Psychological Methods

Regression With Reduced Rank Predictor Matrices: A Model of Trade-Offs

Mark L. Davison, Ernest C. Davenport, Jr, Hao Jia, Ben Seipel, and Sarah E. Carlson

Online First Publication, July 4, 2022. http://dx.doi.org/10.1037/met0000512

CITATION

Davison, M. L., Davenport, E. C., Jr, Jia, H., Seipel, B., & Carlson, S. E. (2022, July 4). Regression With Reduced Rank Predictor Matrices: A Model of Trade-Offs. *Psychological Methods*. Advance online publication. http://dx.doi.org/10.1037/met0000512

AMERICAN PSYCHOLOGICAL ASSOCIATION

Psychological Methods

ISSN: 1082-989X

https://doi.org/10.1037/met0000512

Regression With Reduced Rank Predictor Matrices: A Model of Trade-Offs

Mark L. Davison[1], Ernest C. Davenport Jr[1], Hao Jia[1], Ben Seipel[2], and Sarah E. Carlson[3]
[1] Department of Educational Psychology, University of Minnesota
[2] School of Education, California State University, Chico
[3] Department of Learning Sciences, Georgia State University

Abstract
A regression model of predictor trade-offs is described. Each regression parameter equals the expected change in Y obtained by trading 1 point from one predictor to a second predictor. The model applies to predictor variables that sum to a constant T for all observations; for example, proportions summing to $T = 1.0$ or percentages summing to $T = 100$ for each observation. If predictor variables sum to a constant T for all observations and if a least squares solution exists, the predicted values for the criterion variable Y will be uniquely determined, but there will be an infinite set of linear regression weights and the familiar interpretation of regression weights does not apply. However, the regression weights are determined up to an additive constant and thus differences in regression weights $\beta_v - \beta_{v*}$ are uniquely determined, readily estimable, and interpretable. $\beta_v - \beta_{v*}$ is the expected increase in Y given a transfer of 1 point from variable $v*$ to variable v. The model is applied to multiple-choice test items that have four response categories, one correct and three incorrect. Results indicate that the expected outcome depends, not just on the student's number of correct answers, but also on how the student's incorrect responses are distributed over the three incorrect response types.

Translational Abstract
A regression model of predictor trade-offs is described. Each regression parameter equals the expected change in Y obtained by trading 1 point from one predictor to a second predictor. The model applies to predictor variables that sum to a constant T for all observations; for example, proportions summing to $T = 1.0$ or percentages summing to $T = 100$ for each observation. The model is designed for the study of decisions involving trade-offs, compositional variables, and contrasts between pairs of predictor coefficients.

Keywords: regression, compositional variables, trade-offs, selected response items, decision theory

Supplemental materials: https://doi.org/10.1037/met0000512.supp

In some research applications, a researcher may wish to use predictor variables that sum to the same constant T for each observation. However, if the predictors sum to a constant the predictor matrix is not full column rank, and therefore the regression coefficients are not uniquely determined. Given an infinite set of predictor coefficient vectors, one is hard pressed to choose any one set for interpretation. A common solution, what we call the classical solution, is to drop one of the predictors. However, the regression weight for a given predictor will vary depending on which other predictor is dropped. In what follows, we show that these parameter estimates can still be meaningful.

The most common example of predictors summing to a constant T are compositional variables, variables that describe the make-up of an observation in terms of proportions or percentages. For instance, the composition of a school might be described in terms of three variables: the proportion of girls, the proportion of boys, and the proportion of nonbinary students. Or it might be described in terms of six variables: the percent of American Indians, the percent of Asians, the percent of Blacks, the percent of Hispanics, the percent of Whites, and the percent of Other. To examine how diversity relates to school achievement, a researcher might be interested in regressing a criterion variable (e.g., achievement test scores) onto a vector of compositional predictor variables:

$$Y_p = \boldsymbol{\beta} x_{\boldsymbol{p}}^{\boldsymbol{T}} + a + e_p \quad (1)$$

where Y_p is the criterion variable (e.g., achievement test score) for observation p ($p = 1, \ldots, P$), $\boldsymbol{\beta}$ is a row vector of V ($v = 1, \ldots, V$) predictor weights, $\boldsymbol{x}_p = (x_{p1}, \ldots, x_{pV})$ is a row vector of V

Mark L. Davison https://orcid.org/0000-0003-3656-9672
The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grants R305A140185 and R305A190393 to the University of Oregon and Grant R305A180417 to the California State University, Chico. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

Data for this example will be made available in online supplementary material.

Correspondence concerning this article should be addressed to Mark L. Davison, Department of Educational Psychology, University of Minnesota, 56 East River Drive, Minneapolis MN 55455, United States. Email: MLD@umn.edu

compositional predictor variables, and e_p is a residual. For purposes of hypothesis tests, the residuals e_p are assumed normal, independent, and identically distributed $N(0, \sigma^2(e))$. If the predictor variables refer to categories that are mutually exclusive and exhaustive, there exists a total score T such that, for every observation p, the predictor vector satisfies the constraint

$$\sum_{v=1}^{v=V} x_{pv} = T \tag{2}$$

For proportions, $T = 1$, and for percentages $T = 100$. This constraint implies that any one of the predictor variables x_{pv} is linearly dependent on the remaining $V - 1$ predictors.

Compositional variables are not the only example of variables summing to a constant for each observation. Ipsative measures are a second example. If measures are ipsatized, the predictor scores of each observation have been centered around the observation's predictor mean so that the predictor scores of every individual sum to .0. Ipsative measures are associated with a more person-centered approach. Instead of comparing each person's score to that of other people, each person's predictor score is interpreted relative to the same person's other predictor scores. In the ipsative approach, scores are given a within person, configural interpretation, rather than a between person, normative interpretation (Davison et al., 2022; Davison & Davenport, 2002; Wiernik et al., 2020). In our example, we illustrate application to another common kind of data in psychology and education, multiple-choice item responses.

