2-14-2012 12:00 AM

Confidence Intervals for Comparison of the Squared Multiple Correlation Coefficients of Non-nested Models

Li Tan Jr.
The University of Western Ontario

Supervisor
Dr. G.Y. Zou
The University of Western Ontario Joint Supervisor
Dr. John Koval
The University of Western Ontario

Graduate Program in Biostatistics
A thesis submitted in partial fulfillment of the requirements for the degree in Master of Science

Confidence Intervals for Comparison of the Squared Multiple Correlation Coefficients of Non-nested Models

(Spine Title: CIs for Comparison of the Squared Multiple Correlations of Non-nested Models)

(Thesis Format: Monograph)

by

Li Tan, MSc.

Graduate Program in Epidemiology & Biostatistics

Submitted in partial fulfillment

of the requirements for the degree of

Master of Science

School of Graduate and Postdoctoral Studies

The University of Western Ontario

London, Ontario

February 2012

THE UNIVERSITY OF WESTERN ONTARIO

SCHOOL OF GRADUATE AND POSTDOCTORAL STUDIES

CERTIFICATE OF EXAMINATION

Joint Supervisor	Examiners
Dr. Guangyong Zou	Dr. Allan Donner
Joint Supervisor	Dr. Yun-Hee Choi
Dr. John J. Koval	Dr. Serge B. Provost

The thesis by

Li Tan

entitled

Confidence Intervals for Comparison of the Squared Multiple Correlation Coefficients of Non-nested Models

is accepted in partial fulfillment of the

requirements for the degree of

Master of Science

Date _____

_____

Chair of the Thesis Examination Board

ABSTRACT

Multiple linear regression analysis is used widely to evaluate how an outcome or response variable is related to a set of predictors. Once a final model is specified, the interpretation of predictors can be achieved by assessing the relative importance of predictors.

A common approach to predictor importance is to compare the increase in squared multiple correlation for a given model when one predictor is added to the increase when another predictor is added to the same model.

This thesis proposes asymmetric confidence-intervals for a difference between two correlated squared multiple correlation coefficients of non-nested models. These new procedures are developed by recovering variance estimates needed for the difference from asymmetric confidence limits for single squared multiple correlation coefficients. Simulation results show that the new procedure based on confidence limits obtained from the two-moment scaled central F approximation performs much better than the traditional Wald approach. Two examples are used to illustrate the methodology. The application of the procedure in dominance analysis and commonality analysis is discussed.

KEYWORDS: Coefficient of determination; Multiple correlation coefficient; Dominance analysis; Commonality analysis.

ACKNOWLEDGMENTS

I thank my two advisors Professor Guangyong Zou and Professor John Koval for their continuous guidance, encouragement and advice throughout my graduate studies and the number of hours they spent with me on this research. It has been a privilege and a pleasure to work with mentors with such wisdom and graciousness.

I thank all the friends that I have met over my many years at the University of Western Ontario. I thank all of the Department of Epidemiology & Biostatistics faculty and staff members.

Finally, my very special thanks to my farther, mother and brother for their love and support while I decided to begin another master's degree for awhile. This thesis is dedicated to my husband, Jun, who always has confidence in me and whose encouragement is essential for the completion of my thesis, and also my lovely son, Albert.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION

Multiple linear regression model is one of the most frequently used tools for evaluating how an outcome or response variable is related to a set of predictors. To quantify the performance of the model, the coefficient of determination is commonly used. All statistical computer packages provide values of R^2 automatically, but without mentioning statistical inference for its population parameter (ρ^2). It is well-known that testing $\rho^2 = 0$ can be achieved using an F-test. However, confidence interval construction for ρ^2 is rarely mentioned even though confidence intervals are more informative. The primary goal of this thesis is to develop inference procedures for quantifying the importance of predictors using confidence intervals for changes in ρ^2. Specifically, we focus on the increase in ρ^2 for a given model when one predictor is added as compared to the increase in ρ^2 for the same model when another predictor is added.

1.1 Inferences for a single squared multiple correlation

The coefficient of determination has several definitions. Generally, R^2 is defined as the proportion of "variability" (measured by the sum of squares) in a data set accounted for by a multiple regression model (e.g., Steel and Torrie, 1960, pg. 187, 287). This interpretation is usually presented at the conclusion of a multiple regression analysis. R^2 is also defined as the sample squared correlation coefficient between the response variable and its corresponding predicted value from the regression model (e.g., Cohen et al., 2003).

Based on the definition, R^2 is a widely used goodness-of-fit statistic for the overall

performance of a multiple regression model. A coefficient of determination can represent a measure of how well the regression line approximates the observed data points. It lies between 0 and 1. The closer it is to 1, the better is the linear relationship between the response variable and predictors. The closer it is to 0, the worse is the linear relationship. The correlation coefficient of 0 indicates no linear relationship between variables, although nonlinear relationship may exist. However, there also exists some controversy regarding R^2 as a goodness of fit statistic (Hagquist and Stenbeck, 1998). One argument is that the value of R^2 always increases even when a non-predictive regressor is added in a linear regression model; this can be dealt with by adjusting the R-squared. By including a penalty for the number of predictors in a model, the adjusted R-squared increases only if the added predictor improves the model more than would be expected by chance (Ezekiel, 1930). Another argument is that correlation does not imply causation, since correlation between two variables may exist due to common causes, confounding variables or coincidences (Aldrich, 1995). Moreover, even if the causal relationships between the outcome and predictors in two regression models are identical, the value of R^2 may differ greatly between different samples. R^2 is regarded as more meaningful as a point estimate of population ρ^2 only for a data set with random regressors (Helland, 1987). For a model with random regressors, the accuracy of R^2 depends on not only the sample size, but also the assumed distribution of predictors.

Besides evaluating the overall performance of a multiple regression model, R-squared also can be served as a general measure of determining the relative importance of predictors in multiple regression analysis (Budescu, 1993). After having built a multiple regression model with a chosen set of predictors, one may want to know a relatively important subset of predictors, or rank the predictors according to their contributions in predicting the outcome. Hence, it is an important issue in multiple regression analysis to choose an intuitive and meaningful index of importance for any predictor. There were three classes of measures of importance including slope-based measures such as regression coefficients, standardized regression coefficients and the products of the mean of a predictor and the corresponding

regression coefficient, correlational measures such as the correlation, the squared correlation and the squared partial correlation (The coefficient of partial correlation is defined as the correlation coefficient between two sets of variables keeping a third set of variable constant.), and measures based on a combination of the regression coefficients and the correlations such as the product of the correlation between a predictor and the outcome and the corresponding standard regression coefficient (Azen and Budescu, 2003; Budescu, 1993). However, all these measures do not offer an intuitive and universal interpretation of importance which leads to different orderings of the predictors' importance and confusion on the meaning of importance. Budescu (1993) suggested that an appropriately general measure of importance should satisfy the following three conditions: "(a) Importance should be defined in terms of a variable's 'reduction of error' in predicting the outcome; (b) The method should allow for direct comparison of relative importance instead of relying on inferred measures; (c) Importance should reflect a variable's direct effect, total effect and partial effect." According to these criteria, Budescu (1993) developed a new methodology – dominance analysis, in which a predictor is considered to be dominant or more important than another predictor if its additional contribution in the prediction of the response variable defined as the squared semipartial correlation (i.e., the difference between two squared multiple correlations from nested models), is greater than the competitor's for all possible subset models. In a word, one can identify the relative importance of predictors through a series of pairwise comparisons of squared multiple correlations from all submodels.

1.1.1 Fixed and random regressors

Depending on whether regressors are fixed or random, inference procedures for ρ^2 are different. A key distinction with respect to ρ^2 between fixed-score and random-score multiple regression models is that ρ^2 is constant for fixed-score regression models, while population ρ^2 can be made arbitrarily small and large when regressors are random.

For a multiple regression model having p fixed predictors and sample size n, statistical inferences on population squared multiple correlations are relatively easy. In partic-

ular, when testing for the null hypothesis $H_0 : \rho^2 = 0$, one usually construct a statistic $\frac{R^2/p}{(1-R^2)/(n-p-1)}$. Under H_0, the statistic follows a central F distribution with degrees of freedom p and $n-p-1$. When $\rho \neq 0$, the statistic follows a noncentral F distribution with degrees of freedom p and $n-p-1$ and a noncentrality parameter.

In this study, attention is restricted to inference procedures for models with random regressors. Since inference for ρ^2 arising from fixed regressors can be easily done on the basis of the noncentral F distribution with degrees of freedom independent of ρ^2. For random regressors, the same central F-distributed statistic can be used for testing significance, because when the null hypothesis is true, the sampling distribution for R^2 does not change for fixed or random regressors (Smithson, 2003). However, this is not a case if we are interested in confidence intervals for ρ^2 in models with random regressors. Since for nonzero cases, the noncentrality parameter is a function of regressors, the unconditional distribution of R^2 highly depends on the assumed distributions of those regressors. Hence, when constructing confidence intervals for ρ^2, many statistical inference procedures suitable for fixed regressors only can provide conditional confidence intervals on particular values of random regressors observed in the sample. In fact, since the random regressors themselves take account of sampling error, the confidence intervals for ρ^2 from fixed regressors models are much narrower than those from random regressors models (Smithson, 2003). Hence, a unconditional confidence interval-based inference for ρ^2 in models with random regressors is preferred.

1.1.2 Point estimation for a single squared multiple correlation

It is well known that R^2 is defined as the ratio of regression sum of squares and the total sum of squares. According to this definition, one can calculate the value of R^2 from the fitted model. Almost all statistical softwares automatically provide the value of R-squared for a multiple regression model.

Under the assumption that the joint distribution of the outcome and predictors is multivariate normal, R^2 is approximately unbiased as a sample size approaches to infinity

(Hagquist and Stenbeck, 1998). However, R^2 usually has a positive bias as an estimate of ρ^2, especially when the number of regressors are moderate or large. Many researchers have developed various alternatives that reduce the bias (see Raju et al., 1997). A commonly used estimator is the R^2 adjusted by replacing the sum of squares with the mean square by (Ezekiel, 1930). Alf and Graf (2002) compared R^2 and eight other estimates of ρ^2, which includes Smith's estimator (Ezekiel, 1929), an estimator proposed by Wherry (1931), the adjusted R^2 (Ezekiel, 1930), a unbiased estimator (Olkin and Pratt, 1958) and its two approximate versions proposed by Pratt and presented in Claudy (1978) and by Herzberg (1969), a empirically based estimate (Claudy, 1978), and the maximum likelihood estimate (Alf and Graf, 2002). The first two have similar modifications to the adjusted R^2 by including a penalty for the number of predictors in a model, and the other five estimators involving the unbiased estimate provided by Olkin and Pratt (1958) are derived from models with random regressors. Alf and Graf (2002) found that: 'the adjusted R^2 (Ezekiel, 1930) was unbiased only when ρ^2 is 0'; Olkin and Pratt (1958)'s estimator is unbiased over all the values of the sample size n and ρ^2, but involves a complex hypergeometric function; and the remaining estimators (including R^2 and a maximum likelihood estimate of ρ^2 derived from its exact density function) were biased for all the values of n and ρ^2.

