2011

Standardized Regression Coefficients as Indices of Effect Sizes in Meta-Analysis

Rae Seon Kim

THE FLORIDA STATE UNIVERSITY

COLLEGE OF EDUCATION

STANDARDIZED REGRESSION COEFFICIENTS AS INDICES OF EFFECT SIZES IN META-ANALYSIS

By

RAE SEON KIM

A Dissertation submitted to the
Department of Educational Psychology and Learning Systems

in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Degree Awarded:
Summer Semester, 2011

The members of the committee approve the dissertation of RAE SEON KIM defended on June 30, 2011.

_____
Betsy Jane Becker
Professor Directing Dissertation

_____
Fred Huffer
University Representative

_____
Yanyun Yang
Committee Member

_____
Insu Paek
Committee Member

Approved:

_____
Betsy Jane Becker, Chair, Department of Educational Psychology and Learning Systems

_____
Marcy Driscoll, Dean, College of Education

The Graduate School has verified and approved the above-named committee members.

I dedicate this to my mother, Kyungsoon Ju

ACKNOWLEDGEMENTS

I begin my acknowledgements by thanking my academic advisor, Dr. Betsy Jane Becker. She encouraged and supported me on my life and various aspects of research. This dissertation could not have been completed without her help and advice. I cannot find an appropriate word to express my deepest thanks to her. She has always supported and given me direction when I was lost in my research and had hard times in my life during the Ph.D program. I think of her as my academic mother. I have learned so many things her during the past six years. I am so happy and honored that you were my advisor.

I would like to express my gratitude to my dissertation committee members, Dr Yanyun Yang, Dr. Fred Huffer, and Dr. Insu Paek. I am very thankful for their valuable advice, comments, and suggestions for my progress in this program.

Also I would like to express my sincere appreciation to Dr. Akihito Kamata. He supported and gave guidance on various researches until now. Thank you so much for giving me many opportunities to work with you.

Next, I would like to acknowledge my former advisor in Korea, Dr. Jeong-soo Park. He was my first academic advisor during my bachelor and master programs. I am deeply thankful for his continuing interest and valuable advice.

I extend many sincere thanks to my colleagues and friends, Dr. Ying Zhang, Dr. Hirotaka Fukuhara, Dr. Kuzey Bilir, Dr. Soyeon Ahn, Dr. Nansook Yu, Dr. Bernd Weiss, Dr. Qian, Liu, Dr. Ariel Aloe, Christoper Thompson, Seung-Jin Lee, Kyunghwa Cho, Sanghyun Jeon, Haiyan Wu, Sangwook Park, Esra Kocyigit, Jin Koo, Jiyeo Yun, Dr. Soojeong Ingrisone, Dr. James Ingrisone, Sunha Lee, and Kyungok Kim.

Finally, I am grateful to my family, my father, Eun-gyu Kim, my sister, Rae-young Kim, my lovely nephew, Chan Kang, my lovely niece, Honey Kang, my brother, Tae-woo Kim, my brother-in-law, Jin-sun Kang, my father-in-law, Wan Lim. Most importantly, I extend my love and gratitude to my husband, Youngkwon Lim and my precious and adorable daughter, Kayla Lim for their endless love, patience, and support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

When conducting a meta-analysis, it is common to find many collected studies that report regression analyses, because multiple regression analysis is widely used in many fields. Meta-analysis uses effect sizes drawn from individual studies as a means of synthesizing a collection of results. However, indices of effect size from regression analyses have not been studied extensively. Standardized regression coefficients from multiple regression analysis are scale free estimates of the effect of a predictor on a single outcome. Thus these coefficients can be used as effect–size indices for combining studies of the effect of a focal predictor on a target outcome.

I begin with a discussion of the statistical properties of standardized regression coefficients when used as measures of effect size in meta-analysis. The main purpose of this dissertation is the presentation of methods for obtaining standardized regression coefficients and their standard errors from reported regression results. An example of this method is demonstrated using selected studies from a published meta-analysis on teacher verbal ability and school outcomes (Aloe & Becker, 2009). Last, a simulation is conducted to examine the effect of multicollinearity (intercorrelation among predictors), as well as the number of predictors on the distributions of the estimated standardized regression slopes and their variance estimates. This is followed by an examination of the empirical distribution of estimated standardized regression slopes and their variances from simulated data for different conditions. The estimated standardized regression slopes have larger variance and get close to zero when predictors are highly correlated via the simulation study.

CHAPTER ONE

INTRODUCTION

In nearly every field, large quantities of empirical research studies have been conducted over the decades. It is very common for researchers to have similar research questions. Meta-analysis is a technique to combine studies with similar research questions to increase precision and assess the generalizability of results (Cohn & Becker, 2003; Glass, 1976). Glass (1976) defined meta-analysis as "the statistical analysis of a large collection of analysis results from individual studies for the purpose of integrating the findings." (Glass, 1976, p.3).

Meta-analysis uses effect sizes drawn from individual studies. In order to investigate associations among variables of interest, *r* family effect-size indices, specifically Pearson correlation coefficients, are widely used in meta-analysis research (Borenstein, Hedges, Higgins, & Rothstein, 2009; Hedges & Olkin, 1985; Stankowich & Blumstein, 2005). Pearson correlation coefficients represent the bivariate relationship between two variables without controlling for the effects of other variables.

Many researchers have argued for combining correlation coefficients via various approaches (e.g., Becker, 1992, 1995; Hedges & Olkin, 1985; Hunter & Schmidt, 1990, 1994; Shadish & Haddock, 1994). Hedges and Olkin (1985) introduced synthesis methods for simple correlation coefficients. Becker (1992, 1995) introduced methods for combining correlation matrices using a generalized least squares estimation approach for the case of multivariate meta-analysis.

Multiple regression analysis is widely used in primary studies in education, economics, social science, and to a lesser extent in medical research (Armitage, Berry, & Matthews, 2002;

Cohen, Cohen, West, & Aiken, 2003; Howell, 2010; Kieffer, Reese, & Thompson, 2001). Multiple regression analysis allows for the control of the effects of multiple variables. Thus research on indices for combining effects of studies using multiple regression analysis is needed. However, methods for synthesis of regression slopes have rarely been studied (exceptions include Aloe, 2009; Becker & Wu, 2007). In this paper, I will investigate the combination of regression slopes and propose the standardized regression slope as a metric of effect size in meta-analysis.

It is well-known that the standardized regression coefficient in a bivariate regression model is the same as the bivariate correlation coefficient between the independent and dependent variables. In multiple regression analysis, standardized regression coefficients are scale free estimates and are related to correlation coefficients, but the relationship is much more complex than in bivariate regression. For example, consider a regression model with two independent variables, $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$, where y_i is the score on the dependent variable of the i^{th} subject, x_{1i} and x_{2i} are the values of the independent variables for the i^{th} subject, β_0, β_1, and β_2 are population regression coefficients, and ε_i is a residual term, often assumed to be normally distributed with mean of zero and constant variance. The associated standardized regression model is $y_i^* = \beta_1^* z_1 + \beta_2^* z_2 + \varepsilon_i^*$, where β_1^* and β_2^* are the standardized regression coefficients in population. The error term, ε_i^*, is assumed to be normally distributed with mean of zero and variance of σ^{*2}. The least squares estimates of the standardized regression slopes are

$$b_1^* = \frac{r_{Y1} - r_{12} r_{Y2}}{1 - r_{12}^2} \text{ and } b_2^* = \frac{r_{Y2} - r_{12} r_{Y1}}{1 - r_{12}^2},$$

where r_{Y1} is the simple correlation coefficient between Y and x_1, r_{Y2} is the simple correlation coefficient between Y and x_2, and r_{12} is the simple correlation coefficient between x_1 and x_2. The

standardized regression coefficient for the first independent variable, b_1^*, is a function of all the correlation coefficients among the variables. When the intercorrelation between the two independent variables is zero (i.e., $r_{12} = 0$), the standardized regression coefficient, b_1^*, is equal to the correlation coefficient, r_{Y1}. For multiple regression models with more predictors, these formulas are more complex, but the simplification that $b_1^* = r_{Y1}$ holds if all the intercorrelation values among the predictors are zero.

Purpose of the Dissertation

The purpose of this study is to illustrate how the standardized slope can be an effect-size index in meta-analysis, as well as to discuss its strengths and limitations.

I will examine and compare the interpretations of various available indices of regression results. Furthermore, I will focus on the statistical properties of standardized regression coefficients as measures of effect size in meta-analysis. In addition, a method will be presented for obtaining the standard errors of standardized regression slopes. A practical meta-analysis example is used to illustrate combining standardized slopes drawn from a literature on teacher verbal ability and school outcomes (Aloe & Becker, 2009). Finally, a simulation study is conducted in order to examine the effect of multicollinearity and the number of predictors on the distributions of the estimated standardized regression slopes and their variance estimates.

The Organization of the Dissertation

I begin with a brief literature review. This is followed by a method for obtaining standardized slope estimates and a derivation of their variance will be presented. Next, I will illustrate the use of meta-analysis techniques for combining regression slopes. A simulation study will be presented to examine the empirical distributions of the estimated standardized regression slopes and their variance estimates. Finally I will discuss the advantages and limitations of using the standardized regression slope for combining effects.

CHAPTER TWO

LITERATURE REVIEW

In this section I will summarize the methods for issues in combining regression coefficients (Becker & Wu, 2007; Greenland, Maclure, Schlesselman, Poole, & Morgenstern, 1991; Greenland, Schlesselman, & Criqui, 1986; Peterson & Brown, 2005). Next I will discuss the proposed indices for regression slopes in multiple regression models (Aloe, 2009; Greenwald, Hedges, & Laine, 1996; Stavig, 1977). The applied meta-analysis literature on regression coefficients will be reviewed (Paul, Lipps, & Madden, 2006; Yin, Schmidt, & Besag, 2006), and a method of combining *t* statistics from regression slopes will be discussed (Stanley, Doucouliagos, & Jarrell, 2008; Stanley & Jarrell, 1989, 2005). Finally, the partial effect size for dummy slopes in regression models will be discussed (Keef & Roberts, 2004).

Combining Regression Coefficients

Becker and Wu (2007) addressed the issue of synthesizing regression slopes. They described some existing methods for summarizing regression slopes including summaries of raw slopes or *t* statistics of slopes, iterative least squares regression methods, and weighed least squares methods for summarizing regression slopes, as well as a multivariate Bayesian approach. In addition, they discussed a new synthesis approach based on generalized least squares (GLS) estimation using raw regression coefficients.

The authors considered combining all regression coefficients from individual studies to take into account the effects of predictors on a target outcome. However, raw regression coefficients depend on the scale of variables. If variables from collected studies have the same measure or same scale, it is possible to apply this method. In general, the standardized measure for effect size is required for combining regression coefficients.

The GLS estimation is given by $\hat{\boldsymbol{\beta}}' = (\mathbf{W}' \boldsymbol{\Sigma}^{-1} \mathbf{W})^{-1} \mathbf{W}' \boldsymbol{\Sigma}^{-1} \mathbf{b}$, where $\mathbf{W}$ represents a design matrix which contains zeros and ones to identify coefficients from each study, $\boldsymbol{\Sigma}$ is a blockwise diagonal matrix which contains the variance-covariance matrices of the regression coefficients in each study, and $\mathbf{b}$ represents the stack of reported regression coefficients from the collected studies. The proposed GLS method requires covariances and variances among regression coefficients in each study, which are often challenging to obtain.

Becker and Wu (2007) also discussed problems in synthesizing regression slopes in terms of scaling of variables and differences in additional independent variables across studies. I review these issues here.

Scaling Issue

An essential property of most measures of effect size is that they are scale free, which means the magnitudes of effect sizes are comparable across studies. The Pearson correlation coefficient is a scale free index, therefore the correlation coefficient is widely used as an effect size to represent associations among variables. The magnitudes of estimated unstandardized slope parameters in multiple regression depend on the scales of the predictor and outcome variables. However, the standardized slope is interpreted as the estimated number of standard

deviations of change in the dependent variable for one standard deviation unit change in the independent variable, controlling for other independent variables. This index can be compared across studies in much the same way that standardized mean difference effects are compared.

Standardized Regression Slopes

Peterson and Brown (2005) investigated the empirical relationship between simple correlation coefficients and standardized regression slopes. In this study 1,504 standardized regression coefficients and correlation coefficients from published articles in behavioral journals were collected. The authors provided the estimated slope of the regression of the standardized regression slope on the simple correlation coefficient, and found a strong relation between simple correlation coefficients and standardized regression slopes. The main focus of this study was to combine correlation coefficients in meta-analysis, specifically dealing with the case of reporting standardized regression coefficients when correlation coefficients are not reported. The authors proposed a formula to impute the standardized regression slope when studies have no information on correlation coefficients. However, this paper does not discuss computational methods for effect-size variances.

Greenland et al. (1986) and Greenland et al. (1991) criticized the use of the standardized regression coefficient as a measure of effect size, especially for logistic-regression-analysis results. They argued that the range of standard deviations of variables in multiple regression analysis is wide across studies. Another problem they identified, also discussed by Becker and Wu (2007), was that the covariates in multiple regression models can vary across studies. Their argument thus was that comparisons of standardized slopes were not meaningful because the

standard deviations of variables are different from study to study. Specifically, they claim that the interpretation of standardized slopes in biological or public-health contexts is not meaningful because the population standard deviations of predictors may vary. Their argument applies to logistic regression analysis because the outcome variable is dichotomous. Thus, the standard deviation of dichotomous outcome is not meaningful when interpreting the standardized regression slope. In this dissertation, the outcome variable is assumed to be continuous, which allows for meaningful interpretation of the standardized regression coefficients.

Other Indices of Regression Slopes

Semi-standarized and half-standardized regression coefficients also have been proposed as effect-size measures. Semi-standarized regression slopes provide the effects of standardized predictors on unstandardized outcome variables, or the effects of unstandardized predictors on standardized outcome variables. The half-standardized regression coefficient represents the effect of an unstandardized predictor on a standardized outcome variable, and is the one of the semi-standarized regression slopes provided in Stavig (1977). The semi-partial correlation index has been proposed as an effect-size index in multiple regression analysis (Aloe, 2009). The semi-partial correlation index represents the unique effect of the predictor on the outcome variable, partialling out the effects of other predictors in the regression model.

