Psychological Bulletin
1978, Vol. 85, No. 2, 254-266

# Weighting Predictors in Linear Models: Alternatives to Least Squares and Limitations of Equal Weights

Robert M. Pruzek and Bruce C. Frederick
Department of Educational Psychology and Statistics
State University of New York at Albany

The problem of how to weight several predictors of a criterion in a linear model is discussed. It is argued that the recently recommended practice of replacing least squares weights ($\beta_j$s) with equal or simplified weights of standardized predictors will frequently result in serious, systematic errors of prediction. In particular, Wainer's equal weights theorem is shown to have little relevance to a wide range of prediction problems because, contrary to his claims, his assumptions are highly restrictive.

A comprehensive algebraic system is described that makes explicit the interdependence between least squares regression coefficients and the set of correlations between all pairs of predictors as well as between each predictor and the criterion. Alternative sample-based estimators of population $\beta_j$ values are discussed. It is argued that certain alternative $\beta_j$ estimators can be expected to work particularly well when sample sizes are small and when predictor variables have substantial intercorrelations with one another. A general strategy is proposed for conducting applied prediction studies to take into account one's prior beliefs or knowledge about covariances or correlations in the joint predictor–criterion system.

A number of articles have recently appeared that have compared different weighting schemes for linear prediction models (Dawes & Corrigan, 1974; Einhorn & Hogarth, 1975; Schmidt, 1971). Interestingly, and perhaps surprisingly, all of these authors have found that simplified weighting of standardized predictors may be as good as or even superior to least squares weighting under rather general conditions.

Wainer (1976) has recently offered a further analysis of this problem that is of particular interest, since his arguments were mathematical and his conclusions very general. He concluded that "when you are interested solely in prediction, it is a very rare situation that calls for regression weights which are unequal" (p. 216). In particular, Wainer provided a theorem that shows that under certain conditions, which he clearly believes to be unrestrictive, the coefficients in multiple regression models can be replaced with equal weights with small loss in predictive accuracy.

In this article we show that the equal weights theorem has little relevance to a wide range of prediction problems because, despite Wainer's arguments, his assumptions are highly restrictive. Moreover, of those authors who have previously written about this topic, none have clearly indicated when unit weights are apt to work poorly or what alternatives other than equal (or simplified) weights may be appropriate when sample least squares prediction is apt to be undesirable. Because multiple predictor studies are so often used in the behavioral sciences, it seems important to be as clear as possible about such matters.

The aim of this article is not only to qualify previous authors' conclusions regarding simplified weighting schemes. Newly developed correspondences between common factor analysis and least squares regression are discussed that make explicit the interdependence between $\beta_j$

Requests for reprints should be sent to Robert M. Pruzek, Education 236, Department of Educational Psychology and Statistics, State University of New York, Albany, N.Y. 12222.

weights and joint predictor–criterion correlations or covariances. Using these theoretical findings, it is shown that alternatives to ordinary least squares regression estimators can be constructed that will frequently yield greater precision of future predictions than will ordinary least squares $\beta_j$ estimators. Finally, a general strategy is suggested for designing and conducting applied prediction studies, a strategy that seems to be in harmony with a general Bayesian viewpoint about statistical inference. This strategy involves prespecification of beliefs about structural (correlational) relationships in the joint predictor–criterion system at the time when a study is being planned and/or before an analysis has been started.

Wainer's equal weights theorem begins with the assumptions that all predictors are mutually uncorrelated (not just linearly independent) and that population (standardized) regression coefficients are uniformly distributed in the interval (.25, .75).[1] With these restrictions Wainer argued that if equal weights are used for $k$ standardized predictors instead of least squares weights, the expected loss in criterion variance is $k/96$. The author also concluded that if the predictors are positively correlated with one another, the expected loss in predictive accuracy will be even smaller, as long as the (standardized) regression weights are positive and uniformly distributed over an interval with a range of .5. In the next two sections we question the propriety of each of Wainer's assumptions as well as the appropriateness of his general conclusions.

## Limitations Imposed by Assuming Predictors To Be Uncorrelated

Wainer assumed in his proof of the equal weights theorem that all predictors are mutually orthogonal to one another, that is, uncorrelated. If $\mathbf{R}_x$ represents the matrix of predictor intercorrelations, this means that $\mathbf{R}_x = \mathbf{I}$, an identity matrix. Wainer believed that this assumption is unimportant, however, since he stated that the expected loss in accuracy of prediction is less than $k/96$ if predictors are all positively correlated. Nevertheless we argue below that the theorem is unlikely to be meaningful except when $\mathbf{R}_x$ approximates an identity matrix, so that this case is important to distinguish from the more general case of correlated predictors. It is particularly useful to recognize that if $\mathbf{R}_x = \mathbf{I}$ for the predictor intercorrelations, then severe limits are placed on the number of predictor variables, $k$, which may be observed.

It is easily demonstrated that when least squares regression is carried out using mutually uncorrelated predictors, then standardized least squares regression coefficients are identical to the individual predictor–criterion correlation coefficients, that is, $\rho_{yj} = \beta_j$. Further, squared multiple correlations (*SMC*s) can always be computed as the sum of squares of these squared coefficients:

$$SMC = \sum \rho_{yj}^2 = \sum \beta_j^2.$$

Because $SMC = \sum_{j=1}^{k} \rho_{yj}^2 \leq 1.00$, the average of $k$ squared predictor–criterion correlation coefficients cannot exceed the fraction $1/k$.

In the case of Wainer's assumption for $\rho_{yj}$s (or $\beta_j$s) that all "sample" values are within the interval (.25, .75), it is clear that the maximum number of predictors that may be accommodated in a single sample is $k = 1/(.25)^2 = 16$, in the wholly improbable event that all $\rho_{yj}$ values are exactly .25. If $k$ predictors are selected using the interval (.25, .75) for $\beta_j$ so that the average of $k$ values is .50, then the maximum $k$ occurs for all $\rho_{yj} = .50$. In this case, $k$ drops to $1/(.50)^2 = 4$. But if the sample $\rho_{yj}$s vary around .5, as it is most reasonable to expect in the context of this theorem, then $k \leq 3$. Wainer's statement that it "is no real restriction" (p. 215) to assume that $\beta_j$ values are in the restricted interval (.25, .75) is simply not credible.