Another way to compute the R^2 of a regression model requires the sample simple correlations among the dependent variable and predictors within the model. According to the definition (Pearson and Filon, 1898), a multiple correlation coefficient can be mathematically represented as a function of simple and/or partial correlation coefficients, while partial correlations are also functions of simple correlations. Correspondingly, as a sample squared multiple correlation coefficient, R^2 can be represented as a function of sample simple correlation coefficients. This relationship also can be represented in matrix form. That is, R^2 can be represented as a function of several determinants of matrices of the sample simple correlations between variables included in the model (Olkin and Siotani, 1976). These formulas based on correlation matrices are easily used to estimate ρ^2 of a regression model with small number of predictors.

1.1.3 Confidence interval estimation for a single squared multiple correlation

Several confidence interval procedures for a single ρ^2 in random scores models have appeared in the literature. These procedures include a Wald-type confidence interval, a confidence interval based on Fisher's R^2-to-z transformation, a confidence interval derived from the exact density of R^2, asymptotic confidence intervals based on various approximations to the density of R^2, and bootstrap confidence intervals which do not assume multivariate normality for the outcome and regressors.

With an estimator of the asymptotic variance of R^2 provided by Wishart (1931), the Wald method can be used to construct a symmetric confidence interval of ρ^2. However, the forced symmetry of the Wald-type confidence interval is questionable, since the sampling distribution of R^2 is skewed and converges very slowly to normality (Algina, 1999). Furthermore, this Wald-type confidence interval constructed by using ordinary R^2 and its variance estimate may be out of range of 0 to 1. Meanwhile, if one uses the adjusted R^2 and its variance estimate instead of ordinary R^2 and its variance estimate, since the adjusted R^2 has a much bigger variance estimate than R^2 does, the resulting lower Wald-type confidence limit for ρ^2 may be negative, which is inconsistent with that of the associated F test.

Fisher's z transformation which is commonly used for inferences on simple correlations has been applied to make inferences on squared multiple correlations. The confidence limits for ρ^2 can be represented as a monotone increasing function of confidence limits of Fisher's z statistic, which is assumed to be approximately normally distributed. However, Fisher's statistic, since it takes only non-negative values, has a more positive skewed distribution than R^2 (Algina, 1999). Therefore, it is not recommended that Fisher's transformation is used for constructing confidence intervals for ρ^2.

Under the multivariate normality assumption, Fisher (1928) derived the exact density function of R^2. Due to the complex form of the density, it is impossible to derive an analytical confidence interval for ρ^2 from the density, although lower confidence limits for ρ^2 in particular cases based on the exact density has been tabulated (see Kramer, 1963;

Lee, 1972). Various asymptotic confidence intervals for ρ^2 have been proposed through approximating a transformed variable $R^2/(1-R^2)$ to be scaled central F distributed (Gurland, 1968; Helland, 1987), scaled noncentral F distributed (Lee, 1971), and relocated and rescaled central F distributed (Lee, 1971). The comparison between the exact and three approximate density functions of R-square shows that the scaled noncentral F approximation which requires the estimation of the noncentrality parameter performs much better than that of the relocated and rescaled central F approximation, while slightly better than that of the scaled central F approximation.

Bootstrap methods first proposed by Efron (1979) may be thought to provide a universal solution to inference, especially for those estimators having an unknown or complicated distribution. The bootstrap aims to use computer-based re-sampling to approximate the sampling distribution of the estimate of the parameter. Common bootstrap methods for constructing confidence intervals are the percentile (Schenker, 1985), the bias-corrected (BC) (Efron, 1981), the bias-corrected and accelerated (BCa) (Efron, 1987), bootstrap-t (Efron, 1979), and the approximate bootstrap confidence intervals (ABC) (Efron and Tibshirani, 1993). However, there are limitations to these bootstrap methods (Efron, 1987; Efron and Tibshirani, 1993). For example, the percentile interval performs well only for unbiased statistics having a symmetric sampling distribution; the BC method requires the existence of a normalizing transformation and stabilized variance, and is applicable only for large samples; the validity of the BCa method highly depends on the accuracy of estimating an extra acceleration constant; the bootstrap-t method requires an accurate estimate of the standard error of the statistic and often yields a very wide confidence interval; the ABC is applicable for smoothly defined parameters in exponential families and also requires an estimate of the nonlinearity parameter. Furthermore, nonparametric methods like jackknifing and bootstrapping are more sensitive to the sample size than valid parametric methods. A cautionary example shows that all jackknife and bootstrap confidence intervals of ρ^2 perform too poorly to be trusted for a data set with small sample size and relatively large number of predictors (SAS, 2010). In addition, all these bootstrap confidence inter-

vals require intensive computations, hence, as Efron (1988) stated, the bootstrap can be an alternative method only when there is no any suitable parametric methods.

In summary, the scaled central F approximation (Gurland, 1968; Helland, 1987) and the scaled noncentral F approximation (Lee, 1971) may be good choices to construct confidence interval for a single squared multiple correlation coefficient. Furthermore, the confidence intervals based on these two approximations are easily obtained by respectively implementing existing statistical package programs, such as SAS PROC CANCORR, a SAS macro or an SPSS syntax provided by (Zou, 2007).

1.2 Inferences for a difference between two squared multiple correlations

As mentioned before, a squared multiple correlation is an important measure for quantifying the overall performance of a model, and also for quantifying the importance of a set of predictors. Explicitly, comparisons of R^2s from all possible linear regression submodels involving various sets of predictor, are made to identify a relatively important subset of predictors, or rank the predictors according to their contributions in predicting the response variable. Both are the final objectives of two popular research fields (Azen and Budescu, 2003; Hedges and Olkin, 1981), dominance analysis (Budescu, 1993) and commonality analysis (Kerlinger and Pedhazur, 1973).

The exact sampling distribution of a single R^2 is already highly complex, so we may have to resort to the joint asymptotic distribution theory of two or more squared multiple correlations. For models with random regressors, the joint distribution of simple correlations or functions of correlations (e.g., squared multiple correlation) among the response variable and regressors highly depends on the distribution of the depended variable and independents.

When the outcome and regressors are assumed to be multivariate normal, most literatures have presented the asymptotic joint distributions of functions of correlations including the vector of sample simple correlations between variables (Pearson and Filon, 1898), de-

terminants of correlation matrices (Olkin and Siotani, 1976), squared multiple correlations (Hedges and Olkin, 1981), any sets of partial and/or multiple correlations (Hedges and Olkin, 1983), and so on. Each of these joint distributions is asymptotically multivariate normal by applying the central limit theorem. However, these asymptotic joint distributions are only applicable for very large samples, since the exact joint distribution of two correlated R^2s is probably still extremely skewed.

Without the multivariate normality assumption for a vector variate of interest, Steiger and Hakstian (1982) presented an asymptotic joint distribution of simple correlations among variables involving the kurtosis of a vector of variables. In particular, when the joint distribution of variables belongs to elliptical family, the asymptotic variance-covariance matrix of correlations can be obtained through multiplying the asymptotic variance-covariance matrix derived under the multivariate normality assumption (Olkin and Siotani, 1976) by a relative kurtosis. This relative kurtosis is equal to 1 when the variables have a multivariate normal distribution. The coefficient of multivariate kurtosis can be estimated by using the algorithm provided by Mardia and Zemroch (1975). According to this asymptotic theory, Steiger and Browne (1984) presented a series of chi-square tests for linear combinations of partial correlations, independent or correlated multiple correlations, and canonical correlations, with or without the assumption of multivariate normality. However, this asymptotic theory is complicated except for particular cases, for example, elliptical families.

It has been known that confidence intervals can provide us a quantitative measure of an effect, not just a qualitative impression, that is, whether the effect is statistically significant. Hence, we are interested in not only whether the differences between two squared multiple correlations is positive or negative, but also the confidence intervals for the difference.

The relationships between any two regression models can be classified into three categories: independent, nested (i.e., all predictors within a model are a part of the predictors of another model), and overlapped but non-nested (i.e., two models have a common subset of predictors). Hence, inferences for differences between two R^2s should be made in the following three cases.

A comparison of R^2s from two models without any common predictors or from two independent populations is the simplest case due to the independence between two R^2s. Based on inference procedures for a single R^2, one can easily extend to inferences for differences between two independent population ρ^2s. Because a single R^2 has a positively skewed density, the distribution of a difference between two independent R^2s is probably still skewed. Hence, the Wald-type interval estimation for the difference of two independent R^2s presented in Olkin and Finn (1995) still has its inherent deficiency (see Algina, 1999), which forces a confidence interval to be symmetric for those parameters whose sample estimates having skewed distributions. Later, Chan (2009) described a bootstrap confidence interval about a difference between two independent R^2s which involves intensive computations. Zou (2007) proposed simple and direct confidence interval construction for a difference between two independent R^2s by using a recovered variance estimate from confidence limits for each R^2. The rationale behind this method is termed the Method of Variance Estimates Recovery (MOVER) (Zou, 2008). Given confidence limits for each of two parameters and their correlation coefficient, as a general approach, the MOVER can be used to construct confidence intervals for the difference, the sum or the ratio of those two parameters. This approach takes into account the skewness of some sampling distributions, so that the MOVER performs well for a wide range of parameters in terms of both coverage rate and interval width, even for small to moderate sample sizes.

Other two cases including the comparisons of two R^2s from nested or non-nested models need to consider the correlation between two R^2s. Olkin and Siotani (1976) derived asymptotic covariance estimates between any two simple correlation coefficients. Using these asymptotic covariance estimates, Olkin and Finn (1995) suggested a Wald-type confidence interval for differences between two correlated squared multiple correlations or a squared partial and a squared multiple correlation. Consider the difference between two squared multiple correlations: first, any difference between two R^2s can be represented as a function of sample simple correlation coefficients, since a multiple correlation can be written as a function of simple correlations; second, given the variance-covariance matrix

for the sample simple correlation coefficients involved in the function, one can obtain the asymptotic variance estimate of the difference by using the delta method; finally, according to the central limit theorem, one can construct a Wald-type confidence interval for this difference between two correlated R^2s. Graf and Alf (1999) improved Olkin and Finn (1995)'s approaches for further simplification. Through a simulation study, Azen and Sass (2008) examined the performance of the Wald-type confidence interval for differences between two R^2s from non-nested models proposed by Olkin and Finn (1995) and found that this asymptotic confidence interval is acceptable in terms of coverage rate only for large sample size (> 200).

