Contribution of individual variables to the regression sum of squares

Contribution of Individual Variables to the Regression Sum of Squares

Zillur R Shabuz[a,b] and Paul H Garthwaite[a*]

[a] *The Open University, Milton Keynes, MK7 6AA,UK*

[b] *University of Dhaka, Dhaka, Bangladesh*

Abstract

In applications of multiple regression, one of the most common goals is to measure the relative importance of each predictor variable. If the predictors are uncorrelated, quantification of relative importance is simple and unique. However, in practice, predictor variables are typically correlated and there is no unique measure of a predictor variable's relative importance. Using a transformation to orthogonality, new measures are constructed for evaluating the contribution of individual variables to a regression sum of squares. The transformation yields an orthogonal approximation of the columns of the predictor scores matrix and it maximizes the sum of the covariances between the cross-product of individual regressors and the response variable and the cross-product of the transformed orthogonal regressors and the response variable. The new measures are compared with three previously proposed measures through examples and the properties of the measures are examined.

Keywords: Dominance analysis; Orthogonal counterparts; Relative importance; Relative weights; Rotation invariance; Transformation to orthogonality

*Corresponding author. e-mail: paul.garthwaite@open.ac.uk;

Address: School of Mathematics and Statistics, Open University, Milton Keynes, MK7 6AA,UK

Tel: +44 1908 655974

1. Introduction

An important question that statistical consultants and researches commonly face after conducting a multiple regression analysis is which variable contributes most to predict or explain the criterion variable. For example, a chemist may raise the question of the relative importance of temperature and concentration in determining the rate of reaction. The term *importance* is recognized in the literature as having various possible meanings. A predictor may be considered important if the corresponding regression parameter is statistically significant. A second definition judges a predictor as more or less important on the basis of its practical impact on the response. It has been argued that the question of relative importance is even more common than the question of statistical significance (e.g. Healy (1990)).

Numerous methods have been proposed for evaluating the relative importance of regressors. The two most obvious methods are the beta weight method, which simply looks at the beta coefficients of variables that have been standardized to have variances of one, and the zero-order correlation method, which looks at the correlation between individual variables and the response. General statistical packages automatically include these statistics in their output from a regression or correlation analysis, making them convenient to use. However, as noted by Lipovetsky and Conklin (2015), multicollinearity can make these measures practically meaningless since, for example, high collinearity can change signs and inflate the values of regression coefficients in comparison with pair correlations between the regressors and response. Other evaluation methods include product measures (Pratt, 1987), usefulness (Darlington, 1968), structure coefficients (Courville and Thompson, 2001), dominance analysis (Budescu, 1993), orthogonal counterparts (Gibson, 1962; Johnson, 1966), relative weight analysis (Johnson, 2000), Shapley value regression (Lipovetsky and Conklin, 2001) and random forests (Liakhovitski et al., 2010). When predictors are uncorrelated, these measures lead to the same result and have the desirable property that their measures of the individual contributions of the predictor variables sum to R^2, the proportion of the variation in the response that the regressors explain. However, the measures can give quite different results for correlated regressors.

Reviews of work on relative importance are given in Johnson and LeBreton (2004), Nathans et al. (2012), Grömping (2007), Lipovetsky and Conklin (2010) and Stadler et al. (2017), and a good older overview is given by Darlington (1968). As noted by Johnson and LeBreton (2004), there is no unique solution to the problem of evaluating relative importance, so identifying good measures must be based on the logic behind their development, their properties and shortcomings, and the apparent sensibility of the results they yield.

In this paper we develop new measures of relative importance and compare them with well-regarded alternatives. The new measures are based on transformations that yield orthogonal variables that are closely related to the original regressors. In consequence they have much in common with the orthogonal counterparts measure proposed by Gibson (1962) and the relative weights measure of Johnson (2000). The main difference is that the new measures use the values of both the regressors *and* the response in determining the transformation, while the measures of Gibson (1962) and Johnson (2000) ignore the response when determining the transformation and use only the values of the regressors. Intuitively, there should be benefits in letting the response influence the transformation, as the purpose of the transformation is to help evaluate the relationship between regressors and the response.

The new measures proposed here are compared with the orthogonal counterparts measure (Gibson, 1962), and the relative weights measure (Johnson, 2000) and also with the dominance analysis measure proposed by Budescu (1993). The relative weights measure and dominance analysis are widely recommended procedures for estimating the relative importance of predictor variables (Tonidandel and LeBreton, 2010; Nathans et al., 2012). Comparison is made through examples and by examining theoretical properties.

In Section 2 we describe the measures of Gibson (1962), Johnson (2000) and Budescu (1993) and add insights into these measures. In Section 3 we describe the new measures and in Section 4 they are compared with other measures. Some of the measures have a *rotation invariance property*, whereby an orthogonal rotation can be applied to some variables without affecting the relative weights assigned to un-rotated variables. The rotation invariance property is described in Appendix A. Concluding comments are given in Section 5.

2. Three Measures of Relative Importance

In this section we describe the orthogonal counterparts measure, the relative weights measure and the dominance analysis measure. The orthogonal counterparts measure and the relative weights measure each form the basis of new measures that we propose in Section 3.

We assume that the response, Y, and regressors $X_1, \ldots, X_k$ are related through the regression equation

$$Y|\,\boldsymbol{X} = \beta_0 + \beta_1 X_1 + \ldots + \beta_k X_k + \epsilon, \tag{1}$$

where $\boldsymbol{X} = (X_1, \ldots, X_k)^T$ and ϵ is random error and has variance σ^2. We suppose there are n data, so that the model can be written in matrix form as:

$$\boldsymbol{y}|\,\mathbf{X} = \beta_{\mathbf{0}}\mathbf{1} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{2}$$

where $\mathbf{1}$ is an $n \times 1$ vector of 1's, $\boldsymbol{y}$ is an $n \times 1$ vector of responses, $\mathbf{X} = (\boldsymbol{x}_{\mathbf{1}}, \ldots, \boldsymbol{x}_{\mathbf{k}})$ is an $n \times k$ matrix of known values of $X_1, \ldots, X_k$, $\boldsymbol{\beta}$ is a $k \times 1$ vector of regression coefficients (whose values are unknown) and $\boldsymbol{\epsilon}$ is an $n \times 1$ vector of independent random errors. The coefficient β_0 is irrelevant for regressors' relative importance, so, to simplify notation, throughout this article we assume that Y and $X_1, \ldots, X_k$ have been centered to have sample means of 0. Then the least squares estimate of $\boldsymbol{\beta}$ is $\widehat{\boldsymbol{\beta}} = (\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-\mathbf{1}}\mathbf{X}^{\mathbf{T}}\boldsymbol{y}$ and $\mathrm{var}(\widehat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-\mathbf{1}}$.

2.1 Orthogonal Counterparts (OC) measure

Gibson (1962) and R. M. Johnson (1966) suggested a method for obtaining a set of orthonormal predictors that are closely related on a one-to-one basis with the original set of predictors. The new predictors can be considered as 'orthogonal counterparts' to the original regressors. To approximate the relative importance of the original predictors, the response variable is regressed on the new orthonormal variables. The proportion of the predictable variance in the response that is accounted for by each orthogonal counterpart can be taken as the importance measure of the original regressors. Details are as follows.

Suppose a set of orthonormal $n \times 1$ vectors $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_k$ must be chosen to

$$\text{maximise } \sum_{i=1}^{k} \boldsymbol{x}_i^T \boldsymbol{z}_i \tag{3}$$

and let $\mathbf{A}$ is the symmetric square-root of $\mathbf{X}^{\mathbf{T}}\mathbf{X}$. (That is, $\mathbf{A}$ is a symmetric matrix whose diagonal elements are positive and $\mathbf{AA} = \mathbf{X}^{\mathbf{T}}\mathbf{X}$.) Then, putting

$\mathbf{Z} = (\boldsymbol{z_1}, \ldots, \boldsymbol{z_k})$, it can be shown [see, for example, Garthwaite et al. (2012)] that

$$\mathbf{Z} = \mathbf{X}\mathbf{A}^{-\mathbf{1}}. \tag{4}$$

Each column of $\mathbf{Z}$ has a mean of 0 since each column of $\mathbf{X}$ has a mean of 0.

Gibson (1962) and Johnson (1966) assume that the X_i have been standardised to each have the same sample variance, when the maximisation in (3) is equivalent to

$$\text{maximise} \sum_{i=1}^{k} \text{cor}(X_i, Z_i), \tag{5}$$

where 'cor' denotes sample correlation, and so it is also equivalent to

$$\text{minimise} \sum_{i=1}^{k} \tilde{\boldsymbol{\epsilon}}_i^T \tilde{\boldsymbol{\epsilon}}_i, \tag{6}$$

where $\tilde{\boldsymbol{\epsilon}}_i$ is the residual when X_i is regressed on a single predictor with sample values $\boldsymbol{z}_i$. Based on (5), Gibson (1962) describes $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_k$ as *"the set of orthogonal factors ... having the highest degree of one-to-one correspondence with the correlated predictors."* Based on (6), $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_k$ are the best-fitting orthogonal representation of $\mathbf{X}$ (Johnson, 2000) and are termed the 'orthogonal counterparts' of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$. .