If the interval for $\beta_j$ values is permitted to go to (.0, .5) (see Footnote 1) then it will generally be possible to have more orthogonal predictors. But even in this case $k$ cannot routinely be expected to be large. If all "sample" $\beta_j$ values varied uniformly around the midpoint of this interval, .25, then through use of the relationship *Variance* = $(Range)^2/12$ for a uniform distribution (cf. Hays, 1973, p.

[1] Although Wainer (1976) used the specific interval (.25, .75), it was clear from his discussion and proof that the range of coefficients is actually restricted only to the range $(a, a + .5)$ for $0 \leq a \leq .5$ (see also Wainer & Thissen, 1976).

247), it can be shown that $k$ will equal 12. In practice, under Wainer's orthogonality assumption, it would be unreasonable to expect $k$ to be much larger than 12, even if his least restrictive interval is assumed.

Given the restricted interval for $\beta_j$ values, these limits on the number of orthogonal predictors, $k$, are absolute; they hold for any empirically obtained numbers, not just "samples," in the sense in which Wainer used that term. As valid predictors are added beyond the above limits, they must be correlated with previous ones. Yet as soon as predictor variables become mutually correlated with one another, the $\beta_j$ values cannot be assumed to be distributed independently of one another; their interdependencies are directly related to the configurations of correlations in the joint predictor–criterion matrix for any given population. We examine this point in more detail below.

## Limitations Imposed by Assuming a Uniform Distribution of Regression Coefficients

The key assumption for Wainer's equal weights theorem is that $\beta_j$ values are taken to be uniformly distributed across a restricted interval, specified in Wainer (1976) as (.25, .75). The theorem requires, in effect, that actual $\beta_j$s be thought of as randomly sampled from a uniformly distributed population of potential $\beta_j$s, defined such that all predictor intercorrelations are fixed at zero for each sample. This assumption seems to have little precedent in statistical literature, and indeed, most of what may initially appear as an elegant simplicity about this theorem is owed to this uniformity assumption. However, uniformity is generally highly restrictive for either orthogonal or nonorthogonal predictor systems, thus making the entire theorem empty of broad implications for practice.

As one begins to examine the uniformity assumption, a particularly curious finding is that the expected range of $k$ "sample" $\beta_j$ values is a simple function of $k$, namely, E (*Range*) = (*Interval width*) $[(k-1)/(k+1)]$ (cf. Keeping, 1962, p. 203). If the interval width is .5, as is assumed for the equal weights theorem, then for $k = 2$, E(*Range*) = .167; for $k = 3$, it is .25, and for $k = 5$, the expected range of $\beta_j$ values is .333. Only when $k$ grows large does the expected sample range begin to approach its theoretical limit of .5. Thus, if predictors were mutually uncorrelated so that $k$ must be small and the $\beta_j$s were uniformly distributed over an interval of width .5, then Wainer is in effect assuming that the "sample" $\beta_j$ values are very homogeneous. It should be no surprise that the expected loss in accuracy of prediction is small under the assumption that relatively homogeneous weights are to be replaced with equal weights!

As the number of predictors is increased, we have already noted that predictors will generally have to be mutually intercorrelated. Moreover, the equal weights theorem itself says that the loss in predictive accuracy from replacing true $\beta_j$ weights with equal weights grows linearly with increases in $k$.

A further point is that when predictor variables are permitted to be intercorrelated, then the $\beta_j$ values are not restricted in principle to the interval $(-1.00, 1.00)$, as is true for uncorrelated predictors. With correlated predictors, standardized regression coefficients have a theoretical maximum range of infinity, so that a .5 range restriction may be described as infinitesimal in relation to the possible range of $\beta_j$ values.

The major reason why it is difficult to generalize from the equal weights theorem is that the model on which it is based is defined on a space of $\beta_j$ values that cannot in general exist. If real data, with correlated predictors, are used to generate $\beta_j$ values (for either statistical samples or populations) then the specific interdependence among these $\beta_j$s must generally be taken into account before one can know how well equal weights will work in relation to idealized weights. In particular, when there are correlations of different sizes among the various pairs of predictors and the criterion, then $\beta_j$ values will tend to have substantially greater variance than will the predictor–criterion correlations. If predictor intercorrelations are heterogeneous, then bimodality, or tendencies toward U-shaped distributions of $\beta_j$s, will be common. Thus, to assume uniformity for $k$ $\beta_j$ values is to make a strong assumption, one that, for many practical prediction problems with correlated predictors, is unlikely to be met.

One should not in general be optimistic about the predictive usefulness of equal weights, at least not on the basis of Wainer's theorem or arguments. We provide more concrete evidence on this point in the next section.

The reader may now recall that Wainer was not the first, nor were his arguments the only ones that led to the conclusion that equal weights in place of $\beta_j$ values would often be more than satisfactory. Indeed there are situations in which it may be advantageous to use equal weights or, more generally, simplified weights ($-1$, 0, and 1) in sample prediction problems. In the following section we examine a comprehensive algebraic system that makes explicit the interdependence between joint predictor–criterion correlations and least squares regression coefficients. Through applications of certain principles that are based on some recently developed correspondences between common factor analysis and regression analysis, it will be possible not only to see when simplified weights may be effective but also to develop alternatives to sample least squares regression that may often be expected to be superior to both simplified and least squares weights.

## Some Relationships Between Least Squares Regression and Common Factor Analysis

A recent synthesis of factor analytic and regression algebra serves to illuminate several general characteristics of various regression systems for those typical situations in which a complex of several random independent variables is used to predict one or more criterion variables. Lawley and Maxwell (1973) developed most of the basic relationships between factor analysis and regression, although certain unpublished papers provide useful supplements. (See Browne, Note 1; Pruzek, Walker, Frederick, Huba, & Sherry, Note 2; Scott, Note 3.) Our aim in this section is to highlight certain principles that result from the recent synthesis and to use these principles to identify situations in which either least squares procedures or unit weights can and cannot be expected to work well.

We focus first on common factor analyses of correlation matrices for joint predictor–criterion systems, where it will be assumed that the related regression equations are in standardized metric. Using correlational metric does not reduce the generality of the results, because the factoring methods to which we refer are properly described as "scale free"—meaning that analyses of correlation and arbitrarily scaled covariance matrices are systematically related to one another through a diagonal scaling matrix. We examine what happens in terms of interpredictability of variables in a single battery or complex, starting from prototypical common factor coefficients that can always be used to produce off-diagonals of any empirically obtainable correlation matrix.

