Simpler Isn't Better: Ordinary Least Squares Regression Weights Make Better Predictions than Simple Alternative Weights in Realistic Simulations

By

Michael Cader Nelson

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Psychology

January 31, 2018

Nashville, Tennessee

Approved:

Joseph L. Rodgers, Ph.D.
Andrew J. Tomarken, Ph.D.
David J. Lubinski, Ph.D.
Bethany Rittle-Johnson, Ph.D.

To Marcie:

The strongest, most accomplished person I know.

ACKNOWLEDGEMENTS

I am indebted to my advisors, past and present, including my original advisor, Dr. David Cordray, and my co-advisors, Dr. Joseph Rogers and Dr. Andrew Tomarken. Dave, you showed the path and blessed me with opportunity and insight. Joe and Andy, your advice and guidance went far beyond what anyone would hope for a student you did not recruit. Your faith in me helped me to keep faith in myself, and ultimately got me across a finish line that seemed to never stop receding. If I can uphold the examples of these men in my scholarship and friendship, I can't help but be successful.

To my fellow quant students, there are few relationships that could match the warmth with which we supported have supported one another, as scholars and as friends. From inside jokes to impassioned debates to baby showers and weddings, you have been the only family members who could truly understand this part of my journey. We lifted one another to meet every challenge, no matter how high or deep. I especially acknowledge Cori Perkins, who made clear to me from the first day that I met her that she was a genuinely good person, and also an awful person—but in a great way. And who, through her conquest of obstacles and illnesses of every sort, became my hero.

Of my family—Phil and Janice, Brian and Scott, Tom and Kathy—I can only say that you saw more in me and wanted more for me than I ever did myself. My every worthy quality is due to your nurturing and encouragement, and my every fault is despite it. This is infinitely true of my amazing wife, who has truly earned every epithet: "long-suffering," "a blessing," "a saint." You are my best friend, and I am your biggest fan. I love you.

TABLE OF CONTENTS

Page

DEDICATION ii

ACKNOWLEDGMENTS iii

LIST OF TABLES v

LIST OF FIGURES vi

Chapter

I. Introduction 1

Alternative regression weights 2
Simulation study by Dana and Dawes (2004) 5
Critiquing Dana and Dawes (2004) 7
The present study 9

II. Overall Study Methods 11

III. Study 1 19

Study 1 Methods 19
Study 1 Results 20

IV. Study 2 30

Study 2 Methods 30
Study 2 Results 32

V. Study 3 38

Study 3 Methods 38
Study 3 Results 39

VI. Discussion 48

APPENDICES

Appendix Page

A. R Code for Simulations........55

B. Study 1 Matrices and Arrays of Mean Validated R Values........61

C. Study 2 Matrices and Arrays of Mean Validated R Values........67

D. Study 3 Matrices and Arrays of Mean Validated R Values........71

E. Adaptation of Maxwell's Formula for Estimating R2 from Model Parameters........77

REFERENCES........78

LIST OF TABLES

Table Page

1A. Comparison of Study 1 Results for 3 Predictors and R = .4 25

1B. Comparison of Study 1 Results for 3 Predictors and R = .7 26

1C. Comparison of Study 1 Results for 5 Predictors and R = .4 26

1D. Comparison of Study 1 Results for 5 Predictors and R = .7 27

2A. Comparison of Study 2 Results for 3 Predictors and R = .4 35

2B. Comparison of Study 2 Results for 3 Predictors and R = .7 35

2C. Comparison of Study 2 Results for 5 Predictors and R = .4 35

2D. Comparison of Study 2 Results for 5 Predictors and R = .7 36

3A. Comparison of Study 3 Results for 3 Predictors and R = .4 45

3B. Comparison of Study 3 Results for 3 Predictors and R = .7 45

3C. Comparison of Study 3 Results for 5 Predictors and R = .4 46

3D. Comparison of Study 3 Results for 5 Predictors and R = .7 47

LIST OF FIGURES

Figure Page

1. Simulation Designs 17

2. Relative performance of weight types as a function of equality of validities........... 21

3. Relative performance of weight types as a function of magnitude of validities....... 33

4. Relative performance of weight types with one or all but one predictor larger than the others, and validities positively or negatively associated 41

CHAPTER I

INTRODUCTION

More than a decade ago, Dana and Dawes (2004) published the results of a simulation study in which they compared the predictive accuracy of ordinary least squares (OLS) regression weights with alternative regression weights. Their results led Dana and Dawes to issue an unusually strong recommendation for when researchers ought to abandon OLS weights in their analyses and use equal or correlation weights instead: "[OLS] Regression coefficients should not be used for predictions unless error is likely to be extremely small by social science standards or sample sizes will be larger than 100 observations to predictors. In other words, *[OLS] regression coefficients should almost never be used for social science predictions*" (emphasis added, p. 328).

Any reader acquainted with the social science literature of the past decade will realize that few researchers have heeded Dana and Dawes's advice, with a few notable exceptions in fields such as Industrial-Organizational Psychology (e.g., Graefe et al, 2013; Hattrup, 2012), decision theory (e.g., Dana and Davis-Stober, 2016; German et al, 2016), and program evaluation (e.g., Greer, et al, 2016; Smolkowski and Cummings, 2015). Nevertheless, the result is that Dana and Dawes's (2004) study has been cited more than 100 times, including recent citations in major peer-reviewed journals (e.g., *Behavior Research Methods, Journal of Personality Assessment, Psychological Methods*). Their conclusions continue to be reviewed and promoted by alternative regression weight advocates. For example, Dana and Davis-Stober (2016) cite Dana and Dawes (2004) alongside earlier studies, and remark: "When making predictions using the ubiquitous linear regression model, decades of research have shown that when sample sizes are limited or measurement is noisy, the equal weighting of predictors is likely to perform as well as, or better, than standard regression estimates" (p. 66). Surprisingly, however, no study has attempted to replicate or extend Dana and Dawes (2004) and thus their conclusions have remained unchallenged in the literature.

Methodological practice and literature, therefore, are commonly at odds. The true state of the world is either that Dana and Dawes's forceful recommendation is correct and applied researchers have stubbornly failed to adopt superior statistical methods, or that Dana and Dawes's argument does not hold (at least in some circumstances) and awaits the identification of one or more flaws that limit their recommendations. In either case, a re-evaluation and update to the literature is due. The present study revisits their simulation, from its conceptual basis to its methods to its results, implements some important revisions, and provides evidence through new simulations that favors continued skepticism of alternatives to OLS regression weights.

Alternative regression weights

One of the most frequent and impactful ways that quantitative methods researchers translate their work for real-world application is by providing statistical models as a predictive tool for use by other researchers, practitioners and policymakers. The parameters in these models are sometimes estimated from large and reliable databases but more often are drawn from smaller convenience samples prone to high levels of statistical noise. The rationale for using alternative regression weights in these models rather than ordinary least squares (OLS) weights rests largely upon a great limitation of OLS weights: weights estimated from a single sample are calibrated to fit not only characteristics of the data that are descriptive of the population parameters (and therefore useful for prediction) but also characteristics that are specific to the sample (and therefore irrelevant noise adding error to predictions). The problem grows as we try to get information about more parameters from fewer subjects: As the ratio of subjects in the calibration sample to predictors decreases and/or the predictability of the model (i.e., the population coefficient of multiple correlation, *R*) diminishes[1], the upward bias of *R* will grow substantially (Dana and Dawes, 2004). This phenomenon will lead to shrinkage of the *R* when the model is applied to new samples, a threat to predictive validity.

We often consider calculating adjusted or shrunken *R* as a solution, but adjusted *R* is dependent upon the same unique sample characteristics. Moreover, adjusted *R* only describes how much the present estimated model resembles the population model, not how well the present model will make predictions with future, independent samples. The ideal adjustment would alter the actual values of the model coefficients themselves, thus directly adjusting future predictions. This is exactly what alternative weighting achieves.

The suggestion that alternative weighting methods may be preferable in some circumstances to OLS weighting in regression dates back at least seventy-five years (Wilks, 1938), but the more mathematically principled approach of using OLS clearly won out. Wainer (1976) hoped to bring attention to earlier discussion of the weaknesses of OLS regression (Gulliksen, 1950; Rozeboom, 1966; Goldberger, 1968; Schmidt, 1971) by proving mathematically that equal weights yield superior predictability compared with OLS weights across samples in a wide range of cases.[2] His argument and conclusion were largely rejected

[1] The coefficient of multiple correlation is the square root of the more familiar coefficient of determination, R^2. I reference the former here and throughout to be consistent with the methods and reported results of Dana and Dawes (2004).
[2] Wainer's robust championing of unit weights provoked some criticism, prompting him to publish a brief paper (Wainer, 1978) in which he acknowledged minor errors and potential limitations but also asserted that his argument was unaffected in the vast majority of realistic scenarios.

(see, in particular, Laughlin, 1978), but the factors he identified as making equal weights and unit weights[3] so attractive remain undisputed: beyond their simplicity, they are independent of the sample data and so cannot be influenced by the idiosyncrasies of the sample.

Additional support for unit and equal weights has been provided through empirical and simulation studies (Dawes and Corrigan, 1974; Dawes, 1979; Raju, Bilgic, Edwards & Fleer, 1999; Dana and Dawes, 2004). More recent work has described unit weighting as one of an infinite variety of "constrained estimators," where other examples include rank weighting and "take the best" weighting (Davis-Stober, Dana and Budescu, 2010; Davis-Stober, 2011)[4]. The performance of each type of constrained estimator is determined by the underlying factor structure of the data, although unit weights have the lowest maximum mean squared error and so may be preferable when the precise structure is not known.

Correlation weights (the zero-order correlations between each predictor and outcome, also called validity weights) likewise have been demonstrated to produce regression models with coefficients of multiple determination (R^2) less likely to shrink when applied outside of the initial sample (Campbell, 1974; Dana and Dawes, 2004; Davis and Sauser, 1991; Goldberg, 1972; Dunnette and Borman, 1979) and may produce more stable predictions (as assessed by the root mean squared error) than an OLS regression model (Finch et al., 2011). Waller and Jones (2010) attempted to lay a foundation for a mathematical theory of correlation weights and when they are

[3] Unit weights are either -1, 0, or +1, whereas equal weights are all of the same sign. Wainer (1978) showed that including weak predictors in an equal-weights model can decrease predictability and so recommends their exclusion, essentially assigning a weight of 0. For the remainder of the paper, the term "unit weight" includes equal weights.

[4] To be clear, while these are clever and generally useful studies, I don't describe them fully because they don't quite apply to the issues at hand. First, I will include correlation weights in my simulation, but these studies are only helpful for choosing among the different types of *fixed* weights. Second, their methods assume knowledge of the underlying factor structure, a much stronger requirement than correctly specifying the regression model.

likely to be more or less predictive than OLS weights, but this preliminary work was in terms of population regression models. A general solution for when correlation weights will be preferable to OLS weights for a given set of predictors and criterion was not found. Their work did underline the complexity of the problem of selecting a priori the most efficient weight type: With little effort, they were able to construct a set of independent variables that yielded a coefficient of determination of .99 for a particular criterion using either OLS or correlation weights, whereas the same predictors with a second criterion saw a coefficient of only .09, even as the predictability of OLS weights remained unchanged. Surprisingly, these two criteria were correlated .98. Many other alternative weighting methods for regression exist, but unit weighting and correlation weighting are the most widely studied and most frequently used (Dana and Dawes, 2004; Waller and Jones, 2010; Davis-Stober, 2011).

Simulation study by Dana and Dawes (2004)

Dana and Dawes (2004) designed their simulation after conducting several resampling studies of five public datasets. As their metric for comparing the performance of the different types of regression weights, they used *validated R*, which is the correlation between criterion estimates generated with the sample weights and the population weights. For each data set, they drew repeated samples, calculated the *validated R* for the resulting regression weights for each sample, then reported the average of *validated R* across all samples[5]. They concluded that OLS

[5] Validated R, as calculated by Dana and Dawes (2004), is a function of the estimated regression weights from a sample, and the population weights and predictor correlation matrix. The parameter values used for the empirical data sets were estimated from the total data, while those used in the Monte Carlo simulation were specified for each condition of the simulation. See Chapter II for a full description of the formula and its use.

regression performs best relative to alternatives[6] when the population R (ρ) was large, and the size of ρ had a much greater impact than did the ratio of predictors to sample size (possibly due to differences between the models as specified and the true "models" underlying the real-world data). When this ratio and ρ were small, unit and correlation weights always performed best, and there was little loss of predictability in general even when these factors increased and OLS regression performed best.

Their Monte Carlo design used simulated data with multiple levels of sample size, number of predictors, population error and multicollinearity. The weights compared were OLS, correlation, unit and two other alternative weights suggested in the literature.[7] Performance was compared on the basis of average validated R and also reported in terms of estimated ρ. One difference in the results compared with the performance of weights using empirical data sets was that rho no longer had a greater impact on weight performance than did sample size, likely because all models used in the Monte Carlo study were correctly specified by design.

A perfect regression model would have a validated R equal to the population R[8], indicating that it is yielding as much information about the criterion as the predictors can possibly reveal. Comparing OLS weights with correlation and unit weights, Dana and Dawes (2004) found that, on average, correlation weights produced a larger mean validated R when the population predictability was greater than .6, and both OLS and correlation weights were bested

[6] These were correlation weights, unit weights, "take the best" weights (Gigerenzer and Todd, 1999) and the OLS model with the best Mallows Cp (Mallows, 1973).
[7] These were "take the best," which uses only the predictor most highly correlated with the criterion, and ridge regression. Ridge regression reduces multicollinearity and shrinks R by adding a small amount k to the diagonal values in the predictor correlation matrix. Neither method consistently performed better than OLS, equal and correlation weights in the simulation, and neither factored into the conclusions or recommendations of the study.
[8] The population R is the maximum value of the validated R, otherwise we would be claiming that the model is divining more information from the sample than actually exists in the population.

by unit weights when $\rho < .4$, *even with as many as 100 subjects per predictor*. These findings led to the following practical recommendation:

> [OLS] Regression coefficients should not be used for predictions unless error is likely to be extremely small by social science standards [rho > .6] or sample sizes will be larger than 100 observations to predictors. In other words, regression [OLS] coefficients should almost never be used for social science predictions. Simple alternatives will usually yield better predictions. (Dana and Dawes, 2004, p. 328)

Critiquing Dana and Dawes (2004)

As noted above, the Dana and Dawes (2004) study has not been replicated or extended, and subsequent references to their findings have been generally supportive. Nonetheless, I will describe several apparent weaknesses and ambiguities in their study design that bolster the argument for replication and reassessment. First, a major obstacle to full understanding of their methods is that the description of the design and requisite assumptions upon which it depends is lacking precise detail. For example, Dana and Dawes stated that their procedure for simulating random data were an adaptation of the methods introduced by Wherry et al. (1965), yet the methods described in Wherry et al. were far too simple to produce the results in Dana and Dawes. Wherry's method took as input a single set of factor loadings, whereas Dana and Dawes simulated results across a continuous range of values for each predictor, collinearity, and predictability. The adaptation made for their study was not described in sufficient detail to be fully understood or replicated.

Dana and Dawes reported that the patterns in their results were not "sensitive" (p. 326) to differences in absolute sample size (but were dependent on the ratio of sample size to number of predictors) and variation in predictor collinearity (pairwise correlations among predictors) and validity (pairwise correlations between the criterion and each predictor). However, they did not present to the reader any evidence illustrating such lack of impact, only stating that including these factors would not have changed their conclusions. This lack of information prevents the reader from making his or her own assessment of the data or of the authors' judgment concerning the data, an especially important point considering that the data that were summarized in the paper averaged across these factors. Dana and Dawes (2004) may thus have inadvertently obscured important variation between or interactions among factors, a problem compounded by their reporting these means without standard deviations or other indications of variability[9]

Samples assessed in the study were simulated from population correlation matrices with factor loadings sampled from m-dimensional Dirichlet distributions (where m is the number of regression model parameters), and error was scaled using a random draw from a uniform distribution of values, (0,.5). In this way, Dana and Dawes (2004) were able to address questions of regression weight performance across a continuous spectrum of all possible populations and all possible samples from those populations, maximizing the generalizability of their findings. An implicit assumption of this approach, however, is that all conceivable relationships among variables are equally likely. In reality, researchers usually choose their study variables from a much narrower range of possible joint distributions (with the possible exception of extreme

[9] The overall impact of both population predictability and sample size was compared using effect sizes, which account for the standard deviation, but only on the omnibus level. That is, it appears that they tested whether there were any significant differences in weight performance across all the levels of a factor (e.g., all ratios of subjects to predictors) but did not compute pairwise contrasts between levels of a factor (e.g., between 5 subjects per predictor and 100 subjects per predictor).

datamining), using some a priori knowledge of the respective parameters, and in any event correlations among social science variables tend not to demonstrate such broad ranges of variability.

The present study

Research on alternative weights has continued, from using fungible weights to assess coefficient sensitivity (Waller, 2008) to using random weights as the baseline for quantifying the relative information content of a regression model (Davis-Stober and Dana, 2013). Most of this work has been either analytic/theoretical, or attempts to characterize the geometry underlying alternative regression weight performance (rather than generating practical guidance). This leaves Dana and Dawes (2004) as the only concrete guide for alternative weights, but their analysis had a major drawback: they simulated weight performance across almost the entire spectrum of possible regression studies, then averaged results across factors. These results were the basis for conclusions that, although true in the theoretical universe of all possible studies, could still problematic for the narrower statistical space in which an actual researcher operates. A more realistic simulation design could reasonably consider a more limited range, with focus on plausible parameter values rather than the entire universe of values. The present study modifies and simplifies their simulation design by narrowing the range of each factor (i.e., fewer factor levels) so that I can describe the variability in results rather than averaging across cells. This allows me to directly examine the impact of specific patterns of parameter values on the results, varying the equality, magnitude and spread of collinearities and validities. Should the results of my simulations vary meaningfully from those of Dana and Dawes, this finding would suggest that a more skeptical view of alternative regression weights as a replacement for OLS weights may be appropriate.

CHAPTER II

OVERALL STUDY METHODS

Dana and Dawes (2004) may have obscured some of their important findings by averaging across the effects of variation in validities, as well as across variation in predictor collinearities, within their population correlation matrices. They reported that these factors did not significantly impact weight performance in their results, but across such a wide range of values for all the simulation parameters, significant differences with respect to these factors could easily have been overshadowed or even cancelled out. A second issue is that, by reporting mean validated *R*'s but not their standard errors, Dana and Dawes (2004) omitted an important measure of relative precision. Even if alternative weights do provide an advantage in predictability over OLS weights in the long run, the variability of that advantage could diminish its impact in the short run.

The present study addressed the latter concern by including confidence intervals around the mean validated *R*. The result shows not only the average validated *R* yielded across all samples drawn from a given population but also the range in which the validated *R* would fall ninety-five percent of the time. Assessing the relationship between weight performance and population correlation matrix parameters is more complex, requiring the systematic variation of both the set of predictor validities (i.e., correlations between each predictor and the criterion) and their collinearities (i.e., the correlations among predictors) of simulated populations. For example, validities may be more or less equal to one another, as may collinearities; the average magnitude of validities and collinearities may be higher or lower; and the distribution of validities may be either a) evenly spaced, b) skewed so that the validities are positively associated with collinearities (i.e., the predictor with the greatest validity also has the greatest multicollinearity with other predictors, and the predictor with the least validity also has the least multicollinearity with other predictors), or c) skewed so that the validities are negatively associated with collinearities. A thorough examination is only possible if the simulated samples are drawn from populations that reflect multiple levels of each factor in combination.

