PSYCHOMETRIKA—VOL. 73, NO. 4, 691–703
DECEMBER 2008
DOI: 10.1007/S11336-008-9066-Z

FUNGIBLE WEIGHTS IN MULTIPLE REGRESSION

NIELS G. WALLER

UNIVERSITY OF MINNESOTA

Every set of alternate weights (i.e., nonleast squares weights) in a multiple regression analysis with three or more predictors is associated with an infinite class of weights. All members of a given class can be deemed *fungible* because they yield identical *SSE* (sum of squared errors) and R^2 values. Equations for generating fungible weights are reviewed and an example is given that illustrates how fungible weights can be profitably used to evaluate parameter sensitivity in multiple regression.

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

1. Introduction

Given a criterion variable, y, and a vector variate $\boldsymbol{x}$ of length p, least squares multiple regression (MR) seeks the linear combination of $\boldsymbol{x}$ that minimizes the sum of squared errors (SSE) and maximizes the correlation ($r_{y\widehat{y}_{\boldsymbol{b}}}$) between y and $\widehat{y}_{\boldsymbol{b}}$, where $\widehat{y}_{\boldsymbol{b}} = \boldsymbol{b}'\boldsymbol{x}$ ($\boldsymbol{b}$ is a $p \times 1$ vector of weights). When $\boldsymbol{x}$ is full-rank, the so-called least squares (LS) vector, $\boldsymbol{b}$, is unique. Stated otherwise, if $\widehat{y}_{\boldsymbol{a}}$ denotes a linear combination of $\boldsymbol{x}$ using alternate weights $\boldsymbol{a}$, $\boldsymbol{a} \neq \boldsymbol{b}$ ($\widehat{y}_{\boldsymbol{a}} = \boldsymbol{a}'\boldsymbol{x}$) then, for any $\boldsymbol{a}$, $SSE_{\boldsymbol{b}} < SSE_{\boldsymbol{a}}$ and $r_{y\widehat{y}_{\boldsymbol{b}}} > r_{y\widehat{y}_{\boldsymbol{a}}}$ in the sample from which $\boldsymbol{b}$ has been calculated (Rencher, 2000, p. 238).

Although LS weights are optimal in the sense described above, it is well known that other weights may perform nearly as well as LS weights in many data sets (Goldberger, 1968; Green, 1977; Hoerl & Kennard, 1970; Koopman, 1988; Rozeboom, 1979; Tukey, 1948; Wilks, 1938). For instance, in small to moderate size samples, equal weights or simple validity coefficients ($\boldsymbol{a} = \boldsymbol{r}_{\boldsymbol{x}y}$) often work remarkably well with standardized predictors. Moreover, in cross-validation samples, these so-called alternate weights frequently outperform LS weights (Dana & Dawes, 2004; Dawes & Corrigan, 1974; Einhorn & Hogarth, 1975; Raju, Bilgic, Edwards, & Fleer, 1997; Schmidt, 1971). Commenting on this point, Wainer once quipped that when estimating coefficients in linear models "it don't make no never mind." (1976, p. 213) Although several researchers have taken issue with the generality of Wainer's comment (for corrections and extensions of Wainer's thesis, see Grove, 2001; Keren & Newman, 1978; Laughlin, 1978; Pruzek & Fredrick, 1978; Rozeboom, 1979; Wainer, 1978), one point is undeniable. In many samples, alternate regression weights (e.g., equal weights, rounded weights, correlation weights) perform surprisingly well.

Each set of alternate weights is associated with unique $SSE_{\boldsymbol{a}}$ and $R^2_{\boldsymbol{a}}$ values (where $R^2_{\boldsymbol{a}} = r^2_{y\widehat{y}_{\boldsymbol{a}}}$). Interestingly, however, in problems with three or more predictors, each pair of $SSE_{\boldsymbol{a}}$ and $R^2_{\boldsymbol{a}}$ values is associated with an infinite number of alternate weights. Indeed, a primary goal of this paper is to demonstrate that every set of alternate weights in a multiple regression analysis (with 3 or more predictors) is associated with an uncountably infinite class of weights. In the following, I describe the members of each such class as being *fungible* because they yield identical

The author wishes to thank Drs. Robyn Dawes, William Grove, Markus Keel, Leslie Yonce, Joe Rausch, the editor, and three anonymous reviewers for helpful comments on earlier versions 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

SSE_a and R_a^2 values (hereafter abbreviated as SSE and R^2). In this article, I show how to generate fungible weights for various R^2 values and how the examination of fungible weights can be an informative component of a model sensitivity analysis. Finally, using data on the prediction of graduate school performance (Kuncel, Hezlett, & Ones, 2001), I illustrate how fungible weights can be used to gauge parameter sensitivity in multiple regression.

In the remainder of the paper, I adopt the following notation and conventions. I assume that all variables have been standardized. Although many authors have argued against variable standardization in MR (Achen, 1982; Bring, 1994; Darlington, 1990; Greenland, Schlesselman, & Criqui, 1986; Kim & Ferree, 1981; King, 1986), social scientists often work with standardized scores because their variables lack metrical meaning. Moreover, working with standardized scores simplifies the mathematics of the paper without incurring a loss of generality (all results can be easily expressed for nonstandardized scores). Finally, I adopt the notation of Abadir and Magnus (2002) such that vectors are represented by bold-italic lowercase letters ($\boldsymbol{b}$), matrices are denoted by bold-italic uppercase letters ($\boldsymbol{R}$), and scalars and random variables are denoted by italic lower case letters (c).

