

December 2014

Inferential Procedures for Dominance Analysis Measures in Multiple Regression

Shuwen Tang

University of Wisconsin-Milwaukee

Follow this and additional works at: <https://dc.uwm.edu/etd>



Part of the [Educational Assessment, Evaluation, and Research Commons](#), [Quantitative Psychology Commons](#), and the [Statistics and Probability Commons](#)

Recommended Citation

Tang, Shuwen, "Inferential Procedures for Dominance Analysis Measures in Multiple Regression" (2014). *Theses and Dissertations*. 645.

<https://dc.uwm.edu/etd/645>

This Dissertation is brought to you for free and open access by UWM Digital Commons. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of UWM Digital Commons. For more information, please contact open-access@uwm.edu.

INFERENTIAL PROCEDURES FOR DOMINANCE ANALYSIS MEASURES
IN MULTIPLE REGRESSION

by

Shuwen Tang

A Dissertation Submitted in

Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

in Educational Psychology

at

The University of Wisconsin-Milwaukee

December 2014

ABSTRACT

INFERENTIAL PROCEDURES FOR DOMINANCE ANALYSIS MEASURES IN MULTIPLE REGRESSION

by

Shuwen Tang

The University of Wisconsin-Milwaukee, 2014
Under the Supervision of Professor Razia Azen

In order to better interpret a selected multiple regression model, researchers are often interested in whether a predictor is significantly more important than another or not. This study investigates the performance of the Normal-Theory based (asymptotic) confidence interval and bootstrap confidence intervals for predictors' dominance relationships using both normal and non-normal data. The results show that asymptotic confidence interval method is adequate to make inferences for comparing two general dominance measures when the distribution is multivariate normal or slightly non-normal and when the effect size is no less than 0.15 and the sample size is at least 100. However, the bootstrap confidence interval methods are preferred over the asymptotic confidence interval when the data are considerably non-normal (e.g., skew > 0.75 , or $|\text{kurtosis}| > 1.2$). The choice among standardized, percentile and bias-corrected bootstrap confidence intervals is based on the properties of the real data set, like sample size and distribution. An empirical demonstration and appropriate interpretation are also provided.

Table of Contents

List of Tables.....	v
List of Figures	vii
Chapter 1. Introduction	1
Chapter 2. Literature Review	6
2.1 Relative importance measures	7
2.1.1 The most common importance measures: r and β	8
2.1.2 Importance measures based on combining r and β	12
2.1.3 Importance measures based on R^2	17
2.2 Summary	19
2.3 Dominance Analysis	20
2.3.1 General procedure	20
2.3.2 Complete, conditional, and general dominance	22
2.4 Statistical inference about importance measures	26
2.4.1 Asymptotic confidence interval.....	27
2.4.2 Bootstrap confidence interval.....	29
Chapter 3. Methodology.....	36
3.1 Simulation	36
3.1.1 General procedures.....	36
3.1.2 Specific conditions	42
3.2 Empirical Demonstration	48
Chapter 4. Results	51
4.1 Simulations	52
4.1.1 Normal Distribution Conditions	52
4.1.2 Non-normal: Skewed distribution conditions.....	67
4.1.3 Non-normal: Leptokurtic and platykurtic distribution conditions.....	76
4.2 Empirical Example: Payment Prediction	85
Chapter 5. Conclusions and Discussion	91
5.1 Summary	91
5.2 Limitations and Future Directions	98
5.3 Conclusion	101

References	103
Curriculum Vitae.....	111

List of Tables

Table 2.1 Numerical example of Dominance Analysis in the population using a four-predictor model	24
Table 3.1 Population correlation matrices used in simulation study	43
Table 3.2 Differences between two general dominance measures in population	45
Table 4.1 Type I error rate with different sample sizes	52
Table 4.2 Power for different sample sizes	56
Table 4.3 Averaged accuracy with different sample sizes	60
Table 4.4 Averaged confidence interval widths with different sample sizes	63
Table 4.5 Type I error rate when sample size = 300	68
Table 4.6 Type I error rate when sample size = 100	69
Table 4.7 Power when sample size = 300	70
Table 4.8 Power when sample size = 100	71
Table 4.9 Averaged accuracy when sample size = 300	72
Table 4.10 Averaged accuracy when sample size = 100	73
Table 4.11 Averaged confidence interval widths when sample size = 300	75
Table 4.12 Averaged Confidence interval widths when sample size = 100	75
Table 4.13 Type I error rate when sample size = 100	77
Table 4.14 Type I error rate when sample size = 100	77
Table 4.15 Power when sample size = 300	79
Table 4.16 Power when sample size = 100	80
Table 4.17 Averaged accuracy when sample size = 300	81
Table 4.18 Averaged accuracy when sample size = 100	82
Table 4.19 Averaged Confidence interval widths when sample size = 300	84
Table 4.20 Averaged Confidence interval widths when sample size = 300	84

Table 4.21 Descriptive statistics of the variables in empirical example	86
Table 4.22 The correlation matrix for the empirical cases study	87
Table 4.23 Several important measures for the empirical cases study	87
Table 4.24 Dominance Analysis for the empirical case study	88
Table 4.25 Statistical significance test of general dominance measures	90

List of Figures

Figure 3.1 Histograms of six non-normal distributions	48
Figure 4.1 Type I error rate with different sample size.....	53
Figure 4.2 Type I error rate and effect size when sample size = 300	54
Figure 4.3 Type I error rate and effect size when sample size = 100	54
Figure 4.4 Type I error rate and effect size when sample size = 30	55
Figure 4.5 Power for different sample sizes	56
Figure 4.6 Power and effect size when sample size = 300	57
Figure 4.7 Power and effect size when sample size = 100	58
Figure 4.8 Power and effect size when sample size = 30	59
Figure 4.9 Averaged accuracy with different sample sizes.....	60
Figure 4.10 Accuracy and effect size when sample size = 300	61
Figure 4.11 Accuracy and effect size when sample size = 100	61
Figure 4.12 Accuracy and effect size when sample size = 30	62
Figure 4.13 Averaged confidence interval widths with different sample sizes	63
Figure 4.14 Averaged confidence interval widths when sample size = 300.....	64
Figure 4.15 Averaged confidence interval widths when sample size = 100.....	64
Figure 4.16 Averaged confidence interval widths when sample size = 30.....	65
Figure 4.17 Bias of bootstrap estimates to population parameters with different sample sizes.....	66
Figure 4.18 Bias of bootstrap estimates to sample parameters with different sample sizes	66
Figure 4.19 Standardized bias of sample to population parameters with different sample sizes.....	67
Figure 4.20 Type I error rate when sample size = 300	68
Figure 4.21 Type I error rate when sample size = 100	69

Figure 4.22 Power when sample size = 300.....	70
Figure 4.23 Power when sample size = 100.....	71
Figure 4.24 Averaged accuracy when sample size = 300	73
Figure 4.25 Averaged accuracy when sample size = 100	74
Figure 4.26 Averaged confidence interval width when sample size = 300	75
Figure 4.27 Averaged confidence interval width when sample size = 100	76
Figure 4.28 Type I error rate when sample size = 300	77
Figure 4.29 Type I error rate when sample size = 100	78
Figure 4.30 Power for when sample size = 300	79
Figure 4.31 Power for when sample size = 100	80
Figure 4.32 Averaged accuracy when sample size = 300	82
Figure 4.33 Averaged accuracy when sample size = 100	83
Figure 4.34 Averaged confidence interval width when sample size = 300	84
Figure 4.35 Averaged confidence interval width when sample size = 100	85
Figure 4.36 Prediction Model	86
Figure 5.1 Asymptotic confidence intervals of D_{43} in Matrix 1 when $N = 300$	93
Figure 5.2 Asymptotic confidence intervals of D_{43} in Matrix 1 when $N = 100$	93
Figure 5.3 Asymptotic confidence intervals of D_{43} in Matrix 1 when $N = 30$	94
Figure 5.4 The bootstrap distribution of D_{43} and D_{42} in Matrix 3 when skew = 1 and $N = 100$	96

Chapter 1. Introduction

Multiple regression (MR) is an extremely useful statistical tool for the analysis of data in a variety of disciplines. By adopting MR analysis, researchers can capture the nature and estimate the magnitude of the relationship between several explanatory variables and a response variable. When fitting the observed data into a MR model, the obtained regression coefficients allow researchers to predict the criterion value from the value of predictors and test a particular theory regarding the contribution of predictors to determining the criterion values. In this document, the term “predictors” is used to represent the explanatory/independent variables, “criterion” refers to the response/dependent variable, and “contribution” represents the predictive power or the amount of total variability in the criterion explained by the predictor(s) in a given model.

In order to better interpret a selected MR model, researchers are often interested in comparing the predictors in terms of their contributions to the overall predictive effect and figuring out the relatively most “important” and “meaningful” predictors. If one predictor contributes more than another, we say this predictor is more important than the other in a given model. For example, people may be trying to determine the relative importance of supervisor support and co-worker support as predictors of job satisfaction (Tang, Siu, & Cheung, 2014); students' perceptions of teaching attitude and course content in determining course rating (Ting, 2000); immigrants' in-group orientation and natives' segregation orientation when predicting perceived discrimination (Brenick, Titzmann, Michel, & Silbereisen, 2012); or intuitive and systematic cognitive style when

predicting creativity (Kim, Choi, & Park, 2012). In each of these cases all of the predictors may be of theoretical importance, but one may still wish to rank order them and find out which predictor could be put in first place relative to the others. For this purpose, an intuitive and feasible measure of the relative importance of predictors in a MR model is needed.

When all predictors are uncorrelated, the question of their relative importance can be answered by some traditional measures, such as the zero-order correlations or standardized regression coefficients, which are automatically reported by most statistical packages and will result in the same rank ordering of importance for uncorrelated predictors. However, in most empirical studies, where predictors are at least somewhat related to each other, the situation becomes more complicated because the predictor's contribution is greatly contingent on different contexts. The zero-order correlation measures contribution in a simple regression, without considering any other predictors, while the standardized regression coefficient considers contribution in the full model, including and controlling for all other predictors. When predictors are correlated with each other, it is possible to have one predictor that is by itself highly associated with the criterion but has a relatively small standardized regression coefficient. For example, suppose that there are two predictors highly correlated to the criterion and highly correlated with each other. A small standardized regression coefficient could be obtained for either predictor by controlling for a large amount of shared contribution with the other predictor. Hence, the aforementioned simple indices may produce incomplete and possibly inaccurate results and conclusions when predictors are inter-correlated, since

they fail to fully consider the complexity of what is measured and meant by importance (Budescu, 1993; Darlington, 1968; Kruskal & Majors, 1989; Ward, 1962).

In the context of multiple correlated predictors, researchers need to clearly define what is meant by relative importance and rely on more sophisticated techniques that better reflect this definition. Several approaches have been proposed to gain insight into the assessment of relative importance. Each of them provides different perspectives and sometimes yields different rank orderings of the predictors regarding their importance. Among the more modern measures for relative importance, Dominance Analysis offers an intuitive and informative assessment of predictor contributions (Azen & Budescu, 2003; Azen & Budescu, 2006; Budescu, 1993; Budescu & Azen, 2004). Dominance Analysis is based on an examination of each predictor's contribution by itself and in the presence of all possible combinations of the remaining predictors. The predictors can then be compared to each other based on their contributions, resulting in the establishment of dominance relationships with respect to relative importance. In recent years, there has been growing interest in the use of Dominance Analysis for predictor comparisons in terms of their relative importance, in fields such as applied psychology (Colquitt, LePine, Piccolo, Zapata, & Rich, 2012; Miller, Konopaske, & Byrne, 2011; Morrow, McGonagle, Dove-Steinkamp, Walker, Marmet, & Barnes-Farrell, 2010) and education (Mellard, Anthony, & Woods, 2012).

Moreover, an interesting question researchers may wish to address is whether a predictor is significantly more important than another; in other words, whether its dominance over another predictor in terms of relative importance exceeds chance levels. For example, when one obtain a difference of 0.3 between the sample dominance measures of two predictors, one may wish to determine whether it is a "real" difference (greater than 0) in the population. Therefore, it is important to propose a reliable and feasible procedure for statistical significance testing for Dominance Analysis. To the best of my knowledge, no research exists that fully answered this research question yet. The possible difficulty of addressing this problem may lie in that we do not know the exact sampling distribution of the dominance measures. A technique named "bootstrapping" is frequently applied to make inferences about statistics when the sampling distribution is unknown (Efron, 1979, 1982; Efron & Tibshirani, 1986, 1993), and may be feasible in this case. Using this approach, researchers can estimate the distribution of the dominance measure differences, construct confidence intervals, and then make statistical inferences.

Previous studies (e.g., Hedges & Olkin, 1983; Azen & Sass, 2008) have shown that it is often adequate for asymptotically based methods to make inference about predictors' additional contributions when the data are normally distributed. This study aims to adopt the bootstrapping procedure to determine the significance of a difference between the dominance measures from two predictors: firstly, compare the performance of the asymptotic confidence interval and bootstrap confidence interval in a simulation study using normal data; secondly, extend the simulations to non-normal data; lastly, make

recommendations for its use based on the results, and provide empirical researchers with a demonstration of this analysis and its appropriate interpretation.

Chapter 2. Literature Review

The standardized regression model, with n observations on p predictors, can be represented as follows:

$$Y_j = \sum_{i=1}^p \beta_i X_{ij} + e_j \quad (1)$$

where $i = 1, \dots, p$ (the number of predictors in the model), $j = 1, \dots, n$ (the number of observations in the data set), X_i represents the i^{th} predictor, Y_j represents the value on the criterion for the j^{th} observation, β_i is the standardized regression coefficient associated with X_i , and e_j represents the error term, usually called the “residual”. In this representation all the observed Xs and Ys are standardized to have a mean of zero and a standard deviation of 1, so the model intercept would be 0 and is thus not included in this equation.

The overall regression effect is commonly estimated by the squared multiple correlation coefficient, R^2 , which measures the proportion of total variance in the criterion that is accounted for by the predictors in the regression model. By definition, it can be calculated from the following formula:

$$R^2 = \frac{SST - SSE}{SST} = 1 - \frac{SSE}{SST} \quad (2)$$

where SST is the total sum of squares and represents the overall variability of the criterion, Y , whereas SSE is the error sum of squares and represents the variability of the random error components, e_j . Removing the error term from the total variability, the remaining variability is the explained proportion by the regression model.

2.1 Relative importance measures

Researchers are often interested in understanding and determining the importance of each predictor in their regression model. This predictor importance analysis allows for testing predictions, refining theory, and fully understanding the relationship between predictor and criterion as well as the inter-predictor relationships. Every predictor in a regression model is assumed to contribute to the overall predictive effect. When examining a predictor's contribution, there are three main levels of effect people should be concerned with: direct effect, partial effect, and total effect (Budescu, 1993). The **direct effect** represents the independent contribution of each predictor for predicting the criterion, in isolation from all other predictors. The **partial effect** shows the explanatory ability of each predictor in a particular subset or subsets of predictors except the full model (i.e., conditional on a subset of the other predictors). The **total effect** reflects the contribution of each predictor in the full model (i.e., controlling for all other predictors).

Numerous indices have been proposed for evaluating predictor importance. Different measures may address the issue of predictor importance based on different levels of predictive effect. The current section provides an overall review by presenting and comparing different techniques for assessing predictor importance. Specifically, this review aims to show how each measure quantifies predictor importance in a regression model, what research questions each measure can address, and when researchers should select a particular measure to fit their goals.

2.1.1 The most common importance measures: r and β

Zero-order correlation coefficients (r 's) and standardized regression coefficients (β 's) are very commonly utilized and are provided by default in the output of all popular statistical software programs. Along with their increasing use as measures of importance, however, an increasing number of articles have questioned the utility of these traditional importance measures (Courville & Thompson, 2001; Nathans, Oswald & Nimon, 2012; Nimon, Gavrilova, & Roberts, 2010; Zientek, Capraro, & Capraro, 2008). These critiques are mainly focused on the overreliance and misinterpretation/misuse of these measures.

The **zero-order correlation coefficient** reflects the magnitude and direction of the bivariate relationship between a particular predictor and the single criterion, symbolized as r_{X_iY} and calculated using the following formula:

$$r_{X_iY} = \frac{\text{cov}_{X_iY}}{S_{X_i}S_Y} = \frac{\sum_{j=1}^n (X_{ij} - \bar{X}_i)(Y_j - \bar{Y})}{n-1} / S_{X_i}S_Y \quad (3)$$

$$= \frac{\sum_{j=1}^n X_{ij}Y_j}{n-1} \quad (\text{when } X_i \text{ and } Y \text{ are standized}) \quad (4)$$

The zero-order correlation coefficient quantifies predictor importance without considering the contributions of other predictors in the regression model. It only focuses on the extent to which the targeted predictor and criterion are related, and ignores the effect of other predictors in the model. Therefore, the zero-order correlation coefficient measures a predictor's direct effect on the criterion.

The **standardized regression coefficient** shows the rate of change in the predicted criterion as a function of a unit (i.e., a standard deviation) change in a predictor while the other predictors are held constant. It examines the predictor's total effect, based on the full model. The regression coefficients are estimated based on the linear least squares approach, which minimizes the sum of squared errors between the observed dependent variable values and the values predicted by the regression equation including all the predictors (i.e., $\sum_{j=1}^n e_j^2$, Pedhazur, 1997). In the calculation process, the regression coefficients are obtained by taking the contributions of all the predictors into account. The more predictors in the model, the more complicated the computation that is required. Taking the two-predictor model as an example:

$$\beta_1 = \frac{r_{YX_1} - (r_{YX_2})(r_{X_1X_2})}{1 - (r_{X_1X_2})^2} \quad (5)$$

and

$$\beta_2 = \frac{r_{YX_2} - (r_{YX_1})(r_{X_1X_2})}{1 - (r_{X_1X_2})^2} \quad (6)$$

From the formulas (5 and 6) above, it can be seen that the standardized regression coefficient of a predictor is calculated based on its relationship with Y (e.g., r_{YX_1}), the relationships between all the other predictors and Y (e.g., r_{YX_2}), as well as the inter-predictor relationship (e.g., $r_{X_1X_2}$).

When predictors are uncorrelated (i.e., $r_{X_iX_j} = 0$), each of them contributes to the overall predictive power independently. In such a case, each predictor's direct effect, partial effect and total effect are identical, and their standardized regression coefficient equals its zero-order correlation coefficient with the criterion. Either of the two coefficients is sufficient to represent predictor importance and even rank order the predictors in terms of their importance to predicting the criterion. It is not necessary to turn to more complicated measures since these traditional measures should fully capture what is meant by "importance". In this case the sum of all the predictors' squared standardized regression coefficients or zero-order correlation coefficients equals the overall regression effect (R^2).

When the selected predictors are inter-correlated, as in most realistic cases, they may overlap in accounting for the variance of the dependent variable. In such a case, the regression coefficients and the interpretations arising from them are context specific in that they can change dramatically with the addition or the deletion of a single predictor in the model. Further, the standardized regression coefficient for a given predictor no longer equals the zero-order correlation and the discrepancy may be quite large and even opposite in sign. According to a review by Courville and Thompson (2001), based on published articles in the *Journal of Applied Psychology* from 1987 to 1998, 94% of the collected articles contained at least one discrepancy between the standardized regression coefficient and correlation coefficients in regards to the rank ordering of the predictive power (or importance) of the predictors. This is not surprising because essentially the two measures address the predictor importance from two different perspectives; that is, the zero-order correlation only concerns the contribution of a particular predictor by itself while the standardized regression coefficient concerns a predictor's partial contribution in the presence of all other predictors. It is possible that one may discover that a predictor with a near-zero regression coefficient actually has the highest zero-order correlation with the criterion, or that a predictor weakly associated with criterion by itself turns out to have a substantial regression coefficient in a MR model. Given two related predictors that share the explained variance of a criterion with each other, it is possible that both of the regression coefficients would be underestimated after controlling for the shared variance, thereby resulting in inaccurate conclusions on the importance of the two predictors. For example, suppose that supervisor support and co-worker support are moderately correlated, and both of them are crucial predictors for job satisfaction base on theories.

Since they provide redundant information about job satisfaction, once supervisor support is in the MR model, co-worker support does not have much to contribute, and vice versa. This is why both of supervisor support and co-worker support may have small beta values, even though they both are important to predicting job satisfaction and, individually, highly correlated with job satisfaction.

It is also possible that one predictor would have a large regression coefficient only because it cancels out one or more remaining predictors' irrelevant variance, but does not contribute to the criterion substantially. These predictors, called "suppressors", are not directly related to the criterion but contribute to the regression equation through their relationship with other predictors (Tzelgov & Henik, 1991). For example, suppose that school SES and school districts are targeted as predictors of student achievement. The presence of school districts in the model will help to explain some irrelevant variances of school SES and thereby push up the total model contribution. In this case, school districts may have a large regression coefficient even though it is negligibly correlated with student achievement. Therefore, it is not appropriate for researchers to only rely on regression coefficients or zero-order correlation coefficient to address the issue of whether a predictor is important or not (Courville & Thompson, 2001).

2.1.2 Importance measures based on combining r and β

Product Measure. Just as the name implies, the product measure (P_i) is the product of a particular predictor's regression coefficient and its zero-order correlation coefficient with the criterion:

$$P_i = r_{X_iY} \times \beta_i \quad (7)$$

This measure was proposed by Pratt (1987) to combine the information from the regression coefficient and the zero-order correlation coefficient. The product measures of all the predictors in a model sum up to the squared multiple correlation coefficient (R^2) for the regression model. In this way, the product measure partitions the model regression effect and enables the rank ordering of the predictors, no matter whether they are correlated or not.

However, a major problem is that a predictor's product measure could be negative when the regression coefficient and the zero-order correlation coefficient are opposite in sign. The sign of the regression coefficient or the zero-order correlation coefficient only indicates the direction of the relationship between the predictor and the criterion, but does not affect the magnitude of its contribution. For example, a predictor with a negative regression coefficient or a negative zero-order correlation coefficient may also account for a large amount of variance in the criterion. However, it is awkward or even meaningless to interpret a negative value of the product measure when it is considered as

the proportion of total explained variance, since the contribution of the particular predictor would be subtracted from the total prediction effect rather than adding to it. In addition, this measure is simply a mathematical product and lacks a meaningful conceptual interpretation.

Structure coefficient. Courville and Thompson (2001) and Thompson (2006) suggest reporting structure coefficients (Cooley & Lohnes, 1971) to indicate the predictors' importance. The structure coefficient shows the bivariate correlation between a predictor and the predicted value of criterion (i.e., predicted by all the predictors), $\hat{y} = \sum_{i=1}^p \beta_i X_i$. The structure coefficient measures the direct effect of the predictors' contribution regardless of all other predictors. The structure coefficient (r_s) for a predictor (x_i) could be obtained by the following formula:

$$r_s = r_{x_i \hat{y}} \quad (8)$$

which is the Pearson correlation between a predictor (x_i) and the predicted value of the criterion (\hat{y}). This measure can also be calculated as the zero-order correlation between a predictor (x_i) and the criterion (y) divided by the full model's multiple correlation coefficient (R):

$$r_s = r_{x_i y} / R \quad (9)$$

The major difference between the structure coefficient and the zero-order correlation coefficient is that the structure coefficient examines the correlation between a predictor and the predicted value of the dependent variable (\hat{y}) instead of the observed value of the dependent variable (y). The structure coefficient indirectly considers the effect of the other predictors since they are all used to calculate \hat{y} (or R , in equation 9). However, when looking at Equation 9, the denominator of all structure coefficients is the same for predictors in the same model since there is only one R for a given model. Thus, the rank ordering of predictors based on their structure coefficients would not be different from those based on zero-order correlations as a measure of importance. Conceptually, both of them are only measuring the direct effect but ignoring the joint effect of predictors on the shared variance of the criterion, and they would not account for the inter-relationships among predictors.

Relative Weight. Relative weights analysis (Johnson, 2000; Johnson, 2004; LeBreton, Ployhart, & Ladd, 2004) has been proposed to evaluate relative importance in the context of multiple correlated predictors. In general, relative weight analysis approaches the problem of relative importance by adopting a principal components perspective.

First, it mathematically transforms the original predictors (X_i) into a new set of orthogonal “counterparts” ($Z_k, k = 1, \dots, p$) that are most highly correlated with the original predictors, and then the resulting standardized regression coefficients (β_k) for this new predictor set are obtained. Second, regressing each of the original predictors (X_i) on the new set of orthogonal predictor (Z_k), and then a new set of standardized regression coefficients (λ_{ik}) is obtained. Finally, the products of the squares of the two standardized regression coefficients are summed up to the relative weight:

$$RW_i = \sum_{k=1}^p \beta_k^2 \lambda_{ik}^2 \quad (10)$$

The RW_i values are the estimates of relative importance (called relative weights) for each predictor. β_k represents the individual contribution of predictor to criterion while λ_{ik} reflects the predictors’ joint contribution. From this point of view, relative weights take the predictor’s direct effect and total effect together into account. The relative weights of all predictors in a model add up to the model’s R^2 , thereby the percentage of total predictive variance could be computed by the predictor’s relative weight divided by the R^2 of the model.

However, this measure provides nothing specific to show the partial effect. Similar to the product measure, relative weights rely on a mathematical transformation, and it is hard to interpret their values conceptually. A recent reanalysis (Thomas, Zumbo, Kwan, &

Schweitzer, 2014) pointed out the theoretical flaw of the derivation of relative weights method and showed it can result in inaccurate inferences. They suggested using dominance measures instead as a variable importance metric for multiple linear regression, which is discussed in Section 2.3.