Predictor variables that sum to a constant pose several problems. First, the variables may not have a multivariate normal distribution, although this is not a problem if errors are independently and identically distributed $N(0, \sigma^2(e))$. Second, given the constraint in Equation 2, there is no unique solution for the regression weights. As shown below, the weights and intercept are only determined up to an additive constant k. Third, given the constraint in Equation 2, the $(P \times V)$ matrix of predictors is not of full column rank, and $\boldsymbol{X}^T\boldsymbol{X}$ does not have an inverse. Therefore, the parameters are inestimable using any solution that requires the inverse of $\boldsymbol{X}^T\boldsymbol{X}$ (e.g., ordinary least squares [OLS]). Finally, the standard interpretation of regression weights does not apply, because its premises violate the constraint in Equation 2. The standard interpretation is that β_v is the expected increase in Y given a 1-point increase in predictor X_v holding the other predictor variables constant. This definition envisions two predictor vectors differing by 1 point on one variable v; for example, $(X_1, \ldots, X_v, \ldots, X_V)$ and $(X_1, \ldots, X_v + 1, \ldots, X_V)$, and states that β_v will be the difference in the expected value of Y for these two vectors. However, these two vectors cannot both have the same sum T, so at least one of them must violate the fundamental premise of the constraint stated in Equation 2. Therefore, the standard interpretation of a regression weight is based on a comparison that is counterfactual given the constraint on predictors.

One approach to these problems is what we call the classical approach because it is, in our experience, the most common. It involves dropping one predictor variable, here called the reference variable– say variable v^*– to create a reduced $(P \times V - 1)$ predictor matrix $\boldsymbol{X}_{V-1}$. If $\boldsymbol{X}_{V-1}$ is of full column rank, then for the regression of Y onto $\boldsymbol{X}_{V-1}$, the regression weights are uniquely determined and can be obtained using OLS regression because $\boldsymbol{X}_{V-1}^T\boldsymbol{X}_{V-1}$ will have an inverse. The weights are uniquely determined for a variable v given a reference variable v^*, but the set of regression weights will vary depending on the choice of reference variable. For each reference variable one gets a different set of regression weights. How can one justify the choice of one set of weights over the others; especially when each set may give a different impression of the relationship between the predictor variables as well as the relationship of the predictors relative to the criterion? In what follows, we have a new derivation that leads to a redefinition, reevaluation, and reinterpretation of parameters that are meaningful.

The classical solution is not the only proposed solution. The various alternatives involve one or more of the following steps: transforming predictors (log ratios, Box Cox transformation, Box & Cox, 1964), reducing the number of predictors to resolve the linear dependency (e.g., isometric log ratios), and/or placing constraints on regression weights to resolve the linear dependency (centered log ratios). For discussions of alternatives, see Aitchison (1982), Aitchison and Bacon-Shone (1984), Aitchison et al. (2000), Chen et al. (2017), Egozcue et al. (2003), Greenacre and Grunsky (2019), Hron et al. (2012), and Smithson and Broomell (2022). All of these alternatives suffer from a common problem: the magnitudes of the regression coefficients are difficult, if not impossible, to interpret for one or more of the following reasons. First, after transformation, the predictors are not in natural units, complicating interpretation of both the predictors and the regression coefficients. Second, after transformation one or more predictors may be a composite of several predictors, again complicating interpretation of both the predictors and the regression coefficients. Third, there may be more than one way to reduce the number of variables or transform the variables such that each procedure yields different regression coefficients leading to the question "which set of coefficients should we interpret?" For instance, the isometric log transformation requires an ordering of the variables and every possible ordering yields a different set of regression coefficients. For the alternatives, there is no way to interpret the magnitudes of regression weights consistent with the fact that the model is defined only for predictor vectors summing to the constant T.[1] Those that involve the log transformation are undefined when the argument of the log function is nonpositive, as happens when percentages or proportions equal 0 or ipsative scores are negative.

In what follows, we show that coefficients in the reduced form models have an interpretation lacking in prior approaches. The remainder of this paper is concerned with three issues: the uniqueness of estimates with one variable deleted, invariance of coefficients over choice of deleted variable, and interpretation of regression weights in the reduced form model with one predictor deleted. To lay the groundwork for these issues, the next section takes up the uniqueness properties of regression weights in the full model of Equation 1 when predictors are subject to the constraint in Equation 2. We refer to the regression model in Equation 1 as the parent model. We refer to each model in which one predictor

[1] In some cases the interpretation is limited to whether the predictor matters, that is, whether the coefficient is zero or not. Magnitudes other than zero are uninterpretable beyond saying that the predictor is associated with an effect.

is dropped as a child model. The children are viewed as offspring of the parent model. The interpretation of regression weights in the child models is based on the weights in the parent model.

The main goal of this article is to provide a method to estimate weights relative to Equation 1 that have meaning. Note that these weights would be inestimable in OLS regression with the constraint in Equation 2 and when estimable there would be an infinite set of weights where the choice of any one set is mathematically unjustifiable. A benefit of our approach is that we use OLS regression which is well-known and easily implemented. Furthermore, our use of OLS regression will provide weights in the metric of the predictors that are consistent, unbiased, and minimum variance (Fox, 2016, pp. 109–110). A final benefit of our approach will be the new meaning of the hypothesis test for each individual regression weight.

Uniqueness of Regression Weights in Parent and Child Models

Theorem 1: If $E(Y_p \mid \boldsymbol{x}_p) = \boldsymbol{\beta}\boldsymbol{x}_p^T + a$, then $E(Y_p \mid \boldsymbol{x}_p) = \boldsymbol{\beta}^*\boldsymbol{x}_p^T + a^*$ if and only if there exists a constant k such that $\boldsymbol{\beta}^* = \boldsymbol{\beta} + k\mathbf{1}$ and $a^* = a - kT$ for all $\boldsymbol{x}$ satisfying Equation 2.

Here $\mathbf{1}$ is a row vector of 1s. It will be shown that if $E(Y_p \mid \boldsymbol{x}_p) = \boldsymbol{\beta}\boldsymbol{x}_p^T + a$, then $E(Y_p \mid \boldsymbol{x}_p) = \boldsymbol{\beta}^*\boldsymbol{x}_p^T + a^*$ for all $\boldsymbol{x}$ satisfying Equation 2 and all $\boldsymbol{\beta}^* = \boldsymbol{\beta} + k\mathbf{1}$ and $a^* = a - kT$. It is shown in the Appendix that if $(\boldsymbol{\beta}^*, a^*)$ does not have the form $\boldsymbol{\beta}^* = \boldsymbol{\beta} + k\mathbf{1}$ and $a^* = a - kT$, then there will exist at least one value of $\boldsymbol{x}$ satisfying Equation 2 for which $E(Y_p \mid \boldsymbol{x}_p) = \boldsymbol{\beta}\boldsymbol{x}_p^T + a \neq \boldsymbol{\beta}^*\boldsymbol{x}_p^T + a^*$.