2. Parameter Sensitivity in Multiple Regression

In multiple regression analysis, we can quantify *parameter variability* by computing (via likelihood or bootstrap methods) the covariance matrix of the estimated regression coefficients (Rencher, 2000, p. 135). This information is meaningful when the data conform to the assumptions of our inferential model (e.g., random sampling from a well-defined population). Alternatively, in any sample, we can quantify *parameter sensitivity* by substituting nonoptimal (i.e., alternate) weights into a regression equation (Green, 1977; Tate & Bryant, 1986) and then reexamining the model fit. Optimal (LS) weights are said to be *insensitive* when slightly different weights produce similar fit indices (e.g., SSE and R^2 values). They are deemed *sensitive* when small changes to the LS weights produce large changes in model fit. The topic of parameter sensitivity in multiple regression has been skillfully addressed by many authors (e.g., Goldberger, 1968; Green, 1977; Koopman, 1988; Rozeboom, 1979; Wilks, 1938). In the following paragraphs, I summarize some important findings from these authors as a prelude to our discussion of fungible weights. Let us first formalize the problem.

Let $\boldsymbol{x} = [x_1, x_2, \ldots, x_p]'$ denote a vector valued set of predictors with correlation matrix $\boldsymbol{R}$. Then using previously defined terms (where $\boldsymbol{b}$ denotes the LS weights and $\boldsymbol{a}$ denotes a set of alternate weights), linear composites of $\boldsymbol{x}$ can be defined as follows:

$$\widehat{y}_b = \boldsymbol{b}'\boldsymbol{x},$$
$$\widehat{y}_a = \boldsymbol{a}'\boldsymbol{x}.$$

We can quantify the similarity of these composites by directly correlating their scores or by using the following equation from the algebra of linear composites (see Green, 1977, p. 273; Tatsuoka, 1971, Chap. 5),

$$r_{\widehat{y}_a \widehat{y}_b} = \frac{\boldsymbol{a}'\boldsymbol{R}\boldsymbol{b}}{(\boldsymbol{a}'\boldsymbol{R}\boldsymbol{a})^{1/2}(\boldsymbol{b}'\boldsymbol{R}\boldsymbol{b})^{1/2}}. \tag{1}$$

Using (1) is preferable when we want to explore the similarities between weight vectors. For instance, in an important paper on parameter sensitivity in multiple regression, Koopman (1988) showed how consideration of (1) could be used to answer the following questions.

1. Can a composite that is very similar to the original composite (produced from the LS weights) be produced by weights that are very different from the original weights?

2. Can a composite that is very different from the original composite be produced by weights that are similar to the original weights?

To answer these questions, Koopman operationalized weight similarity as the cosine between vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ in (1). He then showed how numerical methods could be used to locate a vector $\boldsymbol{a}$ with either the maximum or minimum cosine with a given $\boldsymbol{b}$ under the constraint that $r_{\widehat{y}_a \widehat{y}_b}$ equals a user defined value. Finally, using an empirical data set, Koopman convincingly demonstrated—at least for his example—that similar composites could indeed be produced with strikingly dissimilar weights. On the other hand, he also showed that very different composites could not be produced with similar weights.

Koopman's paper is important because, among other things, it demonstrates how to estimate two weight vectors of particular interest from a class of fungible weights. Unfortunately, few applied researchers have taken advantage of this work, perhaps because Koopman's method requires the numerical optimization of a seemingly complex equation. Whatever the reason, as shown in the next section, less complex equations can be used to locate *all* weight vectors in a fungible class of weights. To help researchers calculate these weights, the Appendix includes *R* code (R Development Core Team, 2007) that will carry out the necessary calculations.

3. Calculating Fungible Weights

Our first task is to demonstrate that a set of fungible weights is populated by innumerable weight vectors. To calculate these vectors, and to populate the set, let us consider how (1) can be simplified into a more manageable form. Assume $\boldsymbol{R}$, the correlation matrix of $\boldsymbol{x}$, is full rank. Then

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

where $\boldsymbol{V}$ is a $p \times p$ ($p \geq 3$) matrix of orthonormal eigenvectors (such that $\boldsymbol{V}'\boldsymbol{V} = \boldsymbol{V}\boldsymbol{V}' = \boldsymbol{I}$) and $\boldsymbol{\Lambda}$ is the associated $p \times p$ diagonal matrix of eigenvalues. By substituting (2) into the numerator of (1), we find that $\boldsymbol{b}$ and $\boldsymbol{a}$ can be mapped into unit length vectors $\boldsymbol{u}$ and $\boldsymbol{k}$, where

$$\boldsymbol{u} = \frac{\boldsymbol{\Lambda}^{1/2}\boldsymbol{V}'\boldsymbol{b}}{(\boldsymbol{b}'\boldsymbol{R}\boldsymbol{b})^{1/2}}, \tag{3}$$

and

$$\boldsymbol{k} = \frac{\boldsymbol{\Lambda}^{1/2}\boldsymbol{V}'\boldsymbol{a}}{(\boldsymbol{a}'\boldsymbol{R}\boldsymbol{a})^{1/2}}, \tag{4}$$

such that

$$r_{\widehat{y}_a \widehat{y}_b} = \boldsymbol{k}'\boldsymbol{u}. \tag{5}$$

Importantly, starting from (5), given any unit length vector $\boldsymbol{u}$ we can easily find an appropriate, unit length vector $\boldsymbol{k}$. Specifically, let

$$\boldsymbol{k} = r\boldsymbol{u} + \boldsymbol{U}\boldsymbol{z}\left(1 - r^2\right)^{1/2}, \tag{6}$$

where r is shorthand for $r_{\widehat{y}_a \widehat{y}_b}$, $\boldsymbol{U}$ is any $p \times (p-1)$ orthonormal matrix such that $\boldsymbol{U}'\boldsymbol{U} = \boldsymbol{I}$ and $\boldsymbol{U}'\boldsymbol{u} = \boldsymbol{0}$, and $\boldsymbol{z}$ is a $(p-1) \times 1$ normalized random vector such that $\boldsymbol{z}'\boldsymbol{z} = 1$. Note that $\boldsymbol{U}$ is easily constructed using the Gram–Schmidt method (for details see Carroll & Green, 1997, p. 106). Under these conditions, it is easily shown that

$$\boldsymbol{k}'\boldsymbol{u} = r = \left(r\boldsymbol{u}' + \left(1 - r^2\right)^{1/2}\boldsymbol{z}'\boldsymbol{U}'\right)\boldsymbol{u}. \tag{7}$$