Consider the joint matrix $\mathbf{R}$ of predictor–criterion correlations of order $p \times p$, with $p = k + 1$. Suppose that $\mathbf{R}$ satisfies the common factor model for $m$ factors, so that $\mathbf{R} - \mathbf{U}^2 = \mathbf{FF}'$, where $\mathbf{U}^2$ is the (diagonal) matrix of uniqueness variances, and $\mathbf{F} = \{a_{jr}\}$ is a $p \times m$ matrix of common factor coefficients. Further, let $\mathbf{a}_r$ be any column vector in $\mathbf{F}$. This model will necessarily hold for some value of $m \leq p - 1$, although $\mathbf{U}^2$ may not be uniquely defined for some $\mathbf{R}$ matrices, particularly if $m$ is large (see Lawley & Maxwell, 1971).

The least squares regression coefficients for predicting each variable from the others can be shown to be wholly determined by the common factor coefficients. If there is only $m = 1$ common factor, so that $\mathbf{F} = \mathbf{a}_1$, and if $\mathbf{B}$ represents the $p \times p$ matrix with zero diagonals whose nonzero column entries are the $k = (p - 1)$ least squares regression coefficients for predicting each variable from all others, then $\mathbf{B} = \mathbf{U}^{-2}\mathbf{a}_1\mathbf{a}'_1\mathbf{U}^{-2}\mathbf{D}_1 + \mathbf{D}_2$ where $\mathbf{D}_1$ and $\mathbf{D}_2$ are appropriately chosen diagonal matrices. $\mathbf{D}_1$ scales columns of $\mathbf{U}^{-2}\mathbf{a}_1\mathbf{a}'_1\mathbf{U}^{-2}$, and $\mathbf{D}_2$ merely consists of the negatives of the diagonals of $\mathbf{U}^{-2}\mathbf{a}_1\mathbf{a}'_1\mathbf{U}^{-2}\mathbf{D}_1$, so that diagonals of $\mathbf{B}$ are zero (see Lawley & Maxwell, 1973).

Whenever the number of factors, $m$, equals unity, the off-diagonals of the Rank 1 product $\mathbf{U}^{-2}\mathbf{a}_1\mathbf{a}'_1\mathbf{U}^{-2}$ are proportional in any column to the standardized least squares regression coefficients for predicting one variable from the remaining $p - 1$ others. More simply, for any criterion variable $j_c$, the $\beta_j$s for the predictors will necessarily be proportional to the quotients $a_{j1}/(1 - a_{j1}^2)$ if $m = 1$, $j \neq j_c$.

The above facts lead to some especially

Table 1
*Regression Parameters for Two Hypothetical One-Factor Populations*

| Variable | Factor loadings | Least squares regression coefficients | | | | | | Variance accounted for[a] | |
|---|---|---|---|---|---|---|---|---|---|
| | | Predictor | | | | | | Method | |
| | Factor 1 | 1 | 2 | 3 | 4 | 5 | 6 | Least squares | Equal weights |
| | | | | Population A | | | | | |
| 1 | .98 | | .39 | .24 | .17 | .13 | .11 | .92 | .91 |
| 2 | .95 | .59 | | .14 | .10 | .08 | .06 | .88 | .86 |
| 3 | .92 | .52 | .21 | | .09 | .07 | .06 | .83 | .81 |
| 4 | .89 | .49 | .19 | .12 | | .06 | .05 | .78 | .76 |
| 5 | .86 | .46 | .18 | .11 | .08 | | .05 | .72 | .71 |
| 6 | .83 | .44 | .17 | .11 | .08 | .06 | | .67 | .67 |
| | | | | Population B | | | | | |
| 1 | .98 | | .55 | .39 | .04 | .03 | .03 | .90 | .77 |
| 2 | .95 | .72 | | .20 | .02 | .02 | .01 | .88 | .73 |
| 3 | .93 | .65 | .26 | | .02 | .02 | .01 | .84 | .70 |
| 4 | .50 | .30 | .12 | .08 | | .01 | .01 | .24 | .22 |
| 5 | .45 | .27 | .11 | .08 | .01 | | .01 | .20 | .18 |
| 6 | .40 | .24 | .09 | .07 | .01 | .01 | | .16 | .14 |

*Note.* Each row under the heading "Least squares regression coefficients" contains the standardized weights $\beta_j$ for predicting one of the six variables from the $p - 1 = 5$ others, that is, each is a row vector of the matrix $\mathbf{B}'$ (not $\mathbf{B}$).

[a] Computed as the square of the correlation between the variable chosen as the criterion and a weighted composite of the $p - 1 = 5$ others.

straightforward generalizations about the predictability of any single variable from several others. If $p$ is moderate to large, then the equally weighted composite of all variables, the first centroid of $\mathbf{R}$, will typically be similar to any one of the equally weighted composites of $(p - 1)$ standardized variables. Moreover, as $p$ is increased, the first centroid of $\mathbf{R}$ is known to converge toward the (first) common factor of $\mathbf{R} - \mathbf{U}^2$. The larger the value of $p$, the closer the correspondences. When the least squares regression coefficients for predicting any variable from all others can be reproduced from a single common factor and the $(p - 1)$ order equal weights composites are all very similar to this common factor, then equal weights will be virtually as good as $\beta_j$ weights. This result corresponds almost directly to that of Wilks (1938).

Even if all correlations are positive, $p$ is moderate to large, and a single factor can exactly reproduce $\mathbf{R} - \mathbf{U}^2$, it does not, however, necessarily follow that all the $(p - 1)$ order equal weights composites are very similar to this common factor. Differences between equal weights composites and least squares composites can be substantial if the coefficients in $\mathbf{a}_1$ are sufficiently heterogeneous. Thus, equally weighted composites of $(p - 1)$ variables will not necessarily be as good as ideally weighted composites based on $\beta_j$ weights. The first two hypothetical examples that follow should clarify this distinction.

In Table 1, structural configurations among two sets of six variables are specified using two vectors of common factor coefficients. In the case of Population A, the various $a_j$ values are quite homogeneous. The reader may wish to verify in this example that for any criterion variable, the five $\beta_j$ weights are proportional to the quotients $a_{j1}/(1 - a_{j1}^2)$. For this population system, even with a relatively small number of variables, replacing least squares weights with equal weights can be seen from the final two columns to result in small loss in predictive accuracy, regardless of which vari-

able is chosen as the criterion. Note that this is true despite the fact that for each criterion in Population A, the $\beta_j$ values vary substantially among themselves. Most practical problems in which the single-factor model is relevant will involve $a_{j1}$s which, like these, do not vary greatly among themselves.