2.1.3 Importance measures based on R^2

Squared semi-partial correlation coefficient. Darlington (1990) recommended the use of the squared semi-partial correlation coefficient of a predictor as an indicator of its relative predictive power. The squared semi-partial correlation coefficient for a predictor (X_i) refers to the increase in R^2 when X_i is added to the regression model (consisting of all other predictors). It can be symbolized as ΔR^2 , and usually stands for the contribution of X_i to predicting the criterion after controlling for all other predictors in the model. Take a four-predictor model as an example. The additional contribution of the predictor X_4 to the model consisting of the other three predictors, X_1 , X_2 , and X_3 , is the increase in R^2 that occurs when X_4 is added to the three-predictor model. Specifically, the squared semi-partial correlation coefficient of $\Delta R_{X_4}^2$ could be obtained as follows:

$$\Delta R_{X_4}^2 = R_{Y \cdot X_1 X_2 X_3 X_4}^2 - R_{Y \cdot X_1 X_2 X_3}^2 \quad (11)$$

where $R_{Y \cdot X_1 X_2 X_3 X_4}^2$ is the R^2 of the full model consisting of all four predictors and $R_{Y \cdot X_1 X_2 X_3}^2$ is the R^2 of the subset model consisting of X_1 , X_2 , and X_3 .

If one's question concerns which predictor affects the criterion most strongly in terms of explained variation after controlling for the influence of all other predictors, the squared semi-partial correlation coefficient can be utilized to answer this. Therefore, it is a measure of total effect. Again, it only represents the contribution of predictor in the full model and neglects its contributions with other possible combinations of the remaining predictors (partial effect) or by itself (direct effect).

Commonality coefficient. This measure was developed in the late 1960's to provide a broader understanding of regression effects by partitioning the total explained variance into non-overlapping components (Mayeske et. al, 1969; Mood, 1969, 1971; Newton & Spurrell, 1967). It categorizes the regression effect into two forms: unique effects and common effects. Unique effects refer to the explanatory ability that can be attributed to a single predictor. It is essentially the squared semi-partial correlation of the predictor, measuring the total effect. Common effects reveal the shared variance of the criterion that is jointly explained by two or more predictors, measuring the partial effect. The sum of all the commonality coefficients (including the unique effects and common effects) in a regression model is the squared multiple correlation coefficient (R^2). This property provides information regarding the predictors' proportional contribution to the overall regression effect.

This approach has the same problem as the product measure in that a potential negative commonality coefficient will make interpretation difficult. However, negative values of commonality coefficients can be used to identify and quantify the extent and nature of a suppression effect (Amado, 1999). If a predictor is involved in several subsets with a negative common effect, the predictor may contribute to the explained variance by suppressing the irrelevant variance of other predictors. The summation of all negative common effects reveals the total amount of suppression in the model (Nathans, Oswald & Nimon, 2012).

2.2 Summary

Currently, there is no universal agreement on the definition of predictor importance. People generally consider importance as the relationship between a predictor and criterion or the contribution of a predictor to the prediction of the criterion. However, this is not necessarily an accurate or complete definition since the contribution would be different across different contexts in the analysis (i.e., if a different set of predictors is involved). In other words, the amount of a criterion's variance explained by one predictor would be affected by the relationship between the remaining predictor(s) and criterion as well as the relationships among the predictors given that the predictors are correlated. An inaccurate or incomplete understanding of what is meant by a predictor's "importance" could easily lead to misinterpretation and misuse of its corresponding measures (Azen & Budescu, 2003; Courville & Thompson, 2001).

In this study, the concept of “relative importance” was adopted as defined by Budescu (1993) and further refined by Azen and Budescu (2003, p. 134): “a predictor’s importance reflects its contribution in the prediction of the criterion at a given level of analysis.” Here, the level of analysis means the specific subset of predictors that are included in the regression model. The comparison between two predictors’ importance could be based on different combinations of specific predictors. From this point of view, the definition is intuitive, straightforward, and general as it fully consider all possible bases for comparison.

Furthermore, the current study chose to use Dominance Analysis to evaluate relative importance because it has several desired properties: 1) it is the only importance analysis to date that explicitly calculates and uses the direct effect, partial effect and total effect; 2) it is comprehensive and intuitive in that the results from every step are meaningful and easy to understand; 3) it provides a broad picture and plentiful information to address the issue of predictor importance at three levels of analysis: complete dominance, conditional dominance, and general dominance.

2.3 Dominance Analysis

2.3.1 General procedure

In Dominance Analysis, one predictor is considered as more important (dominant) than another in a given model when it increases the model’s R^2 more than another. For

example, in a model with four predictors, if the additional contribution of X_3 to the model consisting of X_1 and X_2 is more than that of X_4 to the same model, or

$$(R_{Y \cdot X_1 X_2 X_3}^2 - R_{Y \cdot X_1 X_2}^2) - (R_{Y \cdot X_1 X_2 X_4}^2 - R_{Y \cdot X_1 X_2}^2) \quad (12)$$

$$= R_{Y \cdot X_1 X_2 X_3}^2 - R_{Y \cdot X_1 X_2 X_4}^2 > 0 \quad (13)$$

It could be concluded that X_3 is relatively a more importance predictor than X_4 to the base model consisting of X_1 and X_2 . Mathematically, the result is just the difference between two squared multiple correlations since the base model $R^2 (R_{Y \cdot X_1 X_2}^2)$ is canceled out, as shown in Equation 13. It should be noted that if one compare two predictors only in the full model, the difference of their additional contributions is identical to the difference of their squared semi-partial correlation coefficients.

The dominance relationship could change across different contexts (i.e., base or subset models) in the analysis. Therefore, Dominance Analysis extends the use of the squared semi-partial correlation coefficient (ΔR^2) by comparing predictors across all relevant subsets of the full model. That is, it uses all possible sub-models that are comprised of every possible combination of the predictors.

Dominance Analysis does not serve the purpose of model selection or eliminating predictors. It allows for comparing and ranking ordering the predictor variables once the

correct model has already been identified. This method is particularly well suited for those situations when a priori ordering of variables cannot necessarily be justified by theory (Johnson & LeBreton, 2004).

2.3.2 Complete, conditional, and general dominance

There are three levels of dominance that can be achieved between each pair of predictors in Dominance Analysis: complete dominance, conditional dominance, and general dominance. One predictor is said to completely dominate another predictor if its dominance holds across all possible subset models (that do not include the two predictors under comparison). Back to the four-predictor model, for example, complete dominance (of X_3 over X_4) is achieved if the additional contribution of X_3 is more than that of X_4 to the null model, the model consisting of X_1 , the model consisting of X_2 , and the model consisting of both X_1 and X_2 , respectively, as follows:

$$R_{Y \cdot X_3}^2 - R_{Y \cdot X_4}^2 > 0 \quad (\text{model size}=1) \quad (14)$$

$$R_{Y \cdot X_1 X_3}^2 - R_{Y \cdot X_1 X_4}^2 > 0 \text{ and } R_{Y \cdot X_2 X_3}^2 - R_{Y \cdot X_2 X_4}^2 > 0 \quad (\text{model size}=2) \quad (15)$$

$$R_{Y \cdot X_1 X_2 X_3}^2 - R_{Y \cdot X_1 X_2 X_4}^2 > 0 \quad (\text{model size}=3) \quad (16)$$

However, if the additional contributions are inconsistent in favoring the same predictor across all subset models, then complete dominance is undetermined while weaker levels of dominance may still be achieved. If a predictor's averaged additional contribution

within each model size is greater than that of another predictor, then the first predictor is said to conditionally dominate the latter. Here, the model size is indicated by the number of predictors included in a given model. If a predictor's averaged additional contribution is greater for some model sizes but not for all, then conditional dominance between the two predictors cannot be established. In this case, general dominance can still be achieved if the average of a predictor's additional contribution over all possible model sizes is greater than that of another predictor. Thus, the general dominance measure for a predictor essentially measures the weighted average of differences between the R^2 of models that include it and the R^2 of models that exclude it at the same model size. Again take the example of four-predictor regression model. The general dominance measure for X_4 can be calculated as follows:

$$G_{X_4} = 1/4\{[R_{Y \cdot X_4}^2 - 0] \quad (17)$$

$$+ [(R_{Y \cdot X_1 X_4}^2 - R_{Y \cdot X_1}^2) + (R_{Y \cdot X_2 X_4}^2 - R_{Y \cdot X_2}^2) + (R_{Y \cdot X_3 X_4}^2 - R_{Y \cdot X_3}^2)]/3 \quad (18)$$

$$+ [(R_{Y \cdot X_1 X_2 X_4}^2 - R_{Y \cdot X_1 X_2}^2) + (R_{Y \cdot X_1 X_3 X_4}^2 - R_{Y \cdot X_1 X_3}^2) + (R_{Y \cdot X_2 X_3 X_4}^2 - R_{Y \cdot X_2 X_3}^2)]/3 \quad (19)$$

$$+ [(R_{Y \cdot X_1 X_2 X_3 X_4}^2 - R_{Y \cdot X_1 X_2 X_3}^2)] \quad (20)$$

The statistic G_{X_i} , which is a quantitative measure of general dominance indicating the overall average contribution of each predictor in the model, could be decomposed into three components: the direct effect, partial effect, and total effect (LeBreton et al., 2004). In the formulas above, the R^2 difference shown in Equation 17 represents the direct effect

of X_4 (i.e., the base model for comparison is the null model), the R^2 differences shown in the Equation 18 and 19 represent the partial effect components (i.e., the base models for comparison are a variety of sub-models), and the R^2 difference shown in the Equation 20 reflects the total effect of X_4 on predicting Y (i.e., the full model). The individual general dominance measures of all predictors sum up to the total predictive effect, R^2 . In this sense, one can obtain the percentage that each predictor contributes out of the overall regression effect by using its general dominance measure divided by R^2 .

To make the presentation more concrete, a numerical example is included here to illustrate the Dominance Analysis procedure. This example used Matrix 3, shown in Table 3.1, as the population correlation matrix. The additional contributions of predictors in every subset model are shown in Table 2.1, as well as the averaged additional contributions within each model size and the overall averaged additional contributions.

Table 2.1 Numerical example of Dominance Analysis in the population using a four-predictor model

Subset model (\mathbf{x})	$\rho^2_{\mathbf{xy}}$	Additional Contribution			
		X_1	X_2	X_3	X_4
$k = 0$ average	.00	.09	.25	.49	.49
X_1	.09		.16	.47	.47
X_2	.25	.00		.24	.24
X_3	.49	.07	.00		.09

Subset model (\mathbf{x})	$\rho^2_{\mathbf{xY}}$	Additional Contribution			
		X_1	X_2	X_3	X_4
X_4	.49	.07	.00	.09	
$k = 1$ average		.05	.06	.27	.27
$X_1 X_2$.25			.32	.32
$X_1 X_3$.56		.02		.20
$X_1 X_4$.56		.02	.20	
$X_2 X_3$.49	.09			.10
$X_2 X_4$.49	.09		.10	
$X_3 X_4$.58	.18	.01		
$k = 2$ average		.12	.02	.21	.21
$X_1 X_2 X_3$.58				.18
$X_1 X_2 X_4$.58			.18	
$X_1 X_3 X_4$.76		.00		
$X_2 X_3 X_4$.59	.17			
$k = 3$ average		.17	.00	.18	.18
$X_1 X_2 X_3 X_4$.77				
Overall average		.11	.08	.29	.29

Note: $\rho^2_{\mathbf{xY}}$ represents the squared multiple correlation of model \mathbf{x} . The column labeled X_i shows the additional contribution of X_i to the model \mathbf{x} in the corresponding row. A blank means that an additional contribution is not applicable.

In this case, it can be concluded that X_3 and X_4 perform equally well at all three dominance levels, although this is very rare in realistic cases. Both X_3 and X_4 completely

dominate X_2 and X_1 , because their additional contributions are larger than those of X_2 and X_1 across all possible subset models. However, there is no complete dominance or conditional dominance established between X_1 and X_2 , since the averaged additional contribution of X_1 is greater than that of X_2 when model sizes = 2 but smaller when model sizes = 1. The last row provides the general dominance measures for each predictor, which is the mean of the four averaged additional contributions within each model size. Note that the sum of the four general dominance measures ($.11 + .08 + .29 + .29 = .77$) is equal to the R^2 of the full model. Based on the general dominance measures, one can rank order the four predictors as $X_4 = X_3 > X_1 > X_2$.

Note that complete dominance implies conditional dominance, and conditional dominance implies general dominance (if $p > 2$) (Razia & Budescu, 2003). Although general dominance is the weakest level of the three, it is also the easiest to achieve. Therefore, the current study focuses on statistical inference for the general dominance measures.

2.4 Statistical inference about importance measures

In empirical studies, the conclusions about predictor importance are purely based on sample data. Though the truth in population is unknown, one can make statistical inference about populations using data drawn from the population of interest by random sampling.

In the comparison of two predictors regarding their relative importance in a given MR model, a confidence interval for the difference between their general dominance measures would not only provide the results of hypothesis testing about the difference, but also the degree to which the two measures differ. In general, if the $(1-\alpha)$ % confidence interval does not include the population parameter tested under the null hypothesis, it can be concluded that the null hypothesis is rejected at the α significance level. In this case, to establish general dominance, the null hypothesis would be that the difference between the general dominance measures of two predictors is zero. A confidence interval would allow researchers to both test this hypothesis and determine an interval estimate for the magnitude of the difference that exists in the population.

2.4.1 Asymptotic confidence interval

In the past three decades, a series of methods were suggested related to the significance testing of two squared multiple correlation coefficients. It began with the asymptotic joint distribution of commonality components derived by Hedges and Olkin (1983). They proposed a way to calculate the corresponding asymptotic covariance matrix for each commonality component. Beyond that, Olkin and Finn (1995) proposed a method to construct asymptotic confidence intervals for simple, partial and multiple correlation coefficients. The construction of an asymptotic confidence interval relies on the assumption that these squared multiple correlation coefficients are normally distributed. This work was further simplified and generalized to the contexts when there are more than two independent variables in the MR model (Alf & Graf, 1999; Graf & Alf, 1999).

Specifically, let Model A and Model B be two MR models predicting the same single criterion Y . Model A has a set of predictors and their predicted criterion is the weighted sum of these predictors, represented by \hat{y}_A . The multiple correlation coefficient of Model A is essentially the zero-order correlation coefficient between \hat{y}_A and Y , denoted by $r_{Y\hat{y}_A}$. Similarly, in Model B, the predicted criterion is notated as \hat{y}_B , and its multiple correlation coefficient is represented by $r_{Y\hat{y}_B}$. It is possible to have overlap between the set of predictors in Model A and that in Model B. In this case, according to Alf and Graf (1999), the $100(1-\alpha)$ % asymptotic confidence interval (based on n observations) for the difference between the two squared multiple correlations is:

$$r_{Y\hat{y}_A}^2 - r_{Y\hat{y}_B}^2 \pm z_{\alpha/2} \sqrt{Var(r_{Y\hat{y}_A}^2 - r_{Y\hat{y}_B}^2)} \quad (21)$$

$$\text{where } Var(r_{Y\hat{y}_A}^2 - r_{Y\hat{y}_B}^2) = \frac{4r_{Y\hat{y}_A}^2(1 - r_{Y\hat{y}_A}^2)}{n} + \frac{4r_{Y\hat{y}_B}^2(1 - r_{Y\hat{y}_B}^2)}{n} \quad (22)$$

$$- \frac{8r_{Y\hat{y}_A}r_{Y\hat{y}_B} \left[\frac{1}{2}(r_{\hat{y}_A\hat{y}_B} - r_{Y\hat{y}_A}r_{Y\hat{y}_B}) \left(1 - r_{Y\hat{y}_A}^2 - r_{Y\hat{y}_B}^2 - \frac{r_{Y\hat{y}_B}^2}{r_{Y\hat{y}_A}^2} \right) + \frac{r_{Y\hat{y}_B}^3}{r_{Y\hat{y}_A}^3} \right]}{n}, \quad (23)$$

$$\text{and } r_{\hat{y}_A\hat{y}_B} = \frac{\sum(\beta_i \sum r_{ij} \beta_j)}{r_{Y\hat{y}_A}r_{Y\hat{y}_B}}. \quad (24)$$

In the last formula, β_i is the regression coefficient for variable i in Model A, while β_j is the regression coefficient for variable j in Model B, and the summation is across all predictors in the two models combined. Besides the use of z value in the confidence interval construction, the asymptotic variance of the difference between the two squared

multiple correlations rely heavily on the large-sample theory and normal theory (Olkin & Siotani, 1976; Hedges & Olkin, 1983).

However, there are some difficulties and potential problems if people want to adopt this approach to construct the confidence interval for the difference in magnitude of two general dominance measures. Most importantly, the above asymptotic procedure relies on a normal approximation and large sample assumption, but the sampling distribution of the difference in magnitude of two general dominance measures is unknown. Plus, the large sample assumption is not always satisfied in empirical studies. A simulation study (Azen & Sass, 2008) showed that a sample size of 300 may not be sufficient to achieve adequate power when the difference between two squared multiple correlation coefficients is smaller than 0.1, by using the asymptotic method to compare the squared multiple correlations for hypothesis testing purposes. Therefore, we turn to another technique that frees the researchers from parametric assumptions: the bootstrap.

2.4.2 Bootstrap confidence interval

The bootstrap is a nonparametric approach for statistical significance testing based on intensive computer-based resampling (Efron, 1979, 1982; Efron & Tibshirani, 1986, 1993; Davison & Hinkley, 1997). Johnson (2001) and Carpenter and Bithell (2000) also provided some elementary introduction to the bootstrap. It can be applied to different domains of statistical inference such as hypothesis testing, estimation of standard errors,

and construction of confidence intervals. In general, the bootstrap procedure is as follows:

1. Draw a resample of size N by independent random sampling with replacement from an observed data set of N cases. In this resample, some of the original N cases may be replicated and some of the cases may not be selected.
2. Perform the statistical analysis and estimate the sample statistic, denoted by $\hat{\theta}$, based on the resample in step 1. Preserve the values of the statistics of interest.
3. Repeat step 1 and step 2 a large number of times, B , and then build up a bootstrap distribution of the statistic of interest, $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_B$.

There are three types of confidence intervals one can construct from the bootstrap distribution of the test statistics. Among the three, the standardized confidence interval can be obtained as

$$[\hat{\theta}^* - z_{\alpha/2} \hat{\sigma}^*, \hat{\theta}^* + z_{\alpha/2} \hat{\sigma}^*] \quad (25)$$

$$\text{where } \hat{\theta}^* = \frac{1}{B} \sum_{b=1}^B \hat{\theta}_b \quad (26)$$

$$\text{and } \hat{\sigma}^* = \sqrt{\frac{1}{B-1} \sum_{b=1}^B (\hat{\theta}_b - \hat{\theta}^*)^2} \quad (27)$$

Here $\hat{\theta}^*$ is the mean of the bootstrap distribution, $\hat{\sigma}^*$ is the standard deviation of the bootstrap distribution, and $z_{\alpha/2}$ is the standard normal deviate corresponding to the chosen confidence level $100(1-\alpha)$ % or Type I error rate, α (Robertson, 1991).

This approach is analogous to its parametric counterpart in that it also assumes the studied statistic is normally distributed. In other words, the standard normal confidence interval is valid on the basis of the argument that $(\hat{\theta}_b - \hat{\theta}^*)/\hat{\sigma}^*$ has asymptotically a standard normal distribution as $N \rightarrow \infty$ and $B \rightarrow \infty$. Thus, it uses the bootstrap distribution to estimate the standard error but fails to fully take advantage of the bootstrap's nonparametric property because it still relies on the normal assumption to construct the confidence interval. In other words, the only difference between this method and the asymptotic method to constructing confidence intervals is that in the bootstrap procedure the mean and the standard error are estimated based on the bootstrapped sampling distribution. It should be noted that this method requires fewer bootstrap replications than the percentile bootstrap confidence interval and the bias-corrected bootstrap confidence interval, which will be described next.

The percentile bootstrap confidence interval estimates the percentile points of the confidence interval empirically from the observed bootstrap distribution of the statistic. Specifically, the bootstrap estimates of the test statistic are sorted in ascending order and the new ordered estimates are represented by $\hat{\theta}[1] \leq \hat{\theta}[2] \leq \dots \leq \hat{\theta}[B]$. The desired $100(1-\alpha)$ % bootstrap confidence interval endpoints are empirically located at the

particular percentiles in the bootstrapped sampling distribution. Specifically, the $B_{\frac{\alpha}{2}}$ percentile and $B(1 - \frac{\alpha}{2})$ percentile are, respectively, the lower and upper end points of the interval.

The percentile method does not require any assumptions about the distribution of the studied statistic. It allows for an asymmetric confidence interval around the expected value of the statistic if it has a skewed distributed in the population. The percentile method is quite straightforward, intuitive, and easy to perform without any complex analytical formulas. Although the number of bootstrapping replications should be large enough to assure the accuracy of the estimation, this should not be problematic given modern data processing technology.

However, the percentile method assumes that the bootstrapped sampling distribution is unbiased for the sample distribution of the studied statistic, which still puts some restrictions on the method. Therefore, Efron (1987) proposed a bias-corrected bootstrap confidence interval to adjust for possible bias in the bootstrapped estimator($\hat{\theta}_b$). Here, a biasing constant for adjusting the bootstrapped sampling distribution is introduced, notated as z_0 , which could be considered as the standard normal deviate corresponding to the probability of the bootstrapped sampling statistic ($\hat{\theta}_b$) that lies below the sampling statistic ($\hat{\theta}$):

$$\Phi(z_0) = Prob \{ \hat{\theta}_b \leq \hat{\theta} \} \quad (28)$$

where Φ is the cumulative normal distribution function. For this confidence interval, the upper and lower interval endpoints correspond to standard normal deviates of $2z_0 - z_{\alpha/2}$ and $2z_0 + z_{\alpha/2}$. Specifically, the lower boundary of the bias-corrected confidence interval equals to the value of $\hat{\theta}^*$ at the $[\{\Phi(2z_0 + z_{\alpha/2})\} \times 100]$ percentile, and the upper boundary equals to the value of $\hat{\theta}^*$ at the $[\{\Phi(2z_0 + z_{1-\alpha/2})\} \times 100]$ percentile. It is specifically designed to deal with asymmetric or skewed sampling distributions. For example, if 40% of the bootstrap estimates of the test statistic is less than or equals to the test statistic from parent sample, a corresponding z score would be: $z_0 = -0.2533$. Given that $z_{0.975} = 1.96$ and $z_{0.025} = -1.96$, the resulting endpoints of bias corrected confidence interval are the percentiles for $z = -0.2533 \times 2 - 1.96 = -2.4666$ and $z = -0.2533 \times 2 + 1.96 = 1.4534$, which are the 1st and 93rd percentiles of the bootstrapped sampling statistic and are used for the lower boundary and upper boundary, respectively.

Azen and Budescu (2003) addressed the issue of statistical inference of dominance measures using a different approach. They recoded the dominance measure as 1 (X_i dominants X_j), 0 (X_j dominants X_i), or 0.5 (dominance cannot be established between X_i and X_j), and used reproducibility to show the stability and robustness of dominance patterns. Specifically, reproducibility is the percentage of B bootstrap resamples that reproduced the same dominance pattern observed in the parent sample. However, this manual recoding of the continuous dominance measures into categorical variables will

cause information loss. For example, the difference between the two dominance measures $D_{uv} = 0.01$ and $D_{uv} = 0.50$ are indicative of very different situations, but both of them would be treated as 1; however, $D_{uv} = 0.01$ and $D_{uv} = -0.01$ both produce similar dominance measures but they would be treated as 1 and 0, respectively. Here, D_{uv} represents the difference between the general dominance measures of X_u and X_v .

In this study, the asymptotic normal, parametric bootstrap, percentile bootstrap and bias-corrected bootstrap confidence interval are all employed to determine whether there is a significant difference between two general dominance measures, given that the exact distribution of general dominance measures is unknown. Besides examining normally distributed data, the current study also introduced several non-normal distributions in a simulation study. Sometimes, the empirical data may fail to satisfy the normality assumption (for example, due to the outliers, being truncated, or being limited by the floor/ceiling effects). Hence, it is important to check and compare the performance of the four proposed inference methods when the data distribution is normal and non-normal.

It is expected that the percentile and bias-corrected bootstrap confidence interval approaches would outperform methods that assume normality when sample sizes are small, or data are non-normal (Chan, 2009). It would also be interesting to compare the performance of these two confidence intervals. If the bias-corrected bootstrap confidence interval outperforms its percentile counterpart, then the assumption that the bootstrapped

sampling distribution is unbiased for the distribution of the studied statistic is violated and a bias correction is necessary.

Chapter 3. Methodology

3.1 Simulation

3.1.1 General procedures

A series of simulation studies were conducted to examine the empirical performance of the four confidence interval approaches (asymptotic confidence interval, standardized bootstrap confidence interval, percentile bootstrap confidence interval and bias-corrected bootstrap confidence interval) for making statistical inferences regarding the difference between two general dominance measures, under different conditions. Specifically, the study examined how the magnitude of the dominance effect size, sample size and different distributions affects the performance of the proposed approaches. The four confidence interval approaches were compared in four aspects: type I error rate, power, accuracy, and confidence interval width. Recommendations were provided for empirical researchers based on the performance under different scenarios.

In significance testing for Dominance Analysis, the null hypothesis can be stated as: in the population the general dominance measure of a particular predictor is equal to the dominance measure of another predictor. If the confidence interval for the difference between the general dominance measures from two predictors does not include 0, the dominance measure of one particular predictor is significantly different from the dominance measure of the other predictor. Therefore, the null hypothesis would be rejected.

