Psychological Bulletin
1978, Vol. 85, No. 2, 247–253

Comment on "Estimating Coefficients in Linear Models: It Don't Make No Nevermind"

James E. Laughlin
University of South Carolina

In a recent issue of *Psychological Bulletin,* Wainer claimed that it is rare when weighting coefficients in linear prediction models need be other than equal. To support this assertion he formulated and attempted to prove the *equal weights theorem.* It is shown, however, that this theorem is in error and that the loss in explained variance when replacing optimal least squares weights by equal coefficients is twice as great as contended. An alternative formulation and proof of the theorem is developed that (a) states the correct loss in explained variance, (b) is not dependent on any assumed distributional form for the optimal least squares weights, and (c) deals more relevantly with loss in relative predictive accuracy. The practical implications of the original and corrected equal weights theorems are briefly discussed.

In recent years there has been an increase in research attention focused on methods for estimating weighting coefficients in linear prediction models. The interest in this problem seems to parallel an increasing awareness that commonly used least squares estimates of population regression coefficients are quite unstable when based on small or moderate samples of data. Although several different methods have been proposed for deriving weights that are not as sensitive to sampling fluctuations, a major portion of recent research interest has centered on equal weighting schemes. This research has primarily indicated that in many circumstances the initial loss of predictive accuracy incurred when switching from sample least squares weights to equal coefficients is far outweighed by an accompanying large decrease in sampling error. This consistent finding can be little disputed (e.g., see Beckwith & Lehmann, 1973; Dawes & Corrigan, 1974; Einhorn & Hogarth, 1975; Lawshe & Shucker, 1959; Lehmann, 1971; Schmidt, 1971, 1972; Trattner, 1963; Wesman and Bennett, 1959; Fischer, Note 1).

In a recent issue of this journal, however, Wainer (1976) attempted to develop this argument supporting the adoption of equal weighting schemes one strong step further. In particular, Wainer purported to prove that in many circumstances "almost no loss in accuracy" (p. 213) is realized when optimal least squares coefficients are replaced by equal weights. If this assertion is true and is coupled with the knowledge that equal weighting schemes exhibit virtually no sampling error,[1] then one could develop a powerful argument supporting the use of equal weighting schemes in almost all situations. It would indeed be hard to conceive of any other weighting method leading to almost optimally accurate predictions while exhibiting no sampling error. In Wainer's own words, it would be "a very rare situation that called for regression weights which were unequal" (p. 216).

My purpose for commenting on the arguments presented in Wainer's article are threefold. First, I show that the equal weights theorem, which Wainer developed to prove the

The author wishes to thank Charles Lewis for his invaluable comments during the preparation of this article.

Requests for reprints should be sent to James E. Laughlin, Department of Psychology, University of South Carolina, Columbia, South Carolina 29208.

[1] There is indeed no sampling error for relative equal weights (e.g., all weights set to unity). These weights are appropriate when relative rather than absolute prediction is important.

"almost no loss" assertion cited above, is in error and that the loss in explained variance is at least twice as great as he concluded. Second, I present an alternative formulation and proof of the equal weights theorem that (a) indicates the correct loss in explained variance, (b) is not tied to any distributional form assumed for the optimal least squares weights, and (c) deals more relevantly with loss in relative predictive accuracy. Third, I briefly comment on the practical implications of the original equal weights theorem and the corrected version detailed here.

Error in the Equal Weights Theorem

Wainer (1976) indicated that the loss in variance explained when switching from a set of K optimal least squares weights to a set of equal coefficients can be measured by

$$\text{Loss} = (\boldsymbol{\beta} - a\mathbf{1})'\mathbf{R}(\boldsymbol{\beta} - a\mathbf{1}) = \boldsymbol{\gamma}'\mathbf{R}\boldsymbol{\gamma}, \quad (1)$$

where $\boldsymbol{\beta}$ represents a $K \times 1$ column vector of population least squares weights, a indicates the common equal weight, $\mathbf{1}$ is a $K \times 1$ column vector of 1s, and $\mathbf{R}$ represents the intercorrelations of the predictors in the population. When the K predictor measures are uncorrelated, as is a requirement of the equal weights theorem, then this loss in explained variance can be expressed simply as the sum of squared differences between the least squares weights β_j and the common equal weight a. Wainer showed that in order to determine the expected loss in variance explained when switching $\boldsymbol{\beta}$ to $a\mathbf{1}$, it is only necessary to (a) determine the expected loss in changing any one β_j to a and then (b) multiply the resultant loss by K.