1.3 Objective of the thesis

The poor performance of existing confidence interval procedures for differences between two correlated squared multiple correlation coefficients may lead us to wrong conclusions on the relative importance of predictors. The main reason is that the existing Wald-type confidence interval constructions proposed by Olkin and Finn (1995) ignore the potential skewness of the sampling distribution for a comparison of R^2s. Furthermore, this procedure based on the central limit theorem has been shown to be accepted only for large samples (> 200).

The objective of this study is to provide a simple and efficient inference procedure for differences between two correlated R^2s from non-nested models. In particular, under the multivariate normality assumption, we propose a closed-form confidence interval for the comparison of the changes in R^2 when each of two predictors is added to a model with some essential predictors.

Inspired by the good performance of the MOVER proposed by Zou (2008) applied in constructing a confidence interval about a difference between two independent ρ^2 (Zou, 2007), our proposed procedure will employ the MOVER to construct a confidence interval for a difference between two ρ^2s from non-nested models. The MOVER does not force

confidence intervals to be symmetric and also not require the normality assumption of sample estimate of parameter of interest, and thus may improve the performances of the existing confidence intervals for differences between two correlated ρ^2s from non-nested models.

The performance of the proposed confidence interval will also be evaluated and compared to that of the existing Wald-type confidence interval provided by Olkin and Finn (1995).

1.4 Organization of the thesis

In this study, Chapter 2 reviews background on determining the relative importance of predictors, as well as literatures regarding inferences procedures for a single squared multiple regression and differences between two squared multiple regression. Chapter 3 first describes the MOVER proposed in the paper of Zou (2008), then presents a new confidence interval construction for differences between two correlated squared multiple correlations from non-nested models. In Chapter 4, a simulation study compares the performance of our proposed confidence interval to that of the existing Wald-type confidence interval proposed by Olkin and Finn (1995). Finally, Chapter 5 provides some discussion and suggests future work.

Chapter 2

LITERATURE REVIEW

2.1 Introduction

As outlined in Chapter 1, changes in squared multiple correlation may be used to evaluate the relative importance of predictors in multiple linear regression models. In this chapter, we first review predictor importance in the context of multiple linear regressions, then describe some main approaches for constructing confidence intervals for a single R^2 and for differences between two independent and correlated R^2s.

2.2 Predictor importance in multiple linear regressions

With data of a sample of a response variable and a large set of potential predictors, we may build a model using a two-stage process (see Azen et al., 2001; Budescu, 1993). First, we identify a subset of predictors that can adequately describe the relationship between the dependent variable and predictors. This stage is usually referred to as model selection. Once a model is selected, we may proceed to interpret the model by comparisons of predictor importance.

In the model selection stage, there are two general approaches termed *explanation* and *prediction* (Pedhazur, 1982). The explanation approach may also be regarded as the conception of causation. Based on previous theory or substantive research, those predictors which are conceived to be associated with the dependent variable are identified and then included in the model. The prediction approach aims to find the most predictive model, regardless of the underlying mechanism of how the predictors affect the response. A variety of selection principles have been suggested for this purpose, such as the Akaike Informa-

tion Criterion (AIC) (Akaike, 1974) and the Bayesian Information Criterion (BIC) (Akaike, 1977).

Having built a model with a chosen set of predictors, we may then assess a relatively important subset of predictors, or rank the predictors according to their contributions in predicting the response variable. They are the final objective of dominant analysis and commonality analysis, respectively (Azen and Budescu, 2003; Hedges and Olkin, 1981).

2.2.1 Dominant analysis

In order to identify the most important set of predictors from p potential independent variables, we can compare all $2^p - 1$ regression models with each model involving a subset of p predictors and a dependent variable in terms of certain indices. These indices include simple correlations, regression weights, partial and semi-partial correlations, and squared multiple correlations. Among them, the squared multiple correlation R^2 for a multiple regression model is shown to preferable (Budescu, 1993).

Budescu (1993) proposed dominance analysis as a new approach to evaluate the relative importance of predictors in multiple linear regression. Here, dominance is a pairwise relationship, indicating that one predictor dominates another if it is more useful than its competitor in all regressions. Explicitly, given a model with a set of essential predictors, it is defined that a set of additional potential predictors is more important than the other sets, if it increases the model's R^2 more than the other sets do. After estimating the sample squared multiple correlations, R^2s, for all $2^p - 1$ regression models, then comparing all these R^2s, we can identify the relative important set of predictors. In particular, suppose that in a model with X_1 and X_2 as essential predictors, to determine the dominance of X_3 and X_4, we can compare the R^2 increase when X_3 is added to the model to that when X_4 is added.

2.2.2 *Commonality analysis*

The objective of commonality analysis is to partition the variance of the response variable accounted for into the unique and combined contributions a predictor makes. It assesses the individual and collective effects of a set of predictors on a single dependent variable, such as the individual and collective effects of school and social background on educational achievement. Its description can be found in many textbooks on regression analysis in the social sciences, such as Kerlinger and Pedhazur (1973).

After the equality of educational opportunity report (Coleman et al., 1966) created the controversy that there was no general methods of assessing the relative importance of correlated predictors, many papers dealt with methods of assessing or describing the relative contribution of correlated predictors. Commonality analysis was first advocated by Mood (1971) as a tool for developing learning models. It has been used in educational research such as school effect studies (Pedhazur, 1975), teaching studies (Dunkin, 1978). Later, Newton and Spurrell (1969*a*,*b*) developed this technique in industrial sciences and called it "element analysis".

Commonality analyses can be illustrated as follows. Suppose we are given the squared multiple correlation of a model with the response variable Y and two predictors X_1 and X_2, ${\rho_{Y\cdot X_1X_2}}^2$, then we can partition ${\rho_{Y\cdot X_1X_2}}^2$ into three parts: γ_1, γ_2 and γ_{12}. We have

$$\begin{aligned}\gamma_j &= \text{the unique contribution of } X_j \text{ to } {\rho_{Y\cdot X_1X_2}}^2,\ j=1,2,\\ \gamma_{12} &= \text{the common contribution of } X_1 \text{ and } X_2 \text{ to } {\rho_{Y\cdot X_1X_2}}^2,\end{aligned}$$

where the last term γ_{12} is called the commonality of X_1 and X_2. The above definitions lead to the following system of equations:

$$\begin{aligned}{\rho_{Y\cdot X_1X_2}}^2 &= \gamma_1+\gamma_2+\gamma_{12}\\ {\rho_{YX_1}}^2 &= \gamma_1+\gamma_{12},\\ {\rho_{YX_2}}^2 &= \gamma_2+\gamma_{12},\end{aligned}$$

where ${\rho_{YX_1}}^2$ and ${\rho_{YX_2}}^2$ are squared multiple correlation of a model predicting Y from a single predictor X_1 or X_2, respectively.

It was suggested that the rule of selecting predictors should be based on large unique components and small commonalities (Newton and Spurrell, 1969*a*,*b*; Mood, 1971). Hence, from the perspective of commonality analysis, a series of comparisons of R^2s is an important tool of determining predictor importance (Hedges and Olkin, 1981).

2.3 Inference procedures for a single ρ^2

Before presenting procedures for the comparison of R^2s, we describe inference procedures for a single ρ^2. In particular, we start with point estimators of a single population squared multiple correlation ρ^2, including R^2 and other estimators which are designed to reduce the potential bias of R^2 as an estimator of population parameter ρ^2. This parameter also can be estimated in terms of simple and/or partial correlations, which is easily calculated by using a handy calculator for small number of predictors, given the correlation matrix for a dependent variable and predictors. We then introduce five procedures for constructing a confidence interval for ρ^2. These procedures include the Wald-type symmetric confidence interval based on the asymptotic variance, a confidence interval based on the Fisher's exact density of ρ^2 (Fisher, 1928), confidence intervals based on various approximated density of ρ^2, a bootstrap confidence interval, and a confidence interval developed by Gurland (1968) and Helland (1987).

2.3.1 Point estimator of a single ρ^2

Consider a multiple linear regression model

$$\mathbf{y} = \mathbf{1}\beta_0 + \mathbf{x}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

where $\mathbf{y} = (y_1, y_2, \ldots, y_n)'$, a $n \times 1$ vector $\mathbf{1} = (1, 1, \ldots, 1)'$, β_0 and $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)'$ are unknown, $\boldsymbol{\epsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)'$ is a vector of independent errors with mean 0 and variance σ^2

(When regressors are random, the mean and variance of ϵ are conditional on $\mathbf{x}$), and each row of $n \times p$ matrix $\mathbf{x}$ is a sample of a vector consisting of p predictors $\mathbf{X} = (X_1, X_2, \ldots X_p)$,

Let $\bar{Y}$ and $\bar{\mathbf{X}}$ respectively be the sample mean of dependent variable Y and the vector $\mathbf{X}$, and let $\widetilde{\mathbf{x}} = \mathbf{x} - \mathbf{1}\bar{\mathbf{X}}$, then the least square estimators for β_0 and $\boldsymbol{\beta}$ are

$$\begin{aligned} \widehat{\beta}_0 &= \bar{Y} - \bar{\mathbf{X}}\widehat{\boldsymbol{\beta}}, \\ \widehat{\boldsymbol{\beta}} &= \left(\widetilde{\mathbf{x}}'\widetilde{\mathbf{x}}\right)^{-1}\widetilde{\mathbf{x}}'\mathbf{y}, \end{aligned}$$

and the predicted vector is given by

$$\widehat{\mathbf{y}} = \mathbf{1}\widehat{\beta}_0 + \mathbf{x}\widehat{\boldsymbol{\beta}}.$$

The population multiple correlation coefficient ρ^2 can be defined as the correlation between Y and $\mathbf{X}\boldsymbol{\beta}$, since $\mathbf{X}\boldsymbol{\beta}$ is the linear combination of variables $X_1, X_2, \ldots X_p$ that has maximal correlation with Y (Pearson, 1912). Since $Y = \beta_0 + \mathbf{X}\boldsymbol{\beta} + \varepsilon$ and ε is independent of $\mathbf{X}$, according to the definition, ρ^2 can be written as

$$\begin{aligned} \rho^2 &= \frac{[\mathrm{cov}(Y, \mathbf{X}\boldsymbol{\beta})]^2}{\mathrm{var}(Y)\mathrm{var}(\mathbf{X}\boldsymbol{\beta})} \\ &= \frac{\mathrm{var}(\mathbf{X}\boldsymbol{\beta})}{\mathrm{var}(Y)}. \end{aligned}$$

When regressors are random, we usually assume that $\mathbf{X}$ has a multivariate normal distribution with mean μ_x and variance-covariance matrix Σ_x.

$$\rho^2 = \frac{\boldsymbol{\beta}'\Sigma_x\boldsymbol{\beta}}{\sigma^2 + \boldsymbol{\beta}'\Sigma_x\boldsymbol{\beta}}. \tag{2.1}$$

