bottom of each page of Table 3 are given, in addition to the original sample size N_0, the full-rank multiple correlations for the three samples represented by that page; these are denoted by R_0, R_1 and R_2 for the original sample, first new sample, and second new sample, respectively.

Since the criterion variable (as well as the predictors) was normalized before the computations were carried out, the total squared errors of prediction are comparable from sample to sample as well as from method to method and rank to rank. Expressed in normal deviates, the criterion mean is zero and the sum of squares is one. Thus if a prediction of zero were made for each case, without ever going to the trouble of computing regression weights, the total squared errors of prediction would be one. Since, for example, the total squared errors of prediction using the full-rank weights from an original sample of size 75 are greater than one in both new samples, it appears that this particular regression equation is worse than useless. Yet for this same sample the rank-1 errors for method 3 of .737 and .655 are actually lower than either of the full-rank errors obtained for the sample of 390 cases, which were .767 and .745. In general, it may be seen that the lower-rank errors obtained with method 3 using small original samples compare favorably, or at least not unfavorably, with the full-rank errors obtained using large original samples. A similar trend may be noted, though not so clearly, with regard to weight-validities.

Table 4 was prepared from Table 3 in a manner analogous to the preparation of Table 2 from Table 1. Here, of course, only one criterion variable is involved, and the comparisons are made with respect to total squared errors of prediction as well as to weight-validities. For the larger original-sample sizes, the outcomes of the comparisons are not appreciably affected by the index of accuracy used. For the smaller sizes, however, the total squared errors of prediction tend to favor method 3 over the other methods and the lower ranks over the higher to a greater extent than do the weight-validities. In the present series of samples, just as in the preceding series, method 3 appears to be definitely superior to the other methods. And even for the largest original-sample sizes, method 3 appears preferable to the full-rank system.

It appears that method 3 could be used to considerable advantage in

TABLE 4

Comparison Between Four Reduced-Rank Methods With Respect to Weight-Validities and Total Squared Errors of Prediction for a Single Criterion

Sample Size	Methods	Index	Number of ranks for which index is superior to other methods				Number of ranks for which index is superior to full-rank method			
			W_1	ψ_1	W_2	ψ_2	W_1	ψ_1	W_2	ψ_2
	1		2.33	2.33	.25	0.	6.5	6.5	10.5	11
435	2		3.33	3.83	.25	0.	3.5	5.	8.5	10
	3		21.5	21.	26.75	27.5	18.	20.	27.5	28
	5		.83	.83	.75	.5	0.	0.	17.5	19.5
	1		1.33	1.	0.	0.	8.	7.5	2.	2.
390	2		2.33	2.	.33	.5	18.5	18.5	6.5	8.
	3		19.	20.5	14.33	15.5	24.	24.5	20.5	22.5
	5		5.33	4.5	13.33	12.	19.	17.	24.	25.
	1		.83	.5	2.5	1.5	12.	10.	20.5	24.
345	2		3.83	3.5	3.5	4.	10.	9.5	20.	22.
	3		20.83	21.5	20.5	22.	18.	21.	27.	27.
	5		2.5	2.5	1.5	.5	5.	4.	20.5	21.5
	1		1.	1.	2.	2.	6.5	5.	6.	6.
300	2		0.	0.	3.	3.	12.5	12.	11.5	11.5
	3		24.	24.	18.	18.	27.	27.	20.	20.
	5		3.	3.	5.	5.	20.5	20.5	16.	16.5
	1		1.	1.	2.5	2.	11.5	13.5	16.5	14.
255	2		2.	2.	10.5	8.	13.	14.5	19.5	20.5
	3		23.	23.	4.	6.5	24.	24.	8.	21.
	5		2.	2.	11.	11.5	14.5	14.5	21.5	27.
	1		.33	.33	2.	1.5	3.	5.5	8.	13.5
210	2		.33	1.33	2.	1.5	5.5	9.5	9.5	14.5
	3		21.	22.	21.5	22.5	24.	24.	25.	26.
	5		6.33	4.33	2.5	2.5	18.5	21.5	14.5	14
	1		4.33	4.5	0.	0.	7.	8.5	18.5	20.5
165	2		3.83	5.5	1.	1.	4.	6.5	19.5	24.
	3		11.5	14.5	26.5	26.5	15.	21.	27.	27.
	5		8.33	3.5	.5	.5	22.5	22.	6.5	8.
	1		1.	1.	1.5	0.	15.	19.5	19.5	20.
120	2		0.	0.	2.5	2.	11.	13.	14.5	16.
	3		26.5	26.5	22.5	24.5	27.	27.	24.	26.
	5		.5	.5	1.5	1.5	16.	17.5	23.5	25.5
	1		5.33	0.	0.	0.	18.	27.5	26.	23.5
75	2		3.33	1.	1.5	0.	17.5	27.	28.	26.5
	3		18.5	26.5	26.	27.5	20.5	27.	28.	28.
	5		.83	.5	.5	.5	12.	25.	28.	26.5
	1		0.	0.	0.	0.	27.	28.	27.	28.
30	2		9.	0.	2.	0.	28.	28.	28.	28.
	3		17.5	26.5	25.	26.5	28.	28.	28.	28.
	5		1.5	1.5	1.	1.5	25.5	25.	24.	26.

