PSYCHOMETRIKA—VOL. 75, NO. 1, 58–69
MARCH 2010
DOI: 10.1007/S11336-009-9127-Y

CORRELATION WEIGHTS IN MULTIPLE REGRESSION

NIELS G. WALLER AND JEFF A. JONES

UNIVERSITY OF MINNESOTA

A general theory on the use of correlation weights in linear prediction has yet to be proposed. In this paper we take initial steps in developing such a theory by describing the conditions under which correlation weights perform well in population regression models. Using OLS weights as a comparison, we define cases in which the two weighting systems yield maximally correlated composites and when they yield minimally similar weights. We then derive the least squares weights (for any set of predictors) that yield the largest drop in R^2 (the coefficient of determination) when switching to correlation weights. Our findings suggest that two characteristics of a model/data combination are especially important in determining the effectiveness of correlation weights: (1) the condition number of the predictor correlation matrix, $\boldsymbol{R}_{\boldsymbol{xx}}$, and (2) the orientation of the correlation weights to the latent vectors of $\boldsymbol{R}_{\boldsymbol{xx}}$.

Key words: multiple regression, correlation weights, alternate weights, parameter sensitivity.

During the past forty years, numerous studies have documented the effectiveness of so-called "alternate" (non-OLS weights) regression weights (Goldberger, 1968; Hoerl & Kennard, 1970; Schmidt, 1971; Wainer, 1976, 1978; Wesman & Bennett, 1959). Commenting on this work, Wainer once suggested that when choosing between OLS and alternate weights, "it don't make no never mind" (Wainer, 1976). More recently, Dana and Dawes (2004) proclaimed that "[OLS] regression is rarely useful for prediction in most social science contexts" (p. 317). Although provocative, both views reflect a growing consensus among methodologists (e.g., Claudy, 1972; Einhorn & Hogarth, 1975; Green, 1977; Keren & Newman, 1978; Laughlin, 1978; Pruzek & Fredrick, 1978; Raju, Bilgic, Edwards & Fleer, 1997; Rozeboom, 1979; Schmidt, 1971; Wainer, 1976, 1978) that alternate regression weights—such as unit weights, rounded weights, and correlation weights ($\boldsymbol{r}_{\boldsymbol{x}y}$)—can often be profitably used in linear prediction and selection models.

In this paper we examine the use of correlation weights in population regression models. Although prior work (Campbell, 1974; Claudy, 1972; Dana & Dawes, 2004; Davis & Sauser, 1991; Goldberg, 1972; Marks, 1966) supported the effectiveness of correlation weights in sample data, much of this work relied on Monte Carlo simulations or reanalyses of existing data sets. A general theory on the use of correlation weights in linear prediction has yet to be proposed. The object of this paper is to sketch the initial strands of such a theory by describing the requisite conditions for correlation weights to perform well in population models.

The paper is divided into four sections. In Section 1, we show when correlation and OLS weights will yield maximally correlated composites. In Section 2, we show when the weights will be maximally separated in Euclidean space. In Section 3, we discuss weight sensitivity, which we define as the loss in $R_{\boldsymbol{b}}^2$ when switching from OLS to correlation weights, and we provide equations for locating the maximally sensitive OLS weights for any predictor correlation matrix. Finally, in Section 4, we describe two examples that illustrate the geometry of correlation weights in linear regression. To set bounds on our discussion, we assume throughout that (1) all variables have been standardized and that we are working with (2) a correctly specified population model. Although the use of standardized regression weights ($\boldsymbol{b}$) is somewhat controversial (Bring, 1994;

The authors would like to express their appreciation to Drs. Will Grove, Bob Pruzek, Scott Vrieze, and Steve Nydick for helpful comments on a draft of this article.

Requests for reprints should be sent to Niels G. Waller, Department of Psychology, University of Minnesota, N218 Elliott Hall, 75 East River Road, Minneapolis, MN, 55455 USA. E-mail: nwaller@umn.edu

Greenland, Schlesselman & Criqui, 1986), this is the appropriate metric when discussing correlation weights. To motivate the model, let y denote a random variable of criterion scores, let $\boldsymbol{x}$ denote a $p \times 1$ random vector of predictor variables, and let e denote a random variable of model errors. Then, when all variables have been standardized to have means of zero and unit variances and when $y = \boldsymbol{b}'\boldsymbol{x} + e$, the model coefficients, $\boldsymbol{b}$, and predictor/criterion correlations, $\boldsymbol{r}_{\boldsymbol{x}y}$, are linear functions of one another. Specifically,

$$\boldsymbol{b} = \boldsymbol{R}_{\boldsymbol{xx}}^{-1}\boldsymbol{r}_{\boldsymbol{x}y}, \tag{1}$$

where $\boldsymbol{b}$ is a $p \times 1$ vector of standardized regression weights, $\boldsymbol{R}_{\boldsymbol{xx}}$ is a $p \times p$ matrix of predictor correlations ($\boldsymbol{R}_{\boldsymbol{xx}} = \boldsymbol{E}[\boldsymbol{xx}']$), and $\boldsymbol{r}_{\boldsymbol{x}y}$ is a $p \times 1$ vector of correlation weights (i.e., criterion correlations; $\boldsymbol{r}_{\boldsymbol{x}y} = E[\boldsymbol{x}y]$). Hereafter we omit subscripts if the meaning of a term is obvious (e.g., $\boldsymbol{b} = \boldsymbol{R}^{-1}\boldsymbol{r}$).

1. For a Given Set of Predictors, When Will $\boldsymbol{b}$ and $\boldsymbol{r}$ Produce Maximally Correlated Composite Scores?