Let $\boldsymbol{k}$ be a V- length row vector all of whose elements equal k. To show that $E(Y_p \mid \boldsymbol{x}_p) = \boldsymbol{\beta}\boldsymbol{x}_p^T + a = \boldsymbol{\beta}^*\boldsymbol{x}_p^T + a^*$ if $\boldsymbol{\beta}^* = \boldsymbol{\beta} + \boldsymbol{k}$ and $a^* = a - kT$, the proof begins by adding $(\boldsymbol{k} - \boldsymbol{k})$ to the regression weight vector on the right side of the following equation:

$$E(Y_p \mid \boldsymbol{x}_p) = \boldsymbol{\beta}\boldsymbol{x}_p^T + a \tag{3a}$$

$$= (\boldsymbol{\beta} + \boldsymbol{k} - \boldsymbol{k})\boldsymbol{x}_p + a \tag{3b}$$

$$= (\boldsymbol{\beta} + \boldsymbol{k})\boldsymbol{x}_p + a - \boldsymbol{k}\boldsymbol{x}_p \tag{3c}$$

Because $\boldsymbol{k}\boldsymbol{x}_p = k\sum_v x_v = kT$, (where T is the total of the predictors which is constant) then

$$E(Y_p \mid \boldsymbol{x}_p) = \boldsymbol{\beta}^*\boldsymbol{x}_p^T + a^* \tag{3d}$$

where $\boldsymbol{\beta}^* = \boldsymbol{\beta} + \boldsymbol{k}$ and $a^* = a - kT$. Thus, if $E(Y_p \mid \boldsymbol{x}_p)$ have a linear form, there will be an infinite number of solutions for the regression weights and intercept any two of which are related through an additive constant k. Let $\boldsymbol{S}$ be the set of all such parameter vectors $(\boldsymbol{\beta}^*, a^*)$.

If any two solutions for the regression weights $\boldsymbol{\beta}^*$ are related by an additive constant, the differences between the regression weights are uniquely determined. That is, if $\boldsymbol{\beta}$ and $\boldsymbol{\beta}^*$ are two solution vectors related by an additive constant k, then for all pairs of predictors (v, v*),

$$\beta_v - \beta_{v^*} = (\beta_v + k) - (\beta_{v^*} + k) = \beta_v^* - \beta_{v*}^*. \tag{4}$$

Even though there is an infinite set of solution vectors $(\boldsymbol{\beta}^*, a^*)$, for any pair of predictors (v, v*), the difference between regression weights is unique for all solution vectors in set $\boldsymbol{S}$. For Equation 1, given the constraint in Equation 2, the regression weight vector is defined only up to an additive constant, but differences between pairs of regression weights are uniquely determined. Given that the difference between any two regression weights is the same for all sets of coefficient vectors, these differences are mathematically meaningful.

Differences between weights are not only uniquely determined, they are interpretable. Because β_v is the expected change in Y given a 1-unit increase in X_v. and $-\beta_{v*}$ is the expected change in Y given a 1-unit decrease in the reference variable X_{v*}, the difference $\beta_v - \beta_{v*}$ is the expected change in Y given a transfer of 1 point from variable X_{v*} to X_v. The premise of this interpretation does not violate the constraint in Equation 2. That is, if the predictor variable $(X_1, \ldots, X_v, \ldots X_{v*} \ldots, X_V)$ satisfies the constraint, then transferring one point from variable X_{v*} to X_v will result in the predictor vector $(X_1, \ldots, X_v + 1, \ldots, X_{v*} - 1, \ldots X_V)$ which also satisfies the constraint.

Because the regression weights in the parent model are not uniquely determined, it can be said that the OLS solution is not uniquely determined. The unqualified statement that "the solution is not uniquely determined" is true, but involves several errors of omission because there are several aspects of the model that are uniquely determined or nearly so. First, the regression coefficients are almost uniquely determined in that they are determined up to an additive constant. Second, while there is an infinite set of least squares coefficient vectors $(\boldsymbol{\beta}^*, a^*)$, in set $\boldsymbol{S}$, for a given predictor vector $\boldsymbol{x}$, all coefficient vectors in $\boldsymbol{S}$ will return the same predicted values of Y. Thus, the predicted values Y' are uniquely determined. Since the predicted values are uniquely determined, the multiple correlation $R = r(Y, Y')$ is likewise uniquely determined. Because the predicted values are uniquely determined, the sum of squared errors *SSE*, the sum of squares regression *SSR*, and the sum of squares total *SST* will all be uniquely determined. While the regression weights are not uniquely determined, many aspects of the solution are: regression coefficient differences $\beta_v - \beta_{v*}$, Y', R, *SSE*, *SSR,* and *SST*. In light of these determinacies and the interpretability of differences between regression weights, we turn to a reconsideration of the meaning for regression weights in the classical approach that involves dropping one predictor.

Estimating Differences in Parent Model Regression Weights

If expressed in summation notation, the $E(Y_p \mid \boldsymbol{x_p})$ has the following form:

$$E(Y_p \mid \boldsymbol{x_p}) = \sum_v \beta_v x_p + a \tag{5}$$

According to the Theorem, a constant k can be added to each regression weight without altering the equality so long as there is a compensating change in the intercept. Let $k = -\beta_{v*}$, the negative of the regression weight for one variable in Equation 5. Then, by adding and subtracting $-\beta_{v*}$ to each regression weight in Equation 5 and applying some algebra, Equation 5 becomes

$$E(Y_p \mid \boldsymbol{x_p}) = \sum\nolimits_v (\beta_v - \beta_{v*})x_p + a - (-\beta_{v*})T \tag{6a}$$

$$= \sum_{v \neq v*} (\beta_v - \beta_{v*}) x_p + a + \beta_{v*} T \quad (6b)$$

$$= \sum_{v \neq v*} \beta_{vv*} x_p + a^* \quad (6c)$$

where $\beta_{vv*} = (\beta_v - \beta_{v*}), a^* = a + \beta_{v*} T$, and the summation in Equations 6b and 6c includes only the $V - 1$ predictors $v \neq v^*$. The term involving $v = v^*$ (reference variable) drops out, because the regression coefficient on that term is $(\beta_{v*} - \beta_{v*})$. If the reduced predictor matrix $\boldsymbol{X}_{V-1}$ of Equation 6c is of full column rank, the coefficients β_{vv*} and a^* can be estimated using ordinary least squares regression by regressing Y onto the reduced predictor matrix $\boldsymbol{X}_{V-1}$. Since $\beta_{vv*} = (\beta_v - \beta_{v*})$, each of the regression weight estimates can be interpreted as an estimate of the expected increase in Y given a transfer of one unit from the reference variable v^* to variable v. Hereafter for β_{vv*} we will refer to v and v^* as the receiving and sending variables respectively, since the interpretation of β_{vv*} is based on the expected increase in Y if one point is sent from v^* and received by variable v.