In the remainder of this paper we will assume that Y and each X variable have been standardised to have unit length. That is, $\boldsymbol{y}^T\boldsymbol{y} = \boldsymbol{x}_i^T\boldsymbol{x}_i = 1$ for $i = 1, \ldots, k$. Let $\widehat{\boldsymbol{\beta}}_Z = (\widehat{\beta}_{z1}, \ldots, \widehat{\beta}_{zk})^T$ denote the vector of regression coefficients from regressing Y on $\boldsymbol{Z}$, so $\widehat{\boldsymbol{\beta}}_Z = (\mathbf{Z}^\mathbf{T}\mathbf{Z})^{-\mathbf{1}}\mathbf{Z}^\mathbf{T}\boldsymbol{y} = \mathbf{Z}^\mathbf{T}\boldsymbol{y}$. Then $\widehat{\beta}_{zi}$ is called the beta weight of Z_i ($i = 1, \ldots, k$) and the squared beta weight, $\widehat{\beta}_{zi}^2$, is the variation in Y that is explained by Z_i. Hence the squared beta weights are a natural measure of the relative importance of the Z variables. Each Z variable is paired with an X variable, and the Orthogonal Components (OC) measure takes these squared beta weights as a measure of the importance of the X variables, defining the relative importance of X_i as $\widehat{\beta}_{zi}^2$. The sum of these importance weights equals the variation in Y that is explained by a multiple linear regression with $X_1, \ldots, X_k$ as the independent variables (or, equivalently, with $Z_1, \ldots, Z_k$ as the independent variables).

J. W. Johnson (2000) argues that the OC measure can assign relative weights that are inappropriate when the original X variables are highly correlated, and gives examples where some variables are assigned weights that seem too low. However, the OC measure appears to give sensible weights to the X variables when the correlations between variables are not high. Also, recent work by Garthwaite and Koch (2016) implies that the OC measure has an attractive 'rotation invariance' property. When some variables have strong collinearities, they can be transformed into non-collinear variables via orthogonal rotation of coordinate axes. Only axes corresponding to variables involved in the collinearities need to be rotated, and Garthwaite and Koch (2016) show that the rotation has no effect on the Z-variables that correspond to un-rotated axes. The predictable variation in Y is also unaffected by the rotation, so the OC measure has the property that the relative importance is unchanged for those X variables associated with un-rotated axes. (Further detail is given in Appendix A.) This has the following implications for the OC measure.

- Sometimes collinear variables can be transformed into meaningful variables that are not collinear through a rotation of the axes associated with them. This can lead to relative weights that are a transparently reasonable representation of the importance of the different variates. Moreover, the relative weights are unchanged for those variables that are not involved in the rotation.

- Since axes *could* be rotated to remove collinearities without affecting the relative weights of the other variables, *multicollinearities do not affect the relative weights that the OC measure gives to variables not involved in the collinearities.*

Garthwaite et al. (2012) suggest a criterion for choosing $\boldsymbol{z}_1, \dots, \boldsymbol{z}_k$ that is similar, but not identical, to the criterion in (3): Choose $\tilde{\boldsymbol{z}}_1, \dots, \tilde{\boldsymbol{z}}_k$ to

$$\text{maximise} \sum_{i=1}^{k} (\boldsymbol{x}_i^T \tilde{\boldsymbol{z}}_i)^2 \tag{7}$$

under the constraint that they are orthonormal vectors and $\boldsymbol{x}_i^T \tilde{\boldsymbol{z}}_i > 0$ for $i = 1, \dots, k$. They refer to the transformation from $\boldsymbol{x}_1, \dots, \boldsymbol{x}_k$ to the resulting $\tilde{\boldsymbol{z}}_1, \dots, \tilde{\boldsymbol{z}}_k$ as the cos-square transformation. It has an attractive *duplicate invariance property.* Suppose the set of vectors $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_i\}$ is increased by adding the set of vectors $\{\boldsymbol{x}_{i+1}, \dots, \boldsymbol{x}_k\}$ where each of the vectors $\boldsymbol{x}_{i+1}, \dots, \boldsymbol{x}_k$ is identical to $\boldsymbol{x}_i$. With the cos-square transformation, this duplication of $\boldsymbol{x}_i$ has no effect on the transformed values of $\boldsymbol{x}_1, \dots, \boldsymbol{x}_{i-1}$ (i.e $\tilde{\boldsymbol{z}}_1, \dots, \tilde{\boldsymbol{z}}_{i-1}$ are unchanged.)

Thus, for example, if X_1 and X_2 are measurements of, say, a patients blood pressure before and after a meal, then the two variables will be very highly correlated.

The duplicate invariance property means that whether one or both blood pressure measurements are included in the regression model has little impact on those Z variables that are paired with the other X variables. As the orthogonal Z variables that maximize (3) will generally be very similar to those that maximize (7), we might expect the OC measure to usually be fairly insensitive to variable duplication.

2.2 Relative Weights (RW) measure

The Relative Weights (RW) measure of J. W. Johnson (2000) is based on the same Z variables that are calculated for the OC measure. That is, $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_k$ are the orthonormal variables that maximize $\sum_{i=1}^{k} \boldsymbol{x}_i^T \boldsymbol{z}_i$ and, as they are orthogonal, the relative importance of Z_i in predicting Y is clearly $\widehat{\beta}_{zi}^2$. However, while the OC measure simply takes $\widehat{\beta}_{zi}^2$ as the relative importance of X_i, the measure of Johnson (2000) takes into account all the correlations between the X and Z variables. From the criterion that determines the Z variables, the correlation between X_i and Z_i should be high, but this correlation could still be well below 1 if the X variables display collinearities or high intercorrelations. Also, X_i might not be the only X variable that has a marked correlation with Z_i.

Let λ_{ij} denote the correlation between X_i and Z_j. The transformation from X to Z has the unexpected property that $\lambda_{ij} = \lambda_{ji}$ for all i, j (see, for example, Johnson (1966)). This leads to the useful consequence that $\sum_{i=1}^{k} \lambda_{ij}^2 = \sum_{j=1}^{k} \lambda_{ij}^2 = 1$. The RW measure divides the relative importance of Z_j amongst the X variables according to the square of their correlations with Z_j, so the relative importance weight that X_i derives from Z_j is $\lambda_{ij}^2 \widehat{\beta}_{zj}^2$. (This indeed partitions the relative importance of Z_j, as $\sum_{i=1}^{k} \lambda_{ij}^2 \widehat{\beta}_{zj}^2 = \widehat{\beta}_{zj}^2$.) The full relative importance weight of X_i is obtained by summing the relative importance weights that it derives from all the Z variables. Thus, under the RW measure, the relative importance of X_i is given by

$$\text{RW of } X_i = \sum_{j=1}^{k} \lambda_{ij}^2 \widehat{\beta}_{zj}^2. \tag{8}$$

The λ_{ij}^2 in equation (8) may be regarded as the squares of regression coefficients rather than the squares of correlation coefficients, as

$$\mathrm{E}(X_i \mid \boldsymbol{Z}) = \lambda_{i1} Z_1 + \cdots + \lambda_{ik} Z_k \tag{9}$$

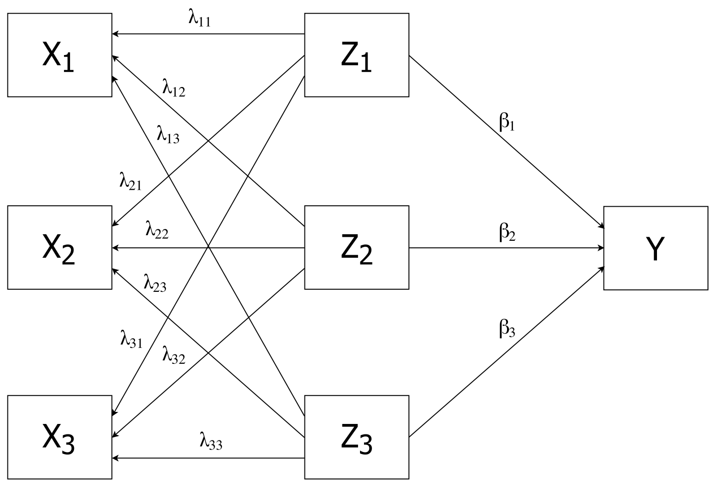

Figure 1 Relationships between the X, Z and Y variables for three regressors when Y is regressed on the Z variables and each X variable is regressed on the Z variables

when X_i is regressed on $Z_1, \ldots, Z_k$. When Johnson (2000) proposed the RW measure he used the regression model in equation (9) to motivate its construction. However, we prefer to view the λ_{ij}^2 as squared correlations because correlation is a symmetric relationship while the regression in equation (9) is a one-directional relationship and shows how the Z variables determine X_i. When viewed as a regression, the relationships between the X, Z and Y variables is illustrated in Figure 1 and shows no direct link between the X and Y variables. When the λ_{ij}^2 are viewed as squared correlations, the links between the X and Z variables are two-directional associations, thus giving links from the X variables to Y.

Applications in which the RW measure has been used are reported in Johnson and LeBreton (2004) and Krasikova et al. (2011). Part of the attraction of the RW measure is that it typically gives similar results to the dominance analysis measure of Budescu (1993), even though Budescu's measure and the RW measure are calculated in very different ways. As Johnson (2000, p.15) suggests, "it is encouraging that two measures that have very different definitions and calculations produce very similar solutions", and Johnson and LeBreton (2004, p.251) argue that the closeness of results indicates that the two measures are measuring the same construct. We next describe the dominance analysis measure.

2.3 Dominance Analysis (DA) measure

The Dominance Analysis (DA) measure evaluates the importance of a regressor X_i by considering the increase in R^2 that results from adding X_i to regression submodels, where R^2 is the proportion of the variation in Y that is explained by the regression. For correlated regressors, the increase in R^2 will generally depend upon which regressors are in the submodel before X_i is added. The DA measure considers all the different submodels that could be formed from every possible subset of X variables that excludes X_i. It defines the weight (relative importance) of X_i as the average increase in R^2 from adding X_i to each of these submodels.

The DA measure was proposed by Budescu (1993) and is sometimes referred to as the 'general dominance measure'. It is equivalent to a measure developed by Lindeman et al. (1980). The measure is well-regarded. For example, Johnson (2000, p.4) writes that "The average increase in R^2 associated with the presence of a variable across all possible models is a meaningful measure that fits the definition of relative weight". At the same time, it could be argued that the importance of a regressor in a particular regression model should not be determined by its importance in smaller submodels, but by its importance in the full model. As noted earlier, the DA measure and the RW measure are generally very similar in the weights they assign to variables, although an example in Section 4 illustrates that this is not always the case.