It is well known that with orthogonal predictors, correlation weights are equivalent to standardized OLS weights (i.e., when $\boldsymbol{R} = \boldsymbol{I}$, $\boldsymbol{b} = \boldsymbol{r}$). What is less well known is that with p predictors (and $\boldsymbol{R}$ full rank), there are $2p$ additional cases in which the two weighting systems yield perfectly correlated composite scores. These situations can be characterized as follows. Let

$$\boldsymbol{R}^{-1} = \boldsymbol{V}\boldsymbol{\Lambda}^{-1}\boldsymbol{V}', \tag{2}$$

where $\boldsymbol{V}$ is a $p \times p$ orthonormal matrix of latent vectors, and $\boldsymbol{\Lambda}$ is the $p \times p$ diagonal matrix of latent values of $\boldsymbol{R}$. Then

$$\boldsymbol{b} = \boldsymbol{V}\boldsymbol{\Lambda}^{-1}\boldsymbol{V}'\boldsymbol{r}. \tag{3}$$

It is easily proved that composites that have been constructed with $\boldsymbol{b}$ and $\boldsymbol{r}$ will be perfectly correlated whenever $\cos(\boldsymbol{b}, \boldsymbol{r}) = 1.00$. Consideration of (3) reveals that $\cos(\boldsymbol{b}, \boldsymbol{r}) = 1.00$ whenever $\boldsymbol{r}$ is collinear with an eigenvector ($\boldsymbol{v}_i$) of $\boldsymbol{R}$. More formally, if $\boldsymbol{r}$ is a scalar multiple of $\boldsymbol{v}_i$ such that

$$\boldsymbol{r} = c\,\boldsymbol{v}_i, \tag{4}$$

then

$$\begin{aligned}\boldsymbol{b} &= \boldsymbol{V}\boldsymbol{\Lambda}^{-1}\boldsymbol{V}'c\boldsymbol{v}_i \\ &= \frac{c}{\lambda_i}\boldsymbol{v}_i \\ &= \lambda_i^{-1}\boldsymbol{r}.\end{aligned} \tag{5}$$

When (5) is true, $\cos(\boldsymbol{b}, \boldsymbol{r}) = \cos(\lambda_i^{-1}\boldsymbol{r}, \boldsymbol{r}) = 1.00$. To find c, simply substitute (4) and (5) into $R_{\boldsymbol{b}}^2 = \boldsymbol{b}'\boldsymbol{r}$ (where $R_{\boldsymbol{b}}^2$ equals the squared multiple correlation between the criterion and the predictors) such that

$$R_{\boldsymbol{b}}^2 = \frac{c^2}{\lambda_i}. \tag{6}$$

Then, by simple algebra

$$c = \left(R_{\boldsymbol{b}}^2\lambda_i\right)^{1/2}. \tag{7}$$

Importantly, these results indicate that correlation and OLS weights will produce linearly related composites whenever $\widehat{y}$ (where $\widehat{y} = \boldsymbol{b}'\boldsymbol{x}$, i.e., the OLS composite) is collinear with a principal component of $\boldsymbol{x}$ (where the ith principal component of $\boldsymbol{x}$ equals $\boldsymbol{x}'\boldsymbol{v}_i$). Because the sign of an eigenvector is indeterminate, this implies that for any (full rank) $\boldsymbol{R_{xx}}$, there are $2p$ sets of potential predicted scores in which $|\cos(\boldsymbol{b}, \boldsymbol{r})| = 1.00$.

2. Measuring the Similarity of $\boldsymbol{b}$ and $\boldsymbol{r}$

A well-known measure of vector similarity is the vector cosine (Koopman, 1988). In this section we use $\cos(\boldsymbol{b}, \boldsymbol{r})$ to quantify the similarity between $\boldsymbol{b}$ and $\boldsymbol{r}$ and we describe equations for locating the minimum extremum of $\cos(\boldsymbol{b}, \boldsymbol{r})$ for a given $\boldsymbol{R}$ (the maximum extremum was discussed in the previous section). To formalize the problem, note

$$\cos(\boldsymbol{b}, \boldsymbol{r}) = \frac{\boldsymbol{b}'\boldsymbol{r}}{|\boldsymbol{b}||\boldsymbol{r}|}, \tag{8}$$

where for any vector $\boldsymbol{a}$, $|\boldsymbol{a}| = (\boldsymbol{a}'\boldsymbol{a})^{1/2}$. Computationally, we can simplify our task by redefining $\boldsymbol{r}$ and $\boldsymbol{b}$ in terms of the latent vectors ($\boldsymbol{V}$) of $\boldsymbol{R}$. Specifically, let

$$\boldsymbol{r} = \boldsymbol{V}\boldsymbol{w}, \tag{9}$$

where $\boldsymbol{w} = \boldsymbol{V}'\boldsymbol{r}$. Thus,

$$\boldsymbol{b} = \boldsymbol{V}\boldsymbol{\Lambda}^{-1}\boldsymbol{w}. \tag{10}$$

Recalling that $R^2_{\boldsymbol{b}} = \boldsymbol{b}'\boldsymbol{r}$, we have

$$R^2_{\boldsymbol{b}} = \boldsymbol{w}'\boldsymbol{\Lambda}^{-1}\boldsymbol{w}. \tag{11}$$

Substituting (9)–(11) into (8) yields

$$\cos(\boldsymbol{b}, \boldsymbol{r}) = \frac{\boldsymbol{w}'\boldsymbol{\Lambda}^{-1}\boldsymbol{w}}{(\boldsymbol{w}'\boldsymbol{\Lambda}^{-2}\boldsymbol{w})^{1/2}(\boldsymbol{w}'\boldsymbol{w})^{1/2}}, \tag{12}$$

an equation with p unknowns in $\boldsymbol{w}$. To minimize (8) we simply minimize (12) with respect to $\boldsymbol{w}$ subject to the constraint $\boldsymbol{w}'\boldsymbol{\Lambda}^{-1}\boldsymbol{w} - R^2_{\boldsymbol{b}} = 0$. More formally, let