The total, regression and error sums of squares are respectively presented by

$$\begin{aligned} \mathrm{SST} &= \sum_{i=1}^{n} (y_i - \bar{Y})^2 = (\mathbf{y} - \mathbf{1}\bar{Y})'(\mathbf{y} - \mathbf{1}\bar{Y}), \\ \mathrm{SSR} &= \sum_{i=1}^{n} (\widehat{y}_i - \bar{Y})^2 = (\widehat{\mathbf{y}} - \mathbf{1}\bar{Y})'(\widehat{\mathbf{y}} - \mathbf{1}\bar{Y}) = \widehat{\boldsymbol{\beta}}'\widetilde{\mathbf{x}}'\widetilde{\mathbf{x}}\widehat{\boldsymbol{\beta}}, \\ \mathrm{SSE} &= \mathrm{SST} - \mathrm{SSR} = (\mathbf{y} - \widehat{\mathbf{y}})'(\mathbf{y} - \widehat{\mathbf{y}}) = \mathbf{y}'\mathbf{y} - \widehat{\mathbf{y}}'\widehat{\mathbf{y}}, \end{aligned}$$

and the coefficient of determination is defined by

$$R^2 = \frac{\text{SSR}}{\text{SST}}.$$

Because

$$\widehat{\sigma}^2 = \frac{\text{SSE}}{n-p}$$

is a unbiased estimator of σ^2 and

$$\mathbf{S_x} = \frac{1}{n-1}\widetilde{\mathbf{x}}'\widetilde{\mathbf{x}} = \frac{1}{n-1}(\mathbf{x}-\mathbf{1}\bar{\mathbf{X}})'(\mathbf{x}-\mathbf{1}\bar{\mathbf{X}})$$

is an unbiased estimator of Σ_x, then R^2 becomes

$$R^2 = \frac{(n-1)\widehat{\boldsymbol{\beta}}'\mathbf{S_x}\widehat{\boldsymbol{\beta}}}{(n-p)\widehat{\sigma}^2 + (n-1)\widehat{\boldsymbol{\beta}}'\mathbf{S_x}\widehat{\boldsymbol{\beta}}}.$$

Due to the similarity of this representation and the formula (2.1), Helland (1987) concluded that only for random regressors, can R^2 be treated as a estimator of ρ^2. This conclusion is based on the multivariate normality assumption for random regressors, so it is not suitable for models with fixed regressors. Although most articles described the R^2 in computer programmes as an estimator of population ρ^2, Helland (1987) clearly pointed out the precondition of this usage.

R^2 has a positive bias as an estimator of population ρ^2, especially when the number of regressors are moderate or large. To correct for the bias in R^2, at least eight alternative estimators have been proposed in the literature.

1. Smith's estimator $\widehat{\rho}_S^2$ (Ezekiel, 1929) has a form of

$$\widehat{\rho}_S^2 \quad = \quad 1 - \left(1 - R^2\right)\frac{n}{n-p}.$$

   $\widehat{\rho}_S^2$ shrinks a lot, but it increases as more predictors are added.

2. By replacing σ^2 and Σ_x in the formula (2.1) with their respective unbiased estimators $\widehat{\sigma}^2$ and $\mathbf{S_x}$, Wherry (1931) defined $\widehat{\rho}_W^2$ as

$$\widehat{\rho}_W^2 \quad = \quad 1 - \left(1 - R^2\right)\frac{n-1}{n-p}.$$

However, this replacement does not make $\widehat{\rho}^2_W$ unbiased as an estimator of ρ^2; it has been shown that $\widehat{\rho}^2_W$ overestimates ρ^2 (Wherry, 1931).

3. Ezekiel (1930) proposed adjusting R^2 by replacing sum of squares with mean squares in ordinary R^2, giving

$$R^2_{adj} = 1-\left(1-R^2\right)\frac{n-1}{n-p-1}.$$

Huberty and Mourad (1980) found that R^2_{adj} adequately estimated ρ^2. Hence, R^2_{adj} is a commonly used alternative of R^2, and is provided by most statistical computer packages.

4. Based on a random regressor model, Olkin and Pratt (1958) provided an unbiased estimator of ρ^2 as

$$\widehat{\rho}^2_{OP} = 1-\left(1-R^2\right)\frac{n-3}{n-p-1}F\left[1,1;\frac{n-p+1}{2};(1-R^2)\right],$$

where $F(\cdot)$ is the hypergeometric function. This estimator can be written in a closed form as

$$\widehat{\rho}^2_{OP} = 1-\left(1-R^2\right)\frac{n-3}{n-p-1}\left[1+\frac{2(1-R^2)}{n-p+1}+\frac{8(1-R^2)^2}{(n-p-1)(n-p+3)}\right].$$

5. By approximating to the hypergeometric function F, Pratt developed an estimator ρ^2_P, which is also presented in Claudy (1978) as

$$\widehat{\rho}^2_P = 1-\left(1-R^2\right)\frac{n-3}{n-p-1}\left[1+\frac{2(1-R^2)}{n-p-2.3}\right].$$

Claudy (1978) found that ρ^2_P underestimates ρ^2.

6. Herzberg (1969) also gave another approximate estimator

$$\widehat{\rho}^2_H = 1-\left(1-R^2\right)\frac{n-3}{n-p-1}\left[1+\frac{2(1-R^2)}{n-p+1}\right].$$

Comparing to $\widehat{\rho}^2_{OP}$, Claudy (1978) found that ρ^2_H overestimates ρ^2.

7. Claudy (1978) developed an empirical estimate as

$$\widehat{\rho}^2_{CL} = 1-\left(1-R^2\right)\frac{n-4}{n-p-1}\left[1+\frac{2(1-R^2)}{n-p+1}\right],$$

while Raju et al. (1997) found that $\widehat{\rho}^2_{CL}$ is biased for estimating ρ^2.

8. For fixed n, p and R^2, Alf and Graf (2002) obtained the maximum likelihood estimate of ρ^2, $\widehat{\rho}^2_{MLE}$, by maximizing the exact density function of R^2 (Fisher, 1928).

Alf and Graf (2002) compared R^2 and other eight versions of estimates of ρ^2. They found that the estimator provided by Olkin and Pratt (1958) performed well over a wide range of values for n and ρ^2; the adjusted estimator for ρ^2 (Ezekiel, 1930) was unbiased only when the population ρ^2 was 0, and the remaining estimators including R^2 and $\widehat{\rho}^2_{MLE}$ were biased for all the values of n and ρ^2. The above evaluations are based on the assumption that the joint distribution of independent and dependent variables is multivariate normal. If the normality assumption is violated, the evaluations become very complex (see Drasgow and Dorans, 1982).

The multiple correlation coefficient can be represented as a function of simple and partial correlation coefficients. Therefore, once we know the correlation matrix for a response variable and predictors, we can estimate any multiple correlation coefficient.

Let $\rho_{Y\cdot X_1X_2\ldots X_p}$ be a population multiple correlation, which measures the strength of the linear relationship between a dependent variable Y and a set of independent variables $X_1, X_2, \ldots, X_p$. In other words, it also can be explained as the maximum correlation coefficient between Y and all linear combinations of $X_1, X_2, \ldots, X_p$ (Pearson, 1912). Squared multiple correlations can then be mathematically represented as functions of simple and/or partial correlations (Pearson, 1912; Olkin and Siotani, 1976). For example, the coefficients of multiple correlation involving the first two or three variables, denoted as $\rho_{Y\cdot X_1X_2}$ and $\rho_{Y\cdot X_1X_2X_3}$ respectively, are given by

$$\begin{aligned} 1-{\rho_{Y\cdot X_1X_2}}^2 &= \left(1-{\rho_{YX_1}}^2\right)\left(1-{\rho_{YX_2\cdot X_1}}^2\right) \qquad (2.2)\\ 1-{\rho_{Y\cdot X_1X_2X_3}}^2 &= \left(1-{\rho_{YX_1}}^2\right)\left(1-{\rho_{YX_2\cdot X_1}}^2\right)\left(1-{\rho_{YX_3\cdot X_1X_2}}^2\right) \end{aligned}$$

where $\rho_{YX_2 \cdot X_1}$ and $\rho_{YX_3 \cdot X_1X_2}$ are two partial correlations. A partial correlation is also referred to as adjusted correlation. For example, $\rho_{YX_2 \cdot X_1}$ is the correlation of variables Y and X_2, after removing the association of each with X_1, given by

$$\rho_{YX_2 \cdot X_1} = \frac{\rho_{YX_2} - \rho_{YX_1}\rho_{X_2X_1}}{\sqrt{(1-{\rho_{YX_1}}^2)(1-{\rho_{X_2X_1}}^2)}} \tag{2.3}$$

$\rho_{YX_3 \cdot X_1X_2}$ is the partial correlation of variables Y and X_3 keeping variables X_1 and X_2 constant, then

$$\begin{aligned} \rho_{YX_3 \cdot X_1X_2} &= \frac{\rho_{YX_3 \cdot X_2} - \rho_{YX_1 \cdot X_2}\rho_{X_3X_1 \cdot X_2}}{\sqrt{(1-{\rho_{YX_1 \cdot X_2}}^2)(1-{\rho_{X_3X_1 \cdot X_2}}^2)}} \\ &= \frac{\rho_{YX_3 \cdot X_1} - \rho_{YX_2 \cdot X_1}\rho_{X_3X_2 \cdot X_1}}{\sqrt{(1-{\rho_{YX_2 \cdot X_1}}^2)(1-{\rho_{X_3X_2 \cdot X_1}}^2)}} \end{aligned}$$

Combining equations 2.2 and 2.3, we have

$$\begin{aligned} {\rho_{Y \cdot X_1X_2}}^2 &= {\rho_{YX_1}}^2 + {\rho_{YX_2 \cdot X_1}}^2 \left(1-{\rho_{YX_1}}^2\right) \\ &= \frac{{\rho_{YX_1}}^2 + {\rho_{YX_2}}^2 - 2\rho_{YX_1}\rho_{YX_2}\rho_{X_1X_2}}{1-{\rho_{X_1X_2}}^2} \end{aligned}$$

The relationships among correlations also can be easily represented in matrix form (Olkin and Siotani, 1976). Let $\Sigma(S) = \Sigma(s_1, s_2, \ldots, s_m)$ and $\widehat{\Sigma}(S) = \widehat{\Sigma}(s_1, s_2, \ldots, s_m)$ represent the population and sample correlation matrices for variables $X_{s_1}, \ldots, X_{s_m}$, respectively. Let ${\rho_{Y \cdot X_\alpha}}^2$ be the population squared multiple correlation between Y and a set of variables with subscripts α; ${\rho_{YX_1 \cdot X_\alpha}}^2$ be the population squared partial correlation between Y and X_1 for fixed variables with subscripts α; ${\rho_{YX_\alpha \cdot X_\beta}}^2$ be the population squared partial-multiple correlation between Y and a set of variables with subscripts α while keeping a set of variables with subscripts β fixed, then

$$\begin{aligned} {\rho_{Y \cdot X_\alpha}}^2 &= 1 - \frac{|\Sigma(0,\alpha)|}{|\Sigma(\alpha)|} \\ {\rho_{YX_1 \cdot X_\alpha}}^2 &= \frac{|\Sigma(0,1,\alpha)|}{\sqrt{|\Sigma(0,\alpha)| \cdot |\Sigma(1,\alpha)|}} \\ {\rho_{YX_\alpha \cdot X_\beta}}^2 &= \frac{|\Sigma(\beta)| \cdot |\Sigma(0,\alpha,\beta)|}{|\Sigma(\alpha)| \cdot |\Sigma(0,\beta)|} \end{aligned}$$

where α and β are two sets of subscripts, $|\Sigma(\alpha)|$ is the determinant of the population correlation matrix for variables with subscripts α, $|\Sigma(0,\alpha)| = \left|(\rho_{ij})\right|$ is the determinant of the population correlation matrix for the response variable Y and variables with subscripts α. Similar definitions for other determinants of population correlation matrices.