This study adopted the bootstrapping procedure (Efron, 1979) by sampling with replacement a large number of times from an existing data set. This is a frequently used technique for drawing inferences from a sample with an unknown distribution to a population. The specific steps for evaluating the performance of the procedure in the simulation study are as follows:

STEP 1. Generate a random sample of n observations as a “parent sample”, drawn from a distribution with a given population correlation matrix, and obtain the differences between the general dominance measures of every two predictors (e.g., X_u and X_v), denoted as $\widehat{D}_{uv} = \widehat{G}_u - \widehat{G}_v$, based on the “parent sample”;

STEP 2. Construct the asymptotic confidence interval about each dominance difference, \widehat{D}_{uv} ;

STEP 3. Use nonparametric (bootstrap) sampling with replacement to create a large number (i.e., $B = 400$) of bootstrapped resamples from the “parent sample”, and calculate the general dominance measures, for all predictors in each bootstrapped resample;

STEP 4. Calculate the differences between any two general dominance measures in the bootstrap resamples, $\widehat{D}_{uv}(b) = \widehat{G}_u(b) - \widehat{G}_v(b)$, and build up a corresponding bootstrap distribution of the general dominance measure differences sorted from lowest to highest based on all of the bootstrap resamples, $\widehat{D}_{uv}(1) < \widehat{D}_{uv}(2) < \dots < \widehat{D}_{uv}(B)$;

STEP 5. Construct the 95% standard normal confidence interval as $[\hat{D}_{uv}^* - 1.96\hat{\sigma}^*, \hat{D}_{uv}^* + 1.96\hat{\sigma}^*]$, using the mean (\hat{D}_{uv}^*) and standard deviation ($\hat{\sigma}^*$) of the bootstrap distribution;

STEP 6. Construct a 95% percentile confidence interval around each difference between two dominance measures by identifying the values corresponding to the 2.5th percentile (i.e., $\hat{D}_{uv}(2.5B)$) and the 97.5th percentile (i.e., $\hat{D}_{uv}(97.5B)$) of the bootstrap distribution;

STEP 7. Construct a 95% bias-corrected confidence interval: first obtain the biasing constant z_0 such that $\Phi(z_0) = Prob\{\hat{D}_{uv}(i) \leq \hat{D}_{uv}\}$ and then use z_0 to adjust the bootstrapped sampling distribution. The lower boundary of the bias-corrected confidence interval equals to the value of $\hat{D}_{uv}(i)$ at the $[\{\Phi(2z_0 + z_{\alpha/2})\} \times 100]$ percentile, and the upper boundary equals to the value of $\hat{D}_{uv}(i)$ at the $[\{\Phi(2z_0 + z_{1-\alpha/2})\} \times 100]$ percentile.

If the confidence interval for the difference does not include zero, the predictor with the higher dominance measure significantly dominates the other. These steps were repeated 100 times for each condition (listed below) to calculate the percentage of times the null hypothesis is rejected when the null hypothesis is true in the population (Type I error rate) and when the null hypothesis is not true in population (power).

Type I error rate

To evaluate the Type I error rate, the null hypothesis rejection rate was examined when the null hypothesis is true in the population (e.g., Matrices 1 and 2 of Table 3.1, where $D_{43} = 0$). For these cases, the proportion of the 100 replications in which the 95% confidence interval did not include the value of zero represents the Type I error rate. Specifically, this rejection rate is expected to be 5% in the null case if 95% confidence intervals are used to make statistical inference; in other words, 5% of the confidence intervals are expected to exclude the true value (i.e., zero). Bradley (1978) suggests that the most liberal acceptable deviation from the expected probability in the null case should be $.5\alpha$ to 1.5α , or .025 to .075 when $\alpha = .05$. For the procedure to be adequate, the Type I error rate is expected to be close to 0.05.

Power

The non-null case is the case where the null hypothesis is not true in the population. In this study, non-null cases are those in which the general dominance difference is not zero in the population. For those cases, the rejection rate represents the power of the null hypothesis test. To evaluate power, the proportion of the 100 replications in which the 95% confidence interval did not include the value of zero was recorded. In this study, there were 45 non-null cases in total. Power of 0.8 or above is the minimum acceptable statistical level for a proposed procedure (Cohen, 1988).

Accuracy

To evaluate whether the confidence intervals provide an accurate estimate of the dominance measure difference in the population, the proportion of 100 replications in which the 95% confidence interval around the dominance difference included the true (population) value (D_{uv}) was examined. It would be ideal that the accuracy rate is 0.95 when α level was set at 0.05. According to suggestions by Bradley (1978), $1-1.5\alpha$ and $1-0.5\alpha$ were used as lenient criteria to evaluate the $100(1-\alpha)$ % intervals. Specifically, empirical probabilities that fall outside the range of .925 to .975 are considered undesirable in the current study. If the accuracy rate is greater than 0.95, it is considered as “conservative”; if it is less than 0.95, it is considered as “liberal”. In general, conservative intervals are preferred over liberal ones (Smithson, 2003).

Confidence interval width

The confidence interval width was computed by taking the difference between the upper boundary and the lower boundary of the confidence interval. On the basis of a fairly good accuracy, the narrower the confidence interval, the more precise and informative it is when estimating the unknown population parameters. The width of the confidence interval depends to a large extent on the variability of the measures: the greater the standard deviation, the wider the confidence interval.

Bias

The extent to which the dominance differences produced by the bootstrapping procedure were unbiased estimates of their corresponding population parameters was also examined. Specifically, the magnitude of standardized bias was computed as the difference between the averaged bootstrapping estimate ($\overline{\hat{D}_{uv}^*}$) across all the 100 replications and their corresponding population parameters (D_{uv}), divided by the standard deviation of the bootstrapping estimates, $S(\hat{D}_{uv}^*)$, across all 100 replications.

$$\text{Standardized Bias}_{(pop)} = \frac{\overline{\hat{D}_{uv}^*} - D_{uv}}{S(\hat{D}_{uv}^*)} \quad (29)$$

Analogous to a Cohen's d , this approach measures the standardized distance between the bootstrapping estimate and the population parameter. According to the guidelines for Cohen's d regarding the effect size of this distance (Cohen, 1988), standardized bias that is less than 0.2 is considered as small and acceptable. Similarly, another standardized bias was also computed to indicate the distance between the bootstrapping estimates and the parent sample estimates (See Equation 30).

$$\text{Standardized Bias}_{(parent)} = \frac{\overline{\hat{D}_{uv}^*} - \overline{\hat{D}_{uv}}}{S(\hat{D}_{uv}^*)} \quad (30)$$

where $\overline{\hat{D}_{uv}}$ is the averaged parent sample estimates across 100 replications

For reference, the standardized bias between the parent sample estimates and population parameters are also calculated, using the difference between the averaged parent sample estimate ($\overline{\hat{D}_{uv}}$) across all the 100 replications and their corresponding population

parameters (D_{uv}), divided by the standard deviation of the parent sample estimates, $S(\hat{D}_{uv})$, across all 100 replications.

$$\text{Standardized Bias}_{(parent)} = \frac{\overline{\hat{D}_{uv}^*} - \overline{\hat{D}_{uv}}}{S(\hat{D}_{uv}^*)} \quad (31)$$

3.1.2 Specific conditions

Normal distribution

The conditions included 8 correlation patterns and 3 sample sizes: $n = 30, 100$, or 300 , in a fully crossed design. The rationale to pick 30 as the lowest level of sample size is because this number is a rule of thumb in the field of social science for an adequate sample size, considering the central limit theorem. The largest sample size was selected as 300 because previous studies (e.g., Hedges & Olkin, 1983; Azen & Sass, 2008) have shown that it is often adequate for asymptotically based inferential methods when the data are normally distributed. As the geometric mean between 30 and 300 , 100 is also a common level of sample size in behavioral studies. The 8 correlation matrices used in the data generation of the simulation study are shown in Table 3.1, and represent 3 correlation patterns between predictors and criterion combined with 4 inter-predictor correlation patterns.

The patterns for correlations with the criterion include: 1. increasingly correlated with two equally high correlation coefficients (matrices 1-3); 2. all equally moderately correlated (matrix 4); 3. increasingly correlated without equal correlation coefficients

(matrices 5-8). The patterns for the correlations among predictors include: 1. uncorrelated (matrix 5); 2. equally low correlated (matrix 1 and 6); 3. equally high correlated (matrix 2 and 7); 4. randomly moderate correlated (matrix 3 and 8). According to the common rule of thumb (Hinkle, Wiersma, & Jurs, 2003), the correlation coefficients of .3, .5, and .7 represent low, moderate and high correlations, respectively. These population correlation matrices were selected to represent and produce a variety of dominance measure differences, ranging from 0 to .48, as shown in Table 3.2. All data were generated by SAS 9.3 (SAS Institute, 2002).

Table 3.1 Population correlation matrices used in simulation study

	ρ_{XiY}	ρ_{XiXj}			
	Y	X ₁	X ₂	X ₃	X ₄
Matrix 1					
Y	1				
X ₁	.3	1			
X ₂	.5	.3	1		
X ₃	.7	.3	.3	1	
X ₄	.7	.3	.3	.3	1
Matrix 2					
Y	1				
X ₁	.3	1			
X ₂	.5	.7	1		
X ₃	.7	.7	.7	1	
X ₄	.7	.7	.7	.7	1
Matrix 3					
Y	1				
X ₁	.3	1			
X ₂	.5	.5	1		

X ₃	.7	.4	.6	1	
X ₄	.7	.5	.4	.6	1

Matrix 4

Y	1				
X ₁	.5	1			
X ₂	.5	.5	1		
X ₃	.5	.4	.6	1	
X ₄	.5	.5	.4	.6	1

Matrix 5

Y	1				
X ₁	.1	1			
X ₂	.3	0	1		
X ₃	.5	0	0	1	
X ₄	.7	0	0	0	1

Matrix 6

Y	1				
X ₁	.1	1			
X ₂	.3	.3	1		
X ₃	.5	.3	.3	1	
X ₄	.7	.3	.3	.3	1

Matrix 7

Y	1				
X ₁	.1	1			
X ₂	.3	.7	1		
X ₃	.5	.7	.7	1	
X ₄	.7	.7	.7	.7	1

Matrix 8

Y	1				
X ₁	.1	1			
X ₂	.3	.5	1		
X ₃	.5	.4	.6	1	

X_4 .7 .5 .4 .6 1

Note: $\rho_{X_i Y}$ represents the correlation coefficient between predictor X_i and criterion Y ;
 $\rho_{X_i X_j}$ represents the correlation coefficient between predictor X_i and another predictor X_j ;

Table 3.2 Difference between two general dominance measures in population

	D_{41}	D_{42}	D_{43}	D_{31}	D_{32}	D_{21}
Matrix 1	0.30	0.20	0.00	0.30	0.20	0.09
Matrix 2	0.18	0.20	0.00	0.18	0.20	-0.03
Matrix 3	0.24	0.18	0.03	0.21	0.15	0.06
Matrix 4	-0.01	0.00	0.01	-0.01	-0.01	-0.01
Matrix 5	0.48	0.40	0.24	0.24	0.16	0.08
Matrix 6	0.38	0.37	0.24	0.15	0.13	0.02
Matrix 7	0.32	0.44	0.34	-0.02	0.11	-0.12
Matrix 8	0.35	0.37	0.28	0.06	0.08	-0.02

Note: D_{uv} represents the difference between the general dominance measures of X_u and X_v .

Non-normal distribution

In terms of the variables' distributions, the simulation study extended the proposed procedure to examine the effect of non-normality on the results. It may be interesting to compare the performance of the proposed dominance inference methods under different population distributions, since the distribution of the difference of two general dominance measures may change if the data distribution changed.

The conditions included 4 correlation patterns (Matrices 1-4 in Table 3.1), 6 non-normal sample distributions, and 2 sample sizes ($n = 100$ and 300) in a crossed design. To generate the non-normal distributed data, first a very large multivariate normal sample ($N = 1,000,000$) was generated based on the correlation matrix, and it was then transformed to be non-normally distributed with a desired degree of skew and kurtosis, according to Fleishman's approach (1978). This method applies a cubic transformation to a standard normal distribution to obtain a non-normal distribution with specific skewness and kurtosis. If Y is a standard normal random variable, a new non-normal random variable (i.e., $Y_{\text{non-normal}}$) with could be computed by the polynomial:

$$Y_{\text{non-normal}} = -b + aY + bY^2 + cY^3 \quad (32)$$

Given the values of cubic coefficients (a , b , and c), the skewness and kurtosis can be calculated. Therefore, a root-finding method could be employed to solve the inverse problem and figure out the corresponding cubic coefficients for specified skewness and kurtosis. The current study used the SAS program provided by Wicklin (2013) to find out the cubic coefficients for transformation. These transformed N observations were treated as the population and the parent sample was randomly drawn from it for the simulations. It should be noted that the population parameters for the difference between two general dominance measures varied a little bit around the corresponding values showed in Table 3.2 after transformation. This approach has been widely used to generate non-normal distributed data in simulation studies (Finch, West & MacKinnon, 1997; Hau & Marsh, 2004; Sharma, Durvasula, & Dillon, 1989).

Here, the skewness of the distribution of N data points (X_1, X_2, \dots, X_N) could be calculated using the following formula:

$$skewness = \frac{\sum_{i=1}^N (X_i - \bar{X})^3}{(N-1)S^3} \quad (33)$$

where \bar{X} is the mean, S is the standard deviation, and N is sample size. The skewness for a symmetric distribution, including normal distribution, is zero. The greater the magnitude of skewness, the more skewed the distribution. The formula for kurtosis is as follows:

$$kurtosis = \frac{\sum_{i=1}^N (X_i - \bar{X})^4}{(N-1)S^4} - 3 \quad (34)$$

This is often referred to as “excess kurtosis” in that using this formula the normal distribution has a kurtosis of zero. Negative values for the kurtosis indicate data that are flat or “platykurtic” and positive values indicate data that are peaked or “leptokurtic”.

To obtain different non-normal distributions, the degree of skewness and kurtosis was manipulated systematically. Specifically, skew was fixed at either 0.5, 0.75 or 1 while holding kurtosis at 0, and kurtosis was fixed at either -1.2 (uniform distribution), 1.2 (logistic distribution), or 3 (Laplace distribution) while holding skew at 0. Figure 3.1 illustrate the distributions of the different conditions generated. The same transformation procedure was applied to Y and X 's, so that their distributions had the same skewness and kurtosis.

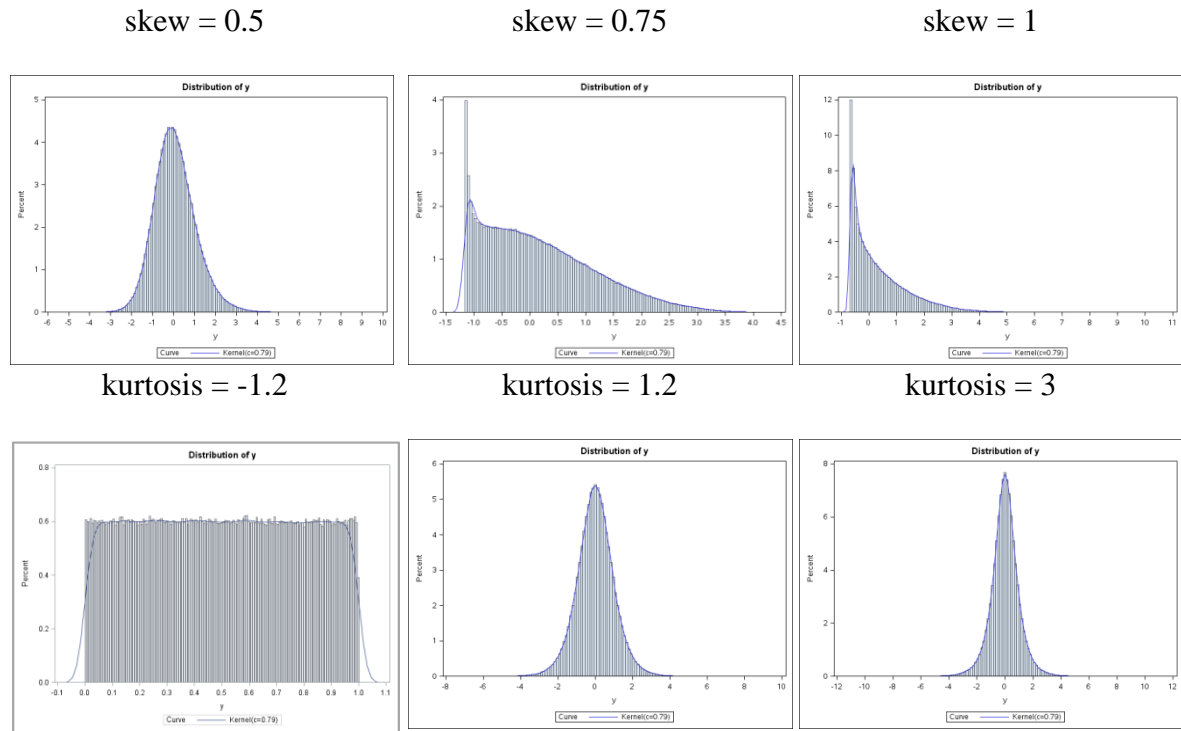


Figure 3.1 Histograms of six non-normal distributions

3.2 Empirical Demonstration

To make the study more concrete, an empirical example was presented to illustrate how to apply the proposed procedure and demonstrate how to interpret the results. Wisconsin Electric Power Company (We Energies) is a utility company, providing electrical service for over one million customers. According to the state law, there is a six-month moratorium per year when service for residential customers cannot be disconnected. During this period, customers can stop paying their bills but still receive the service, which in-turn dramatically increases overall outstanding debt. Therefore, effectively managing customer debt is a very crucial topic to We Energies.

As a collection strategy, some of the customers, who are two months past due, are randomly selected to receive an outbound call and have an option to contact with a customer service representative. If it is known which customers are most likely to pay and pay more, the debt collection efforts can be better targeted and the efficiency of the call center can be increased. As a result, a prediction model is built to identify and compare the key predictors of the amount of customers' payment, and thereby target customers likely to make larger payments after an outbound call.

The study focused on the customers who got outbound calls, and the outcome variable was the payment amount by these customers. The purpose of the current project is to rank order the potential predictors of payment amount in terms of their contribution and find out the most important predictor(s) to maximize the customers' payment amount. The data set was provided by We Energies, combining customer contact data and monthly billing data. Customer contact data details out the date when an "important notice" is obtained, which means the customer is two months past due, and when an outbound call occurred. For this example, a sample of customers was selected who received an important notice on July 1st and outbound call on July 7th ($N = 448$). Monthly billing data contains detailed customer and account related information, including the criterion, last payment amount (Y), and many potential predictors. In this example, four predictors were selected: the age of the account (X_1), total amount of arrears (X_2), and risk score (X_3), and age of oldest arrears (X_4). The age of the account is measured by the active days of the account so far; the amount of arrears is measured by the account's total outstanding payment amount; risk score is an account-level indicator evaluating the customer's risk

level; and the age of oldest arrears is measured by the days since the first time an arrear occurred for the account.

Chapter 4. Results

In this section, the results of both the simulation study and the empirical case study are presented. The simulation results are presented in two parts: results for the normal distribution conditions and for the non-normal distribution conditions. In the normal distribution cases, data was generated using a multivariate-normal population distribution based on 8 matrices (Matrix 1-8, see Table 3.1) and 3 sample sizes ($n = 30, 100$ or 300) for a total of 24 conditions. In the non-normal distribution cases, data was generated using variety of non-normal population distributions, consisting of 4 (Matrix 1-4) $\times 2$ (Sample size = 100 and 300) $\times 3$ (Skew = $.0.5, 0.75$ and 1) $\times 3$ (Kurtosis = $-1.2, 1.2$ and 3) = 72 conditions. A normal distribution could also be considered as a special case (skew = 0 and kurtosis = 0). Therefore, the simulation results of Matrix 1-4 in normal distribution section were also included here for the purpose of comparisons. To evaluate the performance of the proposed four inferential approaches (asymptotic confidence interval, standardized bootstrap confidence interval, percentile bootstrap confidence interval and bias-corrected bootstrap confidence interval) under different conditions, Type I error rate, power, accuracy and confidence interval width are discussed in these two sets of simulations.

The empirical demonstration is presented to show the application of the proposed and recommended inferential approaches in a practical scenario, using real data from a local utility company. In this study, a four-predictor regression model is used to predict

customer payment amount. Several common measures of importance as well as general Dominance Analysis and its statistical inference results are presented.

4.1 Simulations

4.1.1 Normal Distribution Conditions

Type I Error rate

There are three cases in the simulation (D_{43} in Matrix 1, D_{43} in Matrix 2, and D_{42} in Matrix 4) where the null hypothesis is true. The rejection rates were averaged across the three null cases. The results are shown in Table 4.1 and Figure 4.1 for normally distributed data with different sample sizes.

Table 4.1 Type I error rate with different sample sizes

Procedure	N = 300	N = 100	N = 30
Asymptotic	0.05	0.01	0.05
Standardized_B	0.05	0.02	0.05
Percentile_B	0.05	0.02	0.05
BiasCorrected_B	0.05	0.03	0.09

In Figure 4.1, the line represents 0.05, the expected Type I error rate. The averaged type I error rate is considered to be acceptable if it is close to 0.05. It can be seen that the Type I error rates are all acceptable for each of the four confidence interval approaches when

sample size equals 30 or 300, with the exception of an overly high Type I error rate for bias-corrected bootstrap confidence interval when sample size was 30. The Type I error rate seems too low across all procedures when sample size was 100.

The details within each sample size are shown in Figure 4.2 – 4.4, where the acceptable range, $[0.025, 0.075]$, is shaded. When sample size = 300, all the false rejection rates are quite close to the specified level (i.e., 0.05). The false rejection rates of asymptotic confidence interval, standardized and percentile bootstrap confidence interval run too conservative when sample size = 100, but within the acceptable range when sample size = 30. The false rejection rate for the bias-corrected bootstrap approach is always greater than that for other three methods, which is acceptable when sample size is 100, but greater than the upper limit of the acceptable range when the sample size is 30.

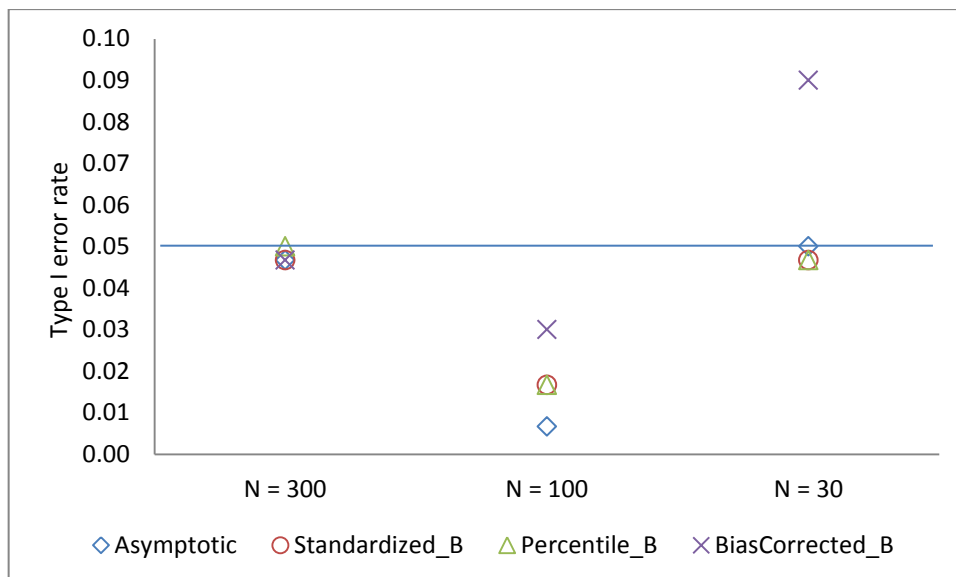


Figure 4.1 Type I error rate with different sample size

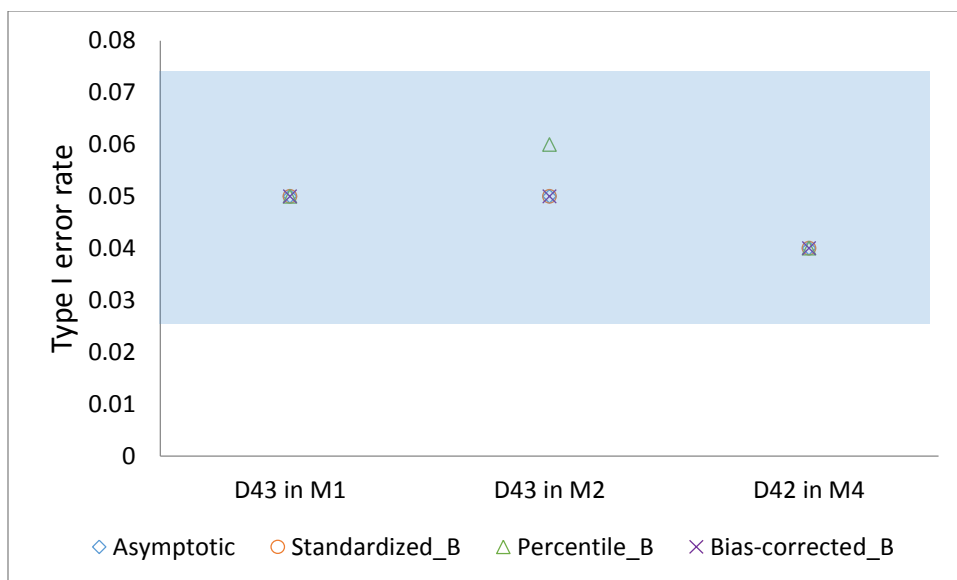


Figure 4.2 Type I error rate and effect size when sample size = 300

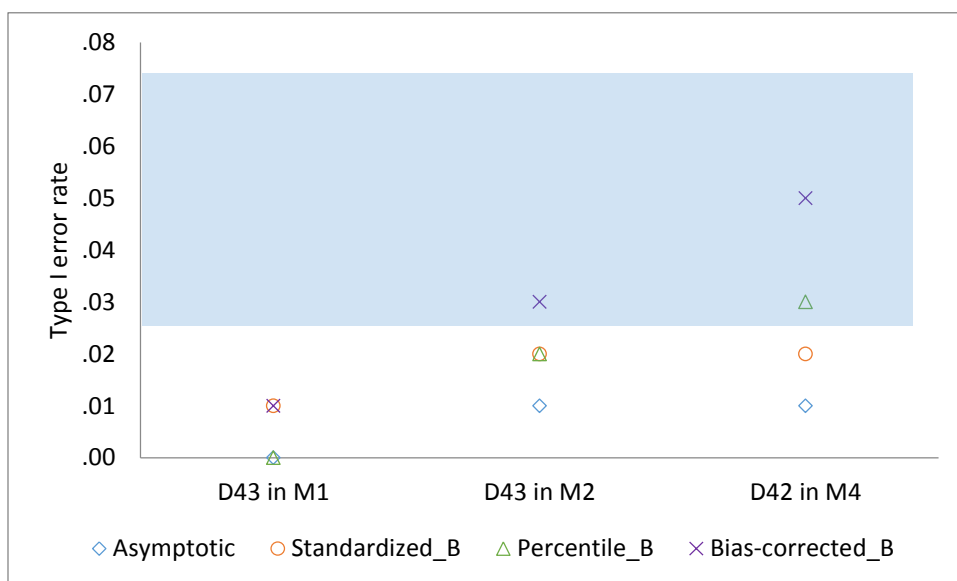


Figure 4.3 Type I error rate and effect size when sample size = 100

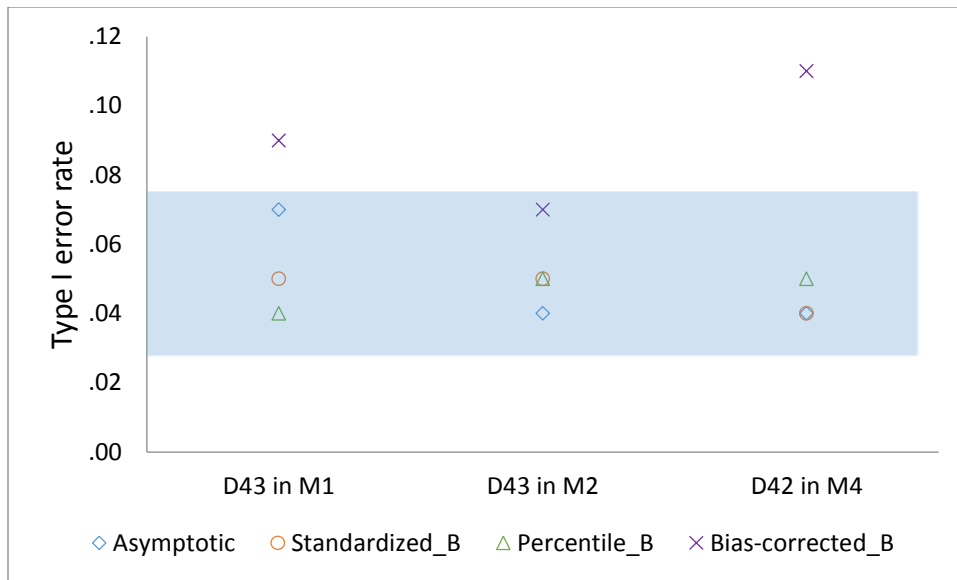


Figure 4.4 Type I error rate and effect size when sample size = 30

Power

The rejection rates were averaged across all 45 non-null cases. The results are shown in Table 4.2 and Figure 4.5 for normal distributions with different sample sizes. It can be seen that the averaged power rates are almost the same across the four confidence interval approaches when sample size is 100 or 300. The bias-corrected bootstrap confidence interval seems a little higher than the other three when sample size is 30. Power did not reach acceptable rates with any of the procedures for sample sizes of 30 or 100, but did with all procedures when sample size was 300.