$$f = \frac{\boldsymbol{w}'\boldsymbol{\Lambda}^{-1}\boldsymbol{w}}{(\boldsymbol{w}'\boldsymbol{\Lambda}^{-2}\boldsymbol{w})^{1/2}(\boldsymbol{w}'\boldsymbol{w})^{1/2}} + L\big(\boldsymbol{w}'\boldsymbol{\Lambda}^{-1}\boldsymbol{w} - R^2_{\boldsymbol{b}}\big), \tag{13}$$

where L is a Lagrange multiplier, and all other terms have been previously defined. The ith element of the gradient of f with respect to $\boldsymbol{w}$ and L can be written as

$$\begin{aligned}\frac{\partial f}{\partial w_i} = {} & \frac{2w_i}{\boldsymbol{\Lambda}_{[i,i]}(\boldsymbol{w}'\boldsymbol{\Lambda}^{-2}\boldsymbol{w})^{1/2}(\boldsymbol{w}'\boldsymbol{w})^{1/2}} \\ & - \frac{(\boldsymbol{w}'\boldsymbol{\Lambda}^{-1}\boldsymbol{w})w_i}{\boldsymbol{\Lambda}^2_{[i,i]}(\boldsymbol{w}'\boldsymbol{\Lambda}^{-2}\boldsymbol{w})^{3/2}(\boldsymbol{w}'\boldsymbol{w})^{1/2}} \\ & - \frac{(\boldsymbol{w}'\boldsymbol{\Lambda}^{-1}\boldsymbol{w})w_i}{(\boldsymbol{w}'\boldsymbol{\Lambda}^{-2}\boldsymbol{w})^{1/2}(\boldsymbol{w}'\boldsymbol{w})^{3/2}} + \frac{2Lw_i}{\boldsymbol{\Lambda}_{[i,i]}}\end{aligned} \tag{14}$$

and

$$\frac{\partial f}{\partial L} = \left(\boldsymbol{w}'\boldsymbol{\Lambda}^{-1}\boldsymbol{w} - R_b^2\right). \tag{15}$$

Four analytic solutions minimize (13). In each solution, the elements of $\boldsymbol{w}$ take on the following values:

$$w_i = \begin{cases} \pm\frac{\sqrt{2}}{2}\sqrt{R_{\boldsymbol{b}}^2\lambda_1}, & i = 1, \\ 0, & i \neq 1 \text{ or } p, \\ \pm\frac{\sqrt{2}}{2}\sqrt{R_{\boldsymbol{b}}^2\lambda_p}, & i = p. \end{cases} \tag{16}$$

When w_i is defined as above, f achieves a global minimum. This can be proved by noting that the first term of (13) can be expressed as a quadratic form and thus f cannot have a saddle point when $\boldsymbol{R}$ is full rank. In the previous section we proved that f reaches a maximum when $\boldsymbol{b}$ is collinear with an eigenvector of $\boldsymbol{R}$. f reaches a local minimum when any two elements of w, say w_i and w_j (other than $i = 1$ and $j = p$), equal $\pm\frac{\sqrt{2}}{2}\sqrt{R_{\boldsymbol{b}}^2\lambda_k}$, $k = i$ or j, and all remaining terms equal zero.

Notice in (16) that the positive ratio of the first and last elements of $\boldsymbol{w}$ equals an important quantity in regression diagnostics. Namely

$$\kappa = \left|\frac{w_1}{w_p}\right| = \sqrt{\frac{\lambda_1}{\lambda_p}},$$

where κ is the condition number of $\boldsymbol{R_{xx}}$. This is the first of many examples in which κ will play an important role. For a second example, notice in (16) that any vector minimizing (13) must lie in a plane defined by the first and last eigenvectors of $\boldsymbol{R_{xx}}$ (i.e., the eigenvectors that are paired with the largest and smallest eigenvalues of $\boldsymbol{R_{xx}}$). Moreover, when $\boldsymbol{r}$ is so defined (i.e., by (9) and (16)),

$$|\boldsymbol{r}| = \left(\frac{1}{2}R_{\boldsymbol{b}}^2(\lambda_1 + \lambda_p)\right)^{1/2}, \tag{17}$$

$$|\boldsymbol{b}| = \left(\frac{1}{2}R_{\boldsymbol{b}}^2\left(\lambda_1^{-1} + \lambda_p^{-1}\right)\right)^{1/2}, \tag{18}$$

and

$$\cos(\boldsymbol{b}, \boldsymbol{r})_{\min} = \frac{2}{(2 + \frac{\lambda_1}{\lambda_p} + \frac{\lambda_p}{\lambda_1})^{1/2}} \tag{19a}$$

$$= \frac{2}{(2 + \kappa^2 + \kappa^{-2})^{1/2}}. \tag{19b}$$

The previous equation merits further consideration for at least two reasons. First, it demonstrates that $\cos(\boldsymbol{b}, \boldsymbol{r})_{\min}$ can be expressed in terms of a single parameter: κ. Second, it shows that in the limit, as $\kappa \to \infty$, $\cos(\boldsymbol{b}, \boldsymbol{r})_{\min} \to 0$. This last point is illustrated in Figure 1, where it can be seen that even moderate condition numbers can be associated with widely separated weight vectors, $\boldsymbol{b}$ and $\boldsymbol{r}$.