either of two situations. The first would be where, for a given original-sample size, one wanted the greatest accuracy of prediction obtainable. The other would be where, for a given accuracy of prediction, one wanted to use the smallest possible original sample. In order actually to compute the coefficients for a reduced-rank prediction equation, however, one has, of course, to select the particular rank to be used. To provide some indication as to how satisfactory the statistics $\hat{W}$ and $\hat{\psi}$ would be for this purpose, they are computed for the original samples of Table 3 using (46) and (96), respectively. They were computed only for method 3, since the other methods are dependent on the criterion observations for order of selection, contrary to the assumptions used in deriving the above statistics. These estimated values for weight-validities and total squared errors of prediction are given in Table 5. To facilitate comparisons, the obtained values from Table 3 are reproduced in the adjacent columns. At the bottom of each page are given the original-sample size and the full-rank multiple correlations for the two cross-validation samples. The multiple correlation and the estimated population correlation, from (32), in the original sample are given for each rank. The column headed $\hat{\alpha}$ is an estimate of the standard error of $\hat{\psi}$, and may be derived as follows. We let a be a column vector composed of the elements of z_2 and z_3 in (91). Then we may write

$$\hat{\psi} = \frac{N + L}{N - L} a'a, \tag{135}$$

where the elements a_i of a are independently distributed with mean zero and variance σ^2. The variance of $a'a$ will be

$$\operatorname{Var}(a'a) = E[(a'a)^2] - [E(a'a)]^2. \tag{136}$$

Under the reduced-rank hypothesis, $a'a$ will be simply the error sum of squares in the original sample, so that from (71), the second term on the right of (136) will be

$$[E(a'a)]^2 = [(N - L)\sigma^2]^2 = (N - L)^2\sigma^4. \tag{137}$$

Expanding the first term on the right of (136), we obtain

$$E[(a'a)^2] = (N - L)E(a_i^4) + (N - L)(N - L - 1)E(a_i^2a_j^2), \qquad i \neq j. \tag{138}$$

Since the a_i are independent, we have

$$E(a_i^2a_j^2) = E(a_i^2)E(a_j^2) = \sigma^4, \qquad i \neq j. \tag{139}$$

If the elements of the criterion vector, y, are assumed to be normally distributed, the elements of a, being linear combinations of the criterion observations, will also be normally distributed. Thus we have (Cramér, 1946, p. 212):

$$E(a_i^4) = 3\sigma^4. \tag{140}$$

TABLE 5
Estimated and Obtained Measures of Accuracy of Prediction Using Method of Largest Principal-Axes Factors

Ranks	R_0	R_c	$\hat{\alpha}$	$\hat{W}$	W_1	W_2	$\hat{\psi}$	ψ_1	ψ_2
1	539	538	048	536	582	488	712	663	763
2	549	546	048	543	596	487	705	647	764
3	549	545	048	540	596	487	708	647	765
4	550	544	048	538	599	491	711	643	760
5	558	551	048	543	603	499	705	638	753
6	559	550	048	542	608	503	707	633	749
7	568	558	048	548	619	513	700	619	738
8	568	556	048	545	620	513	703	617	738
9	568	555	048	543	620	515	706	617	737
10	568	554	049	540	620	516	709	618	736
11	571	555	049	540	613	514	709	625	738
12	571	554	049	537	613	514	712	625	738
13	571	552	049	534	613	514	716	625	738
14	571	551	049	532	614	511	718	624	741
15	578	557	049	536	613	507	714	625	747
16	583	561	049	539	617	518	711	620	734
17	584	561	049	538	617	514	713	619	739
18	590	566	049	543	624	525	708	611	729
19	593	568	049	543	622	526	707	613	728
20	594	567	049	542	629	521	709	604	734
21	609	582	048	556	614	491	693	623	776
22	611	583	048	557	619	500	693	617	767
23	615	586	048	558	617	496	691	620	774
24	617	587	048	558	621	492	692	616	780
25	619	588	048	558	615	488	692	623	787
26	619	587	048	556	615	486	695	624	789
27	622	589	048	557	610	478	694	630	797
28	625	590	048	558	608	472	693	633	805
29	626	591	049	557	613	472	694	626	806
		$N_0 = 435$			$R_1 = 684$			$R_2 = 582$	

Decimal point preceding each entry has been omitted.