Before it is possible to evaluate the expected loss numerically, assumptions must be made concerning the population distribution from which the β_js are expected to arise. In the equal weights theorem it is assumed that all possible regression weights will fall within the continuous interval (.25, .75) and that the probability of observing a β_j at different points in this interval is constant. Given this population form, Wainer contended that the expected loss in switching any β_j to a is given by

$$\text{E (Loss)} = 2\int_0^{.25} \gamma^2 d\gamma = \frac{2(1/4)^3}{3} = \frac{1}{96}, \quad (2)$$

where $\gamma = \beta_j - a$, and a is .5. It should be noted that the integral in Equation 2 is multiplied by 2 because it represents only one half of the symmetric interval over which the loss γ^2 is evaluated. However, using Equation 2 to express the expected loss for any β_j implies that the density of γ is 1. This is clearly in error. The density of γ is equal to the density of β_j and, under the conditions of the equal weights theorem, is given by $f(\gamma) = 2$ (so that the total area under the (.25, .75) rectangular distribution is 1). Hence, the appropriate integral that expresses the loss expected when switching any one β_j to a is

$$\text{E (Loss)} = 2\int_0^{.25} 2\gamma^2 d\gamma = \frac{4(1/4)^3}{3} = \frac{1}{48}. \quad (3)$$

Over K uncorrelated predictor measures the expected loss is then $K/48$, not $K/96$. Thus, the loss in explained variance is twice as great as Wainer contended.

In listening to the comments of colleagues on Wainer's article, it is clear that there is also some confusion regarding what is actually proven in this equal weights theorem. It is important to note that the above proof does not show that the loss will be $K/48$ when switching a set of K optimal weights uniformly spread out over the (.25, .75) interval to .5. What is instead shown can be summarized loosely as follows: Consider all possible sets of K optimal weights that can be obtained when each weight is allowed to take on a value anywhere within the (.25, .75) interval. If we were to determine for each of these sets the loss in predictive accuracy when switching the K optimal weights to .5, then the average loss across sets would be $K/48$. Note that this average loss (i.e., expected loss) does not simply consider the loss for a set of K weights uniformly spread out over the (.25, .75) interval. It also considers, for example, the loss incurred for a set of K weights that are all equal to .5 (where there would be no loss). Hence, the loss to expect for K weights that are allowed to vary over the (.25, .75) range is much different from the actual loss incurred when K uniformly spread out weights are switched to .5. This difference is more concretely demonstrated in Table 1.

Interestingly, if we were to adopt a guideline for determining what constituted an un-

Table 1
A Comparison of Expected and Actual Losses in Explained Variance

Number of predictors	Optimal weights[a]	Actual loss (%)[b]	Expected loss (%)[c]
2	.05, .55	12.5	4.17
3	.05, .3, .55	12.5	6.25
4	.05, .22, .38, .55	13.9	8.33
5	.05, .175, .30, .425, .55	15.6	10.42
6	.05, .15, .25, .35, .45, .55	17.5	12.50
7	.05, .13, .22, .30, .38, .47, .55	19.4	14.58
8	.05, .12, .19, .26, .34, .41, .48, .55	21.4	16.67

[a] For convenience, the optimal weights presented are equally spread out over a (.05, .55) interval.
[b] Entries refer to the actual loss in explained variance when switching the specified set of optimal weights to equal coefficients.
[c] Entries refer to the expected loss, assuming each weight is uniformly distributed over some positive interval of length .5.

acceptable loss in accuracy, such as the one expressed by Green (Note 2), namely, that a loss of 4% in total variance is serious, then it is clear after viewing Table 1 that there would be relatively few situations in which a switch to equal weighting coefficients would be acceptable. It is important to remember that this is a rejection of changing from an optimal weights model to an equal weights model in the population of observations and not necessarily a rejection to adopt equal weights as a practical alternative to sample least squares coefficients. It has already been noted that in applied situations an equal weighting scheme may be preferable due to its insensitivity to sampling error.