Yet, as previous simulation studies have implied, it is computationally impractical to simulate thousands of pre-defined populations and then analyze the results for each population matrix, all without averaging across matrix characteristics. The alternative approach used in the present study is to surrender some of the generalizability afforded by simulating thousands of different populations, but achieving a closer examination of specific, plausible population structures. The parameter values in the population correlation matrices have been restricted to a plausible range for a typical study. The different levels of each simulation condition use parameter values that are at either end of that plausible range. For example, weight performance is compared when validities are nearly equal and when they are very different, as opposed to examining results for hundreds of populations with validities across a continuous range between equality and extreme inequality.

An advantage of this approach is that it can be matched to realistic research scenarios, where a researcher is working with a particular population, often with characteristics that can be partially anticipated a priori. The rationale for selecting each specific population matrix is given below, but each is a variation on a matrix with average validities near .3 and all collinearities averaging .3. This combination is consistent with what has been found to be the most common magnitude for correlations in the social sciences (c.f., Maxwell, 2000). An additional constraint on validities is that none are less than .1, the minimum used by Waller and Jones (2011). Smaller validities are more likely to result in suppression effects (i.e., the multiple R is increased by including an independent variable that is significantly correlated with other predictors but not with the criterion), which were specifically excluded by Dana and Dawes (2004) from their simulations.

When specifying population correlation matrices for simulation studies and generating sample correlation matrices, it is important to assess whether matrices are mathematically admissible. There are constraints on correlation matrices such that not all patterns of correlations that one might specify can actually be observed (Marsaglia and Olgin, 1984; Holmes, 1991). To solve this problem, I use a formula adapted by Andrew Tomarken (personal communication) from one first presented in Maxwell (2000): although the original formula could be used to calculate R^2 from population correlation matrices, the adapted formula below allows one to specify the parameter values for predictor collinearities and R^2 and analytically derive validities that satisfy certain requirements.

$$\boldsymbol{\rho}_{XY} = \mathbf{v}\sqrt{\frac{R^2}{\mathbf{v}'\mathbf{R}_{XX}^{-1}\mathbf{v}}}$$

where $\boldsymbol{\rho}_{\mathbf{xy}}$ is the vector of criterion-predictor correlations, R^2 is the squared coefficient of multiple determination, $\mathbf{R}_{\mathbf{xx}}$ is the population predictor correlation matrix, and **v** is a vector of values equivalent to the ratios of predictor validities (e.g., if the second of two predictors correlates half as much as the first with the criterion, the values in **v** would be 1 and .5). In this way, I insure the population model for each regression is valid and has the characteristics I require for testing my research questions.

As an example, if I wish to generate a design matrix with three predictors, the population R^2 = .5, all predictors correlate .5 with each other, and all predictors correlate equally with the criterion (the equal validities condition), I must solve for:

$$\boldsymbol{\rho}_{XY} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \sqrt{\frac{.5}{\begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1.5 & -.5 & -.5 \\ -.5 & 1.5 & -.5 \\ -.5 & -.5 & 1.5 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}}} = \begin{bmatrix} .58 \\ .58 \\ .58 \end{bmatrix}$$

In this case, each of the predictors has a validity equal to the square root of one third, or approximately .58, and the final correlation matrix for Y, X1, X2, and X3 is:

$$\begin{bmatrix} 1 & & & \\ .58 & 1 & & \\ .58 & .50 & 1 & \\ .58 & .50 & .50 & 1 \end{bmatrix}$$

where the first column and row are validities and the interior values are collinearities.

For most of my simulations, I need validities to have a target average value (e.g., .3) and to adhere to target ratios (e.g., the second predictor should be twice as large as the first, while the third is three times as large). Due to the constraints placed on valid matrices, these target values are only approximated in the population matrices that are used in the simulations. This difference between targeted and generated sets of parameters is not a problem, however. Because the parameter values still represent relative extremes within a plausible range, they are effective in testing the impact of realistic parameter variability on regression weight performance. (See Appendix E for further details.)

Once population correlation matrices are selected for each combination of factors, samples are simulated using the method originally described by Kaiser and Dickman (1962). The advantages of this method are, first, that it is a widely known and frequently used simulation method, second, that it only requires as input a single correlation matrix to produce simulated sample datasets, and third, it is kernel of the method Dana and Dawes (2004) used in their simulation (albeit with complex extensions). I am thus able to create data with the specific properties I wish to test. All of these steps in the simulation are incorporated into the R programs in Appendix A, which accomplishes the following:

1. Takes as input the desired characteristics for a population predictor-criterion correlation matrix (i.e., population multiple *R*, the number of independent variables, and the population inter-correlations among predictors);
2. Generates a population predictor-criterion correlation matrix using the equation adapted from Maxwell (2000);
3. Generates data representing 10,000 independent and random samples from the population;
4. Saves population and sample correlation matrices, and confirms that they are admissible correlation matrices (i.e., they are positive definite);
5. For each sample, calculates the estimated regression coefficients for each type of weight (OLS, correlation and equal weights);
6. Calculates the validated *R* statistic for every set of weights in every sample;
7. Repeats the above steps for each desired sample size; and
8. Plots the results (average validated R by weight type and sample size) for comparison with the results given by Dana and Dawes (2004), with the addition of confidence intervals around validated *R* (calculated as the interval between the 2.5 and 97.5 percentiles from the simulated distribution).

The method and results are divided into three studies, each involving a different set of population correlation matrices that assess the effects of a different set of simulation factors. Study 1 compares results for combinations of equal and unequal collinearities and validities. Study 2 compares higher and lower magnitudes of validities. Study 3 compares variations of validity skewness and association with collinearities. Each study also includes levels of factors varied by Dana and Dawes (2004): validated R (.4 or .7), number of predictors (three or five), and number of observations per predictor (5, 10, 15, 20, 30, 50, 75, or 100). The number of levels of validated R and predictors was reduced from their study for computational tractability, and in order to focus on the factors not included. See Figure 1 for a complete diagram of the design of all three studies.

As in Dana and Dawes (2004), I use *validated R* as the measure of the performance of regression weights. The maximum value of *validated R* for a regression model is the population *R*, which signifies how much covariance in the population can potentially be described by the model coefficients. A set of weights with a *validated R* nearer to the population *R* therefore can be interpreted as outperforming a set with a lower value. *Validated R* is calculated using the following formula:

$$validated\ R = \frac{\mathbf{w}'\mathbf{v}}{\sqrt{\mathbf{w}' \sum \mathbf{w}}}$$

where **w** is a vector of regression weights produced by the simulation, **v** is the vector of population correlations between predictors and criterion (i.e., true validities), and $\sum$ is the population correlation matrix among predictors (i.e., true collinearities). Because the only empirical estimates in the equation are the regression weights, and equal regression weights are

Study 1 design: 2 predictors X 2 *R* X 2 Collinearities X 2 Validities = 16 graphs, eight ratios of sample size to #predictors per graph = 128 total cells

Three Predictors								Five Predictors							
R = .4				*R* = .7				*R* = .4				*R* = .7			
Equal Collinearity		Unequal Collinearity		Equal Collinearity		Unequal Collinearity		Equal Collinearity		Unequal Collinearity		Equal Collinearity		Unequal Collinearity	
Equal Validity	Unequal Validity	Equal Validity	Unequal Validity	Equal Validity	Unequal Validity	Equal Validity	Unequal Validity	Equal Validity	Unequal Validity	Equal Validity	Unequal Validity	Equal Validity	Unequal Validity	Equal Validity	Unequal Validity

Study 2 design: 2 #predictors X 2 *R* X 2 validity magnitudes = 8 graphs, eight ratios of sample size to #predictors per graph = 64 total cells

Three Predictors				Five Predictors			
R = .4		*R* = .7		*R* = .4		*R* = .7	
Lower validities	Higher validities	Lower validities	Higher validities	Lower validities	Higher validities	Lower validities	Higher validities

Study 3 design: 2 #predictors X 2 *R* X 2 validity spreads = 8 graphs, eight ratios of sample size to #predictors per graph = 64 total cells

Three Predictors								Five Predictors							
R = .4				*R* = .7				*R* = .4				*R* = .7			
1 dominant validity		2 dominant validities		1 dominant validity		2 dominant validities		1 dominant validity		4 dominant validities		1 dominant validity		4 dominant validities	
PA*	NA*	PA*	NA*	PA*	NA*	PA*	NA*	PA*	NA*	PA*	NA*	PA*	NA*	PA*	NA*

*PA = The progression of validities is positively associated with predictors' respective multicollinearities, so that the predictor with the largest validity has the largest collinearity, and validities increase with multicollinearities.

NA = The progression of validities is negatively associated with predictors' respective multicollinearities, so that the predictor with the largest validity has the smallest collinearity, and validities decrease as multicollinearities increase.

Figure 1. Simulation Designs.

always the same by definition, the estimate of validated *R* for equal weights will have no variability and therefore no confidence interval when graphed.

Results in all studies are graphed in a manner similar to Dana and Dawes (2004), with cases per predictor on the x-axis and mean validated *R* on the y-axis. One notable difference, however, is that the cases per predictor in their graphs are equally spaced along the x-axes, whereas my x-axes are ratio scales and thus reflect the actual magnitude of the cases per predictor values. The reason for this change is that the original scales give the impression of a more dramatic improvement in weight performance as a function of sample size than was actually the case. Another difference in the graphs is the addition of 95 percent confidence bars for each mean, calculated empirically from simulations, to indicate relative variability of weight performance.

CHAPTER III

STUDY 1

Study 1 Methods

The initial study was designed to evaluate how varying validities and collinearities between equality and inequality can impact relative weight type performance when all variables correlate around .3. I created a set of population matrices with validities that are either equal or unequal, and with collinearities that are either equal or unequal. Crossing these factors, in combination with two levels of population multiple *R* (.4 and .7, representing lower and higher predictabilities) and two levels of number of predictors (three and five), resulted in sixteen distinct populations. (Tables representing the design cells are below; their respective population matrices, mean validated R values, and graphed results are in Appendices II-IV.) For each population, I simulated 10,000 samples for each of the eight levels of cases per predictor (5, 10, 15, 20, 30, 50, 75, 100), or 16 populations X 8 cases per predictor X 10,000 replications = 1,280,000 samples total.

As examples from Study 1, the matrices used for the equal validities and equal collinearities (left), and unequal weights and unequal collinearities (right), with three predictors and R = .4 are shown below (leftmost column of each matrix contains validities):

	Y	X_1	X_2	X_3
Y	1.00			
X_1	0.29	1.00		
X_2	0.29	0.30	1.00	
X_3	0.29	0.30	0.30	1.00

	Y	X_1	X_2	X_3
Y	1.00			
X_1	0.13	1.00		
X_2	0.26	0.20	1.00	
X_3	0.38	0.30	0.40	1.00

It should also be noted that the validities in Study 1, when unequal, were generated to be close to evenly distributed (e.g., [.13, .26, .38]) and the predictor with the largest validity also had the highest multicollinearity. This means that results in this phase are biased toward correlation weights, although not to great effect[10]. The impact of varying the distribution of validities, both in terms of spread and collinearity, is investigated in Study 3.

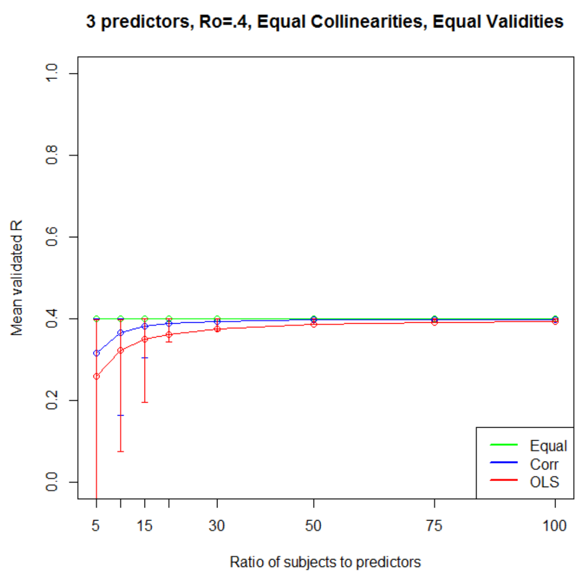
Study 1 Results

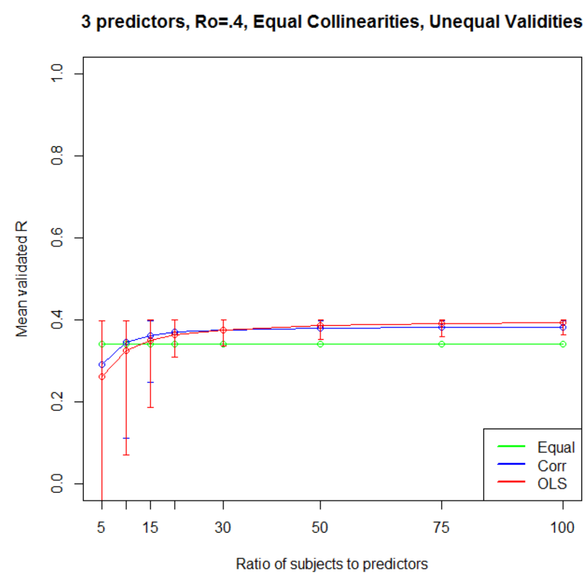
Overall, graphs of study 1 results reveal that varying population validities has far more impact on weight performance than varying collinearities. In each graph, equal weights (green) dominate correlation (blue) and OLS (red) weights whenever the validities are also equal, regardless of the other factors (sample size, number of predictors, validated R, and collinearity). This makes sense given that equal weights have zero error, while weights estimated from the sample will always have some error and so would never be all equal. When validities are unequal, the performance of equal weights is much worse, and they are never the best choice when there are at least ten subjects per predictor.

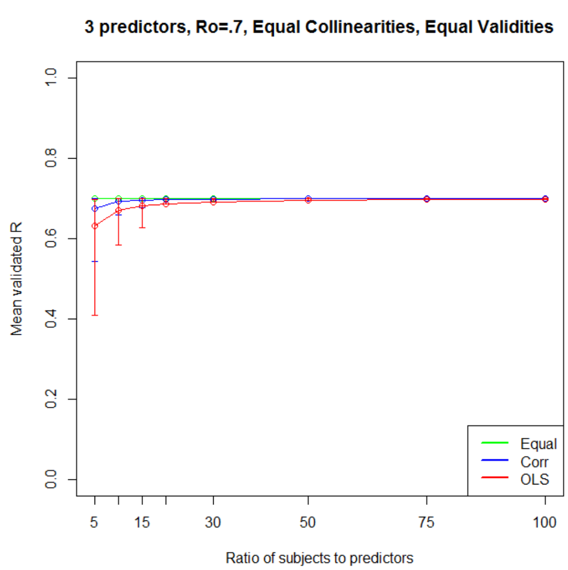
The result most consistent with Dana and Dawes (2004) is that correlation weights perform better than OLS weights at the very lowest ratio of 5 subjects per predictor at the lower predictability of $R = .4$, and frequently perform at least as well as OLS weights when there are 10 or 15 subjects per predictor. Even at the higher predictability of $R = .7$, the advantage of OLS weights is small and its confidence intervals tend to be wider compared with those of correlation

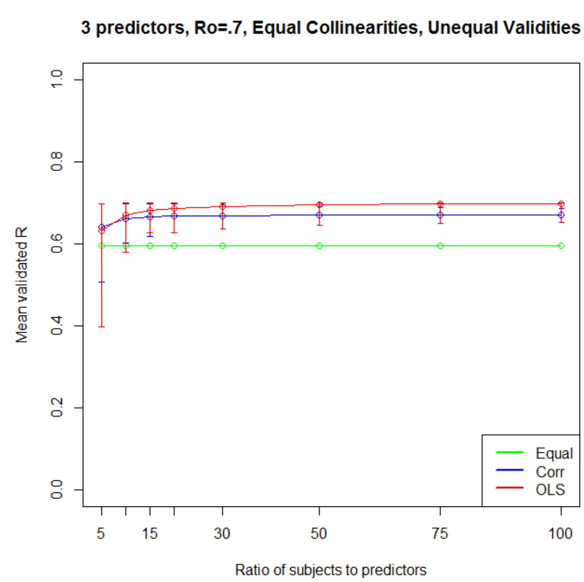
[10] When I reversed collinearity for these models, the same sets of factors favored correlation weights over OLS weights, but to a lesser extent.

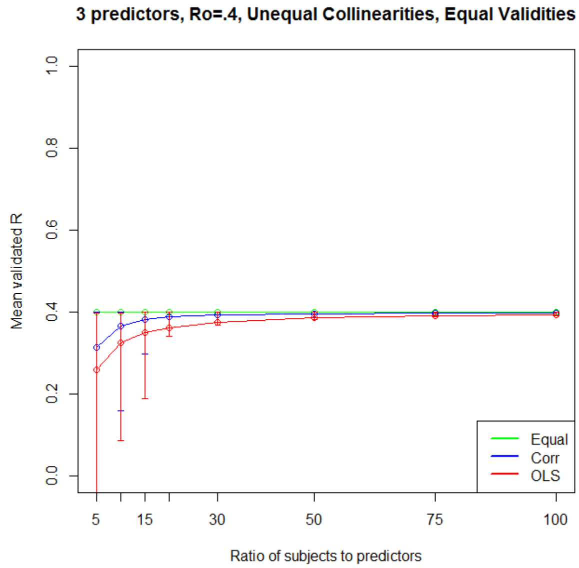
weights. Of course, these results should be given a limited interpretation: the design of Study 1 represents only one way of varying model parameters, while other designs might yield different results. Fifteen subjects per predictor is far fewer than the 100 subjects per predictor standard issued by Dana and Dawes (2004), however, and OLS weights ultimately perform as well as or better than correlation weights.