Having a suitable $\boldsymbol{k}$ in hand, we are ready to find $\boldsymbol{a}$. First, define

$$\widetilde{\boldsymbol{a}} = \boldsymbol{V}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{k}. \tag{8}$$

In (8), $\widetilde{\boldsymbol{a}}$ yields the appropriate correlation in (1), but it does not yield the minimum *SSE* of all weight vectors that are collinear with $\widetilde{\boldsymbol{a}}$. To simultaneously minimize the *SSE*, we multiply $\widetilde{\boldsymbol{a}}$ by a constant s (i.e., $\boldsymbol{a} = s\widetilde{\boldsymbol{a}}$), where

$$s = \frac{\boldsymbol{r}'_{\boldsymbol{x}y}\widetilde{\boldsymbol{a}}}{\widetilde{\boldsymbol{a}}'\boldsymbol{R}\widetilde{\boldsymbol{a}}}. \tag{9}$$

Now, after reexpressing s, we have shown that

$$\boldsymbol{a} = \left(\boldsymbol{r}'_{\boldsymbol{x}y}\boldsymbol{V}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{k}\right)\boldsymbol{V}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{k}. \tag{10}$$

At this point, it is useful to add subscripts to $\boldsymbol{a}$, $\boldsymbol{k}$, and $\boldsymbol{z}$. Obviously, there is an infinite number of vectors, $\boldsymbol{z}_i$, that can be plugged into (7). Thus, there is also an infinite number of vectors $\boldsymbol{k}_i$ in which $\boldsymbol{k}'_i\boldsymbol{u} = r_{\widehat{y}_a\widehat{y}_b}$. Each $\boldsymbol{k}_i$ can be (nonlinearly) mapped into an $\boldsymbol{a}_i$ by (10). Moreover, given $r_{\widehat{y}_a\widehat{y}_b}$, each $\boldsymbol{a}_i$ so defined will satisfy

$$r_{\widehat{y}_a\widehat{y}_b} = \frac{\boldsymbol{a}'_i\boldsymbol{R}\boldsymbol{b}}{(\boldsymbol{a}'_i\boldsymbol{R}\boldsymbol{a}_i)^{1/2}(\boldsymbol{b}'\boldsymbol{R}\boldsymbol{b})^{1/2}},$$

and

$$R^2_{\boldsymbol{a}_i} = r^2_{y\widehat{y}_a} = \boldsymbol{a}'_i\boldsymbol{R}\boldsymbol{a}_i. \tag{11}$$

From previous definitions, it can be shown that the expected mean and covariance matrix of $\boldsymbol{a}$ take on simple forms. Namely,

$$\mathrm{E}(\boldsymbol{a}) = r^2\boldsymbol{b}, \tag{12}$$

and

$$\mathrm{Cov}(\boldsymbol{a}) = \frac{1}{p-1}R^2_{\boldsymbol{b}}r^2\left(1-r^2\right)\boldsymbol{G}\boldsymbol{G}', \tag{13}$$

where $\boldsymbol{G} = \boldsymbol{V}\boldsymbol{\Lambda}^{-1/2}\boldsymbol{U}$.

In the next section, we will see that a consideration of (11) will help us understand the geometry of fungible weights. Before discussing this geometry, it is worth noting that the weight vectors identified by Koopman (1988) as being of particular interest will be among the vectors produced by (7) and (10) and can be found via methods that are described later.

4. The Geometry of Fungible Weights

One of the more interesting findings from the previous section was the discovery that the correlation between two weighted, linear composites of $\boldsymbol{x}$ can be represented by the cosine (or inner product) of two unit length vectors, $\boldsymbol{k}$ and $\boldsymbol{u}$, as defined above. We also learned that there are an infinite number of vectors $\boldsymbol{k}_i$ $(i = 1, 2, \ldots, \infty)$ such that $r_{\widehat{y}_a\widehat{y}_b} = \boldsymbol{k}'_i\boldsymbol{u}$. Combining these facts will help us to understand the geometry of fungible weights. To visualize this geometry, we will consider an example in $\mathbb{R}^3$ (3-dimensional space).

Imagine a cone suspended in space (see Mulaik, 1972, pp. 328–329, for another example of how a cone can represent the geometry of correlational indeterminancy). Further imagine that the tip of the cone is located at O the origin of an $\mathbb{R}^3$ Cartesian coordinate system. It may be helpful to think of a sugar cone (minus the ice cream). Now place a very thin straw into the center of the cone, such that at the cone's mouth, the straw is an equal distance from all sides of the cone. The straw represents a (3×1) vector $\boldsymbol{u}$ and the walls of the cone represent the infinite $\boldsymbol{k}_i$ as defined by (6). Recall that by construction, all $\boldsymbol{k}_i$ have identical cosines with $\boldsymbol{u}$, and thus the straw is in the precise center of the cone.