Semi-standardized Partial Regression Coefficient

Stavig (1977) introduced two types of semi-standardized regression coefficients as a function of standardized or unstandardized slopes and the standard deviation of the independent

or dependent variable, specifically, $b_{Sj} = b_j^* S_Y = b_j S_j$ or $b_{Sj}^{'} = b_j^* / S_j = b_j / S_Y$, where b_j^* is a standardized regression slope for j^{th} predictor, b_j is the raw slope in a simple regression model, and S_Y and S_j are the standard deviations for dependent and independent variables, respectively. The semi-standarized regression slopes, b_{Sj}, represent the effects of standardized predictors on unstandardized outcome variables, and $b_{Sj}^{'}$ represent the effects of unstandardized predictors on standardized outcome variables holding constant other predictors in the model. The author directly applied this formula to slopes from multiple regression models. The author compared simple correlations, standardarized and unstandardized coefficients with an example from Anderson, Rosh, and McClary (1973). Stavig (1977) did not provide the standard error of the semistandardized slope.

Half-standardized Partial Regression Coefficient

Greenwald et al. (1996) proposed the half-standardized partial regression coefficient as a measure of effect size. The half-standardized partial regression coefficient was computed by dividing the unstandardized regression slope by the standard deviation of the dependent variable, defined as $\beta_{H_j} = b_j / S_Y$. The half-standardized slope indicates the change in standard-deviation units on the outcome for one unit change of the predictor controlling for other predictors. The half-standardized slope is one of the semi-standarized regression slopes introduced by Stavig (1977) as described earlier. In this study, half-standardized regression slopes were calculated from 60 collected studies. Two meta-analyses with combined significance testing and half-standardized partial regression coefficients were conducted with the median effect size as the

final representation of the combined effect magnitude. Thus, the authors did not compute a weighted average of the effects to take into account the precision of effect size. The variance of effect sizes represents the precision of effect size in each study. This study did not discuss the precision of the effect sizes in the collected studies.

Semi-partial Correlation Index

Aloe (2009) proposed the semi-partial correlation index for synthesizing slopes in multiple regression models. The index he proposed was

$$r_{sp} = t\sqrt{\frac{1-R^2}{n-p-1}},$$

where t is the t statistic for the focal predictor, R^2 is the variance explained by the model, n is the total sample size, and p is the number of predictors. The semi-partial correlation index represents the unique effect of the focal predictor on the target outcome partialling out the effects of other predictors in the model. This index does not include the common effect shared with other predictors on the outcome. Thus, when the number of predictors is increased, the values of the semi-partial correlation index tend to be smaller. A comparison of the semi-partial correlation with the standardized regression slope, while varying the number of predictors and intercorrelations among predictors, is presented in Chapter III.

Aloe (2009) derived a formula of the estimated variance of the semi-partial correlation using the delta method,

$$Var(r_{sp}) = \frac{R_Y^4 - 2R_Y^2 + R_{Y(p)}^2 + 1 - R_{Y(p)}^4}{n},$$

where R_Y^2 represents the proportion of variance explained by the *p* predictor regression model, $R_{Y(p)}^2$ represents the proportion of variance explained by the (*p*-1) predictor regression model (excluding the p^{th} predictor in the first model), which can be obtained by the formula $R_{Y(p)}^2 = R_Y^2 - r_{sp}^2$, and *n* is the total sample size. Since this formula is complicated an alternative way of obtaining the variance of the semi-partial correlation index is presented in the next chapter.

Aloe (2009) also pointed out that the standardized regression slope was one possible index in multiple regression models, but he mentioned that "one of the major shortcomings of synthesizing standardized slopes is that generally primary researchers report beta weights without standard errors." (Aloe, 2009, p. 10). This weakness is addressed in the next chapter.

Summarizing t Statistics

Stanley and Jarrell (1989, 2005, and 2008) argued for the synthesis of multiple regression slopes. The model used in their article was

$$b_j = \beta + \sum_{a=1}^{P} \alpha_a W_{ja} + u_j, \quad j = 1, 2, \cdots, k,$$

where b_j is the reported estimated slope for the target variable from the j^{th} study, β is the true value of that slope, W_{ja} are study characteristic variables (typically called meta-independent variables), α_a is the effect of the a^{th} study characteristic in study *j*, and *k* is the total number of studies. Stanley and Jarrell also introduced the idea of integrating *t* statistics for the slopes from economics research studies (dividing the slope by its standard error). They argued that meta-

regression analysis controls for variation from the different predictors in the studies. The meta-regression model is

$$t_j = \beta\left(\frac{1}{Se_j}\right) + \sum_{a=1}^{P} \frac{\alpha_j W_{ja}}{Se_j} + \frac{u_j}{Se_j}, \quad j = 1, 2, \cdots, k,$$

where Se_j is the standard error of the slope.

The authors mentioned that the "*t*-statistic is a standardized measure of the critical parameter of interest" (Stanley & Jarrell, 1989, p. 304). However, Becker and Wu (2007) criticized this remark, noting that "they did not say what the parameter of interest is. Clearly *t* is not an estimator of β" (Becker &Wu, 2007, p. 418). Also, combinations of the *t* statistics would not explain the magnitude of the effect of interest. The value of the *t* statistic represents statistical significance for the null hypothesis about the slope parameter. In other words, the *t* statistic cannot tell us the extent of the effect of the focal predictor on the outcome variable.

Partial Standardized Mean Difference Effect Size

Keef and Roberts (2004) introduced another partial effect size from the multiple regression model. In their work, the slope represents the standardized mean difference between two groups, so the predictor of interest in the primary studies is a dummy variable. Thus, their multiple regression model was specifically an analysis-of-covariance (ANCOVA) model. The proposed partial effect size is computed by dividing the dummy slope from the model by the square root of the estimated error variance (the mean square error of the ANCOVA model). The dummy slope represents the adjusted mean difference between two groups (for example, experimental and control groups) controlling for other variables or covariates in the model. Keef

and Roberts focused on this standardized function of the dummy slope to represent the *d*-type effect size, but did not discuss effect sizes for slopes of continuous variables.

Applied Meta-analyses Using Regression Coefficients

Two research papers described below conducted meta-analyses which combined regression coefficients. Both accessed raw data to obtain regression coefficients for each study. Yin et al. (2006) combined standardized regression coefficients and Paul et al. (2006) combined intercepts and raw slopes.

Yin et al. (2006) used standardized slopes as effect sizes in synthesizing studies, but it is not clear what standard errors they used. They investigated trends in student achievement tests over a 3- or 4 year period across 17 urban school districts. The achievement score and time interval variables were standardized. They also calculated the standardized slopes for those two variables across districts. Standardized slopes were computed with raw data for each district and used to obtain the inverse variance weighted mean. Yin et al. did not discuss how to calculate the standard error of the standardized regression slope from reported regression results.

Paul et al. (2006) conducted a meta-analysis with a synthesis of regression slopes and intercepts from bivariate regression models. They summarized the unstandardized slopes and intercepts from 126 studies of the relation between deoxynivalenol content of harvested wheat grain and Fusarium head blight index. The focal predictor and the target outcome had same scale of measurement across the collected studies. From p.14, before conducting the meta-analysis, the

authors estimated intercepts and slopes in a simple regression model from 126 studies. Next overall mean effect sizes in terms of average slope and average intercept under the random-effects model were estimated. However, in general, unstandardized raw regression coefficients depend on the scales of variables. Last, they conducted moderator analyses with study-characteristic variables under the mixed-effects model.

CHAPTER THREE

METHODS

In this section, the estimation of standardized regression coefficients in the cases of two or three predictor regression models is addressed. The ordinary least squares estimators of the standardized regression parameters are presented. In addition, a method is presented for obtaining the standard errors of standardized regression slopes. An investigation of the difference between the standardized regression slope in a two-predictor model and the simple correlation coefficient is presented. Finally I compare the standardized regression slope with the semi-partial correlation while varying the number of predictors and intercorrelations among predictors.

The Case of the Two Predictor Regression Model

When two independent variables are included in a multiple regression model, the model can be defined as $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$, $i = 1, 2, \cdots, n$, where y is the dependent variable, x_1 and x_2 are independent variables, β_0 is the intercept parameter, β_1 and β_2 are slope parameters in the population, and ε_i is the error term assumed to be normally distributed with mean of zero and constant variance. The subscript i represents the i^{th} individual and n is the sample size. The parameter estimates are denoted by b_0, b_1, and b_2, respectively.

To obtain standardized regression coefficients, we can transform the variables into standardized form. From the multiple regression equation with two predictors, the mean of the

dependent variable is obtained as $\bar{y} = \beta_0 + \beta_1 \bar{x}_1 + \beta_2 \bar{x}_2$. The mean-deviation form for the above regression model, subtracting $\bar{y}$, is

$$y_i - \bar{y} = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i - (\beta_0 + \beta_1 \bar{x}_1 + \beta_2 \bar{x}_2)$$

$$= \beta_1 (x_{1i} - \bar{x}_1) + \beta_2 (x_{2i} - \bar{x}_2) + \varepsilon_i .$$

Then we divide by the standard deviation of *y* on both sides, and multiply each term on the right side by 1 in the form of $\frac{S_{x_j}}{S_{x_j}}$, to obtain

$$\frac{y_i - \bar{y}}{S_y} = \left(\beta_1 \frac{S_{x_1}}{S_y} \right) \frac{x_{1i} - \bar{x}_1}{S_{x_1}} + \left(\beta_2 \frac{S_{x_2}}{S_y} \right) \frac{x_{2i} - \bar{x}_2}{S_{x_2}} + \frac{\varepsilon_i}{S_y} .$$

This equation can be denoted by $y_i^* = \beta_1^* z_1 + \beta_2^* z_2 + \varepsilon_i^*$. The transformed variables are $y_i^* = (y_i - \bar{y}) / S_y$, $z_{1i} = (x_{1i} - \bar{x}_1) / S_{x_1}$, and $z_{2i} = (x_{2i} - \bar{x}_2) / S_{x_2}$, where $\bar{y}$, $\bar{x}_1$, and $\bar{x}_2$ are the respective means of each variable, and S_y, S_{x_1}, and S_{x_2} are the respective standard deviations defined as follows:

$$S_y = \sqrt{\frac{\sum (y_i - \bar{y})^2}{n-1}} , \; S_{x_1} = \sqrt{\frac{\sum (x_{1i} - \bar{x}_1)^2}{n-1}} , \; S_{x_2} = \sqrt{\frac{\sum (x_{2i} - \bar{x}_2)^2}{n-1}} .$$

The transformed variables, y^*, z_1, and z_2 are standardized versions of each variable and have means equal to zero and standard deviations equal to one. The transformed error term is $\varepsilon_i^* = \varepsilon_i / S_y$.

The matrix form of the regression model with two predictors is $\mathbf{Y}^* = \mathbf{Z}\boldsymbol{\beta}^* + \boldsymbol{\varepsilon}^*$, where the matrices of $\mathbf{Y}^*$, $\mathbf{Z}$, $\boldsymbol{\beta}^*$, and $\boldsymbol{\varepsilon}^*$ are defined as

$$\mathbf{Y}^* = \begin{bmatrix} y_1^* \\ y_2^* \\ \vdots \\ y_n^* \end{bmatrix},\ \mathbf{Z} = \begin{bmatrix} z_{11} & z_{12} \\ z_{21} & z_{22} \\ \vdots & \vdots \\ z_{n1} & z_{n2} \end{bmatrix},\ \boldsymbol{\beta}^* = \begin{bmatrix} \beta_1^* \\ \beta_2^* \end{bmatrix}, \text{ and } \boldsymbol{\varepsilon}^* = \begin{bmatrix} \varepsilon_1^* \\ \varepsilon_2^* \\ \vdots \\ \varepsilon_n^* \end{bmatrix}.$$

The least squares estimator of the slope parameter, $\boldsymbol{\beta}^*$, is obtained by $\mathbf{b}^* = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}^*$.

The $\mathbf{Z}'\mathbf{Z}$ matrix is given by

$$\mathbf{Z}'\mathbf{Z} = \begin{bmatrix} z_{11} & z_{21} & \cdots & z_{n1} \\ z_{12} & z_{22} & \cdots & z_{n2} \end{bmatrix} \begin{bmatrix} z_{11} & z_{12} \\ z_{21} & z_{22} \\ \vdots & \vdots \\ z_{n1} & z_{n2} \end{bmatrix}$$

$$= \begin{bmatrix} \sum z_{i1}^2 & \sum z_{i1}z_{i2} \\ \sum z_{i1}z_{i2} & \sum z_{i2}^2 \end{bmatrix}.$$

The element in the first row and first column of $\mathbf{Z}'\mathbf{Z}$ is

$$\sum z_{i1}^2 = \sum \left(\frac{x_{i1} - \bar{x}_1}{S_{X_1}} \right)^2 = \frac{\sum (x_{i1} - \bar{x}_1)^2}{S_{X_1}^2}$$

$$= \frac{(n-1)\sum (x_{i1} - \bar{x}_1)^2 / (n-1)}{S_{X_1}^2} = \frac{(n-1)S_{X_1}^2}{S_{X_1}^2}$$

$$= n - 1.$$

This shows that the second diagonal element of $\mathbf{Z}'\mathbf{Z}$ is n-1 as well. The off-diagonal elements of $\mathbf{Z}'\mathbf{Z}$ are equal to

$$\sum z_{i1}z_{i2} = \sum\left(\frac{x_{i1}-\bar{x}_1}{S_{X_1}}\right)\left(\frac{x_{i2}-\bar{x}_2}{S_{X_2}}\right) = \frac{\sum(x_{i1}-\bar{x}_1)(x_{i2}-\bar{x}_2)}{S_{X_1}S_{X_2}}$$

$$= \frac{(n-1)\sum(x_{i1}-\bar{x}_1)(x_{i2}-\bar{x}_2)/(n-1)}{S_{X_1}S_{X_2}}$$

$$= (n-1)\,r_{12},$$

where r_{12} is the simple correlation coefficient between X_1 and X_2.