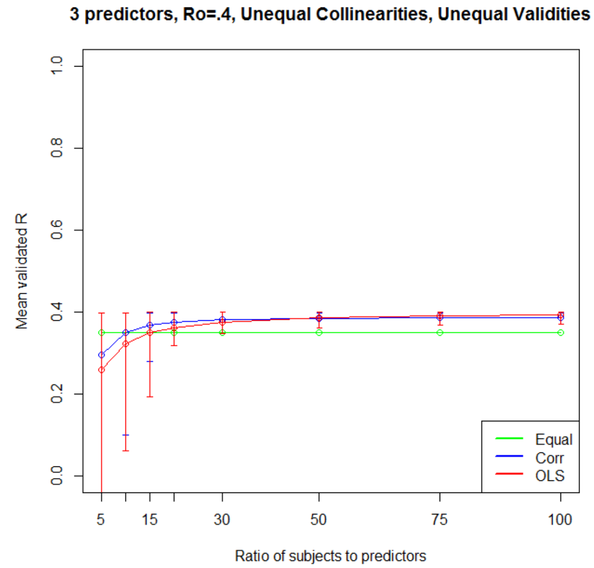


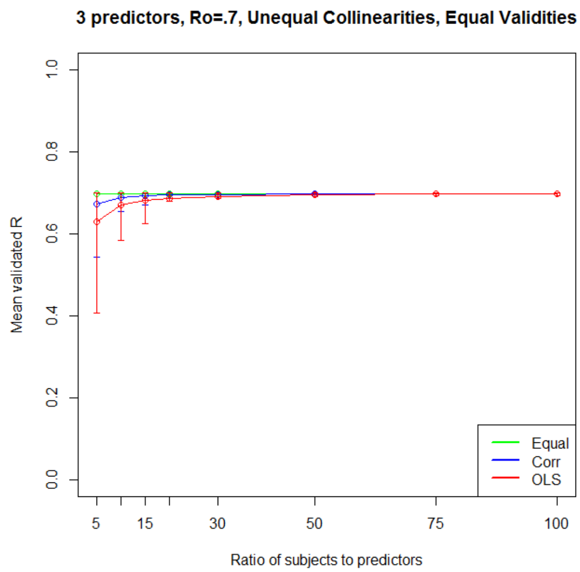


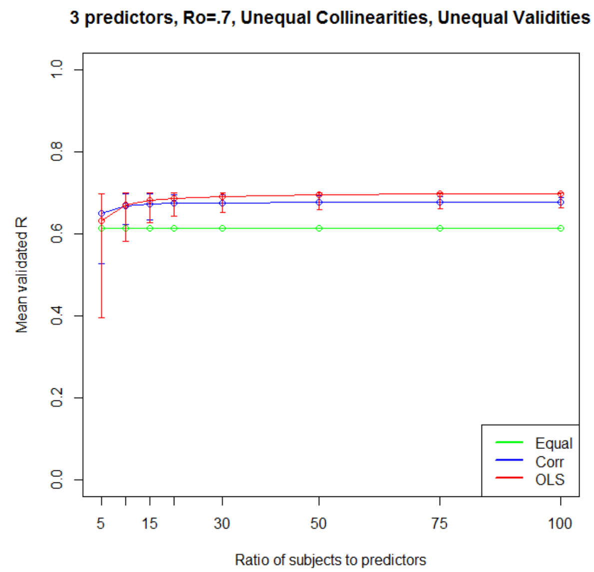




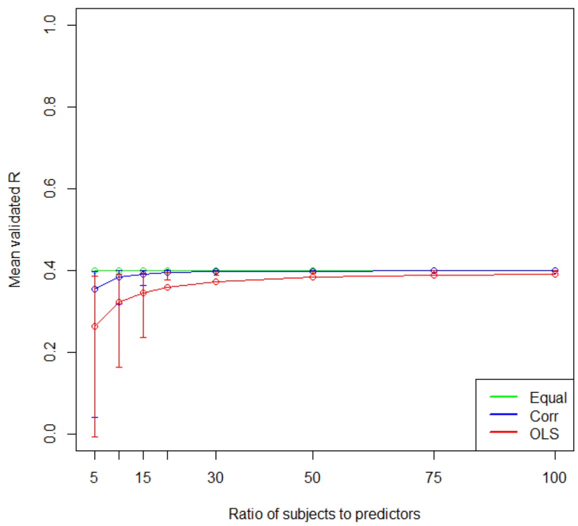

3 predictors, Ro=.4, Unequal Collinearities, Equal Validities
Mean validated R
0.0
0.2
0.4
0.6
0.8
1.0
5
15
30
50
75
100
Ratio of subjects to predictors
Equal
Corr
OLS


3 predictors, Ro=.4, Unequal Collinearities, Unequal Validities
Mean validated R
0.0
0.2
0.4
0.6
0.8
1.0
5
15
30
50
75
100
Ratio of subjects to predictors
Equal
Corr
OLS

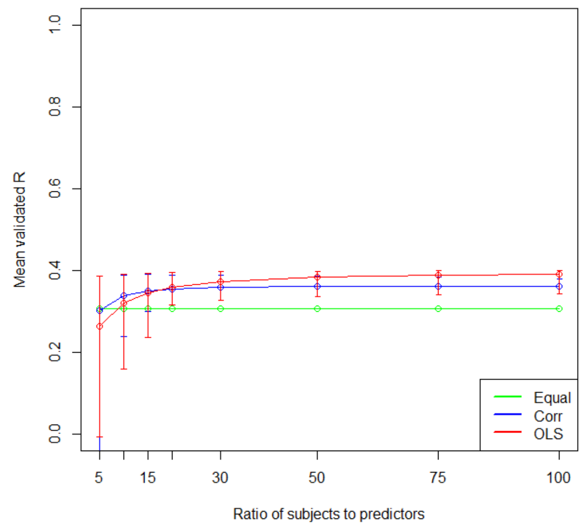

3 predictors, Ro=.7, Unequal Collinearities, Equal Validities
Mean validated R
0.0
0.2
0.4
0.6
0.8
1.0
5
15
30
50
75
100
Ratio of subjects to predictors
Equal
Corr
OLS


3 predictors, Ro=.7, Unequal Collinearities, Unequal Validities
Mean validated R
0.0
0.2
0.4
0.6
0.8
1.0
5
15
30
50
75
100
Ratio of subjects to predictors
Equal
Corr
OLS

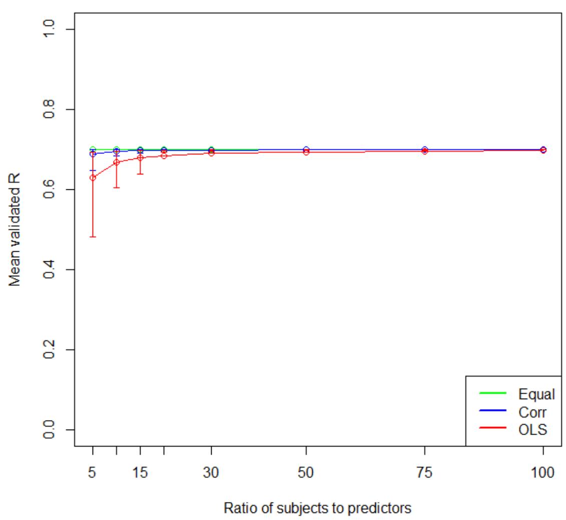
5 predictors, Ro=.4, Equal Collinearities, Equal Validities



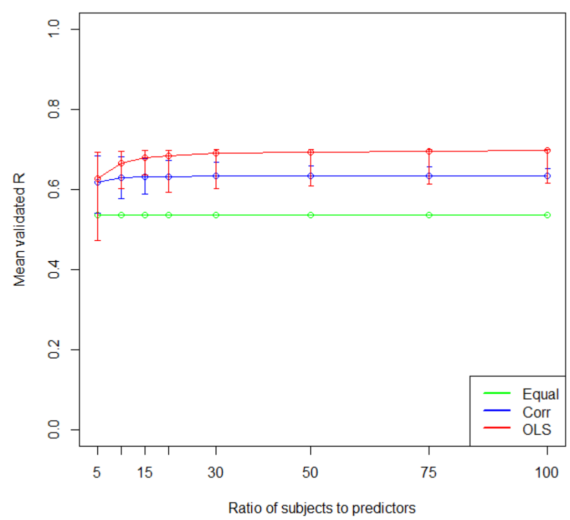
5 predictors, Ro=.4, Equal Collinearities, Unequal Validities



5 predictors, Ro=.7, Equal Collinearities, Equal Validities



5 predictors, Ro=.7, Equal Collinearities, Unequal Validities



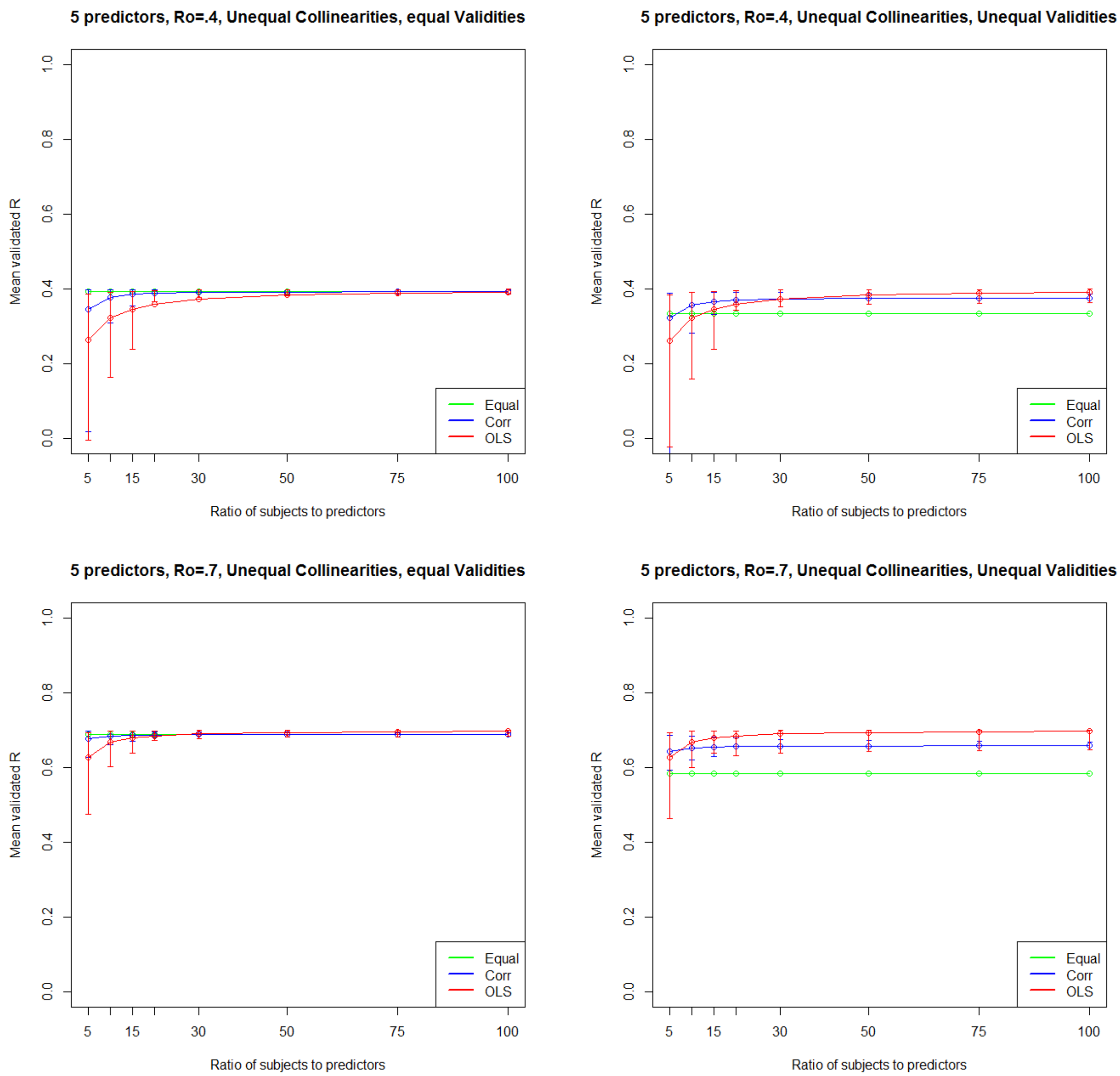


Figure 2. Relative performance of weight types as a function of equality of validities.

It is useful to compare these results with the equivalent results from Dana and Dawes (2004), to evaluate how their results and recommendations would serve a researcher constructing regression models based on the population characteristics used for my simulations. To that end, I developed a table format that overlays my results and theirs. Tables 1A through 1D present Study 1 results in terms of which weight type performed best for a given set of parameters (green cells) superimposed over the equivalent data from Dana and Dawes (2004) (cells with “D&D”). Note that, because Dana and Dawes (2004) did not include as many factors in their simulations, their results are repeated within the tables for each combination of factors. In other words, each table takes a set of their simulations and disaggregates the results into four levels of the validities and collinearities factors in my design.

For example, in Table 1A where models have three predictors and R = .4, my results indicate that equal weights perform best for all ratios of subjects to predictor whenever validities are equal, irrespective of collinearities (the first and third scenarios in the table); whereas unequal validities favor different weight types (equal, then correlation, then OLS) as the ratio increases. Because Dana and Dawes (2004) did not disaggregate their results but rather averaged across these levels of the parameters, their results predict the same pattern in each of the four scenarios in Table 1A, a pattern that unsurprisingly resembles an average of my results (and which provides a validity check on both sets of results).

Table 1A

Comparison of Study 1 Results for 3 Predictors and R = .4

3 predictors, R=.4, equal collinearities, equal validities								
	5m	10m	15m	20m	30m	50m	75m	100m
Equal	D&D	D&D	D&D					
Correlation				D&D	D&D	D&D	D&D	D&D
OLS								
3 predictors, R=.4, equal collinearities, unequal validities								
	5m	10m	15m	20m	30m	50m	75m	100m
Equal	D&D	D&D	D&D					
Correlation				D&D	D&D	D&D	D&D	D&D
OLS								
3 predictors, R=.4, unequal collinearities, equal validities								
	5m	10m	15m	20m	30m	50m	75m	100m
Equal	D&D	D&D	D&D					
Correlation				D&D	D&D	D&D	D&D	D&D

OLS								
3 predictors, R=.4, unequal collinearities, unequal validities								
	5m	10m	15m	20m	30m	50m	75m	100m
Equal	D&D	D&D	D&D					
Correlation				D&D	D&D	D&D	D&D	D&D
OLS								

Table 1B

Comparison of Study 1 Results for 3 Predictors and R = .7

3 predictors, R=.7, equal collinearities, equal validities								
	5m	10m	15m	20m	30m	50m	75m	100m
Equal	D&D							
Correlation		D&D	D&D					
OLS				D&D	D&D	D&D	D&D	D&D
3 predictors, R=.7, equal collinearities, unequal validities								
	5m	10m	15m	20m	30m	50m	75m	100m
Equal	D&D							
Correlation		D&D	D&D					
OLS				D&D	D&D	D&D	D&D	D&D
3 predictors, R=.7, unequal collinearities, equal validities								
	5m	10m	15m	20m	30m	50m	75m	100m
Equal	D&D							
Correlation		D&D	D&D					
OLS				D&D	D&D	D&D	D&D	D&D
3 predictors, R=.7, unequal collinearities, unequal validities								
	5m	10m	15m	20m	30m	50m	75m	100m
Equal	D&D							
Correlation		D&D	D&D					
OLS				D&D	D&D	D&D	D&D	D&D

Table 1C

Comparison of Study 1 Results for 5 Predictors and R = .4

5 predictors, R=.4, equal collinearities, equal validities								
	5m	10m	15m	20m	30m	50m	75m	100m
Equal	D&D	D&D	D&D					
Correlation				D&D	D&D	D&D	D&D	D&D
OLS								
5 predictors, R=.4, equal collinearities, unequal validities								
	5m	10m	15m	20m	30m	50m	75m	100m

Equal	D&D	D&D	D&D					
Correlation				D&D	D&D	D&D	D&D	D&D
OLS								
5 predictors, R=.4, unequal collinearities, equal validities								
	5m	10m	15m	20m	30m	50m	75m	100m
Equal	D&D	D&D	D&D					
Correlation				D&D	D&D	D&D	D&D	D&D
OLS								
5 predictors, R=.4, unequal collinearities, unequal validities								
	5m	10m	15m	20m	30m	50m	75m	100m
Equal	D&D	D&D	D&D					
Correlation				D&D	D&D	D&D	D&D	D&D
OLS								

Table 1D

Comparison of Study 1 Results for 5 Predictors and R = .7

5 predictors, R=.7, equal collinearities, equal validities								
	5m	10m	15m	20m	30m	50m	75m	100m
Equal	D&D							
Correlation		D&D	D&D					
OLS				D&D	D&D	D&D	D&D	D&D
5 predictors, R=.7, equal collinearities, unequal validities								
	5m	10m	15m	20m	30m	50m	75m	100m
Equal	D&D							
Correlation		D&D	D&D					
OLS				D&D	D&D	D&D	D&D	D&D
5 predictors, R=.7, unequal collinearities, equal validities								
	5m	10m	15m	20m	30m	50m	75m	100m
Equal	D&D							
Correlation		D&D	D&D					
OLS				D&D	D&D	D&D	D&D	D&D
5 predictors, R=.7, unequal collinearities, unequal validities								
	5m	10m	15m	20m	30m	50m	75m	100m
Equal	D&D							
Correlation		D&D	D&D					
OLS				D&D	D&D	D&D	D&D	D&D

Overall, the results from the previous and present studies diverge frequently, and the manner of divergence varies between conditions. Results are most similar in tables and graphs when both the conditions of very small samples and lower predictabilities apply, although differences are small. Results begin to diverge as the number of subjects per predictor rise above 5 or 10, still well within the range where Dana and Dawes (2004) consider alternative weights most useful. In the tables as in the graphs, equal weights perform best, but the results tend to favor OLS weights with unequal validities. Results from Dana and Dawes (2004), in contrast, generally lie between these two extremes, frequently favoring correlation weights. Their results are consistent with their practice of simulating samples from population matrices with the full range of possible validity and collinearity values, then averaging across all validated *R*'s. In other words, their recommendations are based on weight performance in the *average* population, from which any specific population may deviate considerably.

We can also quantify how often the weight type favored by Dana and Dawes's (2004) results failed to perform best in this study, signified in the table as red text. Across all 128 combinations of factor levels, the Dana and Dawes (2004) analysis suggested an incorrect dominant weight type compared to results from the current simulation in 78 cases, or 60.9 percent of the time. The greatest number of discrepancies (22 of 32) occurred when there were five predictors and $R = .4$, and the fewest (17 of 32) occurred with three predictors and $R = .4$. For the three smallest sample sizes, when one would be most likely to seek the benefits of alternative weights, 47.9 percent (23 of 48) of cases differed when the equality of validities and collinearities was taken into account. Of course, these sorts of calculations ignore the fact that the size of the differences in weight performance often are essentially negligible, as is demonstrated by the confidence intervals in the graphed results. Nonetheless, a researcher choosing the best weight type for her/his analysis might easily infer from extant literature on alternative weights that the results should be parsed in this way.

In summary, equal weights dominate the other weight types when validities are equal, but otherwise tend to be dominated by correlation and OLS weights. Collinearity equality had little impact in comparison with validity equality, perhaps because the constraints of average $r = .3$ limited how different the unequal collinearities could be. Although correlation weights are more precise when samples are at their smallest, the difference is slight. Finally, Dana and Dawes's (2004) results appear to be most applicable for populations with validities that are not equal but also not too unequal; relative weight performance can be quite different at either extreme, a conclusion that cannot be ascertained from their results because of their averaging approach.

CHAPTER IV

STUDY 2

Study 2 Methods

Study 1 results established that varying validities, even while holding their average constant, can impact the performance of alternative weights relative to OLS weights. Given that perfectly (or even approximately) equal validities are unlikely in the real world, the design for Study 2 assumes unequal validities but varies in other factors. Study 1 also showed that the pattern of collinearities matters less than validities, at least for what I have defined as a typical population correlation matrix with average collinearities of around $r = .3$. Accordingly, for Study 2, I fix the values of collinearities for every simulation, using more realistic unequal correlations. Otherwise, the simulation methods remain the same as in Study 1. Most factors and their levels (eight ratios of subjects per predictor, predictabilities of .4 or .7, three or five predictabilities) are also unchanged.

The factor examined in Study 2 is the average magnitude of validities. This is yet another factor that was not considered in the results reported by Dana and Dawes (2004): while their simulation sampled across the widest possible range of magnitudes for this and other parameters, the results were averaged across that range. I hypothesize that, as a consequence of their averaging across different magnitudes of average validities, the performance of different weight types in Study 2 frequently deviates from their predictions.