Now set aside the cone and visualize each $\boldsymbol{k}_i$ as a point in space, rather than as a vector (or directed line segment). From this perspective, the infinite collection of points, $\boldsymbol{k}_i$, will form a circle. It is interesting to ask how this circle is transformed when $\boldsymbol{k}_i$ is mapped into $\boldsymbol{a}_i$. To understand this mapping, it necessary to review two properties of $\boldsymbol{a}_i$. First, when $\boldsymbol{a}_i$ is constructed by the methods reviewed above, then the quadratic form $\boldsymbol{a}_i'\boldsymbol{R}\boldsymbol{a}_i$ is constant. Specifically,

$$r^2_{y\widehat{y}_{\boldsymbol{a}}} = \boldsymbol{a}_i'\boldsymbol{R}\boldsymbol{a}_i. \tag{14}$$

From a geometrical perspective, (14) implies that $\boldsymbol{a}_i$ must terminate on the surface of an ellipsoid because $\boldsymbol{R}$ is a full rank, symmetric matrix and $\boldsymbol{x}'\boldsymbol{A}\boldsymbol{x} = c_1$ (where c_1 is a constant and $\boldsymbol{A}$ is a positive definite, symmetric matrix) is the generalized equation for an ellipsoid.

A second property concerns a bilinear form. Namely,

$$c_2 = \boldsymbol{a}_i'\boldsymbol{R}\boldsymbol{b}, \tag{15}$$

where c_2 is constant. In general, bilinear forms such as (15) define hyperplanes. When $\boldsymbol{a}_i$ and $\boldsymbol{b}$ lie in $\mathbb{R}^3$, (15) defines a plane. Thus, by proving that c_2 is indeed a constant, we can prove that $\boldsymbol{a}_i$ lies at the intersection of a plane with an ellipsoid. In other words, we will prove that the circle defined by $\boldsymbol{k}_i$ maps into an ellipse defined by $\boldsymbol{a}_i$ (because a plane intersecting with an ellipsoid defines an ellipse, see Ferguson, 1979, for justification of this claim). A proof can be sketched as follows.

By construction,

$$r_{y\widehat{y}_{\boldsymbol{a}}} = (\boldsymbol{a}_i'\boldsymbol{R}\boldsymbol{a}_i)^{1/2}, \tag{16}$$

$$r_{y\widehat{y}_{\boldsymbol{b}}} = (\boldsymbol{b}'\boldsymbol{R}\boldsymbol{b})^{1/2}, \tag{17}$$

and

$$r_{\widehat{y}_{\boldsymbol{a}}\widehat{y}_{\boldsymbol{b}}} = \frac{\boldsymbol{a}_i'\boldsymbol{R}\boldsymbol{b}}{(\boldsymbol{a}_i'\boldsymbol{R}\boldsymbol{a}_i)^{1/2}(\boldsymbol{b}'\boldsymbol{R}\boldsymbol{b})^{1/2}}. \tag{18}$$

Equations (16)–(18) imply

$$r_{\widehat{y}_{\boldsymbol{a}}\widehat{y}_{\boldsymbol{b}}} = \frac{\boldsymbol{a}_i'\boldsymbol{R}\boldsymbol{b}}{r_{y\widehat{y}_{\boldsymbol{a}}} r_{y\widehat{y}_{\boldsymbol{b}}}},$$

which obviously equals

$$r_{\widehat{y}_{\boldsymbol{a}}\widehat{y}_{\boldsymbol{b}}}\, r_{y\widehat{y}_{\boldsymbol{a}}}\, r_{y\widehat{y}_{\boldsymbol{b}}} = \boldsymbol{a}_i'\boldsymbol{R}\boldsymbol{b}. \tag{19}$$

We have shown that (19) is constant for any problem, and thus that the fungible weights of three predictors will define an ellipse in $\mathbb{R}^3$. In higher dimensional problems, similar logic can be used to demonstrate that fungible weights will lie on a $(p-1)$-dimensional hyper ellipsoid in $\mathbb{R}^p$.

Notice that when $\boldsymbol{R} = \boldsymbol{I}$ (i.e., the case of orthogonal predictors), $\boldsymbol{a}_i$ will be collinear with $\boldsymbol{k}_i$ and $\boldsymbol{b}$ (and thus $r_{\boldsymbol{x}y}$) will be collinear with $\boldsymbol{u}$. This implies that for a given class of fungible weights, all $\boldsymbol{a}_i$ will have identical cosines with $\boldsymbol{b}$. Interestingly, it also implies that with orthogonal predictors there are an infinite number of solutions to Koopman's equations for locating the $\boldsymbol{a}_i$ that yields the maximum or minimum cosine with the LS weights (because $\cos \measuredangle \boldsymbol{a}_i O \boldsymbol{b}$ is constant for all i).

Before leaving this section, let us consider how the aforementioned geometry might illuminate the topic of parameter sensitivity. Our example will require us to return to the cone in $\mathbb{R}^3$. Visualize two vectors, $\boldsymbol{k}_i$ and $\boldsymbol{k}_j$, that lie on opposite sides of the cone. How far apart are the vectors? By considering (7), we can infer that wide cones are produced by fungible vectors having dissimilar corresponding coefficients whereas narrow cones are produced by vectors with similar coefficients. This suggests that an informative measure of parameter sensitivity might be $\cos \measuredangle \boldsymbol{k}_i O \boldsymbol{k}_j$, a value that is easily determined by recalling that the angle between $\boldsymbol{k}_i$ and $\boldsymbol{k}_j$ is twice the angle between $\boldsymbol{k}_i$ (or $\boldsymbol{k}_j$) and $\boldsymbol{u}$. Thus, using the formula for the cosine of a double angle (see Guttman, 1955, (44), for a similar application), we find that

$$\cos \measuredangle \boldsymbol{k}_i O \boldsymbol{k}_j = 2 r^2_{\widehat{y}_a \widehat{y}_b} - 1. \tag{20}$$

Alternatively, as shown in the next section, this is equivalent to

$$\cos \measuredangle \boldsymbol{k}_i O \boldsymbol{k}_j = 2 \frac{r^2_{y\widehat{y}_a}}{r^2_{y\widehat{y}_b}} - 1. \tag{21}$$

Later, we will consider additional ways in which fungible weights can be used to evaluate parameter sensitivity in regression coefficients.