The most commonly stated criticism of the DA measure is that it is computationally demanding. This is because there are $(2^k - 1)$ regression models that should be fitted in order to evaluate the relative importance of each variable. When Lindeman et al. proposed his relative importance measure in 1980, fitting models with all combinations of variables was only practical when the number of variables was fewer than 5 or 6. Since then, advances in computer power has substantially increased that number, so currently it takes only 0.28 seconds to fit all submodels of 12 regressors using software developed by Grömping et al. (2006). However, it was not possible to use Gromping's software with models containing 25 regressors and with 20 regressors the general dominance measure could not be calculated. Moreover, the number of variables that are included in regression models has increased substantially, especially with interest in 'big data'. One possibility is to examine a sample of submodels rather than examining all possible submodels. This approach has proved effective when using Shapley value regression to measure relative importance (Conklin et al., 2004), another measure where, in principle, all possible submodels should be examined. Simulations we have conducted suggest that the DA measure can be well-approximated by examining 500 random sequences for entering variables into the regression model.

3 New Measures of Relative Importance

Three new measures are proposed here. All are based on transformations that yield orthogonal variables – the first and third are very similar to the OC measure of Gibson (1962) and R. M. Johnson (1966); the second is very similar to the RW measure of J. W. Johnson (2000). The main difference is that the new measures use transformations that are determined by cross-products of the X and Y variables, rather than ignoring Y in choosing the transformation. The third new measure uses weights to alter the balance of the different cross-products when forming orthogonal variables.

The estimated regression coefficient $\hat{\boldsymbol{\beta}}$ for regressing Y on $\boldsymbol{X} = (X_1, \ldots, X_k)^T$ is

$$\hat{\boldsymbol{\beta}} = (\mathbf{X^T X})^{\mathbf{-1}}\mathbf{X^T}\boldsymbol{y} \tag{10}$$

and the regression sum of squares (RegSS) is

$$\boldsymbol{y}^T\mathbf{X}(\boldsymbol{X}^\mathbf{T}\mathbf{X})^{\mathbf{-1}}\mathbf{X^T}\boldsymbol{y}. \tag{11}$$

Let $(y_1, \ldots, y_n)^T = \boldsymbol{y}$ and let $\mathbf{Y}$ be an $n \times n$ diagonal matrix with diagonal elements $y_1, \ldots, y_n$. The RegSS can also be rewritten as:

$$\mathbf{1}^T\mathbf{Y}\mathbf{X^T}(\mathbf{X^T X})^{-1}\mathbf{X^T}\mathbf{Y}\mathbf{1} \tag{12}$$

where $\mathbf{1}$ is a vector of ones.

Both the OC and RW measures construct orthogonal vectors $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_k$ that corresponds closely to the original predictors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ on a one-to-one basis. The way the $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_k$ are chosen ignores the values of Y, even though the reason for constructing $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_k$ is to partition the RegSS. With our new measures, a set of orthogonal vectors $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_k$ is chosen so that $\mathbf{Y}\boldsymbol{w}_\mathbf{i}$ is closely related to $\mathbf{Y}\boldsymbol{x}_\mathbf{i}$. Suppose Y is regressed on $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_k$ and that w_{ij} is the jth component of $\boldsymbol{w}_i$. Then $\boldsymbol{w}_i$'s contribution to the RegSS from the jth sample is $(y_j w_{ij})^2$. Our new measures take $(y_j w_{ij})^2$ as a first estimate of the contribution of X_i to the RegSS from the jth sample. (The OC and RW measures equivalently take $(y_j z_{ij})^2$ as a first estimate of X_i's contribution to the RegSS from the jth sample, where z_{ij} is the jth component of $\boldsymbol{z}_i$.) Hence, as $y_j w_{ij}$ is the jth component of $\mathbf{Y}\boldsymbol{w}_\mathbf{i}$, it is appropriate to focus on $\mathbf{Y}\boldsymbol{w}_\mathbf{1}, \ldots, \mathbf{Y}\boldsymbol{w}_\mathbf{k}$ in the criterion for choosing $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_k$. It is for this reason that we want $\mathbf{Y}\boldsymbol{w}_\mathbf{i}$ and $\mathbf{Y}\boldsymbol{x}_\mathbf{i}$ to be closely related.

We also want $\mathbf{W} = (\boldsymbol{w}_\mathbf{1}, \ldots, \boldsymbol{w}_\mathbf{k})$ to be a linear transformation of $\mathbf{X} = (\boldsymbol{x}_\mathbf{1}, \ldots, \boldsymbol{x}_\mathbf{k})$, so that regression models with $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_k$ as explanatory variables and with $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ as explanatory variables give identical predictions, residuals and regression sums of squares. Hence, analogous to equation (3), we choose $\mathbf{W}$ so that $\sum_{i=1}^k (\mathbf{Y}\boldsymbol{w}_\mathbf{i})^\mathbf{T}(\mathbf{Y}\boldsymbol{x}_\mathbf{i})$ is maximized subject to the constraints that $\mathbf{W}^\mathbf{T}\mathbf{W} = \mathbf{I}_\mathbf{k}$ and $\mathbf{W} = \mathbf{XC}$ for some $k \times k$ non-singular matrix $\mathbf{C}$. The following result is proved in Appendix B.

Theorem 1. *Under the constraints that* $\mathbf{W} = \mathbf{XC}$ *and* $\mathbf{W}^\mathbf{T}\mathbf{W} = \mathbf{I}_\mathbf{k}$*, the value of* $\mathbf{W} = (\boldsymbol{w}_\mathbf{1}, \ldots, \boldsymbol{w}_\mathbf{k})$ *that maximizes* $\sum_{i=1}^k (\mathbf{Y}\boldsymbol{w}_\mathbf{i})^\mathbf{T}(\mathbf{Y}\boldsymbol{x}_\mathbf{i})$ *is*

$$\mathbf{W} = \mathbf{X}(\mathbf{X}^\mathbf{T}\mathbf{X})^{-\mathbf{1/2}}\mathbf{G}, \tag{13}$$

where

$$\mathbf{G} = \boldsymbol{\Psi}(\boldsymbol{\Psi}^\mathbf{T}\boldsymbol{\Psi})^{-\mathbf{1/2}} \tag{14}$$

and

$$\boldsymbol{\Psi} = (\mathbf{X}^\mathbf{T}\mathbf{X})^{-\mathbf{1/2}}\mathbf{X}^\mathbf{T}\mathbf{YYX}. \tag{15}$$

We should note that the X variables are standardised but $||\mathbf{Y}\boldsymbol{x}_\mathbf{i}||$ typically varies with i. Hence the X variables are given equal importance in maximising $\sum_{i=1}^k \boldsymbol{x}_i^T \boldsymbol{z}_i$ (as with the OC and RW measures) but here, in maximising $\sum_{i=1}^k (\mathbf{Y}\boldsymbol{w}_\mathbf{i})^\mathbf{T}(\mathbf{Y}\boldsymbol{x}_\mathbf{i})$, X_i is given greater importance when $||\mathbf{Y}\boldsymbol{x}_\mathbf{i}||$ is large than when it is small. This has the benefit that those X variables that are most highly correlated with Y are given greater weight when choosing the $\boldsymbol{w}_i$. (We could scale the X variables so that $||\mathbf{Y}\boldsymbol{x}_\mathbf{i}||$ is the same for each X_i, but that would lose this benefit.)

3.1 First new measure (NM1)

In the same way that the OC measure views $\boldsymbol{z}_i$ as the counterpart of $\boldsymbol{x}_i$ $(i = 1, \ldots, k)$, our first New Measure (NM1) views $\mathbf{Y}\boldsymbol{w}_\mathbf{i}$ as the counterpart of $\mathbf{Y}\boldsymbol{x}_\mathbf{i}$ $(i = 1, \ldots, k)$. The RegSS when Y is regressed on $\boldsymbol{w}_i$ is $\{\mathbf{1}^T\mathbf{Y}\boldsymbol{w}_\mathbf{i}\}^\mathbf{2} = (\boldsymbol{y}^\mathbf{T}\boldsymbol{w}_\mathbf{i})^\mathbf{2}$. As $\{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_k\}$

are a set of orthonormal vectors, $(\boldsymbol{y}^T\boldsymbol{w}_i)^2$ is the RegSS both when Y is regressed on $\boldsymbol{w}_1, \ldots, \boldsymbol{w}_k$ and when Y is regressed on $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$. NM1 defines the relative importance of X_i as

$$\text{NM1: Relative importance of } X_i = (\boldsymbol{y}^T\boldsymbol{w}_i)^2. \tag{16}$$

Like the OC measure, NM1 has a rotation invariance property. Specifically, if an orthogonal rotation is applied to some of the X variables, the relative importance of the other X variables is unchanged if relative importance is measured using NM1. This result is proved in Appendix A, where further detail of rotation invariance is given. As with the OC measure, it means that collinearities do not affect the relative importances that NM1 gives to variables not involved in the collinearities.

3.2 Second new measure (NM2)

While NM1 allocates all the RegSS of W_j to X_j, our second new method, NM2, divides the RegSS of W_j between the X variables according to their association with W_j. As $\mathbf{Z} = \mathbf{X}(\mathbf{X^TX})^{-\mathbf{1/2}}$, from equation (13) we have that $\mathbf{W} = \mathbf{ZG}$. (This is an attractive representation of $\mathbf{W}$ because $\mathbf{Z}$ is a set of orthonormal vectors and $\mathbf{G} = (\boldsymbol{g}_\mathbf{1}, \ldots, \boldsymbol{g}_\mathbf{k})$ is an orthogonal matrix.) Thus,

$$\boldsymbol{w}_j = \mathbf{Z}\boldsymbol{g}_\mathbf{j}, \tag{17}$$

As noted in Section 3, $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_k$ correspond closely to $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_k$ on a one-to-one basis, so $\boldsymbol{w}_j$ should generally be highly correlated with $\mathbf{X}\boldsymbol{g}_\mathbf{j}$. Also, as $\boldsymbol{g}_i$ and $\boldsymbol{g}_j$ are orthogonal for $i \neq j$, typically $\boldsymbol{w}_j$ will not be closely associated with $\mathbf{X}\boldsymbol{g}_\mathbf{i}$ for $i \neq j$.