As before, population matrices for Study 2 were created by specifying values for R and the collinearities, resulting in the desired properties of lower or higher magnitude validities. As an example, below are the actual lower magnitude average validities (left) and higher magnitude average validities (right) with R = .4 and three predictors. In practice, the lower validities are not that different from the higher validities, due to the constraints of average validity size and model predictability (this is much less of an issue when R = .7). However, the important point is that the results for each population are reported separately instead of being aggregated so that we can see any impact of this factor on weight performance. If we find that the best choice of weights changes even when the average magnitude of validities is only varied by such a small amount, then there are even greater implications for abandoning OLS weights.

	Y	X_1	X_2	X_3
Y	1.00			
X_1	0.13	1.00		
X_2	0.26	0.20	1.00	
X_3	0.38	0.30	0.40	1.00

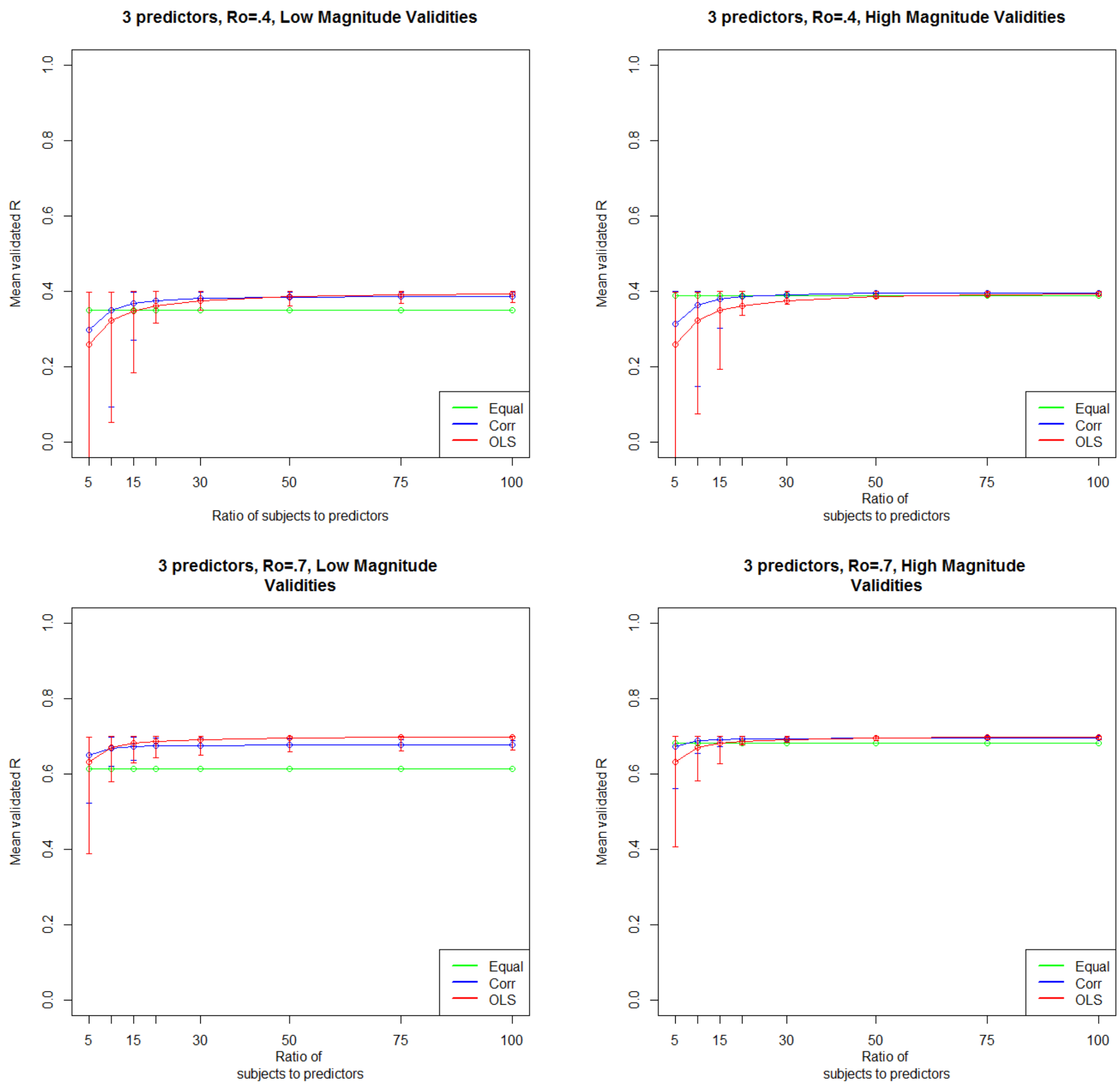
	Y	X_1	X_2	X_3
Y	1.00			
X_1	0.21	1.00		
X_2	0.28	0.20	1.00	
X_3	0.36	0.30	0.40	1.00

As you can see, the larger magnitude set of validities on the right are closer to equality than those on the left. This is because the constraints I have placed on the parameters (unequal validities that are approximately equally spaced, averaging near .3, and R = .4) strictly limit the possible range of validities. Increasing the validities' average magnitude while maintaining these constraints necessarily squeezes their range up against that ceiling. As a result, the low and high magnitude conditions in Study 2 can also be understood as a variation on the unequal and equal validity conditions in Study 1, although unlike Study 1, the matrices represent degrees of inequality with no scenario of perfect equality.

Study 2 Results

The greatest impact of varying the average magnitude of validities (and thus their relative equality) was for simulations with population multiple R = .4. As shown in Figure 3, higher average validities favored equal weights over OLS and correlation weights only for the lowest ratios of observations per predictor and then only barely, and it is rapidly matched or overtaken by correlation weights and then OLS weights at higher ratios. Once again, models of populations with lower validities slightly favored correlation weights at the lowest subject ratios, this time at ratios of 20 to 30 subjects or fewer per predictor, and OLS weights at higher ratios, both beating equal weights handily. A multiple R of .7 appears to be sufficiently large to negate equal weights' advantage even when average validities are high; instead, correlation weights dominated OLS weights (and had narrower confidence intervals) and equal weights at lower ratios, then were surpassed by OLS weights for higher ratios (Figure 3).

If we consider the correspondence between magnitude and degree of equality, these results echo Study 1 results: equal weights perform best when validities are closer to equality, although the fact that validities in Study 2 are never particularly close to equality means that their advantage is small and they are eventually dominated in many cases. This pattern is less pronounced with the higher predictability of .7, because population validities are spread across a wider range whether their mean is high or low. Also, as in Study 1 results, confidence intervals around both the correlation and OLS weights are wide, and while correlation weight confidence intervals were narrower than OLS intervals, the magnitude of the errors still dwarfs the magnitude of improvement in validated R.


3 predictors, Ro=.4, Low Magnitude Validities
Mean validated R
Ratio of subjects to predictors
Equal
Corr
OLS
3 predictors, Ro=.4, High Magnitude Validities
Mean validated R
Ratio of subjects to predictors
Equal
Corr
OLS
3 predictors, Ro=.7, Low Magnitude Validities
Mean validated R
Ratio of subjects to predictors
Equal
Corr
OLS
3 predictors, Ro=.7, High Magnitude Validities
Mean validated R
Ratio of subjects to predictors
Equal
Corr
OLS

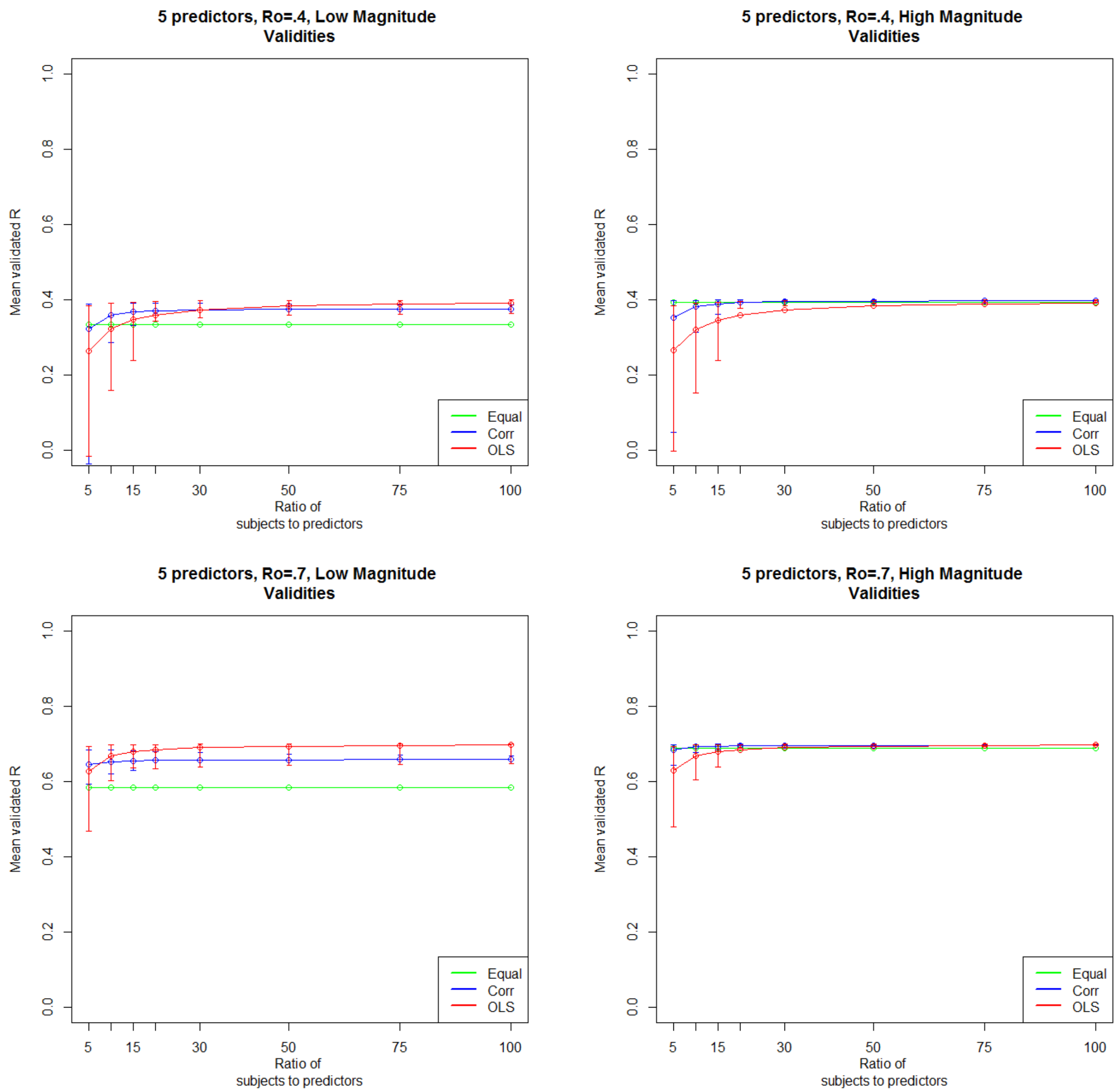


Figure 3. Relative performance of weight types as a function of magnitude of validities.

Tables 2A – 2D, comparing Study 2 results (green cells) with those of Dana and Dawes (2004) supports the hypothesis that their predictions cannot be relied upon when results are broken out by different levels of the factor. Their results nearly agree with mine when predictability is .4 and validities are higher, conflicting only at a ratio of 20 subjects per predictor. This is the point at which they say one ought to stop using equal weights and start using correlation weights, so disagreement here is consequential. Of course, the graphs show that the improvement gained by using alternatives to OLS weights is very small and rapidly diminishes with larger samples and validities anyway, just as in Study 1.

Table 2A

Comparison of Study 2 Results for 3 Predictors and R = .4

3 predictors, R=.4, low magnitude validities								
	5m	10m	15m	20m	30m	50m	75m	100m
Equal	D&D	D&D	D&D					
Correlation				D&D	D&D	D&D	D&D	D&D
OLS								
3 predictors, R=.4, high magnitude validities								
	5m	10m	15m	20m	30m	50m	75m	100m
Equal	D&D	D&D	D&D					
Correlation				D&D	D&D	D&D	D&D	D&D
OLS								

Table 2B

Comparison of Study 2 Results for 3 Predictors and R = .7

3 predictors, R=.7, low magnitude validities								
	5m	10m	15m	20m	30m	50m	75m	100m
Equal	D&D							
Correlation		D&D	D&D					
OLS				D&D	D&D	D&D	D&D	D&D
3 predictors, R=.7, high magnitude validities								
	5m	10m	15m	20m	30m	50m	75m	100m
Equal	D&D							
Correlation		D&D	D&D					
OLS				D&D	D&D	D&D	D&D	D&D

Table 2C

Comparison of Study 2 Results for 5 Predictors and R = .4

5 predictors, R=.4, low magnitude validities								
	5m	10m	15m	20m	30m	50m	75m	100m
Equal	D&D	D&D	D&D					
Correlation				D&D	D&D	D&D	D&D	D&D
OLS								
5 predictors, R=.4, high magnitude validities								

	5m	10m	15m	20m	30m	50m	75m	100m
Equal	D&D	D&D	D&D					
Correlation				D&D	D&D	D&D	D&D	D&D
OLS								

Table 2D

Comparison of Study 2 Results for 5 Predictors and R = .7

5 predictors, R=.7, low magnitude validities								
	5m	10m	15m	20m	30m	50m	75m	100m
Equal	D&D							
Correlation		D&D	D&D					
OLS				D&D	D&D	D&D	D&D	D&D
5 predictors, R=.7, high magnitude validities								
	5m	10m	15m	20m	30m	50m	75m	100m
Equal	D&D							
Correlation		D&D	D&D					
OLS				D&D	D&D	D&D	D&D	D&D

The tables confirm that correlation weights consistently outperform OLS weights only at ratios of 30 subjects per predictor or below, given a population *R* of .4. This is in contrast with Dana and Dawes's (2004) recommendation that OLS weights not be used when population R is below .6 or the number of subjects per predictor is 100. One might be tempted to take these results to suggest a narrower version of their rule of thumb: equal weights should always be used when *R* is less than or equal to .4 and there are 5 subjects or fewer per predictor, and correlation weights should always be used when *R* is at or below .4 and there are 10 to 30 subjects per predictor. That conclusion is overreaching given that this study's design, as mentioned above, sacrifices generalizability for specificity. It is designed to show that there are plausible populations for which alternative regression weights perform in ways that Dana and Dawes (2004) did not anticipate; it cannot show that the results for some plausible populations generalize to all others. At best, Study 2 results support the application of such a rule of thumb only when criterion and predictor parameters are known to have a pattern of correlations similar to the those in the matrices used in Study 2. This is a level of information about parameters that, for most studies and most parameters, is highly improbable. As Study 3 will show, there are other patterns with different consequences for choice of weights.

CHAPTER V

STUDY 3

Study 3 Methods

The designs of the preceding two studies both compared weight performance between populations with different distributions of validities, but always spread evenly. The Study 3 design differs in that it compares weight type performance in populations with validities that are skewed (each vector has one validity much smaller than the others or much greater than the others), and that are either positively associated with collinearities (predictors with larger validities also have larger collinearities) or negatively associated (predictors with larger validities have smaller collinearities).

Results from Study 1 and especially Study 2 imply that correlation weights often may outperform OLS weights, but the validities in those simulations were all positively associated with the collinearities, a condition more likely to favor correlation weights. The reason for this is that correlation weights and OLS weights will be correlated when the association is positive, potentially telling the same story with fewer parameter estimates, but this is not so when the association is negative. Take, for example, when the first predictor in a model has a much higher correlation with the criterion than do the other predictors. If validities and collinearities are positively associated, then the largest correlation weight will go to the predictor that both contributes the most information about the criterion (highest validity) and that best accounts for information from all the other predictors in the model (highest collinearities). On the other hand, if we reverse the association, the low validity predictors are now less redundant with the first predictor, yet they receive the exact same correlation weights as before. In contrast, OLS weights would change to reflect this difference. The Study 3 design therefore also includes a factor reflecting whether the predictor with the highest multicollinearity has either the highest or lowest validity.

To illustrate, below are the actual population matrices used for the simulations with one dominant validity and validities positively associated with collinearities (left), and two dominant validities and validities negatively associated with collinearities (right), with *R* = .4 and three predictors:

	Y	X_1	X_2	X_3
Y	1.00			
X_1	0.13	1.00		
X_2	0.20	0.20	1.00	
X_3	0.40	0.30	0.40	1.00

	Y	X_1	X_2	X_3
Y	1.00			
X_1	0.33	1.00		
X_2	0.28	0.20	1.00	
X_3	0.11	0.30	0.40	1.00

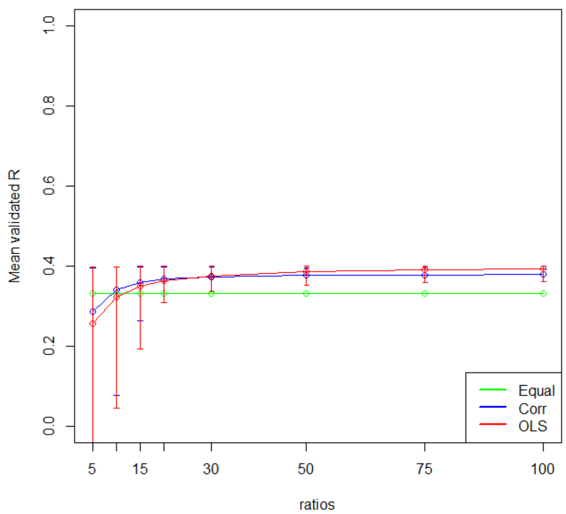
Study 3 Results

In general, equal weights performed best for the smallest ratios when the multiple *R* was smaller, but worst for all ratios when it was larger. Their performance compared to correlation and OLS weight performance was largely unaffected by number of predictors, validity-collinearity association, or whether the outlying validity value was larger or smaller than the others. Thus, consistent with other findings, patterns of validities only impacted equal weight performance as they approached or deviated from equality (as in Studies 1 and 2); when degree of equality is not manipulated (as in Study 3), results for equal weights resemble those of Dana and Dawes (2004).

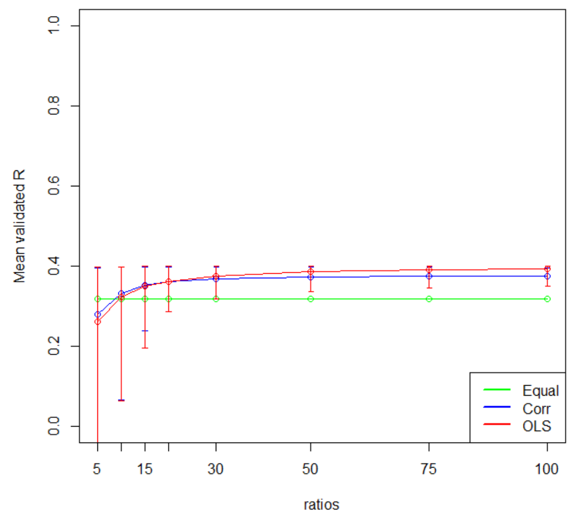
As predicted, correlation weights gave their best performance across conditions where the highest multicollinearities coincide with the highest validities, but even then, they were bested by equal weights at the lowest ratios when multiple $R = .4$. When correlation weights did overtake the other weight types, its advantage in terms of mean validated *R*'s was quite small. As expected, OLS weights almost invariably outperformed alternative weights as the number of subjects per predictor rose, especially with the higher predictability of multiple $R = .7$. Even at the lower predictability of .4, OLS sometimes performed best at ratios as low as fifteen subjects per predictor when validities and collinearities were negatively associated. Conversely, OLS was never outperformed when predictability and sample size were high. In practice, this means that underestimating the population predictability of the model is very likely to lead to overestimating the performance of alternative weights.