TABLE 5 (Cont.)

Estimated and Obtained Measures of Accuracy of Prediction Using Method of Largest Principal-Axes Factors

Ranks	R_0	R_c	$\hat{\alpha}$	$\hat{W}$	W_1	W_2	$\hat{\psi}$	ψ_1	ψ_2
1	545	544	051	542	481	502	706	770	749
2	554	550	050	547	507	516	701	743	734
3	554	549	050	544	506	514	704	744	736
4	562	555	050	549	518	535	699	732	714
5	562	554	050	546	519	535	702	731	714
6	568	559	050	550	530	548	698	720	700
7	571	560	050	550	519	533	698	731	716
8	578	565	050	553	518	535	694	733	715
9	585	571	050	558	518	516	689	733	737
10	586	571	050	556	516	514	692	737	740
11	587	571	050	555	514	508	693	740	747
12	587	569	051	552	517	510	697	736	746
13	588	568	051	549	518	510	700	736	746
14	588	567	051	546	518	509	703	736	746
15	591	569	051	547	522	518	703	731	738
16	592	568	051	545	516	523	705	738	732
17	595	570	051	546	528	536	704	724	717
18	605	579	051	554	522	542	695	730	713
19	607	579	051	553	525	540	697	728	715
20	610	582	051	554	502	535	696	755	722
21	611	581	052	552	495	533	699	763	723
22	611	579	052	550	497	532	702	761	725
23	614	582	052	551	496	530	701	762	727
24	615	581	052	549	497	525	703	761	733
25	619	583	052	550	496	515	702	766	745
26	619	582	052	548	495	516	705	767	743
27	619	581	052	545	492	517	709	771	743
28	619	579	053	542	491	516	712	772	743
29	619	578	053	539	495	515	715	767	745
		$N_0 = 390$			$R_1 = 646$			$R_2 = 638$	

TABLE 5 (Cont.)

Estimated and Obtained Measures of Accuracy of Prediction Using Method of Largest Principal-Axes Factors

Ranks	R_0	R_c	$\hat{\alpha}$	$\hat{W}$	W_1	W_2	$\hat{\psi}$	ψ_1	ψ_2
1	598	596	049	595	511	531	646	747	723
2	601	598	049	595	524	535	646	732	718
3	605	600	049	596	518	530	645	741	724
4	622	616	048	610	516	524	628	753	735
5	625	618	048	611	523	530	627	742	727
6	627	618	048	610	527	536	629	735	720
7	628	618	048	608	530	536	630	731	721
8	630	618	049	607	534	535	632	726	722
9	630	617	049	604	535	535	636	726	722
10	633	619	049	605	537	532	635	724	727
11	634	619	049	603	536	531	637	727	730
12	641	624	049	608	534	513	631	735	756
13	643	625	049	607	531	506	633	738	768
14	643	623	049	604	530	505	636	739	768
15	652	631	049	612	549	516	628	716	759
16	654	633	049	612	546	507	627	722	771
17	658	635	049	613	543	510	626	726	766
18	659	635	049	611	541	513	628	728	762
19	659	633	049	609	541	512	632	728	763
20	661	634	049	609	553	519	632	712	752
21	664	636	049	609	547	521	632	719	750
22	666	637	050	610	550	518	632	716	756
23	666	636	050	607	552	519	635	712	755
24	668	637	050	607	548	518	636	717	757
25	673	641	050	610	535	509	632	732	769
26	673	639	050	607	535	509	636	732	769
27	674	639	050	606	535	503	638	732	778
28	675	639	051	604	532	497	640	737	788
29	676	638	051	602	533	502	642	736	782

$N_0 = 345$ $R_1 = 649$ $R_2 = 630$

TABLE 5 (Cont.)