Equation 6c is here called a child of the original regression Equation 1 or Equation 5. The parent model with V predictors has V reduced form child models each with $V - 1$ predictors. Each child model is fitted by dropping one predictor from the parent model. Moreover, each child model involves dropping a different predictor variable. If the predictors of the child models are of full column rank, the regression weights within each child model are uniquely determined estimates of $(V - 1)$ differences $(\beta_v - \beta_{v*})$.

For a given set of V predictors, there are $V^*(V - 1)$ permutations of the predictors taken two at a time, because there are $V^*(V - 1)$ differences $(\beta_v - \beta_{v*})$. For a given child equation, $V - 1$ of these quantities will be estimated by the regression coefficients β_{vv*}. There will be V child equations each of which yields a different set of $V - 1$ differences $[\beta_{vv*} = (\beta_v - \beta_{v*})]$. The sets of regression weight estimates for the several children are mutually exclusive and exhaustive, so that the estimate of a difference $\beta_{vv*} = (\beta_v - \beta_{v*})$ appears in one and only one of the child equations. Taken together, the V child models each containing estimates of $(V - 1)$ differences, yield estimates of all $V^*(V - 1)$ differences $\beta_{vv*} = (\beta_v - \beta_{v*})$.

For a given value of v^*, the weights in the regression equation will be uniquely determined if $\boldsymbol{X}_{V-1}$ is of full column rank. However, the regression weight for a given predictor v (β_{vv*}) will vary depending on the sending variable v^* because $\beta_{vv*} = (\beta_v - \beta_{v*})$. A given predictor will appear in several child models, ($V - 1$ of the V child models to be exact) and its regression weight will vary systematically across those child models as the reference variable v^* varies. In words, the increase in the expected value of Y, given that one unit is transferred to variable v, depends on from which variable v^* that unit is transferred. As a result, variable v will have a different weight in each child model in which it appears, but this does not impair the interpretability of those weights, because the changes are readily interpreted in terms of the varying sending variables.

The familiar t-statistic can be used to test the hypothesis about regression weights in a reduced model: $H_0 : \beta_{vv*} = 0$. Given that $\beta_{vv*} = \beta_v - \beta_{v*}$, that same t-statistic provides a test of the hypothesis about weights in the full model: $H_0 : \beta_v - \beta_{v*} = 0$. The hypothesis test addresses the question of whether transferring one unit from v to v^* has an effect on the criterion variable.

Example: Reading Achievement Test Items

Our example is from a multiple-choice reading test (online supplementary materials). Whereas most multiple-choice tests have two kinds of answers, correct and incorrect, each item on the *Multiple-choice Online Causal Comprehension Assessment* (MOCCA) has three types of valid responses, one correct and two different incorrect responses. The edition of MOCCA used in this study has 40 items. Each MOCCA item consists of a seven-sentence paragraph with one sentence missing. From three alternatives, the student must choose the sentence to fill in the missing sentence that best completes the story. One answer, the causally coherent inference (CCI), is the correct answer. One incorrect answer, the paraphrase (PAR), is simply a paraphrase of earlier information in the story that adds no new information and does not advance or complete the story line. The second incorrect answer, the elaboration (ELA) involves an inference that injects new information based on background knowledge about the text, but does not complete the story line. If a student does not complete the test, then some items have a fourth response type labeled not attempted (NA).

For a given respondent, let X_1 be the number of CCI responses, X_2 be the number of PAR responses, X_3 be the number of ELA responses, and X_4 be the number of NA responses. Because CCI, PAR, ELA, and NA are mutually exclusive and exhaustive categories, $X_1 + X_2 + X_3 + X_4 = 40$ for each student. In this example, we use these four predictors to predict the scale score on a subsequent statewide test, the SBAC English Language Arts Test (Smarter Balanced Assessment Consortium, 2015, 2016). The regression used here is OLS linear regression. The respondents are 285 fourth graders from a convenience sample of fourth graders in schools scattered across the United States. Students completing less than 10 items were excluded. Here we are interested in the question of whether each of the three noncorrect answers (PAR, ELA, and NA) are equally indicative of proficiency.

Let Y_p be the total score that person p with predictor vector (X_1, X_2, X_3, X_4) will get on the SBAC English Language Arts Test. In our OLS linear regression model, the parent model is

$$Y_p = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + a \quad (7)$$

with the constraint that

$$X_1 + X_2 + X_3 + X_4 = 40 \quad \text{for all} \quad p.$$

The regression weights in Equation 7 are determined only up to an additive constant k. There are an infinite number of regression coefficient vectors $(\beta_1, \beta_2, \beta_3, \beta_4, a)$ that will fit the data equally well. Four child models with uniquely determined regression weights are those obtained by deleting one of the four predictors:

$$Y_p = \beta_{21} X_2 + \beta_{31} X_3 + \beta_{41} X_4 + a_1 \quad (8a)$$

$$Y_p = \beta_{12} X_1 + \beta_{32} X_3 + \beta_{42} X_4 + a_2 \quad (8b)$$

$$Y_p = \beta_{13} X_1 + \beta_{23} X_2 + \beta_{43} X_4 + a_3 \quad (8c)$$

$$Y_p = \beta_{14} X_1 + \beta_{24} X_2 + \beta_{34} X_3 + a_4 \quad (8d)$$

In this notation, β_{vv*} refers to the regression weight on variable v in the child equation with variable $v*$ deleted, and a_{v*} refers to the intercept in the child equation with predictor $v*$ deleted. Each coefficient β_{vv*} equals a difference between two coefficients in the parent equation: $\beta_{vv*} = \beta_v - \beta_{v*}$.

The regression coefficient estimates for the four models are in Table 1. Column 1 of Table 1 contains coefficients for the model with the CCI predictor deleted. Columns 2–4 correspond to the models with predictor variables PAR, ELA, and NA deleted, respectively. Each row label refers to the receiving predictor in the model, and the column label refers to the sending (deleted) variable. For each of the four models, the R^2 are the same, because the child models are four reparameterizations of the same model, the parent model of Equation 7.