Equal Weights Theorem: Revision 1

It would be useful here to formulate and prove an alternative expression for the equal weights theorem stating the correct loss in explained variance when switching from optimal to equal weights. In particular, it would be informative for this revised equal weights theorem to reflect two basic changes in the form of the original theorem presented by Wainer. First, it would be more appropriate for the revised theorem to deal with the actual loss in predictive accuracy when switching K optimal weights to equal coefficients, instead of the loss expected across all possible sets of K weights that are restricted to fall within some interval. The notion of expected loss is simply not very meaningful in the present context. That is, for many different values of K and distributions of optimal weights that might be considered, there will be a large number of different sets of weights evaluated in the expected loss that will simply not be plausible sets of optimal weights under the linear regression model. Thus, for example, when it is assumed that each weight will fall within the (.25, .75) range, and K is greater than 4, over half of the sets of weights evaluated in the expected loss will lead to a multiple correlation greater than 1. Even when $K = 2$, it is impossible to obtain two regression weights equal to .75, since it suggests a multiple correlation of 1.06. Thus, knowing the expected loss in predictive accuracy is not as informative as knowing the general form for the actual loss in explained variance when switching any K weights to equal coefficients.

Second, it would also be more appropriate for the revised theorem to consider the loss in relative predictive accuracy, rather than the loss in absolute accuracy as does the original theorem. As Wainer himself pointed out, relative prediction "is the most typical kind of problem" (p. 216) found in the behavioral sciences. Measuring the loss in terms of relative accuracy would thus be more informative to the audience of this journal.

Given this consideration, I adopt the following alternative measure to represent the loss in relative predictive accuracy:

$$\text{Loss} = R_\beta^2 - R_\text{E}^2 = \boldsymbol{\beta}'\mathbf{R}\boldsymbol{\beta} - \frac{(\mathbf{1}'\mathbf{R}\boldsymbol{\beta})^2}{\mathbf{1}'\mathbf{R}\mathbf{1}}, \quad (4)$$

where R_β^2 is the squared multiple correlation in the population of observations, and R_E^2 is the squared population validity for the equal weights model (see Appendix for a more complete justification of this loss measure). This measure simply represents the difference in proportional variance explained between the optimal and equal weighting schemes. Note that Equation 4 can also represent the loss in absolute accuracy if it is assumed that (a) all variables are standardized and (b) the common weight is the optimal equal regression coefficient.

Considering the basic changes in form mentioned above, the revised equal weights theorem can be stated as follows:

When K uncorrelated predictor variables $\mathbf{x}_i$ ($i = 1, \ldots, K$) with zero means and unit variances are used in a linear regression model to predict a criterion variable $\mathbf{y}$ that is also scaled to zero mean and unit variance, then the proportional loss in variance explained when switching from optimal to equal weighting coefficients is $K\sigma_\beta^2$, where σ_β^2 represents the variance of the K optimal weights. When only one predictor is included in the linear model, no loss in explained variance is realized.

Proof

The proof of this revised equal weights theorem is straightforward. First, we know that the loss in explained variance when switching any K optimal weights to equal coefficients can generally be measured by Equation 4. In the equal weights theorem, however, we are restricting attention to those sets of predictors that are uncorrelated (i.e., $\mathbf{R} = \mathbf{I}$). Thus, substituting $\mathbf{I}$ for $\mathbf{R}$ in Equation 4 allows the loss to be expressible as

$$\text{Loss} = \boldsymbol{\beta}'\mathbf{I}\boldsymbol{\beta} - \frac{(\mathbf{1}'\mathbf{I}\boldsymbol{\beta})^2}{\mathbf{1}'\mathbf{I}\mathbf{1}} = \boldsymbol{\beta}'\boldsymbol{\beta} - \frac{(\mathbf{1}'\boldsymbol{\beta})^2}{\mathbf{1}'\mathbf{1}}. \quad (5)$$