NM2 divides the RegSS of W_j between $X_1, \ldots X_k$ to reflect the squares of the sample correlations between $\mathbf{X}\boldsymbol{g}_\mathbf{i}$ and $\boldsymbol{w}_j$ $(i = 1, \ldots, k)$. Let r_{ij} denote the sample correlation between $\mathbf{X}\boldsymbol{g}_\mathbf{i}$ and $\boldsymbol{w}_j$. It is readily shown that

$$r_{ij} = \frac{\boldsymbol{g}_i^T(\mathbf{X^TX})^{\mathbf{1/2}}\boldsymbol{g}_\mathbf{j}}{[\boldsymbol{g}_i^T\mathbf{X^TX}\boldsymbol{g}_\mathbf{i}]^{\mathbf{1/2}}}. \tag{18}$$

The proportion of W_j's RegSS that NM2 attributes to X_i is $r_{ij}^2 / \sum_{\ell=1}^k r_{\ell j}^2$, so NM2 defines the relative importance of X_i as:

$$\text{NM2: Relative importance of } X_i = \sum_{j=1}^{k} \frac{r_{ij}^2(\boldsymbol{y}^T\boldsymbol{w}_j)^2}{\sum_{\ell=1}^k r_{\ell j}^2}. \tag{19}$$

If X_i has low correlations with the other X variables, the NM1 and NM2 measures will give similar relative importance to X_i. However the relative importances that they assign to X_i can differ markedly if X_i is highly correlated with some of the X variables. This will be seen in Section 4.

3.3 Third new measure (NM3)

If an X variable has a small regression coefficient in the multiple regression of y on all the X variables, then dropping that variable from the regression model can be attractive. With the NM1 and NM2 measures (and also the OC and RW measures), the orthogonal counterparts of all variables can change markedly if any X variables are discarded, which might be undesirable in some situations. Our third new measure, NM3, takes account of the size of regression coefficients when forming orthogonal counterparts, so that the inclusion or exclusion of variables with small regression coefficients has little effect on the orthogonal counterparts of other variables.

As in equation (10), let $\hat{\boldsymbol{\beta}}$ denote the estimated regression coefficient for regressing Y on $\boldsymbol{X} = (X_1, \ldots, X_k)^T$ and put $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \ldots, \hat{\beta}_k)^T$. While NM1 and NM2 choose $\mathbf{W} = (\boldsymbol{w}_\mathbf{1}, \ldots, \boldsymbol{w}_\mathbf{k})$ to maximize $\sum_{i=1}^k (\mathbf{Y}\boldsymbol{w}_\mathbf{i})^\mathbf{T}(\mathbf{Y}\boldsymbol{x}_\mathbf{i})$, with NM3 we choose $\mathbf{W}^\# = (\boldsymbol{w}_\mathbf{1}^\#, \ldots, \boldsymbol{w}_\mathbf{k}^\#)$ to maximize $\sum_{i=1}^k |\hat{\beta}_i| \, (\mathbf{Y}\boldsymbol{w}_\mathbf{i}^\#)^\mathbf{T}(\mathbf{Y}\boldsymbol{x}_\mathbf{i})$. Thus, with NM3, the importance of the correlation between $(\mathbf{Y}\boldsymbol{w}_\mathbf{i}^\#)$ and $(\mathbf{Y}\boldsymbol{x}_\mathbf{i})$ depends upon the size of $\hat{\beta}_i$.

If we let $\boldsymbol{x}_i^\# = |\hat{\beta}_i| \, \boldsymbol{x}_i$, then $\sum_{i=1}^k |\hat{\beta}_i| \, (\mathbf{Y}\boldsymbol{w}_\mathbf{i}^\#)^\mathbf{T}(\mathbf{Y}\boldsymbol{x}_\mathbf{i}) = \sum_{\mathbf{i=1}}^\mathbf{k} (\mathbf{Y}\boldsymbol{w}_\mathbf{i}^\#)^\mathbf{T}(\mathbf{Y}\boldsymbol{x}_\mathbf{i}^\#)$, and the maximization problem is analogous to the maximisation problem addressed in Theorem 1. Put $\mathbf{X}^\# = (\boldsymbol{x}_\mathbf{1}^\#, \ldots, \boldsymbol{x}_\mathbf{k}^\#)$. Then replacing $\mathbf{W}$ with $\mathbf{W}^\#$ and $\mathbf{X}$ with $\mathbf{X}^\#$ in equations (13)-(15) yields the value of $\mathbf{W}^\#$ that maximizes $\sum_{i=1}^k |\hat{\beta}_i| \, (\mathbf{Y}\boldsymbol{w}_\mathbf{i}^\#)^\mathbf{T}(\mathbf{Y}\boldsymbol{x}_\mathbf{i})$.

NM3 views $(\mathbf{Y}\boldsymbol{w}_\mathbf{i}^\#)$ as the counterpart of $(\mathbf{Y}\boldsymbol{x}_\mathbf{i}^\#)$ $(i = 1, \ldots, k)$ and evaluates the relative importance of X_i as the value of R^2 when Y is regressed on $\boldsymbol{w}_i^\#$. Thus,

$$\text{NM3: Relative importance of } X_i = (\boldsymbol{y}^T \boldsymbol{w}_i^\#)^2. \tag{20}$$

The NM3 and the DA measures are the only measures we examine that explicitly use the multiple regression of Y on $\boldsymbol{X} = (X_1, \ldots, X_k)^T$. Using this regression model seems sensible, since the purpose of the measures is to evaluate the contribution of each variable to this regression.

4 Examples

In this section we apply the measures of relative importance to several datasets. In Section 4.1 we examine straightforward application of the measures, using three datasets that have clear structures. In Section 4.2 we examine how relative importance changes under orthogonal rotation of some variables and under variable selection.

4.1 Fixed models

Each dataset consists of 1000 data drawn from a multivariate normal distribution with a mean vector of zeros and variance-covariance matrix $\mathbf{\Sigma}$, where $\mathbf{\Sigma}$ varies with the dataset. The first component of a datum is the response, y, and the other components are the explanatory variables, $x_1, \ldots, x_k$. We first describe each dataset by giving the sample correlation matrix $\widehat{\mathbf{R}}$, the multiple regression model that relates Y to the explanatory variables, the value of R^2 for that regression, and the regression coefficients for univariate regressions when Y is regressed separately on one x-variable at a time. We also note salient features of the dataset. After this brief description of the datasets, we tabulate the relative importance that the different measure allocate to each variable. The results are then discussed.

Example 1.

In this first dataset, Y correlates highly with X_1 and its correlations with X_2 and X_3 are much lower. Also, X_1 has much the biggest regression coefficient in a multiple regression of Y on X_1, X_2. There is marked correlation between the X variables.

The sample correlation matrix is

$$\begin{array}{cc} & \begin{array}{cccc} Y & X_1 & X_2 & X_3 \end{array} \\ \widehat{\mathbf{R}} = & \left(\begin{array}{cccc} 1.000 & 0.847 & 0.419 & 0.382 \\ 0.847 & 1.000 & 0.701 & 0.697 \\ 0.419 & 0.701 & 1.000 & 0.483 \\ 0.382 & 0.697 & 0.483 & 1.000 \end{array}\right) \begin{array}{c} Y \\ X_1 \\ X_2 \\ X_3 \end{array} \end{array}$$

The fitted standardized multiple regression model is:

$$\hat{Y} = 1.380X_1 - 0.351X_2 - 0.411X_3 \qquad (R^2 = 0.865)$$

and the univariate regression models are

$$\hat{Y} = 0.847X_1, \quad \hat{Y} = 0.419X_2, \quad \text{and } \hat{Y} = 0.382X_3.$$

Example 2.

There are just two explanatory variables in this dataset. The Y variable is highly correlated with X_1 but uncorrelated with Z_1. Together, X_1 and X_2 give a multiple regression equation that predicts Y perfectly.

The sample correlation matrix is

$$\widehat{\mathbf{R}} = \begin{array}{c} \\ \\ \\ \end{array}\begin{array}{ccc} Y & X_1 & X_2 \\ \end{array}$$

$$\widehat{\mathbf{R}} = \begin{pmatrix} 1.000 & 0.893 & 0.450 \\ 0.893 & 1.000 & 0.803 \\ 0.450 & 0.803 & 1.000 \end{pmatrix} \begin{array}{c} Y \\ X_1 \\ X_2 \end{array}$$

The fitted standardized multiple regression model is:

$$\hat{Y} = 1.499X_1 - 0.755X_2 \qquad (R^2 = 1.00)$$

and the univariate regression models are

$$\hat{Y} = 0.893X_1, \quad \text{and } \hat{Y} = 0.450X_2.$$

Example 3.

In this example, Y is almost a perfect linear function of the last five X variables $(X_2, \ldots, X_6)$ while Y is more highly correlated with X_1 than the other X variables. Also, the largest correlations between the X variables are the correlations involving X_1.