To better understand the practical implications of these findings, we now consider the effect on $R_{\boldsymbol{b}}^2$ of using correlation weights when $\boldsymbol{b}$ and $\boldsymbol{r}$ are maximally separated. Let $R_{\boldsymbol{w}}^2$ denote the

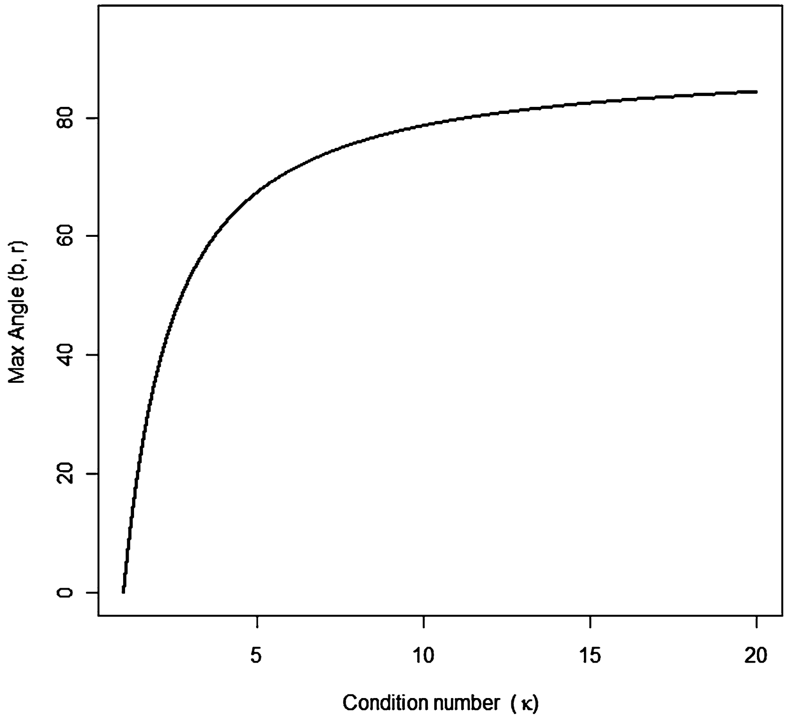

FIGURE 1.
The influence of the predictor condition number on the maximum angle between OLS and correlation weights.

squared multiple correlation that results from using the maximally separated correlation weights. From previous considerations, it can be shown (c.f. Tatsuoka, 1988, p. 45) that

$$R_w^2 = \frac{(w_1^2 + w_p^2)^2}{w_1^2\lambda_1 + w_p^2\lambda_p}. \tag{20}$$

After making appropriate substitutions,

$$R_w^2 = \frac{(0.5R_b^2\lambda_1 + 0.5R_b^2\lambda_p)^2}{0.5R_b^2\lambda_1^2 + 0.5R_b^2\lambda_p^2}, \tag{21}$$

an equation that reduces to

$$R_w^2 = 0.5R_b^2\frac{(\lambda_1 + \lambda_p)^2}{\lambda_1^2 + \lambda_p^2}. \tag{22}$$

Alternatively, (22) can be written as a function of κ,

$$R_w^2 = 0.5R_b^2\frac{1 + \kappa^{-4} + 2\kappa^{-2}}{1 + \kappa^{-4}}. \tag{23}$$

When expressed in this form, it is easily shown that

$$\lim_{\kappa\to\infty} R_w^2 = 0.5R_b^2. \tag{24}$$

Taking stock of these results, we have shown that when $\boldsymbol{b}$ and $\boldsymbol{r}$ are maximally separated, the use of correlation weights may result in the loss of up to 50% of the predictive power of a model. Of course, few data sets will approximate the precise requirements of these draconian results (e.g., extremely large values of κ). Nevertheless, these findings are useful in that they clearly set the bounds of a worst-case scenario. Practically speaking, however, perhaps a more informative bound for a given $\boldsymbol{R_{xx}}$ is the quantity $\max(R^2_{\boldsymbol{b}} - R^2_{\boldsymbol{r}})$. In the next section we show how this index can be easily calculated.

3. Calculating $\max(R^2_{\boldsymbol{b}} - R^2_{\boldsymbol{r}})$ for a Given $\boldsymbol{R_{xx}}$

To breathe some life into this topic, imagine a researcher who is interested in the predictive power of a test battery but who is not particularly interested in the individual regression weights of the final equation. This situation might arise in applied psychology when an individual uses a fixed test battery to predict diverse criteria. For instance, an I/O psychologist might use a standard test battery to predict multiple and diverse job performance criteria. In this setting, before using correlation weights in a prediction model, our I/O psychologist would be well advised to calculate $\max(R^2_{\boldsymbol{b}} - R^2_{\boldsymbol{r}})$ for typical values of $\boldsymbol{R_{xx}}$.

When solving this problem, our task is made easier by working with scaled correlation weights. Specifically, scale the weights to minimize the sum of squared errors around the regression line. The calculus reveals that to minimize $Q = \sum_{j=1}^{N}(y_j - s\boldsymbol{r}'\boldsymbol{x}_j)^2$ (where summation occurs over cases)

$$s = \frac{\boldsymbol{r}'\boldsymbol{r}}{\boldsymbol{r}'\boldsymbol{R}\boldsymbol{r}}. \tag{25}$$

Next, define $\boldsymbol{r}$ in terms of the orthonormal basis $\boldsymbol{V}$, where $\boldsymbol{R} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}'$, such that

$$\boldsymbol{r} = \boldsymbol{V}\boldsymbol{\omega} \tag{26}$$

and

$$\boldsymbol{\omega} = \boldsymbol{V}'\boldsymbol{r}.$$

Then, by substitution,

$$s = \frac{\boldsymbol{\omega}'\boldsymbol{\omega}}{\boldsymbol{\omega}'\boldsymbol{\Lambda}\boldsymbol{\omega}}, \tag{27}$$

$$R^2_{\boldsymbol{b}} = \boldsymbol{\omega}'\boldsymbol{\Lambda}^{-1}\boldsymbol{\omega}, \tag{28}$$

and

$$R^2_{\boldsymbol{r}} = s^2\boldsymbol{\omega}'\boldsymbol{\Lambda}\boldsymbol{\omega}, \tag{29}$$

where $R^2_{\boldsymbol{r}}$ denotes the coefficient of determination that results from using correlation weights. In these forms, it is easily shown that the desired weights are obtained by maximizing

$$g = R^2_{\boldsymbol{b}} - R^2_{\boldsymbol{r}} - L\big(\boldsymbol{\omega}'\boldsymbol{\Lambda}^{-1}\boldsymbol{\omega} - R^2_{\boldsymbol{b}}\big), \tag{30}$$