Next, note that the various matrix expressions in Equation 5 can algebraically be written as

$$\boldsymbol{\beta}'\boldsymbol{\beta} = \sum_{j=1}^{K} \beta_j^2, \quad \mathbf{1}'\boldsymbol{\beta} = \sum_{j=1}^{K} \beta_j = K\bar{\beta}, \quad \text{and} \quad \mathbf{1}'\mathbf{1} = K, \quad (6)$$

where $\bar{\beta}$ is the mean of the K optimal weights. Substituting these algebraic equivalents into Equation 5 then leads the loss to be expressible as

$$\text{Loss} = \sum_{j=1}^{K} \beta_j^2 - \frac{(K\bar{\beta})^2}{K} = \sum_{j=1}^{K} \beta_j^2 - K\bar{\beta}^2, \quad (7)$$

which is well known to be $K\sigma_\beta^2$.

The loss in predictive accuracy is thus solely dependent upon the spread of the optimal weights. If the β_js are spread out equally over some positive interval of length .5, then the losses when switching to equal coefficients are as given in the "Actual loss" column of Table 1.

Note also that when only one predictor measure is employed, $\sigma_\beta^2 = 0$, and there is no loss in relative predictive accuracy. That is, we can switch the one β_j to any other positive value and still retain the same correlation between predicted and actual criterion scores.

Applicability of the Equal Weights Theorem

One of the main conditions of the equal weights theorem is that the set of K predictors be mutually uncorrelated. Unfortunately, it is rare when such a condition can be met in practice. Hence, unless some general statements can be formulated concerning the loss in explained variance as $\mathbf{R}$ increasingly departs from $\mathbf{I}$, the applicability of the equal weights theorem will be extremely limited. Wainer suggested that the loss in predictive accuracy "can be *diminished considerably* [italics added] when the $\mathbf{x}_i$s are not independent" (p. 214).[2] If this assertion is true, then we could consider the loss expressed in the equal weights theorem (where it is assumed that $\mathbf{R} = \mathbf{I}$) to represent an upper bound for the loss in accuracy as $\mathbf{R}$ increasingly departs from $\mathbf{I}$. However, although Wainer's assertion is in a sense technically correct, it is for the most part highly misleading.

Consider that any set of K positive regression weights could have been based on one of many different sets of intercorrelated predictors (i.e., one of many different $\mathbf{R}$s). Then, point-

[2] It is quite curious for Wainer to conclude that "almost no loss" in explained variance (when $\mathbf{R} = \mathbf{I}$) can be "considerably diminished" when there is intercorrelation among predictors. This is clearly misleading. How can almost no loss be considerably diminished?

ing to such works as those of Wilks (1938), Gulliksen (1950, chap. 20), and Ghiselli (1964, chapt. 10), it can be shown that when switching to equal coefficients the loss in explained variance will be smaller when the optimal weights are based on a more highly intercorrelated set of predictors. Thus, Wainer's assertion would appear substantiated.

Note, however, that as we consider some set of optimal weights to be based on different possible **R**s, we are correspondingly considering these weights to be based on different possible sets of predictor–criterion correlations. Further, note that when the correlations in **R** get higher, so also will the predictor–criterion correlations in order for the same K positive weights to remain optimal. This is where the misleading part of Wainer's assertion can be recognized. As the range of correlations in **R** is allowed to increase, it requires an increasingly unrealistic range of predictor–criterion correlations to retain the same weights as optimal. In relation to Wainer's original theorem, as **R** is allowed to depart from **I**, it becomes increasingly unrealistic to expect the optimal weights to fall within the (.25, .75) interval. To do so would require the predictor–criterion correlations to fall within an interval much higher than (.25, .75), even when K and the predictor intercorrelations are small. Several examples of this are presented in Table 2. Thus, although technically correct, it serves little practical utility to assert that the loss in explained variance will be diminished for a given distribution of optimal weights when **R** is different from **I**.