Table 4.2 Power for different sample sizes

Power	N = 300	N = 100	N = 30
Asymptotic	0.80	0.69	0.43
Standardized_B	0.80	0.68	0.39
Percentile_B	0.80	0.69	0.40
BiasCorrected_B	0.81	0.71	0.48

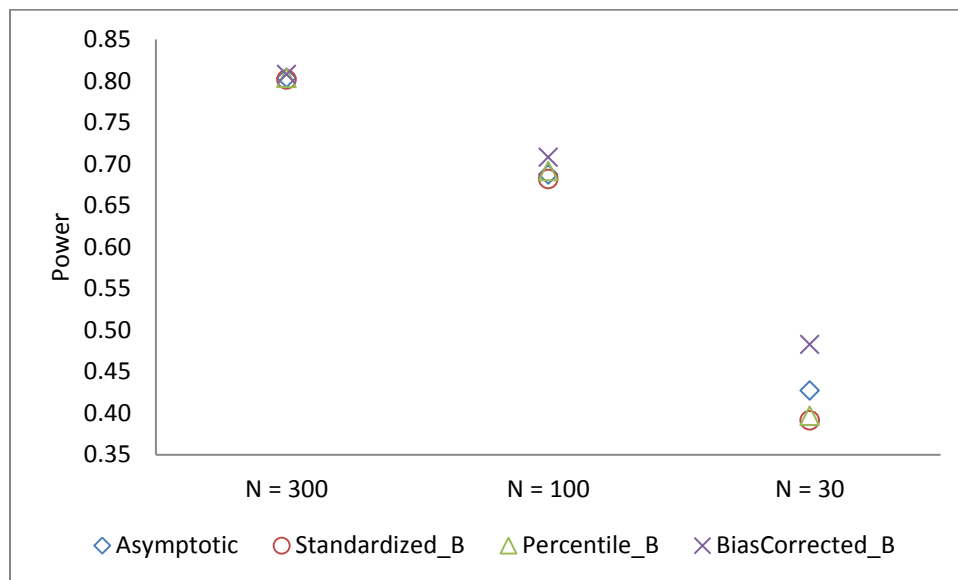


Figure 4.5 Power for different sample sizes

By further examining the relationship between power and effect size within each sample size, the required sample size and effect size for the proposed procedure to obtain an adequate power level (i.e., 0.8) could be determined, as highlighted in Figures 4.6 - 4.8. For a sample size of 300, the power is above .8 when the absolute value of general dominance difference is above 0.05 for all four confidence interval approaches. For a

sample size of 100, the power reaches .8 when the absolute value of general dominance difference is above 0.15. When sample size is 30, only the bias-corrected bootstrap confidence interval and percentile bootstrap confidence interval reached the desired power level of 0.8, and only when the general dominance difference was above 0.35.

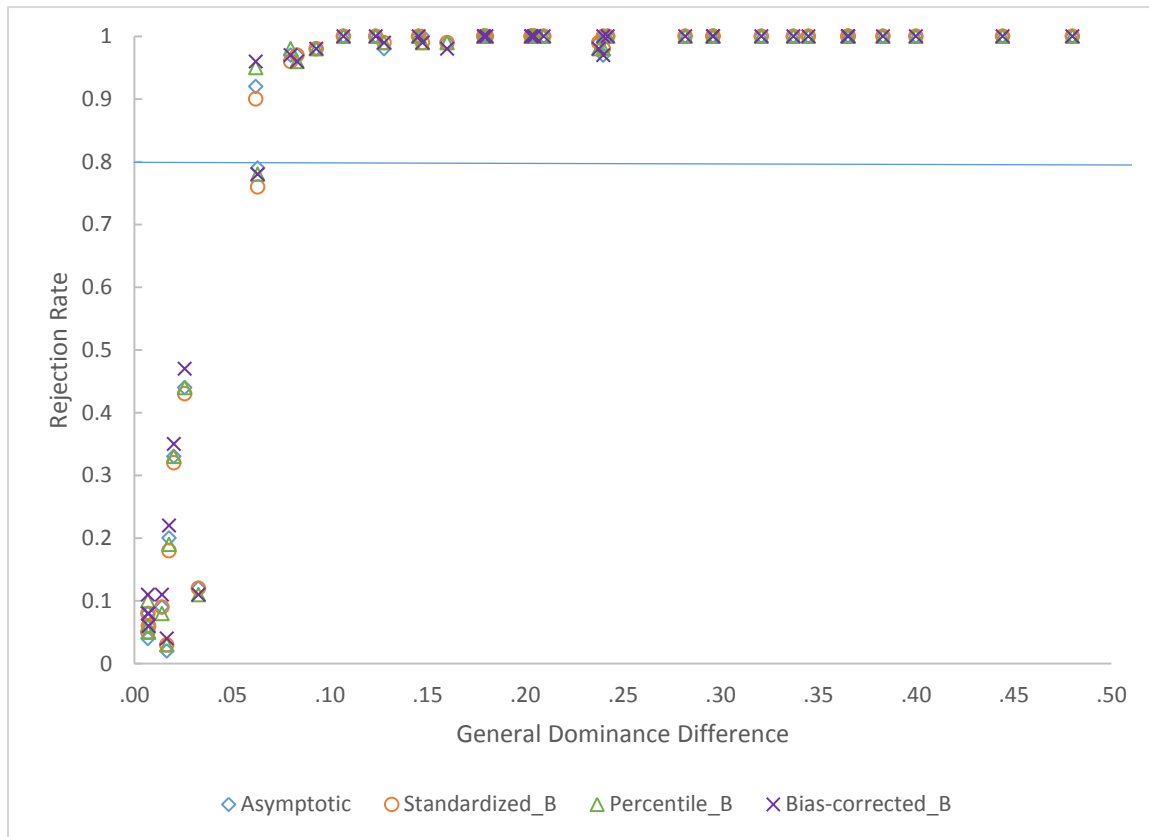


Figure 4.6 Power and effect size when sample size = 300

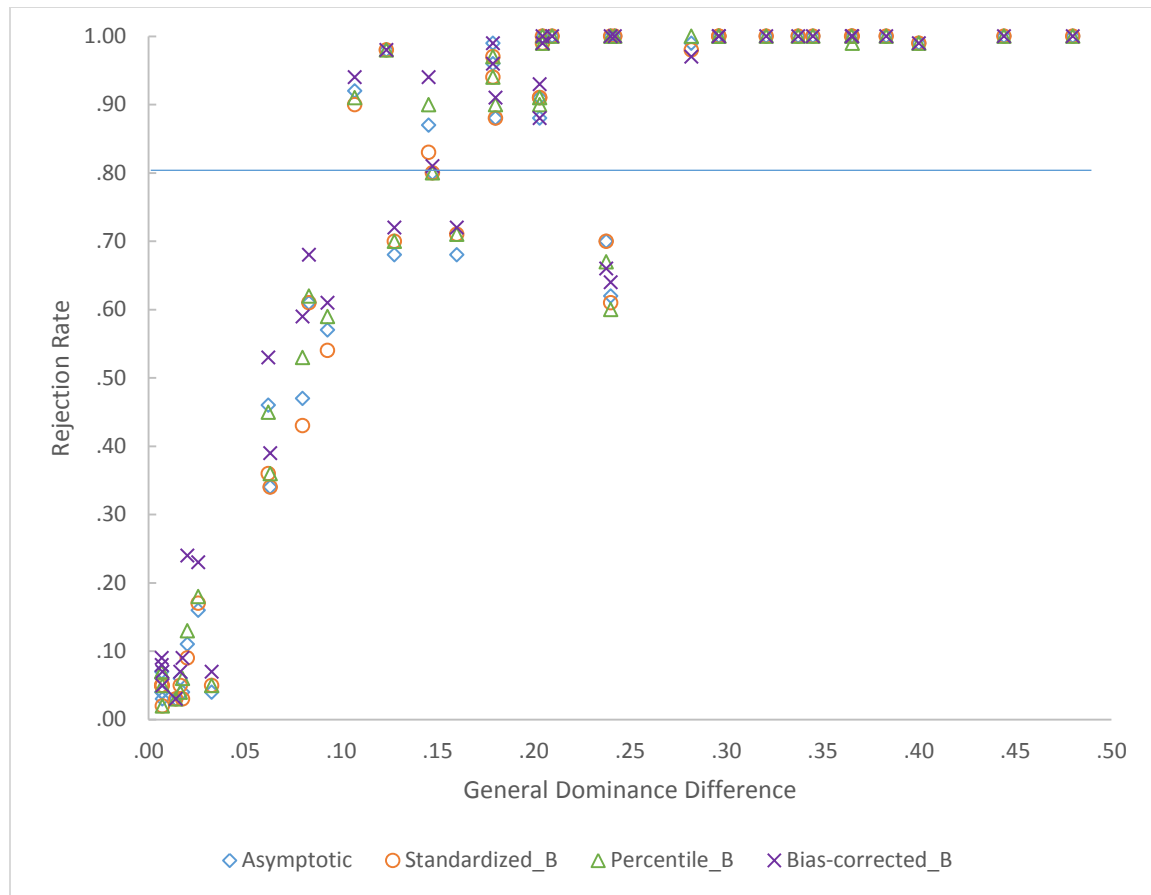


Figure 4.7 Power and effect size when sample size = 100

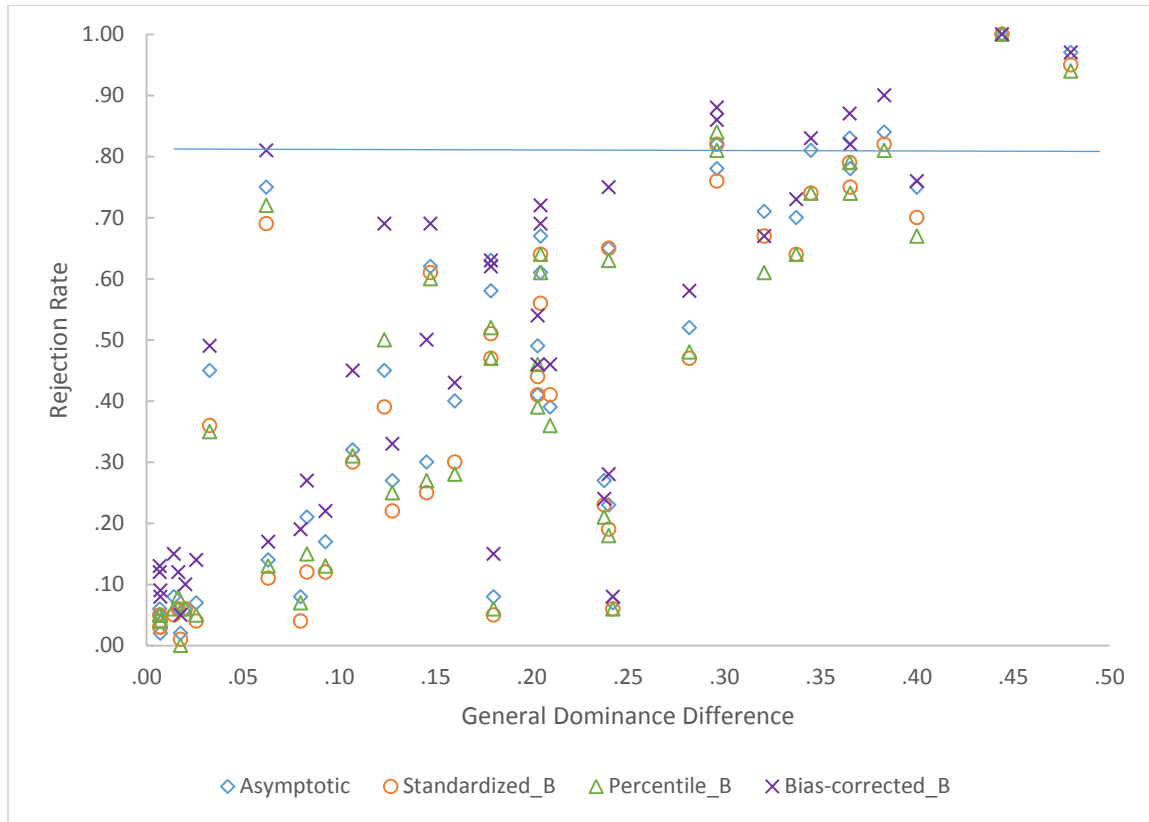


Figure 4.8 Power and effect size when sample size = 30

Accuracy

For each effect size (general dominance difference), one accuracy rate was computed by the proportion of the 100 replications in which the 95% confidence interval included the population value of the general dominance difference. Table 4.3 and Figure 4.9 show the averaged accuracy rates across all 48 effect sizes. For example, when sample size equals 300, the accuracy rate of percentile bootstrap confidence intervals is averaged to 0.93 across all of the effect sizes. For the procedure to be adequate, the averaged accuracy is expected to be close to 0.95, which is indicated by the line in Figure 4.9. All the results

are quite close to the nominal level (e.g., 0.95) except for the bias-corrected bootstrap confidence interval when sample size is 30.

Table 4.3 Averaged accuracy with different sample sizes

Averaged Accuracy	N = 300	N = 100	N = 30
Asymptotic	0.94	0.96	0.93
Standardized_B	0.93	0.96	0.94
Percentile_B	0.93	0.96	0.94
BiasCorrected_B	0.93	0.95	0.91

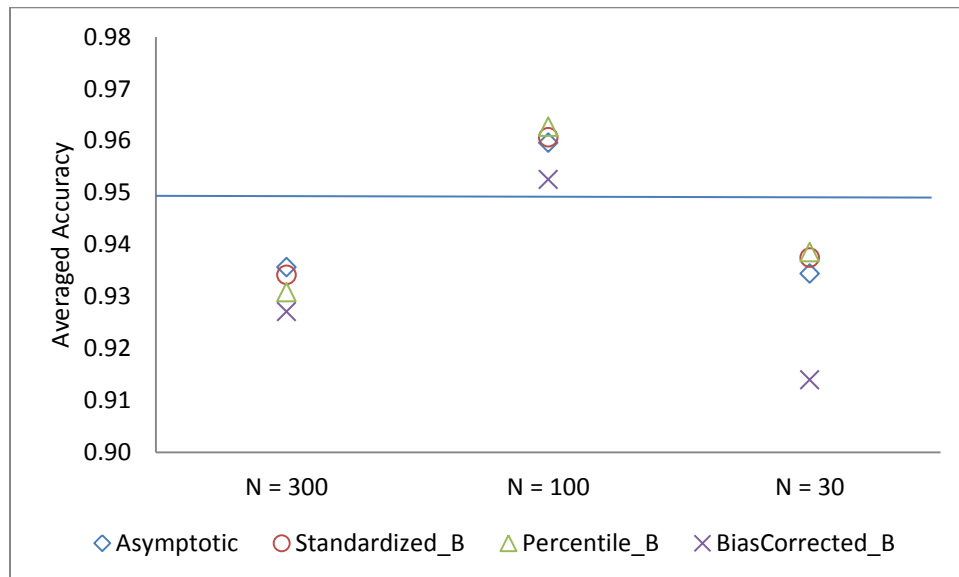


Figure 4.9 Averaged accuracy with different sample sizes

From the details shown in Figures 4.10 - 4.12, it can be concluded that the bias-corrected bootstrap confidence interval underperform the other three approaches when sample size is very small (e.g., $N = 30$). The acceptable range, $[\text{.925}, \text{.975}]$, is shaded in Figures 4.10 - 4.12.

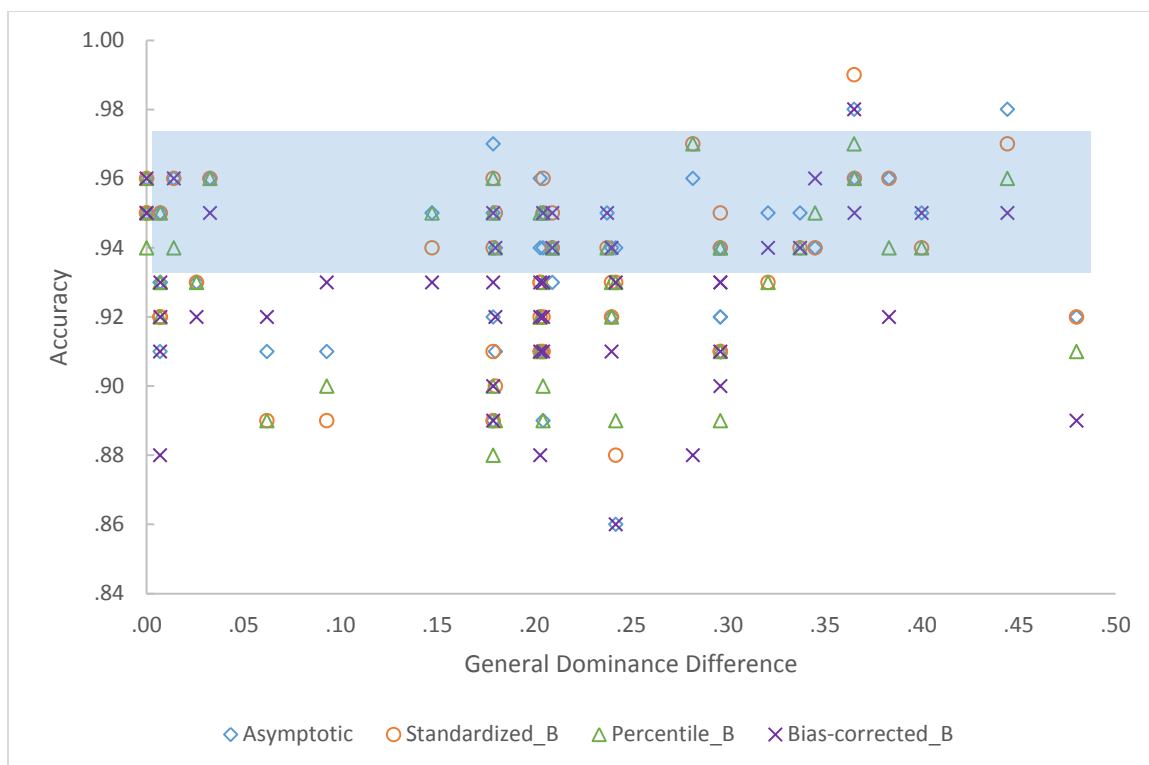


Figure 4.10 Accuracy and effect size when sample size = 300

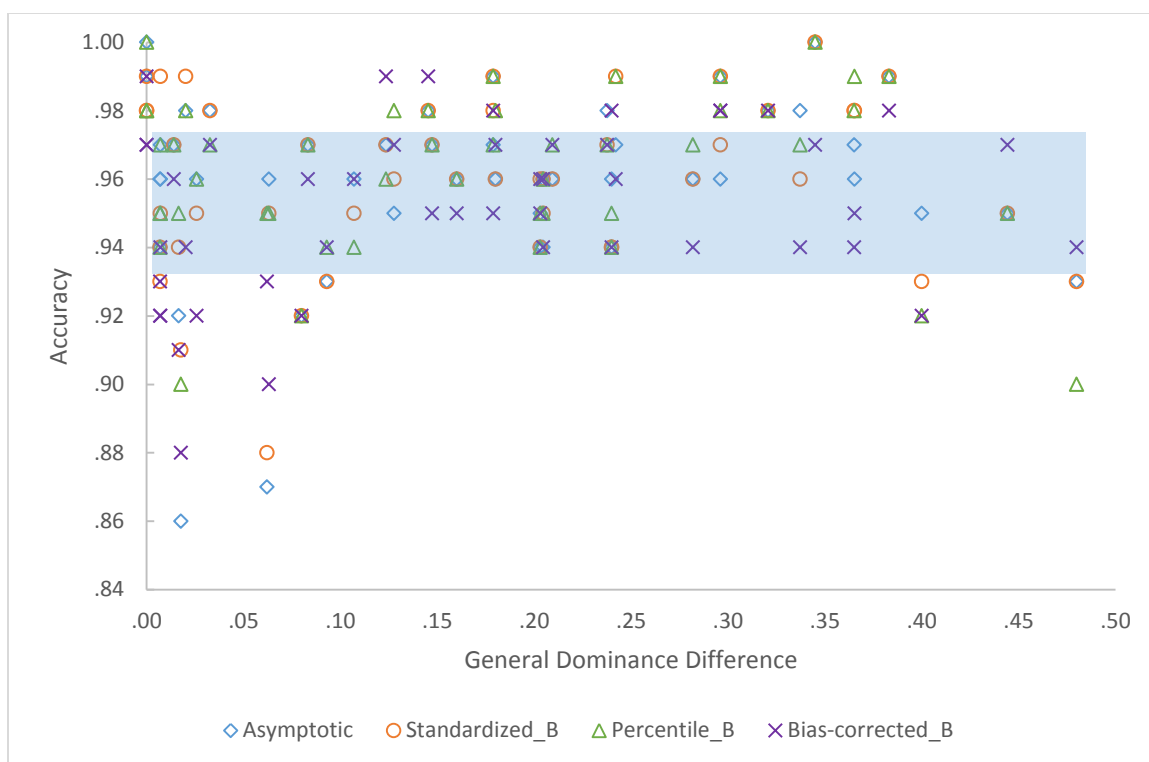


Figure 4.11 Accuracy and effect size when sample size = 100

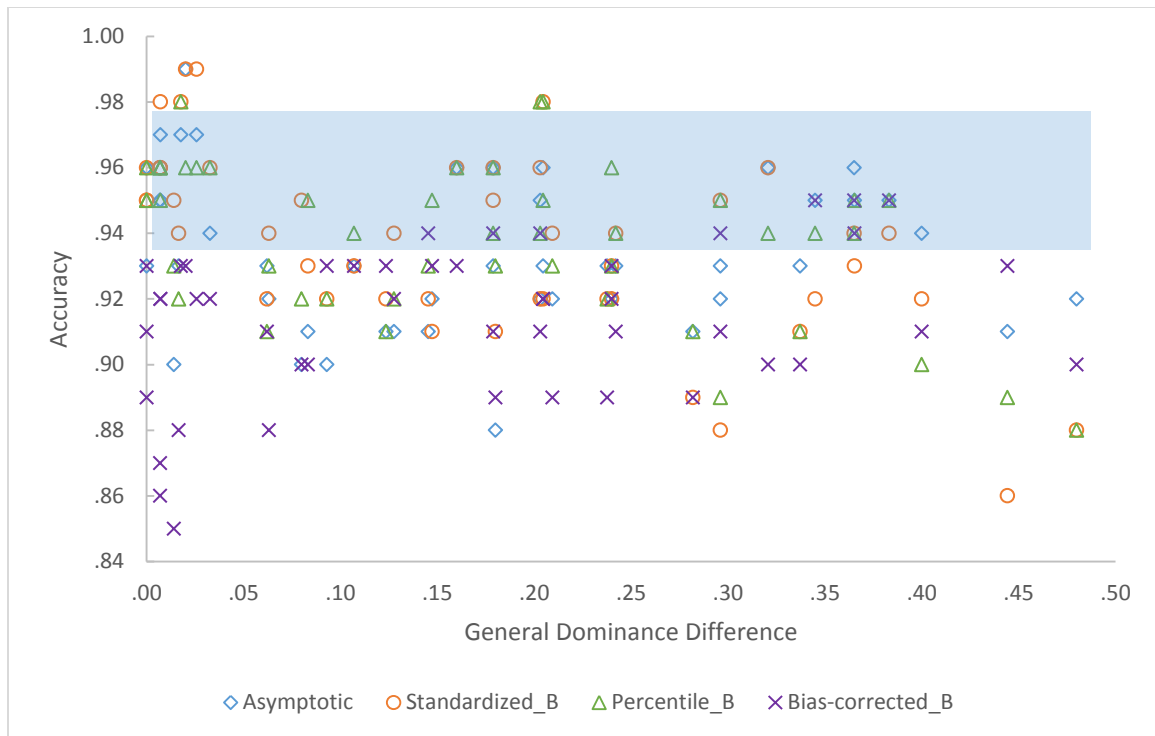


Figure 4.12 Accuracy and effect size when sample size = 30

Confidence interval width

Table 4.4 and Figure 4.13 show the averaged confidence interval width across all 48 effect sizes and Figure 4.14 – 4.16 show the details within each sample size. As expected, the confidence interval widths increase as sample size decreases. It could also be seen that the confidence interval widths are almost identical across the four confidence interval approaches within three sample sizes.