Substituting the elements, the $\mathbf{Z'Z}$ matrix is

$$\mathbf{Z'Z} = \begin{bmatrix} \sum z_{i1}^2 & \sum z_{i1}z_{i2} \\ \sum z_{i1}z_{i2} & \sum z_{i2}^2 \end{bmatrix}$$

$$= \begin{bmatrix} (n-1) & (n-1)\,r_{12} \\ (n-1)\,r_{12} & (n-1) \end{bmatrix}$$

$$= (n-1)\begin{bmatrix} 1 & r_{12} \\ r_{12} & 1 \end{bmatrix},$$

The inverse of the $\mathbf{Z'Z}$ matrix is

$$(\mathbf{Z'Z})^{-1} = \frac{1}{(n-1)(1-r_{12}^2)}\begin{bmatrix} 1 & -r_{12} \\ -r_{12} & 1 \end{bmatrix}.$$

Now we shift our attention to the $\mathbf{Z}'\mathbf{Y}^*$ matrix.

$$\mathbf{Z}'\mathbf{Y}^* = \begin{bmatrix} z_{11} & z_{21} & \cdots & z_{n1} \\ z_{12} & z_{22} & \cdots & z_{n2} \end{bmatrix}\begin{bmatrix} y_1^* \\ \vdots \\ y_n^* \end{bmatrix}$$

$$= \begin{bmatrix} \sum z_{i1}y_i^* \\ \sum z_{i2}y_i^* \end{bmatrix}.$$

The first element of $\mathbf{Z}'\mathbf{Y}^*$ is

$$\sum z_{i1} y_i^* = \sum \left(\frac{x_{i1} - \bar{x}_1}{S_{X_1}} \right) \left(\frac{y_i - \bar{y}}{S_Y} \right) = \frac{\sum (x_{i1} - \bar{x}_1)(y_i - \bar{y})}{S_{X_1} S_Y}$$

$$= \frac{(n-1) \sum (x_{i1} - \bar{x}_1)(y_i - \bar{y}) / (n-1)}{S_{X_1} S_Y}$$

$$= (n-1)\, r_{Y1}.$$

Similarly, the second element of $\mathbf{Z}'\mathbf{Y}^*$ is

$$\sum z_{i2} y_i^* = (n-1)\, r_{Y2}\,.$$

Here r_{Y1} is the simple correlation coefficient between Y and X_1, and r_{Y2} is the simple correlation coefficient between Y and X_2.

After substituting these two matrices into the equation for the least squares estimator, the estimated standardized coefficients are

$$\mathbf{b}^* = (\mathbf{Z}'\mathbf{Z})^{-1} \mathbf{Z}'\mathbf{Y}^*$$

$$= \frac{1}{(n-1)(1-r_{12}^2)} \begin{bmatrix} 1 & -r_{12} \\ -r_{12} & 1 \end{bmatrix} \begin{bmatrix} (n-1) r_{Y1} \\ (n-1) r_{Y2} \end{bmatrix}$$

$$= \frac{1}{(1-r_{12}^2)} \begin{bmatrix} r_{Y1} - r_{12} r_{Y2} \\ r_{Y2} - r_{12} r_{Y1} \end{bmatrix}.$$

Thus,

$$b_1^* = \frac{r_{Y1} - r_{12} r_{Y2}}{1 - r_{12}^2}\,, \text{ and } b_2^* = \frac{r_{Y2} - r_{12} r_{Y1}}{1 - r_{12}^2}\,.$$

The regression coefficients b_1^* and b_2^* are called standardized regression coefficients.

Variance-Covariance Matrix for Standardized Slopes

The Case of the Two Predictor Regression Model

The variance and covariance matrix of the least squares slope estimator is

$$Var(\mathbf{b}^*) = (\mathbf{Z}'\,\mathbf{Z})^{-1}\hat{\sigma}^{*2}$$

$$= \frac{1}{(n-1)(1-r_{12}^2)}\begin{bmatrix} 1 & -r_{12} \\ -r_{12} & 1 \end{bmatrix}\hat{\sigma}^{*2}$$

$$= \begin{bmatrix} \dfrac{\hat{\sigma}^{*2}}{(n-1)(1-r_{12}^2)} & \dfrac{-r_{12}\hat{\sigma}^{*2}}{(n-1)(1-r_{12}^2)} \\ \dfrac{-r_{12}\hat{\sigma}^{*2}}{(n-1)(1-r_{12}^2)} & \dfrac{\hat{\sigma}^{*2}}{(n-1)(1-r_{12}^2)} \end{bmatrix},$$

where $\hat{\sigma}^{*2}$ is the variance of the error terms in the standardized regression model. The standard errors of the standardized regression coefficients in the case of two independent variables are obtained by taking the square roots of the diagonal elements of the above matrix. Therefore

$$SE(b_1^*) = SE(b_2^*) = \frac{\hat{\sigma}^{*2}}{(n-1)(1-r_{12}^2)}\,.$$

Most statistical computer packages such as SAS, SPSS, R, or MINITAB do not provide the standard errors of standardized regression coefficients. However, if the mean squared error and intercorrelation among *X*s are given for the two-predictor model, the standard errors of standardized slopes can be calculated directly.

The Case of the Three Predictor Regression Model

The multiple regression model with three predictors is written as $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \varepsilon_i$, $i = 1, 2, \cdots, n$. The standardized transformed model is $y_i^* = \beta_1^* z_1 + \beta_2^* z_2 + \beta_3^* z_3 + \varepsilon_i^*$. The matrix form of the regression model with three predictors is $\mathbf{Y}^* = \mathbf{Z}\boldsymbol{\beta}^* + \boldsymbol{\varepsilon}^*$. The least squares estimator of the slope parameter is again obtained by $\mathbf{b}^* = (\mathbf{Z'Z})^{-1}\mathbf{Z'Y}^*$. Now $\mathbf{Z}$ is the matrix of standardized predictors defined by

$$\mathbf{Z} = \begin{bmatrix} z_{11} & z_{12} & z_{13} \\ z_{21} & z_{22} & z_{22} \\ \vdots & \vdots & \vdots \\ z_{n1} & z_{n2} & z_{n3} \end{bmatrix},$$

The product $\mathbf{Z'Z}$ is given by

$$\mathbf{Z'\,Z} = \begin{bmatrix} \sum z_{i1}^2 & \sum z_{i1}z_{i2} & \sum z_{i1}z_{i3} \\ \sum z_{i1}z_{i2} & \sum z_{i2}^2 & \sum z_{i2}z_{i3} \\ \sum z_{i1}z_{i3} & \sum z_{i2}z_{i3} & \sum z_{i3}^2 \end{bmatrix}$$

$$= (n-1)\begin{bmatrix} 1 & r_{12} & r_{13} \\ r_{12} & 1 & r_{23} \\ r_{13} & r_{23} & 1 \end{bmatrix}$$

$$= (n-1)\hat{\mathbf{\Lambda}}.$$

The inverse of the $\mathbf{Z'Z}$ matrix is given by

$$(\mathbf{Z}'\,\mathbf{Z})^{-1} = \frac{1}{(n-1)\,|\,\hat{\mathbf{\Lambda}}\,|}\begin{bmatrix} r_{23}^2-1 & r_{12}-r_{13}r_{23} & r_{13}-r_{12}r_{23} \\ r_{12} & r_{13}^2-1 & r_{23}-r_{12}r_{13} \\ r_{13}-r_{12}r_{23} & r_{23}-r_{12}r_{13} & r_{12}^2-1 \end{bmatrix},$$

where $|\,\hat{\mathbf{\Lambda}}\,|$ is the determinant of $\hat{\mathbf{\Lambda}}$ which is

$$|\,\hat{\mathbf{\Lambda}}\,| = r_{12}^2 + r_{13}^2 + r_{23}^2 - 2r_{12}r_{13}r_{23} - 1.$$

The variance-covariance matrix of the least squares slope estimator, *Var*($\mathbf{b}^*$) is

$$(\mathbf{Z}'\,\mathbf{Z})^{-1}\hat{\sigma}^{*2}$$
$$= \frac{\hat{\sigma}^{*2}}{(n-1)(r_{12}^2 + r_{13}^2 + r_{23}^2 - 2r_{12}r_{13}r_{23} - 1)}\begin{bmatrix} r_{23}^2-1 & r_{12}-r_{13}r_{23} & r_{13}-r_{12}r_{23} \\ r_{12} & r_{13}^2-1 & r_{23}-r_{12}r_{13} \\ r_{13}-r_{12}r_{23} & r_{23}-r_{12}r_{13} & r_{12}^2-1 \end{bmatrix}$$

The diagonal elements of the above matrix are the variances of the least squares standardized slope estimators. (3.1)

Alternative Ways of Obtaining the Standard Error

Three possible alternative ways of obtaining the standard error of standardized slope are addressed in this section.

Using t Statistics for Slopes

The null hypothesis for statistical testing for standardized regression coefficients is given by $H_0 : \beta_j^* = 0,\ \ j = 1, 2, \cdots, p$. This can be tested using *t* statistics, specifically

$$t_j = \frac{b_j^*}{\mathrm{SE}(b_j^*)},\ \ j = 1, 2, \cdots, p.$$

Also, the statistical tests for the raw regression coefficients are

$$t_j = \frac{b_j}{\mathrm{SE}(b_j)},\ \ j = 1, 2, \cdots, p.$$

The *t* statistics for both standardized and unstandardized slopes are the same, specifically $t_j = \frac{b_j^*}{\mathrm{SE}(b_j^*)} = \frac{b_j}{\mathrm{SE}(b_j)}$. Cohen et al. (2003) stated that "all partial coefficients (partial, semi-partial, standardized slope, and raw slope) for X_i must share the same *t* (or *F*) value." (p. 112). Statistical packages such as SAS, SPSS, and MINITAB report a slope test *t*-statistic value along with the raw and standardized estimated slopes. The standard error for the standardized slope can thus be obtained by algebraic manipulation of the *t* statistic. Specifically, because

$t_j = \frac{b_j^*}{SE(b_j^*)}$, then $t_j \times SE(b_j^*) = b_j^*$, thus $SE(b_j^*) = \frac{b_j^*}{t_j}$.

This approach can be applied to the regression model with any number of predictors.

Using Standard Deviations of Variables and the SE of the Raw Slope

The estimated unstandardized regression slope for the j^{th} predictor is given by $b_j = b_j^* \frac{S_Y}{S_{X_j}}$ in the general regression model. The estimated standardized regression slope is obtained by $b_j^* = b_j \frac{S_{X_j}}{S_Y}$, when unstandardized regression slopes and standard deviations of the dependent and the predictor variables are given. The standard error of b_j^* is given by $SE(b_j^*) = SE(b_j)\frac{S_{X_j}}{S_Y}$ (Cohen et al., 2003, p.109) under the large sample condition.

Using the Variance Inflation Factor

The standard error of the standardized regression slope for the j^{th} predictor in the *p*-predictor regression model is given by Cohen et al. (2003, p.112) as

$$SE(b_j^*) = \sqrt{\frac{1-R^2}{(1-R_j^2)(n-p-1)}},$$

where R^2 is the proportion of variance explained by the *p*-predictor regression model, *p* is the number of predictors in the model, and *n* is the total sample size. Also in the above equation, R_j^2 represents the squared multiple correlation between X_j and the set of all other predictors. The Variance Inflation Factor (VIF) for diagnosing multicollinearity for the j^{th} predictor in multiple regression model is given by $VIF_j = \frac{1}{1-R_j^2}$ (Howell, 2010). The standard error formula can be rewritten as a function of the VIF :

$$SE(b_j^*) = \sqrt{\frac{1-R^2}{(1-R_j^2)(n-p-1)}}$$

$$= \sqrt{\frac{1}{1-R_j^2}}\sqrt{\frac{1-R^2}{n-p-1}}$$

$$= \sqrt{VIF_j}\sqrt{\frac{1-R^2}{n-p-1}}.$$

When a study reports the Variance Inflation Factor, R^2, n, and p, the standard error of standardized slope can be obtained by the above formula. However, the VIF is rarely reported in primary studies.

From the formula above, the relation between the standardized slope and semi-partial correlation index can also be shown. The standard error of the standardized slope can be written as $SE(b_j^*) = b_j^* / t_j$ as shown in the previous section. If we plug in b_j^* / t_j into the above equation, the above equation can be rewritten as $\frac{b_j^*}{t_j} = \sqrt{VIF_j}\sqrt{\frac{1-R^2}{n-p-1}}$. Then move t_j to the right part of the equation, then it turned out $b_j^* = \sqrt{VIF_j}\left(t_j\sqrt{\frac{1-R^2}{n-p-1}}\right)$. Finally, the standardized slope is the square root of the VIF multiplied by the semi-partial correlation (r_{sp_j}), specifically, $b_j^* = r_{sp_j}\sqrt{VIF_j}$.

In addition, it can be shown that the standard error of the standardized slope can be written as the standard error of the semi-partial correlation times the square root of the VIF which is $SE(b_j^*) = SE(r_{sp_j})\sqrt{VIF_j}$. Because the *t* statistic is same for all partial coefficients, the *t*

statistic for the semi-partial correlation is $t_j = r_{sp_j} / SE(r_{sp_j})$ (Cohen et al., 2003). The semi-partial correlation formula contains the *t* statistic, which is $r_{sp} = t\sqrt{\frac{1-R^2}{n-p-1}} = t \times SE(r_{sp})$, thus the standard error of r_{sp} is $\sqrt{\frac{1-R^2}{n-p-1}}$.

Summary of the Relation among Raw and Standardized Slopes, and Semi-partial Correlations

In summary, the standardized regression slope is related to the raw slope and the semi-partial correlation. The relation among these three values can be written as $b^* = b\frac{S_X}{S_Y} = r_{sp}\sqrt{VIF}$, and the relations among the standard errors of these estimates can be written as $SE(b^*) = SE(b)\frac{S_X}{S_Y} = SE(r_{sp})\sqrt{VIF}$ under the large sample condition.

Investigating the Difference between the Standardized Slope and Correlation Coefficient

As described earlier, the standardized slope for the first independent variable (X_1) is $b_1^* = \frac{r_{Y1} - r_{12}r_{Y2}}{1-r_{12}^2}$ when two independent variables are in the regression model. Figure 3.1 shows the difference between the standardized slope (b_1^*) and bivariate correlation coefficient (r_{Y1}) between *Y* and X_1 as a function of the correlation (r_{12}) between X_1 and X_2, and the correlation

between Y and X_2 (r_{Y2}). The values in the figure are obtained directly from the equation for the difference, which is $r_{Y1} - b_1^* = r_{Y1} - \frac{r_{Y1} - r_{12}r_{Y2}}{1 - r_{12}^2}$. The selected values for r_{Y1}, r_{12}, and r_{Y2} are .1, .3, .5, and .7. The vertical axis values in the figure represent the difference between r_{Y1} and b_1^*, and the horizontal axis represents the intercorrelation of the two independent variables. Different symbols are plotted for each value of the correlation of Y and X_2 (denoted as r_{Y2}).