To illustrate the interpretation of the regression coefficients, consider Column 1. The first entry is .00 in row CCI (the receiving variable) and column CCI (the sending variable). The .00 indicates that the increase in Y_p is .00 if one response from the sending variable CCI is transferred to the receiving variable CCI. All of the diagonal elements in Table 1 equal .00, because they represent the expected increase in Y_p obtained by transferring one response from a predictor to itself. The second element in Column 2 is −8.810, the increase (actually a decrease) in Y_p associated with the transfer of one response to the PAR predictor (row Paraphrase) from the CCI (column CCI) predictor. The third element, −6.943, the decrease in Y_p associated with transferring one response to the ELA predictor from the CCI predictor. Finally, −4.864 is the decrease in Y_p associated with transferring one response to the NA predictor from the CCI predictor. Except for the first element, all of the coefficients in model C are negative, suggesting that reducing the CCI predictor (correct response) by one unit decreases the predicted Y_p, but the size of the decrease varies depending on whether the unit goes to the PAR, ELA, or NA predictor (It appears that being correct is better than the alternatives as it leads to a higher expected value for the criterion, the SBAC English Language Arts Test).

The intercept under model C is 2,545.733. Given that the regression weight for the CCI predictor is .000, the intercept of 2,545.733 is the total score associated with a response vector in which all 40 responses are CCI responses: that is, the response vector $(X_1 = 40, X_2 = 0, X_3 = 0, X_4 = 0)$. In general, the intercept in a child model is equal to the predicted value for a response vector in which all responses fall in the sending predictor category.

Table 1
Linear Regression Weights for Four Models of Proficiency Varying in Sending Predictor Variable

Receiving predictor	Model and sending predictor			
	Model *C*	Model *P*	Model *E*	Model *NA*
Correct	0.000	8.810**	6.943**	4.864**
Paraphrase	−8.810**	0.000	−1.867	−3.945**
Elaboration	−6.943**	1.867	0.000	−2.078
No attempt	−4.864**	3.945**	2.078	0.000
Intercept	2,545.733**	2,193.342**	−2,268.024*	2,351.162**
R^2	0.530**	0.530**	0.530**	0.530**

Note. Model *C* = correct as sending predictor; Model *P* = paraphrase as sending predictor; Model *E* = elaboration as sending predictor; Model *NA* = not attempted as sending predictor.
* $p < .05$. ** $p < .01$.

In child models, the regression coefficient associated with a predictor will vary depending on which other predictor variable is deleted. For instance, the CCI variable has regression weight estimates of .000, 8.810, 6.943, or 4.864 in the four models of Table 1. This simply reflects the fact that the increase in Y_p associated with a one unit increase in the CCI predictor varies depending on whether that unit comes from the PAR, ELA, or NA predictor. Variation across the rows of Table 1 reflects the variation in Y_p associated with the different sending predictors given a fixed receiving predictor. Likewise, variation along the columns of Table 1 reflects the variation in Y_p associated with different receiving predictors given a fixed sending predictor.

The 4 × 4 matrix at the top of Table 1 has three interesting features characteristic of such matrices. First, any two columns are related by an additive constant *k*. That is corresponding elements in any two columns will be related by the same constant. For instance, if we subtract the first element in Column 1 (.000) from the first element in Column 2 (8.810), we get 8.810. If we subtract the second elements (−8.810 and .000) we also get 8.810. Similarly, corresponding elements in any two rows will differ by an additive constant (within rounding error).

Second, the 4 × 4 matrix at the top of Table 1 is a skew symmetric matrix. That is, for any two variables (*v*, *v**), $\beta_{vv*} = -\beta_{v*v}$. Corresponding elements in the upper and lower halves of the matrix are equal in magnitude but opposite in sign. This can be explained in terms of the regression coefficients in the parent equation. $\beta_{vv*} = \beta_v - \beta_{v*}$ whereas $\beta_{v*v} = \beta_{v*} - \beta_v$. For instance, the element in Row 2, Column 1 is −8.810, whereas the element in Row 1, Column 2 is 8.810. The element in Row 2, Column 4 is −3.945, whereas the element in Row 4, Column 2 is 3.945.

When there are *V* predictor variables, there are *V**(*V* – 1) permutations of *V* things taken two at a time, each permutation corresponding to a different $\beta_v - \beta_{v*}$. From the parent model, there are *V* child models, each corresponding to the deletion of a different predictor variable. Within each of the child models, there are *V* – 1 coefficients estimated. Each of the *V**(*V* – 1) permutations appears in one and only one child model, so that there are *V* models each with *V* – 1 of the permutations. In the off-diagonal elements of the 4 × 4 matrix of Table 1, each of the 4*(4 – 1) permutations $\beta_v - \beta_{v*}$ appears once and only once.

The significance tests in Table 1 are tests of the null hypothesis that $H_0 : \beta_{vv*} = \beta_v - \beta_{v*} = 0$. If the hypothesis cannot be rejected, the data fail to support the statement that the transfer of 1 point from predictor *v* to *v** changes the expected value of the criterion variable in the population. If the null can be rejected, then the statement receives support from the data.

In Model C, the regression weights are all significantly different from zero ($p < .01$) and negative, indicating that the regression weight for the correct answer CCI differs significantly from that for the PAR, ELA, and NA. Transferring one response from the correct response to any of the other categories is associated with a significant decline in proficiency. But are all of the incorrect answers created the same? The only significant difference between incorrect responses is that between PAR and NA (see Model P or Model NA). Transferring a point from PAR to NA (Model PAR) leads to a significant increase in the total score and, conversely, transferring a point from the NA to PAR (Model NA) leads to a significant decline. NA is more indicative of proficiency than is

PAR (it appears that not responding is more indicative of proficiency than paraphrasing).

The results above suggest that it may be overly simplistic to think of multiple-choice responses as falling into just two categories, correct or incorrect. The results above would suggest that a person whose 40 responses contained 35 correct answers and five paraphrase responses would not have the same expected outcome as someone with the same number of correct responses but five not attempted items. Supplied response items, either multiple-choice or rating scales, contain a set of mutually exclusive and exhaustive response categories, and usually there are more than just two categories. For example, for a multiple-choice item there are correct answers, incorrect answers, omitted items, and not reached items. Recent research has divided responses for an item into four categories: fast correct, slow correct, fast incorrect, and slow incorrect (De Boeck et al., 2017; Partchev & De Boeck, 2012; Su & Davison, 2019).[2] Other research suggests four different categories based on rapid guessing: fast effortful, slow effortful, fast noneffortful, and slow noneffortful (Rios, 2021; Wise & DeMars, 2009). There are new response formats, such as drag-and-drop or point-and-click. For supplied response items, including rating scales, criterion-related validity involves the study of how the composition of the person's test item response vector relates to the criterion variable. The results above illustrate a more fine-grained analysis of the criterion-related validity question. Results indicate that the student's outcome Y depends, not just on the number correct, but also on how incorrect responses are distributed over the incorrect response categories. In addition to the studies listed above, there is a literature in developmental psychology making use of the types of errors chosen using multiple choice item formats. For example, Powell (1968) found that wrong answers are not randomly distributed and that there is additional information in the type of wrong answers chosen.