With population multiple $R = .4$, correlation weights went from performing slightly better than OLS weights at lower ratios to essentially tying when the associations between validities and collinearities were switched from positive to negative. Once again, confidence intervals for both types of weights were wide, but this time the OLS and correlation weight intervals were nearly equal, meaning that even when correlation weights performed better on average they would also frequently perform worse.

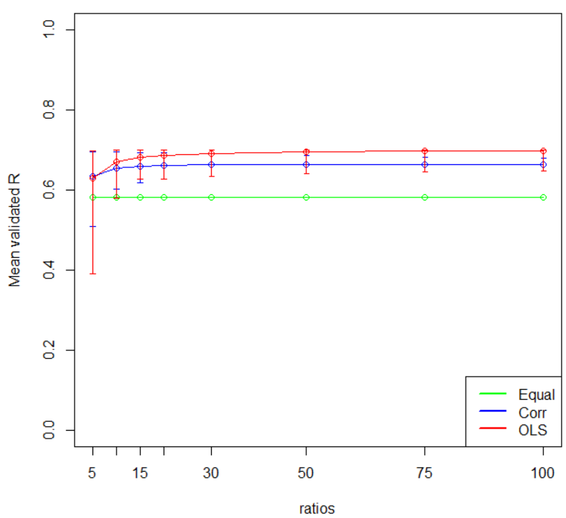
3 predictors, Ro=.4, One Dominant Validity, Positive Association



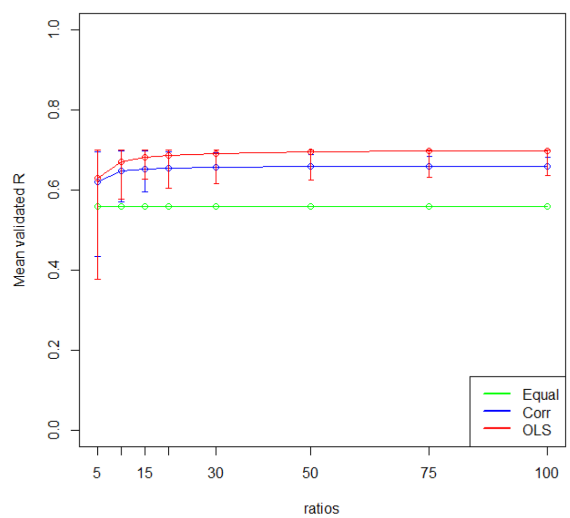
3 predictors, Ro=.4, One Dominant Validity, Negative Association

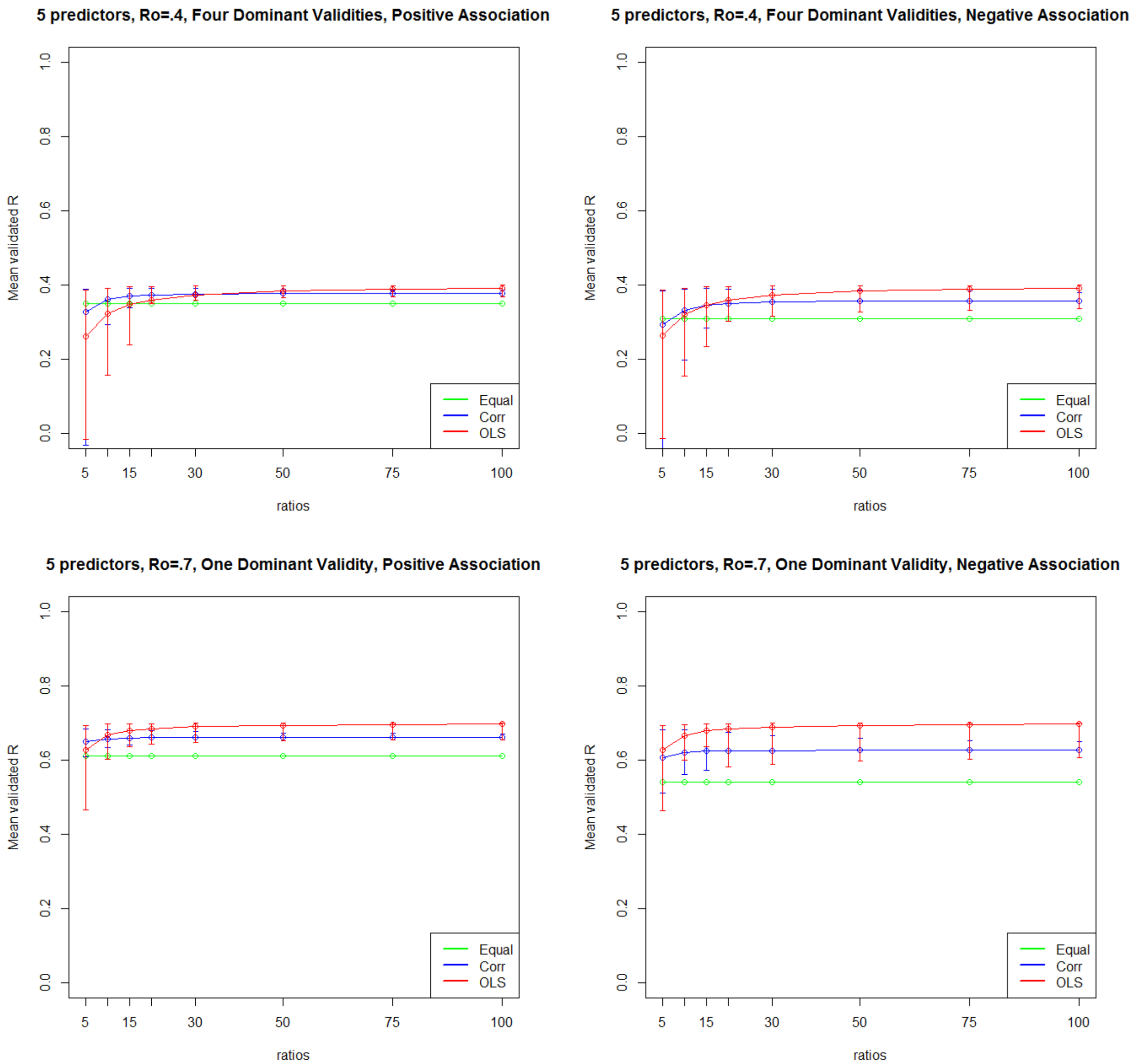


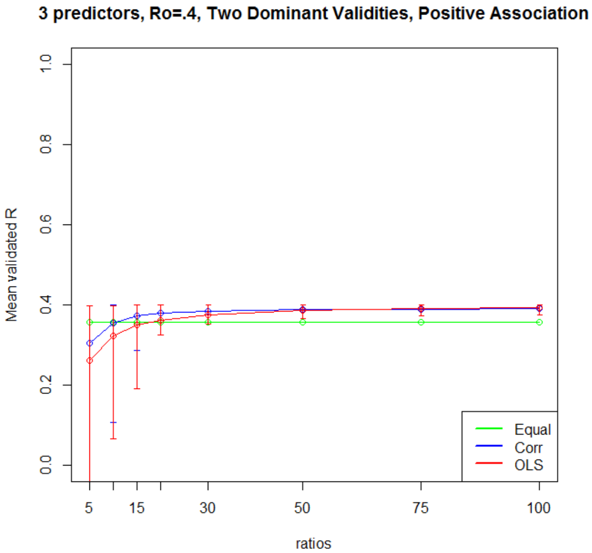
3 predictors, Ro=.7, One Dominant Validity, Positive Association

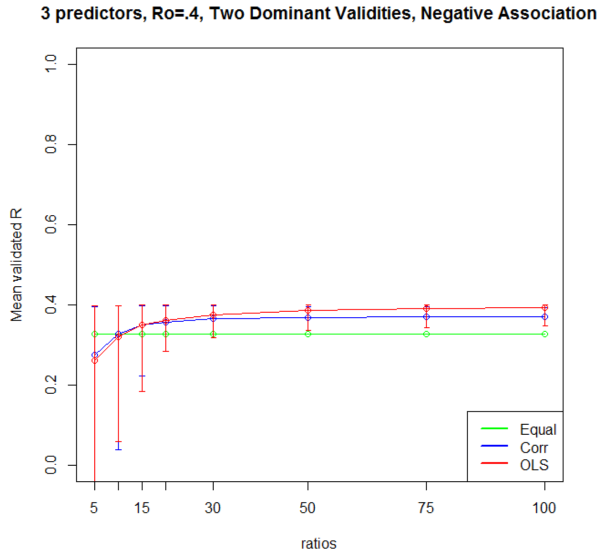


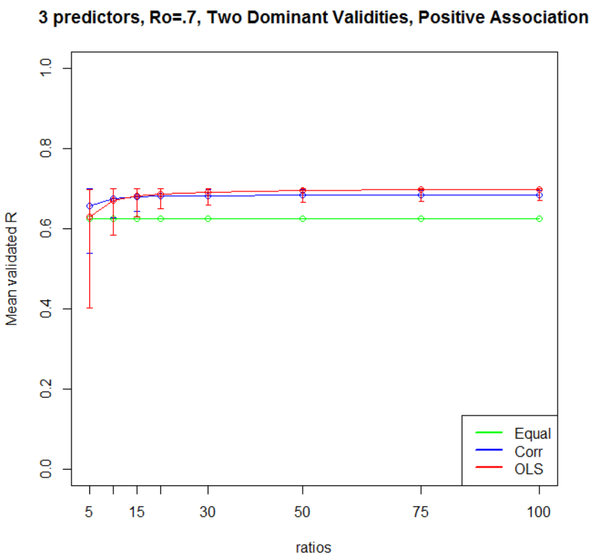
3 predictors, Ro=.7, One Dominant Validity, Negative Association

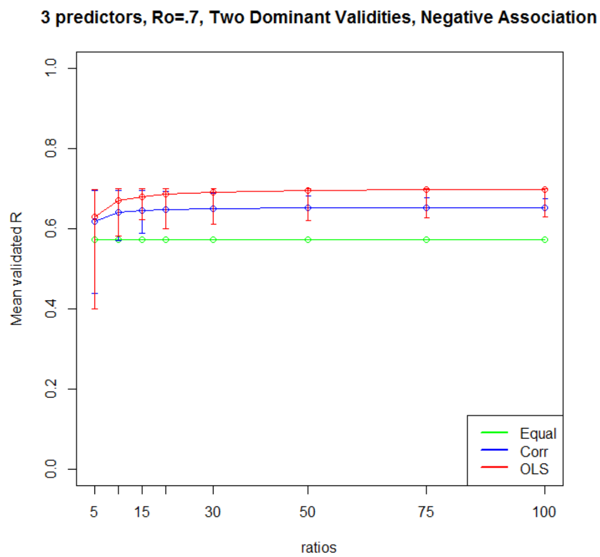



5 predictors, Ro=.4, Four Dominant Validities, Positive Association
5 predictors, Ro=.4, Four Dominant Validities, Negative Association
5 predictors, Ro=.7, One Dominant Validity, Positive Association
5 predictors, Ro=.7, One Dominant Validity, Negative Association
Mean validated R
ratios
Equal
Corr
OLS
0.0
0.2
0.4
0.6
0.8
1.0
5
15
30
50
75
100


3 predictors, Ro=.4, Two Dominant Validities, Positive Association
Mean validated R
0.0
0.2
0.4
0.6
0.8
1.0
5
15
30
50
75
100
ratios
Equal
Corr
OLS


3 predictors, Ro=.4, Two Dominant Validities, Negative Association
Mean validated R
0.0
0.2
0.4
0.6
0.8
1.0
5
15
30
50
75
100
ratios
Equal
Corr
OLS


3 predictors, Ro=.7, Two Dominant Validities, Positive Association
Mean validated R
0.0
0.2
0.4
0.6
0.8
1.0
5
15
30
50
75
100
ratios
Equal
Corr
OLS


3 predictors, Ro=.7, Two Dominant Validities, Negative Association
Mean validated R
0.0
0.2
0.4
0.6
0.8
1.0
5
15
30
50
75
100
ratios
Equal
Corr
OLS

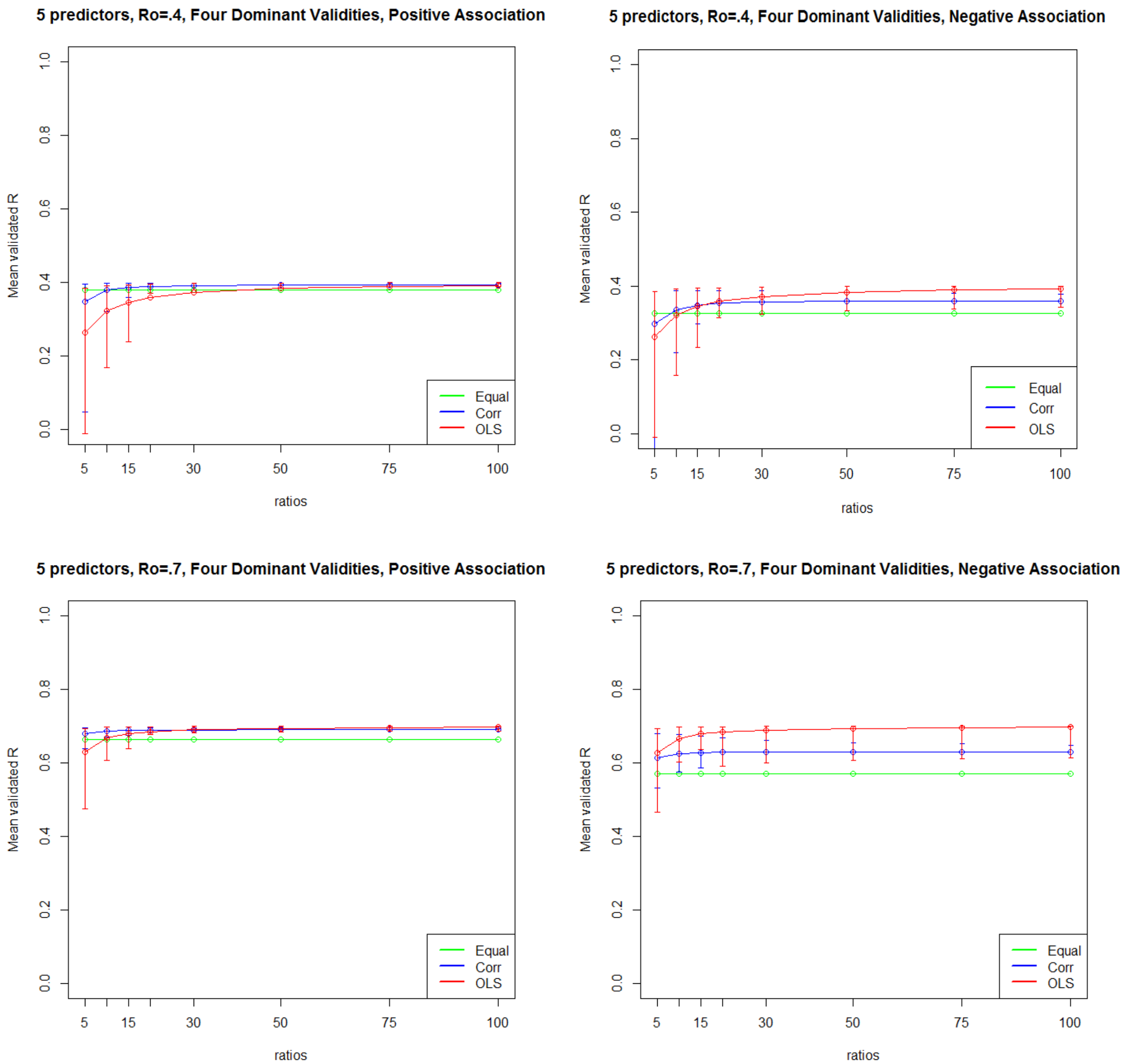


Figure 4. Relative performance of weight types with one or all but one predictor larger than the others, and validities positively or negatively associated.

Exactly half of the 128 combinations of factors in Study 3 yielded results that were inconsistent with Dana and Dawes's (2004) findings as to which weight type performed best. As can be seen in the tables below, most of the inconsistencies were due to the relative performance of correlation weights, and then mostly due to the negative association between validities and collinearities. For matrices with positive associations, there was disagreement only 37.5 percent of the time, whereas results for matrices with negative associations differed between studies in 62.5 percent of cases. For example, in Table 3A, Dana and Dawes (2004) underestimated the performance of OLS weights at higher ratios in all four scenarios, but they also underestimated correlation weight performance at lower ratios when the association between validities and collinearities was negative.

Table 3A

Comparison of Study 3 Results for 3 Predictors and R = .4

3 predictors, R=.4, one dominant, positively associated validities								
	5m	10m	15m	20m	30m	50m	75m	100m
Equal	D&D	D&D	D&D					
Correlation				D&D	D&D	D&D	D&D	D&D
OLS								
3 predictors, R=.4, two dominant, positively associated validities/collinearities								
	5m	10m	15m	20m	30m	50m	75m	100m
Equal	D&D	D&D	D&D					
Correlation				D&D	D&D	D&D	D&D	D&D
OLS								
3 predictors, R=.4, one dominant, negatively associated validities/collinearities								
	5m	10m	15m	20m	30m	50m	75m	100m
Equal	D&D	D&D	D&D					
Correlation				D&D	D&D	D&D	D&D	D&D
OLS								
3 predictors, R=.4, two dominant, negatively associated validities/collinearities								
	5m	10m	15m	20m	30m	50m	75m	100m
Equal	D&D	D&D	D&D					
Correlation				D&D	D&D	D&D	D&D	D&D
OLS								

Table 3B

Comparison of Study 3 Results for 3 Predictors and R = .7

3 predictors, R=.7, one dominant, positively associated validities/collinearities								
	5m	10m	15m	20m	30m	50m	75m	100m
Equal	D&D							
Correlation		D&D	D&D					
OLS				D&D	D&D	D&D	D&D	D&D

3 predictors, R=.7, two dominant, co positively associated validities/collinearities								
	5m	10m	15m	20m	30m	50m	75m	100m
Equal	D&D							
Correlation		D&D	D&D					
OLS				D&D	D&D	D&D	D&D	D&D
3 predictors, R=.7, one dominant, negatively associated validities/collinearities								
	5m	10m	15m	20m	30m	50m	75m	100m
Equal	D&D							
Correlation		D&D	D&D					
OLS				D&D	D&D	D&D	D&D	D&D
3 predictors, R=.7, two dominant, negatively associated validities/collinearities								
	5m	10m	15m	20m	30m	50m	75m	100m
Equal	D&D							
Correlation		D&D	D&D					
OLS				D&D	D&D	D&D	D&D	D&D

Table 3C

Comparison of Study 3 Results for 5 Predictors and R = .4

5 predictors, R=.4, one dominant, positively associated validities/collinearities								
	5m	10m	15m	20m	30m	50m	75m	100m
Equal	D&D	D&D	D&D					
Correlation				D&D	D&D	D&D	D&D	D&D
OLS								
5 predictors, R=.4, four dominant, positively associated validities/collinearities								
	5m	10m	15m	20m	30m	50m	75m	100m
Equal	D&D	D&D	D&D					
Correlation				D&D	D&D	D&D	D&D	D&D
OLS								
5 predictors, R=.4, one dominant, negatively associated validities/collinearities								
	5m	10m	15m	20m	30m	50m	75m	100m
Equal	D&D	D&D	D&D					
Correlation				D&D	D&D	D&D	D&D	D&D
OLS								
5 predictors, R=.4, four dominant, negatively associated validities/collinearities								
	5m	10m	15m	20m	30m	50m	75m	100m
Equal	D&D	D&D	D&D					
Correlation				D&D	D&D	D&D	D&D	D&D
OLS								

Table 3D

Comparison of Study 3 Results for 5 Predictors and R = .7

5 predictors, R=.7, one dominant, positively associated validities/collinearities								
	5m	10m	15m	20m	30m	50m	75m	100m
Equal	D&D							
Correlation		D&D	D&D					
OLS				D&D	D&D	D&D	D&D	D&D
5 predictors, R=.7, four dominant, positively associated validities/collinearities								
	5m	10m	15m	20m	30m	50m	75m	100m
Equal	D&D							
Correlation		D&D	D&D					
OLS				D&D	D&D	D&D	D&D	D&D
5 predictors, R=.7, one dominant, negatively associated validities/collinearities								
	5m	10m	15m	20m	30m	50m	75m	100m
Equal	D&D							
Correlation		D&D	D&D					
OLS				D&D	D&D	D&D	D&D	D&D
5 predictors, R=.7, four dominant, negatively associated validities/collinearities								
	5m	10m	15m	20m	30m	50m	75m	100m
Equal	D&D							
Correlation		D&D	D&D					
OLS				D&D	D&D	D&D	D&D	D&D

CHAPTER VI

DISCUSSION

In the present study, I have revisited the conclusions of Dana and Dawes (2004) using a design that focuses on variation in population predictor validities and collinearities, factors overlooked in their study. Specifically, I have examined the relative performance of OLS, equal, and correlation regression weights as I varied the set of predictor validities' equality, magnitude, and spread relative to predictors' collinearities. In contrast to their design, which simulated a great variety of populations with parameter values across their entire possible range, my design limited each parameter value to a narrow, plausible range and then simulated populations with parameters across that range.