Table 4.4 Averaged confidence interval widths with different sample sizes.

	N = 300	N = 100	N = 30
Asymptotic	0.13	0.21	0.38
Standardized_B	0.12	0.21	0.38
Percentile_B	0.12	0.21	0.38
BiasCorrected_B	0.12	0.21	0.38

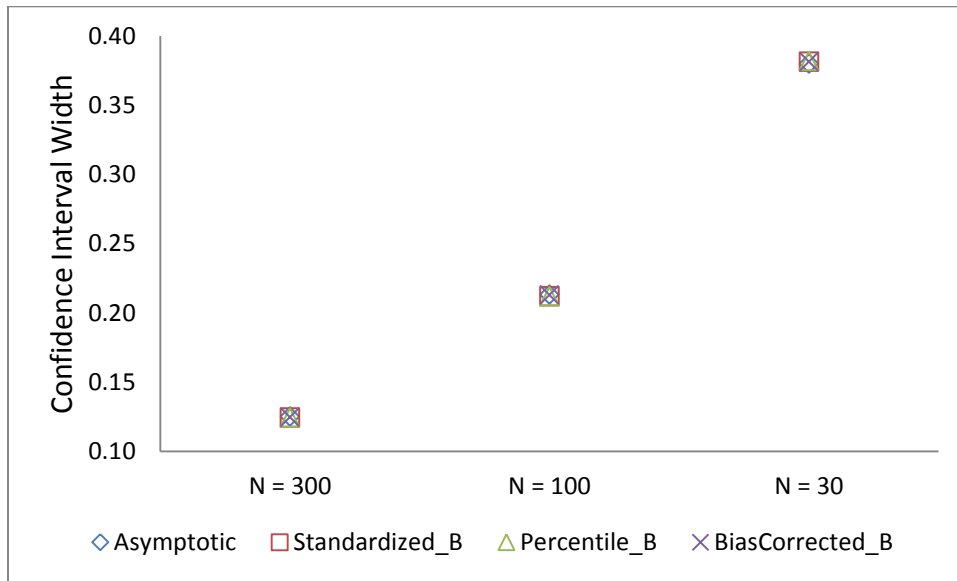


Figure 4.13 Averaged confidence interval widths with different sample sizes

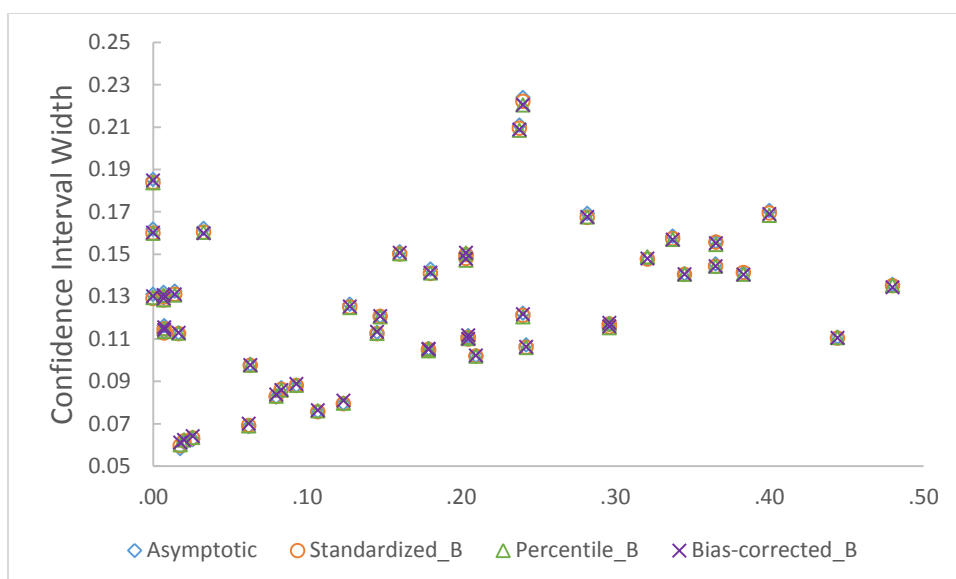


Figure 4.14 Averaged confidence interval widths when sample size = 300

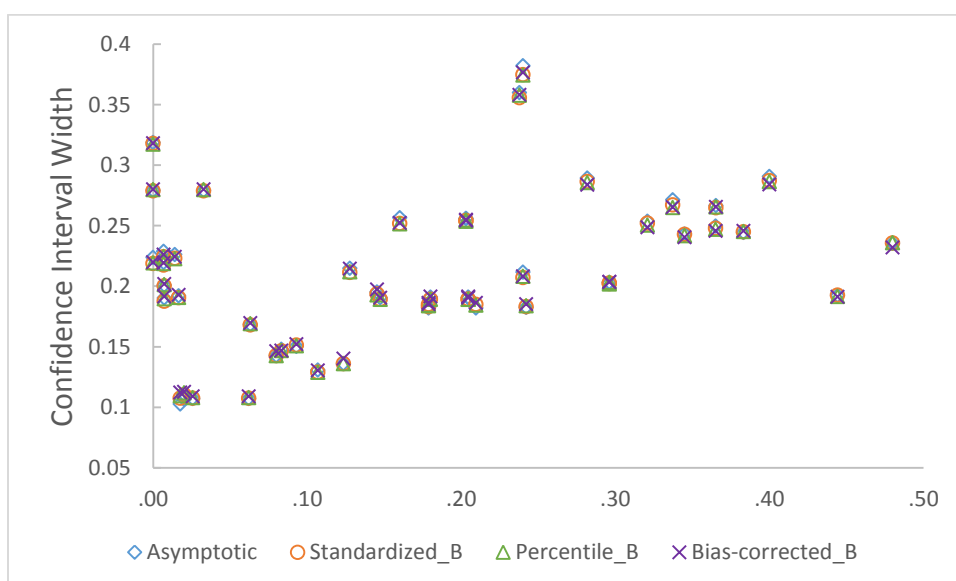


Figure 4.15 Averaged confidence interval widths when sample size = 100

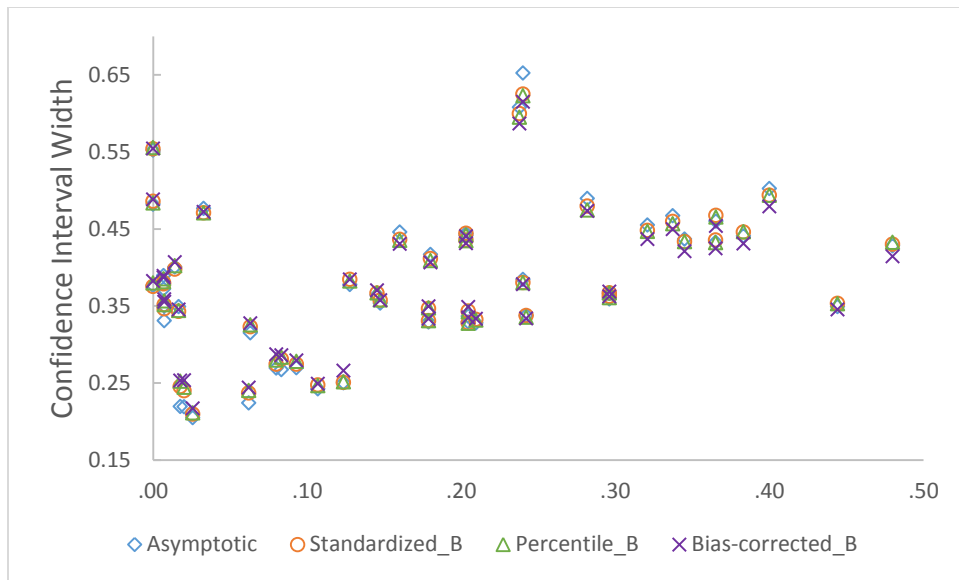


Figure 4.16 Averaged confidence interval widths when sample size = 30

Standardized Bias

From Figure 4.17, it can be seen that the standardized bias of general dominance difference is consistently controlled within $[-0.2, 0.2]$ when the sample size is 100 and 300, but relatively larger when sample size is very small (e.g., $N = 30$) or when the effect size is large (e.g., $>.40$).

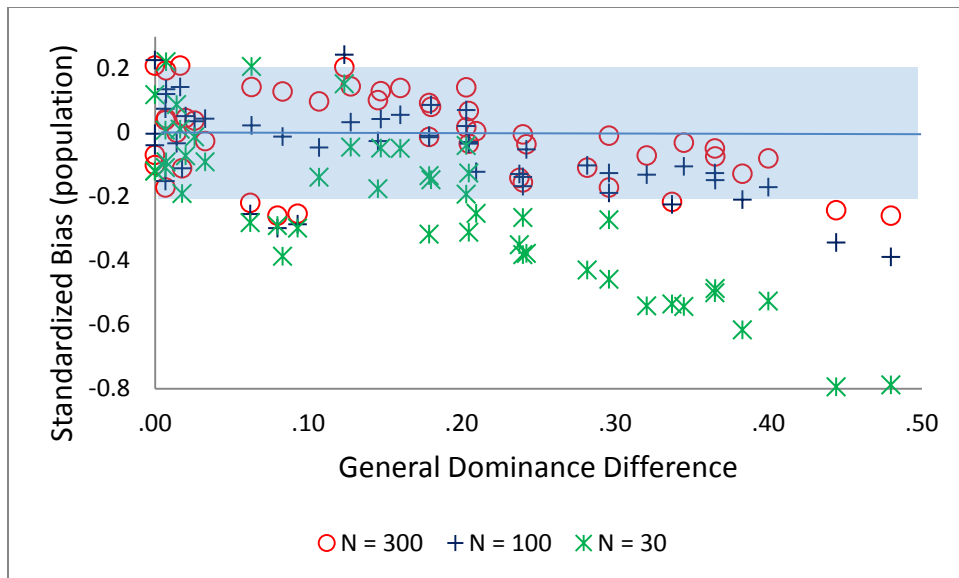


Figure 4.17 Bias of bootstrap estimates to population parameters with different sample sizes

As shown in Figure 4.18, almost all of the standardized bias fell within $[-0.2, 0.2]$ except a few outliers when sample size is 30.

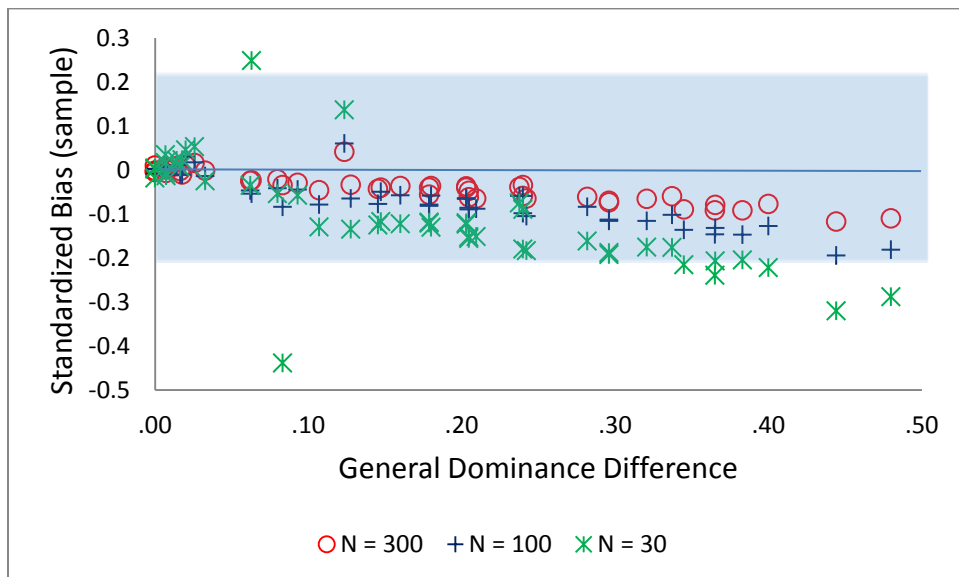


Figure 4.18 Bias of bootstrap estimates to sample parameters with different sample sizes

For reference, the standardized bias between the parent sample estimates and population parameters are also presented, shown in Figure 4.19.

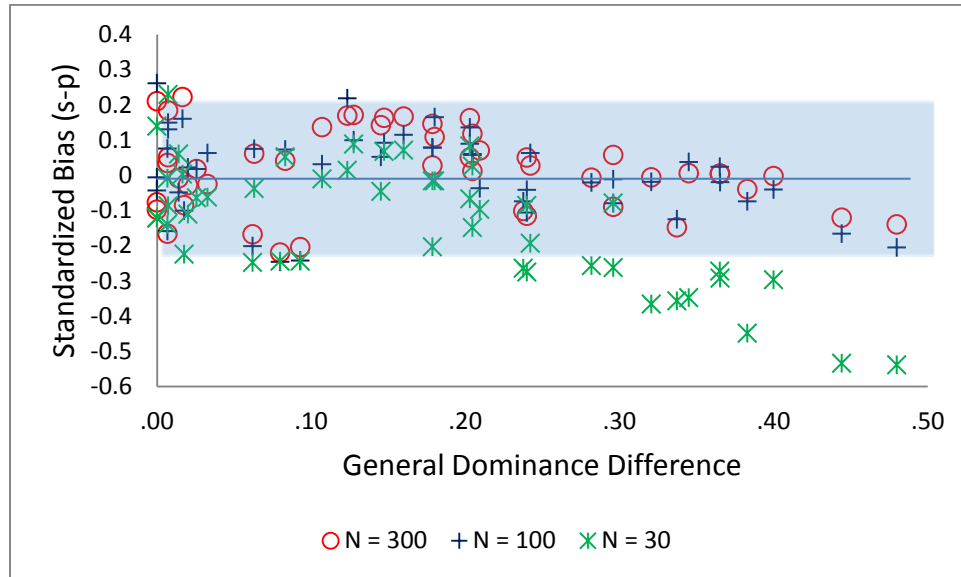


Figure 4.19 Standardized bias of sample to population parameters with different sample sizes

4.1.2 Non-normal: Skewed distribution conditions

Type I Error rate

The averaged Type I Error rates across three null cases are shown in Table 4.5 and Figure 4.20 (when sample size = 300) and Table 4.6 and Figure 4.21 (when sample size = 100) for non-normal distributed data with different degrees of skew (S). In Figure 4.20 and Figure 4.21, it can be seen that the Type I error rates for the standardized and percentile bootstrap approaches are relatively close to 0.05 across a variety of skewed distributions and both sample sizes, but the type I error rates for the asymptotic confidence interval get further from 0.05 as the distribution gets more skewed and the sample size gets smaller.

The type I error rates for the bias-corrected bootstrap confidence interval consistently fell within the acceptable range (i.e., [0.025, 0.075]) when sample size was 300 but was inflated when sample size was 100.

Table 4.5 Type I error rate when sample size = 300

	$S = 0$	$S = 0.5$	$S = 0.75$	$S = 1$
Asymptotic	0.05	0.04	0.08	0.08
Standardized_B	0.05	0.03	0.06	0.03
Percentile_B	0.05	0.03	0.06	0.04
BiasCorrected_B	0.05	0.03	0.07	0.05

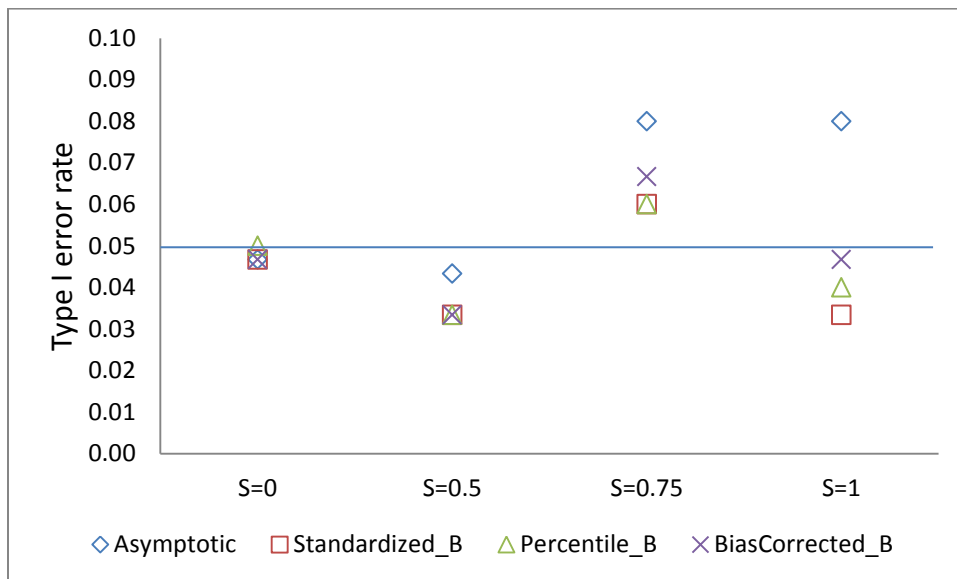


Figure 4.20 Type I error rate when sample size = 300

Table 4.6 Type I error rate when sample size = 100

	S = 0	S = 0.5	S = 0.75	S = 1
Asymptotic	0.01	0.08	0.09	0.12
Standardized_B	0.02	0.06	0.08	0.07
Percentile_B	0.02	0.07	0.09	0.07
BiasCorrected_B	0.03	0.09	0.12	0.09

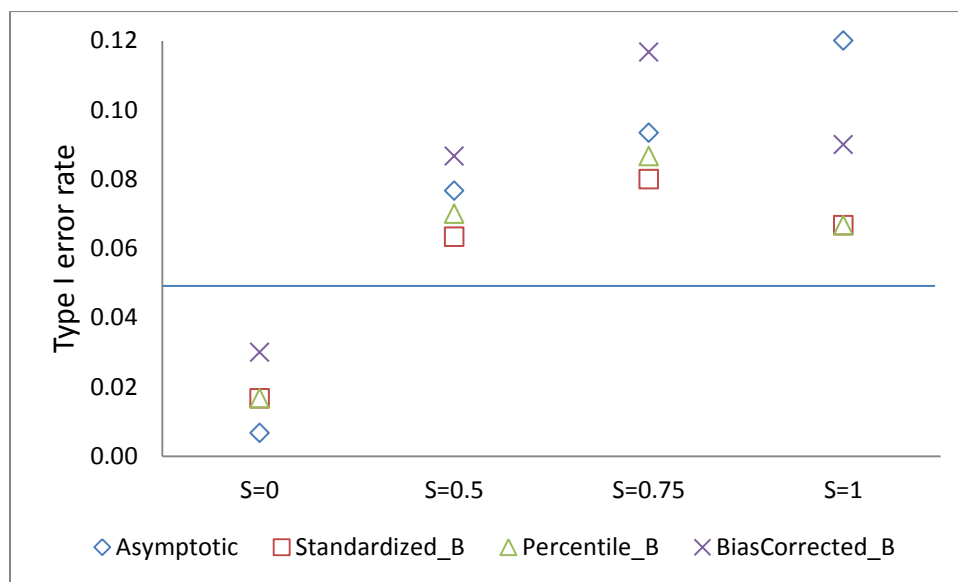


Figure 4.21 Type I error rate when sample size = 100

Power

The averaged power across all non-null cases is shown in Table 4.7 and Figure 4.22 (when sample size = 300) and Table 4.8 and Figure 4.23 (when sample size = 100) for non-normal distributed data with different degrees of skew. Specifically, the averaged power rates remain at about 0.7 when sample size is 300 across different confidence

interval methods and different levels of skew, and is below 0.65 when sample size is 100. In general, as the distribution gets more skewed, the power decreases. When sample size is smaller and the distribution is more skewed, the performance of the bias-corrected bootstrap confidence interval and the asymptotic confidence interval are slightly better than the other two, but power is still too low.

Table 4.7 Power when sample size = 300

	S = 0	S = 0.5	S = 0.75	S = 1
Asymptotic	0.70	0.70	0.70	0.69
Standardized_B	0.70	0.70	0.69	0.65
Percentile_B	0.71	0.70	0.69	0.66
BiasCorrected_B	0.71	0.70	0.70	0.67

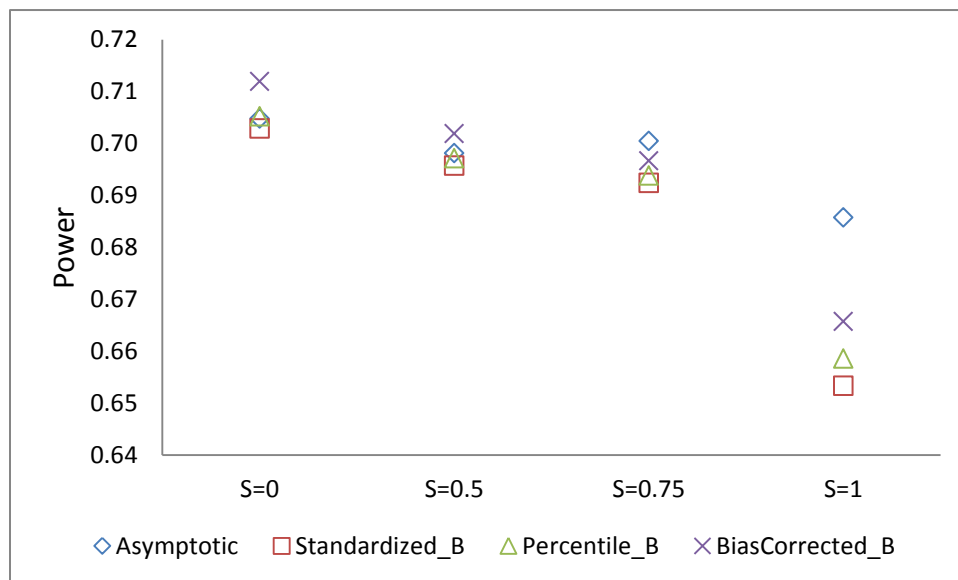


Figure 4.22 Power when sample size = 300

Table 4.8 Power when sample size = 100

	$S = 0$	$S = 0.5$	$S = 0.75$	$S = 1$
Asymptotic	0.61	0.60	0.59	0.56
Standardized_B	0.61	0.59	0.57	0.48
Percentile_B	0.62	0.60	0.57	0.51
BiasCorrected_B	0.63	0.62	0.60	0.54

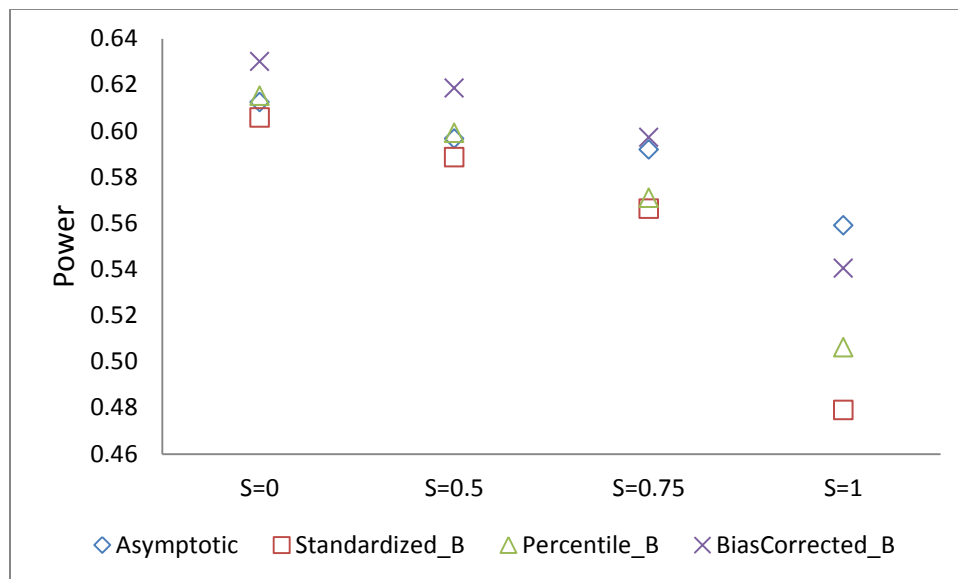


Figure 4.23 Power when sample size = 100

Accuracy

Table 4.9 and Figure 4.24 (when sample size = 300) as well as Table 4.10 and Figure 4.25 (when sample size = 100) show the averaged accuracy rates across all effect sizes for data with skewed distributions. For example, when sample size equals 300 and the

parameter of skew equals 1, an average of 95% of the percentile bootstrap confidence intervals included the true (population) dominance difference value. As presented in Figure 4.24, the averaged accuracy for the standardized, percentile, and bias-corrected bootstrap confidence interval approaches are close to 0.95 across a variety of skewed distributions when sample size = 300, but the averaged accuracy for the asymptotic confidence interval are relatively lower as the distribution gets more skewed. As the sample size gets smaller and the distribution more skewed (e.g., $N = 100$, $\text{Skew} = 1$, shown in Figure 4.25), the bias-corrected bootstrap and asymptotic confidence intervals become less accurate. Especially, the averaged accuracy drops below 0.9 for the asymptotic confidence interval when parameter of skew is 1.