The sample correlation matrix is

$$\begin{array}{ccccccc} Y & X_1 & X_2 & X_3 & X_4 & X_5 & X_6 \end{array}$$

$$\widehat{\mathbf{R}} = \begin{pmatrix} 1.000 & 0.805 & 0.681 & 0.669 & 0.702 & 0.702 & 0.698 \\ 0.805 & 1.000 & 0.581 & 0.572 & 0.601 & 0.605 & 0.598 \\ 0.681 & 0.581 & 1.000 & 0.352 & 0.361 & 0.392 & 0.386 \\ 0.669 & 0.572 & 0.352 & 1.000 & 0.372 & 0.384 & 0.365 \\ 0.702 & 0.601 & 0.361 & 0.372 & 1.000 & 0.428 & 0.422 \\ 0.702 & 0.605 & 0.392 & 0.384 & 0.428 & 1.000 & 0.400 \\ 0.698 & 0.598 & 0.386 & 0.365 & 0.422 & 0.400 & 1.000 \end{pmatrix} \begin{array}{c} Y \\ X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \\ X_6 \end{array}$$

The fitted standardized multiple regression model is:

$$\hat{Y} = 0.010X_1 + 0.277X_2 + 0.266X_3 + 0.272X_4 + 0.261X_5 + 0.269X_6 \qquad (R^2 = 0.936)$$

and the univariate regression models are

$$\hat{Y} = 0.805X_1, \quad \hat{Y} = 0.681X_2, \quad \hat{Y} = 0.669X_3$$
$$\hat{Y} = 0.702X_4, \quad \hat{Y} = 0.702X_5, \quad \text{and } \hat{Y} = 0.698X_6.$$

Results from the three examples.

The relative importance given to each variable by the six different measures are given for each example in Table 1. Advocates of the RW measure argue that one of its strengths is that it generally gives similar results to the DA measure. Table 1 shows that this was also the case for our examples, but the table shows that the NM2 measure also gave similar results to DA. Indeed, for Examples 1 and 3 the relative importances assigned by DA are a little closer to those of NM2 than to those of RW. The results of the other measures (OC, NM1 and NM3) are often fairly similar to each other, especially those of OC and NM3, as in Examples 1 and 2. At the same time, NM1 is notably similar to DA in example 3, while NM3 gives radically different results to all other measures in that example.

In each of the examples, at least one variable's contribution to predicting Y was small but it was correlated with variables that were better predictors. Then the variable's relative importance was generally higher when measured by RW, DA or NM2 than when measured by OC, NM1 or NM3. This can be seen in Example 1, where X_2 and X_3 are poor predictors, and in Example 2, where X_2 is a poor predictor. The NM1 and NM3 measures, though conceptually quite similar to the OC measure, can give evaluations that are clearly more sensible than those of the OC measure. This is illustrated in Example 2, where the OC measure evaluates the relative importance of X_1 as 100% and the relative importance of X_2 as 0%. This is inappropriate, since X_1 on its own cannot explain all the variation in Y, while the combination of X_1 and X_2 *can* explain all the variation in Y, clearly showing that X_2 contributes usefully to the multiple regression model. The NM1 and NM3 measures evaluate the contribution of X_2 as small, but non-zero. The larger values

Table 1: Relative importances given by the orthogonal counterparts (OC), relative weights (RW) and dominance analysis (DA) measures and by three new measures (NM1, NM2 and NM3) in Examples 1–3

	OC	RW	DA	NM1	NM2	NM3
Example 1						
X_1	0.856	0.642	0.665	0.764	0.669	0.864
X_2	0.008	0.115	0.101	0.061	0.105	0.000
X_3	0.002	0.108	0.099	0.040	0.091	0.001
Example 2						
X_1	1.000	0.798	0.798	0.930	0.830	0.989
X_2	0.000	0.202	0.202	0.070	0.170	0.011
Example 3						
X_1	0.124	0.137	0.170	0.172	0.159	0.001
X_2	0.163	0.160	0.152	0.143	0.147	0.184
X_3	0.155	0.153	0.145	0.146	0.149	0.177
X_4	0.167	0.164	0.158	0.160	0.162	0.199
X_5	0.163	0.160	0.155	0.160	0.161	0.187
X_6	0.165	0.162	0.156	0.155	0.157	0.188

given to X_2 by the RW, DA and NM2 measures are perhaps a better reflection of X_2's contribution, since on its own X_2 explains 20.3% of the variation in Y.

In Example 3, it is arguable whether X_1 is useful for predicting Y. On the one hand, X_1 makes little contribution to the multiple regression model while, on the other hand, it is the best univariate predictor of Y. NM3 gives X_1 a relative importance that is close to 0, which might be considered appropriate in view of the multiple regression model. Other measures give it a much higher relative importance; indeed, DA and NM1 evaluate it as the most important predictor which, to the writers, seems inappropriate. Example 3 also shows that the RW and DA measures are not always in close agrement: while DA evaluates X_1 as the most important variable in the regression model, RW evaluates it as the least important.

4.2 Orthogonal rotation and variable selection

Two examples are examined in this section. In the first, two of the explanatory variables are highly correlated and we consider both the model with the original variables and the model that results from rotating the correlated variables. Measures of relative importance are applied to both models and their differences are examined. In the second example, one variable has a regression coefficient that does not differ significantly from 0 (at the 5% level of significance). We examine how dropping this variable from the model effects the relative importances of the other variables.

Example 4. Orthogonal rotation

The Longley dataset (Longley, 1967) is well-used as an example of highly collinear regression. The dataset contains annual values of various US macroeconomic variables for the years 1947-1962. Here we use five of its variables: *npe* (number of thousands of people employed), GNP_1 (GNP price deflator), GNP_2 (GNP in millions of dollars), *npue* (number of thousands of unemployed people) and *npa* (number of people in the armed forces). We take *npe* as the response variable and initially take the other four variables as the explanatory variables.

The following is the sample correlation matrix for these variables:

$$\widehat{\mathbf{R}} = \begin{array}{c} \begin{array}{ccccc} npe & GNP_1 & GNP_2 & npue & npa \end{array} \\ \begin{pmatrix} 1.000 & 0.971 & 0.984 & 0.502 & 0.457 \\ 0.971 & 1.000 & 0.992 & 0.621 & 0.465 \\ 0.984 & 0.992 & 1.000 & 0.604 & 0.446 \\ 0.502 & 0.621 & 0.604 & 1.000 & -0.177 \\ 0.457 & 0.465 & 0.446 & -0.177 & 1.000 \end{pmatrix} \end{array} \begin{array}{c} \\ npe \\ GNP_1 \\ GNP_2 \\ npue \\ npa \end{array}$$

The fitted standardized multiple regression model is:

$$\widehat{npe} = 0.173GNP_1 + 0.998GNP_2 - 0.227npue - 0.109npa, \tag{21}$$

for which $R^2 = 0.986$.

The correlation matrix shows that there is a strong collinearity between two of the explanatory variables, GNP_1 and GNP_2. Collinearity can radically affect the values of parameter estimates and will inflate their variances. Transforming variables to remove collinearity is consequently attractive and here we replace GNP_1 and GNP_2 by the variables

$$X_1 = (GNP_1 + GNP_2)/\sqrt{2} \quad \text{and} \quad X_1 = (GNP_1 - GNP_2)/\sqrt{2}.$$

Table 2: Relative importances of variables before and after rotation

	OC	RW	DA	NM1	NM2	NM3
Relative importances before rotation						
GNP_1	0.400	0.390	0.390	0.527	0.361	0.088
GNP_2	0.526	0.417	0.411	0.371	0.393	0.888
npue	0.023	0.099	0.104	0.046	0.139	0.005
npa	0.036	0.079	0.081	0.042	0.092	0.004
Relative importances after rotation						
X_1	0.922	0.682	0.687	0.891	0.550	0.967
X_2	0.004	0.014	0.015	0.007	0.183	0.004
npue	0.023	0.161	0.156	0.046	0.146	0.004
npa	0.036	0.128	0.128	0.042	0.107	0.011

This is equivalent to multiplying the original variables by the orthogonal rotation matrix,

$$\mathbf{\Gamma} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

The new variables X_1 and X_2 are uncorrelated.

Regressing *npe* on the transformed set of variables gives the equation

$$\widehat{npe} = 1.681X_1 - 0.054X_2 - 0.227npue - 0.109npa. \tag{22}$$

Theory implies that the regression coefficients of the unrotated components (*npue* and *npa*) should be unchanged – comparison of equations (21) and (22) shows that this is indeed the case. Also, the R^2 value is again 0.986. However, with some measures of relative importance, the importances of *npue* and *npa* in the pre-rotation model (equation (21)) will differ from their importances in the post-rotation model (equation (22)). This can be seen in Table 2, where the relative importances given by our six measures of importance are presented.

In line with theory, the table shows that the relative importances given by the OC and NM1 measures to *npue* and *npa* are unchanged by the rotation of GNP_1

and GNP_2. With the other measures, the relative importances given to *npue* and *npa* do change, though the degree of change varies with the measure. With NM3 the importance values change by a large proportion (e.g. from 0.004 to 0.011), though the changes are small in absolute terms. With the RW and DA measures the changes are quite large - noticeably larger (at least three times larger) than with the NM2 measure. Interestingly, values given by the NM2 measure are straddled by the before/after values given by the RW and DA measures, and are quite close to the averages of the before/after values given by both the RW measure and the DA measure. For example, the RW measure gives before/after values of 0.079 and 0.128 to *npa*, and their average is relatively close to the values 0.092 and 0.107 that NM2 gives to *npa*.