When the intercorrelation between independent variables (r_{12}) is .1, the difference between the estimated standardized slope (b_1^*) and the correlation coefficient (r_{Y1}) is close to zero. As the intercorrelation r_{12} increases, the magnitude of the difference between the standardized slope b_1^* and the correlation coefficient r_{Y1} deviates from zero, except in some conditions. Even though a high correlation between two predictors (e.g., $r_{12} = .7$) exists, the difference between b_1^* and r_{Y1} is close to zero in certain conditions. For example, the condition with $r_{12} = .7$, $r_{Y1} = .7$, and $r_{Y2} = .5$ shows that b_1^* is close to r_{Y1} (the difference between them is zero). Another example of this situation is the condition where $r_{12} = .7$, $r_{Y1} = .1$, and $r_{Y2} = .1$.

The correlation between Y and X_2 (r_{Y2}) is another factor that affects the difference between the standardized slope, b_1^*, and the correlation coefficient, r_{Y1}. A strong relationship between Y and X_2 (e.g., $r_{Y2} = .7$) produces large differences between b_1^* and r_{Y1} across values of r_{12} and r_{Y1}. As shown in Figure 3.1, the differences for $r_{Y2} = .7$ are plotted as triangles; these are always the largest differences in each plot.

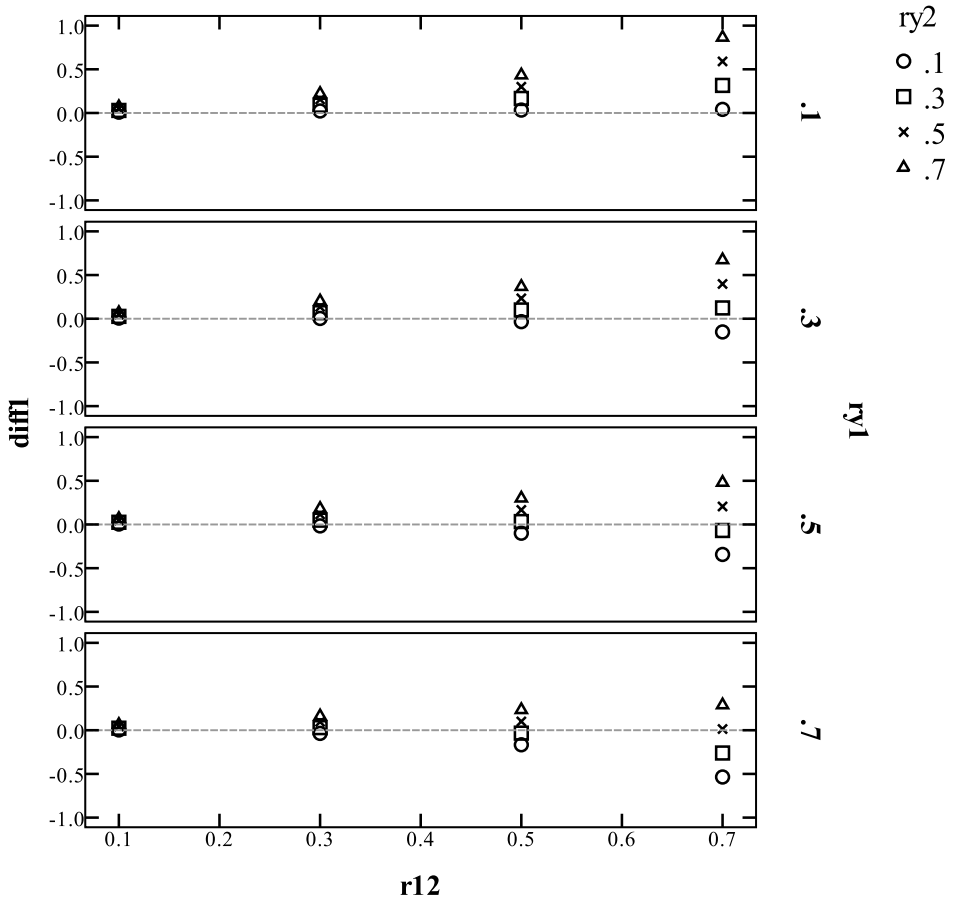

Figure 3.1. The differences between standardized slopes (b1*) and correlation coefficients (ry1) as a function of ry2 and r12 for the two predictor model

Comparison of the Standardized Regression Slope and the Semi-partial Correlation

In this section, I compare the standardized regression slope (b^*) with the semi-partial correlation (r_{sp}) while varying the number of predictors and intercorrelations among predictors. The number of predictors, p, will be set to $p = 2$, 5 and 10, while the intercorrelations, ρ, among predictors vary from $\rho = .1$ to $\rho = .9$ by .1 increments. The correlation between the outcome and each predictor is set to .4. The comparison is conducted by direct computation from the equations described in this chapter.

Two Predictor Model

A regression model with two independent variables is $y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \varepsilon_i$, where y_i is the score on the dependent variable of the i^{th} subject, x_{1i} and x_{2i} are the values of the independent variables for the i^{th} subject, β_0, β_1, and β_2 are population regression coefficients, and ε_i is a residual term, often assumed to be normally distributed with mean of zero and constant variance. The associated standardized regression model is $y_i^* = \beta_1^* z_1 + \beta_2^* z_2 + \varepsilon_i^*$, where β_1^* and β_2^* are the standardized regression coefficients in the population. The error term, ε_i^*, is assumed to be normally distributed with mean of zero and variance of σ^{*2}. The least squares estimator of the standardized regression slope for the first predictor is

$b_1^* = \dfrac{r_{Y1} - r_{12} r_{Y2}}{1 - r_{12}^2}$, where the correlation matrix is $\mathbf{r} = \begin{bmatrix} 1 & r_{Y1} & r_{Y2} \\ r_{Y1} & 1 & r_{12} \\ r_{Y2} & r_{12} & 1 \end{bmatrix}$.

The semi-partial correlation coefficient is obtained by $r_{sp} = \frac{r_{Y1} - r_{12}r_{Y2}}{\sqrt{1-r_{12}^2}}$. The ratio of the standardized regression slope to the semi-partial correlation is

$$\frac{b^*}{r_{sp}} = \frac{1}{\sqrt{1-r_{12}^2}} = \sqrt{\frac{1}{1-r_{12}^2}} = \sqrt{VIF} \ .$$

Comparison Results

Table 3.1 and Figure 3.2 show the comparison of the standardized regression slope and the semi-partial correlation while varying the number of predictors and intercorrelations. The first column in Table 3.1 represents the number of predictors, the second column is the intercorrelation among predictors, the third column is the standardized regression slope, the fourth column is the semi-partial correlation coefficient, the fifth column shows the difference between the standardized regression slope and the semi-partial correlation coefficient, the sixth column represents the ratio of β^* and ρ_{sp}, and the last column represents the squared ratio which is the same as the VIF (variance inflation factor).

Figure 3.2 plots the values found in Table 3.1. The Y-axis represents the standardized regression slope and the semi-partial correlation and the X-axis represents the intercorrelation values. Three separate panels correspond to the number of predictors. As shown in Figure 3.2, the semi-partial correlation coefficient dramatically decreases across higher intercorrelations for all panels. This plot is for one condition of ρ_{YX} = .4, and similar patterns could be made for other ρ_{YX} values except location of values.

Table 3.1. *Comparison of the Standardized Regression Slope and the Semi-partial Correlation*

p	$\rho(x_i,x_j)$	β^*	ρ_{sp}	$\beta^* - \rho_{sp}$	β^* / ρ_{sp}	VIF
2	0.1	0.364	0.362	0.002	1.005	1.010
	0.2	0.333	0.327	0.007	1.021	1.042
	0.3	0.308	0.294	0.014	1.048	1.099
	0.4	0.286	0.262	0.024	1.091	1.190
	0.5	0.267	0.231	0.036	1.155	1.333
	0.6	0.250	0.200	0.050	1.250	1.563
	0.7	0.235	0.168	0.067	1.400	1.961
	0.8	0.222	0.133	0.089	1.667	2.778
	0.9	0.211	0.092	0.119	2.294	5.263
5	0.1	0.286	0.281	0.004	1.016	1.032
	0.2	0.222	0.211	0.011	1.054	1.111
	0.3	0.182	0.164	0.018	1.111	1.234
	0.4	0.154	0.130	0.024	1.188	1.410
	0.5	0.133	0.103	0.030	1.291	1.667
	0.6	0.118	0.082	0.036	1.435	2.059
	0.7	0.105	0.064	0.041	1.649	2.719
	0.8	0.095	0.047	0.048	2.012	4.048
	0.9	0.087	0.031	0.056	2.836	8.043
10	0.1	0.211	0.205	0.005	1.026	1.053
	0.2	0.143	0.133	0.010	1.077	1.161
	0.3	0.108	0.094	0.014	1.146	1.313
	0.4	0.087	0.070	0.016	1.234	1.522
	0.5	0.073	0.054	0.019	1.348	1.818
	0.6	0.063	0.042	0.021	1.505	2.266
	0.7	0.055	0.032	0.023	1.736	3.014
	0.8	0.049	0.023	0.026	2.124	4.512
	0.9	0.044	0.015	0.029	3.002	9.011

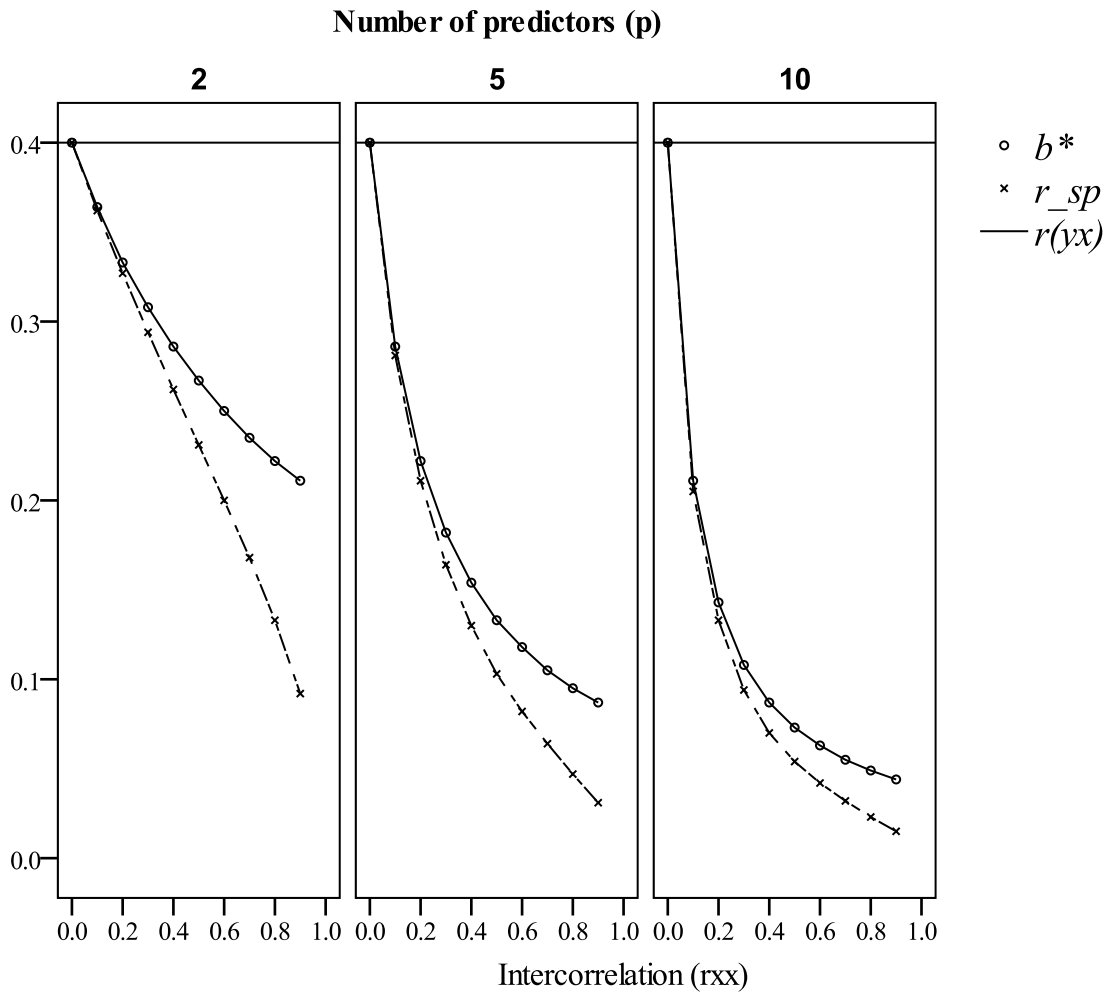

Figure 3.2. Comparison of the Standardized Regression Slope and the Semi-partial Correlation

CHAPTER FOUR

EXAMPLE

This section illustrates methods for combining standardized regression slopes using data from a published meta-analysis article. A practical meta-analysis example is illustrated for combining standardized slopes drawn from the study of teacher verbal ability and school outcomes (Aloe & Becker, 2009).

Data Description

The example data set includes standardized slopes from Aloe and Becker (2009). They synthesized studies of the relationship between teachers' verbal ability and school outcomes using semipartial and bivariate correlation indices. In their meta-analysis, eight studies reported regression analysis results and five of those reported standardized regression coefficients.

Table 4.1 shows the list of reported standardized slopes and *t* statistics. The standardized slopes represent the relationship between teacher's verbal ability score and student achievement. Bowles (1970), Ferguson (1991), and Murnane (1981) reported standardized coefficients and corresponding *t* statistics. Bowles (1970) reported results from three different samples. Ferguson (1991) presented six results from different grade levels. The standard errors of the reported standardized coefficients from Bowles (1970), Ferguson (1991), and Murnane (1981) were obtained from the standardized coefficients divided by their *t* statistics. Smith (1972) reported only standardized coefficients, and also reported that the slope was not significant in the multiple regression. The value of the *t* statistic for Smith (1972) was arbitrarily set to 0.1 for a

conservative choice. Cornett (1969) reported standardized regression coefficients without *t* statistics. This study reported the results from two independent samples (elementary and secondary schools), and reported enough information to allow calculation of the standard errors of the standardized regression coefficients. That is, the full correlation matrix, sample size, and standard deviations of dependent and independent variables were given. The standard errors of the standardized regression coefficients in Cornett's (1969) study were calculated using the variance equation for the case of three predictors, equation (3.1), in the model as shown in the previous section.