Table 4.9 Averaged accuracy when sample size = 300

	S = 0	S = 0.5	S = 0.75	S = 1
Asymptotic	0.94	0.95	0.92	0.90
Standardized_B	0.93	0.95	0.94	0.95
Percentile_B	0.93	0.95	0.94	0.95
BiasCorrected_B	0.93	0.95	0.94	0.94

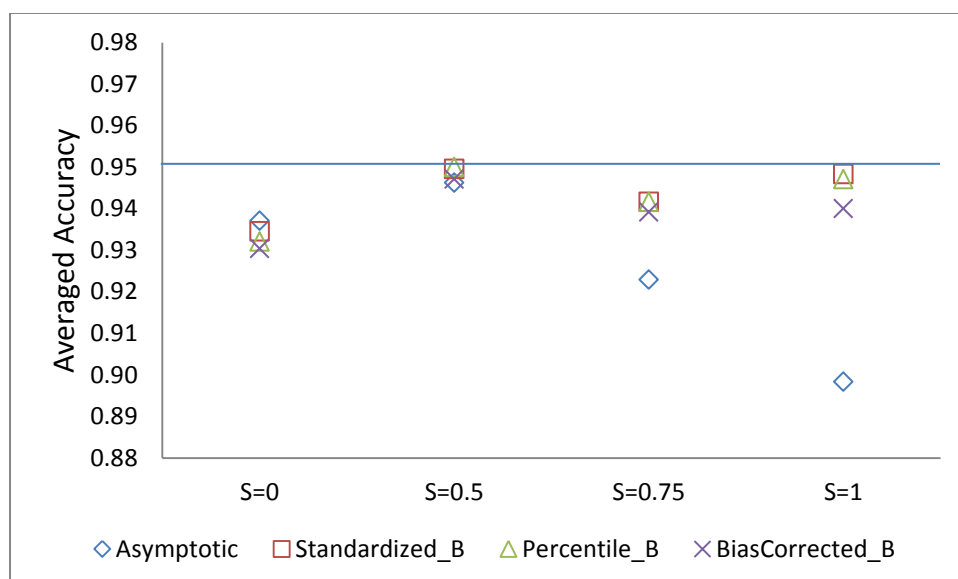


Figure 4.24 Averaged accuracy when sample size = 300

Table 4.10 Averaged accuracy when sample size = 100

	S = 0	S = 0.5	S = 0.75	S = 1
Asymptotic	0.96	0.92	0.94	0.89
Standardized_B	0.96	0.93	0.95	0.94
Percentile_B	0.96	0.93	0.95	0.94
BiasCorrected_B	0.95	0.92	0.93	0.92

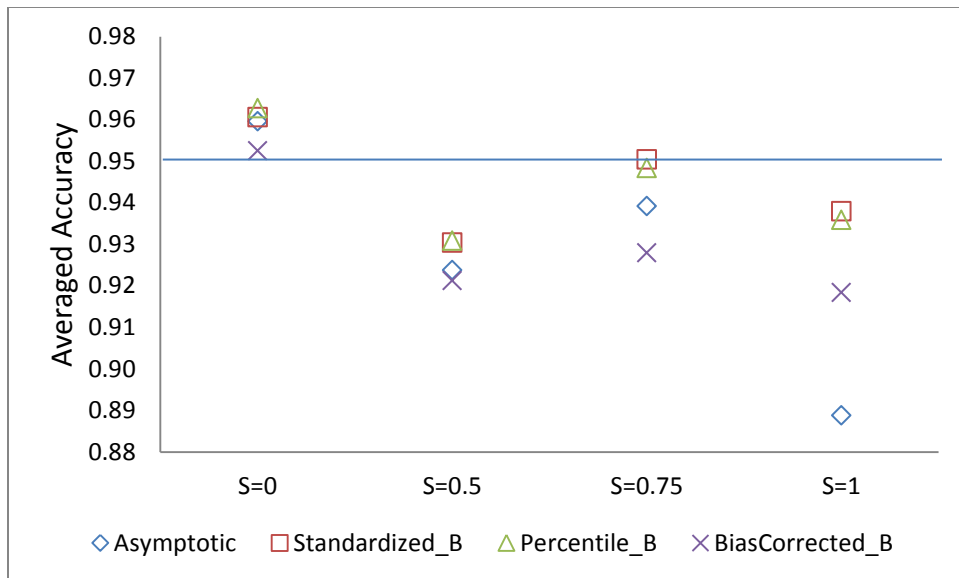


Figure 4.25 Averaged accuracy when sample size = 100

Confidence interval width

Table 4.11 and Figure 4.26 (when sample size = 300) as well as Table 4.12 and Figure 4.27 (when sample size = 100) show the averaged confidence interval widths across all effect sizes for data with skewed distributions. It can be seen from the following two figures that with the increase of skewness, the widths of the asymptotic confidence intervals stays almost the same while the widths of the three bootstrap confidence intervals go up for both of sample sizes.

Table 4.11 Averaged confidence interval widths when sample size = 300

	S = 0	S = 0.5	S = 0.75	S = 1
Asymptotic	0.12	0.12	0.12	0.13
Standardized_B	0.12	0.13	0.13	0.15
Percentile_B	0.12	0.13	0.13	0.15
BiasCorrected_B	0.12	0.13	0.13	0.16

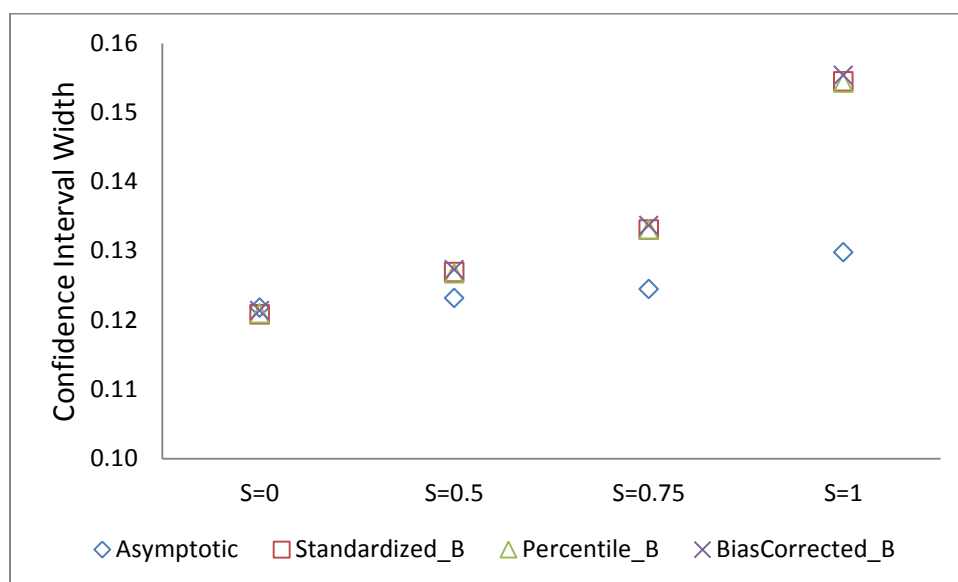


Figure 4.26 Averaged confidence interval width when sample size = 300

Table 4.12 Averaged Confidence interval widths when sample size = 100

	S = 0	S = 0.5	S = 0.75	S = 1
Asymptotic	0.21	0.21	0.21	0.22
Standardized_B	0.21	0.21	0.23	0.26
Percentile_B	0.21	0.21	0.23	0.26
BiasCorrected_B	0.21	0.21	0.23	0.26

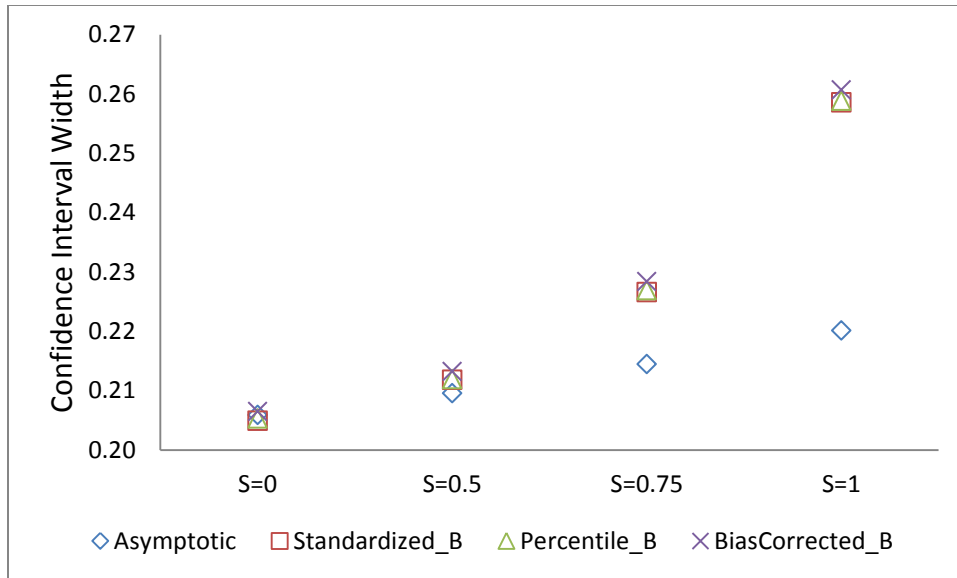


Figure 4.27 Averaged confidence interval width when sample size = 100

4.1.3 Non-normal: Leptokurtic and platykurtic distribution conditions

Type I Error rate

The averaged type I error rates across three null cases are shown in Table 4.13 and Figure 4.28 (when sample size = 300) and Table 4.14 and Figure 4.29 (when sample size = 100) for non-normal distributed data with different degrees of kurtosis. In Figure 4.28 and Figure 4.29, it can be seen that generally the Type I error rates for the standardized and percentile bootstrap confidence intervals are closer to 0.05 than the other two approaches across most of the leptokurtic and platykurtic distributions examined and for the different sample sizes. Specifically, the type I error rates for the asymptotic and bias-corrected bootstrap confidence intervals are inflated when kurtosis equals 3.

Table 4.13 Type I error rate when sample size = 100

	K = 0	K = -1.2	K = 1.2	K = 3
Asymptotic	0.05	0.06	0.06	0.09
Standardized_B	0.05	0.03	0.05	0.05
Percentile_B	0.05	0.02	0.06	0.05
BiasCorrected_B	0.05	0.04	0.06	0.08

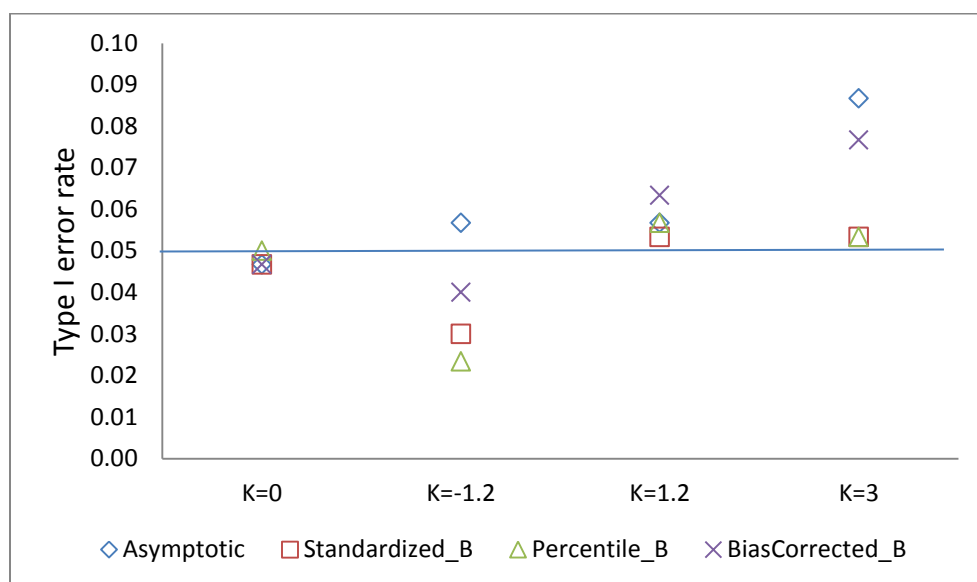


Figure 4.28 Type I error rate when sample size = 300

Table 4.14 Type I error rate when sample size = 100

	K = 0	K = -1.2	K = 1.2	K = 3
Asymptotic	0.01	0.06	0.07	0.08
Standardized_B	0.02	0.05	0.07	0.08
Percentile_B	0.02	0.05	0.07	0.08
BiasCorrected_B	0.03	0.06	0.09	0.10

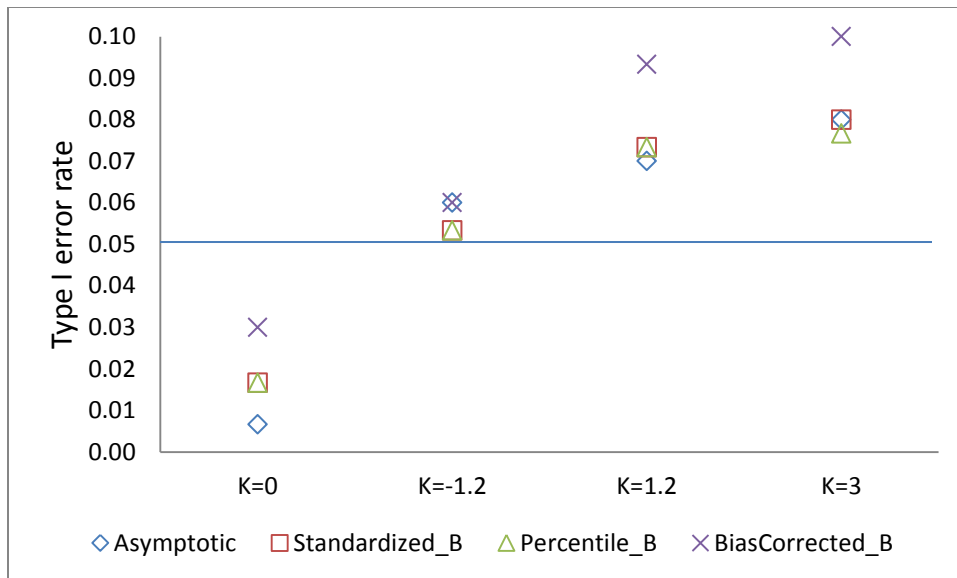


Figure 4.29 Type I error rate when sample size = 100

Power

The averaged power across all non-null cases is shown in Table 4.15 and Figure 4.30 (when sample size = 300) and Table 4.16 and Figure 4.31 (when sample size = 100) for non-normal distributed data with different degrees of kurtosis. Specifically, the averaged power rates remain at about 0.7 when sample size is 300 across different confidence interval methods and different levels of skew, and are below 0.65 when sample size is 100. When sample size is 100, the averaged power rate of bias-corrected bootstrap confidence interval is a little higher than the others.

Table 4.15 Power when sample size = 300

	K = 0	K = -1.2	K = 1.2	K = 3
Asymptotic	0.70	0.68	0.69	0.71
Standardized_B	0.70	0.67	0.69	0.70
Percentile_B	0.71	0.67	0.69	0.71
BiasCorrected_B	0.71	0.67	0.70	0.71

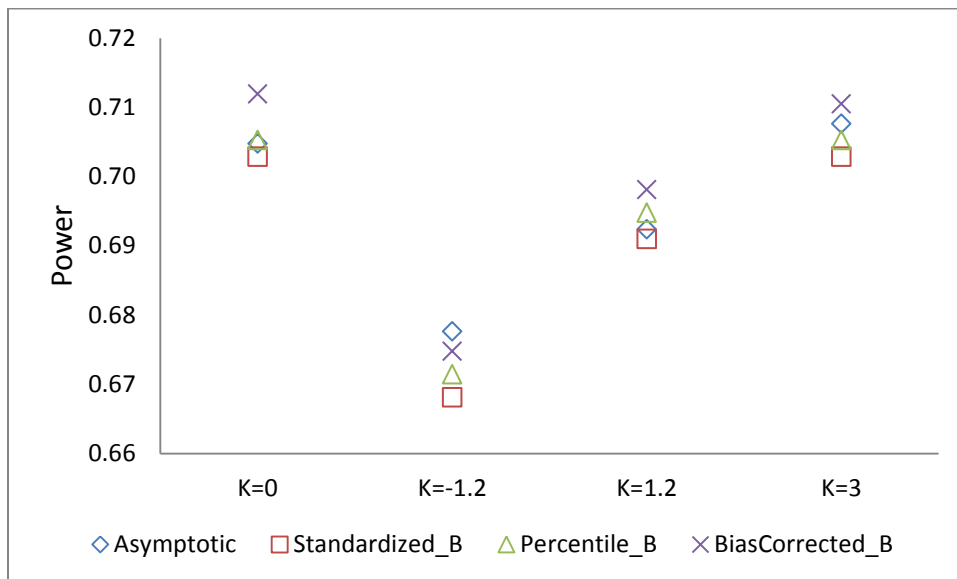


Figure 4.30 Power for when sample size = 300

Table 4.16 Power when sample size = 100

	K = 0	K = -1.2	K = 1.2	K = 3
Asymptotic	0.61	0.56	0.61	0.59
Standardized_B	0.61	0.54	0.60	0.58
Percentile_B	0.62	0.56	0.61	0.59
BiasCorrected_B	0.63	0.57	0.63	0.61

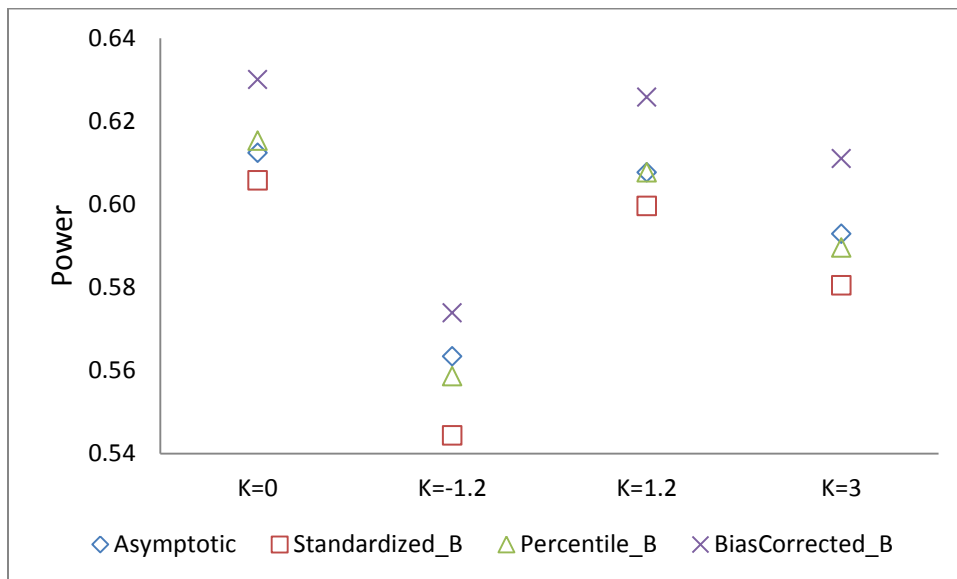


Figure 4.31 Power for when sample size = 100

Accuracy

Table 4.17 and Figure 4.32 (when sample size = 300) as well as Table 4.18 and Figure 4.33 (when sample size = 100) show the averaged accuracy rates across all effect sizes

for data with leptokurtic and platykurtic distributions. For example, when sample size equals 300 and the parameter of kurtosis equals 3, an average of 94% of the percentile bootstrap confidence intervals included the true dominance difference value. As presented in the following two figures, the averaged accuracy for the standardized and percentile bootstrap confidence interval approaches are close to 0.95 across different kurtosis parameters and sample sizes, but the averaged accuracy for asymptotic confidence interval and bias-corrected bootstrap confidence interval are relatively lower when the distribution is leptokurtic.

Table 4.17 Averaged accuracy when sample size = 300

	K = 0	K = -1.2	K = 1.2	K = 3
Asymptotic	0.94	0.95	0.93	0.92
Standardized_B	0.93	0.95	0.95	0.94
Percentile_B	0.93	0.95	0.95	0.94
BiasCorrected_B	0.93	0.94	0.95	0.93

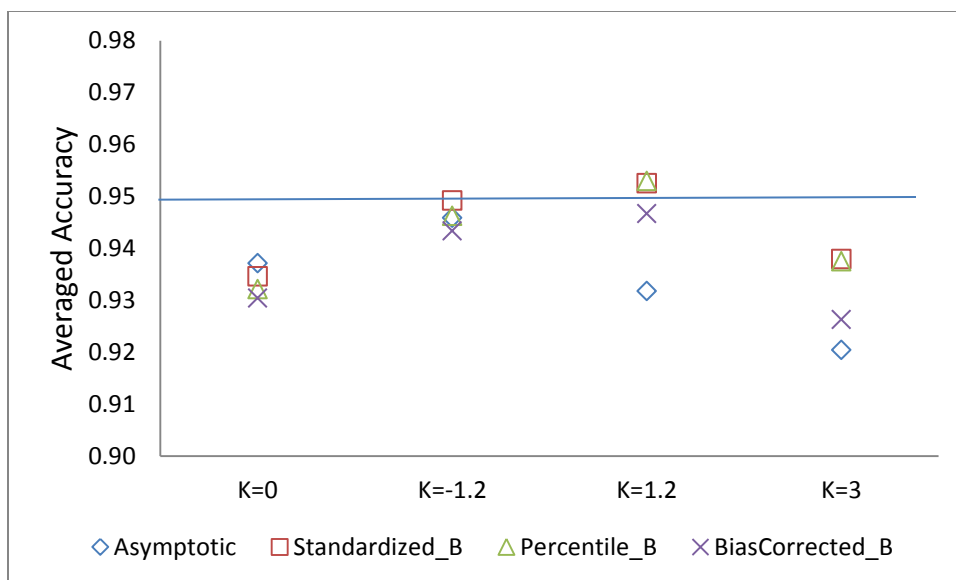


Figure 4.32 Averaged accuracy when sample size = 300

Table 4.18 Averaged accuracy when sample size = 100

	K = 0	K = -1.2	K = 1.2	K = 3
Asymptotic	0.96	0.93	0.92	0.92
Standardized_B	0.96	0.93	0.94	0.93
Percentile_B	0.96	0.93	0.94	0.94
BiasCorrected_B	0.95	0.91	0.93	0.91

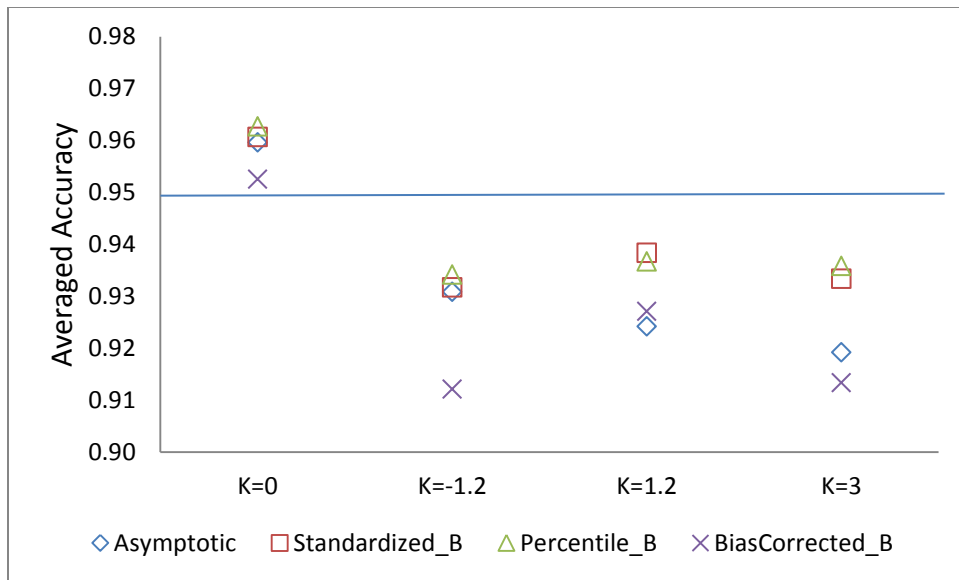


Figure 4.33 Averaged accuracy when sample size = 100

Confidence interval width

Table 4.19 and Figure 4.34 (when sample size = 300) as well as Table 4.20 and Figure 4.35 (when sample size = 100) show the averaged confidence interval widths across all effect sizes for data with leptokurtic and platykurtic distributions. It could be seen from the following two figures that when kurtosis equal 0 and 1.2, the width of all confidence interval are almost identical and stay the same, but when kurtosis equal -1.2 and 3, all the widths increase and the width of the asymptotic confidence interval is smaller than the three bootstrap confidence intervals for both of sample sizes.