Example 5. Variable selection

Wood (1973) presents data from a process variable study of a petroleum refinery unit. The dependent variable (Y) is the octane value of the petroleum produced and there are four independent variables: three relate to feed composition (X_1, X_2, X_3) and the fourth relates to process conditions (X_4). Eighty-two observations were taken, giving the following sample correlation matrix:

$$\widehat{\mathbf{R}} = \begin{array}{c} \begin{array}{ccccc} Y & X_1 & X_2 & X_3 & X_4 \end{array} \\ \begin{pmatrix} 1.000 & -0.870 & 0.392 & -0.638 & 0.629 \\ -0.870 & 1.000 & -0.589 & 0.449 & -0.337 \\ 0.392 & -0.589 & 1.000 & -0.298 & 0.161 \\ -0.638 & 0.449 & -0.298 & 1.000 & -0.722 \\ 0.629 & -0.337 & 0.161 & -0.722 & 1.000 \end{pmatrix} \end{array} \begin{array}{c} \\ Y \\ X_1 \\ X_2 \\ X_3 \\ X_4 \end{array}$$

After centring the variables, regression of Y on the four independent variables gave

$$\widehat{Y} = -0.824X_1 - 0.172X_2 - 0.097X_3 + 0.309X_4, \quad (R^2 = 0.905) \tag{23}$$

as the regression model. There is clear evidence that X_1, X_2 and X_4 should be included in the regression model ($p < 0.0002$ for each of these three variables) but whether X_3 should be included is debatable. The null hypothesis that the regression coefficient for X_3 is zero is rejected only at significance level 0.07. Omitting X_3 from the model gives the regression equation

$$\widehat{Y} = -0.841X_1 - 0.163X_2 + 0.372X_4, \quad (R^2 = 0.902). \tag{24}$$

Table 3: Relative importances of variables before and after omitting X_3

	OC	RW	DA	NM1	NM2	NM3
Relative importances before omitting X_3						
X_1	0.593	0.515	0.518	0.488	0.439	0.685
X_2	0.012	0.066	0.064	0.009	0.051	0.060
X_3	0.121	0.146	0.150	0.160	0.178	0.000
X_4	0.180	0.179	0.173	0.250	0.238	0.161
Relative importances after omitting X_3						
X_1	0.640	0.570	0.578	0.604	0.542	0.665
X_2	0.016	0.075	0.075	0.017	0.072	0.074
X_4	0.246	0.257	0.248	0.281	0.287	0.162

The top half of Table 3 displays the real importance assigned to the different X-variables by the different measures when all four X-variables are included in the regression model. Surprisingly, all but one of the measures gives X_3 a higher relative importance than X_2, even though X_3 is the variable whose inclusion in the model is tenuous. The NM3 measure is the exception. It gives X_3 a relative importance of 0.0, which concords fully with the inference that X_3 can reasonably be omitted from the regression model.

The lower half of Table 3 shows the relative importances assigned to X_1, X_2 and X_4 after X_3 has been omitted from the model. In the whole of the table, the RW and DA measures are strikingly similar in all their evaluations. It is also the case that all measures evaluate X_1 as the most important variable and X_2 as the second most important (both before and after omitting X_3). In other respects though, there is limited agreement across measures. For example, NM1 and NM2 agree quite closely in their evaluations of X_1 and X_4, but NM1 is similar to OC in its evaluation of X_2, while NM2's evaluations of X_2 are similar to those of RW, DA and NM3.

With most measures, the relative importance of X_3 is far greater than the difference between the R^2 values of the models in equations (23) and (24). Hence, with those measures the omission of X_3 must substantially increase the relative importance of at least one X-variable. As X_3 has higher absolute correlation with X_4 than with X_1 or X_2, it might be anticipated that omitting X_3 would increase the relative importance of X_4 more than that of X_1 or X_2. This is indeed the case for the OC, RW and DA measures, but not for NM1, NM2 or NM3. It seems then, that the effects on relative importance of omitting a variable are somewhat unpredictable and can vary markedly with the choice of measure.

5 Conclusion

Six measures for evaluating the relative importance of predictor variables in a regression have been examined. From the examples presented in Section 4, it is clear that usually there is some consensus between them – variables given a high relative importance by one measure are usually given a high relative importance by other measures, and similarly for low relative importance. At the same time, in each example there were differences between the measures in their evaluations, and some differences were substantial.

The following correlation matrix combines results from Tables 1–3 to give an overview of the similarity between the different measures. It gives the correlation between each pair of measures for the contributions recorded in Table 1 and the top halves of Tables 2 and 3. (The lower halves of Tables 2 and 3 are ignored to avoid double-counting of data.)

$$
\begin{array}{c c}
 & \begin{array}{cccccc} OC & RW & DA & NM1 & NM2 & NM3 \end{array} \\
\begin{array}{c} OC \\ RW \\ DA \\ NM1 \\ NM2 \\ NM3 \end{array} &
\begin{pmatrix}
1.000 & 0.981 & 0.982 & 0.978 & 0.981 & 0.932 \\
0.981 & 1.000 & 0.999 & 0.976 & 0.988 & 0.903 \\
0.982 & 0.999 & 1.000 & 0.977 & 0.990 & 0.897 \\
0.978 & 0.976 & 0.977 & 1.000 & 0.984 & 0.845 \\
0.981 & 0.988 & 0.990 & 0.984 & 1.000 & 0.889 \\
0.932 & 0.903 & 0.897 & 0.845 & 0.889 & 1.000
\end{pmatrix}
\end{array}
$$

The correlations between methods are very high in general, and the correlation between the RW and DA methods is especially high (0.999), showing the high concordance between these two methods that has been found in previous studies (Krasikova et al., 2011; Johnson, 2000). The other striking feature of the correlations is the comparatively low correlation between NM3 and each of the other methods (never exceeding 0.94), indicating that NM3 gives a distinctive perspective on the contributions of variables.

Occasionally, common sense shows that an evaluation is unreasonable. For instance, in Example 2 the OC measure evaluated the relative importance of X_1 as 100% and that of X_2 as 0%. This is clearly inappropriate, as all the variation in Y could not be explained by X_1 on its own, but could be explained by the combination of X_1 and X_2. Often though, the evaluations of the different measures all seem reasonable and how to choose between them is not clear-cut, because there are no known 'correct' evaluations with which to make comparison. As noted by Johnson and LeBreton (2004, page 240), "Because there is no unique mathematical solution to the problem [of evaluating relative importances], these indices [measures] must be evaluated on the basis of the logic behind their development, the apparent sensibility of the results they provide, and whatever shortcomings can be identified." Properties of the different measures and features of the data set should also be taken into account.

The following arguments favour different measures.

1. The DA and RW measures have been the most widely recommended measures in recent years, partly because they typically give similar evaluations, suggesting that there is an underlying construct that they both appraise. The examples presented here support that rationale, as they give further evidence that the two measures generally give similar results – there is only one case (variable X_1 in Example 3) where the DA and RW evaluations differ appreciably. The RW measure is simpler and easier to implement than the DA measure.

2. The OC and NM1 measures have the rotation invariance property so, with either measure, multicollinearities have little affect on the relative weights given to variables not involved in the collinearities.

3. In constructing the new measures (NM1–3), the aim was to improve upon the OC and RW measures by letting Y influence the transformation to orthogonality, rather than determining the transformation from just the values of the regressors. This was motivated by the observation that the transformation's purpose is to help evaluate the relationship between Y and the regressors, so

both should be taken into account in forming the transformation. On that basis, NM1 is to be preferred over OC, since in other respects the construction of the two measures are very similar. Similarly for NM2 and RW.

4. In the examples, a feature of NM3 is that it gave low relative importance to variables that might reasonably be omitted from the regression, which could be considered an attractive characteristic. In Example 5, for instance, it gave X_3 a relative importance of 0.000 while other measures gave it a relative importance of 0.121 or more. Similarly, in Example 3, predictions of Y are not improved by including X_1 in the regression model, but only NM3 gave X_1 a low evaluation.

Taking account of the above points, the NM3 measure is recommended when there are some independent variables whose inclusion or exclusion from the regression model is debateable. When it is clear which independent variables should feature in the regression, but there is high multicollinearity among a subset of them, then the NM1 measure is recommended because it has the rotation invariance property, though its choice in preference to the OC measure (which also has the rotation invariance property) is close. In other circumstances, one of the RW, DA and NM2 measures should be used and we recommend the RW measure – the three measures are likely to give very similar evaluations of relative importance and the RW measure is widely recommended for its simplicity and ease of use (Tonidandel and LeBreton, 2010).

The new measures presented here and ideas behind them could be adapted to give other measures of potential value. In particular, any of the OC, RW and NM2 measures could be modified to use regression coefficients as weights when forming orthogonal counterparts, in the same way that NM3 is derived from NM1. The weighting scheme could also be generalised to use the $(|\hat{\beta}_i|)^\alpha$ as weights (where the β_i are the multiple regression coefficients). Setting α equal to 0 would correspond to 'no weighting', and increasing α would increase the importance of the weighting.

Acknowledgments

We are grateful to a referee whose constructive comments led to clear improvements in the paper.

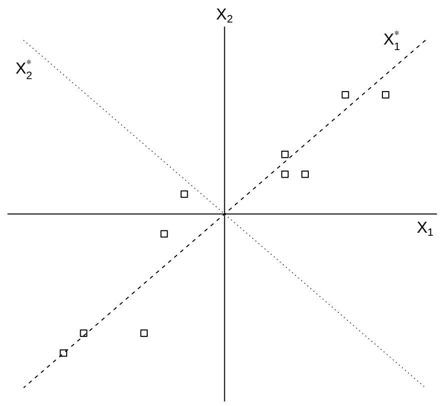

Figure 2(a) Points before rotation with X_1, X_2 as axes.

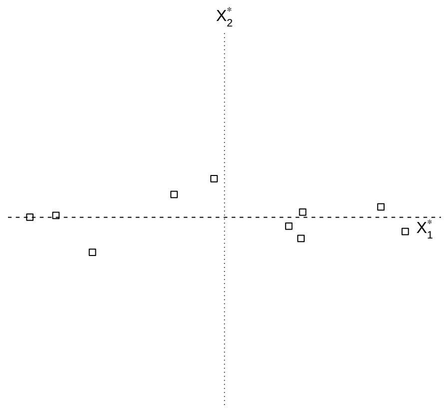

Figure 2(b) Points after rotation with X_1^*, X_2^* as axes.

Appendix A: Rotation invariance property

An orthogonal rotation of axes X_1, X_2 to axes X_1^*, X_2^* is illustrated in Figures 2(a) and 2(b). In Figure 2(a), the positions of 10 points (x_1, x_2) are plotted and new axes X_1^* and X_2^* are shown. The new axes are obtained by rotating the original axes X_1 and X_2 (by 45^o in this case). Figure 2(b) shows the same 10 points, but drawn with X_1^* and X_2^* as the horizontal and vertical axes. It can be seen that rotation of axes changes the correlation between variables: Figure 2(a) shows that the points are highly correlated when expressed in terms of X_1 and X_2, while Figure 2(b) shows that the correlation is low when the points are expressed in terms of X_1^* and X_2^*. Consequently, orthogonal rotation can be used to remove or reduce collinearity between variables.