Heretofore, we have formulated our problem from the perspective of (1). That is, we have considered the generation of weights that yield a specific $r_{\widehat{y}_a \widehat{y}_b}$ under the constraint that the sum of squared errors (when using the alternate weights) is minimal. Sometimes, we might wish to formulate the problem from a slightly “different angle” (pun intended) and generate weights that yield a specific $r_{y\widehat{y}_a}$ under the constraint that $r_{y\widehat{y}_a} < r_{y\widehat{y}_b}$. As shown below, this is easily done once we recognize the similarity of the two problems. To understand this similarity, recall that

$$r_{y\widehat{y}_a} = \frac{\text{cov}_{y\widehat{y}_a}}{\text{sd}_y \text{sd}_{\widehat{y}_a}}. \tag{22}$$

Equation (22) can be simplified by recognizing that $\text{cov}_{y\widehat{y}_a}$ equals $\text{cov}_{\widehat{y}_a \widehat{y}_b}$ because e, the LS residuals (where $y = \widehat{y} + e$), lies outside of the predictor space. Therefore, when $\boldsymbol{x}$ and y are standardized,

$$r_{y\widehat{y}_a} = \frac{r_{\widehat{y}_a \widehat{y}_b} \text{sd}_{\widehat{y}_a} \text{sd}_{\widehat{y}_b}}{(1)\text{sd}_{\widehat{y}_a}}$$

which after making appropriate substitutions yields

$$r_{y\widehat{y}_a} = r_{y\widehat{y}_b} r_{\widehat{y}_b \widehat{y}_a}. \tag{23}$$

The three terms in (23) are known or are easily calculated; thus, it is a simple matter to generate weights for either a specific $r_{y\widehat{y}_a}$ or a specific $r_{\widehat{y}_a \widehat{y}_b}$. (Note that (23) also implies that $r^2_{y\widehat{y}_a} = \boldsymbol{a}_i' \boldsymbol{R} \boldsymbol{b}$.) Let us see how this is done.

Imagine that a researcher performs a LS regression and obtains an R^2 of 0.65. Further suppose that she decides that any (appropriately scaled) weight vector that produces an R^2 of 0.64—a reduction of only 1% predicted variance—is “good enough” from a predictive standpoint and that *some* vectors satisfying this criterion might merit theoretical consideration (and evaluation in cross-validation samples). How should our hypothetical researcher proceed?

TABLE 1.
Six fungible weight vectors for the GRE data.

	$\boldsymbol{a}_1$	$\boldsymbol{a}_2$	$\boldsymbol{a}_3$
$\boldsymbol{a}_1$ high	**0.33**	0.12	0.00
$\boldsymbol{a}_2$ high	0.22	**0.23**	0.01
$\boldsymbol{a}_3$ high	0.16	0.07	**0.23**
$\boldsymbol{a}_1$ low	**0.13**	0.14	0.19
$\boldsymbol{a}_2$ low	0.24	**0.03**	0.17
$\boldsymbol{a}_3$ low	0.31	0.19	**−0.04**

To find the needed quantities for (23), let $R_{\boldsymbol{b}}^2$ denote the R^2 using the LS weights ($\boldsymbol{b}$) and let $R_{\boldsymbol{a}}^2$ denote the coefficient of determination using the alternate weights. Define $\theta = R_{\boldsymbol{b}}^2 - R_{\boldsymbol{a}}^2$. In the current example, $\theta = 0.01$. From (23), a little algebra reveals that

$$r_{\widehat{y}_a \widehat{y}_b} = \left(1 - \frac{\theta}{R_{\boldsymbol{b}}^2}\right)^{1/2}. \tag{24}$$

Previously described equations can now be used to generate the fungible weights. With the aid of real data, we will consider a second example in greater detail.

5. Empirical Application: Are GRE Subtests "Worth Their Regression Weights?"

This example uses data from the Graduate Record Exam (GRE) as reported in Kuncel, Hezlett, and Ones (2001). These authors combined data from 82,659 students to assess the usefulness of the GRE for predicting various measures of graduate school performance. Using data from Tables 3 and 8 of their original report, we will consider the prediction of grade point average (GPA) for social science students. The following matrices report the correlations ($\boldsymbol{R_x}$) among the verbal, quantitative, and analytic subtests of the GRE and the correlations ($\boldsymbol{r}_{\boldsymbol{x}y}$) between the GRE subtests and GPA.

$$\boldsymbol{R_x} = \begin{pmatrix} 1.00 & 0.56 & 0.77 \\ 0.56 & 1.00 & 0.73 \\ 0.77 & 0.73 & 1.00 \end{pmatrix}, \qquad \boldsymbol{r}_{\boldsymbol{x}y} = \begin{pmatrix} 0.39 \\ 0.34 \\ 0.38 \end{pmatrix}.$$

The standardized regression weights for these data equal $\boldsymbol{b} = (0.24, 0.14, 0.10)$ producing a model $R^2 = 0.176$. Although these weights are relatively similar, for illustrative purposes, we can imagine that an enthusiastic admissions officer uses these findings to place greater emphasis on the GRE verbal score when selecting social science students. Given the extremely large sample size of this example, the variances of the estimated regression coefficients will be exceedingly small. In other words, parameter variability is small. But what about parameter sensitivity? To investigate this alternative measure of precision, we can compute and evaluate various sets of fungible weights.