where L is a Lagrange multiplier, and all other terms are as previously defined. Making further substitutions,

$$\begin{aligned} g &= \boldsymbol{\omega}'\boldsymbol{\Lambda}^{-1}\boldsymbol{\omega} - s^2\boldsymbol{\omega}'\boldsymbol{\Lambda}\boldsymbol{\omega} - L\big(\boldsymbol{\omega}'\boldsymbol{\Lambda}^{-1}\boldsymbol{\omega} - R^2_{\boldsymbol{b}}\big) \\ &= \boldsymbol{\omega}'\boldsymbol{\Lambda}^{-1}\boldsymbol{\omega} - \left(\frac{\boldsymbol{\omega}'\boldsymbol{\omega}}{\boldsymbol{\omega}'\boldsymbol{\Lambda}\boldsymbol{\omega}}\right)^2 \boldsymbol{\omega}'\boldsymbol{\Lambda}\boldsymbol{\omega} - L\big(\boldsymbol{\omega}'\boldsymbol{\Lambda}^{-1}\boldsymbol{\omega} - R^2_{\boldsymbol{b}}\big), \end{aligned} \tag{31}$$

an equation that can be solved in terms of the p unknowns in $\boldsymbol{\omega}$. Furthermore, it can be proved that g is maximized whenever

$$\omega_i = \begin{cases} \left(\frac{R_{\boldsymbol{b}}^2}{\frac{1}{\lambda_1}+\frac{1}{\lambda_p}}\right)^{1/2}, & i = 1 \text{ or } p, \\ 0.00, & i \neq 1 \text{ or } p, \end{cases} \tag{32}$$

implying that

$$R^2_{\Delta\max} = \max\left(R^2_{\boldsymbol{b}} - R^2_{\boldsymbol{r}}\right) = R^2_{\boldsymbol{b}}\left(1 - \frac{4\lambda_1\lambda_p}{(\lambda_1+\lambda_p)^2}\right), \tag{33}$$

an equation that is also expressible in terms of κ (the condition number of $\boldsymbol{R_{xx}}$),

$$R^2_{\Delta\max} = R^2_{\boldsymbol{b}}\left(1 - \frac{4}{2+\kappa^2+\kappa^{-2}}\right). \tag{34}$$

Two immediate results follow from these solutions. First, at one extreme, when $\kappa = 1.00$, $\boldsymbol{R_{xx}} = \boldsymbol{I}$ (the case of orthogonal predictors) and $R^2_{\Delta\max} = 0.00$. This is simply another way of stating that correlation weights will equal standardized regression weights when working with orthogonal predictors. Second, as κ increases without bound,

$$\lim_{\kappa\to\infty} R^2_{\Delta\max} = R^2_{\boldsymbol{b}}. \tag{35}$$

In other words, under some conditions, correlation weights represent the absolute worst choice of weights. As illustrated in the next section, inspection of the OLS and correlation weights that produce this unfortunate result can yield important insights into the types of criteria that reside in this danger zone.

4. The Geometry of Correlation Weights

Before leaving this topic, we consider two examples that illustrate how the aforementioned results can be used to better understand the limitations of correlation weights in multiple regression. In our first example, we demonstrate an important point that, heretofore, has gone unrecognized. Namely, that small changes in a criterion can have dramatic changes on the effectiveness of correlation weights. We then explain this phenomenon in a second example that illustrates the geometry of correlation weights in a low-dimensional model.

To set the stage for these illustrations, let us take another look at (31)–(33). Notice in (32) that when g reaches a maximum, $\boldsymbol{r_{xy}}$ is an equally weighted composite of the first and last eigenvectors of $\boldsymbol{R_{xx}}$. Let $\boldsymbol{r}_{\Delta\max} = \boldsymbol{r_{xy}}$ at the solution point (i.e., when g is maximized). From previous considerations, it is easily proved that $\boldsymbol{r}_{\Delta\max}$ is situated 45° from the first and last eigenvectors of $\boldsymbol{R_{xx}}$. Moreover, simple algebra reveals that when $\boldsymbol{r_{xy}} = \boldsymbol{r}_{\Delta\max}$, $\boldsymbol{b}_{\Delta\max}$ (the associated OLS weight vector) will approach v_p (the last eigenvector of $\boldsymbol{R_{xx}}$). Specifically, when g reaches a maximum,

$$\cos(\boldsymbol{b}_{\Delta\max}, v_p) = \frac{1}{(1+\kappa^{-4})^{1/2}}, \tag{36}$$

an equation that approaches 1.00 as $\kappa \to \infty$ (recall, however, that when $\cos(\boldsymbol{b}, v_p) = 1.00$, $R^2_{\Delta\max} = 0.00$). These results can be used to show that extremely small changes in a criterion can have substantial effects on the performance of correlation weights. This important point is illustrated in the Appendix, where we report a small R (R Development Core Team, 2007) program to analyze data from 50,000 hypothetical subjects who were administered the Wechsler

Adult Intelligence Scale-III (WAIS; The Psychological Corporation, 1997). The predictor correlations for this example were originally reported in the WAIS-III test manual (Table A.7, p. 224). The R code in the Appendix generates raw data that will perfectly reproduce these correlations. It then generates scores for two criteria that we have labeled y_1 and y_2. Both OLS and correlation weights are used to predict these criteria.

All persons who have seen this example have been surprised by its findings, which can be briefly summarized as follows. The example demonstrates that when predicting y_1 with the simulated intelligence data, the OLS($\boldsymbol{b}$) or correlation weights ($\boldsymbol{r_{xy}}$) are equally effective. In *both* models, the coefficient of determination is 0.99. However, when predicting y_2, the two sets of weights are not equally effective. On the contrary, when y_2 is the criterion, the r-squared for the OLS weights is also 0.99, but it is only 0.09 when using correlation weights. In itself, this result is not remarkable. No one claims that correlation weights are a universally good choice. What makes this finding so surprising is that the correlation between y_1 and y_2 is 0.98! Here we have a clear example in which a small change in the criterion produced a large change in the performance of correlation weights.