Table 4.1. *Example data*

	Author (Year)	b^*	SE(b^*)	*t* statistic
1	Bowles (1970)	0.222	0.031	7.197
2	Bowles (1970)	0.097	0.030	3.193
3	Bowles (1970)	0.210	0.032	6.593
4	Cornett (1969)	-0.090	0.162	-0.555
5	Cornett (1969)	0.030	0.183	0.164
6	Ferguson (1991)	-0.024	0.063	-0.38
7	Ferguson (1991)	0.248	0.041	6.03
8	Ferguson (1991)	0.245	0.040	6.1
9	Ferguson (1991)	0.206	0.035	5.82
10	Ferguson (1991)	0.204	0.036	5.6
11	Ferguson (1991)	0.233	0.034	6.8
12	Murnane (1981)	0.053	0.040	1.341
13	Murnane (1981)	-0.035	0.061	-0.576
14	Murnane (1981)	-0.030	0.064	-0.472
15	Murnane (1981)	-0.353	0.071	-4.953
16	Smith (1972)	0.020	0.200	*NS* (0.1)[a]

Note. [a]The study reported that the slope was nonsignificant. The *t* value was arbitrarily set equal to 0.1.

Four of the five studies in this example presented multiple results. However, as described earlier, those results were from independent samples such as different grade levels or different regions. For further analysis, these sixteen effect sizes are treated as independent, which meets the assumption of meta-analysis.

To synthesize the effects in this example, I first calculate the weighted mean with weights equal to the inverse of the variance of each effect size under the fixed-effects model. Here the effect sizes are the reported standardized slope coefficients, and the variances are the squared standard errors of the standardized coefficients shown in Table 3.1. The weighted mean of the standardized slopes is given by $\bar{b}^* = \frac{\sum b_i^* / v_i}{\sum 1/v_i}$, and equals 0.154. The standard error of the weighed mean is given by $SE(\bar{b}^*) = \sqrt{\frac{1}{\sum \frac{1}{v_i}}} = 0.011$.

Homogeneity Test

The homogeneity test of the standardized slopes is needed. This test can be done by using a large sample chi-square test (Hedges & Olkin, 1985). The *Q* statistic is defined by $Q = \sum \frac{(b_i^* - \bar{b}^*)^2}{v_i}$, where b_i^* is the standardized slope coefficient reported in study *i* and $\bar{b}^*$ is the inverse-variance weighted mean under fixed-effects assumptions (Hedges, 1981; Hedges & Olkin, 1985). Under the null hypothesis, the *Q* statistic is distributed as an asymptotic chi-square with degrees of freedom of $k-1=15$, where *k* is the total number of effect sizes, assuming that

$b_i^* \sim N(\beta_i^*, v_i)$ (Hedges, 1981; Hedges, 1982; Hedges & Olkin, 1985). The null hypothesis of the Q statistic is that the effect sizes arise from the same population, or the effect sizes are homogeneous. The obtained Q statistic for this example is 117.49 with a p value of less than .001. Because the null hypothesis for the Q statistic, $H_0 : \beta_1^* = \beta_2^* = \cdots = \beta_k^* = \beta^*$ or $H_0 : \sigma_{\beta^*}^2 = 0,$ is rejected, it means that these 16 effect sizes are not homogeneous. It also leads to the conclusion that the fixed-effects inverse-sample variance weighted mean of 0.154 is not appropriate. The random effects model can be applied to try to quantify the variation of effect sizes across studies.

Random-effects Model

The weighted mean of the standardized slopes under the random-effects model is 0.10 and the standard error of the random-effects mean is 0.03. The statistical test for this weighted mean under the random-effects model can be obtained as $z = \frac{\bar{b}^*}{SE(\bar{b}^*)} = 3.33$ with a p value of less than .01. The null hypothesis for this test is $H_0 : \beta^* = 0,$ which means the random effects weighted mean of the standardized slopes is not equal to zero in the population. The 95% confidence interval of this weighted mean is from 0.04 to 0.17. This interval does not capture zero, which means the overall weighted mean of the standardized slope is different from zero under the random-effects model. Thus we can say that an average the relation between school outcomes and teachers' verbal ability is nonzero.

Figure 4.1 shows the standardized slopes and 95% confidence intervals under the fixed-effects model from the example studies. The confidence intervals from 16 samples do not seem

to share a common effect. The size of the square box in the confidence interval plot represents the precision of the study; the effect sizes with bigger boxes have larger weight in computing the overall mean. The overall effects under both the fixed-effects and random-effects model are shown at the bottom of the figure and significantly differ from zero because the confidence intervals for both models do not capture zero. These results show that evidence of a weak but nonzero relationship between teacher verbal ability and school outcomes exists when we combine the standardized regression slopes from these 16 samples.

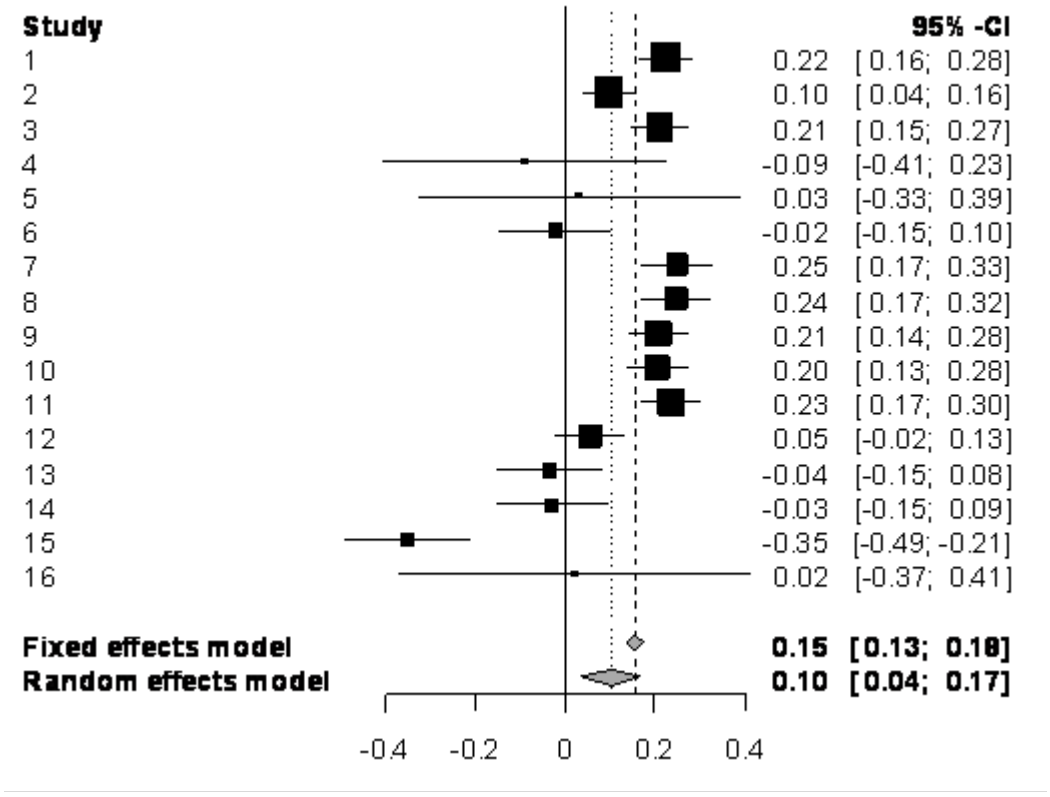

Figure 4.1. Standardized slopes with 95% confidence intervals

CHAPTER FIVE

SIMULATION

In this chapter, a simulation study is presented. The purpose of the simulation is to examine the effect of multicollinearity (intercorrelation among predictors), as well as the number of predictors on the distributions of the estimated standardized regression slopes and their variance estimates. I examine empirical distributions of estimated standardized regression slopes from simulated data for different conditions. Furthermore, I examine the variance of the estimates of the beta-weights when the predictors are collinear. The simulation study was programmed in R 2.12.1 (see the code in Appendix A).

Simulation Conditions

The number of predictors, p, is set to $p = 2$, 5, and 10 in the simulated regression models. The intercorrelation, ρ, among predictors is set to $\rho = .1$, .3, .5, and .8. The correlation between the outcome and each predictor is fixed to .4. The sample sizes are $n = 50$, 100, and 200. The total number of conditions is 36 (3×4×3).

Data Generation

The data is generated from the multivariate standardized normal distribution, $\mathcal{N}(\mathbf{0},\mathbf{\Sigma})$, where $\mathbf{0}$ is a vector with all zeros and $\mathbf{\Sigma}$ is a variance-covariance matrix whose diagonal elements

are 1 and off-diagonal elements are correlations among variables. Each condition is generated for 5,000 independent samples.

Two Predictor Model

The correlation matrix for population in the case of the two predictor model is

$$\boldsymbol{\rho} = \begin{bmatrix} 1 & \rho_{Y1} & \rho_{Y2} \\ \rho_{Y1} & 1 & \rho_{12} \\ \rho_{Y2} & \rho_{12} & 1 \end{bmatrix},$$

where ρ_{Y1} is a population correlation between the outcome and the first predictor, ρ_{Y2} is a population correlation between the outcome and the second predictor, and ρ_{12} is a population intercorrelation between two predictors. For example, the correlation matrices for generating samples for conditions of intercorrelations are

$$\boldsymbol{\rho} = \begin{bmatrix} 1 & .4 & .4 \\ .4 & 1 & .1 \\ .4 & .1 & 1 \end{bmatrix}, \boldsymbol{\rho} = \begin{bmatrix} 1 & .4 & .4 \\ .4 & 1 & .3 \\ .4 & .3 & 1 \end{bmatrix}, \boldsymbol{\rho} = \begin{bmatrix} 1 & .4 & .4 \\ .4 & 1 & .5 \\ .4 & .5 & 1 \end{bmatrix}, \text{ and } \boldsymbol{\rho} = \begin{bmatrix} 1 & .4 & .4 \\ .4 & 1 & .8 \\ .4 & .8 & 1 \end{bmatrix}.$$

An independent random sample from the multivariate standardized normal distribution is generated for each condition. The means for each variable are set to zero and the variances of each variable are set to one. The sample size for each dataset is set to 50, 100, and 200. In each generated sample, standardized regression slopes and their standard errors are estimated.

Five Predictor Model

The correlation matrix for the population in the case of the five predictor model is

$$\boldsymbol{\rho} = \begin{bmatrix} 1 & \rho_{Y1} & \rho_{Y2} & \rho_{Y3} & \rho_{Y4} & \rho_{Y5} \\ \rho_{Y1} & 1 & \rho_{12} & \rho_{13} & \rho_{14} & \rho_{15} \\ \rho_{Y2} & \rho_{12} & 1 & \rho_{23} & \rho_{24} & \rho_{25} \\ \rho_{Y3} & \rho_{13} & \rho_{23} & 1 & \rho_{34} & \rho_{35} \\ \rho_{Y4} & \rho_{14} & \rho_{24} & \rho_{34} & 1 & \rho_{45} \\ \rho_{Y5} & \rho_{15} & \rho_{25} & \rho_{35} & \rho_{45} & 1 \end{bmatrix},$$

where $\rho_{Yj},\ j = 1,2,3,4,5$ is a correlation coefficient between outcome and j^{th} predictor, and $\rho_{jj^*}, j \neq j^*$ are the intercorrelations among the j^{th} and $j^{*\text{th}}$ predictor. In this simulation study, all intercorrelations are set to be equal in each condition. For example, the correlation matrix for the condition with all intercorrelations $\rho_{jj^*} = .3$ is

$$\boldsymbol{\rho} = \begin{bmatrix} 1 & .4 & .4 & .4 & .4 & .4 \\ .4 & 1 & .3 & .3 & .3 & .3 \\ .4 & .3 & 1 & .3 & .3 & .3 \\ .4 & .3 & .3 & 1 & .3 & .3 \\ .4 & .3 & .3 & .3 & 1 & .3 \\ .4 & .3 & .3 & .3 & .3 & 1 \end{bmatrix}.$$

An independent random sample of size *n* from the multivariate standardized normal distribution is generated for each sample-size condition, and standardized regression slopes and their standard errors are estimated.

Ten Predictor Model

A random sample for the ten predictor model is generated with the same procedure described in the two and five predictor cases. Also in each generated sample, the standardized regression slopes and their standard errors are estimated.

Data Evaluation

After generating 5,000 independent random samples from the multivariate normal distribution for each condition, the estimated standardized regression slopes and their standard deviations are averaged. Also, the standard deviations of the empirical set of standardized regression slopes are computed so that it can be compared to the formula of the standard errors of the standardized regression slopes. The histogram plots of empirical the standardized regression slopes and the standard errors is provided. Last, the bias (the difference between the averaged estimated standardized regression slopes and the population value) and mean squared error for the estimated standardized regression slopes is computed.

Bias of the Estimated Standardized Regression Slope

The bias is represented by the difference between the averaged estimated values and the population value. Specifically, the bias of the standardized regression slope (b^*) is obtained by

$$Bias\,(b^*) = E(b^*) - \beta^* = \frac{1}{N}\sum_{i=1}^{N} b_i^* - \beta^*,$$

where $E(b^*)$ is the expected value of estimated standardized regression slope, N is the number of replications of samples, b_i^* is the estimated standardized regression slope for each sample, and β^* is a population value for the standardized regression slope. If the value of bias is 0, then the estimator of the standardized regression slope is unbiased.

Mean Squared Error of the Estimated Standardized Regression Slope

The mean squared error (MSE) is obtained by

$$MSE(b^*) = E\left((b^* - \beta^*)^2\right) = Var(b^*) + \left(Bias(b^*)\right)^2,$$

where $Var(b^*)$ is the variance of the estimated standardized regression slope from replicated samples and $Bias(b^*)$ is the bias defined above. The MSE is decomposed into a sum of the variance of the estimator and the squared bias. The MSE represents the variability of the estimator.

Simulation Results

In each condition, the estimated standardized regression slopes and their standard errors are summarized from 5,000 replicated samples. The population value, β^*, is computed from the population correlation matrix. The bias and MSE are obtained from the generated samples in each condition.

Two Predictor Model

Table 5.1 shows summary statistics for the two predictor model. The first column in Table 5.1 represents the intercorrelations among predictors. The second column is the sample size for each generated sample. The third column is the population value for the standardized

regression slopes. The fourth and fifth column are the empirical mean and the standard deviation of the estimated standardized regression slopes, respectively. The sixth and seventh columns are the empirical mean and the standard deviation of the standard errors of the estimated standardized regression slopes, respectively. The last two columns are the bias and MSE of the estimated standardized regression slopes. The empirical standard deviations of the standardized regression slopes and the average of the standard errors are quite similar from the comparison of sixth and seventh column. The bias of the estimator is close to zero in each condition.