Note that these formula are also suitable for sample correlations by replacing ρ and Σ with sample correlation coefficients r and matrix $\widehat{\Sigma}$, respectively.

2.3.2 Confidence interval construction for a single ρ^2

There are several approaches to constructing confidence interval for a single ρ^2. These approaches can be classified into five categories. Herein, these five categories of approaches are presented respectively as follows. Among these five approaches, a confidence interval developed by Gurland (1968) and Helland (1987) based on a scaled central F approximations to the density of $R^2/(1-R^2)$ performs well.

2.3.2.1 Wald-type method

The variance of R^2 was originally derived by Wishart (1931), given as

$$\operatorname{var}(R^2) = \frac{4\rho^2\left(1-\rho^2\right)^2(n-p-1)^2}{\left(n^2-1\right)(n+3)}.$$

When n is large, the variance can be approximated by

$$\operatorname{var}(R^2) \approx \frac{4}{n}\rho^2\left(1-\rho^2\right)^2\left(1-\frac{2p+5}{n}\right).$$

If $2p+5$ is small relative to n, the expression reduces further to

$$\operatorname{var}(R^2) \approx \frac{4\rho^2\left(1-\rho^2\right)^2}{n}. \tag{2.4}$$

Hence, $100(1-\alpha)\%$ confidence limits for ρ^2 by using the Wald method is given by $R^2 \pm z_{\alpha/2}\sqrt{\widehat{\operatorname{var}}(R^2)}$, where $\widehat{\operatorname{var}}(R^2)$ is the sample estimate of $\operatorname{var}(R^2)$ with ρ^2 replaced by R^2, $z_{\alpha/2}$ is the $100\cdot\alpha/2$ upper percentile point of a standard normal distribution.

From equation (2.4), we can see that when $\rho^2 = 0$, the variance estimate of R^2 is 0, so that the Wald method cannot be used. The method is more suitable for cases with symmetric sampling distributions, but not for ρ^2, since the sampling distribution of R^2 is skewed and converges very slowly to normality (Algina, 1999). Furthermore, this method may give confidence limits out of range of 0 to 1.

2.3.2.2 Fisher's R^2-to-z transformation

Another method of interval estimation for ρ^2 is based on Fisher's z transformation of R^2 rather than r (see Olkin and Finn, 1995). Fisher'z statistic

$$z = \log\left(\frac{1+R}{1-R}\right) = \log\left(\frac{1+\sqrt{R^2}}{1-\sqrt{R^2}}\right)$$

is approximately normally distributed with mean $E(z) = \zeta = \log[(1+\sqrt{\rho^2})/(1-\sqrt{\rho^2})]$ and variance $4/n$ (Algina, 1999). Then the confidence limits of z, l_z and u_z, have forms

$$\log\left(\frac{1+\sqrt{R^2}}{1-\sqrt{R^2}}\right) \pm z_{\alpha/2}\sqrt{4/n}.$$

Since z is a monotone increasing function of R^2, if both l_z and u_z are non-negative then the confidence limits for ρ^2 are given by

$$\left(\frac{\exp(l_z)-1}{\exp(l_z)+1}\right)^2, \quad \left(\frac{\exp(u_z)-1}{\exp(u_z)+1}\right)^2$$

and if $l_z < 0$ then the lower confidence limit for ρ^2 is 0.

Note that the confidence interval based on Fisher's z transformation performs more poorly than others (Algina, 1999). The main reason is that the limiting distribution of Fisher's transformation on R^2 failed to approach normality asymptotically (Gajjar, 1967). As R^2 ranges from 0 to 1, the transformed values of Fisher's z only covers the range from 0 to plus infinity. Therefore, Alf and Graf (1999) pointed out that, "Fisher's z values are severely truncated in the lower tail, resulting in a distribution that is even more positively skewed than is the original distribution of squared multiple correlations" (p. 74). In addition, the approximate variance $4/n$ for z breaks down when ρ^2 is near 0 (Algina, 1999).

2.3.2.3 *Exact method*

Based on the geometrical interpretation of multiple correlations, Fisher (1928) gave the exact density function of R^2, which is later written in the following form (Lee, 1972)

$$\begin{aligned} f_{R^2}(x) &= B(p/2, n_1/2)^{-1} \left(1-\rho^2\right)^{(n-1)/2} (x)^{p/2-1} (1-x)^{n_1/2-1} \\ & \quad F\left((n-1)/2, (n-1)/2; p/2; \rho^2 x\right) \end{aligned} \tag{2.5}$$

where $B(.)$ and $F(.)$ respectively denote the beta and Gaussian hypergeometric functions, $n_1 = n-p-1$, n being the sample size.

We know that both beta and Gaussian hypergeometric functions are complicated integrands or series. There is also no closed formula for calculating the exact cumulative distribution function of R^2. Therefore, the calculation of exact confidence limits requires iterations on computing the percentiles of the integral of density $f_{R^2}(x)$. Based on this distribution, Kramer (1963) and Lee (1972) tabulated only upper percentage points of the distribution of ρ^2 and limited to confidence levels 5% and 1% in particular cases. A stand-alone program, "R2", available at `http://www.statpower.net` , has been developed for obtaining the confidence limits based on this distribution (Steiger and Fouladi, 1992).

2.3.2.4 *Approximation methods based on the density of R^2*

Due to the complexity of the exact density of R^2, many researchers developed various methods of obtaining approximate confidence intervals for ρ^2, through asymptotically expanding or approximating the density function of R^2. Several asymptotic procedures have appeared in the literature, including a two-moments scaled central F approximation proposed by Khatri (1966), a three-moments scaled noncentral F approximation and a three-moments relocated and rescaled central F approximation proposed by Lee (1971).

The development is based on $\frac{R^2}{1-R^2}$ denoted as $\widetilde{R}^2$. The distribution of $\widetilde{R}^2$ can be represented as (Lee, 1971)

$$\widetilde{R}^2 = \frac{(\widetilde{\rho}\chi_{n-1} + z)^2 + \chi^2_{p-1}}{\chi^2_{n-p-1}}$$

where $\widetilde{\rho}$ is the square root of $\widetilde{\rho}^2 = \rho^2/(1-\rho^2)$, n denotes the sample size, z is a standard normal variable, χ_f and χ_f^2 are respectively chi and chi-square distributed variables having f degrees of freedom, and all included variables are independent from each other.

Through expansion of the characteristic function of the variable $(\widetilde{\rho}\chi_{n-1}+z)^2+\chi_{p-1}^2$, the density function of $\widetilde{R}^2$ can be asymptotically represented as a linear combination of non-central F distributed variables, and the density function of R^2 can be rewritten approximately in terms of non-central beta distributed variables (Lee, 1971). Furthermore, the non-central F distribution is reasonably well approximated by the central F distribution. Hence, $\widetilde{R}^2$ may be further approximated in terms of a noncentral or even central F-distributed variable.

2.3.2.4.1 A scaled central F approximation

Khatri (1966) proposed approximating the distribution of $\widetilde{R}^2$ by a scaled noncentral or central F distribution. Later, both Gurland (1968) and Helland (1987) described and examined the scaled central F approximation. The unknown scale coefficient and degree of freedom of the scaled central F variable are determined by fitting the first two moments of the numerator of $\widetilde{R}^2$. Derived from this F approximation, one can obtain an asymptotic confidence interval for the population ρ^2 by using an iterative procedure. The following combines all the findings from the papers of Gurland (1968) and Helland (1987).

Given the multivariate normality assumption for random regressors, $\widetilde{R}^2$ also can be represented as the ratio of a noncentral chi-square distributed variable $\chi_{p-1}^2(\Delta)$ having degrees of freedom $p-1$ and noncentrality parameter Δ, and a central chi-squared distributed variable χ_{n-p-1}^2 having degrees of freedom $n-p-1$ independently of the numerator $\chi_{p-1}^2(\Delta)$, that is,

$$\widetilde{R}^2 = \frac{\chi_{p-1}^2(\Delta)}{\chi_{n-p-1}^2}. \tag{2.6}$$

The noncentrality parameter Δ is given by

$$\Delta = \frac{\beta'\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\beta}{\sigma^2}, \tag{2.7}$$

from which we can see that the noncentrality parameter highly depends on the design matrix $\mathbf{X}$. For fixed regressors, Δ is a constant, and $\widetilde{R}^2$ follows a noncentral F distribution with degrees of freedom $p-1$ and $n-p-1$, and a constant noncentrality parameter Δ. When the regressors are random, Δ is a random variable which has the same distribution as the variable $\widetilde{\rho}^2\chi^2_n$, so the variable $\widetilde{R}^2$ no longer follows a noncentral F distribution, except when $\rho = 0$.

According to the definition of ρ^2 presented in equation (2.1), equation (2.7) becomes

$$\Delta = \frac{\boldsymbol{\beta}'\Sigma_x\boldsymbol{\beta}}{\sigma^2}\chi^2_{n-1} = \frac{\rho^2}{1-\rho^2}\chi^2_{n-1}. \tag{2.8}$$

Inserting equation (2.8) into the numerator of $\widetilde{R}^2$ shown in equation (2.6), one can obtain the exact density function of $\widetilde{R}^2$ as given by Fisher (1928).

Given random regressors, one can easily approximate the numerator of $\widetilde{R}^2$ by a scaled central chi-square distributed variable, that is, $\chi^2_p(\Delta)$ is approximated by $a\chi^2_v$, where the constants

$$a = \frac{(n-1)\widetilde{\rho}^2(\widetilde{\rho}^2+2)+p}{(n-1)\widetilde{\rho}^2+p}$$

and

$$v = \frac{\left[(n-1)\widetilde{\rho}^2+p\right]^2}{(n-1)\widetilde{\rho}^2(\widetilde{\rho}^2+2)+p}$$

are determined by using methods of moment, that is, through equating the first two moments of $a\chi^2_v$ to those of $\chi^2_p(\Delta)$.