When explaining these results to colleagues, we have found it useful to work with a smaller example that allows one to visualize the geometry of the problem in $\mathbb{R}^3$. Our example includes three predictors with the following correlations,

$$\boldsymbol{R_{xx}} = \begin{pmatrix} 1.00 & 0.80 & 0.30 \\ 0.80 & 1.00 & 0.34 \\ 0.30 & 0.34 & 1.00 \end{pmatrix},$$

with $\Lambda = (2.00, 0.80, 0.20)'$ and $\kappa = \sqrt{\frac{2.00}{0.20}} = 3.16$. For the sake of argument, imagine that we are interested in all criterion variables y_i ($i = 1, 2, \ldots, \infty$; where i runs over possible variables, not cases) such that $y_i = \boldsymbol{b}_i'\boldsymbol{x} + e_i$ and $R^2_{\boldsymbol{b}_i} = 0.40$. In other words, for a given set of predictor correlations, we are considering all criteria that can be predicted with an r-squared of 0.40. This constraint implies that

$$\boldsymbol{b}_i'\boldsymbol{R_{xx}}\boldsymbol{b}_i = 0.40, \tag{37}$$

an equation that is easily recognized as a quadratic form. It can be shown (Ferguson, 1979) that quadratic forms in $\mathbb{R}^3$ define ellipsoids in Euclidean space. In terms of our example, this implies that all $\boldsymbol{b}_i$ that satisfy (37) will define the surface of an ellipsoid with each $\boldsymbol{b}_i$ defining a point on this surface. We have found that by controlling the color of these points, one can literally visualize the effectiveness of correlation weights for all possible criteria. This idea is illustrated in Figure 2.

Each color in this figure represents a different amount of predictive loss, R^2_Δ, when using correlation weights in lieu of OLS weights for a given class of criteria. Each class represents the set of y_i that can be optimally (in a least squares sense) predicted with a given $\boldsymbol{b}_i$. To construct these figures, we generated 5 million sets of regression weights using a modified algorithm by Fishman (1996, p. 235) that samples points on the surface of an ellipsoid (defined by (37)). When constructing Figure 2, each point on this surface was colored using the following scheme.

- Blue: $R^2_\Delta \le 0.01$
- Red: $0.01 < R^2_\Delta \le 0.05$
- Orange: $0.05 < R^2_\Delta \le 0.10$
- Cyan: $0.10 < R^2_\Delta \le 0.20$
- Yellow: $0.20 < R^2_\Delta \le 0.30$
- White: $0.30 < R^2_\Delta$.

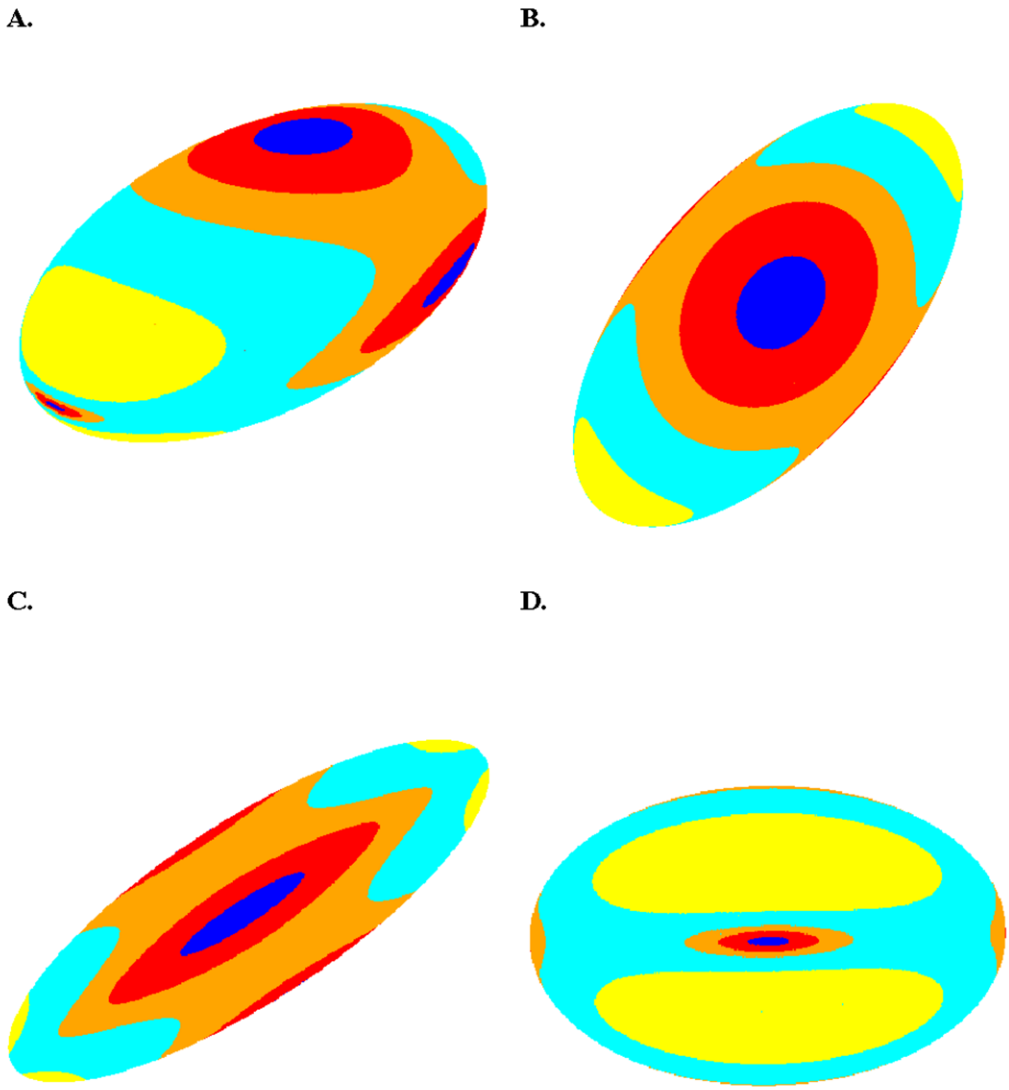

FIGURE 2.
Four views of the effectiveness of correlation weights.