Even for the relatively simple case where $m = 1$, it is possible in principle for equally weighted composites to result in substantial loss of predictive accuracy[2] in relation to least squares composites. If, for example, the $a_{j1}$ values for a given set of variables range from, say, .2 or .3 up to .85 or .95, then it is easy to show that least squares regression coefficients for at least some of the variables can produce composites that are quite different from equally weighted composites. The $a_{j1}$s for hypothetical Population B (Table 1) have accordingly been chosen to be quite heterogeneous. For this hypothetical population system one could incur substantial loss by adopting equal weighting, depending on which variable one was interested in predicting. Indeed, the loss would almost surely be intolerable were variables such as 1, 2, or 3 to be selected as criteria. On the other hand, if Variables 4, 5, or 6 were defined as criteria, then one would be willing to sacrifice the small luxury of least squares weights for the convenience of equal weights.

If, however, equal weights for the predictors are apt to lead to unacceptable loss in predictive accuracy in relation to $\beta_j$s, this does not necessarily mean that one must resort to sample least squares in empirical work. In practice one may be willing to assume that a single common factor model was appropriate for the joint predictor–criterion $\mathbf{R}$, yet be unable to specify $a_{j1}$ values for this system. In such a case the $a_{j1}$ values can be estimated by any one of several possible methods, including centroid factoring. If $m = 1$, and especially if $p$ is substantial, the $a_{j1}$ estimators, $a_{j1}^*$, can be expected to be very stable across independent random samples. Thus, if $\beta_j^*$ values are determined (up to a constant of proportionality) as $a_{j1}^*/(1 - a_{j1}^{*2})$ then these reduced rank $\beta_j^*$ estimators of the population least squares weights may be expected to be systematically superior to sample least squares weights for a given sample size $n$. (Because the correlation between a criterion and a composite is the same as the correlation between a criterion and a constant multiple of the composite, it is always possible to compute the constant of proportionality once relative weights for predictors are available.)

If it cannot be assumed that $m = 1$ for the joint predictor–criterion matrix $\mathbf{R}$, then a more general system is required for relating factor coefficients to regression coefficients. Suppose again that $\mathbf{R}$ satisfies the common factor model for $m$ factors so that $\mathbf{R} - \mathbf{U}^2 = \mathbf{FF}'$ for $\mathbf{F}$ of order $p \times m$. Then consider $\mathbf{U}^{-1}\mathbf{R}\mathbf{U}^{-1} - \mathbf{I} = \mathbf{Q}_m(\mathbf{D}_\lambda^2 - \mathbf{I})_m\mathbf{Q}'_m$, where $\mathbf{Q}_m$ consists of $m$ unit length column eigenvectors and $\mathbf{D}_\lambda^2$ is the diagonal matrix of $m$ corresponding eigenvalues of $\mathbf{U}^{-1}\mathbf{R}\mathbf{U}^{-1}$. Communalities may be defined as nonzero entries in $\mathbf{I} - \mathbf{U}^2$.

In this notation, and where $\mathbf{S}^2 = (\text{diag } \mathbf{R}^{-1})^{-1}$, it can be shown that (cf. Pruzek et al., Note 2)

$$\mathbf{B} = (\mathbf{I} - \mathbf{S}^2\mathbf{U}^{-2}) + \mathbf{U}^{-1}\mathbf{Q}_m(\mathbf{I} - \mathbf{D}_\lambda^{-2})_m\mathbf{Q}'_m\mathbf{U}^{-1}\mathbf{S}^2 \quad (1)$$

or, by moving the diagonal matrix $(\mathbf{I} - \mathbf{S}^2\mathbf{U}^{-2})$ to the left side,

$$\mathbf{B} + \mathbf{S}^2\mathbf{U}^{-2} - \mathbf{I} = \mathbf{U}^{-1}\mathbf{Q}_m(\mathbf{I} - \mathbf{D}_\lambda^{-2})_m\mathbf{Q}'_m\mathbf{U}^{-1}\mathbf{S}^2.$$

In words, this expression means that all the off-diagonal entries in $\mathbf{B}$, the $p$ sets of $(p - 1)$ $\beta_j$ coefficients, can be reproduced from a matrix of Rank $m$. The matrix members of this $\mathbf{B}$ are just the constituent parts of the maximum likelihood common factor matrix $\mathbf{F} = \mathbf{U}\mathbf{Q}_m(\mathbf{D}_\lambda^2 - \mathbf{I})_m^{\frac{1}{2}}$, which is defined such that $\mathbf{FF}' = \mathbf{R} - \mathbf{U}^2$ (cf. Lawley & Maxwell,

---

[2] Most advocates of simplified weighting schemes have taken the *SMC* as their index of predictive accuracy. It may at times be more meaningful, however, to assess predictive accuracy in terms of the standard error of estimate or its chief determinant, the coefficient of alienation, $K = (1 - SMC)^{\frac{1}{2}}$. $K$ is directly interpretable as the standard deviation of standardized criterion scores around the regression surface (or hyperplane) of best fit. If $c$ denotes the reduction in *SMC* due to the use of other than least squares weights for prediction, and $c^*$ denotes the corresponding increase in $K$, the reader may easily verify that relative to $c^*/K$ (the proportionate increase in the coefficient of alienation), $c$ or $c/SMC$ understates the loss in predictive accuracy whenever *SMC* is larger than approximately .5. Moreover, as *SMC* gets closer and closer to unity, the extent to which $c$ understates predictive imprecision grows rapidly more extreme.

1971). This says that any matrix **B** of all the least squares regression coefficients is wholly reproducible from a matrix that is a row-and-column rescaling of the matrix $\mathbf{Q}_m$ of eigenvectors of $\mathbf{U}^{-1}\mathbf{R}\mathbf{U}^{-1}$. Relatively few parameters will be associated with **B** when $m$ is much smaller than $p$. This suggests that effective reduced-rank sample estimators may be defined to replace conventional sample least squares estimators whenever there exists a $\mathbf{U}^2$ for the population **R** such that $\mathbf{R} - \mathbf{U}^2$ has a relatively small rank. Note that as distinguished from common factor analysis, conventional forms of principal component analysis are not particularly useful in developing relations between regression systems and structural covariance methods. Also, it should be made explicit that the methods under consideration here are quite different from so-called principal component regression methods, as discussed, for example, by Herzberg (1969).