Estimated and Obtained Measures of Accuracy of Prediction Using Method of Largest Principal-Axes Factors

Ranks	R_0	R_c	$\hat{\alpha}$	$\hat{W}$	W_1	W_2	$\hat{\psi}$	ψ_1	ψ_2
1	493	490	062	487	561	542	762	691	710
2	524	519	060	515	556	545	735	692	704
3	524	517	060	511	557	548	740	691	701
4	525	516	061	506	558	546	744	689	702
5	552	541	059	531	564	540	719	682	708
6	553	540	059	528	560	539	722	686	710
7	553	538	060	523	562	540	727	685	709
8	559	542	060	525	560	554	725	686	694
9	559	540	060	521	560	554	730	686	694
10	563	542	060	521	547	565	730	701	682
11	564	540	061	518	551	566	734	697	681
12	566	540	061	515	548	569	737	700	677
13	568	540	061	513	547	566	739	701	681
14	568	538	062	510	551	563	743	696	683
15	577	545	062	516	544	571	738	704	675
16	579	546	062	515	546	574	739	702	671
17	583	548	062	515	547	580	739	701	665
18	590	554	062	520	568	585	735	677	659
19	593	554	062	519	560	578	736	686	666
20	593	553	062	515	557	575	741	690	670
21	595	553	063	513	554	571	743	694	674
22	595	550	063	509	553	573	748	694	672
23	596	549	064	506	553	570	752	694	675
24	598	549	064	504	548	572	755	701	673
25	598	546	065	500	547	572	760	702	673
26	604	552	064	504	525	559	755	726	688
27	606	552	065	503	529	551	758	721	697
28	607	550	065	499	529	546	762	722	703
29	608	549	066	496	527	550	766	723	698
		$N_0 = 300$			$R_1 = 664$			$R_2 = 664$	

TABLE 5 (Cont.)

Estimated and Obtained Measures of Accuracy of Prediction Using Method of Largest Principal-Axes Factors

Ranks	R_0	R_c	$\hat{\alpha}$	$\hat{W}$	W_1	W_2	$\hat{\psi}$	ψ_1	ψ_2
1	559	557	061	555	561	473	692	687	780
2	593	588	058	584	592	443	659	650	820
3	593	587	059	580	590	443	663	652	820
4	593	585	059	576	590	443	669	652	819
5	595	584	060	573	597	446	672	644	817
6	596	583	060	570	596	439	676	645	825
7	599	583	061	568	612	446	678	627	817
8	599	581	061	564	611	446	683	627	818
9	601	581	062	562	612	446	686	626	821
10	601	579	062	557	612	447	691	626	821
11	601	577	063	553	614	449	696	624	818
12	602	576	063	550	608	441	700	631	827
13	604	575	064	547	607	436	704	633	832
14	605	574	064	545	616	435	707	622	834
15	607	573	065	542	612	438	711	625	832
16	608	572	065	539	608	441	714	630	830
17	612	575	065	540	608	435	714	631	836
18	618	579	065	542	620	442	711	616	828
19	623	582	065	544	602	434	710	638	841
20	623	580	066	540	600	435	716	640	840
21	626	581	066	539	599	435	717	642	839
22	639	593	065	551	600	443	704	640	833
23	640	593	065	549	589	449	707	654	828
24	641	591	066	545	591	449	712	652	829
25	644	592	066	545	581	452	713	667	828
26	645	592	067	543	573	457	716	676	821
27	645	590	067	538	572	457	722	678	822
28	646	587	068	534	569	456	727	682	823
29	646	586	069	531	576	448	732	673	832
	$N_0 = 225$			$R_1 = 717$			$R_2 = 563$		

TABLE 5 (Cont.)

Estimated and Obtained Measures of Accuracy of Prediction Using Method of Largest Principal-Axes Factors

	R_0	R_c	$\hat{\alpha}$	$\hat{W}$	W_1	W_2	$\hat{\psi}$	ψ_1	ψ_2
1	528	525	071	522	498	571	728	753	676
2	537	530	071	524	506	566	726	745	680
3	538	528	072	519	507	563	731	744	684
4	538	525	072	512	508	563	738	743	683
5	546	531	072	515	498	565	736	754	681
6	583	566	069	550	525	572	699	727	673
7	601	582	067	564	519	570	683	740	677
8	607	586	067	566	522	576	682	736	670
9	607	583	068	561	521	575	688	737	671
10	608	581	069	556	522	573	694	735	673
11	609	580	070	552	521	571	698	737	677
12	611	579	070	549	526	569	702	732	679
13	616	581	070	549	520	552	703	740	701
14	616	579	071	544	519	554	710	743	699
15	632	594	070	558	509	560	694	762	695
16	633	593	071	555	514	561	698	756	695
17	639	596	071	557	500	554	696	774	705
18	647	603	070	562	499	539	691	773	725
19	647	601	071	558	499	538	697	774	727
20	647	598	072	553	499	540	704	773	725
21	651	600	072	553	504	532	704	766	732
22	653	599	073	550	507	542	708	762	720
23	653	597	074	545	506	542	715	764	722
24	658	599	074	546	500	542	714	775	724
25	660	600	074	545	488	539	716	795	730
26	664	602	074	546	490	523	716	794	754
27	665	600	075	541	486	519	722	798	759
28	665	597	076	536	486	520	729	797	758
29	670	600	076	538	493	524	728	789	757

$N_0 = 210$ $R_1 = 605$ $R_2 = 653$

TABLE 5 (Cont.)