The three axes of the ellipsoid point in the directions of the three eigenvectors of the aforementioned correlation matrix. In Panel 2B, our line-of-sight is oriented towards the first eigenvector; in Panel 2C, it is oriented towards the second eigenvector; in Panel 2D, we are facing the third eigenvector. Perhaps the most striking feature of these plots concerns the relative sizes and placements of the blue-colored patches. Recall that OLS vectors are colored blue whenever $R^2_\Delta \leq 0.01$. In other words, a blue-colored point indicates that for a given vector $\boldsymbol{b}_i$, $R^2_{\boldsymbol{b}} - R^2_{\boldsymbol{r}_{xy}} \leq 0.01$. Thus, in an important sense, blue-colored patches represent "neighborhoods of indifference" in that, when predicting relative standing, it makes little difference if one uses the optimal, OLS weights or the alternate, correlation weights.

Significantly, each (scaled) eigenvector of $\boldsymbol{R}_{xx}$ resides in one of these neighborhoods. Of equal importance is the fact that the six neighborhoods are not of equal size. Notice, for instance, that the blue patch that surrounds the first eigenvector is considerably larger than that surrounding the last eigenvector. This pattern will be observed with all correlation matrices that have eigenvalues $\mathbf{\Lambda} = (2.00, 0.80, 0.20)$. (Marsaglia & Olkin, 1984 show how to generate these matrices.) As $\boldsymbol{b}_i$ moves away from a neighborhood of indifference, R^2_Δ increases. Moreover, the rate of increase is highest in the neighborhood of the last eigenvector (Panel 2D) and smallest in the neighborhood of the first eigenvector (Panel 2B). (Recall that the longest axis of the ellipsoid corresponds to the *smallest* eigenvalue of the correlation matrix.) We used this observation to construct the data for the WAIS-III example. Specifically, we constructed y_1 to be collinear with the last principal component of the predictors, producing a situation in which both $\boldsymbol{b}$ and $\boldsymbol{r}_{xy}$ were collinear with the last eigenvector of $\boldsymbol{R}_{xx}$ (and thus, $R^2_\Delta = 0$). This direction (for $\boldsymbol{b}$) was chosen because the blue patch that is centered on the largest axis of this 14-dimensional hyper-ellipsoid is so small that trivial changes to the criterion produce dramatic changes in the effectiveness of correlation weights (recall that y_1 and y_2 correlate 0.98). Other choices that were available to us would have produced similar results.

Nearly three decades ago, Dunnette and Borman proclaimed that the "use of zero-order validity coefficients as weights is an under-utilized practice [and that] under a variety of conditions it seems preferable to other approaches" (Dunnette & Borman, 1979, p. 492). Although a handful of studies have supported this position during the past 40 years (Campbell, 1974; Claudy, 1972; Dana & Dawes, 2004; Davis & Sauser, 1991; Goldberg, 1972; Marks, 1966), heretofore, no one has offered a comprehensive theory of the putative effectiveness of correlation weights in linear prediction and selection models. We hope that by describing the conditions under which correlation weights perform well (and not so well) in population regression models, this work will lay the foundations for such a theory.

Appendix

```
   ##  R Code for Correlation Weights Example
   ##  Authors: Niels G. Waller and Jeff A. Jones
   ###################################################

# Correlations derived from Table A.7 page 224
# from WAIS-III manual
# The Psychological Corporation (1997). WAIS-III
# WMS-III Technical Manual.
# San Antonio, TX: Author.

wais3 <- matrix(
c(1,  .76, .58, .43, .75, .75, .42, .54, .41, .57, .64, .54, .50, .53,
 .76,  1,  .57, .36, .69, .71, .45, .52, .36, .63, .68, .51, .47, .54,
 .58, .57,  1,  .45, .65, .60, .47, .48, .43, .59, .60, .49, .56, .47,
 .43, .36, .45,  1,  .37, .40, .60, .30, .32, .34, .35, .28, .35, .29,
 .75, .69, .65, .37,  1,  .70, .44, .54, .34, .59, .62, .54, .45, .50,
 .75, .71, .60, .40, .70,  1,  .42, .51, .44, .53, .60, .50, .52, .44,
 .42, .45, .47, .60, .44, .42,  1,  .46, .49, .47, .43, .27, .50, .42,
 .54, .52, .48, .30, .54, .51, .46,  1,  .45, .50, .58, .55, .53, .56,
 .41, .36, .43, .32, .34, .44, .49, .45,  1,  .47, .49, .41, .70, .38,
 .57, .63, .59, .34, .59, .53, .47, .50, .47,  1,  .63, .62, .58, .66,
 .64, .68, .60, .35, .62, .60, .43, .58, .49, .63,  1,  .59, .50, .59,
 .54, .51, .49, .28, .54, .50, .27, .55, .41, .62, .59,  1,  .48, .53,
 .50, .47, .56, .35, .45, .52, .50, .53, .70, .58, .50, .48,  1,  .51,
 .53, .54, .47, .29, .50, .44, .42, .56, .38, .66, .59, .53, .51,  1),
  nrow=14,ncol=14)
```

```
## Generate 50,000 cases for WAIS III R matrix
NumSubj<-50000
library(MASS)
set.seed(111)
wais.data <- mvrnorm(n = NumSubj,
                     mu=rep(0,14),
                     Sigma=wais3,
                     empirical = TRUE)

EigVecs <- eigen(wais3)$vectors
EigVals <- eigen(wais3)$values

PC.Weights.1  <- matrix(EigVecs[, 1] * 1/sqrt(EigVals[1]), 14,1)
PC.Weights.14 <- matrix(EigVecs[,14] * 1/sqrt(EigVals[14]),14,1)

PC.Scores.1  <- scale(wais.data %*% PC.Weights.1)
PC.Scores.14 <- scale(wais.data %*% PC.Weights.14)

#yhat.1 is collinear with the last principal component
y1 <- PC.Scores.14 + rnorm(NumSubj,mean=0,sd=.1)

# R^2 from OLS weights
Rsq.OLS.1 <- summary(lm(y1~wais.data))$r.squared

rxy1 <- cor(y1, wais.data)

# R^2 from correlation weights
Rsq.rxy1 <- cor(y1, wais.data%*%t(rxy1) )^2

#yhat.2 is ALMOST parallel with last principal component
y2 <- .99*PC.Scores.14 +
         sqrt(1-.99^2)* PC.Scores.1  +
         rnorm(NumSubj,mean=0,sd=.1)

# R^2 from OLS weights
Rsq.OLS.2 <- summary(lm(y2~wais.data))$r.squared

rxy2<-cor(y2,wais.data)

# R^2 from correlation weights
Rsq.rxy2 <- cor(y2, wais.data%*%t(rxy2) )^2

r.y1.y2 <- cor(y1,y2)

cat("R^2 OLS weights, y1:", round(Rsq.OLS.1,3),"\n")
cat("R^2 rxy weights, y1:", round(Rsq.rxy1,3),"\n\n")

cat("R^2 OLS weights, y2:", round(Rsq.OLS.2,3),"\n")
cat("R^2 rxy weights, y2:", round(Rsq.rxy2,3),"\n\n")

cat("Correlation of y1 and y2:", round(cor(y1,y2),3),"\n")
```

```
#~~~~~~~~OUTPUT ~~~~~~~~~~~~~#
# R^2 OLS weights, y1: 0.99
# R^2 rxy weights, y1: 0.99
#
# R^2 OLS weights, y2: 0.99
# R^2 rxy weights, y2: 0.089
#
# Correlation of y1 and y2: 0.98
```