By using this approximation, we have

$$\begin{aligned}
\widetilde{R}^2 &\approx \frac{a\chi^2_v}{\chi^2_{n-p-1}} \\
&= \frac{av}{n-p-1}F_{v,n-p-1} \\
&= \frac{(n-p-1)\rho^2+p}{(n-p-1)(1-\rho^2)}F_{v,n-p-1},
\end{aligned} \tag{2.9}$$

where $F_{v,n-p-1}$ has a central F distribution with v and $n-p-1$ degrees of freedom. This asymptotic form of $\widetilde{R}^2$ is a monotone increasing function of R^2, so we can obtain the following approximate $100(1-\alpha)\%$ confidence interval for ρ^2,

$$\left[\rho_L^2, \rho_U^2\right],$$

where the lower limit is

$$\rho_L^2 = \frac{(n-p-1)R^2 - \left(1-R^2\right) pF_{1-\alpha/2;v,(n-p-1)}}{(n-p-1)\left[R^2 + (1-R^2)F_{1-\alpha/2;v,(n-p-1)}\right]} \tag{2.10}$$

and the upper limit is

$$\rho_U^2 = \frac{(n-p-1)R^2 - \left(1-R^2\right) pF_{\alpha/2;v,(n-p-1)}}{(n-p-1)\left[R^2 + (1-R^2)F_{\alpha/2;v,(n-p-1)}\right]}, \tag{2.11}$$

in which, $F_{\alpha/2;v,(n-p-1)}$ and $F_{1-\alpha/2;v,(n-p-1)}$ are the $100\alpha/2$ and $100(1-\alpha/2)$ percentile points of the central F distribution with v and $n-p-1$ degrees of freedom, respectively. Although the parameter v is still a function of ρ^2, one can calculate it by applying an iterative procedure with the sample estimate R^2 as the starting value. The inference procedure based on the scaled central F approximation can be implemented by SAS PROC CANCORR with the option SMC (which stands for squared multiple correlations), given a series of observations on the outcome and predictors.

When $\rho_L^2 = \rho_U^2 = 0$, $v = p$ and equation (2.9) reduces to a commonly used statistic for testing the null hypothesis $H_0 : \rho^2 = 0$, namely

$$\widetilde{R}^2 = \frac{p}{(n-p-1)} F_{p,n-p-1},$$

which depends on a central F distributed variable. It implies that the usual test for the hypothesis is consistent with the test found from the asymptotical confidence interval for ρ^2.

From equations (2.10) and (2.11), it follows that when $F_{\alpha/2} < 1 < F_{1-\alpha/2} = 1/F_{\alpha/2}$, the confidence interval $\left[\rho_L^2, \rho_U^2\right]$ always covers the adjusted coefficient of determination R_{adj}^2.

Moreover, it has been shown by Helland (1987) that this approximate confidence interval for ρ^2 performs very well, compared to those from the tabulated results by Lee (1972).

2.3.2.4.2 A scaled noncentral F approximation

Similar to the approach for the scaled central F approximation, Lee (1971) proposed to approximate $\widetilde{R}^2$ by a scaled noncentral F distributed variable or a relocated and rescaled central F distributed variable.

Lee (1971) approximated the numerator of $\widetilde{R}^2$, $(\widetilde{\rho}\chi_{n-1}+z)^2+\chi^2_{p-1}$, by a scaled noncentral chi-squared variable $g\chi^2_w(\lambda)$, where g, w, λ are determined by equating the first three moments of the numerator to those of variate $g\chi^2_w(\lambda)$, having the following forms

$$\begin{aligned} g &= \left[\phi_2-\sqrt{\phi_2^2-\phi_1\phi_3}\right]/\phi_1, \\ w &= \left[\phi_2-2\widetilde{\rho}^2\gamma\sqrt{(n-1)(n-p-1)}\right]/g^2, \\ \lambda &= \widetilde{\rho}^2\gamma\sqrt{(n-1)(n-p-1)}/g^2, \end{aligned}$$

where

$$\gamma^2 = 1/(1-\rho^2)$$

and

$$\phi_j = (n-1)\left(\gamma^{2j}-1\right)+p, \quad j=1,2,3. \tag{2.12}$$

Hence, $\widetilde{R}^2$ may be approximated by a scaled noncentral F variate, which satisfies

$$\widetilde{R}^2 \approx \frac{gw}{n-p-1}F_{w,n-p-1}(\lambda). \tag{2.13}$$

The confidence interval for ρ^2 derived from this scaled noncentral F approximation can be obtained by using a bisection method implemented by statistical software programs. For example, a SAS macro and SPSS syntax for constructing an asymptotic confidence interval based on the scaled noncentral F approximation by Zou (2007) are available at `http://dx.doi.org/10.1037/1082-989x.12.4.399.supp` . The program requires as input only the sample estimate R^2, the sample size and the number of predictors.

2.3.2.4.3 A relocated and rescaled central F approximation

$\widetilde{R}^2$ also can be approximated by a relocated and rescaled central F variate, which satisfies

$$\widetilde{R}^2 \approx c(F_{q,n-p-1} + a), \tag{2.14}$$

where q, a and c are determined by equating the first three moments and can be written as

$$\begin{aligned} q &= \frac{1}{2}(n-p-3)[\sqrt{E/(E-4)} - 1], \\ c &= \frac{q}{(n-p-1)(n-p+2q-3)}(H/K), \\ a &= [(\phi_1/c) - (n-p)]/(n-p-2), \end{aligned}$$

where

$$\begin{aligned} H &= 2\phi_1^3 + 3\phi_1\phi_2(n-p-3) + \phi_3(n-p-3)^2, \\ K &= \phi_1^2 + \phi_2(n-p-3), \\ E &= H^2/K^3 \end{aligned}$$

and $\phi_j, j = 1,2,\ldots$ is defined by equation (2.12).

Similarly, one can construct an asymptotic confidence interval for ρ^2 by using a bisection method based on this relocated and rescaled central F approximation.

Lee (1971) evaluated the accuracy of these three approximations by comparing the absolute differences between each approximation and the exact density of R^2 in some particular cases. These three approximations are the scaled central F approximation presented in equation (2.9) (Gurland, 1968), the scaled noncentral F approximation presented in equation (2.13) and the relocated and rescaled central F approximation presented in equation (2.14) (Lee, 1971). The results showed that the three-moment scaled noncentral F approximation performs better than other two approximations but its evaluation is more complicated due to a noncentrality parameter. The relocated and rescaled central F approximation performs well only when ϕ_1 is large. Hence, the scaled central F approximation seems to be a simple and good procedure to construct an asymptotic confidence interval for a single ρ^2.

2.3.2.5 *Bootstrap method*

The bootstrap methods have also been applied to construct confidence intervals for ρ^2 (e.g., Ohtani, 2000).

The bootstrap is a nonparametric approach obtained by applying a series of intensive computer-based re-sampling to approximate the sampling distribution of an estimate of a parameter of interest θ. Then, a confidence interval for this parameter based on the bootstrap can be constructed by calculating its confidence limits from the bootstrap distribution of the sample estimate of the parameter θ.

There are several bootstrap methods for constructing confidence intervals. The ones in common use are percentile, bias-corrected method (BC), bias-corrected and accelerated (BCa), bootstrap-*t*, and approximate bootstrap confidence intervals (ABC).

As a simple and direct bootstrap method, the $\alpha/2$ and $1-\alpha/2$ percentiles of the bootstrap distribution are respectively taken as the $100(1-\alpha)$ percentile bootstrap lower and upper confidence limits. However, the percentile bootstrap method performs well only for unbiased statistics having a symmetric sampling distribution Schenker (1985).

To correct the percentile interval for median bias, the BC method (Efron, 1981) constructed a confidence interval based on the similarity between the histogram of bootstrap replications of standardized θ and a standard normal density function. This method assumes that there exists a monotonic increasing function g such that $g(\hat{\theta})-g(\theta)$ has a normal distribution with stabilized variance for all θ. It also has been shown that the BC method breaks down even for moderate sample sizes (Schenker, 1985).

As an improvement, Efron (1987) proposed the BCa method, which incorporates an extra constant related to the skewness of the sampling distribution, termed an accelerated constant. The BCa method requires that only the function $g(\cdot)$ to be a normalizing transformation, not necessary variance stabilizing. The performance of the BCa bootstrap confidence interval highly depends on the accuracy of estimating the acceleration. However, there are also no simple approaches to accurately estimating the bias-correction factor

and acceleration constant, even if one can show the existence of the transformation (Shao, 1995). Furthermore, the length of the BCa interval does not increase as the size of the significance level increases (Hall, 1992, pp 134-135, 137).

The bootstrap-t method (Efron, 1979) aims to estimate the percentiles of the standardized θ by bootstrapping. It is conceptually simpler than the BCa method, and also has a better second order properties. But it only works well under an accurate estimate of the standard error of a statistic of interest. It is also numerically unstable, sometimes yielding very long confidence intervals (DiCiccio and Efron, 1996).

The ABC method is an analytic version of BCa, only applicable for smoothly defined parameters in exponential families (Efron and Tibshirani, 1993). It does not involve the bootstrap cumulative distribution function, but it requires an estimate of the nonlinearity parameter.

All these bootstrap confidence intervals require complex and intensive computation, hence, as Efron (1988) stated, the bootstrap can be an alternative method only when there is no suitable parametric methods.

2.4 Inference procedures for differences between two ρ^2s

Comparisons of squared multiple correlations may arise in the following three cases (Alf and Graf, 1999). Case 1 represents comparisons of two independent R^2s; case 2 represents comparisons of two R^2s arising from nested models, and case 3 represents comparisons of two R^2s arising two non-nested models.

Case 1: Examining if a given set of predictors performs equally in two separate populations or groups: ${\rho_I}^2 - {\rho_{II}}^2$. This comparison shows whether a given set of predictors $(X_1, X_2, \ldots, X_k)$ performs equally well in two separate, independent populations.

Case 2: Determining whether an additional predictor provides an significant improvement in prediction of the response: ${\rho_{12}}^2 - {\rho_1}^2$. It means whether an additional variable X_2 provides improvement over X_1 alone in predicting y. Equivalently, the equality of the

squared multiple correlation ${\rho_{12}}^2$ and the squared simple correlation ${\rho_1}^2$ is commonly tested with an F statistic by comparing a full model with a reduced model in regression analysis.

Case 3: Deciding which of two predictors adds significantly more to the model already including a set of predictors: ${\rho_{12}}^2 - {\rho_{13}}^2$. This comparison shows which pair of predictors X_1 and X_2 or X_1 and X_3 is more effective in predicting outcome Y.