A more applicable approach to evaluating the effect on loss as **R** departs from **I** would be as follows: Assume that by orienting predictors properly we can restrict attention to predictors that will correlate positively with the criterion. Then consider that any set of K positive predictor–criterion correlations can be associated with one of many different possible sets of intercorrelated predictors (where each different **R** leads to a different set of optimal weights). The relevant question then is, for a given set of predictor–criterion correlations, can we generally expect either a consistent increase or decrease in the loss of explained variance as the **R** associated with the set is allowed to depart from **I**? Even though it can

Table 2
Examples of Unrealistically High Predictor–Criterion Correlations When $\boldsymbol{R} \neq \boldsymbol{I}$ for Weights Restricted to the (.25, .75) Interval

Optimal weights ($\boldsymbol{\beta}$)	Correlations: Interpredictor ($\mathbf{R}$)	Correlations: Predictor–criterion ($\mathbf{R}_{xy}$)
$\begin{bmatrix} .25 \\ .50 \\ .75 \end{bmatrix}$	$\begin{bmatrix} 1 & & \\ .2 & 1 & \\ .2 & .2 & 1 \end{bmatrix}$	$\begin{bmatrix} .5 \\ .7 \\ .9 \end{bmatrix}$
$\begin{bmatrix} .25 \\ .42 \\ .58 \\ .75 \end{bmatrix}$	$\begin{bmatrix} 1 & & & \\ .1 & 1 & & \\ .1 & .1 & 1 & \\ .1 & .1 & .1 & 1 \end{bmatrix}$	$\begin{bmatrix} .43 \\ .58 \\ .72 \\ .88 \end{bmatrix}$

empirically be shown that the loss will quite often increase, the answer to this question must be no. The direction of change in loss can vary and will depend primarily on the pattern of both interpredictor and predictor–criterion correlations.

It can be shown, however, that once some pattern of correlation exists among predictors, then the higher the correlations in the pattern, the larger will be the loss in explained variance when switching to equal coefficients. Support for this is intuitively straightforward. As the redundancy among predictors increases (through higher predictor intercorrelations), the K optimal weights change so as to minimize the effects of the increased redundancy. Equal weights, on the other hand, are independent of the level of predictor intercorrelations and thus suffer the full extent of any increased redundancy. Hence, the multiple correlation will always decrease less than the equal weights validity as the predictor intercorrelations get higher.

It should be clear that it is not at all obvious or necessarily true that the loss in explained variance will be considerably less when predictors are correlated. The situation is simply more complex than Wainer suggests. More research is clearly needed to identify those data structures that most appropriately lend themselves to an equal weights model.

Concluding Remarks

As was indicated earlier, the practical implications of Wainer's (1976) arguments are quite strong. If the assertion is true that there is no serious loss in accuracy incurred when switching from optimal to equal weights and it is coupled with the knowledge that equal weighting schemes entail virtually no sampling variability, then it would indeed be a rare situation that called for differential weighting coefficients. There would certainly be little point in searching for improved methods of estimating weighting coefficients, since equal weighting schemes would allow no room for improvement.

As detailed above, however, Wainer misstated the loss to expect when equal rather than optimal weights are employed. In many situations the loss in explained variance can in fact be regarded as serious or nonnegligible. Even though this "almost no loss" assertion can be rejected, it should still be recognized that equal weighting coefficients can perform remarkably well, particularly considering that they require no data to derive. They are simply not, however, the final solution to the problem of estimating weighting coefficients in linear prediction models, as Wainer seemed to suggest. It may well be that other weighting methods can be developed that will lead to predictions that are stable and more optimally valid. Indeed, this author has recently reported (Laughlin, Note 3) on an alternative method, which offers a Bayesian compromise between sample least squares and equal weighting coefficients, that has quite consistently led to a more attractive prediction model.

In sum, the extreme stance supporting equal weighting schemes that was taken by Wainer is simply not warranted.

Reference Notes

1. Fischer, G. W. *Four methods for assessing multiattribute utilities: An experimental validation.* University of Michigan, Engineering Psychology Laboratory, 1972.
2. Green, B. F. *Parameter sensitivity in multivariate methods.* Unpublished manuscript, Johns Hopkins University, 1974.
3. Laughlin, J. E. *A Bayesian alternative to the use of equal weighting coefficients to robustify regression.* Manuscript submitted for publication, 1977.