Estimated and Obtained Measures of Accuracy of Prediction Using Method of Largest Principal-Axes Factors

Ranks	R_0	R_c	$\hat{\delta}$	$\hat{W}$	W_1	W_2	$\hat{\psi}$	ψ_1	ψ_2
1	544	540	078	536	587	540	712	658	709
2	544	536	079	528	588	541	721	657	708
3	549	537	080	525	581	543	725	665	705
4	563	548	079	533	591	537	716	652	712
5	582	564	078	546	579	539	703	665	710
6	590	569	078	548	578	534	701	666	716
7	593	569	079	545	582	547	705	661	702
8	608	581	078	555	584	537	695	660	716
9	618	588	078	560	573	538	690	676	717
10	618	585	079	553	574	541	698	675	714
11	639	605	077	573	563	527	677	696	736
12	645	609	077	574	557	522	675	707	743
13	645	606	078	568	559	523	683	704	741
14	646	603	079	562	555	522	691	710	744
15	648	601	080	558	555	529	697	712	735
16	648	598	081	552	554	529	705	713	735
17	648	594	082	545	550	528	713	719	736
18	649	591	084	539	550	529	721	719	735
19	649	588	085	533	552	534	729	717	729
20	650	585	086	527	558	538	737	708	722
21	651	583	087	522	550	536	744	721	725
22	657	586	087	523	543	532	744	733	731
23	657	582	089	516	543	532	753	733	731
24	658	580	090	511	544	527	761	732	741
25	659	578	091	506	539	528	767	740	740
26	659	573	093	499	539	528	777	740	740
27	659	569	094	492	538	529	787	741	738
28	665	573	094	494	551	499	786	727	778
29	666	570	096	487	555	502	794	723	777

$N_0 = 165$ $R_1 = .679$ $R_2 = .646$

TABLE 5 (Cont.)

Estimated and Obtained Measures of Accuracy of Prediction Using Method of Largest Principal-Axes Factors

		R_0	R_c	$\hat{\alpha}$	$\hat{W}$	W_1	W_2	$\hat{\psi}$	ψ_1	ψ_2
Ranks	1	554	549	091	543	546	514	705	703	738
	2	582	572	088	563	536	543	684	718	706
	3	582	568	090	553	536	545	695	718	704
	4	582	563	092	543	534	545	706	720	704
	5	582	557	094	533	534	545	718	721	704
	6	597	568	093	540	541	535	711	713	717
	7	598	563	095	531	542	533	723	712	720
	8	607	569	096	533	521	501	721	743	759
	9	637	598	092	561	512	505	691	754	752
	10	638	594	094	553	516	504	701	750	753
	11	647	600	094	556	509	490	699	759	770
	12	647	595	096	547	509	490	711	759	770
	13	647	590	098	537	509	490	723	759	770
	14	660	601	097	547	520	503	713	752	758
	15	660	596	099	538	522	502	725	749	760
	16	674	609	098	549	506	480	714	775	790
	17	678	608	099	546	496	487	720	789	783
	18	683	610	100	544	479	479	723	816	795
	19	683	605	102	536	479	473	734	816	802
	20	699	622	100	553	484	474	716	817	810
	21	699	617	102	544	485	473	728	816	812
	22	703	617	104	541	475	480	733	835	809
	23	712	624	103	547	478	454	728	838	852
	24	713	622	105	541	482	450	736	833	862
	25	725	633	104	553	466	432	724	863	895
	26	726	630	106	546	462	428	734	870	905
	27	735	638	105	554	448	434	726	904	921
	28	737	635	107	548	441	440	735	917	917
	29	737	630	110	539	443	441	748	914	916
			$N_0 = 120$			$R_1 = 638$			$R_2 = 642$	

TABLE 5 (Cont.)