Discussion and Conclusions

If the predictors add to a constant for all observations, then there are V*(V − 1) uniquely defined regression weight coefficients. However, they are not the coefficients in the ordinary linear model, the parent model. Rather they are coefficients in a family of V child models, each obtained by dropping one predictor (the sending variable) from the parent model. These uniquely defined parameters express the expected increase in the criterion resulting from a transfer of one unit from the sending predictor variable to the receiving predictor variable. The parameters quantify a type of interaction between two predictors, but one that cannot be modeled by a cross-product term $x_v x_{v*}$, because the interaction is a different kind of interaction involving a transfer of points rather than a moderating effect.

Collectively, the child models contain a large number of parameters: V*(V − 1). However, there are only (V − 1) degrees of freedom associated with the full set of V*(V − 1) parameters. From the (V − 1) parameters in any given child model, one can estimate the remaining (V − 1)*(V − 1) parameters from the coefficients of the given child model. Because of the fewer degrees of freedom, estimation may not pose the problems (for example, sample size) commonly associated with a large number of parameters. The large number of parameters raises concerns about the familywise error rate of statistical tests associated with the parameters. Given the dependencies between the parameters, much of the literature on familywise error rate may not apply, and familywise error rates must remain a matter for future research. The issue will need to be examined separately across the various types of regression (for example, OLS, ridge, logit, and so forth) to which the reduced form approach can be applied.

The reduced model of Equation 5 is a trade-off model. Its coefficients are regression weight differences, $\beta_v - \beta_{v*}$. Of special note is that these weight differences are relative to the parent model; values that are inestimable via usual OLS regression. These differences express the expected effect on Y obtained by trading one point from the sending variable $v*$ in exchange for one point on the receiving variable v. The model will be of particular interest in research areas, such as decision theory, where trade-offs are of interest in their own right.

If the researcher is fitting a linear model and the assumptions above do not hold; linearity with errors distributed independently and identically $N(0, \sigma^2(e))$, then the classical solution, fitting a reduced form OLS model does not apply. Nonlinearities may be accommodated by adding higher order terms or applying a generalized linear approach. Robust estimators may improve estimation of parameters, but do not address the violations of assumptions for the statistical tests (Hampel, 1971; Huber, 1973, 1981). There needs to be at least one reduced predictor matrix $\boldsymbol{X}_{V-1}$ of full column rank. Moderated versions of the child models can address the question of whether the effect of a trade-off varies in size across subgroups.

Compositional variables are of interest in many areas of psychology. For instance, group composition is of interest in social psychology, industrial organizational psychology (for example, employee teams), and education (for example, schools and classrooms). The composition of calorie intake (that is, fats versus carbohydrates versus proteins) is of interest as related to weight loss and adherence to diets. The composition of intervention activities is also of interest: for instance, clinical psychology (individual versus in-person group versus virtual group therapy) or education (reading group versus worksheets versus independent reading). Intervention design often involves decisions about allocating client and professional time to various activities that constitute the ingredients of an intervention.

In a reduced form solution, the regression weights are interpretable: β_{vv*} equals the expected increase in the criterion given the transfer of 1 point from the sending variable X_{v*} to the receiving variable X_v. This interpretation is more plausible for percentages than proportions. For proportions, a transfer of 1 point is only plausible if $X_{v*} = 1$ and $X_v = 0$. For predictor vectors, in which the proportion $X_{v*} \neq 1$ or $X_v \neq 0$, transferring one point yields a value of $X_{v*} < 0$ and a value of $X_v > 1$. Such values do not make sense, because they fall outside the range of proportions. For percentages, however, the premise of transferring one point is plausible in all but extreme cases in which $X_{v*} = 0$ or $X_v = 100$. For purposes of interpreting differences in regression weights, it seems more appropriate to base the regression on percentages rather than proportions, although percentages and proportions are perfectly linearly related.

Our results lead to several conclusions. First, the regression weights in the full model are determined only up to an additive constant. However, if there is a solution, differences between regression weights will be uniquely determined, precisely because those

[2] Items are classified as fast or slow prior to any regression analysis. One common way to do so is to classify the response as fast if the response time is less than the median response time and slow if the response time is greater than the median.

weights are determined up to an additive constant. If the reduced predictor matrix X_{V-1} in a child model is of full column rank, the coefficients in that model are uniquely determined and can be interpreted as differences between regression weights in the parent model. The familiar *t*-statistic can be used to test the hypothesis that regression weights in the child model $\beta_{vv*} = 0$ which is tantamount to testing the hypothesis $H_0 : \beta_v - \beta_{v*} = 0$ for pairs of regression weights in the parent model. Covariates (beyond the compositional variables) can be added without fundamentally changing the interpretation of the coefficients for compositional variables, except to make the interpretation conditional on the added covariates. However, issues that complicate interpretation in regression generally or causal modeling more specifically will carry-over into regressions based on reduced form equations (Allison, 1999; Didelez & Stensrud, 2021).

The major advantage of this approach over virtually all of the alternatives is that the magnitudes of regression weights are not only interpretable, but interpretable in the metric of the original predictors, and the interpretation is based on a comparison of expected criterion values for pairs of predictor vectors that both satisfy the constraint of equal sums. Furthermore, predictors can take any number on the real number line; they need not be positive. Methods using log transformations require positive predictors or ratios of predictors, and they yield regression weights whose magnitudes are difficult to interpret at best, even if one ignores the constraint that predictors sum to a constant. Reduced form models do not require specialized software. They can be applied in OLS linear regression, generalized linear regression, robust modeling techniques, and random coefficient models. Reduced form models are uniquely suited to many regression problems due to their breadth of applicability and the interpretability of their coefficients. OLS regression is fairly simple, very popular, and more universally understood. Furthermore, the weights in the metric of the predictors are consistent, unbiased, and minimum variance (Fox, 2016, pp. 109–110).

In many areas of research, there are important questions involving trade-offs (for example, fat versus protein versus carbohydrate calories; stocks versus bonds; eating out versus eating at home, work–life balance). Many of these trade-offs involve how people use their time or their money. Indeed, there are trade-offs between time and money (for example, more free time versus more income). Both people and organizations face trade-offs. Trade-off is a fundamental topic for any decision theory. Reduced form equations provide a statistical model for the analysis of such trade-offs. Reduced form models provide interpretable regression coefficients for any data where the sum of predictors is a constant for all observations, and they are particularly useful for research domains in which trade-offs are of interest in their own right.