Two hypothetical population systems for which $m = 2$ are given in Table 2. The $F$ for Population C is written so that all off-diagonal values in the corresponding correlation matrix are equal to a constant, .40, except for the correlations $\rho_{12} = \rho_{13} = \rho_{23} = .65$. The first eigenvalue (3.27) of the matrix **R** in this example is nearly four times as large as the second (.836). Yet despite this fact and the reasonably homogeneous communalities, it is clear from the final two columns of squared validities that a rather substantial loss in predictive accuracy results from the use of equal weights instead of the distinctive $\beta_j$ values when Variables 1, 2, and 3 are designated as criteria. However, if the equally weighted composite of, say, only Variables 2 and 3 were used to predict Variable 1, then the squared validity for this simplified composite drops only to .518 from .536, the corresponding *SMC*. Thus a simplified composite based on two predictors is more valid here than one based on four predictors. For Criteria 4, 5, and 6, omnibus equal weights are practically as good as least squares weights, despite the differences

Table 2
*Regression Parameters for Two Hypothetical Two-Factor Populations*

| Variable | Factor loadings | | Least squares regression coefficients | | | | | | Variance accounted for[a] | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Factor | | Predictor | | | | | | Method | |
| | 1 | 2 | 1 | 2 | 3 | 4 | 5 | 6 | Least squares | Equal weights |
| | | | | | Population C | | | | | |
| 1 | .73 | .35 | | .35 | .35 | .07 | .07 | .07 | .54 | .47 |
| 2 | .73 | .35 | .35 | | .35 | .07 | .07 | .07 | .54 | .47 |
| 3 | .73 | .35 | .35 | .35 | | .07 | .07 | .07 | .54 | .47 |
| 4 | .27 | .57 | .10 | .10 | .10 | | .20 | .20 | .28 | .27 |
| 5 | .27 | .57 | .10 | 10 | .10 | .20 | | .20 | .28 | .27 |
| 6 | .27 | .57 | .10 | .10 | .10 | .20 | .20 | | .28 | .27 |
| | | | | | Population D | | | | | |
| 1 | .90 | .00 | | .45 | .32 | .18 | −.16 | −.08 | .63 | .30 |
| 2 | .80 | .00 | .55 | | .18 | .10 | −.09 | −.04 | .55 | .26 |
| 3 | .70 | .50 | .31 | .15 | | .39 | .16 | .08 | .64 | .59 |
| 4 | .60 | .60 | .20 | .09 | .43 | | .23 | .11 | .60 | .56 |
| 5 | .00 | .70 | .29 | −.14 | −.30 | .38 | | .16 | .32 | .09 |
| 6 | .00 | .50 | −.17 | −.08 | .17 | .22 | .19 | | .18 | .06 |

*Note.* Each row under the heading "Least squares regression coefficients" contains the standardized weights $\beta_j$ for predicting one of the six variables from the $p - 1 = 5$ others, that is, each is a row vector of the matrix $\mathbf{B}'$ (not **B**).

[a] Computed as the square of the correlation between the variable chosen as the criterion and a weighted composite of the $p - 1 = 5$ others.

among the $\beta_j$s. Knowledge of the particular clustering among these variables is in any case at least moderately effective in reducing predictive inaccuracy. And the generalization follows that knowledge of distinctive clustering within any joint predictor–criterion matrix may be of special value in knowing whether to use ideal $\beta_j$ weights, some particular simplified weights, or perhaps other weighting schemes for the predictors of any single criterion. Brogden (1946) is a useful reference for showing that a relatively small increase in a validity coefficient may at times be of substantial practical value.

For Population D, the factor coefficients depict relatively more highly differentiated structural relations. In this case it is clear that the loss in accuracy associated with using equal instead of least squares weights for predicting all but Variables 3 and 4 is particularly great. In general it is true that suppressor effects are most sharply defined when there are at least two common factors for the joint predictor–criterion matrix and where columns depict uncorrelated clusters of variables and factorially complex variables, those that load on two or more factors. Thus we see that for multivariate populations characterized by relatively distinct clustering of variables, the loss in accuracy associated with equal weights instead of $\beta_j$s may indeed be substantial. Moreover, to the extent that in a given domain of prediction problems, factor matrices like that of Population D are typical, then substantial suppressor effects for at least some variables may be very common. Indeed, Schmidt's (1971) survey of empirical correlation matrices in four behavioral journals over a 10-year period led to a finding of suppressor effects in over 60% of the matrices examined. For applied work it is of course an empirical question as to whether small numbers of factors may be sufficient to account for all covariation or whether factor structures are simple or complex.

One might respond to all of this by saying that simplified weighting of predictors may still be appropriate; just use weights of 1, 0, or $-1$ for the various predictors of any criterion and forget about least squares weights. But if this is to be done, the question is, How does one decide which variables to weight 1, 0, and $-1$? Wainer suggested calculating sample least squares weights and making the weighting decisions on this basis. But such a practice is highly problematic for several reasons. One is that when predictors are even moderately correlated, sample least squares estimators of $\beta_j$s may be highly unreliable unless $n$ is very large. Moreover, as these examples indicate, it could sometimes be difficult to know if equal or simplified weighting is apt to be effective, even if population $\beta_j$s were themselves known. Some of the unfortunate characteristics of the sample least squares weights that led Wainer to suggest that they be replaced with simple weights are the very characteristics that make these $\beta_j$ estimators inadequate for choosing simple weights.

The important point to emphasize is that just because sample least squares weights are unstable does not mean that all sample-based estimators of population $\beta_j$s are unsatisfactory. It has already been shown how $\beta_j$ values may be estimated from common factor coefficients if $m = 1$, and from the preceding algebra it should be clear that analogous $\beta_j$ estimators can be simply derived even when $m > 1$. In the next section we discuss some sample-based alternatives to least squares regression estimators and consider their potential role in applied prediction studies.