We only need to rotate those variables that are involved in collinearities. For example, suppose there is just one collinearity and it involves only the first d of the k explanatory variables. Then axes are rotated using a rotation matrix, $\mathbf{\Gamma}$ say, that has the following block-diagonal form:

$$\mathbf{\Gamma} = \begin{pmatrix} \mathbf{\Gamma}_d & \mathbf{0} \\ \mathbf{0} & \mathbf{I_{k-d}} \end{pmatrix}, \tag{25}$$

where Γ_d is an orthogonal matrix of order d and $\mathbf{I_{k-d}}$ is a $(k-d)$ order identity matrix.

Rotation produces new variables that are linear combinations of the original predictors. The rotation matrix should be chosen in such a way that the variables that are created have meaningful interpretation. For example, if only the first two predictors X_1 and X_2 are responsible for one collinearity then $\mathbf{\Gamma}_d$ can be set as:

$$\mathbf{\Gamma}_d = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix} \tag{26}$$

This rotation creates two meaningful variables, the first one is proportional to $X_1 + X_2$ and the second one is proportional to $X_2 - X_1$.

In terms of the original variables, X_1 and X_2, the ten points in Figure 2 form the data matrix:

$$\begin{pmatrix} -0.48 & -0.42 & -0.24 & -0.18 & -0.12 & 0.18 & 0.18 & 0.24 & 0.36 & 0.48 \\ -0.48 & -0.41 & -0.41 & -0.07 & 0.07 & 0.14 & 0.21 & 0.14 & 0.41 & 0.41 \end{pmatrix}. \tag{27}$$

Post-multiplying this data matrix by $\mathbf{\Gamma}_d$ in equation (26) gives the points in terms of the new variables X_1^* and X_2^*:

$$\begin{pmatrix} -0.68 & -0.59 & -0.46 & -0.18 & -0.04 & 0.22 & 0.27 & 0.27 & 0.55 & 0.63 \\ 0.00 & 0.01 & -0.12 & 0.08 & 0.13 & -0.03 & 0.02 & -0.07 & 0.04 & -0.05 \end{pmatrix}.$$

The sample correlation between X_1 and X_2 is 0.951, while the sample correlation between X_1^* and X_2^* is 0. (The correlation between the sum and difference of two variables that have been standardised to have equal variances is always 0.)

With the majority of measures of importance, rotating some explanatory variables will change the relative importance of *every* variable. However, results in Garthwaite and Koch (2016) show that with the OC measure only the relative importances of variables involved in the rotation are changed – the relative importances are unchanged for those variables that are not involved in the rotation. Theorem 2 (below) shows that NM1 also has this rotation invariance property.

Lemma 1. *If* $\mathbf{H}$ *is a positive-definite matrix and* $\mathbf{\Gamma}$ *is an orthogonal matrix of the same dimension as* $\mathbf{H}$*, then* $(\mathbf{\Gamma}^T\mathbf{H}\mathbf{\Gamma})^{-\mathbf{1/2}} = \mathbf{\Gamma^T}\mathbf{H}^{-\mathbf{1/2}}\mathbf{\Gamma}$.

Proof. $(\mathbf{\Gamma}^T\mathbf{H}^{\mathbf{1/2}}\mathbf{\Gamma}).(\mathbf{\Gamma^T}\mathbf{H}^{\mathbf{1/2}}\mathbf{\Gamma}) = \mathbf{\Gamma^T}\mathbf{H}^{\mathbf{1/2}}(\mathbf{\Gamma}\mathbf{\Gamma^T})\mathbf{H}^{\mathbf{1/2}}\mathbf{\Gamma} = \mathbf{\Gamma^T}\mathbf{H}\mathbf{\Gamma}$, so $(\mathbf{\Gamma}^T\mathbf{H}\mathbf{\Gamma})^{\mathbf{1/2}} = \mathbf{\Gamma^T}\mathbf{H}^{\mathbf{1/2}}\mathbf{\Gamma}$. Hence, $(\mathbf{\Gamma}^T\mathbf{H}\mathbf{\Gamma})^{-\mathbf{1/2}} = (\mathbf{\Gamma^T}\mathbf{H}^{\mathbf{1/2}}\mathbf{\Gamma})^{-\mathbf{1}} = \mathbf{\Gamma}^{-\mathbf{1}}\mathbf{H}^{-\mathbf{1/2}}(\mathbf{\Gamma^T})^{-\mathbf{1}} = \mathbf{\Gamma^T}\mathbf{H}^{-\mathbf{1/2}}\mathbf{\Gamma}$. $\square$

Lemma 2. *Suppose* $\mathbf{X}^* = \mathbf{X}\boldsymbol{\Gamma}$*. Under the constraints that* $\mathbf{W}^*$ *is a linear transformation of* $\mathbf{X}^*$ *and that* $(\mathbf{W}^*)^{\mathbf{T}}\mathbf{W}^* = \mathbf{I_k}$*, the value of* $\mathbf{W}^* = (\boldsymbol{w}_\mathbf{1}^*, \ldots, \boldsymbol{w}_\mathbf{k}^*)$ *that maximises* $\sum_{i=1}^{k}(\mathbf{Y}\boldsymbol{w}_\mathbf{i}^*)^{\mathbf{T}}(\mathbf{Y}\boldsymbol{x}_\mathbf{i}^*)$ *is*

$$\mathbf{W}^* = \mathbf{W}\boldsymbol{\Gamma}, \tag{28}$$

where $\mathbf{W}$ *is defined by equations (13), (14) and (15).*

Proof. Let $\boldsymbol{\Psi}^* = [(\mathbf{X}^*)^{\mathbf{T}}\mathbf{X}^*]^{-\mathbf{1/2}}(\mathbf{X}^*)^{\mathbf{T}}\mathbf{YYX}^*$ and put $\mathbf{G}^* = \boldsymbol{\Psi}^*[(\boldsymbol{\Psi}^*)^{\mathbf{T}}\boldsymbol{\Psi}^*]^{-\mathbf{1/2}}$. Now, $[(\mathbf{X}^*)^{\mathbf{T}}\mathbf{X}^*]^{-\mathbf{1/2}}(\mathbf{X}^*)^{\mathbf{T}} = [\boldsymbol{\Gamma}^{\mathbf{T}}\mathbf{X}^{\mathbf{T}}\mathbf{X}\boldsymbol{\Gamma}]^{-\mathbf{1/2}}\boldsymbol{\Gamma}^{\mathbf{T}}\mathbf{X}^{\mathbf{T}} = \boldsymbol{\Gamma}^{\mathbf{T}}[\mathbf{X}^{\mathbf{T}}\mathbf{X}]^{-\mathbf{1/2}}\boldsymbol{\Gamma}\boldsymbol{\Gamma}^{\mathbf{T}}\mathbf{X}^{\mathbf{T}}$ (from Lemma 1), so

$$[(\mathbf{X}^*)^{\mathbf{T}}\mathbf{X}^*]^{-\mathbf{1/2}}(\mathbf{X}^*)^{\mathbf{T}} = \boldsymbol{\Gamma}^{\mathbf{T}}[\mathbf{X}^{\mathbf{T}}\mathbf{X}]^{-\mathbf{1/2}}\mathbf{X}^{\mathbf{T}}. \tag{29}$$

Hence, $\boldsymbol{\Psi}^* = \boldsymbol{\Gamma}^{\mathbf{T}}[\mathbf{X}^{\mathbf{T}}\mathbf{X}]^{-\mathbf{1/2}}\mathbf{X}^{\mathbf{T}}\mathbf{YYX}\boldsymbol{\Gamma} = \boldsymbol{\Gamma}^{\mathbf{T}}\boldsymbol{\Psi}\boldsymbol{\Gamma}$, where $\boldsymbol{\Psi}$ is defined in equation (15). Thus $\mathbf{G}^* = \boldsymbol{\Gamma}^{\mathbf{T}}\boldsymbol{\Psi}\boldsymbol{\Gamma}[(\boldsymbol{\Gamma}^{\mathbf{T}}\boldsymbol{\Psi}\boldsymbol{\Gamma})(\boldsymbol{\Gamma}^{\mathbf{T}}\boldsymbol{\Psi}\boldsymbol{\Gamma})]^{-\mathbf{1/2}} = \boldsymbol{\Gamma}^{\mathbf{T}}\boldsymbol{\Psi}\boldsymbol{\Gamma}\boldsymbol{\Gamma}^{\mathbf{T}}[\boldsymbol{\Psi}\boldsymbol{\Gamma}\boldsymbol{\Gamma}^{\mathbf{T}}\boldsymbol{\Psi}]^{-\mathbf{1/2}}\boldsymbol{\Gamma}$ (from Lemma 1), so $\mathbf{G}^* = \boldsymbol{\Gamma}^{\mathbf{T}}\boldsymbol{\Psi}[\boldsymbol{\Psi}\boldsymbol{\Psi}]^{-\mathbf{1/2}}\boldsymbol{\Gamma} = \boldsymbol{\Gamma}^{\mathbf{T}}\mathbf{G}\boldsymbol{\Gamma}$, where $\mathbf{G}$ is defined by equation (15). Now the proof of Theorem 1 does not require the fact that the X variables have been standardised to have unit variance. Hence the result of the theorem also applies to $\mathbf{W}^*$ and $\mathbf{X}^*$. It follows that $\mathbf{W}^* = (\mathbf{X}^*)^{\mathbf{T}}[(\mathbf{X}^*)^{\mathbf{T}}\mathbf{X}^*]^{-\mathbf{1/2}}\mathbf{G}^*$ so, from equation (29), $\mathbf{W}^* = \mathbf{X}[\mathbf{X}^{\mathbf{T}}\mathbf{X}]^{-\mathbf{1/2}}\boldsymbol{\Gamma}\mathbf{G}^*$. As $\mathbf{G}^* = \boldsymbol{\Gamma}^{\mathbf{T}}\mathbf{G}\boldsymbol{\Gamma}$, this gives $\mathbf{W}^* = \mathbf{X}[\mathbf{X}^{\mathbf{T}}\mathbf{X}]^{-\mathbf{1/2}}\boldsymbol{\Gamma}\boldsymbol{\Gamma}^{\mathbf{T}}\mathbf{G}\boldsymbol{\Gamma}$, so $\mathbf{W}^* = \mathbf{W}\boldsymbol{\Gamma}$, where $\mathbf{W}$ is defined by equation (13). $\square$