Figure 5.1 shows the empirical distribution of the standardized regression slopes for the two predictor model for all conditions. All histograms in the figure 5.1 appear approximately normal. The column panel represents the sample size condition and the row panel represents the intercorrelations condition. Increases in the spread of histograms correspond with increases in intercorrelations. This indicates that the estimated standardized regression slopes have larger variance when two predictors are highly correlated. We expect the empirical distributions to become narrower as sample sizes become larger. The mean of the estimated standardized regression slopes gets close to zero as the intercorrelations increase.

Figure 5.4 shows the empirical distribution of the standard errors of the standardized regression slopes in case of two predictor model for all conditions. All histograms in the figure 5.4 appear approximately normal. These histograms tend to have a wider spread when intercorrelations are large and sample sizes are small.

Five or Ten Predictor Model

Table 5.2 and Table 5.3 show the summary statistics for the five and ten predictor models. Figure 5.2 and Figure 5.3 show the empirical distributions of the standardized regression slopes for the five and ten predictor models for all conditions. Figure 5.5 and Figure 5.6 show the empirical distribution of the standard errors of the standardized regression slopes for the five and ten predictor model in each condition. The same pattern seen for the case of the two predictor model appeared for five and ten predictor models. The estimated standardized regression slopes have larger variance when predictors are highly correlated for the five and ten predictor models. The empirical distributions become narrower as sample sizes become larger.

The estimated standardized regression slopes are closer to zero as the number of predictors becomes larger, controlling for other conditions. The standard deviations of the estimated standardized regression slopes get large as the number of predictors in the model increase.

Table 5.1. *Summary Statistics for Two Predictor Model*

$\rho(x_i,x_j)$	n	β^*	Mean of b^*	SD of b^*	Mean of SE(b^*)	SD of SE(b^*)	Bias of b^*	MSE of b^*
0.1	50	0.364	0.362	.117	0.123	.009	-0.002	0.014
	100	0.364	0.361	.081	0.086	.005	-0.003	0.007
	200	0.364	0.363	.058	0.060	.002	0.000	0.003
0.3	50	0.308	0.306	.127	0.132	.010	-0.001	0.016
	100	0.308	0.307	.089	0.092	.005	-0.001	0.008
	200	0.308	0.308	.062	0.065	.002	0.000	0.004
0.5	50	0.267	0.262	.146	0.148	.013	-0.005	0.021
	100	0.267	0.264	.100	0.104	.006	-0.002	0.010
	200	0.267	0.267	.070	0.073	.003	0.001	0.005
0.8	50	0.222	0.223	.219	0.220	.027	0.001	0.048
	100	0.222	0.220	.154	0.153	.013	-0.002	0.024
	200	0.222	0.221	.106	0.108	.006	-0.001	0.011

Table 5.2. *Summary Statistics for Five Predictor Model*

$\rho(x_i,x_j)$	n	β^*	Mean of b^*	SD of b^*	Mean of SE(b^*)	SD of SE(b^*)	Bias of b^*	MSE of b^*
0.1	50	0.286	0.287	.101	0.100	.011	0.001	0.010
	100	0.286	0.287	.068	0.068	.005	0.001	0.005
	200	0.286	0.285	.047	0.048	.002	0.000	0.002
0.3	50	0.182	0.181	.135	0.133	.014	-0.001	0.018
	100	0.182	0.183	.091	0.091	.006	0.001	0.008
	200	0.182	0.182	.063	0.063	.003	0.001	0.004
0.5	50	0.133	0.129	.168	0.165	.019	-0.004	0.028
	100	0.133	0.132	.113	0.114	.009	-0.001	0.013
	200	0.133	0.133	.081	0.079	.004	0.000	0.006
0.8	50	0.095	0.101	.276	0.272	.037	0.005	0.076
	100	0.095	0.096	.189	0.187	.017	0.001	0.036
	200	0.095	0.093	.130	0.130	.008	-0.002	0.017

Table 5.3. *Summary Statistics for Ten Predictor Model*

$\rho(x_i,x_j)$	n	β^*	Mean of b^*	SD of b^*	Mean of SE(b^*)	SD of SE(b^*)	Bias of b^*	MSE of b^*
0.1	50	0.211	0.212	.070	0.065	.009	0.002	0.035
	100	0.211	0.211	.046	0.043	.004	0.000	0.038
	200	0.211	0.212	.032	0.030	.002	0.001	0.041
0.3	50	0.108	0.111	.140	0.137	.017	0.003	0.008
	100	0.108	0.111	.090	0.091	.007	0.003	0.009
	200	0.108	0.109	.063	0.063	.003	0.001	0.009
0.5	50	0.073	0.073	.184	0.180	.024	0.001	0.002
	100	0.073	0.071	.119	0.120	.010	-0.001	0.003
	200	0.073	0.070	.082	0.083	.005	-0.003	0.003
0.8	50	0.049	0.058	.305	0.304	.045	0.009	0.001
	100	0.049	0.050	.201	0.202	.019	0.001	0.001
	200	0.049	0.052	.139	0.139	.009	0.003	0.001

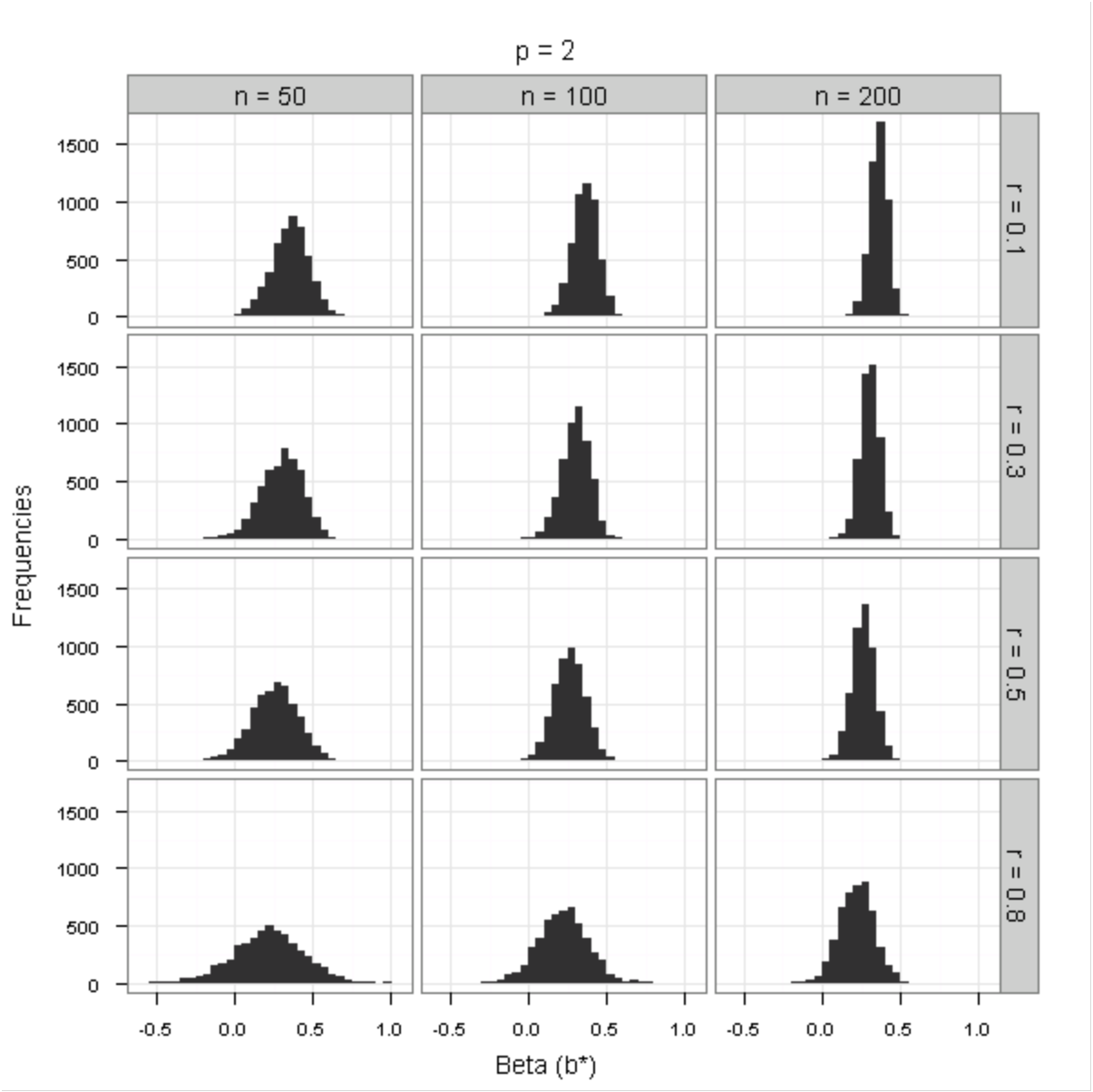

Figure 5.1. Histograms of b* for Two Predictor Model Varying Intercorrelation and Sample Size.

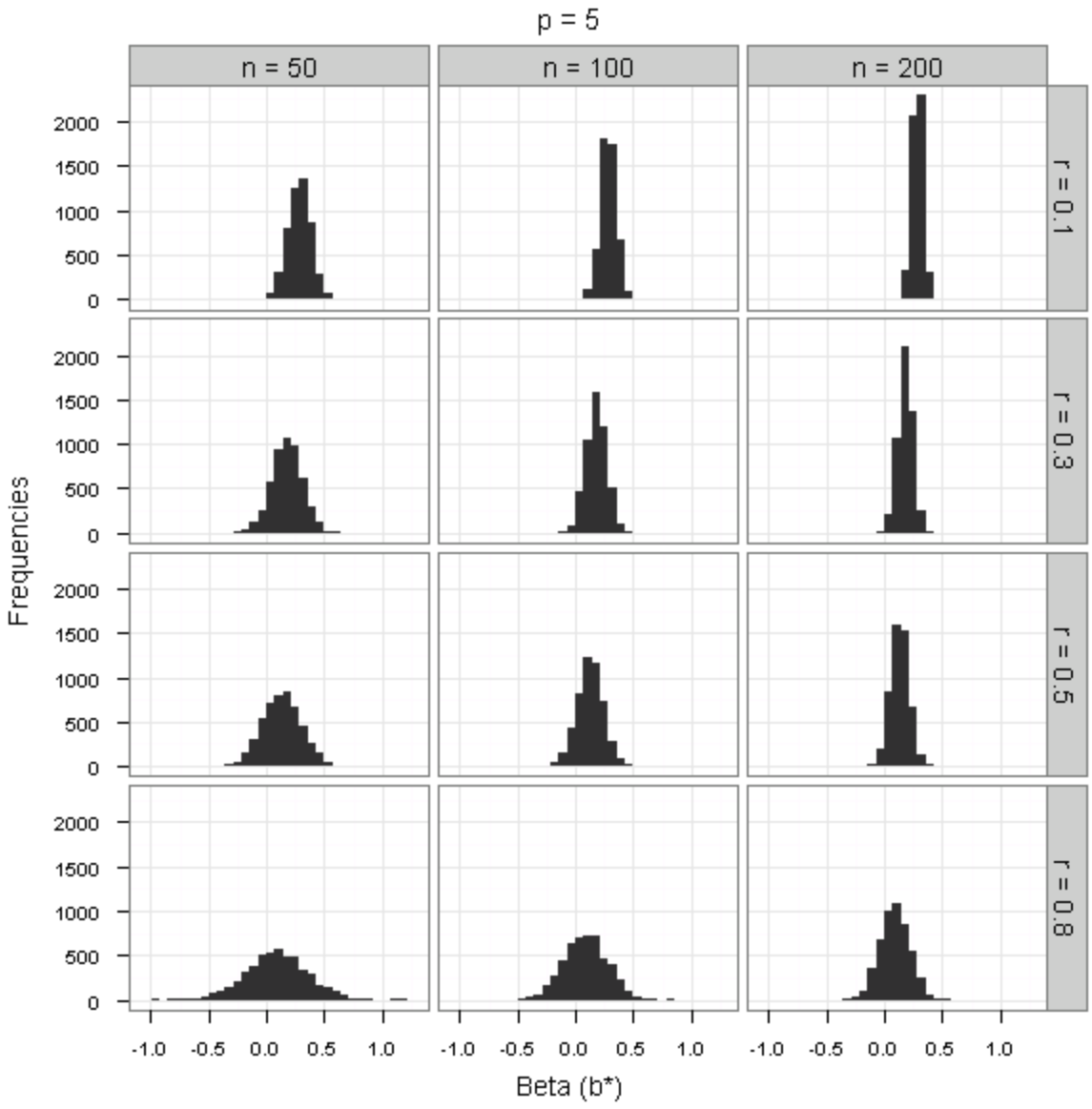

Figure 5.2. Histograms of *b** for Five Predictor Model Varying Intercorrelation and Sample Size.