Table 4.19 Averaged Confidence interval widths when sample size = 300

	K = 0	K = -1.2	K = 1.2	K = 3
Asymptotic	0.12	0.13	0.12	0.12
Standardized_B	0.12	0.14	0.12	0.13
Percentile_B	0.12	0.14	0.12	0.13
BiasCorrected_B	0.12	0.14	0.12	0.13

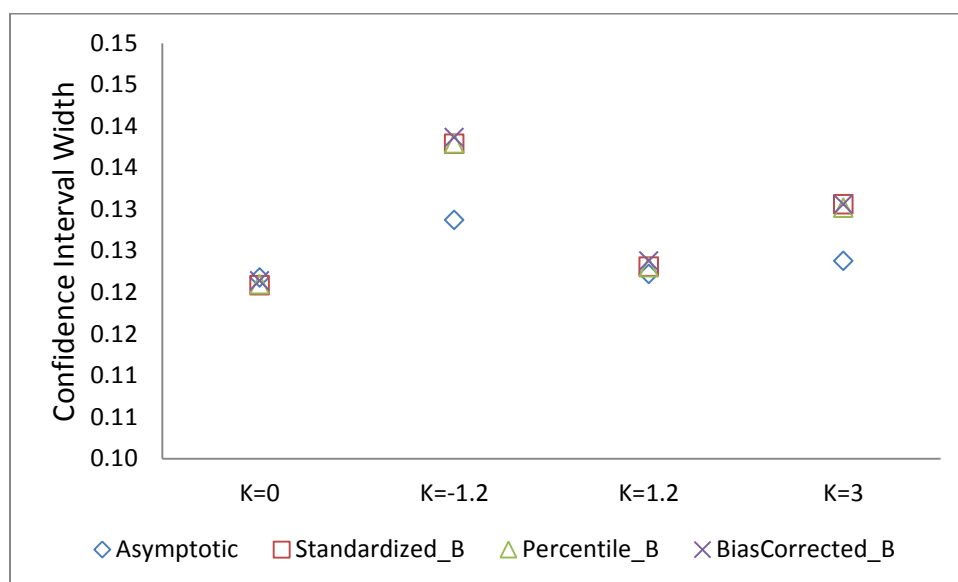


Figure 4.34 Averaged confidence interval width when sample size = 300

Table 4.20 Averaged Confidence interval widths when sample size = 300

	K = 0	K = -1.2	K = 1.2	K = 3
Asymptotic	0.21	0.22	0.21	0.21
Standardized_B	0.21	0.23	0.21	0.22
Percentile_B	0.21	0.23	0.21	0.22
BiasCorrected_B	0.21	0.23	0.21	0.22

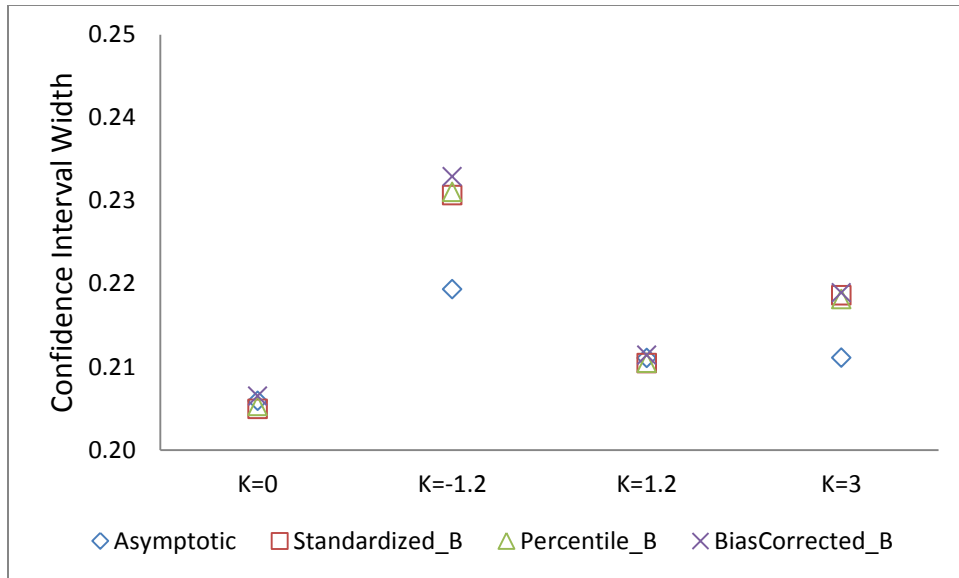


Figure 4.35 Averaged confidence interval width when sample size = 100

4.2 Empirical Example: Payment Prediction

The empirical example used the age of the account, total amount of arrears, risk score, and the age of oldest arrears as predictors of payment amount. The four-predictor model is shown in Figure 4.36 and the descriptive statistics of these variables is shown in Table 4.21. Specifically, it is interesting to know the rank ordering of these predictors and whether there is one predictor that significantly dominates the others in term of its relative importance in predicting payment amount. Such an analysis may be of interest for policy makers in utility company who want to explain the factors that influence customers' payment amount.

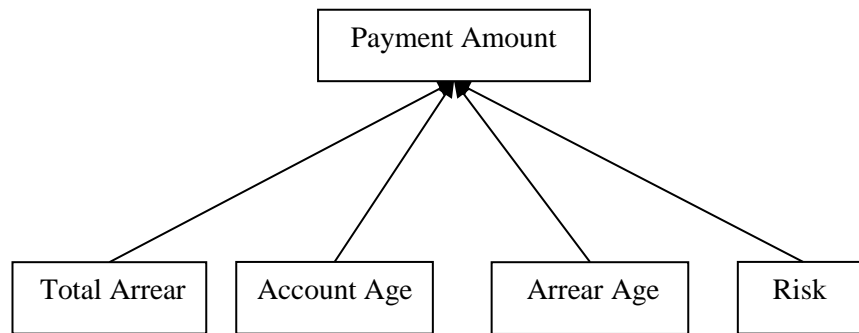


Figure 4.36 Prediction Model

Table 4.21 Descriptive statistics of the variables in empirical example

	Mean	SD	Skew	Kurtosis
X_1	17579.76	2894.28	-1.98	4.52
X_2	465.81	243.23	0.97	2.34
X_3	551.30	136.65	0.20	-0.53
X_4	97.41	49.95	0.22	-0.66
Y	67.42	110.51	1.79	3.04

The correlation matrix is shown in Table 4.22, and the values of several common measures of importance are given in Table 4.23. The various measures indicate age of oldest arrears to be the most important predictor, total amount of arrears to be the second most important predictor, risk score the third, and age of account the least important predictor. The ordering is identical across these measures of importance.

Table 4.22 The correlation matrix for the empirical cases study

	ρ_{XiY}	ρ_{XiXj}			
	Y	X ₁	X ₂	X ₃	X ₄
Y (last payment amount)	1				
X ₁ (age of account)	-.15	1			
X ₂ (total amount of arrears)	-.41	0.10	1		
X ₃ (risk score)	-.49	0.23	0.47	1	
X ₄ (age of oldest arrears)	-.54	0.26	0.51	0.64	1

Table 4.23 Several important measures for the empirical cases study

	<i>r</i>	β	Product Measure	Structure coefficient	Relative Weight	Squared semi-partial
X ₁ (age of account)	-.15	-.01	.00	-.26	.01	.00
X ₂ (total amount of arrears)	-.41	-.14	.06	-.70	.07	.01
X ₃ (risk score)	-.49	-.21	.10	-.84	.11	.02
X ₄ (age of oldest arrears)	-.54	-.34	.18	-.93	.15	.06

The Dominance Analysis results in the sample are shown in Table 4.24. Examining the sample results, it appears that according to the complete dominance measure, total amount of arrears (X_2) dominates age of account (X_1); risk score (X_3) dominates age of account (X_1) and total amount of arrears (X_2); and age of oldest arrears (X_4) dominates all of the other predictors. Thus, the overall ordering is age of oldest arrears (X_4), risk score (X_3), total amount of arrears (X_2), and age of account (X_1). The conditional dominance and general dominance results are identical.

Table 4.24 Dominance Analysis for the empirical case study

Subset model (\mathbf{x})	$\rho^2_{\mathbf{XY}}$	Additional Contribution			
		X_1	X_2	X_3	X_4
$k = 0$ average	.00	.02	.17	.24	.30
X_1	.02		.16	.22	.27
X_2	.17	.01		.11	.15
X_3	.24	.00	.04		.09
X_4	.30	.00	.03	.04	
$k = 1$ average		.00	.08	.12	.17
$X_1 X_2$.18			.10	.14
$X_1 X_3$.24		.04		.09
$X_1 X_4$.30		.03	.03	
$X_2 X_3$.28	.00			.06
$X_2 X_4$.32	.00		.02	
$X_3 X_4$.33	.00	.01		
$k = 2$ average		.00	.03	.05	.10
$X_1 X_2 X_3$.29				.06
$X_1 X_2 X_4$.32			.02	
$X_1 X_3 X_4$.33		.01		
$X_2 X_3 X_4$.34	.00			
$k = 3$ average		.00	.01	.02	.06
$X_1 X_2 X_3 X_4$.35				

Subset model (\mathbf{x})	$\rho^2_{\mathbf{XY}}$	Additional Contribution			
		X_1	X_2	X_3	X_4
Overall average		.01	.07	.11	.16

The proposed four inferential approaches (asymptotic confidence interval, standardized bootstrap confidence interval, percentile bootstrap confidence interval and bias-corrected bootstrap confidence interval) were applied to the empirical data to determine whether the difference between each pair of general dominance measures is significant. The results are shown in Table 4.25. Although the center and width of the four confidence intervals are slightly different, the significance results are identical. Actually, since the data is non-normal distributed (e.g., X_1 , X_2 and Y) and the sample size is greater than 300, the standardized bootstrap confidence interval is most recommended to make inference about the dominance relationships.

It could be concluded that total amount of arrears (X_2) significantly dominates age of account (X_1); risk score (X_3) significantly dominates X_1 (age of account); and age of oldest arrears (X_4) significantly dominates total amount of arrears (X_2) and age of account (X_1). However, the dominance relationships between X_4 and X_3 as well as X_3 and X_2 are not significant. This provides a detailed picture of the prediction model beyond the traditional Dominance Analysis. For example, by comparing the general dominance measures, the traditional Dominance Analysis finds that age of oldest arrears (X_4) dominates all of the other predictors. However, the inferential examination tells us that

the difference between X_4 and X_3 is not statistically significant. In other words, X_4 did not significantly dominate X_3 , though it does significantly dominate X_1 and X_2 .

Table 4.25 Statistical significance test of general dominance measures

	Sample	Asymptotic	Standardized	Percentile	Bias-corrected
	Mean	CI	Bootstrap CI	Bootstrap CI	Bootstrap CI
D41	0.15	[0.10, 0.20]	[0.11, 0.18]	[0.11, 0.18]	[0.12, 0.19]
D42	0.08	[0.03, 0.14]	[0.04, 0.12]	[0.05, 0.12]	[0.05, 0.13]
D43	0.05	[-0.01, 0.10]	[-0.00, 0.09]	[-0.01, 0.09]	[-0.01, 0.09]
D31	0.10	[0.06, 0.14]	[0.07, 0.14]	[0.07, 0.14]	[0.07, 0.14]
D32	0.04	[-0.01, 0.09]	[-0.01, 0.08]	[-0.01, 0.08]	[-0.01, 0.08]
D21	0.06	[0.03, 0.10]	[0.03, 0.09]	[0.04, 0.09]	[0.04, 0.09]

Chapter 5. Conclusions and Discussion

5.1 Summary

When a multiple regression model is fit, researchers are often interested in which predictors matter more. Chapter 2 reviewed and compared several predictor importance measures that are commonly utilized by researchers. Among these statistical methods, Dominance Analysis is one of the most comprehensive and intuitive. The main research question of the current study was whether there is a feasible way to make statistical inference for general dominance measures, and thereby compare the importance of predictors in a multiple regression model.

A series of simulations was conducted using different underlying distribution and sample size conditions to examine and compare the performance (i.e., Type I error rate, power, accuracy, and bias) of four proposed inferential approaches: asymptotic confidence interval, standardized bootstrap confidence interval, percentile bootstrap confidence interval and bias-corrected bootstrap confidence interval. To make the presentation more concrete, an empirical example was used to demonstrate, using real data, how to utilize the inferential approaches when comparing predictor importance and how to interpret the corresponding results.

Among the four proposed inferential approaches, the asymptotic confidence interval is the least computationally intensive and the other three are all rested on the bootstrap

resampling procedures. When the underlying distributions of predictors and criteria are normal, the asymptotic confidence interval is sufficient to use, since it is the most convenient one and provides comparable results to the other three bootstrap confidence interval approaches. Specifically, when the alpha level is set at .05, the type I error rates of those confidence intervals are around .05 and accuracy rate is around .95. As for power, the rates of all four confidence intervals are reasonable (approximately 70% - 80%) when the effect size (i.e., absolute value of general dominance difference) is above 0.05 and the sample size is 300, and when the effect size is above 0.15 and the sample size is 100. However, none of the methods perform well when sample size is extremely small (i.e., $N = 30$). To sum up, the asymptotic confidence interval is adequate and recommended to make inference for Dominance Analysis when the effect size is no less than 0.15 and there are at least 100 multivariate-normally distributed observations, or the effect size is no less than 0.05 and there are at least 300 multivariate-normally distributed observations. For effect size less than 0.05, it is recommended to increase sample size (e.g., 500 or 1000) for a power of 0.8 or above.

Here, the type I error rates for $N = 100$ are too conservative and lower than $N = 300$ and $N = 30$. As an example, the asymptotic confidence intervals of D_{43} in Matrix 1 under different sample sizes are presented in Figures 5.1 - 5.3. From these figures, it could be seen that $N = 100$ has relatively wider interval than $N = 300$ and smaller bias than $N = 30$ (i.e., the center of interval is closer to 0). It is also be conclude from Figure 4.13 and 4.17. In general, narrow interval and great bias will result in easily excluding zero. It is

possible that this combination of wide interval and small bias lead to a tiny possibility of excluding the population value (i.e., 0).

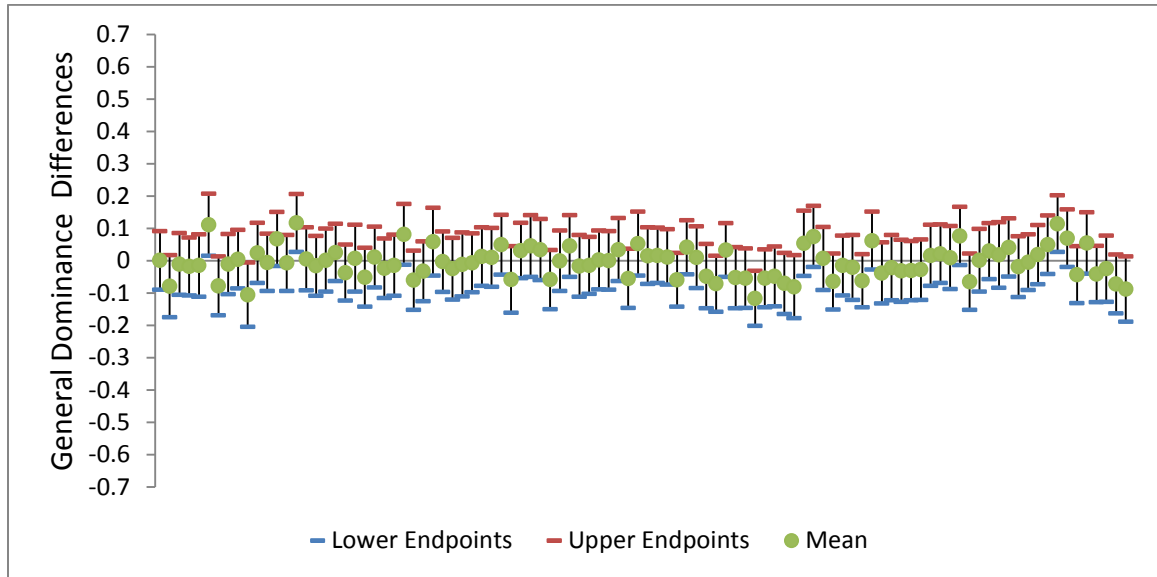


Figure 5.1 Asymptotic confidence intervals of D_{43} in Matrix 1 when $N = 300$

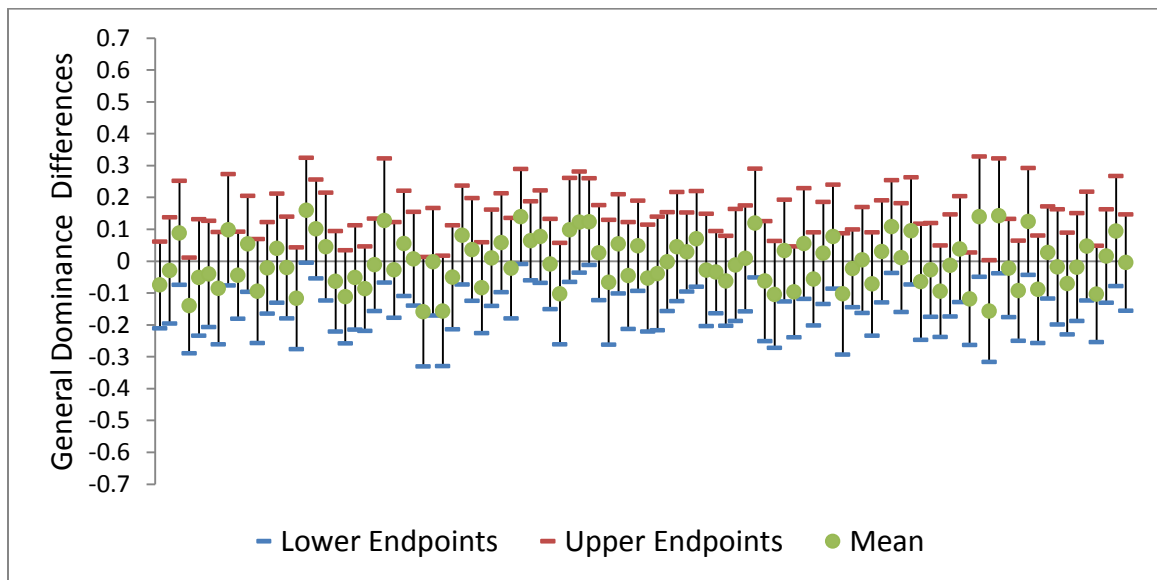


Figure 5.2 Asymptotic confidence intervals of D_{43} in Matrix 1 when $N = 100$

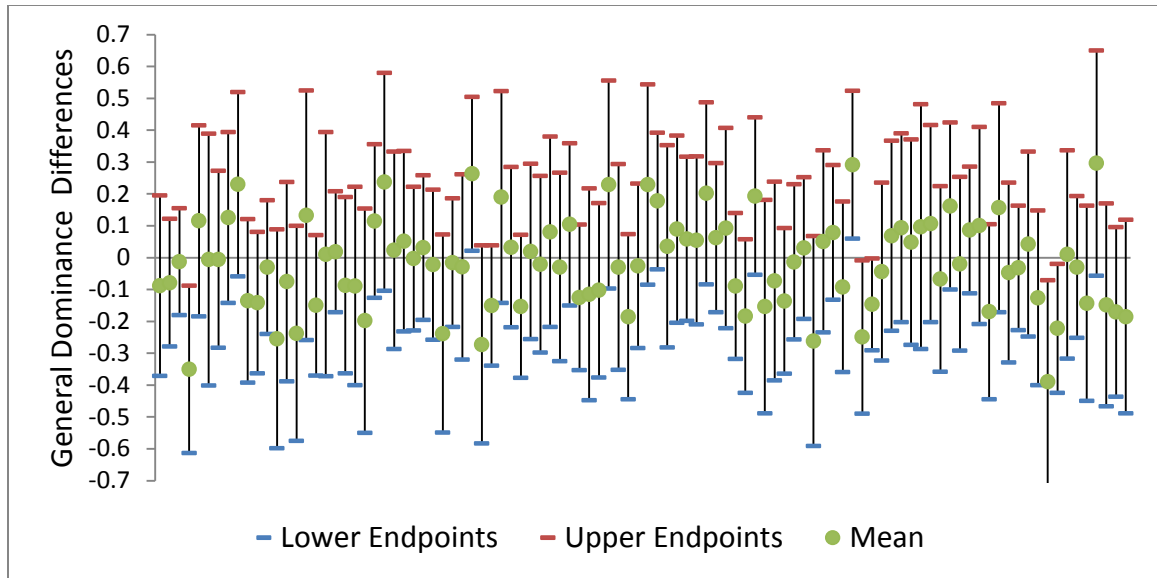


Figure 5.3 Asymptotic confidence intervals of D_{43} in Matrix 1 when $N = 30$

However, when the distributions are non-normal (e.g., skewed, leptokurtic, and platykurtic), the asymptotic method is not robust. Specifically, when the absolute magnitude of skew is equal to or greater than 0.75, or the absolute magnitude of kurtosis is greater than 1.2, the asymptotic confidence interval produced type I error rates greater than 0.08 and accuracy rates below .90. This result is expected since the asymptotic confidence interval relies on the assumption of normality. This is consistent with the results of previous studies (Maydeu-Olivares, Coffman, & Hartmann, 2007) that normal-theory intervals were found to be less accurate when the degree of skew or kurtosis was greater than 1. The reason that the asymptotic confidence interval underperforms its bootstrap counterpart may rely on the estimated variability. From Figure 4.26, 4.27, 4.34, and 4.35, the width of asymptotic confidence intervals stay the same but the widths of bootstrap confidence intervals get wider as the distribution deviates more from normality. The narrower the confidence interval, the easier it is for the interval to exclude the

population parameter, resulting in lower accuracy. For the same reason, it is easier to exclude 0, and thereby produce higher power rate than that of the bootstrap approaches. While not derived here, the evidence in the results suggests that the Normal-Theory based asymptotic confidence interval may underestimate the data variability when data is considerably non-normally distributed.

According to the simulation results, the bootstrap confidence interval approach is recommended in the non-normal scenarios. The results of the standardized bootstrap and the percentile bootstrap approaches always agreed quite closely, no matter how large the sample size was and whether the distribution was normal or non-normal. It is possible that the bootstrapping distribution of the estimate (i.e., the difference of two general dominance measures magnitude) is approximately normally distributed. Figure 1 presents the distribution of D_{43} and D_{42} in Matrix 3 in 100 bootstrap resamples when the skew = 1 and the sample size = 100 as two examples.

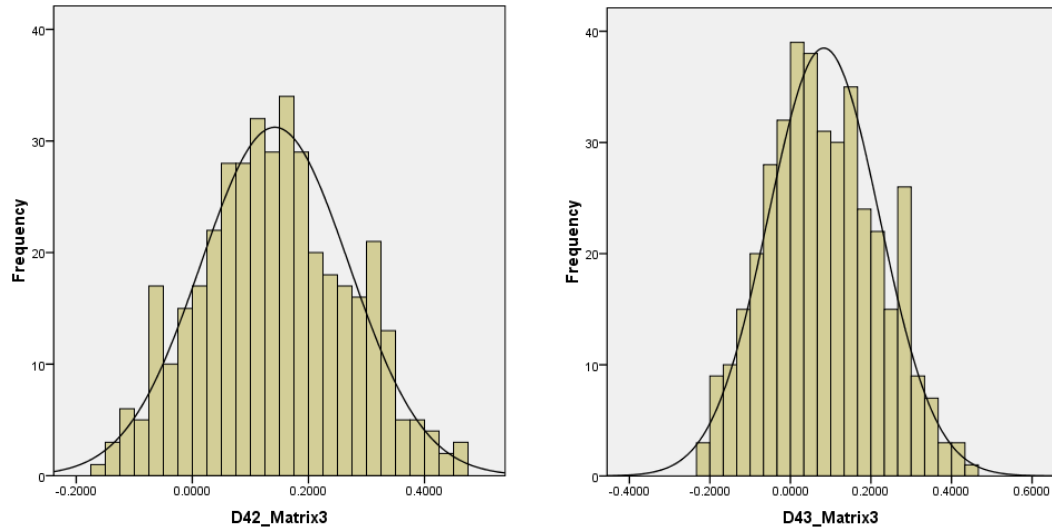


Figure 5.4 The bootstrap distribution of D_{43} and D_{42} in Matrix 3 when skew = 1 and $N = 100$

In addition, the bias-corrected bootstrap confidence interval is somewhat sample-sensitive. It adjusts for possible bias in the bootstrapped estimator and sample estimator. However, if the sample size is small (e.g., equal or less than 100), the sample may be less representative of the population (see the bias in Table 4.17 - 4.19). In these cases, the bias-corrected bootstrap confidence interval is not adequate to make inferences from the sample to the population. Specifically, it produced inaccurate results and a large false rejection rate. This is consistent with Padilla and Divers's (2013) results that the bias-corrected bootstrap confidence interval would be a good choice to estimate coefficient omega, a reliability index whose distribution remains unknown, for sample sizes of 150 or more.

Generally, based on the results of this study, it is recommended that researchers use the bootstrap confidence interval approach to make inference for Dominance Analysis when the assumption of multivariate normality is violated. Between the standardized bootstrap confidence interval, percentile bootstrap confidence interval, and bias-corrected bootstrap confidence interval, the selection depends on the circumstances. The bias-corrected bootstrap confidence interval is preferred when the sample size is at least 300 and when the sample is representative of the unknown population. When the bootstrap distribution of the targeted estimate is approximately normally distributed, it is recommended to choose the standardized bootstrap confidence interval since it requires fewer bootstrap resamples than the percentile and bias-corrected bootstrap confidence interval approaches.