Using the data reported above and the computer code that is reported in the Appendix, I generated 20,000 samples of fungible weights for the set $R^2 = 0.171$. Notice that this value is a mere half a percentage point lower than the $R_{\boldsymbol{b}}^2$ obtained from the optimal (LS) weights.

In Table 1 are six weight vectors from this analysis which were selected from the larger collection of 20,000 fungible vectors because they include the highest and lowest values for each

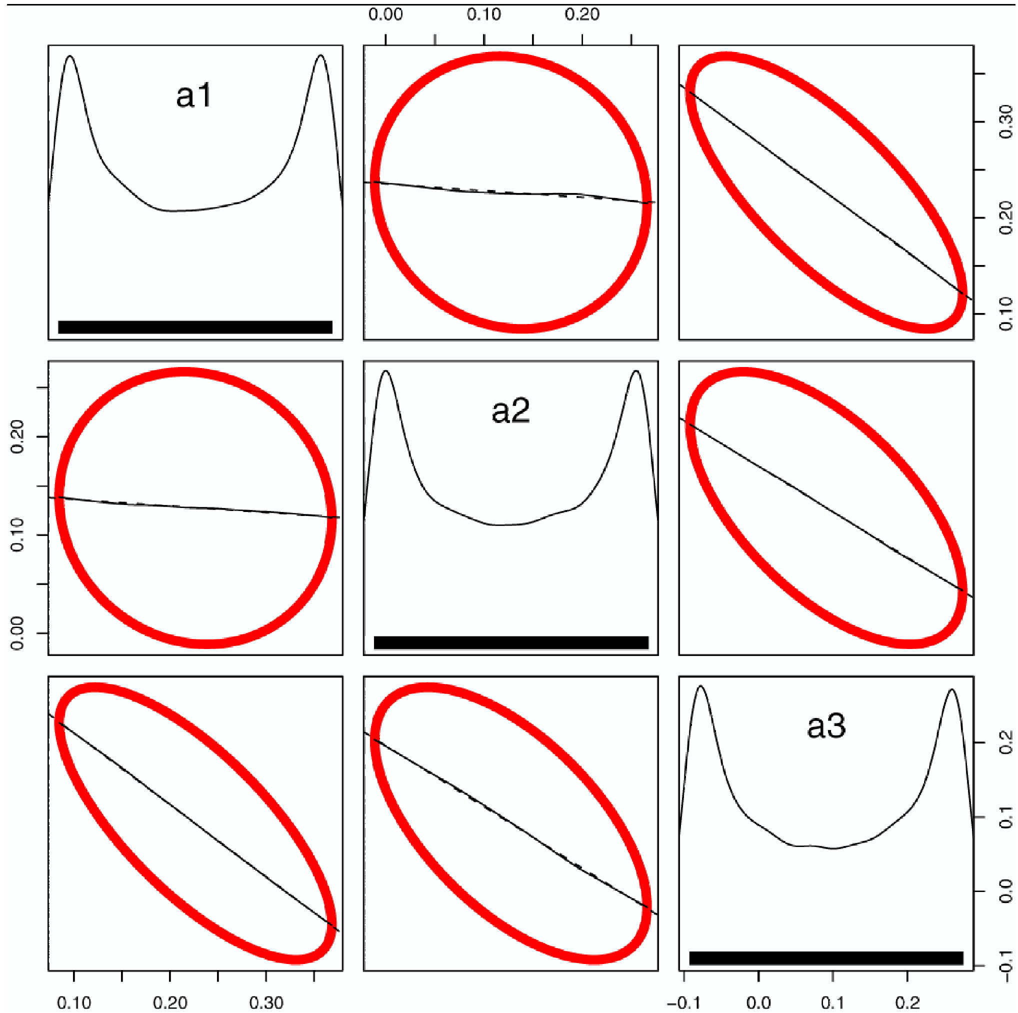

FIGURE 1.
Scatterplot matrix of fungible weights.

coefficient. Scanning this table makes it immediately clear that the rankings of the LS weights are poor indicators of variable importance, since an exceedingly small reduction in predicted variance is consistent with many weight vectors that have different coefficient ranks and quite different coefficient magnitudes.

We can also gauge parameter sensitivity by examining univariate and bivariate plots of the fungible weights. To illustrate this idea, Fig. 1 displays smoothed density plots and scatter plots for the 20,000 weight vectors of the GRE example. Several aspects of this figure warrant discussion. For instance, notice that in the diagonal cells the density plots show marked bimodality. This indicates that the alternate weights typically receive either a high or low value within their distributions. The scatter plots demonstrate that whether a coefficient receives an extreme value is a tightly constrained function of the remaining coefficients. This is shown by the plots (with superimposed LS regression lines) looking more like ellipses than point clouds with associated variation. These plots contain the actual raw data of the fungible weights.

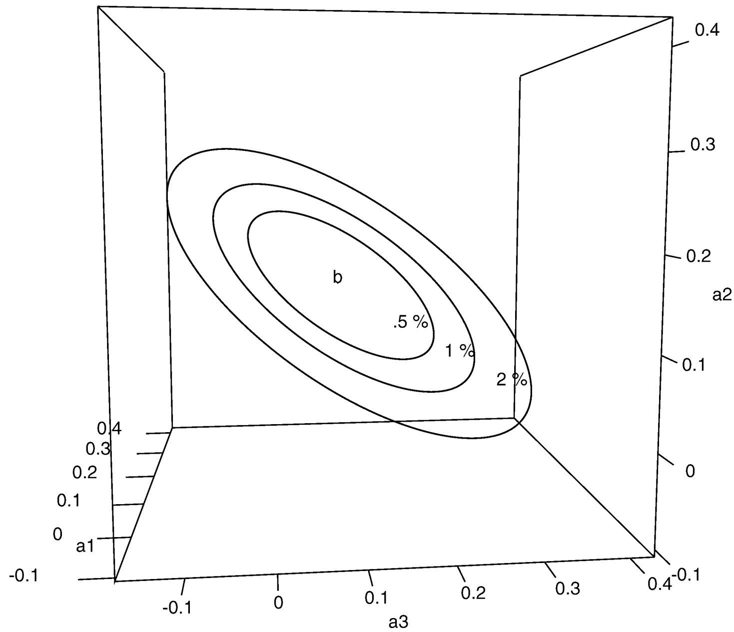

FIGURE 2.
Three classes of fungible weights for the example problem.