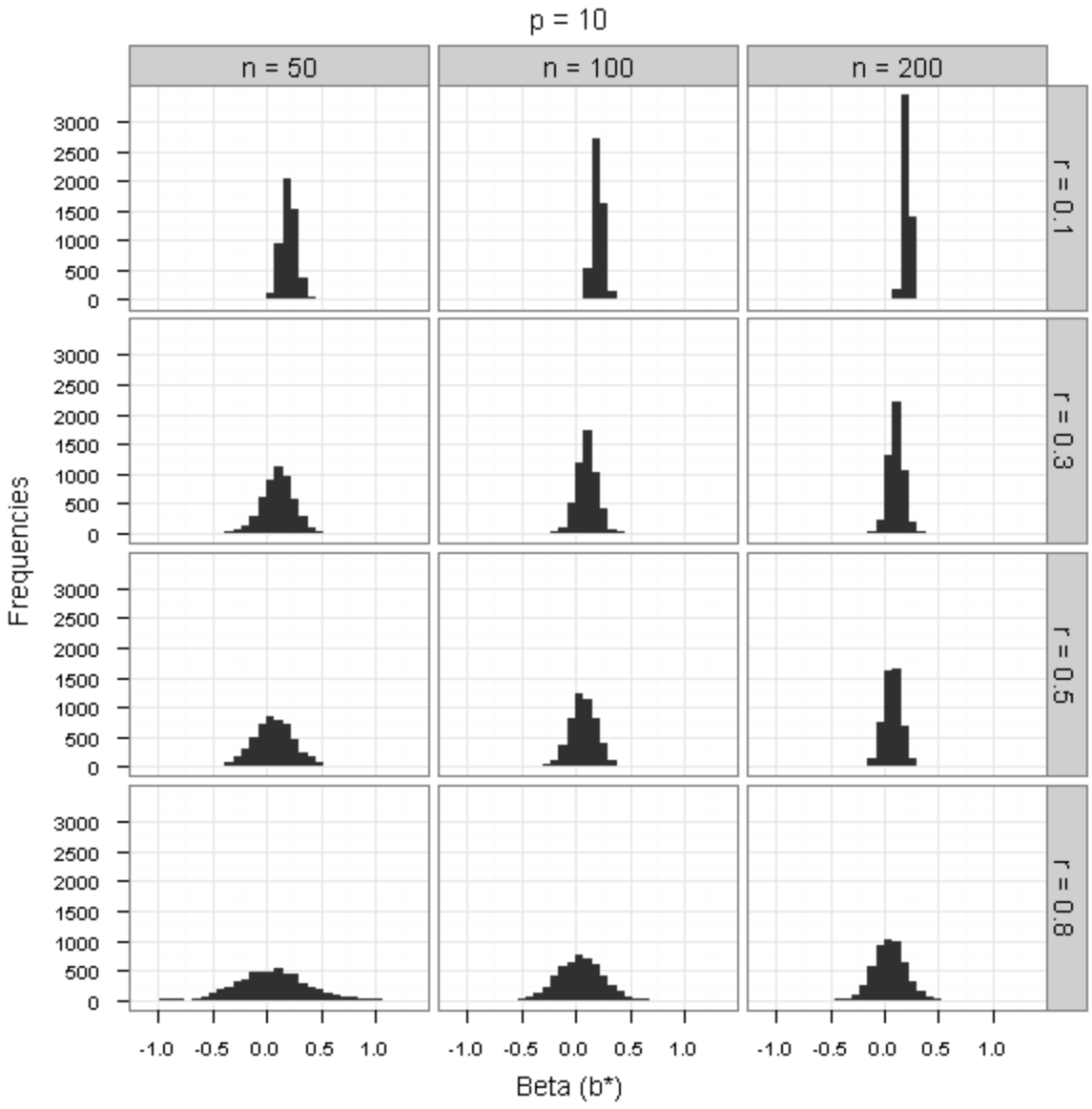

Figure 5.3. Histograms of *b** for Ten Predictor Model Varying Intercorrelation and Sample Size.

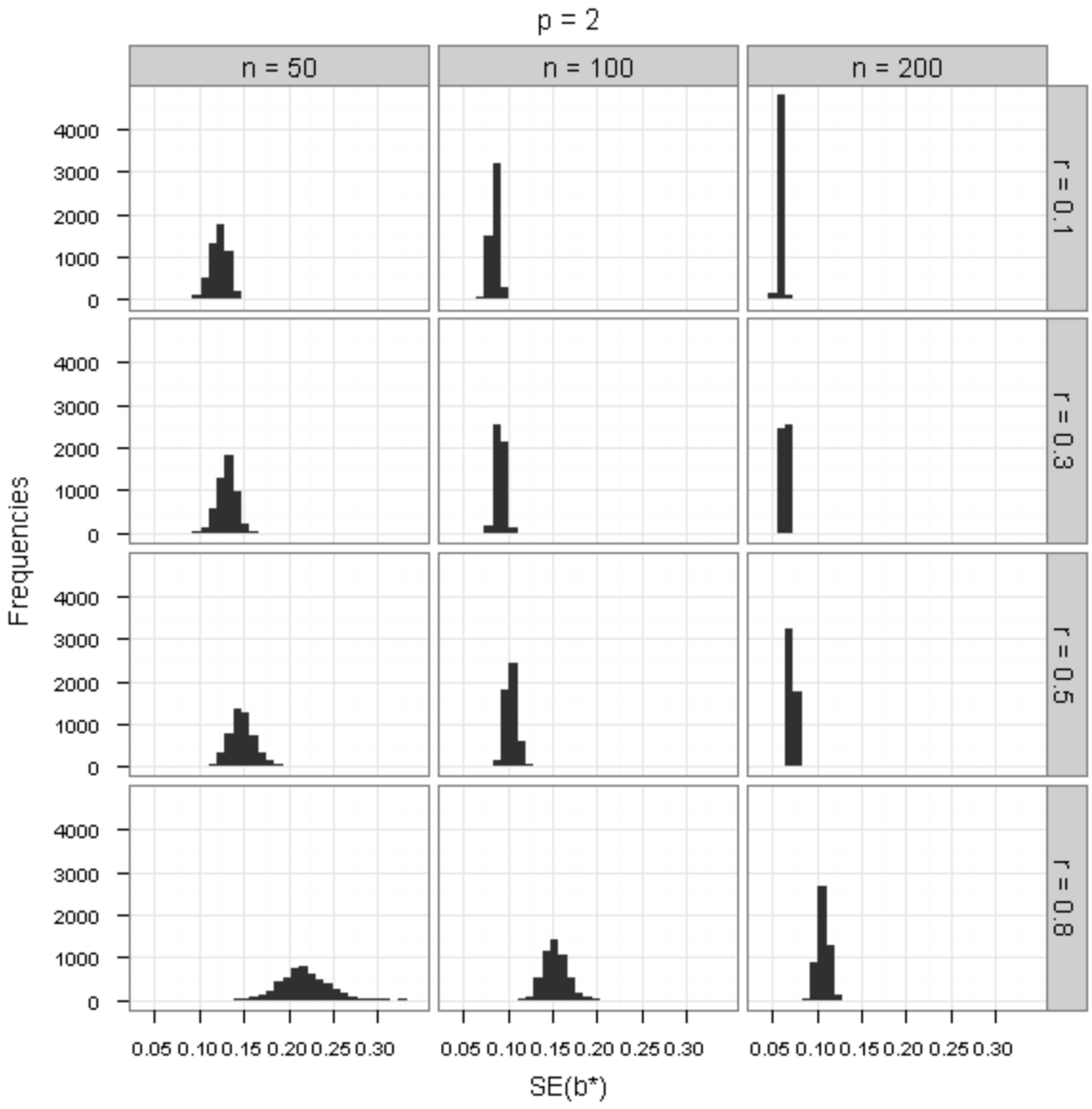

Figure 5.4. Histograms of SE(*b**) for Two Predictor Model Varying Intercorrelation and Sample Size.

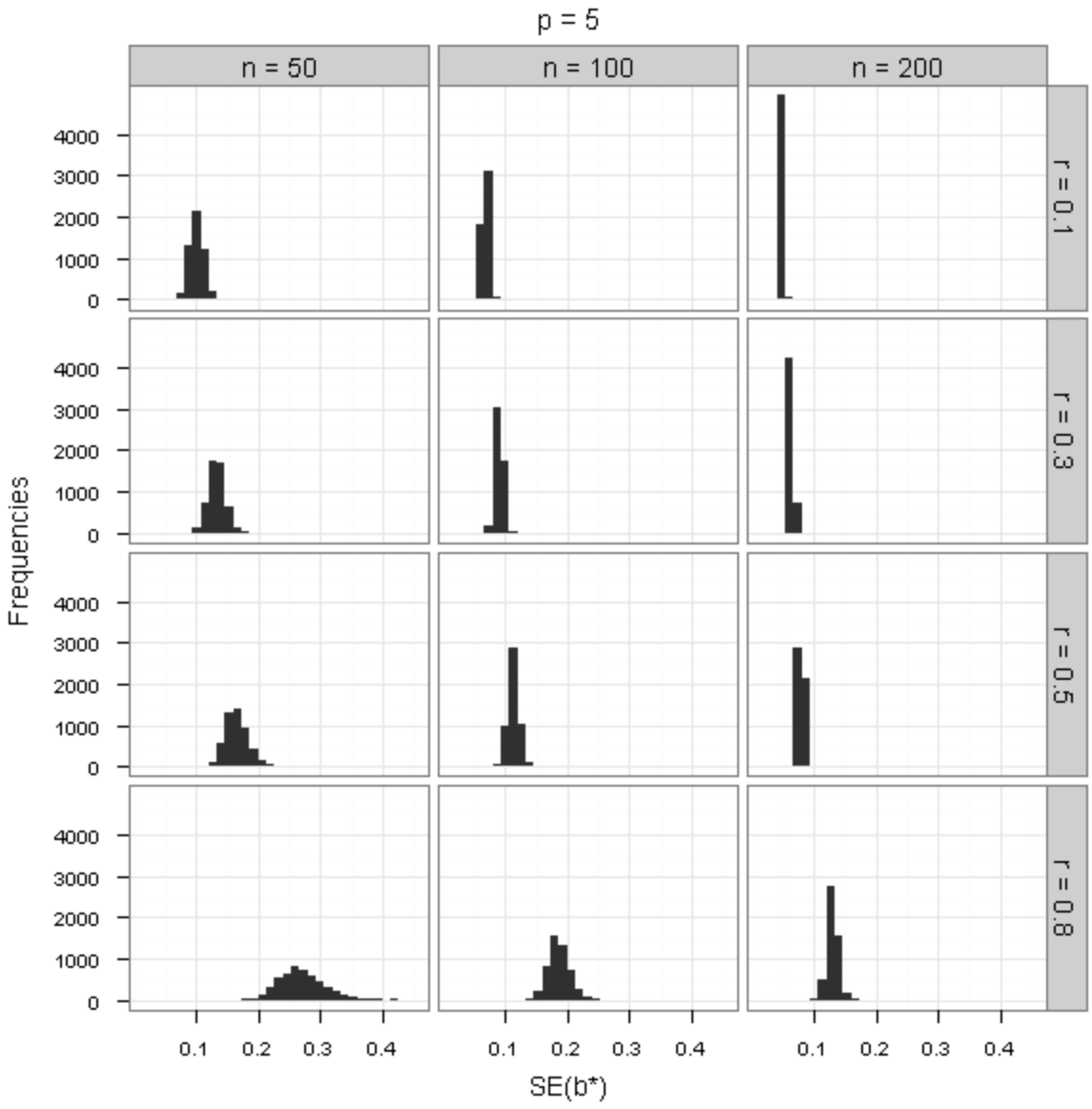

Figure 5.5. Histograms of SE(*b**) for Five Predictor Model Varying Intercorrelation and Sample Size.

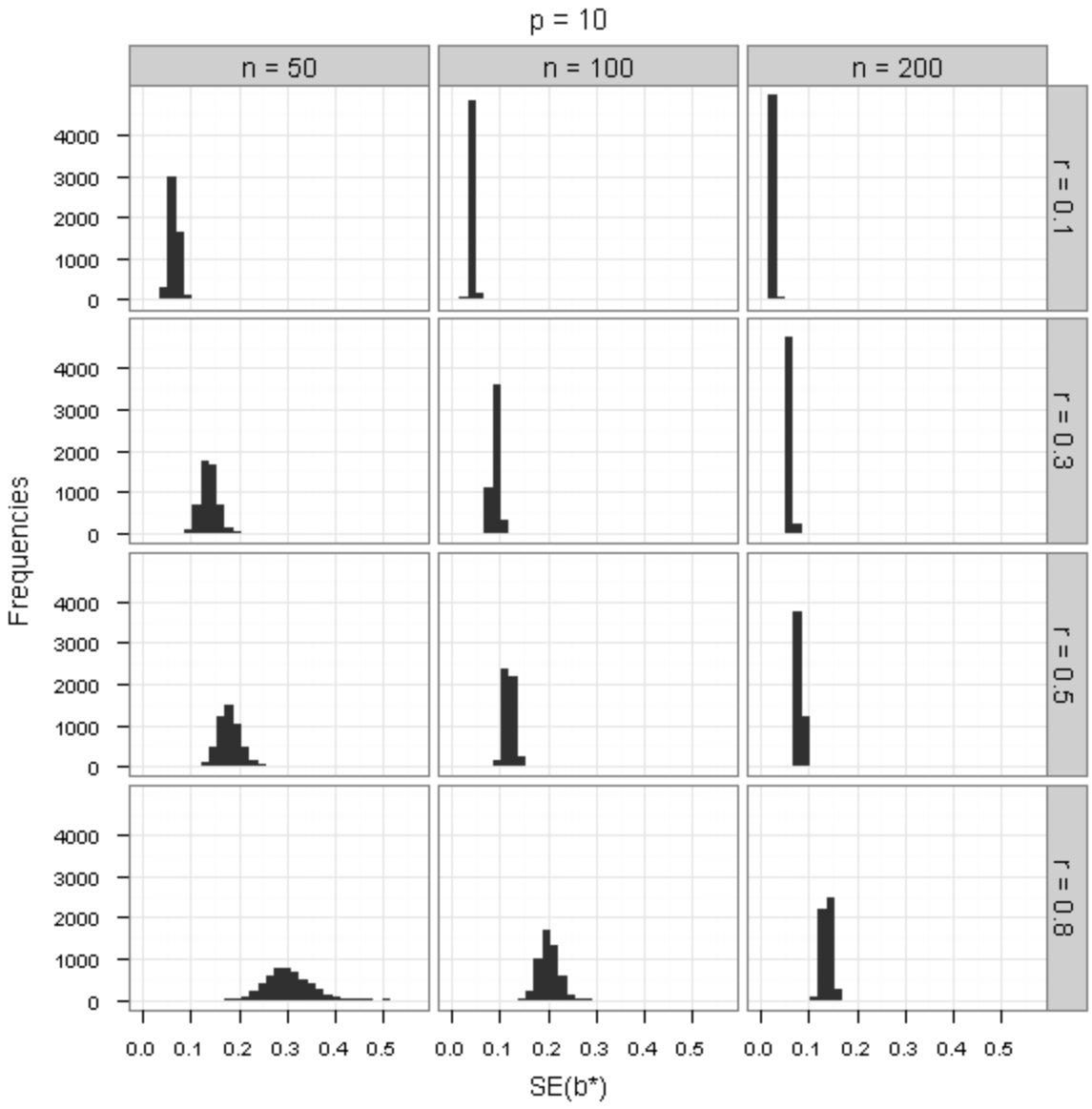

Figure 5.6. Histograms of SE(*b**) for Five Predictor Model Varying Intercorrelation and Sample Size.

CHAPTER SIX

DISCUSSION

In this chapter, I begin with a summary of the ideas proposed in this dissertation. Results from computations and the simulation study were discussed in the previous chapter. Next, practical considerations for using standardized regression slopes as effect-size indices for synthesizing reported regression results from collected studies are discussed. Last, I consider the advantages and limitations of using the standardized regression slope as an effect size for combining effects from regression results.