I found that the type of regression weights that produced the best validated *R* depended on the distribution of validities relative to the distribution of collinearities: more equal validities favored equal weights, while correlation weights performed best when validities were positively associated with collinearities. Equal weights consistently outperformed correlation and OLS weights at the lowest levels of predictability and subjects per predictor. At low to moderate levels of these factors, OLS weights frequently performed as well or better than the alternatives, even when predictability was low and subjects per predictor were few. When correlation weights did outperform OLS weights at these levels, the addition of confidence intervals revealed that the difference was negligible. As expected, OLS weights performed best at higher levels of predictability and subjects per predictor.

Based on the results of these three studies, equal weights should not be expected to be the best weight type when there are more than five subjects per predictor or the population predictability is lower than .4. Correlation weight performance is suspect when there are fewer than twenty subjects per predictor and population multiple *R* is lower than .4. The consequences of using alternative weights under these circumstances could include decreasing validated *R* even as you believe you are increasing it. Note that these observations are not phrased as recommendations: these studies demonstrate several populations that may undermine alternative weight performance compared with OLS weights, but the types of scenarios I have examined are far from exhaustive. It is therefore more accurate to say that the results neither prove nor disprove that alternative weights will be the best choice for regression with few subjects and low predictabilities, although results are suggestive of the patterns identified above.

To summarize results across all three phases of this study, the predictions of Dana and Dawes (2004) as to the performance of alternative weights, in terms of which weight type would perform best in an absolute sense, frequently failed when the values of regression parameters were fixed within a plausible range. Furthermore, the way in which their predictions differed from my findings varied depending on the patterns of validities and collinearities in the population matrices, patterns that are unlikely to be apparent a priori to real researchers. When validities are equal, as in Study 1, they favor equal weights; equal weights give way to correlation weights as validities diverge from equality but remain evenly distributed, as in Study 2; unless, as Study 3 shows, the validities are negatively associated with collinearities, in which case correlation weights may be bested by OLS weights even at lower ratios of subjects to predictors and even when predictability is low.

The results presented above clearly indicate that the alternative weights selected can outperform OLS weights in realistic scenarios, but whether the better choice is correlation weights or equal weights, and where the best choice transitions from one weight type to another, depends on characteristics of the data a researcher is unlikely to know when planning the analysis. OLS weight performance frequently is between that of equal weights and correlation weights, so in many cases we gain the possibility of making better predictions than OLS regression only if we are willing to also risk worse predictions. In short, a researcher who knows enough about the variables of interest to select the best alternative weights with confidence may have the least to learn from the analysis in the first place. Improvement over such knowledge may require a rich dataset with large samples or highly predictable criteria, in which case alternative regression weights almost never outperform OLS weights.

A researcher working under more typical conditions, who chooses to use alternative weights without either strong a priori knowledge or unusually high-quality data, is adding uncertainty to parameter estimates through a process that is supposed to reduce uncertainty, and is more confident of her predictions when she should be less. This directly contradicts Dana and Dawes's (2004) argument that we ought not to use methods that produce predictions that imply greater precision than the data can support. It turns out that using blunter tools to make noisier estimates often increases imprecision. Instead, we should sharpen our interpretations by calculating and reporting confidence intervals around the OLS weights in our models as well around their criterion estimates.

One might argue that it is unnecessary to know in advance the patterns of correlations among all the model parameters, given that the sample at hand can be used to estimate those correlations. For example, a researcher might be tempted to use equal weights only when sample validities are near equal (setting aside the issue of how “near equal” should be defined), correlation weights when validities are not too skewed and are positively associated enough with collinearities (again, however defined), and otherwise use OLS weights. There are two problems with this approach, one methodological and the other ethical. Methodologically, the use of different weights based on sample characteristics is the very definition of overfitting and carries with it all the problems of overfitting (e.g., poor predictions with other samples) combined with the problems of post hoc analysis (e.g., standard errors that are too small and confidence intervals that are too narrow). But worse than increased error in predictions is the bias introduced by post hoc weight selection. If we deviate from OLS weights only when correlation patterns show less variability, then our methods will systematically decrease the variability in the long run, particularly when using equal weights. We will also be routinely (but unknowingly) penalizing individual variables that have high collinearities in the population, or sets of variables with patterns of intercorrelations in the population that would significantly impact model predictions.

The ethical dimension of weight choice relates to the difference between predictive and explanatory regression models. While explanatory models often inform theory or policy, predictive models frequently determine individuals’ access to important resources and major life opportunities. Corporations and governments have long used predictive models to inform decisions about who will be hired or promoted or laid off, while many educational institutions have developed their own predictive models for accepting or rejecting students. Increasingly, the justice system is using predictive models to decide which defendants should get bail and which convicts should be paroled (Dewan, 2015). These models use demographic variables like race, sex, disability and others associated with historically underrepresented populations—populations for which small samples are a concern in many settings. At best, the use of alternative weights may be a threat to face validity in such cases; at worst, the use of a highly consequential formula may consistently ignore or overemphasize personal characteristics of individuals, embedding bias in a process that statistical methods are thought to render more objective.

Dawes (1979) made a slightly different case for what he called improper weights, regression weights chosen independent of the data (e.g., equal weights). He showed that these insensitive weights at the very least provide better predictions compared to predictions from clinical judgment. Clinicians' work would thus benefit from using improper weights, which also have the advantages of simplicity of face validity (for those who distrust methods that intensively manipulate the data). In the modern age, however, simplicity is a weak rationale, as most professionals who work with data are quite familiar with OLS regression, and the software for employing it is inexpensive and user-friendly. Those who continue to eschew actuarial approaches likely do so for philosophical reasons and are therefore unlikely to be persuaded by statistical arguments.

These arguments lead me to conclude that, without a clearer picture of when to use which alternative weights, avoiding OLS regression weights for prediction is difficult to justify, virtually regardless of one's actuarial inclinations. For scientists, alternative regression weights are intriguing but rarely useful; for non-scientists, they are potentially useful but rarely appealing. This is a verdict that could be as sweeping as that of Dana and Dawes (2004) but for the several limitations of the present study. However, there are some important limitations that contextualize the current findings. I had to sacrifice breadth of analysis for specificity, which required making informed assumptions as to what constitutes a typical range for parameters. My conclusions therefore are most applicable for research in domains where these assumptions hold:

where the mean of predictor validities and collinearities is about .3, where patterns of correlations among parameters vary in ways not described by strong theory, and where predictabilities are unlikely to range across the entire spectrum between zero and one.[11] This study could be usefully replicated for typical populations in particular social science fields.

Alternatively, this study's design could be augmented to investigate those scenarios by treating the variable sets in the present simulation as templates, and generating many more variable sets that gradually deviate from each template across combinations of the relevant factors.[12] Evaluating so many data points would be impractical for a person but probably could be managed by a program written to automatically "score" weight performance and output summary results.

The current results would also be bolstered if the same results are obtained with different data and additional weight types. The overall design should be applied to large, publicly available real datasets through resampling simulations, like those conducted by Dana and Dawes (2004) that supported their other results. They also included several other types of weights, all of which underperformed in comparison with equal, correlation and OLS weights. I included only the last three in order to directly test Dana and Dawes's (2004) recommendation to researchers. Future research could also evaluate other weight types, including what they called "more

[11] One could also include as factors violations of design assumptions (e.g., correctly specified models, no suppression) or assumptions of the multiple regression procedure itself (e.g., linearity, normality, homoscedasticity), but the relative robustness of alternative weights to assumption violations is a separate question from their performance when the assumptions hold.

[12] This approach resembles Waller and Jones's (2011) simulation that gradually varied the validities of three predictors, where each set of three validities corresponds to the coordinates of a criterion moving from one endpoint of a three-dimensional arc to the other. However, the proposed design would vary many more factors, whereas their design fixed all other factors.

principled" (p. 325) options like ridge regression. Such weights often are designed to address particular problems with data and theoretically would be exceptions to an "always OLS" rule.

This study will not result in a major change in the typical practices of researchers, but it does bolster those practices against an old and somewhat enduring critique. There may be greater implications for the more general practice of creating predictive composite indices. For example, although standardized tests and commercial measures generally eschew multiple regression for multidimensional item response theory (MIRT) or other latent-variable approaches for creating composite scores from subscales (Yao, 2011; Wall and Li, 2003), this option is not available for ad hoc measures with fewer items or studies with small samples. My results suggest that it would be worthwhile to determine if the use of composite scores risks the same weaknesses as with their regression equivalents (e.g., averaging variables for composites and equal weights for predictors).

A similar question frequently arises in the context of evaluating the implementation and outcomes of complex interventions. Researchers often use measures of implementation fidelity to predict outcomes because the model can be used to 1) support the relationship between treatment and impact and 2) predict outcomes in the program's future and at other sites that may replicate it. As yet, there is no general agreement on how to combine different measures of fidelity or even whether the method should vary depending on program design. Complete models may involve many predictors because interventions commonly have multiple components and each component may be measured in multiple ways. The problem is exacerbated when fidelity is measured for only a subset of sites or participants due to the need to devote scare resources to assessing outcomes. Again, this study's results imply that OLS regression is often the best choice.

APPENDIX A

R Code for Simulations

```
############ Libraries ############

library(semTools)
library(Matrix)

############ WRAPPER FUNCTIONS ############

# Take a standardized multivariate normal sample with n cases and 1+m
# variables (criterion+predictors) and return a vector of m
# correlations between the criterion and each predictor

get.cor.weights=function(samp){return(c(cor(samp)[1,2:length(samp[1,])
]))}

# Take a standardized multivariate normal sample with n cases and 1+m
# variables (criterion+predictors), run the regression analysis, and
# return the beta coefficients

get.ols.weights=function(samp){
vars=length(samp[1,])
return(summary(lm(samp[,1]~samp[,2:vars]))$coefficients[2:vars])}

# Take a standardized multivariate normal sample with n cases and 1+m
# variables (criterion+predictors), confirm eigenvalues of its
# correlation matrix are positive, and return whether it's PD (1) or
# not (0)

is.PD.sample.matrix=function(samp){ifelse(min(eigen(cor(samp),symmetri
c=T,only.values=T)$values) >= 0,return(1),return(0))}

# Take an mX1 vector of weights of one type, along with the population
# correlation matrix, and calculate the validated R's using D&D's
# equation (p. 320, D&D): (w'*v)/sqrt(w'*Sigma*v), where w is the
# vector of weights for the sample, v is the vector of population
# correlations
# between the DV and the p predictors, and Sigma is the population
# matrix of correlations among p predictors

get.valid.Rs=function(weight1,pop.cor){
vars=length(pop.cor[,1])
return((weight1%*%pop.cor[1,2:vars])/sqrt(weight1%*%pop.cor[2:vars,2:v
ars]%*%weight1))}
```

```
############ MAIN FUNCTIONS ############

# Take population values for R-squared, the predictor correlation
# matrix Rxx, and a vector v that scales the correlations among the
# criterion and the predictors pxy; calculate the population values
# for pxy using Maxwell's (2007) equation 7 rearranged to solve for
# pxy, bind pxy to Rxx to create the population correlation matrix
# pop.cor; and return pop.cor if it is PD

make.cor.matrix=function(R2,Rxx,v)
{
t=t(v)%*%solve(Rxx)%*%v
pxy=v%*%sqrt(R2/t)
A1=cbind(pxy,Rxx)
A2=cbind(1,t(pxy))
pop.cor=rbind(A2,A1)
if(min(eigen(pop.cor,symmetric=T,only.values=T)$values) >=
0)return(pop.cor)
}

# Take population correlation matrix pop.cor along with the sample
# size num.cases and the number of samples to be generated from the
# population num.samples, use function kd to apply Kaiser & Dickman's
# method and generate a list of samples from the population, and
# return if population, and return the list only if all samples
# correlation matrices are PD

get.all.samples=function(pop.cor,num.cases,num.samples)
{
scores=lapply(rep(list(pop.cor),num.samples),kd,num.cases,"sample")
if(prod(as.numeric(lapply(scores,is.PD.sample.matrix)))==1)return(lapp
ly(scores,scale))
}

# Take a list of samples and the number of samples in the list, call
# separate wrapper functions for each type of weight to make a 3 lists
# of m weights for each sample, return a num.samples X 3 list where
# each row is the list of weights for one sample and each column is a
# different weight type

get.all.weights=function(scores,num.samples)
{
unit.weights=rep(list(rep(1,(length(scores[[1]][1,])-1)),num.samples)
cor.weights=lapply(scores,get.cor.weights)
ols.weights=lapply(scores,get.ols.weights)
return(cbind(unit.weights,cor.weights,ols.weights))
}

# Take 3 lists weights (one for each weight type) and the population
# correlation matrix, send each list to a function calculating
# validated R, and return a single list containing the validated Rs
# for the 3 types of weights
```

```
get.all.valid.Rs=function(unit.weights,cor.weights,ols.weights,pop.cor
)
{
unit.valid.Rs=lapply(unit.weights,get.valid.Rs,pop.cor)
cor.valid.Rs=lapply(cor.weights,get.valid.Rs,pop.cor)
ols.valid.Rs=lapply(ols.weights,get.valid.Rs,pop.cor)
return(cbind(unit.valid.Rs,cor.valid.Rs,ols.valid.Rs))
}

# Take R-squared, the population predictor correlation matrix, vector
# v that scales the criterion-predictor correlations, the sample size,
# and the number of samples; create the population matrix with
# make.cor.matrix(), create the scores for all samples simultaneously
# using get.all.samples(), create the three types of weights for every
# sample with get.all.weights(), and calculate the validated R's for
# each set of weights using get.all.valid.Rs(); and return the list of
# validated R's

dndSIM=function(R2,Rxx,v,num.cases,num.samples)
{
pop.cor=make.cor.matrix(R2,Rxx,v)
scores=get.all.samples(pop.cor,num.cases,num.samples)
weights=get.all.weights(scores,num.samples)
valid.Rs=get.all.valid.Rs(weights[,1],weights[,2],weights[,3],pop.cor)
return(valid.Rs)
}

############ MAIN PROGRAM CODE ############

# Define the population multiple R, the number of independent
# variables, the population predictor matrix, and the vector v that
# scales criterion-predictor correlations


seed=11111
R=.4
IV=3
Z.0=matrix(c(1,.3,.3,.3,1,.3,.3,.3,1),3,3)
v.0=matrix(c(.3,.3,.3),IV,1)

# For each sample size used by D&D, call dndSIM to return validated
# R's

num.5IV=dndSIM(R2=(R^2),Rxx=Z.0,v=v.0,num.cases=5*IV,
num.samples=10000)
num.10IV=dndSIM(R2=(R^2),Rxx=Z.0,v=v.0,num.cases=10*IV,
num.samples=10000)
num.15IV=dndSIM(R2=(R^2),Rxx=Z.0,v=v.0,num.cases=15*IV,
num.samples=10000)
num.20IV=dndSIM(R2=(R^2),Rxx=Z.0,v=v.0,num.cases=20*IV,
num.samples=10000)
```

```
num.30IV=dndSIM(R2=(R^2),Rxx=Z.0,v=v.0,num.cases=30*IV,
num.samples=10000)
num.50IV=dndSIM(R2=(R^2),Rxx=Z.0,v=v.0,num.cases=50*IV,
num.samples=10000)
num.75IV=dndSIM(R2=(R^2),Rxx=Z.0,v=v.0,num.cases=75*IV,
num.samples=10000)
num.100IV=dndSIM(R2=(R^2),Rxx=Z.0,v=v.0,num.cases=100*IV,
num.samples=10000)

############ GRAPHING CODE ############


## Define number of subjects per predictor (x-axis coordinates)

ratios=c(5,10,15,20,30,50,75,100)


##Create vector of mean validated R's (y-axis coordinates)

#Equal weights means

means1=c(mean(as.numeric(num.5IV[,1])),mean(as.numeric(num.10IV[,1])),
mean(as.numeric(num.15IV[,1])),mean(as.numeric

(num.20IV[,1])),mean(as.numeric(num.30IV[,1])),mean(as.numeric(num.50I
V[,1])),mean(as.numeric(num.75IV[,1])),mean

(as.numeric(num.100IV[,1])))

#Correlation weights means

means2=c(mean(as.numeric(num.5IV[,2])),mean(as.numeric(num.10IV[,2])),
mean(as.numeric(num.15IV[,2])),mean(as.numeric

(num.20IV[,2])),mean(as.numeric(num.30IV[,2])),mean(as.numeric(num.50I
V[,2])),mean(as.numeric(num.75IV[,2])),mean

(as.numeric(num.100IV[,2])))

#OLS weights means

means3=c(mean(as.numeric(num.5IV[,3])),mean(as.numeric(num.10IV[,3])),
mean(as.numeric(num.15IV[,3])),mean(as.numeric

(num.20IV[,3])),mean(as.numeric(num.30IV[,3])),mean(as.numeric(num.50I
V[,3])),mean(as.numeric(num.75IV[,3])),mean

(as.numeric(num.100IV[,3])))


##Create vector of lower confidence limits (250th out of 10000 sorted
values) across subject ratios
```

```
#Correlation weight lower limits

lci2=c(sort(as.numeric(num.5IV[,2]))[250],sort(as.numeric(num.10IV[,2]
))[250],sort(as.numeric(num.15IV[,2]))[250],sort

(as.numeric(num.20IV[,2]))[250],sort(as.numeric(num.30IV[,2]))[250],so
rt(as.numeric(num.50IV[,2]))[250],sort(as.numeric

(num.75IV[,2]))[250],sort(as.numeric(num.100IV[,2]))[250])

#OLS weight lower limits

lci3=c(sort(as.numeric(num.5IV[,3]))[250],sort(as.numeric(num.10IV[,3]
))[250],sort(as.numeric(num.15IV[,3]))[250],sort

(as.numeric(num.20IV[,2]))[250],sort(as.numeric(num.30IV[,2]))[250],so
rt(as.numeric(num.50IV[,2]))[250],sort(as.numeric

(num.75IV[,2]))[250],sort(as.numeric(num.100IV[,2]))[250])


##Create vector of upper confidence limits across subject ratios
(9751st out of 10000 sorted values)

#Correlation weight upper limits

uci2=c(sort(as.numeric(num.5IV[,2]))[9751],sort(as.numeric(num.10IV[,2
]))[9751],sort(as.numeric(num.15IV[,2]))[9751],sort

(as.numeric(num.20IV[,2]))[9751],sort(as.numeric(num.30IV[,2]))[9751],
sort(as.numeric(num.50IV[,2]))[9751],sort

(as.numeric(num.75IV[,2]))[9751],sort(as.numeric(num.100IV[,2]))[9751]
)

#OLS weight upper limits

uci3=c(sort(as.numeric(num.5IV[,3]))[9751],sort(as.numeric(num.10IV[,3
]))[9751],sort(as.numeric(num.15IV[,3]))[9751],sort

(as.numeric(num.20IV[,3]))[9751],sort(as.numeric(num.30IV[,3]))[9751],
sort(as.numeric(num.50IV[,3]))[9751],sort

(as.numeric(num.75IV[,3]))[9751],sort(as.numeric(num.100IV[,3]))[9751]
)

##Define the width of horizontal lines at ends of confidence intervals

epsilon <- 0.5

##Creating initial plot with axis labels
```

```
plot(ratios,means1,xaxt="n",ylim=c(.0,1),col="green",xlab="Ratio of
subjects to predictors",ylab="Mean validated R",title

("3 predictors, Ro=.7, Equal Collinearities, Equal Validities"))


##Drawing line graphs and confidence intervals

#Equal weight line

points(ratios,means1,col="green")
lines(ratios,means1,col="green")

#Correlation weight line

points(ratios,means2,col="blue")
lines(ratios,means2,col="blue")

#Correlation weight intervals

segments(ratios, lci2,ratios, uci2,col="blue")
segments(ratios-epsilon,lci2,ratios+epsilon,lci2,col="blue")
segments(ratios-epsilon,uci2,ratios+epsilon,uci2,col="blue")

#OLS weight line

points(ratios,means3,col="red")
lines(ratios,means3,col="red")

#OLS weight intervals

segments(ratios, lci3,ratios, uci3,col="red")
segments(ratios-epsilon,lci3,ratios+epsilon,lci3,col="red")
segments(ratios-epsilon,uci3,ratios+epsilon,uci3,col="red")

##Create legend

legend("bottomright",c("Equal","Corr","OLS"),lty=c(1,1,1),lwd=c(2.5,2.
5,2.5),col=c("green","blue","red"))

##Create x-axis

axis(side=1,at=ratios,las=1)
```