## Alternative Regression Estimators for Empirical Studies

It has been noted that when there exists a population correlation matrix $\mathbf{R}$ and a diagonal matrix $\mathbf{U}^2$ such that $\mathbf{R} - \mathbf{U}^2$ has relatively small rank, then the number of parametric determinants of all $p$ sets of $(p - 1)$ least squares regression coefficients in the matrix $\mathbf{B}$ can be reduced in relation to the number of parameters that are implicitly involved in ordinary least squares regression. The previous discussion has been unrealistic in having been equally concerned with the prediction of each variable from all others in a population and in having considered only population systems, where sampling variation does not arise. Yet, the principles involved in reducing the number of parametric determinants of all elements in $\mathbf{B}$ also work when considering the prediction of any single variable from all others. The

major statistical idea in this context is that one can use factor analytic procedures to produce sample-based estimators of the possibly small number of parametric determinants of $\beta_j$ values, which in turn can be used to construct estimators of these values. These regression estimators are expected to be advantageous, especially when one's sample size is relatively small. If one has reason to believe that a single factor is likely to be sufficient to account for all off-diagonal covariation in the joint predictor–criterion matrix, then as previously noted, equal weights for all predictors are likely to lead to a trivial loss of predictive accuracy, particularly when there are several predictors. If $m$ is expected to exceed unity, however, there may be no way to avoid serious losses of predictive accuracy if equal weights are used in place of optimal weights. Simplified weights may be satisfactory and indeed should be considered if one has confidence in one's knowledge (or beliefs) of joint predictor–criterion factor structures. But especially when one is only willing to venture the guess that there are relatively few factors in the population system, then reduced-rank regression estimators may serve to considerable advantage in relation to other possible weighting methods. The question of how to choose factoring methods to use in small-sample regression applications should nevertheless be approached with care. This is the next topic for discussion.

As with least squares regression methods, there is a conventional wisdom about factor analytic methods that admonishes applied workers always to use large samples if factor results are to be generalizable. As we note below, however, such a convention is by no means relevant to a wide range of important applications. Some methods, especially maximum likelihood factor analysis, yield notoriously unstable communality estimates for small samples and also result in "improper" solutions (those with at least one nonpositive uniqueness variance) a large portion of the time. But fortunately there exist other common factor methods that are computationally efficient, do not permit improper solutions, and can be adopted for present purposes. The simplest such method is one that was first devised by Whittle (1953), a method that bears a superficial resemblance to conventional principal component analysis. Pruzek (1977) has discussed a simple generalization of Whittle's method that has special relevance for reduced-rank regression applications.

The major idea of the Whittle method or its generalization is to assume (nominally) that population uniqueness variances (or communalities) are known up to a constant of proportionality, say $\theta^2$. Thus one may prespecify that $\mathbf{U}^2 = \theta^2\mathbf{I}$ so that $\mathbf{R} - \theta^2\mathbf{I}$ becomes the assumed form of the population common portions covariance matrix. More generally, if one wished to prespecify differentiated uniqueness variances, say, as nonzero elements of the diagonal matrix $\mathbf{D}_u^2$, then $\mathbf{R} - \theta^2\mathbf{D}_u^2$ would become the starting point for a common factor analysis.

Given a prespecified $p \times p$ diagonal matrix $\mathbf{D}_u^2$ whose nonzero elements represent one's beliefs about population uniqueness variances, at least up to a constant $\theta^2$, then an eigenanalysis of a rescaled sample correlation matrix may be carried out to produce a simple system of common factor analysis. Thus, when $\mathbf{R}_s$ denotes the sample correlation matrix, $\mathbf{R}_{su} = \mathbf{D}_u^{-1}\mathbf{R}_s\mathbf{D}_u^{-1}$ may be resolved as $\mathbf{R}_{su} = \mathbf{Q}\mathbf{D}_\lambda^2\mathbf{Q}'$, where $\mathbf{Q}$ represents the set of all unit length column eigenvectors, and $\mathbf{D}_\lambda^2$ represents the corresponding diagonal matrix of eigenvalues. Either a priori or empirical approaches may be used to estimate $m$, the number of common factors, and $\theta^2$ may be estimated as the average value of "rejected" eigenvalues, those corresponding to the smallest $p - m$ roots of $\mathbf{R}_{su} = \mathbf{D}_u^{-1}\mathbf{R}_s\mathbf{D}_u^{-1}$ (cf. Jöreskog, 1962).

The special virtue of this approach to small-sample factoring is that by prespecifying the diagonal $\mathbf{D}_u^2$ and estimating only the scalar constant $\theta^2$, $p - 1$ fewer parameters are being estimated in relation to virtually all other forms of common factor analysis. A relatively small price may be paid for inaccuracy of prespecified $\mathbf{D}_u^2$ values unless population uniquenesses (or more correctly, their reciprocals) are in fact quite heterogeneous. But of course population uniqueness values will not be known in practice, so that the a priori choice of $\mathbf{D}_u^2$ values will generally merit careful thought.

Given Equation 1 for $\mathbf{B}$ and the eigenanalysis of $\mathbf{R}_{su} = \mathbf{D}_u^{-1}\mathbf{R}_s\mathbf{D}_u^{-1}$, it is possible to show

(Pruzek et al., Note 2) that

$$\mathbf{B}^* = [\mathbf{D}_u^{-1}\mathbf{Q}(\mathbf{I} - \theta^2\mathbf{D}_\lambda^{-2})\mathbf{Q}'\mathbf{D}_u^{-1}\mathbf{S}^2]_0$$

represents a sample-based identity for all $p$ ($p - 1$) least squares regression coefficients, where $[\ ]_0$ denotes setting diagonal elements of the matrix in brackets equal to zero. Thus, the reduced rank estimate of $\mathbf{B}^*$, namely, $\mathbf{B}_m{}^*$, can be written as

$$\mathbf{B}_m{}^* = [\mathbf{D}_u^{-1}\mathbf{Q}_m(\mathbf{I} - \theta^2\mathbf{D}_\lambda^{-2})_m\mathbf{Q}'_m\mathbf{D}_u^{-1}\mathbf{S}^2]_0, \quad (2)$$

where it is to be understood that each of the matrix elements and $\theta^2$ derive from sample information. Equation 2 is easily computed from $\mathbf{R}_{su} = \mathbf{D}_u^{-1}\mathbf{R}_s\mathbf{D}_u^{-1} = \mathbf{Q}\mathbf{D}_\lambda{}^2\mathbf{Q}'$, given that $\theta^2$ equals the average of the $p - m$ smallest eigenvalues in $\mathbf{D}_\lambda{}^2$ and that $\mathbf{S}^2 = (\text{diag } \mathbf{R}_s^{-1})^{-1}$. Computationally, such reduced rank estimators can be obtained at least as efficiently as least squares estimators.