Finally, in examples with three predictors, we can plot the fungible weights in $\mathbb{R}^3$. This idea is illustrated in Fig. 2, which displays fungible weights for three separate classes of weights. The three classes (or sets) have different R^2 values. The smallest ellipse includes weights that yield an $R_{\boldsymbol{a}}^2$ that is 0.5% lower than the LS $R_{\boldsymbol{b}}^2$, whereas the largest ellipse includes weights that yield an $R_{\boldsymbol{a}}^2$ that is 2% lower than the LS $R_{\boldsymbol{b}}^2$. In higher dimensional problems, it may be instructive to create similar plots with predictor subsets.

6. Discussion

Numerous researchers have reminded us that coefficient size is an unreliable index of parameter importance in multiple regression (e.g., Budescu, 1993; Darlington, 1968; Green, Carroll, & DeSarbo, 1978; Grömping, 2007; Johnson, 2000). Yet, in applied settings, regression coefficients are still often ascribed undue importance. One reason why this practice is problematic concerns the performance of alternate weights. Alternate weights, such as unit weights, rounded weights, or simple validity correlations can be very dissimilar from the optimal LS weights yet still perform remarkably well in calibration samples. Moreover, these so-called nonoptimal weights may actually outperform LS weights under cross-validation. These facts have led some researchers to proclaim that "it don't make no never mind" when estimating coefficients in linear models.

In this article, we have seen that every set of alternate weights (with 3 or more predictors) is associated with an uncountably infinite class of weights. All members of a given class yield identical *SSE* and R^2 values and are, therefore, deemed *fungible*. Because fungible weights are easily calculated, they offer a convenient means for evaluating parameter sensitivity in multiple regression and recursive path analysis models. In other words, fungible weights can help you decide whether, in fact, "it do or don't make no never mind."

Appendix: R Code for Fungible Weights

```
## R function: Fungible
## Author: Niels Waller
## March 11, 2008
##
## Input Variables
## R.X   p x p Predictor variable correlation matrix.
## rxy   p x 1 Vector of predictor-criterion correlations.
## r.yhata.yhatb  = correlation between least squares (yhatb)
##            and alternate-weight (yhata) composites.
## sets Number of returned sets of fungible weights.
## print logical, if TRUE then print 5-point summaries
##         of alternative weights
##
## Output Variables
## a   sets x p matrix of fungible weights
## k   sets x p matrix of k weights
## b   p x 1 vector of LS weights
## u   p x 1 vector of u weights
## r.yhata.yhatb  correlation between yhata and yhatb
## r.y.yhatb     correlation between y and yhatb
## cov.a  Expected covariance matrix for a
## cor.a  Expected correlation matrix for a

Fungible <- function(R.X,rxy,r.yhata.yhatb,sets=20,print=TRUE){

 GenU <- function(mat,u){
   ## Generate U matrix via Gram Schmidt
   p <- ncol(mat)
   n <- nrow(mat)
   oData <-matrix(0,n,p+1)
   oData[,1]<-u

   for(i in 2:(p+1)){
     oData[,i] <- resid(lm(mat[,(i-1)]~-1+oData[,1:(i-1)]))
   }

   U<-oData[,2:(p+1)]
   d <- diag(1/sqrt(diag(crossprod(U))))
   U <- U%*%d
   U
 }#end GenU

  NX <- ncol(R.X)
  a.matrix <- k.matrix <- matrix(0,sets,NX)

  #OLS weights
  b <- crossprod(solve(R.X),rxy)
  r <- as.numeric(r.yhata.yhatb)

  VLV <- eigen(R.X)
  V <- VLV$vectors
  L <- diag(VLV$values)
```

```
  Linv.sqrt <- solve( sqrt(L))
  u.star <- t(V)%*%b

  u.circle <- sqrt(L) %*% u.star
  u <- u.circle/ as.numeric(sqrt((t(u.circle) %*% u.circle)))

  r.y.yhatb <- sqrt( (t(b) %*% R.X %*%b) )

  mat <- matrix(rnorm(NX*(NX-1)),NX,NX-1)
  U <- GenU(mat,u)

  for(i in 1:sets){

     z <- rnorm((NX-1))
     z <- z / as.numeric( sqrt( t(z) %*% z))

     k <- r * u + + U %*% z * sqrt(1-r^2)
     k.star <- Linv.sqrt%*%k
     a <- V %*% k.star

# scale a to minimize SSE_a
   s <- (t(rxy) %*% a)/(t(a)%*%R.X%*%a)

   a <- as.numeric(s) * a

   if(i==1) {
     cat("\n\nGenerating alternate weights . . . \n")
     r.yhata.yhatb <- (t(a) %*% R.X %*%b)/( sqrt((t(a)%*%R.X%*%a))*
                sqrt(t(b)%*%R.X%*%b))
   }

   a.matrix[i,] <-a
   k.matrix[i,] <- k
 }
 cat("\n\n")
 cat(" r.yhata.yhatb = ",r.yhata.yhatb,"\n")
 cat(" RSQb = ",round(r.y.yhatb^2,3),"\n")
 cat(" RSQa = ",round((r.yhata.yhatb * r.y.yhatb)^2,3),"\n")
 cat(" Relative loss = RSQb - RSQa = ",
 round( r.y.yhatb^2-(r.yhata.yhatb * r.y.yhatb)^2 ,3),"\n")
 cat(" OLS b = ",t(round(b,3)),"\n\n")
 cat("\n")
 colnames(a.matrix) <- paste("a",1:NX,sep="")

 if(print){
  cat("\nSummary of generated alternate weights\n")
  print( apply(a.matrix,2,summary) )
  cat("\n")
 }

# Compute Expected Moments
  G <- V%*%Linv.sqrt%*%U
  esq <- (1-r^2)
# Expected a
```

```
  mn.a <-  r^2 * b
  cat("\n Expected a \n")
  print(mn.a)
  Ezsq <- 1/(NX-1)

# Expected covariance matrix
  cov.a <- as.numeric(r^2 * r.y.yhatb^2)* Ezsq * esq * G %*%t(G)

  cat("\nExpected Covariance Matrix \n")
  print(cov.a)
  Dmat <- diag(1/sqrt(diag(cov.a)))
  cor.a <- Dmat %*% cov.a %*% Dmat
  cat("\nExpected Correlation Matrix \n")
  print(cor.a)

 list(a = a.matrix ,
      k = k.matrix,
      b = b,
      u = u,
      r.yhata.yhatb=r.yhata.yhatb,
      r.y.yhatb=r.y.yhatb,
      cov.a=cov.a,
      cor.a=cor.a)
 } ## End of Fungible

##--------------------##
##   EXAMPLE
##  GRE/GPA Data
##--------------------##

R.X <- matrix(c(1.00,  .56,  .77,
                 .56, 1.00,  .73,
                 .77,  .73, 1.00),3,3)

rxy <- c(.39, .34, .38)


b <- solve(R.X)%*%rxy
theta <- .01
OLSRSQ <- t(b)%*%R.X%*%b

r.yhata.yhatb <- sqrt( 1 - (theta)/OLSRSQ)

Nsets <- 5000


output <- Fungible(R.X, rxy, r.yhata.yhatb, sets=Nsets, print=TRUE)
```