For empirical researchers, these four confidence interval methods could be applied to their real data to make inference about the dominance in population. The choice among the four methods is based on the properties of the real data set, like sample size and distribution. It is recommended to test the normality and check the skewness and kurtosis of the real data set before making inference. In the example of payment prediction, the traditional Dominance Analysis told us that the age of oldest arrears (X_4) dominates risk score (X_3) when predicting the payment amount (Y), since the general dominance measures of X_4 (.1557) is greater than that of X_3 (.1104). However, it is found that the dominance relationship of X_4 over X_3 is not statistically significant after using the inferential approaches. In other words, though a difference of .0453 was observed between the general dominance measures of X_4 and X_3 based on the sample ($N = 448$), it

is not sufficiently indicative of a difference in the population. The selection of inferential methods depends on the properties of the data, according to the aforementioned general recommendations. Also for the empirical example, as the distribution is non-normal (as shown in Table 1) and the sample size is larger than 300, all three bootstrap confidence intervals are workable here and the standardized bootstrap confidence interval would be the best choice because it requires less computation.

5.2 Limitations and Future Directions

This study has several limitations that should be considered in generalizing its results. First, in the simulation studies, the conditions were set up to control for “noise” and easily interpret the results, but this may not be representative of data in the real world. For example, the parameters of skew are set as 0 when examining the impact of different degrees of kurtosis, and vice versa. Additionally, the population distributions of the criteria and predictors were generated as identical, which is also uncommon in reality. Future research should include more conditions using parameters from real cases, and compare the results with the current.

Secondly, the current study used 400 as the number of bootstrap resamples for all simulation conditions. It might be interesting to add different numbers of bootstrap resamples to determine the minimum required numbers under different conditions for different confidence interval approaches. Generally, the standardized bootstrap confidence interval requires fewer bootstrap resamples than percentile and bias-corrected

bootstrap confidence intervals. According to Efron (1988), 200 resamples is sufficient to obtain bootstrap estimates of standard deviation for standardized bootstrap confidence intervals; but nonparametric confidence intervals demand far more computation (e.g, 1000 resamples). It is possible that the performance of the percentile confidence interval will be enhanced when the number of bootstrap resamples is increased to 1000 or more.

Thirdly, in general, the rejection rate where the null hypothesis is not true (power) is expected to increase as the effect size increases. However, it was found that sometimes similar effect sizes from two different correlation matrices produce very different power rates. For example, D_{31} in Matrix 5 and D_{43} in Matrix 5 had same effect sizes (D_{uv} of 0.2400), but they generated very different levels of power with all of the four methods (around 1.0 and 0.60, respectively, with $n = 100$). The main reason for these odd results may be the great width of confidence interval (see Figure 4.15). The wider the confidence interval, the easier it is for it to include 0, and thereby produce lower power. Specifically, the width of the intervals for D_{31} in Matrix 5 and D_{43} in Matrix 5 are around 0.21 and 0.38 with all of the four methods, respectively. The standardized bias may be another reason for the differences, especially in terms of how well the parent sample represents the population: the larger the bias, the lower the power. Specifically, the standardized bias for D_{31} in Matrix 5 and D_{43} in Matrix 5 are -0.041 and -0.106, respectively. This type of difference exists across a variety of sample sizes and distributions, which is likely due to the correlation patterns. It is possible that a certain correlation pattern may produce large bias and thereby negatively impact the power. Although the current simulation study used 3 correlation patterns between predictors and criterion combined with 4 inter-

predictor correlation patterns, the potential impact of the correlation matrix pattern was not specifically explored. More matrix cases could be included to better understand the effect of the correlation pattern on the results in future studies.

Lastly, multiple comparisons are made within every correlation matrix. Specifically, there are six paired comparisons under a four-predictor regression model. However, this study did not correct the Type I error rate, dividing alpha by the number of comparisons. It is unclear whether and how the correction should be applied in this case. More research is needed on this issue. In addition, the study could include a part of comparing power when controlling for the empirical Type I error rate for each case.

Future studies could also extend the inferential methods for general dominance measures to examine inference for conditional dominance and complete dominance. This will be more complicated since multiple comparisons are needed to determine those dominance relationships. One way to address this issue is to recode the dominance measure as 1 (X_i dominates X_j), 0 (X_j dominates X_i) and 0.5 (dominance cannot be established between X_i and X_j) and use reproducibility to show the stability and robustness of dominance patterns (Azen & Budescu, 2003). However, this recoding of the continuous dominance measures into categorical variables causes information loss. Another future direction is to extend the inferential procedure of general dominance measures from multiple linear regression to other regression models such as multivariate regression, logistic regression, and hierarchical linear regression.

5.3 Conclusion

Despite the limitations, this study provides a contribution to the predictor importance comparison literature. To date, there is no study that thoroughly investigates the performance of the Normal-Theory based (asymptotic) confidence interval and bootstrap confidence intervals for predictors' dominance relationships using both normal and non-normal data. Previous studies (e.g., Azen & Sass, 2008) have addressed the inference procedures about additional conditions when data is multi-normally distributed. This study extended these to non-normal distributed data and focused on general dominance measures.

The obtained results generally support previous findings (e.g., Hedges & Olkin, 1983; Azen & Sass, 2008) that the asymptotic confidence interval method is adequate to make inferences for comparing two general dominance measures when the effect size is no less than 0.15 and the sample size is at least 100 or the effect size is no less than 0.05 and the sample size is at least 300, and when the distribution is multivariate normal or slightly non-normal. However, the bootstrap confidence interval methods are preferred over the asymptotic confidence interval when the data are considerably non-normal (e.g., skew > 0.75, or $|\text{kurtosis}| > 1.2$). Among the three bootstrap confidence interval methods, the standardized and percentile bootstrap confidence intervals consistently provided equally better performance than the bias-corrected bootstrap confidence interval across different sample sizes and degrees of non-normality. Given that the standardized bootstrap confidence interval requires less computation, it is preferred to make inferences for general dominance when the data are non-normally distributed. When the sample size is

300 and above, the bias-corrected bootstrap confidence interval perform almost equally as well as the standardized and percentile bootstrap confidence intervals. Therefore, if one does not wish to make the normality assumption about the population distribution of statistic, then the percentile bootstrap confidence interval for sample size of 100 or more or the bias-corrected bootstrap confidence interval for sample size of 300 or more are good choices. It is hoped that these simulation results can guide empirical researchers when they would like to make inferences regarding predictors' general dominance measures.

Future studies should firstly concentrate on stabilizing and refining the current results by including more conditions (e.g., more correlation matrices, sample sizes, and number of bootstrap resamples). Also, it could be interesting to further investigate the inferential methods for conditional dominance and complete dominance, as well as to extend the current study to multivariate regression, logistic regression, hierarchical linear regression, and so on.

References

- Alf, E. F., & Graf, R. G. (1999). Asymptotic confidence limits for the difference between two squared multiple correlations: A simplified approach. *Psychological Methods*, 4, 70-75.
- Algina, J., & Keselman, H. J. (1999). Comparing squared multiple correlation coefficients: Examination of a confidence interval and a test of significance. *Psychological Methods*, 4, 76-83.
- Amado, A. J. (1999). *Partitioning predicted variance into constituent parts: A primer on regression commonality analysis*. Paper presented at the annual meeting of the Southwest Educational Research Association, San Antonio, TX.
- Azen, R., & Budescu, D. V. (2003). The Dominance Analysis approach for comparing predictors in multiple regression. *Psychological Methods*, 8, 129–148.
- Azen, R., & Budescu, D. V. (2006). Comparing predictors in multivariate regression models: an extension of Dominance Analysis. *Journal of Educational and Behavioral Statistics*, 31, 157–180.
- Azen, R., & Sass, D. A. (2008). Comparing the squared multiple correlation coefficients of non-nested models: An examination of confidence intervals and hypothesis testing. *British Journal of Mathematical and Statistical Psychology*, 61, 163-178.
- Azen, R., & Traxel, N. M. (2009). Using Dominance Analysis to determine predictor importance in logistic regression. *Journal of Educational and Behavioral Statistics*, 34, 319–347.
- Bradley, J. V. (1978). Robustness? *The British Journal of Mathematical & Statistical Psychology*, 31, 144-152.

- Brenick, A., Titzmann, P., Michel, A., & Silbereisen, R. K. (2012). Perceptions of discrimination by young Diaspora migrants: Individual and school-level associations among adolescent ethnic German immigrants. *European Psychologist*, 17(2), 105-109.
- Budescu, D. V. (1993). Dominance Analysis: A new approach to the problem of relative importance of predictors in multiple regression. *Psychological Bulletin*, 114, 542–551.
- Budescu, D. V., & Azen, R. (2004). Beyond global measures of relative importance: Some insights from Dominance Analysis. *Organizational Research Methods*, 7(3), 341–350.
- Carpenter, J., & Bithell, J. (2000). Bootstrap confidence intervals: When, which, what? A practical guide for medical statisticians. *Statistics in Medicine*, 19, 1141–1164.
- Chan, W. (2009). Bootstrap standard error and confidence intervals for the difference between two squared multiple correlation coefficients. *Educational and Psychological Measurement*, 69, 566-584.
- Chatterji, M. (2006). Reading achievement gaps, correlates, and moderators of early reading achievement: Evidence from the Early Childhood Longitudinal Study (ECLS) kindergarten to first grade sample. *Journal of Educational Psychology*, 98(3), 489-507.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Colquitt, J. A., LePine, J. A., Piccolo, R. F., Zapata, C. P., Rich, B. L. (2012) Explaining the justice–performance relationship: Trust as exchange deepener or trust as

- uncertainty reducer? *Journal of Applied Psychology*, 97(1), 1-15.
- Cooley, W. W., & Lohnes, P. R. (1971). *Multivariate data analysis*. New York: Wiley.
- Courville, T., & Thompson, B. (2001). Use of structure coefficients in published multiple regression articles: b is not enough. *Educational and Psychological Measurement*, 61, 229-248.
- Darlington, R. B. (1968). Multiple regression in psychological research and practice. *Psychological Bulletin*, 69, 161-182.
- Darlington, R. B. (1990). *Regression and linear models*. New York: McGraw-Hill.
- Davison, A. C., & Hinkley, D. V. (1997). *Bootstrap methods and their application*. Cambridge, England: Cambridge University Press.
- Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics*, 7, 1-26.
- Efron, B. (1982). *The jackknife, the bootstrap and other resampling plans*. Philadelphia: Society for Industrial and Applied Mathematics.
- Efron, B., & Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science*, 1, 54-77.
- Efron, B. (1987). Better bootstrap confidence intervals. *Journal of the American Statistical Association*, 82, 171-200.
- Efron, B., & Tibshirani, R. (1993). *An introduction to the bootstrap*. New York: Chapman & Hall.
- Fabbris, L. (1980). Measures of predictor variable importance in multiple regression: An additional suggestion. *Quality & Quantity*, 4, 787-792.
- Finch, J. F., West, S. G., & MacKinnon, D. P. (1997). Effects of sample size and

- nonnormality on the estimation of mediated effects in latent variable models. *Structural Equation Modeling*, 4(2), 87-107.
- Fleishman, A. I. (1978). A method for simulating non-normal distributions. *Psychometrika*, 43(4), 521-532.
- Graf, R. G., & Alf, E. F. (1999). Correlations redux: Asymptotic confidence limits for partial and squared multiple correlations. *Applied Psychological Measurement*, 23(2), 116-119.
- Hau, K., & Marsh, H. W. (2004). The use of item parcels in structural equation modelling: Non-normal data and small sample sizes. *British Journal of Mathematical and Statistical Psychology*, 57(2), 327-351.
- Hedges, L. V., & Olkin, I. (1983). Joint distributions of some indices based on correlation coefficients. In S. Karlin, T. Amemiya, & L. A. Goodman (Eds.), *Studies in econometrics, time series, and multivariate statistics* (pp. 437-454). New York: Academic Press.
- Hinkle, D. E., Wiersma, W., & Jurs, S. G. (2003). *Applied Statistics for the Behavioral Sciences* (5th ed.). Boston: Houghton Mifflin.
- Johnson, R. (2001). An introduction to the bootstrap. *Teaching Statistics*, 23, 49-54.
- Johnson, J. W. (2000). A heuristic method for estimating the relative weight of predictor variables in multiple regression. *Multivariate Behavioral Research*, 35, 1-19.
- Johnson, J. W. (2004). Factors affecting relative weights: The influence of sampling and measurement error. *Organizational Research Methods*, 7, 283-299.
- Johnson, J. W., & LeBreton, J. M. (2004). History and use of relative importance indices in organizational research. *Organizational Research Methods*, 7, 238-257.

- Kim, M. J., Choi, J. N., & Park, O. S. (2012). Intuitiveness and creativity in groups: Cross-level interactions between group conflict and individual cognitive styles. *Social Behavior and Personality, 40*, 1419-1434.
- Krasikova, D., LeBreton, J. M., & Tonidandel, S. (2011). *Estimating the relative importance of variables in multiple regression models*. In G. P. Hodgkinson & J. K. Ford (Eds.), *International review of industrial and organizational psychology*. Indianapolis, IN: Wiley.
- Kruskal, W., & Majors, R. (1989). Concepts of relative importance in recent scientific literature. *American Statistician, 43*, 2–6.
- LeBreton, J. M., Ployhart, R. E., & Ladd, R. T. (2004). A Monte Carlo comparison of relative importance methodologies. *Organizational Research Methods, 7*, 258–282.
- Maydeu-Olivares, A., Coffman, D. L. & Hartmann, W. M. (2007). Asymptotically distribution-free (ADF) interval estimation of coefficient alpha. *Psychological Methods, 12*(2), 157–176.
- Mayeske, G. W., Cohen, W. M., Wisler, C. E., Okada, T., Beaton, A. E., Proshek, J. M., Weinfeld, F. D., & Tabler, K. A. (1969). *A study of our nation's schools*. Washington, DC: U.S. Department of Health, Education, and Welfare, Office of Education.
- Mellard, D. F., Anthony, J. L., & Woods, K. L. (2012). Understanding Oral Reading Fluency among Adults with Low Literacy: Dominance Analysis of Contributing Component Skills. *Reading and Writing, 25*(6), 1345-1364.
- Miller, B. K., Konopaske, R., & Byrne, Z. S. (2012). Dominance analysis of two

- measures of organizational justice. *Journal of Managerial Psychology*, 27, 264-282.
- Mood, A. M. (1969). Macro-analysis of the American educational system. *Operations Research*, 17, 770-784.
- Mood, A. M. (1971). Partitioning variance in multiple regression analyses as a tool for developing learning models. *American Educational Research Journal*, 8, 191-202.
- Morrow, S., McGonagle, A., Dove-Steinkamp, M., Walker, Jr., C., Marmet, M., & Barnes-Farrell, J. (2010). Relationships between psychological safety climate facets and safety behavior in the rail industry: A dominance analysis. *Accident Analysis and Prevention*, 42, 1460-1467.
- Nathans, Laura L., Oswald, Frederick L., & Nimon, Kim (2012). Interpreting Multiple Linear Regression: A Guidebook of Variable Importance. Practical Assessment, *Research & Evaluation*, 17(9).1-19.
- Newton, R. G., & Spurrell, D. J. (1967). A development of multiple regression for the analysis of routine data. *Applied Statistics*, 16, 51-64.
- Nimon, K., Gavrilova, M., & Roberts, J. K. (2010). Regression results in human resource development research: Are we reporting enough? In C. Graham & K. Dirani (Eds.), *Proceedings of the Human Resource Development 2010 International Conference* (pp. 803-812), Knoxville, TN: AHRD.
- Olkin, I., & Finn, J. D. (1995). Correlations redux. *Psychological Bulletin*, 118, 155-164.
- Olkin, I., & Siotani, M. (1976). Asymptotic distribution of functions of a correlation matrix. In S. Ikeda (Ed.). *Essays in probability and statistics* (pp. 235-251).

Tokyo: Shinko Tsusho.

- Padilla, M. A. & Divers, J. (2013). Bootstrap Interval Estimation of Reliability via Coefficient Omega. *Journal of Modern Applied Statistical Methods*, 12(1), 78-89.
- Pedhazur, E. J. (1997). *Multiple regression in behavioral research: Explanation and prediction* (3rd ed.). Stamford, CT: Thompson Learning.
- Pratt, J. W. (1987). Dividing the indivisible: Using simple symmetry to partition variance explained. In T. Pukkila & S. Puntanen (Eds.), *Proceedings of the second international Tampere conference in statistics* (pp. 245-260). Tampere, Finland: University of Tampere.
- Robertson, C. (1991). *Computationally intensive statistics*. In P. Lovie & A. D. Lovie (Eds.), *New developments in statistics for psychology and the social sciences* (pp. 49-80). London: The British Psychological Society.
- Sharma, S., Durvasula, S., & Dillon, W. R. (1989). Some results on the behavior of alternate covariance structure estimation procedures in the presence of non-normal data. *Journal of Marketing Research*, 26(2), 214-221.
- Smithson, M. (2003). *Confidence intervals*. Thousand Oaks, CA: Sage.
- Tang, S.W., Siu, O. L., & Cheung, F. (2014). A Study of Work-Family Enrichment among Chinese Employees: The Mediating Role between Work Support and Job Satisfaction. *Applied Psychology: An International Review*, 63(1), 130-150.
- Ting, Kwok-fai (2000). Cross-level effects of class characteristics on students' perceptions of teaching quality. *Journal of Educational Psychology*, 92(4), 818-825.
- Thomas, D. R., Zumbo, B. D., Kwan, E., & Schweitzer, L. (2014). On Johnson's (2000)

Relative Weights Method for Assessing Variable Importance: A Reanalysis. *Multivariate Behavioral Research*, 49, 329-338.

Thompson B. (2006). *Foundations of Behavioral Statistics: An Insight-Based Approach*. New York: Guilford Press.

Tonidandel, S., LeBreton, J. M., & Johnson, J. W. (2009). Determining the statistical significance of relative weights. *Psychological Methods*, 14, 387–399.

Tzelgov, J., & Henik, A. (1991). Suppression situations in psychological research: Definitions, implications, and applications. *Psychological Bulletin*, 109, 524–536.

Ward, J. H. (1962). Comments on “The paramorphic representation of clinical judgment.” *Psychological Bulletin*, 59, 74–76.

Wicklin, R. (2013). *Simulating Data with SAS*, Cary, NC: SAS Institute Inc.

Zientek, L. R., Capraro, M. M., & Capraro, R. M. (2008). Reporting practices in quantitative teacher education research: One look at the evidence cited in the AERA panel report. *Educational Researcher*, 34, 208-216.

Curriculum Vitae

Shuwen Tang

EDUCATION

- | | |
|-----------------------------------|--|
| Doctor of Philosophy
2010-2014 | University of Wisconsin–Milwaukee (UWM), WI
Major: Educational Statistics and Measurement
Minor: Mathematical Statistics
GPA: 3.83/4.0
Dissertation: Inferential Procedures for Dominance Analysis
Measures in Multiple Regression
Advisor: Dr. Razia Azen |
| Master of Philosophy
2008-2010 | Lingnan University (LN), Hong Kong
Major: I/O Psychology
Thesis: Work Support, Work-family Enrichment, Work Demand
and Work Well-being among Chinese Employees: A Study of
Mediating and Moderating Processes
Advisor: Dr. Oi-ling Siu |
| Bachelor of Science
2004-2008 | Hangzhou Normal University (HZNU), P. R. China
Major: Psychology, with Honors
GPA: 3.90/4.0 |

RESEARCH INTERESTS

Multilevel Modeling, Mediation and Moderation analysis, Hierarchical Linear Models, Structural Equation Modeling, Bootstrapping procedure, Item Response Theory, Multidimensionality, Differential Item Functioning, and Computer-based Testing.

PUBLICATIONS

Tang, S.W., Siu, O. L., & Cheung, F. (2014). A Study of Work-Family Enrichment among Chinese Employees: The Mediating Role between Work Support and Job Satisfaction*. *Applied Psychology: An International Review*, 63(1), 130-150.

Cheung, F., Tang, C., & **Tang, S.W.** (2011). Psychological capital as a moderator between emotional labor, burnout, and job satisfaction among school teachers in China. *International Journal of Stress Management*, 18, 348-371.

Zhu, X. B., & **Tang, S.W.** (2009). The relationship of primary school pupils' writing attitude with their writing achievement*. *Psychological Science (Chinese)*, 32(4), 942-945.

CONFERENCE PRESENTATIONS

Tang, S.W., Zeng, W., Walker, C, & Potter, N.(2014). DIF Analysis when Recoding the Polytomous Response as Dichotomous. *Annual Meeting of the National Council on Measurement in Education*, Philadelphia, PA.

Tang, S.W., & Luo, W. (2013). Examining Cross-level Interactions in Multilevel Models: Problems and Solutions. *121st Annual Convention of the American Psychological Association*. Honolulu, Hawaii.

Ning, L., Walker, C, & **Tang, S.W.** (2013). DIF Purification Strategies in MIRT. *Annual Meeting of the National Council on Measurement in Education*, San Francisco, CA.

Azen, R., **Tang, S.W.**, & Budescu, D. V. (2011). Comparing Methods to Evaluate Predictor Importance in Lexicographical Models. *International Meeting of the Psychometric Society*, Hong Kong.

Tang, S.W. (2010). Work Support, Work-family Enrichment and Work Well-being: A Study of Mediating Processes among Chinese Employees. *International Conference on Behavioral, Cognitive and Psychological Sciences*. Paris, France.

RESEARCH IN PROGRESS

The Robustness of SIBTEST in a Multiplicative Multidimensional IRT Model
with Dr. Cindy Walker

DIF Analysis when Recoding the Polytomous Response as Dichotomous
with Dr. Cindy Walker

Examining Cross-level Interactions using Multilevel SEM
with Dr. Wen Luo

Using Conditioning of Total Score Goodness-of-fit Statistics for Testlet Models
with Dr. Bo Zhang

RESEARCH EXPERIENCE

2010-2013 Project Assistant/ statistical consultant,
 Consulting Office for Research and Evaluation

- research design, database work, statistical analysis, instrument development, and software training

2013 “A study of School Climate in Milwaukee Public Schools”
 Supervisor: Dr. Cindy Walker

2012 “Strategies for Motivating and Supporting Online Students”
 Supervisor: Dr. Cindy Walker & Dr. Simone Conceição

2010-2013 “Fostering Opportunities for Tomorrow's Engineers (FORTE) Program”
 (The National Science Foundation's STEM Talent Expansion Program (STEP))
 Supervisor: Dr. Cindy Walker & Dr. John R. Reisel

- 2010 “OIC of America’s federally-funded Healthy Marriage Program (09-10)”
Supervisor: Dr. Cindy Walker
- 2008-2010 Research Assistant, Lingnan University
 - database building, statistical analysis, training assistant, and report drafting
- 2010 “Psychological Approach in Handling Students' Challenging Behavior”
(Education Bureau, Hong Kong) Supervisor: Dr. Oi-ling Siu
- 2009 “Managing Abusive/Violent Customers in Workplace for Managerial Staff”
(The Hong Kong Housing Authority) Supervisor: Dr. Oi-ling Siu
- 2008-2010 “Work Family Balance”
(3-yr international project involving 15,000 employees)
Supervisor: Dr. Oi-ling Siu
- 2009 “Life Writing Research Program”
Supervisor: Dr. Judy Ho

TEACHING EXPERIENCE

Fall 2013, Spring 2014

Teaching Assistant for *EdPsy724 Educational Statistical Methods II*

- laboratory instruction on SPSS & SAS, grading, meeting with students

Fall 2009, Spring 2010

Teaching Assistant for *BEH206 Abnormal Psychology*

- 4 tutorials/week, grading, class materials preparation, meeting with students

Fall 2008, Spring 2009

Teaching Assistant for *BEH103 Introduction to Behavioral Science*

- 5 tutorials/week, grading, tutoring and meeting with students

Fall 2007

Teacher and School Psychology Consultant in Ziyang Primary School

- psychological assessment, crisis intervention, individual and group counseling

INTERNSHIP

Intern-Analytics, Customer Service - Business Effectiveness and Integration

Wisconsin Electric Power Company (We Energies)

- Analyze huge data and figure out innovative solutions that drive efficiency, cut cost and increase customer satisfaction.

AWARDS & HONORS

2010-2013 Travel Award by University of Wisconsin–Milwaukee

2010-2013 Graduate Scholarship by University of Wisconsin–Milwaukee

2008-2010 Full Postgraduate Scholarship by Lingnan University
 2008 The “Fuhui Award” by Fuhui Charity Foundation Limited, Hong Kong

STATISTICAL PROGRAMS

Statistical	•SPSS	•SAS	•R		
Modeling	•Amos	•MPlus	•HLM	•LISREL	
IRT	•SIBTEST	•WinBUGS	•Bilog	•Dimtest	•Scoright
SAS Certified Base Programmer			2011		
SAS Certified Advanced Programmer			2012		

PROFESSIONAL ORGANIZATIONS

American Psychological Association (APA, Division 5)
 American Statistical Association (ASA)
 American Educational Research Association (AERA, Division D)
 National Council on Measurement in Education (NCME)

PROFESSIONAL ACTIVITIES

2013 Reviewer
 National Council on Measurement in Education Graduate Poster Session
 2008-2009 Committee Member
 Research Postgraduate Student Circle in Lingnan University
 2005-2007 President
 Psychology Student Association in Hangzhou Normal University

REFERENCES

Razia Azen, Ph.D.
Associate Professor
Department of Educational
Psychology
University of Wisconsin –
Milwaukee
Phone: (414) 229-4767
azen@uwm.edu

Bo Zhang, Ph.D.
Associate Professor
Department of Educational
Psychology
University of Wisconsin –
Milwaukee
Phone: (414) 688-4690
boz@uwm.edu

Oi-ling Siu, Ph.D.
Professor and Head
Department of Psychology
Lingnan University, Hong Kong
Phone: (852) 2616 7170
siuol@ln.edu.hk

Cindy M. Walker, Ph.D.
Associate Dean of Research and Engaged Scholarship
Professor, Department of Educational Psychology
Director, Consulting Office for Research and
Evaluation
University of Wisconsin – Milwaukee
Phone: (414) 229-5441
cmwalker@uwm.edu

Wen Luo, Ph.D.
Associate Professor
Department of Educational Psychology
Texas A&M University
Phone: (979)845-9250
wluo@neo.tamu.edu