Conclusion

This dissertation proposes to use standardized regression slopes as effect-size indices for combining regression results in meta-analysis. The standardized regression slope in multiple regression analysis represents the effect of a focal predictor on a target outcome controlling for other predictors in the model. Standardized regression coefficients are scale-free estimates of the effect of a predictor on a single outcome. Thus, these coefficients can be used as effect–size indices for combining studies of the effect of a focal predictor on a target outcome in meta-analysis.

The critical part of conducting meta-analysis is to extract effect-size information from reported results. Standardized regression coefficients can be extracted with ease from reported regression results. The variance of an effect-size estimate is commonly used to represent the

precision of the effect size in meta-analysis. This dissertation provides methods for obtaining the standard errors of standardized regression coefficients from regression results with ease.

In addition, I compare the semi-partial correlation index and the standardized regression coefficients in terms of formulas and computational methods. The standardized regression slope is the product of the square root of the VIF and the semi-partial correlation index. Thus, the semi-partial correlation is more sensitive than the standardized regression slope to the multicollinearity among predictors in the model.

This dissertation provides a re-analysis example to illustrate how to combine standardized regression slopes using data from a published meta-analysis. The example data are a sub-sample from a synthesis of studies of teacher verbal ability and school outcomes (Aloe & Becker, 2009). The original meta-analysis reported combined correlation coefficients and semi-partial correlation indices.

Finally, a simulation is provided to examine the effect of multicollinearity and the number of predictors on the distributions of the estimated standardized regression slopes and their variance estimates. The empirical distribution of the estimated standardized regression slopes tends to have a wider spread when intercorrelations are large and sample sizes are small.

Practical Implications

In this section, I examine the research synthesis literature to assess how likely a researcher is to be able to accumulate standardized regression slopes in meta-analysis research.

The example data used in my example (from Aloe & Becker, 2009) had a total of nineteen studies. Eight out of nineteen studies reported regression analysis results. Five of those

eight studies reported standardized regression slopes and four of those five reported associated *t* statistics.

Qu and Becker (2003) and a subsequent ongoing synthesis of studies about the effect of alternative certification programs summarized standard mean differences between traditionally certified teachers and teachers holding different types of alternative certification. Six out of thirty-nine studies reported regression-analysis results. Five of those six studies reported standardized regression slopes with *t* statistics.

Becker and Aloe (2008) synthesized the effect of teacher science knowledge on student science achievement. From the dataset collected in this synthesis, eleven studies reported correlation coefficients and five studies reported regression coefficients along with standard deviations for reported variables.

From the information given in the research synthesis literatures I have just described, sufficient data were available to support the use of the standardized regression slope for combining study results.

Advantages and Limitations

Finally, the advantages and limitations of using the standardized regression slope as an effect size for combining effects are addressed in this section.

The standardized regression slope is a scale-free index and a well-known measure. It is not difficult to derive standardized slopes and their standard errors from regression studies. The standardized regression slope represents the magnitude of effects. These aspects are the advantages.

It is well-known that the standardized regression coefficient in a simple regression model is the same as the correlation coefficient between two respective variables. When we treat the correlation coefficient as the standardized regression coefficient in a simple regression model, it is possible to combine the correlation coefficient with the standardized regression coefficient.

When conducting a meta-analysis, each study may have different control variables. The regression analysis results may vary in terms of the selection of control variables used. This limitation can be examined by moderator analysis to explore how the control variables influence the effects. Another issue concerns multicollinearity, which makes standardized slopes deviate from the correlation. However, the standard errors of the standardized regression slopes are functions of the measures of multicollinearity, which means the standard errors are larger when the predictors are highly correlated. Thus, the standardized regression slopes with multicollinearity contribute less weight when the effects are synthesized.

In this dissertation, the values of correlations among the variables were all positive for the computations and the simulation study. It would be useful to add more conditions for different correlation values.

Lastly, primary studies may lack the information that is necessary to obtain the standardized slope and its standard error. Further research is needed to develop computational approaches for these types of situations.

APPENDIX A

SIMULATION CODE

```
library(MASS)
library(plyr)

## use expand.grid to create df of all conditions
p <- c(2, 5, 10)
n <- c(50, 100, 200)
r <- c(0.1, 0.3, 0.5, 0.8)
conditions0 <- expand.grid(p, n, r)
colnames(conditions0) <- c("p", "n", "r")

## conditions for loop
conditions.loop <- split(conditions0, 1:dim(conditions0)[1])

## conditions for list; i.e. create a list with 36 x 5000 elements
conditions.list <- conditions0[rep(1:36, 5000),]
dim(conditions.list)

## Creates a list of {p, n, mu1, sig}
## lapply is applied on EACH element of the list, here x
## and x itself consists of three elements (p, n, r)

cond2list <- function(x){
  p <- x$p
  n <- x$n
        r <- x$r

        ## create matrix
        sig <- matrix(rep(r, (p+1)^2), ncol = p+1)
        ryx <- 0.4
        sig[,1] <- ryx   # Set corr(Y,x)
        sig[1,] <- ryx   # Set corr(Y,x)
        diag(sig) <- 1

        ## mu
        mu <- rep(0, p+1)

        out <- list(sig = sig, mu1 = mu, p = p, n = n)
        return(out)
```

```
}

sim.beta <- function(x)
{
p <- x$p
n <- x$n
mu1 <- x$mu1
sig <- x$sig

# Data generation from multivariate normal dist.
d1 <- mvrnorm(n=n, mu=mu1, Sigma=sig)

sd.d <- sd(d1) # Standard deviation for each variable
mean.d <- colMeans(d1) # Mean for each variable

# Regression analysis
x.d1 <- d1[,2:(p+1)]
reg1 <- summary(lm(d1[,1]~x.d1))

# regression results estimates
b <- reg1$coefficients

# beta weights(b*1: beta[2])
beta <- b[2,1]*sd.d[2]/sd.d[1]
se.beta <- b[2,2]*sd.d[2]/sd.d[1]

#out1 <- cbind(beta, se.beta)
out1 <- data.frame(beta, se.beta)

return(out1)
}


## foreach ###########################################################

listOfConditions <- llply(conditions.loop, cond2list)
length(listOfConditions)

library(doSMP)
w <- startWorkers(workerCount = 2) ## number of cores
registerDoSMP(w)

#system.time({
nrep <- 5000
out.beta <- data.frame(matrix(ncol = 0, nrow = 0))
out <- foreach(i = 1:nrep, .combine = rbind, .packages = c("plyr", "MASS"))%dopar%{
```

```
            ldply(listOfConditions, sim.beta)
            }
        #})

        conditions.list <- conditions0[rep(1:36, 5000),]

        ## identifying the conditions
        results.beta <-data.frame(out, conditions.list)


        ## !!!!!!!!!!!! RUN THIS BEFORE CLOSING R !!!!!!!!!!!!!!!!!!!!!!!!!!!!
        stopWorkers(w)
        rmSessions(all=TRUE)

        ## Descriptive statistics

        summary.beta <- matrix(unlist(describe.by(results.beta$beta,
list(results.beta$p,results.beta$n,results.beta$r), na.rm = TRUE)),ncol=13,byrow=T)
        summary.se <- matrix(unlist(describe.by(results.beta$se.beta,
list(results.beta$p,results.beta$n,results.beta$r), na.rm = TRUE)),ncol=13,byrow=T)

        colnames(summary.beta) <-
c("varname","n","mean","sd","median","trimmed","mad","min","max","range","skew","kurtosis
","se")
        colnames(summary.se) <- colnames(summary.beta)

        write.table(summary.beta,"summary_beta.xls",sep='\t', row.names = F )
        write.table(summary.se,"summary_se.xls",sep='\t', row.names = F )

        library(ggplot2)

        ### Histogram

        ggplot(aes(x = beta), data = subset(results.beta, p == 2)) + geom_histogram()
+facet_grid(r.label ~ n.label) +
        opts(title ="p = 2")
```

REFERENCES

References marked with an asterisk indicate studies included in the meta-analysis.

Aloe, A. M., & Becker, B. J. (2009). Teacher verbal ability and school outcomes: Where is the evidence? *Educational Researcher*, *38*(8), 612–624.

Aloe, A. M. (2009). *A partial effect size for the synthesis of multiple regression models*. Unpublished doctoral dissertation, Florida State University.

Anderson, H. Roush, S., & McClary, J. (1973). Relationships among ratings, production, efficiency, and the General Aptitude Test Battery scales in an industrial setting. *Journal of Applied Psychology*, *58*, 77-82.

Armitage, P., Berry, G., & Matthews, J. N. S. (2002). *Statistical methods in medical research* (4th ed.). Malden: Wiley-Blackwell.

Becker, B. J. (1992). Using results from replicated studies to estimate linear models. *Journal of Educational Statistics*, *17*, 341-362.

Becker, B. J. (1995). Corrections to "Using results from replicated studies to estimate linear models." *Journal of Educational and Behavioral Statistics*, *20*, 100-102.

Becker, B. J., & Wu, M. J. (2007). The synthesis of regression slopes in meta-analysis, *Statistical Science*, *22*(3), 414-429.

Becker, B. J., & Aloe, A. M. (2008, March). Teacher science knowledge and student science achievement. *Paper presented at the annual meeting of the American Educational Research Association*, New York.

Borenstein, M., Hedges, L. V., Higgins, J., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester: Wiley.

*Bowles, S. S. (1970). Towards an educational production function. In W. L. Hansen (Ed.), *Education, income, and human capital* (pp. 11–70). New York: National Bureau of Economic Research.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*. New York: Wiley.

Cohn, L. D., & Becker, B. J. (2003) How meta-analysis increases statistical power. *Psychological Methods*, *8*(3), 243–253.

*Cornett, J. D. (1969). Effectiveness of three selective admissions criteria in predicting performance of first-year teachers. *Journal of Educational Research*, *62*, 247–250.

*Ferguson, R. F. (1991). Paying for public education: New evidence on how and why money matters. *Harvard Journal on Legislation*, *28*(2), 465–498.

Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, *5*, 3–8.

Greenland, S., Maclure, M., Schlesselman, J, J., Poole, C., & Morgenstern, H. (1991). Standardized regression coefficients: A further critique and review of some alternatives. *Epidemiology*, *2*(5), 387-392.

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

Greenwald, R., Hedges, L. V., & Laine, R. D. (1996). The effect of school resources on student achievement. *Review of Educational Research*, *66*(3), 361-396.

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, *6*, 107-128.

Hedges, L. V. (1982). Fitting categorical models to effect sizes from a series of experiments. *Journal of Educational Statistics*, *7*, 119-137.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

Howell, D. C. (2010). *Statistical methods for psychology* (7th ed.). Pacific Grove, CA: Duxbury.

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings.* New York: Russell Sage Foundation.

Hunter, J. E., & Schmidt, F. L. (1994). Correcting for sources of artificial variation across studies. In H. M. Cooper & L. V. Hedges (Eds.). *The handbook of research synthesis* (pp. 323 - 336). New York: Russell Sage Foundation.

Keef, S. P., & Roberts, L. A. (2004). The meta-analysis of partial effect sizes. *British Journal of Mathematical and Statistical Psychology, 57*(1), 97-129.

Kieffer, K. M., Reese, R. J., & Thompson, B. (2001). Statistical techniques employed in AERJ and JCP articles from 1988 to 1997: A methodological review. *The Journal of Experimental Education*, *69*(3), 280-309.

*Murnane, R. J., & Phillips, B. (1981). What do effective teachers of inner city children have in common? *Social Science Research, 10*, 83-100.

Paul, P. A., Lipps, P. E., & Madden, L. V. (2006). Meta-analysis of regression coefficients for the relationship between Fusarium head blight and deoxynivalenol content of wheat. *Phytopathology*, *96*(9), 951-961.

Peterson, R. A., & Brown, S. P. (2005). On the use of beta coefficients in meta-analysis, *Journal of Applied Psychology*, *90*(1), 175-181.

Qu, Y., & Becker, B. J. (2003, April). Does traditional teacher certification imply quality? A meta-analysis. Paper presented at the meeting of the American Educational Research Association, Chicago, IL.

Shadish, W. R., & Haddock, C. K. (1994). Combining estimates of effect size. In H. M. Cooper & L. V. Hedges (Eds.) *The handbook of research synthesis* (pp. 261 - 282). New York: Russell Sage Foundation.

*Smith, M. S. (1972). Equality of educational opportunity: The basic finding reconsidered. In F. Mosteller, & D. P. Moynihan (Eds.). *On equality of educational opportunity* (pp. 230–342). New York: Random House.

Stanley, T. D., & Jarrell, S.B. (1989). Meta-regression analysis: A quantitative method of literature surveys. *Journal of Economic Surveys, 3*(2), 161-170.

Stanley, T. D., & Jarrell, S.B. (2005). Meta-regression analysis: A quantitative method of *literature* surveys. *Journal of Economic Surveys, 19*(3), 299-308 [Reprinted version of the 1989 paper].

Stankowich, T. & Blumstein, D. T. (2005). Fear in animals: A meta-analysis and review of risk assessment. *Proceedings: Biological Science*, *272*, 2627-2643.

Stanley, T. D., Doucouliagos, C, & Jarrell, S. B. (2008). Meta-regression analysis as the socio-economics of economics research. *Journal of Socio-Economics*, *37*, 276-292.

Stavig, G. R. (1977). The semistandardized regression coefficient, *Multivariate Behavioral Research*, *12*, 255-258.

BIOGRAPHICAL SKETCH

RAE-SEON (SUNNY) KIM

EDUCATION

M.S. Computer Sciences and Statistics, Chonnam National University, Korea, 1998.
Thesis: Fitting Models of Precipitation Data by Markov Chain Dependent Model.
Advisor: Dr. Jeong-Soo Park.

B.S. Statistics, Chonnam National University, Korea, 1995.

EXPERIENCE

Sep.2006 – Present	*Research Assistant* for College of Education, Florida State University.
Jan.2010 – Present	*Statistical and Research Design Consultant*, Florida State University.
Sep.2005–Aug.2009	*Teaching Assistant* for College of Education, Florida State University.
Feb.2000–May.2001	*Administrative Assistant* for Department of Statistics, Chonnam National University, Korea
Aug.1999–Aug.2000	*Lecturer* for Information Computing Institute, Chonnam National University, Korea.
Aug.1999–Dec.1999	*Lecturer* for Dongshin University, Korea.
Jan.1998–Dec.1998	*Research Scientist* for Meteorological Research Institute, Seoul, Korea.
Mar.1996–May.2001	*Statistical Consultant* for Department of Statistics, Chonnam National University, Korea
Mar.1996–Feb.1998	*Research Assistant* for Department of Statistics, Chonnam National University, Korea.