References

Aitchison, J. (1982). The statistical analysis of compositional data. *Journal of the Royal Statistical Society. Series B. Methodological*, *44*(2), 139–177. https://doi.org/10.1111/j.2517-6161.1982.tb01195.x

Aitchison, J., & Bacon-Shone, J. (1984). Log contrast models for experiments with mixtures. *Biometrika*, *71*(2), 323–330. https://doi.org/10.1093/biomet/71.2.323

Aitchison, J., Barceló-Vidal, C., Martín-Fernández, J. A., & Pawlowsky-Glahn, V. (2000). Logratio analysis and compositional distance. *Mathematical Geology*, *32*(3), 271–275. https://doi.org/10.1023/A:1007529726302

Allison, P. (1999). Comparing logit and probit coefficients across groups. *Sociological Methods & Research*, *28*(2), 186–208. https://doi.org/10.1177/0049124199028002003

Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society. Series B. Methodological*, *26*(2), 211–234. https://doi.org/10.1111/j.2517-6161.1964.tb00553.x

Chen, J., Zhang, X., & Li, S. (2017). Multiple linear regression with compositional response and covariates. *Journal of Applied Statistics*, *44*(12), 2270–2285. https://doi.org/10.1080/02664763.2016.1157145

Davison, M. L., & Davenport, E. C., Jr. (2002). Identifying criterion-related patterns of predictor scores using multiple regression. *Psychological Methods*, *7*(4), 468–484. https://doi.org/10.1037/1082-989X.7.4.468

Davison, M. L., Davenport, E. C., & Jia, H. (2022). Linear equality constraints: Reformulations of criterion related profile analysis with extensions to moderated regression for multiple groups. *Psychological Methods*. Advance online publication. https://doi.org/10.1037/met0000430

De Boeck, P., Chen, H., & Davison, M. (2017). Spontaneous and imposed speed of cognitive test responses. *British Journal of Mathematical & Statistical Psychology*, *70*(2), 225–237. https://doi.org/10.1111/bmsp.12094

Didelez, V., & Stensrud, M. J. (2021). On the logic of collapsibility for causal effect measures. *Biometrical Journal*, *64*(2), 235–242. https://doi.org/10.1002/bimj.202000305

Egozcue, J. J., Pawlowsky-Glahn, V., Mateu-Figueras, G., & Barcelo-Vidal, C. (2003). Isometric logratio transformations for compositional data analysis. *Mathematical Geology*, *35*(3), 279–300. https://doi.org/10.1023/A:1023818214614

Fox, J. (2016). *Applied regression analysis and generalized linear models* (3rd ed.). Sage.

Greenacre, M., & Grunsky, E. (2019, January). *The isometric logratio transformation in compositional data analysis: A practical evaluation*. Economics Working Paper Series, Working Paper No. 1627 Universitat Pompeu Fabra. https://econ-papers.upf.edu/ca/paper.php?id=1627

Hampel, F. R. (1971). A general qualitative definition of robustness. *Annals of Mathematical Statistics*, *42*(6), 1887–1896. https://doi.org/10.1214/aoms/1177693054

Hron, K., Filzmoser, P., & Thompson, K. (2012). Linear regression with compositional explanatory variables. *Journal of Applied Statistics*, *39*(5), 1115 –1128. https://doi.org/10.1080/02664763.2011.644268

Huber, P. J. (1973). Robust regression: Asymptotics, conjectures and Monte Carlo. *Annals of Statistics*, *1*(5), 799–821. https://doi.org/10.1214/aos/1176342503

Huber, P. J. (1981). *Robust statistics*. Wiley. https://doi.org/10.1002/0471725250

Partchev, I., & De Boeck, P. (2012). Can fast and slow intelligence be differentiated? *Intelligence*, *40*(1), 23–32. https://doi.org/10.1016/j.intell.2011.11.002

Powell, J. C. (1968). The interpretation of wrong answers from a multiple choice test. *Educational and Psychological Measurement*, *28*(2), 403–412. https://doi.org/10.1177/001316446802800220

Rios, J. A. (2021). Is differential noneffortful responding associated with Type I error in measurement invariance. *Educational and Psychological Measurement*, *81*(5), 957–979. https://doi.org/10.1177/0013164421990429

Smarter Balanced Assessment Consortium. (2015). *English language arts & literacy computer adaptive test (CAT) and performance task (PT) stimulus specifications*. https://www.smarterbalanced.org/wp-content/uploads/2015/08/ELA-Stimulus-Specifications.pdf

Smarter Balanced Assessment Consortium. (2016). *Smarter Balanced Assessment Consortium: 2014-2015 Technical Report*. https://portal.smarterbalanced.org/library/en/2014-15-technical-report.pdf

Smithson, M., & Broomell, S. B. (2022). Compositional data analysis tutorial. *Psychological Methods*. Advance online publication. https://doi.org/10.1037/met0000464

Su, S., & Davison, M. L. (2019). Improving the predictive validity of reading comprehension using response times of correct responses. *Applied*

Measurement in Education, *32*(2), 166–182. https://doi.org/10.1080/08957347.2019.1577247

Wiernik, B. M., Wilmot, M. P., Davison, M. L., & Ones, D. S. (2020). Meta-analytic criterion profile analysis. *Psychological Methods*, *26*(2), 186–209. https://doi.org/10.1037/met0000305

Wise, S. L., & DeMars, C. E. (2009). A clarification of the effects of rapid guessing on coefficient α: A note on Attali's "Reliability of speeded number-right multiple-choice tests." *Applied Psychological Measurement*, *33*(6), 488–490. https://doi.org/10.1177/0146621607304655

Appendix

Proof of Theorem 1

The theorem above states that

Theorem 1: If $E(Y_p \mid \boldsymbol{x}_p) = \boldsymbol{\beta}\boldsymbol{x}_p^T + a$, then $E(Y_p \mid \boldsymbol{x}_p) = \boldsymbol{\beta}^*\boldsymbol{x}_p^T + a^*$ for all $\boldsymbol{x}$ satisfying the constraint $\sum_{v=1}^{v=V} x_{pv} = T$ if and only if there exists a constant k such that $\boldsymbol{\beta}^* = \boldsymbol{\beta} + k\mathbf{1}$ and $a^* = a - kT$.