If $p$ is substantial and $m$ is relatively small, then $\mathbf{B}_m{}^*$ provides a parametrically parsimonious representation of $\mathbf{B}^*$. The idea is that if $p - m$ eigenvalues are in fact quite homogeneous (as they will be if $\mathbf{D}_u{}^2$ is properly chosen), then the elements of the corresponding eigenvectors are known to be statistically unreliable (cf. Morrison, 1976). Reduced-rank estimators simply ignore the sample information that is associated with the rejected eigenvalues and their eigenvectors. Thus, reduced-rank estimators can be expected to be more stable than full-rank (conventional) least squares $\beta_j$ estimators whenever the common factor assumptions for $\mathbf{R} - \theta^2\mathbf{D}_u{}^2$ can be defended. Indeed, the previous algebra for population systems indicates that the eigenvectors associated with the smallest $p - m$ eigenvectors contribute nothing to the matrix $\mathbf{B}$ whenever the population $\mathbf{R} - \mathbf{U}^2$ system is of rank $m$. Pruzek (Note 4) has shown that such reduced-rank estimators constitute a class of random variable ridge regression[3] estimators in which $p - m$ ridge coefficients are employed. There is strong evidence from simulation studies (Pruzek, Note 5) that the reduced-rank regression methods may indeed be expected to lead to greater predictive precision than conventional least squares methods in many real-life situations. A general program has been developed to conduct factor and regression analyses of this general form (Pruzek et al., 1976).

A number of properties of the above reduced-rank estimators should now be made explicit. First, because this presentation has been made in the framework of so-called correlation metric, $\mathbf{B}$ or $\mathbf{B}^*$ have been implicitly defined in terms of standardized regression systems. It will often be important to be able to compute raw-score regression coefficients from such sample-based betas. This can be simply done using available standard deviations as described, for example, by Rozeboom (1966). Although the population analysis is scale free, sample reduced-rank analyses will not necessarily be, unless maximum likelihood factoring is used. Nevertheless, for small samples, Whittle factoring is preferable in this context due to its efficiency relative to scale-free factoring methods.

Second, it is clear that to the extent that communalities (or communality estimates) reflect reliabilities of variables, the present system for reduced-rank regression can be said to be accommodated to unreliabilities of the random variable predictors as well as the criterion variable. Indeed, Lawley and Maxwell (1973) argued that the communality of any variable can be viewed as the upper limit of the squared multiple correlation of any (criterion) variable when using all other variables as predictors. This limit could be reached in principle if the underlying factors could be directly observed, so that there is no measurement error. This argument may have special didactic value for those individuals who sometimes speak as if all imprecision of prediction were a consequence of only measurement error.

Beyond the above properties, reduced-rank regression estimators hold potential for making regression estimators more robust in some situations, particularly when a population covariance system can be characterized as having a relatively low ratio of $m/p$. Distributional irregularities for individual (predictor)

[3] Ridge regression was initially developed by Hoerl and Kennard (1970) for the case of fixed predictors. Pruzek (Note 4) has studied a version of ridge regression for random predictor problems and has generalized the latter ridge regression system for the case in which a common factor model is not expected to hold in the population.

variables seem likely to have little influence on reduced-rank $\beta_j$ estimators, especially if there are several other predictors that measure the same underlying factors. Moreover, as shown by Pruzek et al. (1976, Note 2), the stability of individual (reduced-rank) regression coefficients will tend to increase as the number of predictors is increased for a fixed value of $m$, the number of factors. It should be made explicit that this tendency of reduced-rank regression estimators to improve under such conditions is opposite to what happens for conventional least squares regression estimators, which are known always to become less stable as $p$ is increased for a fixed $m$. In effect, conventional least squares regression estimates are unable to use factorial redundancy properly when it exists. When one's predictors are orthogonal but predictor–criterion correlations are nonzero, then the off-diagonal rank of $\mathbf{R}$ must be $p - 1$. But when predictors and criteria are correlated, and particularly when both are psychometrically sampled from the same domain of content, then the off-diagonal rank of population $\mathbf{R}$s can probably be assumed to be much less than $p - 1$.

## Discussion

This article was initiated, in part, as a response to Wainer (1976). For that reason much of the foregoing has dealt with what we see to be problems with assumptions underlying his equal weights theorem and with distinctions he did not make among distinguishably different prediction problems. Our arguments should nevertheless not be taken as evidence that highly simplified weighting schemes are always inadequate, nor even that for some domains of research, equal weighting of standardized predictors may be generally inappropriate. It has repeatedly been established by investigators such as Rozeboom (1966), Schmidt (1971), Dawes and Corrigan (1974), and Einhorn and Hogarth (1975), as well as Wainer (1976) *that in certain situations* unit weighting may indeed work well relative to idealized least squares weighting. It is important, therefore, that those who use regression methods for prediction have a basis for deciding whether or not their problems are likely to be amenable to simplified weighting schemes.

We believe that special benefits may be apt to follow from attention to prior structural beliefs about relationships among variables in a joint domain of predictors and criteria. The foregoing considerations suggest to us that prediction problems would often be dealt with more effectively if investigators approached each new problem with an aim to establish, on relevant theoretical or empirical grounds, just what the covariance patterns are likely to be among observable predictors and criteria. If these prior considerations lead to a view that correlations among predictors and criteria are reasonably homogeneous, then indeed, equal or simplified weighting schemes may be all that are required for the available predictors of the criterion. But if the prior studies lead instead to the view that some of the correlations among the prospective variables of interest are apt to be considerably different from others, and especially if one common factor is unlikely to be adequate for all variables, then equal or simplified weighting schemes are less likely to prove adequate.

The preceding discussion implies that a general strategy for designing and conducting applied prediction studies may commence from careful a priori consideration of the common factor structure of the joint predictor–criterion covariance system. A particularly efficient way of operationalizing such a strategy may often be to attempt on a priori grounds to write a matrix $\mathbf{F}_0$ of common factor coefficients that describes one's beliefs about the joint matrix of predictor–criterion correlations among prospective observable variables.

Suppose that one were to begin by specifying as few common factors as seemed necessary for $\mathbf{F}_0$ to describe the assumed factorial makeup of the criterion, selecting predictors so as to assure the highest possible determinacy of these factors. Further factors might be added to correspond to what is believed to be further stable covariance among the predictors. Computation of $\mathbf{R}_0 = \mathbf{F}_0\mathbf{F}'_0 + \mathbf{U}_0^2$ for any $\mathbf{F}_0$ would permit a direct assessment of the implied correlational structure for $\mathbf{F}_0$: $\mathbf{U}_0^2 = \text{diag}\ (\mathbf{I} - \mathbf{F}_0\mathbf{F}'_0)$. Given a matrix of hypothetical factor coefficients that an investigator has constructed to be representative of his structural beliefs for the joint predictor–criterion system, the corresponding $\mathbf{R}_0$ might be input to any standard

package program for regression analysis to compare, using transgeneration procedures, the potential virtues of various weighting schemes for any subsets of hypothetical predictors. Simplified or equal weights may be found, in this hypothetical framework, to be provisionally appropriate when a particular type of prior information is available. In fact, if one's prior beliefs about covariance structure happen to be nearly correct, then such rationally—or "clinically"—derived weights could be used to form a single predictor composite for this criterion, so as to obviate the need for any *multiple* regression methods. Indeed, one might expect routinely to be able to find subsets of predictors about which one has high confidence in the appropriateness of a prior linear combinations. Yet for some variables at least, it seems likely that such a priori judgments typically could not be made with impunity.