References

Abadir, K.M., & Magnus, J.R. (2002). Notation in econometrics: A proposal for a standard. *Econometrics Journal*, *5*, 76–90.

Achen, C. (1982). *Interpreting and using regression*. Beverly Hills, CA: Sage Publications.

Bring, J. (1994). How to standardize regression coefficients. *American Statistician*, *48*, 209–213.
Budescu, D.V. (1993). Dominance analysis: A new approach to the problem of relative importance of predictors in multiple regression. *Psychological Bulletin*, *114*, 542–551.
Carroll, J.D., & Green, P.E. (1997). *Mathematical tools for applied multivariate analysis*. San Diego, CA: Academic Press.
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.
Darlington, R.B. (1968). Multiple regression in psychological research and practice. *Psychological Bulletin*, *69*, 161–182.
Darlington, R.B. (1990). *Regression and linear models*. New York: McGraw-Hill.
Dawes, R.M., & Corrigan, B. (1974). Linear models in decision making. *Psychological Bulletin*, *81*, 95–106.
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*, 329–333.
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.
Green, P.E., Carroll, J.D., & DeSarbo, W.S. (1978). A new measure of predictor importance in multiple regression. *Journal of Marketing Research*, *15*, 356–360.
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.
Grömping, U. (2007). Estimators of relative importance in linear regression based on variance decomposition. *The American Statistician*, *61*, 139–147.
Grove, W.M. (2001). Correction and extension of Wainer's "Estimating coefficients in linear models: It don't make no nevermind." Unpublished manuscript.
Guttman, L. (1955). The determinancy of factor score matrices with implications for five other basic problems of common factor theory. *British Journal of Statistical Psychology*, *8*, 65–81.
Hoerl, A.E., & Kennard, R.W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, *12*, 55–67.
Johnson, J.W. (2000). A heuristic method for estimating the relative weight of predictor variables in multiple regression. *Multivariate Behavioral Research*, *35*, 1–19.
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.
Kim, J.O., & Ferree, G. (1981). Standardization in causal analysis. *Sociological Methods and Research*, *10*, 187–210.
King, G. (1986). How not to lie with statistics: Avoiding common mistakes in quantitative political science. *American Journal of Political Science*, *30*, 666–687.
Koopman, R.F. (1988). On the sensitivity of a composite to its weights. *Psychometrika*, *53*, 547–552.
Kuncel, N.R., Hezlett, S.A., & Ones, D.S. (2001). A comprehensive meta-analysis of the predictive validity of the graduate record examinations: Implications for graduate student selection and performance. *Psychological Bulletin*, *127*, 162–181.
Laughlin, J.E. (1978). Comment on "Estimating coefficients in linear models: It don't make no nevermind". *Psychological Bulletin*, *85*, 247–253.
Mulaik, S.A. (1972). *The foundations of factor analysis*. New York: McGraw-Hill.
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.
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.
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.
Rencher, A.C. (2000). *Linear models in statistics*. New York: John Wiley & Sons.
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.
Tate, R.L., & Bryant, J.L. (1986). Parameter sensitivity for discriminant analysis. *Multivariate Behavioral Research*, *21*, 411–427.
Tatsuoka, M.M. (1971). *Multivariate analysis in educational and psychological research*. New York: Wiley.
Tukey, J.W. (1948). Approximate weights. *Annals of Mathematical Statistics*, *19*, 91–92.
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.
Wilks, S.S. (1938). Weighting schemes for linear functions of correlated variables when there is no dependent variable. *Psychometrika*, *3*, 23–40.

Manuscript Received: 12 JUL 2007
Final Version Received: 12 MAR 2008
Published Online Date: 23 JUL 2008