2.4.1 Case 1: Differences between two independent R^2s

Due to independence, Case 1 is the simplest among the three cases. There exists a large amount of literature on statistical inferences for differences between two independent R^2. For example, Olkin and Finn (1995) suggested two inference procedures. Both of them are constructed by using the Wald method through equating the sum of the variance estimates of ${R_I}^2$ and ${R_{II}}^2$ from two independent populations to the variance estimate of ${R_I}^2 - {R_{II}}^2$. For the first method, each variance of ${R_I}^2$ and ${R_{II}}^2$ is estimated by using equation (2.4). Therefore, a Wald-type confidence interval can be constructed for changes in population squared multiple correlation $\Delta\rho^2$ from two separate populations. The second method applies Fisher's z transformation on ${R_I}^2$ and ${R_{II}}^2$. Each transformed variable has a variance of $4/n$, then the variance of the difference between Fisher's z transformation on ${R_I}^2$ and Fisher's z transformation on ${R_{II}}^2$ is $2/n$. However, this method is not applicable for constructing a confidence interval for ${\rho_I}^2 - {\rho_{II}}^2$, and can only serve to test the hypothesis $H_0 : {\rho_I}^2 = {\rho_{II}}^2$, because the difference between two Fisher's z transformations on ρ^2s is not a monotone function of ${\rho_I}^2 - {\rho_{II}}^2$.

Simulation results (Algina and Keselman, 1999) show that the coverage rates of these two confidence intervals were inadequate for the unequal allocation of the sample sizes and the multiple correlation coefficients in two populations. These two methods also required larger sample sizes when the number of predictors are large.

Chan (2009) suggested a confidence interval for this case based on bootstrap methods as an alternative, which performs well for normal data and non-normal data as well, but

involving a large amount of computations. Moreover, the performance of the confidence interval based on bootstrap methods is also not good for small sample sizes (*e.g.*, n=50) (Chan, 2009).

Zou (2007) proposed a simple confidence interval construction for differences between two independent R^2 by using recovery variance estimators differently from the lower and upper limits for each R^2. This approach can account for the skewness of the sampling distribution of R^2. Simulation results have shown that this method performs better than the Wald method (Zou, 2007).

2.4.2 Case 2: Differences between two ρ^2s from nested models

In Case 2, the confidence interval construction for an increase in squared multiple correlations, ${\rho_{\alpha,\beta}}^2 - \rho_\alpha^2$, also termed as squared semi-partial correlations, is discussed previously, in which, α, β denote subsets of subscripts of predictors.

The best way to describe an unknown parameter of interest is to obtain the density function of the parameter. However, when comparing two R^2s, it is hard to obtain the exact joint distribution of the corresponding two population squared multiple correlations, because the exact marginal distribution for a multiple correlation (Fisher, 1928) is already extremely complex. Many researchers provide asymptotic solutions for the joint distribution of multiple correlations. For example, Olkin and Siotani (1976) provided the asymptotic distribution of functions of correlation matrices. Hedges and Olkin (1981) extended their results to obtain the asymptotic joint distribution of all $2^p - 1$ squared multiple correlations. Hedges and Olkin (1983) further obtained the asymptotic joint distributions of any sets of partial, multiple and partial-multiple correlations. From the central limit theorem, all these joint distributions are approximately multivariate normal. However, the sampling distribution of a single R^2 has been known to be so skewed that the asymptotical multivariate normality of the joint distribution of two or more R^2s is suitable only for very large samples.

It has been well known that a confidence interval for a unknown parameter is the most important index when making statistical inferences on this parameter. Olkin and Finn

(1995) suggested Wald-type confidence intervals for differences between two ρ^2s, which is constructed through directly estimating the variance of the difference between R^2s.

Herein, the Wald-type confidence interval for ${\rho_{\alpha,\beta}}^2 - {\rho_\alpha}^2$ is given by

$$ {R_{\alpha,\beta}}^2 - {R_\alpha}^2 \pm z_{\alpha/2}\widehat{\sigma_1}/\sqrt{n}, $$

where σ_1 is the asymptotic standard error of the difference ${R_{\alpha,\beta}}^2 - {R_\alpha}^2$ and $\widehat{\sigma_1}$ is the sample estimate of σ_1 with population correlations replaced by their corresponding sample correlations, given by

$$ \widehat{\sigma_1} \quad = \quad \sqrt{\widehat{\text{var}}\left({R_{\alpha,\beta}}^2 - {R_\alpha}^2\right)}. $$

The standard error estimator $\widehat{\sigma_1}$ can be obtained using the delta method as suggested by Olkin and Finn (1995).

Specifically, assuming that a difference between two R^2s denoted as ΔR^2 is written as a function (f) of a vector of simple correlations related to k predictors $\mathbf{r} = (r_1, r_2, \ldots, r_k, r_{12}, \ldots, r_{1k}, \ldots, r_{k-1,k})$, where r_i is the correlation of y and X_i, and r_{ij} is the correlation of X_i and X_j . That is, $\Delta R^2 = f(\mathbf{r})$. Applying the delta methods,

$$ \text{var}\,(\Delta R^2) = \mathbf{a}\Phi_{1,\ldots,k}\mathbf{a}' $$

where matrix $\Phi_{1,\ldots,k}$ is the variance and covariance matrix of $\mathbf{r}$, vector $\mathbf{a}$ consists of partial derivatives of the form

$$ \mathbf{a} = \frac{df}{d\mathbf{r}} = \left(\frac{\partial f}{\partial r_1}, \frac{\partial f}{\partial r_2}, \ldots, \frac{\partial f}{\partial r_{k-1,k}}\right). $$

Hence, estimating the variance-covariance matrix of a vector of sample simple correlations $\Phi_{1,\ldots,k}$, and calculating the partial derivatives in $\mathbf{a}$, are required.

First, let us consider the estimation of the variance-covariance matrix of sample simple correlations. Pearson and Filon (1898) first derived the asymptotic covariances between any two sample simple correlation coefficients. Similar results have also been derived by

Hotelling (1953) and Olkin and Siotani (1976). These asymptotic variances and covariances among sample correlations are presented as follows: In general, the covariance of two correlations r_{ij} and r_{kl} without any variables in common is given by

$$\begin{aligned}\operatorname{cov}(r_{ij}, r_{kl}) &= \frac{1}{n}\left[\frac{1}{2}\rho_{ij}\rho_{kl}\left({\rho_{ik}}^2+{\rho_{il}}^2+{\rho_{jk}}^2+{\rho_{jl}}^2\right)+\rho_{ik}\rho_{jl}+\rho_{il}\rho_{jk}\right. \\ &\quad \left.-\left(\rho_{ij}\rho_{ik}\rho_{il}+\rho_{ji}\rho_{jk}\rho_{jl}+\rho_{ki}\rho_{kj}\rho_{kl}+\rho_{li}\rho_{lj}\rho_{lk}\right)\right].\end{aligned}$$

For cases with a common variable X_i, the expression reduces to

$$\operatorname{cov}(r_{ij}, r_{ik}) = \frac{1}{n}\left[\frac{1}{2}\left(2\rho_{jk}-\rho_{ij}\rho_{ik}\right)\left(1-{\rho_{ij}}^2-{\rho_{ik}}^2-{\rho_{jk}}^2\right)+\rho_{jk}^3\right],$$

which further reduces to

$$\operatorname{var}(r_{ij}) = \frac{1}{n}\left(1-{\rho_{ij}}^2\right)^2,$$

when $k = j$.

To easily calculate $\mathbf{a}$, Graf and Alf (1999) provided computer programmes having analytic derivatives substituted by numerical derivatives by using the method of numerical differentiation (Scarborough, 1966). Through representing a multiple correlation as a simple correlation of a response variable and a weighted sum of involved predictors, Alf and Graf (1999) further improved Graf and Alf (1999)'s computation procedures by presenting a simpler form of asymptotic confidence limits for $\Delta\rho^2$.

Although Olkin and Finn (1995) noticed that the Wald confidence interval for a single ρ^2 is not applicable due to its forced symmetry, Olkin and Finn (1995)'s procedures still provide a symmetric Wald-type confidence interval for differences between two ρ^2s, while the sampling distribution of ΔR^2 is probably skewed. Furthermore, since Olkin and Finn (1995)'s approaches are based on the central limit theorem, it is not surprising that they found that their procedure is good only for large samples ($n > 200$). Hence, a better interval estimation for ΔR^2 is expected.

2.4.3 *Case 3: Differences between two ρ^2s from non-nested models*

The aboved approach can also be applied to the case of differences between two ρ^2s from non-nested models, i.e., ${\rho_{\alpha,\beta}}^2 - {\rho_{\alpha,\gamma}}^2$, in which α, β and γ denote subsets of subscripts of predictors. The differences can also be represented as a comparison of two squared semi-partial correlations, which satisfies

$$ {\rho_{\alpha,\beta}}^2 - {\rho_{\alpha,\gamma}}^2 = \left({\rho_{\alpha,\beta}}^2 - {\rho_{\alpha}}^2\right) - \left({\rho_{\alpha,\gamma}}^2 - {\rho_{\alpha}}^2\right) $$

Thus, the Wald-type confidence interval for $\Delta\rho^2$ from non-nested models is written as

$$ {R_{\alpha,\beta}}^2 - {R_{\alpha,\gamma}}^2 \pm z_{\alpha/2}\widehat{\sigma_2}/\sqrt{n} $$

where

$$ \widehat{\sigma_2} \;=\; \sqrt{\widehat{\text{var}}\left({R_{\alpha,\beta}}^2 - {R_{\alpha,\gamma}}^2\right)}. $$

Similarly, after representing $\Delta\rho^2$ as a function of simple correlations, the Wald-type confidence interval for $\Delta\rho^2$ can be obtained by applying the procedure proposed by Olkin and Finn (1995) presented in last section.

Azen and Budescu (2003) assessed qualitatively the stability of dominance relationships in terms of ΔR^2 across repeated sampling from the bootstrapping. Azen and Sass (2008) investigated the performance of the asymptotic procedure proposed by Olkin and Finn (1995) for comparing the R^2s from non-nested models. They found that the Wald-type confidence interval for $\Delta\rho^2$ from non-nested models provides the expected coverage rates for large sample sizes ($n > 200$) but not for small and moderate samples.

2.5 Summary

In this review chapter, we started with the basic task of determining the relative importance of predictors in multiple regression models. We pointed out that the basic approach is to construct a good confidence interval for the difference of two population squared multiple correlations ($\Delta\rho^2$) from non-nested models.

Among a large number of confidence interval estimations for a single ρ^2, we showed that a simple and reliable construction is the one based on the density of $R^2/(1-R^2)$ approximated by a scaled central F distribution proposed by Gurland (1968) and Helland (1987).