If say, just the first d of k explanatory variables are rotated, then the rotation matrix $\boldsymbol{\Gamma}$ has the block-diagonal structure in equation (25). Then, from Lemma 2, $(\boldsymbol{w}_{d+1}^*, \ldots, \boldsymbol{w}_k^*) = (\boldsymbol{w}_{d+1}, \ldots, \boldsymbol{w}_k)$. When Y is regressed on $(\boldsymbol{w}_1^*, \ldots, \boldsymbol{w}_k^*)$, the contribution of $\boldsymbol{w}_i^*$ to the RegSS is the same as the RegSS from a univariate regression of Y on $\boldsymbol{w}_i^*$, because $\boldsymbol{w}_1^*, \ldots, \boldsymbol{w}_k^*$ are an orthogonal set of vectors. Under NM1, this RegSS is taken as the relative importance of X_i^* in a regression of Y on $(X_1^*, \ldots, X_k^*)$. Similarly, when Y is regressed on $(X_1, \ldots, X_k)$, NM1 evaluates the relative importance of X_i as the RSS from a univariate regression of Y on $\boldsymbol{w}_i$. As $\boldsymbol{w}_i^* = \boldsymbol{w}_i$ for $i = d+1, \ldots, k$, NM1 has the rotation invariance property given in the following theorem.

Theorem 2. *If an orthogonal rotation is applied to some of the X variables, the relative importance of the other X variables is unchanged if relative importance is measured using NM1.*

Appendix B: Proof of Theorem 1

Preliminary lemma:

Lemma 3. *If* $\mathbf{W} = \mathbf{XC}$ *and* $\mathbf{W^TW} = \mathbf{I_k}$*, then* $\mathbf{W} = \mathbf{X}(\mathbf{X^TX})^{-1/2}\mathbf{G}$ *where* $\mathbf{G}$ *is a* $k \times k$ *orthogonal matrix. The converse also holds: if* $\mathbf{W} = \mathbf{X}(\mathbf{X^TX})^{-1/2}\mathbf{G}$ *and* $\mathbf{G}$ *is an orthogonal matrix, then* $\mathbf{W^TW} = \mathbf{I_k}$.

Proof of Lemma 3. For the first part of the lemma, let $\mathbf{G} = (\mathbf{X^TX})^{1/2}\mathbf{C}$. Then $\mathbf{C} = (\mathbf{X^TX})^{-1/2}\mathbf{G}$ and $\mathbf{W} = \mathbf{X}(\mathbf{X^TX})^{-1/2}\mathbf{G}$. Also $\mathbf{I_k} = \mathbf{W^TW} = \mathbf{G^T}(\mathbf{X^TX})^{-1/2}\mathbf{X^TX}(\mathbf{X^TX})^{-1/2}\mathbf{G} = \mathbf{G^TG}$. This implies that $\mathbf{G}$ is orthogonal, as required. The converse is immediate: if $\mathbf{W} = \mathbf{X}(\mathbf{X^TX})^{-1/2}\mathbf{G}$ and $\mathbf{G}$ is an orthogonal matrix, then $\mathbf{W^TW} = \mathbf{G^T}(\mathbf{X^TX})^{-1/2}\mathbf{X^TX}(\mathbf{X^TX})^{-1/2}\mathbf{G} = \mathbf{G^TG} = \mathbf{I_k}$. □

Proof of Theorem 1. From Lemma 3, $\mathbf{W} = \mathbf{X}(\mathbf{X^TX})^{-1/2}\mathbf{G}$ where $\mathbf{G}$ is an orthogonal matrix. Put $\mathbf{G} = (\boldsymbol{g}_\mathbf{1}, \ldots, \boldsymbol{g}_\mathbf{k})$, so $\boldsymbol{w}_i = \mathbf{X}(\mathbf{X^TX})^{-1/2}\boldsymbol{g}_\mathbf{i}$. Also, define $\boldsymbol{\psi}_i = (\mathbf{X^TX})^{-1/2}\mathbf{X^TYY}\boldsymbol{x}_\mathbf{i}$ for $i = 1, \ldots, k$. Then $\sum_{i=1}^{k}(\mathbf{Y}\boldsymbol{w}_\mathbf{i})^\mathbf{T}(\mathbf{Y}\boldsymbol{x}_\mathbf{i}) = \sum_{i=1}^{k}(\boldsymbol{g}_i^T(\mathbf{X^TX})^{-1/2}\mathbf{X^TY^T})(\mathbf{Y}\boldsymbol{x}_\mathbf{i}) = \sum_{i=1}^{k}\boldsymbol{g}_i^T\boldsymbol{\psi}_i$. As $\mathbf{G}$ is an orthogonal matrix, it is immediate from Theorem 1 in Garthwaite et al. (2012) that $\sum_{i=1}^{k}\boldsymbol{g}_i^T\boldsymbol{\psi}_i$ is maximised when $\mathbf{G} = \boldsymbol{\Psi}(\boldsymbol{\Psi}^\mathbf{T}\boldsymbol{\Psi})^{-1/2}$, where $\boldsymbol{\Psi} = (\boldsymbol{\psi}_\mathbf{1}, \ldots, \boldsymbol{\psi}_\mathbf{k})$. Thus equation (15) defines $\boldsymbol{\Psi}$. □

References

Budescu, D. V. (1993). Dominance analysis: A new aproach to the problem of relative importance of predictors in multiple regression. *Psychological Bulletin*, 114:542–551.

Conklin, M., Powaga, K., and Lipovetsky, S. (2004). Customer satisfaction analysis: Identification of key drivers. *European Journal of Operational Research*, 154:819–827.

Courville, T. and Thompson, B. (2001). Use of structure coefficients in published multiple regression articles: β is not enough. *Educational and Psychological Measurement*, 61:229–248.

Darlington, R. B. (1968). Multiple regression in psychological research and practice. *Psychological Bulletin*, 69(3):161–182.

Garthwaite, P. H., Critchley, F., Anaya-Izquierdo, K., and Mubwandarikwa, E. (2012). Orthogonalization of vectors with minimal adjustment. *Biometrika*, 99(4):787–798.

Garthwaite, P. H. and Koch, I. (2016). Evaluating the contributions of individual variables to a quadratic form. *Australian & New Zealand Journal of Statistics*, 58(1):99–119.

Gibson, W. (1962). Orthogonal predictors: A possible resolution of the Hoffman-Ward controversy. *Psychological Reports*, 11(1):32–34.

Grömping, U. (2007). Estimators of relative importance in linear regression based on variance decomposition. *The American Statistician*, 61:139–147.

Grömping, U. et al. (2006). Relative importance for linear regression in R: the package Relaimpo. *Journal of Statistical Software*, 17(1):1–27.

Healy, M. (1990). Measuring importance. *Statistics in Medicine*, 9(6):633–637.

Johnson, J. W. (2000). A heuristic method for estimating the relative weight of predictor variables in multiple regression. *Multivariate Behavioral Research*, 35(1):1–19.

Johnson, J. W. and LeBreton, J. M. (2004). History and use of relative importance indices in organizational research. *Organizational Research Methods*, 7:238–257.

Johnson, R. M. (1966). The minimal transformation to orthonormality. *Psychometrika*, 31(1):61–66.

Krasikova, D., Lebreton, J. M., and Tonidandel, S. (2011). Estimating the relative importance of variables in multiple regression models. In *International Review of Industrial and Organizational Psychology*. Indianopolis: Wiley.

Liakhovitski, D., Bryukhov, Y., and Conklin, M. (2010). Relative importance of predictors: Comparison of Random Forest with Johnson's relative weights. *Model Assisted Statistics and Applications*, 5:235–249.

Lindeman, R. H., Merenda, P. F., and Gold, R. Z. (1980). *Introduction to Bivariate and Multivariate Analysis.* Scott, Foresman.

Lipovetsky, S. and Conklin, M. (2001). Analysis of regression in game theory approach. *Applied Stochastic Models in Business and Industry*, 17:319–330.

Lipovetsky, S. and Conklin, M. (2010). Meaningful regression analysis in adjusted coefficients Shapley value model. *Model Assisted Statistics and Applications*, 5:251–264.

Lipovetsky, S. and Conklin, W. M. (2015). Predictor relative importance and matching regression parameters. *Journal of Applied Statistics*, 42:1017–1031.

Longley, J. W. (1967). An appraisal of least squares programs for the electronic computer from the point of view of the user. *Journal of the American Statistical Association*, 62(319):819–841.

Nathans, L. L., Oswald, F. L., and Nimon, K. (2012). Interpreting multiple linear regression: A guidebook of variable importance. *Practical Assessment, Research and Evaluation*, 17(9):1–19.

Pratt, J. W. (1987). Dividing the indivisible: Using simple symmetry to partition variance explained. In *Proceedings of the Second International Conference in Statistics*, pages 245–260. Tampere, Finland: University of Tampere.

Stadler, M., Cooper-Thomas, H. D., and Greiff, S. (2017). A primer on relative importance analysis: Illustrations of its utility for psychological research. *Psychological Test and Assessment Modeling*, 59:381–403.

Tonidandel, S. and LeBreton, J. M. (2010). Determining the relative importance of predictors in logistic regression: An extension of relative weight analysis. *Organizational Research Mehods*, 13:767–781.

Wood, F. S. (1973). The use of individual effects and residuals in fitting eqations to data. *Technometrics*, 15:677–695.