To the extent that one's prior structural beliefs do not correspond to reality, empirically derived weights and their associated validities may conceivably be much different than is suggested by the elements of $\mathbf{R}_0$. This is, of course, just a special case of a priori hypothesizing in the context of a particular problem of estimation. Exploring one's data further to learn what they have to say beyond the original questions or beliefs is a hallmark of good data analysis generally, and it is at least as relevant in prediction studies as in other contexts.

The only technical requirement in specifying $\mathbf{F}_0$ would be to make sure that all row sums of squares are less than unity. This insures that the derived, hypothetical $\mathbf{R}_0$ is positive definite and thus satisfies the first requirement of real-data product-moment covariance or correlation matrices. There would be no reason in principle why some of the prospective predictor variables could not be defined as squares or cubes of others, or as cross-products, and so forth. Moreover, even group indicator variables or contrasts (like sex, religion, or experimental treatment) could be used in the specification of rows of $\mathbf{F}_0$. Also, several criteria might be considered simultaneously. Whatever the details, of course, the whole process should be viewed chiefly as a vehicle for sharpening one's thinking about how to define and weight predictors of particular criterion variables. As long as such a strategy is used tentatively, so that the appropriateness of prior beliefs is ultimately judged in light of data, this would seem to be at least a step in the right direction for many prediction problems. It is conceivable that a strategy of this general form might be further developed to make explicit, in a coherent Bayesian context, for example, that any given $\mathbf{R}_0$ is only one of numerous possible population covariance systems. Thus, the strength of one's belief or confidence in a given $\mathbf{R}_0$ might be formally included in one's analysis. Whatever form such formalization may take, it seems crucial to insure that major attention be given to how one specifies the domain of joint predictor–criterion variables—in short, to how all variables are selected or defined in the first place. It is interesting to note that many conventions for test construction and validation may be viewed as special cases of the general strategy that has been outlined above.

A great deal of theoretical statistical work leads to the general conclusion that the strongest inductive inferences one can make arise through careful consideration of prior information and how to express it. It would be surprising if for prediction problems in particular the same principle did not hold.

## Reference Notes


1. Browne, M. W. *A critical evaluation of some reduced-rank regression procedures* (ETS RB 70-21). Princeton, N. J.: Educational Testing Service, 1970.
2. Pruzek, R. M., Walker, N. F., Frederick, B. C., Huba, M. E., & Sherry, J. R. *An integration of regression and factor analysis.* Paper presented at the meeting of the Psychometric Society, Murray Hill, N.J., April 1976.
3. Scott, J. T., Jr. *Combining regression and factor analysis for use in agricultural economics research.* Paper presented at the meeting of the Southern Agricultural Economics Association, Mobile, Ala., 1976.
4. Pruzek, R. M. *On the role of a priori information in linear prediction models.* Paper presented at the meeting of the Psychometric Society, Chapel Hill, N.C., June 1977.
5. Pruzek, R. M. *A comparison of alternative weighting systems in multiple regression analysis.* Paper presented at the meeting of the American Educational Research Association, New York, April 1977.

## References


Brogden, H. E. On the interpretation of the correlation coefficient as a measure of predictive efficiency. *Journal of Educational Psychology*, 1946, *37*, 64–76.

Dawes, R. M., & Corrigan, B. Linear models in decision making. *Psychological Bulletin*, 1974, *81*, 95–106.

Einhorn, H. J., & Hogarth, R. M. Unit weighting schemes for decision making. *Organizational Behavior and Human Performance*, 1975, *13*, 171–192.

Hays, W. L. *Statistics for the social sciences*. New York: Holt, Rinehart & Winston, 1973.

Herzberg, P. A. The parameters of cross-validation. *Psychometrika Monograph Supplement*, 1969, *34*(2, Pt. 2).

Hoerl, A. G., & Kennard, R. W. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 1970, *12*, 55–67.

Jöreskog, K. G. On the statistical treatment of residuals in factor analysis. *Psychometrika*, 1962, *27*, 335–354.

Keeping, E. S. *Introduction to statistical inference*. Princeton, N.J.: Van Nostrand, 1962.

Lawley, D. N., & Maxwell, A. E. *Factor analysis as a statistical method*. New York: American Elsevier, 1971.

Lawley, D. N., & Maxwell, A. E. Regression and factor analysis. *Biometrika*, 1973, *60*, 331–338.

Morrison, D. F. *Multivariate statistical methods* (2nd ed.). New York: McGraw-Hill, 1976.

Rozeboom, W. W. *Foundations of the theory of prediction*. Homewood, Ill.: Dorsey Press, 1966.

Pruzek, R. M. Factor analysis. In J. Belzer et al. (Eds.), *Encyclopedia of computer science and technology*, Vol. 8. New York: Dekker Press, 1977.

Pruzek, R. M., Sherry, J. R., Walker, N. F., & Pfeiffer, R. A. Interactive multivariate analysis generalized for explorations of data—IMAGE. In D. C. Hoaglin & R. E. Welsch (Eds.), *Proceedings of the ninth interface symposium on computer science and statistics*. Boston: Prindle, Weber & Schmidt, 1976.

Schmidt, F. L. The relative efficiency of regression and simple unit predictor weights in applied differential psychology. *Educational and Psychological Measurement*, 1971, *31*, 699–714.

Wainer, H. Estimating coefficients in linear models: It don't make no nevermind. *Psychological Bulletin*, 1976, *83*, 213–217.

Wainer, H., & Thissen, D. Three steps toward robust regression. *Psychometrika*, 1976, *41*, 9–34.

Wilks, S. S. Weighting systems for linear functions of correlated variables when there is no dependent variable. *Psychometrika*, 1938, *3*, 23–40.

Whittle, P. A principal component and least squares method of factor analysis. *Scandinavian Actuarial Journal*, 1953, *35*, 223–239.


Received February 28, 1977 ■