References

Beckwith, N. E., & Lehmann, D. R. The importance of differential weights in multiple attribute models of consumer attitude. *Journal of Marketing Research*, 1973, *10*, 141–145.

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.

Ghiselli, E. E. *Theory of psychological measurement.* New York: McGraw-Hill, 1964.

Gulliksen, H. O. *Theory of mental tests.* New York: Wiley, 1950.

Lawshe, C. H., & Shucker, R. E. The relative efficiency of four test weighting methods in multiple regression. *Educational and Psychological Measurement*, 1959, *19*, 103–114.

Lehmann, D. R. Television show preference: Application of a choice model. *Journal of Marketing Research*, 1971, *8*, 47–55.

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.

Schmidt, F. L. The reliability of differences between linear regression weights in applied differential psychology. *Educational and Psychological Measurement*, 1972, *32*, 879–886.

Tatsuoka, M. M. *Multivariate analysis in educational and psychological research.* New York: Wiley, 1971.

Trattner, M. H. Comparison of three methods for assembling aptitude test batteries. *Personnel Psychology*, 1963, *16*, 221–232.

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

Wesman, A. G., & Bennett, G. K. Multiple regression vs. simple addition of scores in prediction of college grades. *Educational and Psychological Measurement*, 1959, *19*, 243–246.

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

Appendix

The purpose of this appendix is to justify the relative loss measure presented in Equation 4.

It is well known (e.g., see Tatsuoka, 1971, Chap. 3) that for standardized variables, the squared multiple correlation can be obtained by

$$R_\beta^2 = \sum_{j=1}^{K} \beta_j r_{jy}, \tag{A1}$$

where r_{jy} is the jth predictor's correlation with the criterion. In matrix terms, Equation A1 can be expressed as

$$R_\beta^2 = \boldsymbol{\beta}'\mathbf{R}_{xy} = \boldsymbol{\beta}'\mathbf{R}\boldsymbol{\beta}, \tag{A2}$$

where $\mathbf{R}_{xy}$ is the $K \times 1$ vector of predictor–criterion correlations. The right-hand side of Equation A2 is obtained by recognizing that $\mathbf{R}_{xy} = \mathbf{R}\boldsymbol{\beta}$.

Next, note that any set of equal weights will produce predictions that will correlate the same with actual criterion scores. For convenience, assume the equal weights are unity. Then, following Tatsuoka (1971, chap. 5), it is possible to express the covariance between equal weight predictions and actual criterion scores as

$$\text{Cov}\ (Y, \hat{Y}_{\mathrm{E}}) = \mathbf{1}'\mathbf{R}_{xy} = \mathbf{1}'\mathbf{R}\boldsymbol{\beta}, \tag{A3}$$

where the right hand side of Equation A3 again follows from $\mathbf{R}_{xy} = \mathbf{R}\boldsymbol{\beta}$. Again following Tatsuoka, we can determine that the standard deviation of the equal weight predictions is

$$S_{\hat{Y}_{\mathrm{E}}} = (\mathbf{1}'\mathbf{R}\mathbf{1})^{\frac{1}{2}}. \tag{A4}$$

Then, from Equations A3 and A4, and remembering that the variance of actual criterion scores is assumed to be 1, we can write the correlation between equal weight predictions and actual criterion scores as

$$R_{\mathrm{E}} = \frac{\text{Cov}\ (Y, \hat{Y}_{\mathrm{E}})}{S_{\hat{Y}_{\mathrm{E}}} S_Y} = \frac{\mathbf{1}'\mathbf{R}\boldsymbol{\beta}}{(\mathbf{1}'\mathbf{R}\mathbf{1})^{\frac{1}{2}}}. \tag{A5}$$

Squaring Equation A5 and substracting from Equation A2 thus leads to the loss measure expressed in Equation 4 above:

$$\text{Loss} = R_\beta^2 - R_{\mathrm{E}}^2 = \boldsymbol{\beta}'\mathbf{R}\boldsymbol{\beta} - \frac{(\mathbf{1}'\mathbf{R}\boldsymbol{\beta})^2}{(\mathbf{1}'\mathbf{R}\mathbf{1})}.$$

Received November 29, 1976 ■