Estimated and Obtained Measures of Accuracy of Prediction Using Method of Largest Principal-Axes Factors

Ranks	R_0	R_c	$\hat{\alpha}$	$\hat{W}$	W_1	W_2	$\hat{\psi}$	ψ_1	ψ_2
1	520	510	122	501	513	592	749	737	655
2	536	517	123	499	497	591	752	753	656
3	563	537	122	512	485	566	740	768	680
4	604	573	117	544	482	586	707	783	660
5	606	567	121	531	484	588	724	782	657
6	615	569	122	527	491	591	730	774	654
7	634	584	122	537	456	570	721	822	684
8	635	576	126	522	450	568	740	830	687
9	635	567	130	506	451	568	760	830	686
10	637	561	134	494	458	569	777	820	686
11	655	575	134	505	441	557	767	847	703
12	661	574	136	499	432	566	777	863	695
13	689	604	132	529	426	560	745	900	716
14	763	698	108	638	388	532	609	1034	792
15	763	692	112	627	388	533	626	1034	791
16	767	691	115	622	389	520	634	1036	816
17	797	727	105	663	400	537	578	1027	799
18	797	722	109	653	400	536	594	1027	801
19	798	716	113	643	404	540	611	1023	797
20	799	712	117	634	410	543	624	1013	791
21	799	706	121	624	411	543	642	1011	790
22	807	712	121	629	390	540	637	1084	813
23	818	723	120	640	390	544	623	1104	809
24	826	731	120	646	381	516	615	1111	872
25	827	725	124	636	376	512	633	1123	883
26	850	758	113	676	381	504	573	1228	984
27	850	752	118	666	384	504	590	1230	989
28	850	746	123	655	385	505	609	1226	988
29	854	748	125	655	389	491	611	1251	1030
		$N_0 = 75$			$R_1 = 608$			$R_2 = 684$	

TABLE 5 (Cont.)

Estimated and Obtained Measures of Accuracy of Prediction Using Method of Largest Principal-Axes Factors

Ranks	$\hat{R}_0$	R_c	$\hat{W}$	W_1	W_2	$\hat{\psi}$	ψ_1	ψ_2
1	593	573	555	429	411	694	817	838
2	593	552	514	429	411	742	817	837
3	662	613	567	503	545	687	754	709
4	681	617	560	492	537	701	766	718
5	690	610	539	467	529	733	791	725
6	722	634	557	426	479	718	836	779
7	732	628	538	412	473	747	850	786
8	744	626	526	424	477	770	846	790
9	759	629	520	432	509	786	852	758
10	803	683	581	347	413	712	984	909
11	823	701	597	336	442	695	1015	865
12	824	682	564	337	440	750	1012	867
13	830	671	542	317	427	789	1036	878
14	837	661	523	321	427	825	1045	887
15	843	650	501	279	403	866	1117	938
16	845	623	459	270	400	938	1126	943
17	848	592	413	264	419	1018	1152	933
18	853	566	375	257	387	1088	1181	980
19	872	589	398	305	447	1067	1169	957
20	906	682	513	222	411	892	1547	1169
21	911	660	478	196	391	959	1648	1237
22	915	625	426	198	376	1057	1708	1326
23	952	773	627	138	279	712	2427	2056
24	964	805	672	164	274	634	3041	2310
25	972	820	692	143	287	600	4042	3290
26	978	824	694	134	215	598	4934	3873
27	988	875	775	209	164	445	6529	5785
28	995	916	844	196	092	310	9073	8519
29	999	975	951	−085	−077	099	*	*
		$N_0 = 30$		$R_1 = 619$			$R_2 = 672$	

* Value greater than ten.

Putting (139) and (140) in (138), we obtain

$$(141) \qquad E[(a'a)^2] = (N - L)(N - L + 2)\sigma^4.$$

Then, putting (141) and (137) in (136), we may write

$$(142) \qquad \text{Var}\,(a'a) = 2(N - L)\sigma^4.$$

From (135) the variance of $\hat{\psi}$ will be

$$(143) \qquad \alpha^2 = \frac{2(N + L)^2\sigma^4}{N - L}.$$

For an unbiased estimate of α^2 we use (141) and (95) to obtain

$$(144) \qquad \hat{\alpha}^2 = \frac{2(1 - R_L^2)^2(N + L)^2}{(N - L)^2(N - L + 2)}.$$

The values for $\hat{\alpha}$ given in Table 5 were computed from the square root of (144).

In discussing Table 5, we will consider first the 16 new samples corresponding to the original-sample sizes of 120 and up. With a few exceptions, the estimated errors of prediction did not differ from the obtained values by more than one or two times the standard error of the estimate. In the full-rank case, for example, the difference between ψ and $\hat{\psi}$ was less than $\hat{\alpha}$ in eight samples, between $\hat{\alpha}$ and $2\hat{\alpha}$ in six samples, and between $2\hat{\alpha}$ and $3\hat{\alpha}$ in two samples. Ten of the obtained values fell above the estimated and six fell below. Estimates for the lower ranks tended to be more accurate. The weight-validities and their estimates evidently were less variable than the errors of prediction. Though no estimate of the standard error of $\hat{W}$ is available, its accuracy is apparently comparable to that of $\hat{\psi}$. Taking into consideration the variability of the obtained measures of accuracy, both statistics appear to be fairly good estimates of the corresponding expected values, though their standard errors are rather larger than one could wish.