Appendix B

Study 1 Matrices and Arrays of Mean Validated R Values

Equal collinearities and equal validities, R = .4

1.0000000

0.2921187 1.0000000

0.2921187 0.3000000 1.0000000

0.2921187 0.3000000 0.3000000 1.0000000

equal.curve 0.4000000 0.4000000 0.4000000 0.4000000 0.4000000 0.4000000 0.4000000 0.4000000

cor.curve 0.3158637 0.3667404 0.3827165 0.3889442 0.3937787 0.3967898 0.3979321 0.3985152

ols.curve 0.2595590 0.3232200 0.3487953 0.3622271 0.3749508 0.3855525 0.3903106 0.3929288

Equal collinearities and unequal validities, R = .4

1.0000000

0.1242911 1.0000000

0.2485822 0.3000000 1.0000000

0.3728733 0.3000000 0.3000000 1.0000000

equal.curve 0.3403852 0.3403852 0.3403852 0.3403852 0.3403852 0.3403852 0.3403852 0.3403852

cor.curve 0.2903773 0.3441682 0.3614512 0.3698964 0.3759084 0.3795165 0.3812481 0.3818025

ols.curve 0.2610611 0.3239987 0.3489310 0.3625837 0.3752660 0.3854818 0.3906217 0.3928986

Equal collinearities and equal validities, R = .7

1.0000000

0.5112077 1.0000000

0.5112077 0.3000000 1.0000000

0.5112077 0.3000000 0.3000000 1.0000000

equal.curve 0.7000000 0.7000000 0.7000000 0.7000000 0.7000000 0.7000000 0.7000000 0.7000000

cor.curve 0.6761044 0.6917638 0.6951943 0.6965773 0.6978166 0.6987324 0.6991694 0.6993919

ols.curve 0.6315892 0.6702752 0.6810434 0.6861831 0.6909757 0.6946360 0.6964688 0.6973861

Equal collinearities and unequal validities, R = .7

1.0000000

0.2175094 1.0000000

0.4350188 0.3000000 1.0000000

0.6525282 0.3000000 0.3000000 1.0000000

equal.curve 0.5956741 0.5956741 0.5956741 0.5956741 0.5956741 0.5956741 0.5956741 0.5956741

cor.curve 0.6398837 0.6602232 0.6647647 0.6669568 0.6682461 0.6693821 0.6700289 0.6703598

ols.curve 0.6304613 0.6704466 0.6814830 0.6863164 0.6911296 0.6947799 0.6965983 0.6974187

Unequal collinearities and equal validities, R = .4

1.0000000

0.2912236 1.0000000

0.2912236 0.2000000 1.0000000

0.2912236 0.3000000 0.4000000 1.0000000

equal.curve 0.3987743 0.3987743 0.3987743 0.3987743 0.3987743 0.3987743 0.3987743 0.3987743

cor.curve 0.3125404 0.3662640 0.3812851 0.3876491 0.3926109 0.3954078 0.3966831 0.3972585

ols.curve 0.2595489 0.3249604 0.3488975 0.3619278 0.3753647 0.3852901 0.3903340 0.3928674

Unequal collinearities and unequal validities, R = .4

1.0000000

0.1280203 1.0000000

0.2560405 0.2000000 1.0000000

0.3840608 0.3000000 0.4000000 1.0000000

equal.curve 0.3505979 0.3505979 0.3505979 0.3505979 0.3505979 0.3505979 0.3505979 0.3505979

cor.curve 0.2964258 0.3487023 0.3677698 0.3743718 0.3806271 0.3838771 0.3851752 0.3858065

ols.curve 0.2589079 0.3215605 0.3487811 0.3617377 0.3751130 0.3855824 0.3902950 0.3928226

Unequal collinearities and equal validities, R = .7

1.0000000

0.5096413 1.0000000

0.5096413 0.2000000

0.5096413 0.3000000 0.4000000 1.0000000

equal.curve 0.6978551 0.6978551 0.6978551 0.6978551 0.6978551 0.6978551 0.6978551 0.6978551

cor.curve 0.6724258 0.6894243 0.6927781 0.6944088 0.6956329 0.6965849 0.6970215 0.6972107

ols.curve 0.6303976 0.6703849 0.6811776 0.6861619 0.6910539 0.6946375 0.6964696 0.6973553

Unequal collinearities and unequal validities, R = .7

1.0000000

0.2240355 1.0000000

0.4480709 0.2000000 1.0000000

0.6721064 0.3000000 0.4000000 1.0000000

equal.curve 0.6135464 0.6135464 0.6135464 0.6135464 0.6135464 0.6135464 0.6135464 0.6135464

cor.curve 0.6494405 0.6682358 0.6718323 0.6738308 0.6754113 0.6766433 0.6770946 0.6773447

ols.curve 0.6315652 0.6710561 0.6814868 0.6866801 0.6914484 0.6948195 0.6966165 0.6974865

Equal collinearities and equal validities, R = .4

1.00000

0.26533 1.00000

0.26533 0.30000 1.00000

0.26533 0.30000 0.30000 1.00000

0.26533 0.30000 0.30000 0.30000 1.00000

0.26533 0.30000 0.30000 0.30000 0.30000 1.00000

equal.curve 0.4000000 0.4000000 0.4000000 0.4000000 0.4000000 0.4000000 0.4000000 0.4000000

cor.curve 0.3539056 0.3838697 0.3913093 0.3942170 0.3964141 0.3980029 0.3987162 0.3990390

ols.curve 0.2644249 0.3218332 0.3457424 0.3593424 0.3719094 0.3831947 0.3888669 0.3915397

Equal collinearities and unequal validities, R = .4

1.00000000

0.06786463 1.00000000

0.13572925 0.30000000 1.0000000

0.20359388 0.30000000 0.3000000 1.0000000

0.27145850 0.30000000 0.3000000 0.3000000 1.0000000

0.33932313 0.30000000 0.3000000 0.3000000 0.3000000 1.0000000

equal.curve 0.3069293 0.3069293 0.3069293 0.3069293 0.3069293 0.3069293 0.3069293 0.3069293

cor.curve 0.3021518 0.3391802 0.3508009 0.3551621 0.3589154 0.3604387 0.3613039 0.3615311

ols.curve 0.2634291 0.3210328 0.3460484 0.3592528 0.3720831 0.3834273 0.3887012 0.3915491

Equal collinearities and equal validities, R = .7

1.0000000

0.4643275 1.0000000

0.4643275 0.3000000 1.0000000

0.4643275 0.3000000 0.3000000 1.0000000

0.4643275 0.3000000 0.3000000 0.3000000 1.0000000

0.4643275 0.3000000 0.3000000 0.3000000 0.3000000 1.0000000

equal.curve 0.7000000 0.7000000 0.7000000 0.7000000 0.7000000 0.7000000 0.7000000 0.7000000

cor.curve 0.6878615 0.6951657 0.6969206 0.6977849 0.6985683 0.6991573 0.6994386 0.6995878

ols.curve 0.6282458 0.6675082 0.6789813 0.6844647 0.6897090 0.6938703 0.6959486 0.6969917

Equal collinearities and unequal validities, R = .7

1.0000000

0.1187631 1.0000000

0.2375262 0.3000000 1.0000000

0.3562893 0.3000000 0.3000000 1.0000000

0.4750524 0.3000000 0.3000000 0.3000000 1.0000000

05938155 0.3000000 0.3000000 0.3000000 0.3000000 1.0000000 equal.curve 0.5371263 0.5371263 0.5371263 0.5371263 0.5371263 0.5371263 0.5371263 0.5371263

cor.curve 0.6169993 0.6287804 0.6312070 0.6321259 0.6330300 0.6338158 0.6338347 0.6341159

ols.curve 0.6270844 0.6665128 0.6784282 0.6842493 0.6895946 0.6937984 0.6958965 0.6969344

Unequal collinearities and equal validities, R = .4

1.0000000

0.2611463 1.0000000

0.2611463 0.1500000 1.0000000

0.2611463 0.1830000 0.2830000 1.0000000

0.2611463 0.2170000 0.3170000 0.3830000 1.0000000

0.2611463 0.2500000 0.3500000 0.4170000 0.4500000 1.0000000

equal.curve 0.3936929 0.3936929 0.3936929 0.3936929 0.3936929 0.3936929 0.3936929 0.3936929

cor.curve 0.3458096 0.3773913 0.3851181 0.3878788 0.3902285 0.3918023 0.3924400 0.3927739

ols.curve 0.2631933 0.3214953 0.3462462 0.3588587 0.3724195 0.3833394 0.3886619 0.3915948

Unequal collinearities and unequal validities, R = .4

1.00000000

0.07370387 1.00000000

0.14740775 0.15000000 1.0000000

0.22111162 0.18300000 0.2830000 1.0000000

0.29481549 0.21700000 0.3170000 0.3830000 1.0000000

0.36851937 0.25000000 0.3500000 0.4170000 0.4500000 1.0000000

equal.curve 0.3333383 0.3333383 0.3333383 0.3333383 0.3333383 0.3333383 0.3333383 0.3333383

cor.curve 0.3220801 0.3576407 0.3668152 0.3698606 0.3724839 0.3743262 0.3750006 0.3752513

ols.curve 0.2614803 0.3217418 0.3463975 0.3581450 0.3722649 0.3833072 0.3887453 0.3916041

Unequal collinearities and equal validities, R = .7

1.0000000

0.4570061 1.0000000

0.4570061 0.1500000 1.0000000

0.4570061 0.1830000 0.2830000 1.0000000

0.4570061 0.2170000 0.3170000 0.3830000 1.0000000

0.4570061 0.2500000 0.3500000 0.4170000 0.4500000 1.0000000

equal.curve 0.6889626 0.6889626 0.6889626 0.6889626 0.6889626 0.6889626 0.6889626 0.6889626

cor.curve 0.6766586 0.6840718 0.6859376 0.6866702 0.6875059 0.6881273 0.6884188 0.6885825

ols.curve 0.6278189 0.6669777 0.6787585 0.6842906 0.6896219 0.6939294 0.6958871 0.6969438

Unequal collinearities and unequal validities, R = .7

1.0000000

0.1289818 1.0000000

0.2579636 0.1500000 1.0000000

0.3869453 0.1830000 0.2830000 1.0000000

0.5159271 0.2170000 0.3170000 0.3830000 1.0000000

0.6449089 0.2500000 0.3500000 0.4170000 0.4500000 1.0000000

equal.curve 0.5833420 0.5833420 0.5833420 0.5833420 0.5833420 0.5833420 0.5833420 0.5833420

cor.curve 0.6437610 0.6530680 0.6547794 0.6558432 0.6568583 0.6574783 0.6578118 0.6579022

ols.curve 0.6270285 0.6670084 0.6788489 0.6845266 0.6897414 0.6939889 0.6960538 0.6969995

Appendix C

Study 2 Matrices and Arrays of Mean Validated R Values

Low Magnitude Validities, R = .4

1.0000000

0.1280203 1.0000000

0.2560405 0.2000000 1.0000000

0.3840608 0.3000000 0.4000000 1.0000000

unit.curve 0.3505979 0.3505979 0.3505979 0.3505979 0.3505979 0.3505979 0.3505979 0.3505979

cor.curve 0.2977281 0.3491389 0.3671282 0.3740650 0.3806128 0.3838842 0.3851380 0.3857929

ols.curve 0.2596417 0.3219706 0.3482994 0.3618335 0.3756658 0.3854300 0.3902721 0.3929277

High Magnitude Validities, R = .4

1.0000000

0.2132604 1.0000000

0.2843472 0.2000000 1.0000000

0.3554340 0.3000000 0.4000000 1.000000

unit.curve 0.3893584 0.3893584 0.3893584 0.3893584 0.3893584 0.3893584 0.3893584 0.3893584

cor.curve 0.3126870 0.3636311 0.3802645 0.3863961 0.3913833 0.3945978 0.3957474 0.3963177

ols.curve 0.2600166 0.3228993 0.3489080 0.3619142 0.3748278 0.3855868 0.3903278 0.3927335

Low Magnitude Validities, R = .7

1.0000000

0.1357673 1.0000000

0.4073018 0.2000000 1.0000000

0.6788364 0.3000000 0.4000000 1.0000000

unit.curve 0.6135464 0.6135464 0.6135464 0.6135464 0.6135464 0.6135464 0.6135464 0.6135464

cor.curve 0.6487972 0.6680160 0.6719387 0.6738887 0.6753469 0.6764969 0.6769978 0.6773647

ols.curve 0.6310371 0.6707285 0.6815707 0.6865647 0.6911906 0.6949537 0.6965890 0.6974198

High Magnitude Validities, R = .7

1.0000000

0.3732057 1.0000000

0.4976076 0.2000000 1.0000000

0.6220095 0.3000000 0.4000000 1.0000000

unit.curve 0.6813773 0.6813773 0.6813773 0.6813773 0.6813773 0.6813773 0.6813773 0.6813773

cor.curve 0.6721099 0.6876309 0.6912355 0.6925767 0.6940036 0.6949438 0.6954280 0.6956371

ols.curve 0.6319840 0.6708385 0.6811362 0.6863590 0.6909958 0.6948542 0.6965368 0.6973809

Low Magnitude Validities, R = .4

1.00000000

0.07370387 1.00000000

0.14740775 0.15000000 1.0000000

0.22111162 0.18300000 0.2830000 1.0000000

0.29481549 0.21700000 0.3170000 0.3830000 1.0000000

0.36851937 0.25000000 0.3500000 0.4170000 0.4500000 1.0000000

unit.curve 0.3333383 0.3333383 0.3333383 0.3333383 0.3333383 0.3333383 0.3333383 0.3333383

cor.curve 0.3231196 0.3580199 0.3669760 0.3700612 0.3727942 0.3743768 0.3749873 0.3753003

ols.curve 0.2626082 0.3220292 0.3468212 0.3590352 0.3724678 0.3833418 0.3888257 0.3915246

High Magnitude Validities, R = .4

1.0000000

0.1861954 1.0000000

0.2234345 0.1500000 1.0000000

0.2606736 0.1830000 0.2830000 1.0000000

0.2979126 0.2170000 0.3170000 0.3830000 1.0000000

0.3351517 0.2500000 0.3500000 0.4170000 0.4500000 1.0000000

unit.curve 0.3929802 0.3929802 0.3929802 0.3929802 0.3929802 0.3929802 0.3929802 0.3929802

cor.curve 0.3526633 0.3817674 0.3895504 0.3925352 0.3946781 0.3962026 0.3968745 0.3971893

ols.curve 0.2647336 0.3211058 0.3454586 0.3591084 0.3722585 0.3833750 0.3887280 0.3916349

Low Magnitude Validities, R = .7

1.0000000

0.1289818 1.0000000

0.2579636 0.1500000 1.0000000

0.3869453 0.1830000 0.2830000 1.0000000

0.5159271 0.2170000 0.3170000 0.3830000 1.0000000

0.6449089 0.2500000 0.3500000 0.4170000 0.4500000 1.0000000

unit.curve 0.5833420 0.5833420 0.5833420 0.5833420 0.5833420 0.5833420 0.5833420 0.5833420

cor.curve 0.6440979 0.6530275 0.6548612 0.6558621 0.6567325 0.6575426 0.6577894 0.6578831

ols.curve 0.6274718 0.6674326 0.6787826 0.6845171 0.6897384 0.6939676 0.6960076 0.6970173

High Magnitude Validities, R = .7

1.0000000

0.3258419 1.0000000

0.3910103 0.1500000 1.0000000

0.4561787 0.1830000 0.2830000 1.0000000

0.5213471 0.2170000 0.3170000 0.3830000 1.0000000

0.5865155 0.2500000 0.3500000 0.4170000 0.4500000 1.0000000

unit.curve	0.6877153	0.6877153	0.6877153	0.6877153	0.6877153	0.6877153	0.6877153	0.6877153
cor.curve	0.6844807	0.6917754	0.6935779	0.6944599	0.6952071	0.6958536	0.6961437	0.6963032
ols.curve	0.6293681	0.6671737	0.6788613	0.6843380	0.6896907	0.6939116	0.6959564	0.6970102

Appendix D

Study 3 Matrices and Arrays of Mean Validated R Values

One Dominant Validity, Positively Associated Validities and Collinearities, R = .4

1.0000000

0.1325056 1.0000000

0.1987584 0.2000000 1.0000000

0.3975167 0.3000000 0.4000000 1.0000000

unit.curve 0.3326413 0.3326413 0.3326413 0.3326413 0.3326413 0.3326413 0.3326413 0.3326413

cor.curve 0.2861758 0.3396078 0.3595643 0.3669130 0.3726802 0.3764402 0.3776498 0.3783080

ols.curve 0.2577136 0.3218101 0.3492782 0.3627905 0.3751078 0.3855087 0.3904106 0.3929170