References

Bring, J. (1994). How to standardize regression coefficients. *American Statistician*, *48*, 209–213.

Campbell, J.P. (1974). *A Monte Carlo approach to some problems inherent in multivariate prediction with special reference to multiple regression* (Technical Report 2002). Personnel Training Research Program. Washington DC: Office of Navel Research.

Claudy, J.G. (1972). A comparison of five variable weighting procedures. *Educational and Psychological Measurement*, *32*, 311–322.

Dana, J., & Dawes, R.M. (2004). The superiority of simple alternatives to regression for social science predictions. *Journal of Educational and Behavior Statistics*, *29*, 317–331.

Davis, K.R., & Sauser, W.I. (1991). Effects of alternate weighting methods in a policy-capturing approach to job evaluation: A review and empirical investigation. *Personnel Psychology*, *44*, 85–127.

Dunnette, M.D., & Borman, W.C. (1979). Personnel selection and classification systems. *Annual Review of Psychology*, *30*, 477–525.

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

Ferguson, C.C. (1979). Intersections of ellipsoids and planes of arbitrary orientation and position. *Mathematical Geology*, *11*, 329–336.

Fishman, G.S. (1996). *Monte Carlo: Theory, algorithms and applications*. New York: Springer.

Goldberg, L.R. (1972). Parameters of personality inventory construction and utilization: A comparison of prediction strategies and tactics. *Multivariate Behavioral Research Monographs*, *7*(2).

Goldberger, A.S. (1968). *Topics in regression analysis*. New York: Macmillan.

Green, B.F. (1977). Parameter sensitivity in multivariate methods. *Multivariate Behavioral Research*, *12*, 263–287.

Greenland, S., Schlesselman, J.J., & Criqui, M.H. (1986). The fallacy of employing standardized regression coefficients and correlations as measures of effect. *American Journal of Epidemiology*, *123*, 203–208.

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

Keren, G., & Newman, J.R. (1978). Additional considerations with regard to multiple regression and equal weighting. *Organizational Behavior and Human Decision Processes*, *22*, 143–164.

Koopman, R.F. (1988). On the sensitivity of a composite to its weights. *Psychometrika*, *53*, 547–552.

Laughlin, J.E. (1978). Comment on "Estimating coefficients in linear models: It don't make no nevermind." *Psychological Bulletin*, *85*, 247–253.

Marks, M.R. (1966). *Two kinds of regression weights which are better than betas in crossed samples*. Paper presented at the meeting of the American Psychological Association, New York (September).

Marsaglia, G., & Olkin, I. (1984). Generating correlation matrices. *SIAM: Journal of Scientific and Statistical Computing*, *5*, 470.

The Psychological Corporation (1997). *WAIS-III WMS-III technical manual*. San Antonio, TX: Author.

Pruzek, R.M., & Fredrick, B.C. (1978). Weighting procedures in linear models: Alternatives to least squares and limitations of equal weights. *Psychological Bulletin*, *85*, 254–266.

R Development Core Team (2007). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN:3-900051-07-0, URL:http://www.R-project.org

Raju, N.S., Bilgic, R., Edwards, J.E., & Fleer, P.F. (1997). Methodology review: Estimation of population validity and cross-validity, and the use of equal weights in prediction. *Applied Psychological Measurement*, *21*, 291–305.

Rozeboom, W.W. (1979). Sensitivity of a linear composite of predictor items to differential item weighting. *Psychometrika*, *44*, 289–296.

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

Tatsuoka, M.M. (1988). *Multivariate analysis* (2nd ed.). New York: Wiley.

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

Wainer, H. (1978). On the sensitivity of regression and regressors. *Psychological Bulletin*, *85*, 267–273.

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

Manuscript Received: 12 NOV 2008
Final Version Received: 3 APR 2009
Published Online Date: 20 MAY 2009