Of perhaps more significance than the absolute magnitudes of the expected values for ψ and W are the relative magnitudes from one rank to another. As a rough indication of how feasible it would be to base the choice of the rank to be used on $\hat{\psi}$, we may compare the values of ψ corresponding to the rank for which $\hat{\psi}$ was smallest with the full-rank ψ. Again considering only the 16 new samples corresponding to the original-sample size of 120 and above, we see that in 15 of the 16 instances, the reduced-rank weights so chosen gave more accurate predictions than did the full-rank weights. Some of these improvements were, of course, very small. For example, in only 8 of the 16 new samples was the reduction in total squared errors of prediction as large as 4 per cent. The largest reductions were 22.9 per cent and 21.4 per cent, both for weights from the original sample of 120 cases. Just how large the reduction would have to be to attain practical significance is, of course, debatable.

In an effort to evaluate the success of $\hat{\psi}$ as an indicator of the rank corresponding to the lowest expected error of prediction, two comparisons were made. First, it would seem reasonable to require that the total squared errors of prediction for the selected rank be closer to the lowest value obtained in a given sample than to the highest. This is the case, however, in only 9 of the 16 samples. A second comparison, intended to control for variability in the obtained errors of prediction, was made on the basis of the rank orders (from lowest to highest) of these values in the individual samples. For each member of each pair of samples corresponding to a particular original sample, the rank corresponding to rank-order 1 was determined. The rank order in the opposite member of the pair of the error of prediction corresponding to the optimal rank in the first member was then obtained. The average of these 16 rank orders was 7.4, suggesting a fair degree of stability in optimal rank. In contrast to this value, the average rank order of the errors of prediction corresponding to the selected ranks was 12.4. Since, if the ranks had been selected at random, the expected rank order would be 15, it appears that $\hat{\psi}$ does not provide a satisfactory basis for selection. However, a better basis does not appear to be available.

We consider now the results of Table 5 for the original-sample sizes of 75 and 30. For the higher ranks, both estimates appear to break down completely. For the lower ranks, taking into account the large standard errors, the two estimates appear to do about as well as in the larger samples. Because of these large standard errors, however, $\hat{\psi}$ and $\hat{W}$ are not very helpful as guides to the absolute magnitude of the corresponding expected values. If taken as an aid to judgment rather than as an index to be applied blindly, $\hat{\psi}$ in particular might be of value in arriving at an optimal rank. In the original sample of size 30, the lowest value of $\hat{\psi}$ for ranks below 24 occurred for rank 3. Very little judgment is required to select a rank-3 solution in preference to a solution of rank 24 or more on a sample of 30 cases. As it turned out, the optimal rank was in fact 3 in both cross-validation samples. In the original sample of size 75, the alternative to a rank-4 solution would be one of rank 14 or more. For samples of 75 cases an optimal rank of 14 is certainly possible, though unlikely. In any event, it appears that, providing unrealistically low values for higher ranks are ignored, $\hat{\psi}$ is potentially of some value in deciding what rank to use for small samples as well as for large ones.

It will be recalled that in deriving $\hat{\psi}$ and $\hat{W}$, the assumption was made that the factor loadings of the predictor matrix would be constant from sample to sample. Thus the very limited success of these statistics may be due to the failure to take sampling variation of the factor loadings into account. This, of course, could not have been done within the context of regression theory, since there only the criterion variable is considered random. The regression model was selected for this study largely on the basis of its simplicity, but also on the grounds that it is the model generally used in connection with prediction problems. However, it seems likely that an analysis of prediction problems in terms of the multivariate normal model of correlation theory or in terms of some other model where the predictor variables are considered random would lead to more successful estimates of accuracy of prediction than those obtained using regression theory.

CHAPTER 4

SUMMARY AND CONCLUSIONS

The primary concern of this study has been with the possibility of using reduced-rank solutions for regression weights to increase the accuracy of prediction obtainable in future samples. Using regression theory, a general factor model for reduced-rank prediction was developed. It was shown that, if errors in the criterion observations are not to be capitalized upon, the optimal basis for determining a lower-rank solution will be the amount of variance accounted for in the predictor data matrix. Thus the best alternative to reduced-rank methods that seek to obtain the maximum multiple correlation with the criterion would be the method of largest principal-axes factors, as suggested by Horst (1941). Estimates of the weight-validities and total squared errors of prediction to be expected when a particular set of weights is applied in future samples were also derived.