In the text, it was shown that if $E(Y_p \mid \boldsymbol{x}_p) = \boldsymbol{\beta}\boldsymbol{x}_p^T + a$, then $E(Y_p \mid \boldsymbol{x}_p) = \boldsymbol{\beta}^*\boldsymbol{x}_p^T + a^*$ when there exists a constant k such that $\boldsymbol{\beta}^* = \boldsymbol{\beta} + k\mathbf{1}$ and $a^* = a - kT$ for all $\boldsymbol{x}$ satisfying the constraint. Here, we will show that, if there is not a single constant k relating the vectors, $(\boldsymbol{\beta}, a)$ and $(\boldsymbol{\beta}^*, a^*)$, then there will be at least one vector $\boldsymbol{x}$ such that $E(Y_p \mid \boldsymbol{x}_p) = \boldsymbol{\beta}\boldsymbol{x}_p^T + a \neq \boldsymbol{\beta}^*\boldsymbol{x}_p^T + a^*$. Let there be one predictor vector $\boldsymbol{x}_1 = (x_1, \ldots, x_{v'} \ldots, x_{v''}, \ldots x_V)$ such that $E(Y_p \mid \boldsymbol{x}_1) = \boldsymbol{\beta}\boldsymbol{x}_1^T + a = \boldsymbol{\beta}^*\boldsymbol{x}_1^T + a^*$, which means that

$$\sum_v \beta_v x_v + a = \sum_v \beta_v^* x_v + a^* \tag{A1}$$

Here we need only consider the case were there is at least one predictor vector such that $E(Y_p \mid \boldsymbol{x}_p) = \boldsymbol{\beta}\boldsymbol{x}_p^T + a = \boldsymbol{\beta}^*\boldsymbol{x}_p^T + a^*$, because if there is not even one predictor for which $E(Y_p \mid \boldsymbol{x}_1) = \boldsymbol{\beta}\boldsymbol{x}_p^T + a = \boldsymbol{\beta}^*\boldsymbol{x}_p^T + a^*$, then $E(Y_p \mid \boldsymbol{x}_p) = \boldsymbol{\beta}\boldsymbol{x}_p^T + a = \boldsymbol{\beta}^*\boldsymbol{x}_p^T + a^*$ cannot possibly be true for all predictor vectors satisfying the constraint.

Let $\boldsymbol{x}_2 = (x_1, \ldots, x_{v'} + 1 \ldots, x_{v''} - 1, \ldots, x_V)$. $\boldsymbol{x}_2$ and $\boldsymbol{x}_1$ differ only in their elements v′ and v″. Clearly, if $\boldsymbol{x}_1$ satisfies the constraint, $\boldsymbol{x}_2$ will as well. Let $\beta_{v\prime}^* = \beta_v + k_1$ and for v = v″ let $\beta_{v''}^* = \beta_{v''} + k_2, k_1 \neq k_2$. Then

$$\boldsymbol{\beta}^*\boldsymbol{x}_2^T + a^* = \sum_{v \neq v' \cdot v''} \beta_v^* x_v + \beta_{v'}^*(x_{v'} + 1) + \beta_{v\prime\prime}^*(x_{v''} - 1) + a^* \tag{A2}$$

$$= \sum_v \beta_v^* x_v + a^* + \beta_{v\prime}^* - \beta_{v''}^*$$

By a similar line of reasoning

$$\boldsymbol{\beta}\boldsymbol{x}_2^T + a = \sum_v \beta_v x_v + a + \beta_{v'} - \beta_{v''} \tag{A3}$$

If we subtract Equation A2 from Equation A3

$$(\boldsymbol{\beta}^*\boldsymbol{x}_2^T + a^*) - (\boldsymbol{\beta}\boldsymbol{x}_2 + a) = \left(\sum_v \beta_v^* x_v + a^* + \beta_{v'}^* - \beta_{v''}^*\right) - \left(\sum_v \beta_v x_v + a + \beta_{v'} - \beta_{v''}\right)$$

Substituting the left side of Equation A1 for $\sum_v \beta_v^* x_v + a^*$, $\beta_{v'} + k_1$ for β_v^*, and $\beta_{v''} + k_2$ for $\beta_{v''}^*$, yields

$$\left(\boldsymbol{\beta}^*\boldsymbol{x}_2^T + a^*\right) - \left(\boldsymbol{\beta}\boldsymbol{x}_2^T + a\right) = \left(\sum_v \beta_v x_v + a + \beta_{v'} + k_1 - \beta_{v''} - k_2\right) - \left(\sum_v \beta_v x_v + a + \beta_{v'} - \beta_{v''}\right) = k_1 - k_2 \neq 0$$

Thus, one arrives at the conclusion that, for the vector $\boldsymbol{x}_2$, $\boldsymbol{\beta}^*\boldsymbol{x}_2^T + a^*$ is not equal to $\boldsymbol{\beta}\boldsymbol{x}_2 + a$. There exists at least one vector $\boldsymbol{x}_2$ for which $\boldsymbol{\beta}^*\boldsymbol{x}_2^T + a^* \neq \boldsymbol{\beta}\boldsymbol{x}_2^T + a$, and therefore the statement $E(Y_p \mid \boldsymbol{x}_p) = \boldsymbol{\beta}\boldsymbol{x}_p^T + a = \boldsymbol{\beta}^*\boldsymbol{x}_p^T + a^*$ for all vectors $\boldsymbol{x}_p$ satisfying the constraint cannot be true when there is no single constant k relating $(\boldsymbol{\beta}, a)$ to $(\boldsymbol{\beta}^*, a^*)$.

Having shown in the text that the statement "$E(Y_p \mid \boldsymbol{x}_p) = \boldsymbol{\beta}\boldsymbol{x}_p^T + a = \boldsymbol{\beta}^*\boldsymbol{x}_p^T + a^*$ for all vectors $\boldsymbol{x}_p$ satisfying the constraint" is true when there is such a constant k, but that it cannot be true when there is no such constant k, one arrives at the conclusion in the theorem: if $E(Y_p \mid \boldsymbol{x}_p) = \boldsymbol{\beta}\boldsymbol{x}_p^T + a$, then $E(Y_p \mid \boldsymbol{x}_p) = \boldsymbol{\beta}^*\boldsymbol{x}_p^T + a^*$ for all $\boldsymbol{x}$ satisfying Equation 2 if and only if there exists a constant k such that $\boldsymbol{\beta}^* = \boldsymbol{\beta} + k\mathbf{1}$ and $a^* = a - kT$.

Received December 15, 2021
Revision received February 18, 2022
Accepted April 27, 2022 ■