One Dominant Validity, Negatively Associated Validities and Collinearities, R = .4

1.0000000

0.3813012 1.0000000

0.1906506 0.2000000 1.0000000

0.1271004 0.3000000 0.4000000 1.0000000

unit.curve 0.3190722 0.3190722 0.3190722 0.3190722 0.3190722 0.3190722 0.3190722 0.3190722

cor.curve 0.2790435 0.3313539 0.3527280 0.3609847 0.3680640 0.3726569 0.3741052 0.3750345

ols.curve 0.2602713 0.3223425 0.3500845 0.3622169 0.3752020 0.3854417 0.3903544 0.3928253

One Dominant Validity, Positively Associated Validities And Collinearities, R = .7

1.0000000

0.2318848 1.0000000

0.3478271 0.2000000 1.0000000

0.6956543 0.3000000 0.4000000 1.0000000

unit.curve 0.5821223 0.5821223 0.5821223 0.5821223 0.5821223 0.5821223 0.5821223 0.5821223

cor.curve 0.6341561 0.6543631 0.6587662 0.6608397 0.6623491 0.6634812 0.6639921 0.6643030

ols.curve 0.6298076 0.6707193 0.6817341 0.6868910 0.6913616 0.6949861 0.6966753 0.6974981

One Dominant Validity, Negatively Associated Validities and Collinearities, R = .7

1.0000000

0.6672771 1.0000000

0.3336386 0.2000000 1.0000000

0.2224257 0.3000000 0.4000000 1.0000000

unit.curve 0.5583764 0.5583764 0.5583764 0.5583764 0.5583764 0.5583764 0.5583764 0.5583764

cor.curve 0.6202899 0.6469851 0.6527257 0.6550040 0.6567603 0.6582218 0.6588207 0.6592710

ols.curve 0.6291351 0.6700151 0.6811795 0.6865527 0.6911649 0.6949089 0.6965554 0.6974141

Two Dominant Validities, Positively Associated Validities and Collinearities, R = .4

1.0000000

0.1205359 1.0000000

0.3013397 0.2000000 1.0000000

0.3616077 0.3000000 0.4000000 1.0000000

unit.curve 0.3576096 0.3576096 0.3576096 0.3576096 0.3576096 0.3576096 0.3576096 0.3576096

cor.curve 0.3043519 0.3540126 0.3728045 0.3789834 0.3845539 0.3879232 0.3890904 0.3897522

ols.curve 0.2607574 0.3228860 0.3501712 0.3618318 0.3751033 0.3858019 0.3904105 0.3927126

Two Dominant Validities, Negatively Associated Validities and Collinearities, R = .4

1.0000000

0.3305883 1.0000000

0.2754902 0.2000000 1.0000000

0.1101961 0.3000000 0.4000000 1.0000000

unit.curve 0.3269331 0.3269331 0.3269331 0.3269331 0.3269331 0.3269331 0.3269331 0.3269331

cor.curve 0.2753374 0.3271158 0.3488143 0.3574887 0.3647911 0.3690282 0.3707720 0.3713226

ols.curve 0.2609551 0.3208338 0.3489886 0.3615778 0.3750288 0.3854272 0.3902382 0.3926781

Two Dominant Validities, Positively Associated Validities and Collinearities, R = .7

1.0000000

0.2109378 1.0000000

0.5273445 0.2000000 1.0000000

0.6328134 0.3000000 0.4000000 1.0000000

unit.curve 0.6258167 0.6258167 0.6258167 0.6258167 0.6258167 0.6258167 0.6258167 0.6258167

cor.curve 0.6568485 0.6748706 0.6787838 0.6808092 0.6823512 0.6831436 0.6837218 0.6840217

ols.curve 0.6302861 0.6706777 0.6815143 0.6863120 0.6910464 0.6947606 0.6965182 0.6974182

Two Dominant Validities, Negatively Associated Validities and Collinearities, R = .7

1.0000000

0.5785295 1.0000000

0.4821079 0.2000000 1.0000000

0.1928432 0.3000000 0.4000000 1.0000000

unit.curve 0.5721330 0.5721330 0.5721330 0.5721330 0.5721330 0.5721330 0.5721330 0.5721330

cor.curve 0.6172584 0.6416537 0.6462934 0.6485850 0.6504825 0.6514719 0.6518915 0.6521615

ols.curve 0.6302618 0.6700150 0.6801979 0.6856912 0.6908701 0.6946004 0.6964156 0.6973034

One Dominant Validity, Positively Associated Validities and Collinearities, R = .4

1.0000000

0.1287386 1.0000000

0.1716515 0.1500000 1.0000000

0.2145643 0.1830000 0.2830000 1.0000000

0.2574772 0.2170000 0.3170000 0.3830000 1.0000000

0.3862158 0.2500000 0.3500000 0.4170000 0.4500000 1.0000000

unit.curve 0.3493453 0.3493453 0.3493453 0.3493453 0.3493453 0.3493453 0.3493453 0.3493453

cor.curve 0.3269083 0.3610599 0.3697628 0.3726166 0.3752542 0.3768180 0.3773884 0.3777142

ols.curve 0.2622233 0.3219953 0.3468865 0.3590422 0.3724642 0.3833949 0.3888304 0.3915361

One Dominant Validity, Negatively Associated Validities and Collinearities, R = .4

1.0000000

0.3413416 1.0000000

0.2275611 0.1500000 1.0000000

0.1896342 0.1830000 0.2830000 1.0000000

0.1517074 0.2170000 0.3170000 0.3830000 1.0000000

0.1137805 0.2500000 0.3500000 0.4170000 0.4500000 1.0000000

unit.curve 0.3087551 0.3087551 0.3087551 0.3087551 0.3087551 0.3087551 0.3087551 0.3087551

cor.curve 0.2940261 0.3323288 0.3450766 0.3501109 0.3538563 0.3563322 0.3572497 0.3576045

ols.curve 0.2639888 0.3212264 0.3450975 0.3586814 0.3720692 0.3833409 0.3887137 0.3916398

One Dominant Validity, Positively Associated Validities and Collinearities, R = .7

1.0000000

0.2252926 1.0000000

0.3003901 0.1500000 1.0000000

0.3754876 0.1830000 0.2830000 1.0000000

0.4505851 0.2170000 0.3170000 0.3830000 1.0000000

0.6758777 0.2500000 0.3500000 0.4170000 0.4500000 1.0000000

unit.curve 0.6113544 0.6113544 0.6113544 0.6113544 0.6113544 0.6113544 0.6113544 0.6113544

cor.curve 0.6492202 0.6575299 0.6593109 0.6602599 0.6610398 0.6617848 0.6620654 0.6621268

ols.curve 0.6276813 0.6676865 0.6789267 0.6846247 0.6897960 0.6940155 0.6960354 0.6970424

One Dominant Validity, Negatively Associated Validities and Collinearities, R = .7

1.0000000

0.5973478 1.0000000

0.3982319 0.1500000 1.0000000

0.3318599 0.1830000 0.2830000 1.0000000

0.2654879 0.2170000 0.3170000 0.3830000 1.0000000

0.1991159 0.2500000 0.3500000 0.4170000 0.4500000 1.0000000

unit.curve 0.5403214 0.5403214 0.5403214 0.5403214 0.5403214 0.5403214 0.5403214 0.5403214

cor.curve 0.6072231 0.6210588 0.6240256 0.6257769 0.6255501 0.6266644 0.6267954 0.6272150

ols.curve 0.6271853 0.6665696 0.6785640 0.6839434 0.6894747 0.6937586 0.6958938 0.6968989

Four Dominant Validities, Positively Associated Validities and Collinearities, R = .4

1.0000000

0.1301946 1.0000000

0.2169909 0.1500000 1.0000000

0.2603891 0.1830000 0.2830000 1.0000000

0.3037873 0.2170000 0.3170000 0.3830000 1.0000000

0.3471855 0.2500000 0.3500000 0.4170000 0.4500000 1.0000000

unit.curve 0.3794663 0.3794663 0.3794663 0.3794663 0.3794663 0.3794663 0.3794663 0.3794663

cor.curve 0.3480053 0.3790841 0.3864002 0.3889052 0.3911651 0.3926621 0.3934003 0.3937162

ols.curve 0.2627525 0.3222701 0.3462673 0.3591267 0.3722676 0.3833282 0.3888338 0.3915942

Four Dominant Validities, Negatively Associated Validities and Collinearities, R = .4

1.0000000

0.2983460 1.000000

0.2610527 0.150000 1.0000000

0.2237595 0.183000 0.2830000 1.0000000

0.1864662 0.217000 0.3170000 0.3830000 1.0000000

0.1118797 0.250000 0.3500000 0.4170000 0.4500000 1.0000000

unit.curve 0.3260858 0.3260858 0.3260858 0.3260858 0.3260858 0.3260858 0.3260858 0.3260858

cor.curve 0.2986015 0.3362371 0.3482601 0.3532553 0.3559498 0.3583254 0.3590123 0.3595234

ols.curve 0.2618514 0.3206783 0.3454897 0.3589560 0.3716301 0.3830212 0.3887058 0.3916065

Four Dominant Validities, Positively Associated Validities and Collinearities, R = .7

1.0000000

0.2278405 1.0000000

0.3797341 0.1500000 1.0000000

0.4556809 0.1830000 0.2830000 1.0000000

0.5316278 0.2170000 0.3170000 0.3830000 1.0000000

0.6075746 0.2500000 0.3500000 0.4170000 0.4500000 1.0000000

unit.curve 0.6640660 0.6640660 0.6640660 0.6640660 0.6640660 0.6640660 0.6640660 0.6640660

cor.curve 0.6784405 0.6856684 0.6874153 0.6883541 0.6891385 0.6897753 0.6899983 0.6902468

ols.curve 0.6286255 0.6678152 0.6789413 0.6843794 0.6897866 0.6938997 0.6959572 0.6969924

Four Dominant Validities, Negatively Associated Validities and Collinearities, R = .7

1.0000000

0.5221055 1.0000000

0.4568423 0.1500000 1.0000000

0.3915791 0.1830000 0.2830000 1.0000000

0.3263159 0.2170000 0.3170000 0.3830000 1.0000000

0.1957895 0.2500000 0.3500000 0.4170000 0.4500000 1.0000000

unit.curve 0.5706501 0.5706501 0.5706501 0.5706501 0.5706501 0.5706501 0.5706501 0.5706501

cor.curve 0.6138219 0.6254747 0.6276327 0.6281797 0.6292490 0.6297096 0.6300061 0.6301074

ols.curve 0.6265186 0.6667276 0.6783157 0.6838503 0.6894482 0.6936969 0.6958424 0.6969097

Appendix E

Adaptation of Maxwell's Formula for Estimating R^2 from Model Parameters

The purpose of Maxwell's (2000) paper is to provide researchers with statistical tools for estimating the required sample size for obtaining a desired level of power in a regression analysis. As one part of this effort, Maxwell provided (and proved) his equation 7:

$$R^2 = \boldsymbol{\rho}'_{XY}\mathbf{R}_{XX}^{-1}\boldsymbol{\rho}_{XY}$$

where R^2 is the squared coefficient of multiple determination, $\boldsymbol{\rho}_{\mathbf{xy}}$ is the vector of criterion-predictor correlations, and $\mathbf{R}_{\mathbf{xx}}$ is the population predictor correlation matrix. Maxwell provided this formula to allow researchers to estimate the anticipated effect size in a planned study, which in turn could be used in the power analysis.

My study has the advantage of being a simulation, allowing me to select parameter values a priori, but not all combinations of parameter values are actually possible. That is, it is impossible to specify both the coefficient of multiple determination and the predictor-criterion matrix independently and with certainty that the mathematical relationships among regression parameters will hold. Maxwell's equation would be helpful in this regard except that the levels of R^2 are predetermined in my simulation design and therefore cannot be solved for as an unknown. Instead, it is necessary to solve for the values in $\boldsymbol{\rho}_{\mathbf{xy}}$ (which are not part of the design) using a rearranged function (shown to be equivalent in a proof by Andrew Tomarken, personal communication, 2015):

$$\boldsymbol{\rho}_{XY} = \mathbf{v}\sqrt{\frac{R^2}{\mathbf{v}'\mathbf{R}_{XX}^{-1}\mathbf{v}}}$$

where **v** is a vector of values equivalent to the ratios of predictor validities (e.g., if the second of two predictors correlates half as much as the first with the criterion, the values in **v** would be 1 and .5).

As an example, if I wish to generate a design matrix with three predictors, the population R^2 = .5, all predictors correlate .5 with each other, and all predictors correlate equally with the criterion, I must solve for:

$$\boldsymbol{\rho}_{XY} = \begin{bmatrix}1\\1\\1\end{bmatrix}\sqrt{\frac{.5}{\begin{bmatrix}1 & 1 & 1\end{bmatrix}\begin{bmatrix}1.5 & -.5 & -.5\\ -.5 & 1.5 & -.5\\ -.5 & -.5 & 1.5\end{bmatrix}\begin{bmatrix}1\\1\\1\end{bmatrix}}} = \begin{bmatrix}1 & .58 & .58 & .58\\ .58 & 1 & .50 & .50\\ .58 & .50 & 1 & .50\\ .58 & .50 & .50 & 1\end{bmatrix}$$

In this case, each of the predictors has a validity equal to the square root of one third, or approximately .58.

REFERENCES

Campbell, J. P. (1974). *A Monte Carlo approach to some problems inherent in multivariate prediction: With special reference to multiple regression* (No. 2002), Minneapolis, MN: Minnesota University.

Dana, J., & Davis-Stober, C. P. (2016). Rational foundations of fast and frugal heuristics: The ecological rationality of strategy selection via improper linear models. *Minds and Machines*, *26*(1-2), 61-86.

Dana, J., & Dawes, R. M. (2004). The superiority of simple alternatives to regression for social science predictions. *Journal of Educational and Behavioral Statistics*, *29*(3), 317-331.

Davies, P. I., & Higham, N. J. (2000). Numerically stable generation of correlation matrices and their factors. *BIT Numerical Mathematics*, *40*(4), 640-651.

Davis, K. R., & Sauser, W. I. (1991). Effects of alternative weighting methods in a policy-capturing approach to job evaluation: a review and empirical investigation. *Personnel Psychology*, *44*(1), 85-127.

Davis-Stober, C. P. (2011). A geometric analysis of when fixed weighting schemes will outperform ordinary least squares. *Psychometrika*, *76*(4), 650-669.

Davis-Stober, C. P., & Dana, J. (2014). Comparing the accuracy of experimental estimates to guessing: a new perspective on replication and the "Crisis of Confidence" in psychology. *Behavior Research Methods*, *46*(1), 1-14.

Davis-Stober, C. P., Dana, J., & Budescu, D. V. (2010). A constrained linear estimator for multiple regression. *Psychometrika*, *75*(3), 521-541.

Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist*, *34*(7), 571-582.

Dawes, R. M., & Corrigan, B. (1974). Linear models in decision making. *Psychological Bulletin*, *81*(2), 95-106.

Dewan, S. (2015, June 26). Judges replacing conjecture with formula for bail. *The New York Times*. Retrieved from http://www.nytimes.com.

Dunnette, M. D., & Borman, W. C. (1979). Personnel selection and classification systems. *Annual Review of Psychology*, *30*(1), 477-525.

Finch, W. H., Chang, M., Davis, A. S., Holden, J. E., Rothlisberg, B. A., & McIntosh, D. E. (2011). The prediction of intelligence in preschool children using alternative models to regression. *Behavior Research Methods*, *43*(4), 942-952.

German, H., Fortin, M., & Read, D. (2016). Justice judgments: individual self-insight and between-and within-person consistency. *Academy of Management Discoveries*, *2*(1), 33-50.

Gigerenzer, G., & Todd, P. M. (1999). *Simple heuristics that make us smart*. New York: Oxford University Press.

Goldberg, L. R. (1972). Parameters of personality inventory construction and utilization: A comparison of prediction strategies and tactics. *Multivariate Behavioral Research Monographs, 7*, No. 2, Fort Worth, TX: Texas Christian University Press.

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

Graefe, A., Green, K. C., & Armstrong, J. S. (2013). Forecasting. In *Encyclopedia of operations research and management science*, 593-604, USA: Springer.

Greer, A. L., Shinn, M., Kwon, J., & Zuiderveen, S. (2016). Targeting services to individuals most likely to enter shelter: Evaluating the efficiency of homelessness prevention. *Social Service Review*, *90*(1), 130-155.

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

Hattrup, K. (2012). Using composite predictors in personnel selection. *The Oxford handbook of personnel assessment and selection*, 297-319, New York: Oxford University Press.

Holmes, R. B. (1991). On random correlation matrices. *SIAM Journal on Matrix Analysis and Applications*, *12*(2), 239-272.

Kaiser, H.F. & Dickman, K. (1962). Sample and population score matrices and sample correlation matrices from an arbitrary population correlation matrix. *Psychometrika, 27*(2) 179-182.

Laughlin, J. E. (1978). Comment on "Estimating coefficients in linear models: It don't make no nevermind." *Psychological Bulletin*, *85*(2), 247-253.

Mallows, C. L. (1973). Some comments on C_P. *Technometrics*, *15*(4), 661-675.

Marsaglia, G., & Olkin, I. (1984). Generating correlation matrices. *SIAM Journal on Scientific and Statistical Computing*, *5*(2), 470-475.

Maxwell, S. E. (2000). Sample size and multiple regression analysis. *Psychological Methods*, *5*(4), 434-458.

Raju, N. S., Bilgic, R., Edwards, J. E., & Fleer, P. F. (1999). Accuracy of population validity and cross-validity estimation: An empirical comparison of formula-based, traditional empirical, and equal weights procedures. *Applied Psychological Measurement*, *23*(2), 99-115.

Rozeboom, W. W. (1966). *Foundations of the theory of prediction*. Homewood, Illinois: Dorsey Press.

Schmidt, F. L. (1971). The relative efficiency of regression and simple unit predictor weights in applied differential psychology. *Educational and Psychological Measurement , 31*(3), 699-714.

Smolkowski, K., & Cummings, K. D. (2015). Evaluation of diagnostic systems: The selection of students at risk of academic difficulties. *Assessment for Effective Intervention*, *41*(1), 41-54.

Wainer, H. (1976). Estimating coefficients in linear models: It don't make no nevermind. *Psychological Bulletin*, *83*(2), 213-217.

Wainer, H. (1978). On the sensitivity of regression and regressors. *Psychological Bulletin*, *85*(2), 267-273.

Wall, M. M., & Li, R. (2003). Comparison of multiple regression to two latent variable techniques for estimation and prediction. *Statistics in Medicine*, *22*(23), 3671-3685.

Waller, N. G. (2008). Fungible weights in multiple regression. *Psychometrika*, *73*(4), 691-703.

Waller, N. G., & Jones, J. A. (2010). Correlation weights in multiple regression. *Psychometrika*, *75*(1), 58-69.

Waller, N., & Jones, J. (2011). Investigating the performance of alternate regression weights by studying all possible criteria in regression models with a fixed set of predictors. *Psychometrika*, *76*(3), 410-439.

Wherry, R.J., & Naylor, J.C. (1965). Generating multiple samples of multivariate data with arbitrary population parameters *Psychometrika 30*(3), 303-313.

Wilks, S. S. (1938). Weighting systems for linear functions of correlated variables when there is no dependent variable. *Psychometrika*, *3*(1), 23-40.

Yao, L. (2010). Reporting valid and reliable overall scores and domain scores. *Journal of Educational Measurement*, *47*(3), 339-360.