An empirical comparison of five particular reduced-rank methods was carried out, using 29 predictors and with partial replication on five criteria, Weights were computed on samples ranging from 30 to 435 cases. As expected, the method of largest principal-axes factors was markedly superior to the other methods tested. This superiority was quite general, appearing in all samples for some criteria, and in some samples for all criteria. The above finding, together with the very poor showing of the method of smallest principal-axes factors, supports the conclusion regarding the importance of predictor variance accounted for by the lower-rank system. The fact that the largest principal-axes factors tended to give more accurate predictions than did the principal-axes factors having the highest multiple correlation with the criterion suggests the desirability of selecting predictors independently of the criterion observations. The exceptions to this trend for the larger original-sample sizes on some criteria indicates the desirability of developing some sort of statistical test for deciding when the predictor-selection methods using the criterion observations may be advantageously applied.

Although their standard errors were rather large, especially in small samples, the estimates of weight-validity and of total squared errors of prediction to be expected in future samples appeared to be reasonably serviceable as regards absolute magnitude. As to relative magnitude from one rank to another, however, it may be questioned whether a rank chosen on the basis of these estimates would be preferable to a rank chosen at random. As estimates of either absolute or relative magnitude, it seems likely that the statistics derived here could be substantially improved upon if variation in the predictor variables or in their factor loadings were taken into account. Without such improved estimates, the large potential advantages of reduced-rank methods demonstrated here, cannot be fully realized. Thus it would seem well worthwhile to undertake an analysis of prediction problems using a statistical model which, unlike regression theory, treats the predictors as random variables.

Until more efficient methods are developed, it is suggested that a regression equation based on the subset of largest principal-axes factors for which $\hat{\psi}$ is smallest will be the best available. For samples with less than, say, 50 degrees of freedom, this procedure must be supplemented by a subjective process to the extent of ignoring low values of $\hat{\psi}$ for ranks of say, ten or more. Although this procedure leaves considerable room for improvement, the relevant evidence seems sufficiently favorable to warrant further empirical research. At any rate, the strong possibility has been raised that the conventional full-rank weights can almost always be improved upon even in samples of several hundred cases. Such weights, moreover, may give predictions only slightly more accurate than those made from weights obtainable with samples of as few as 30 cases.

REFERENCES

Anderson, T. W. *An introduction to multivariate statistical analysis.* New York: Wiley, 1958.

Cramér, H. *Mathematical methods of statistics.* Princeton: Princeton Univ. Press, 1946.

Davis, F. B. The reliability of component scores. *Psychometrika,* 1945, **10**, 57–60.

Edwards, A. L. *The Edwards Personal Preference Schedule.* New York: Psych. Corp., 1954.

Guttman, L. To what extent can communalities reduce rank? *Psychometrika,* 1958, **23**, 297–308.

Horst, P. (ed.) *The prediction of personal adjustment.* New York: Social Science Research Council Bulletin 48, 1941.

Horst, P. A technique for the development of a differential prediction battery. *Psychol. Monogr.,* 1954, **68**, Whole No. 9.

Horst, P. A technique for the development of a multiple absolute prediction battery. *Psychol. Monogr.,* 1955, **69**, Whole No. 390.

Horst, P. Matrix algebra for social scientists. Technical Report, Office of Naval Research Contract Nonr-477(08) and Public Health Research Grant M-743 (C6), Univer. of Washington, 1961.

Horst, P. and MacEwan, C. Predictor-elimination techniques for determining multiple prediction batteries. *Psychol. Reports,* Monogr. Suppl. 1-V7: 1960.

Kempthorne, O. *Design and analysis of experiments.* New York: Wiley, 1952.

Leiman, J. M. The calculation of regression weights from common factor loadings. Unpublished doctoral dissertation, Univ. of Washington, 1951.

Mood, A. M. *Introduction to the theory of statistics.* New York: McGraw-Hill, 1950.

Shanker, P. The contribution of EPPS scores to differential and multiple absolute academic prediction. Unpublished doctoral dissertation and Technical Report, Office of Naval Research Contract Nonr-477(08) and Public Health Research Grant M-743(C5), Univer. of Washington, 1961.

Sheffé, H. *The analysis of variance.* New York: Wiley, 1959.

Summerfield, A. and Lubin, A. A square-root method of selecting a minimum set of variables in multiple regression. *Psychometrika*, 1951, **16**, 271–284.

Western Data Processing Center. *The WDPC users manual.* Graduate School of Business Administration, Univ. California, Los Angeles, 1961.

Wright, C. E. Relations between normative and ipsative measures of personality. Unpublished doctoral dissertation and Technical Report, Office of Naval Research Contract Nonr-477(08) and Public Health Research Grant M-743(C2), Univer. of Washington, 1957.