In contrast, there is little literature available on inference procedures for $\Delta\rho^2$ from non-nested models. Although Olkin and Finn (1995) proposed a general approach to confidence interval construction for various $\Delta\rho^2$, it has been shown that this asymptotic procedure provides the expected coverage rates only for large sample sizes ($n > 200$). This conclusion is not surprising, because the asymptotic procedure proposed by Olkin and Finn (1995) is derived from the central limit theorem, and uses a Wald-type confidence interval. It has been known that a Wald-type confidence interval performs poorly due to its forced symmetry. Furthermore, the exact joint distribution of squared multiple correlation coefficients is intractable. The asymptotic joint distribution of $\Delta\rho^2$ presented by Hedges and Olkin (1983) is so complex that it is hard to obtain an analytic form of confidence interval estimation for $\Delta\rho^2$.

Therefore, our purpose here is to propose a new method for constructing a confidence interval for $\Delta\rho^2$ from non-nested models. It is expected that the resulting confidence interval procedure will perform well in practical sample sizes. Based on the good performance of the method of variance estimate recovery (MOVER) proposed by Zou and Donner (2008) for other applications, it is expected that MOVER is applicable for providing a better confidence interval estimation on $\Delta\rho^2$ from non-nested models than the Wald method.

Chapter 3

CONFIDENCE INTERVAL FOR A DIFFERENCE BETWEEN TWO SQUARED MULTIPLE CORRELATION COEFFICIENTS FROM NON-NESTED MODELS

3.1 Introduction

As we have discussed in last Chapter, the existing Wald-type confidence interval procedure for changes in R-squared from non-nested models may not perform well, because it enforces symmetry to the sampling distribution for two R^2s even though they are left skewed. To deal with this issue, we use the method of variance estimates recovery, known as the MOVER (Zou, 2008). The MOVER approach recovers variance estimates from the lower and upper limits separately and thus does not require the sampling distribution to be symmetric. We begin this chapter with a description of the MOVER.

3.2 The MOVER

The MOVER is a general approach to constructing confidence intervals for simple functions of parameters, including sums, differences and ratios. In contrast to the Wald-type methods, it relaxes the symmetry assumption for the sampling distribution. The central idea of the MOVER is to obtain the variance estimator of each parameter component separately for the lower and upper confidence limits of the parameter in their corresponding neighborhoods. Hence, this method only requires the confidence limits for each parameter component and the correlation coefficient estimates between any two estimators of the parameter components. Note that, we use the term "MOVER" (method of variance estimates

recovery) named by Zou (2008), rather than other terms such as 'modified large sample' (Burdick and Graybill, 1992), 'square-and-add' (Newcombe, 2011), because the key step is to recover the variances estimates from confidence limits for each of two or more parameter components.

As pointed out by Zou and Donner (2008), ideas similar to the MOVER can be traced back to Howe (1974) who applied the Cornish-Fisher expansion to obtain approximate confidence limits for the sum of two independent normal means. Newcombe (1998) used the same idea for constructing a confidence interval for a difference between two proportions by first obtaining separate confidence limits for single proportions using the Wilson procedure. However, neither these two articles give analytic justification for its general applicability.

Zou (2007) derived a closed-form expression for the confidence interval for a difference between two parameters constructed by using recovery variance estimators separately from the lower and upper limits for each parameter. The results were applied to differences between two correlations and two independent R^2s. Simulation results show that the proposed procedure performs much better than the traditional Wald method in terms of both overall coverage and tail errors.

The MOVER has been extended to constructing confidence intervals for a ratio of two parameters (Li et al., 2010) and any linear combination of parameters (Newcombe, 2011). Here, we summarize the rationale of the MOVER and its extensions and applications as follow.

First, consider the construction of an approximate two-sided $100(1-\alpha)\%$ confidence interval (L,U) for a sum $\theta_1+\theta_2$, given $\widehat{\theta}_i$ and (l_i,u_i), $i=1,2$, as point estimator and a two-sided $100(1-\alpha)\%$ confidence interval for θ_1 and θ_2, respectively, we have according to the central limit theorem,

$$\begin{aligned}(l_i,u_i) &= \widehat{\theta}_i \mp z_{\alpha/2}\sqrt{\widehat{\mathrm{var}}(\widehat{\theta}_i)}, \quad i=1,2,\\ (L,U) &= \widehat{\theta}_1+\widehat{\theta}_2 \mp z_{\alpha/2}\sqrt{\mathrm{var}(\widehat{\theta}_1+\widehat{\theta}_2)}\end{aligned}$$

$$= \quad \widehat{\theta}_1 + \widehat{\theta}_2 \mp z_{\alpha/2} \left[\text{var}\,(\widehat{\theta}_1) + \text{var}\,(\widehat{\theta}_2) + 2\text{cov}\,(\widehat{\theta}_1, \widehat{\theta}_2) \right]^{1/2}, \tag{3.1}$$

where $z_{\alpha/2}$ is the upper $\alpha/2$ quantile of the standard normal distribution.

Conventional Wald-type methods assume that the variance $\text{var}\,(\widehat{\theta}_i)$ is constant for all values of θ_i, so resulting in a symmetric confidence interval. To improve the performance of conventional methods which do not account for the potential asymmetry of the underlying sampling distributions of $\widehat{\theta}_1$ and $\widehat{\theta}_2$, Zou (2008) proposed to obtain the variance estimator $\widehat{\text{var}}\,(\widehat{\theta}_i)$ at the neighborhood of the confidence limits L and U separately. Confidence limits l_i and u_i can be regarded as the minimum and maximum values of $\theta_1 + \theta_2$ that satisfy

$$\frac{[(\widehat{\theta}_1 + \widehat{\theta}_2) - (\theta_1 + \theta_2)]^2}{\widehat{\text{var}}\,(\widehat{\theta}_1 + \widehat{\theta}_2)} < z^2_{\alpha/2}.$$

It is easy to show that the distance between $l_1 + l_2$ and L given by

$$z_{\alpha/2} \left\| \sqrt{\text{var}\,(\widehat{\theta}_1 + \widehat{\theta}_2)} - \left[\sqrt{\text{var}\,(\widehat{\theta}_1)} + \sqrt{\text{var}\,(\widehat{\theta}_2)} \right] \right\|$$

is smaller than that between the point estimator $\widehat{\theta}_1 + \widehat{\theta}_2$ and L, which is given by

$$z_{\alpha/2} \left\| \sqrt{\text{var}\,(\widehat{\theta}_1 + \widehat{\theta}_2)} \right\|.$$

Likewise, $u_1 + u_2$ is closer to U than $\widehat{\theta}_1 + \widehat{\theta}_2$. Therefore, it is reasonable to estimate the variance $\text{var}\,(\widehat{\theta}_i)$ at $\theta_i = l_i$ and $\theta_i = u_i$, $i = 1, 2$, respectively for obtaining L and U.

As shown in Figure 3.1, the variance estimate $\widehat{\text{var}}\,(\widehat{\theta}_i)$ is equal to

$$\widehat{\text{var}}\,(\widehat{\theta}_i) = (\widehat{\theta}_i - l_i)^2 / z^2_{\alpha/2}$$

at $\theta_i = l_i$, and

$$\widehat{\text{var}}\,(\widehat{\theta}_i) = (u_i - \widehat{\theta}_i)^2 / z^2_{\alpha/2}$$

at $\theta_i = u_i$. Furthermore, the correlation between $\widehat{\theta}_1$ and $\widehat{\theta}_2$ can be estimated as

$$\widehat{\rho} = \frac{\text{cov}\,(\widehat{\theta}_1, \widehat{\theta}_2)}{\sqrt{\widehat{\text{var}}\,(\widehat{\theta}_1)\widehat{\text{var}}\,(\widehat{\theta}_2)}} \tag{3.2}$$

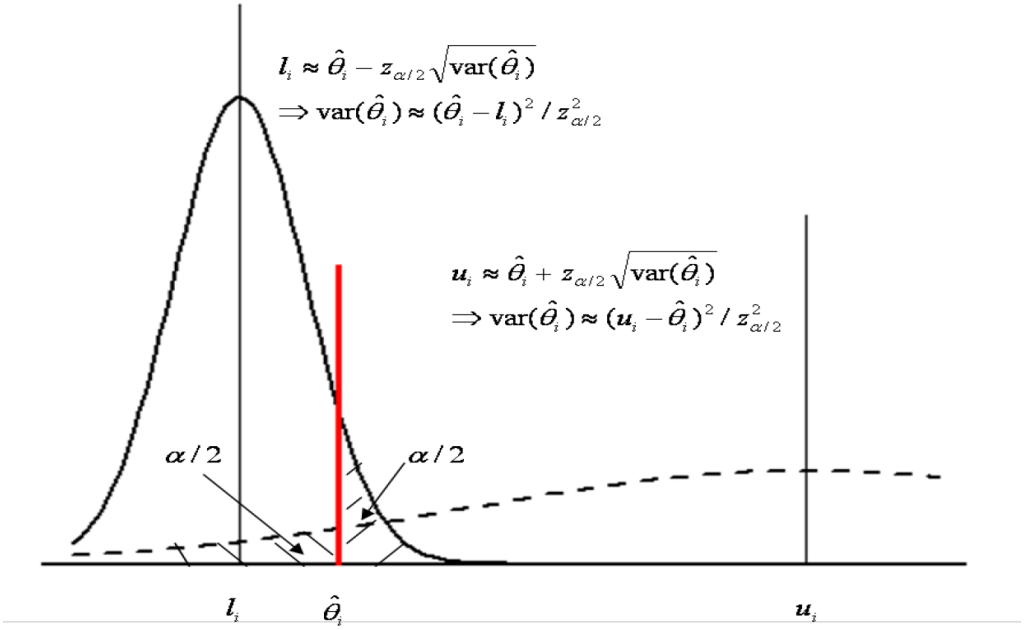


Figure 3.1: Confidence limits l_i and u_i for θ_i. The point estimator $\widehat{\theta}_i$ is the $100\cdot(1-\alpha/2)$ quantile of the normal distribution with mean l_i and variance $(\widehat{\theta}_i - l_i)^2/z^2_{\alpha/2}$, and is also the $100\cdot\alpha/2$ quantile of the normal distribution with mean u_i and variance $(u_i - \widehat{\theta}_i)^2/z^2_{\alpha/2}$.

Substituting the corresponding variance estimates at $\theta_1 = l_1$ and $\theta_2 = u_2$ for L and at $\theta_1 = u_1$ and $\theta_2 = l_2$ for U into equations (3.1) and (3.2), we have

$$\begin{aligned} L &= \widehat{\theta}_1 + \widehat{\theta}_2 - z_{\alpha/2}\sqrt{\widehat{\text{var}}(\widehat{\theta}_1) + \widehat{\text{var}}(\widehat{\theta}_2) + 2\widehat{\rho}\sqrt{\widehat{\text{var}}(\widehat{\theta}_1)\widehat{\text{var}}(\widehat{\theta}_2)}} \\ &= \widehat{\theta}_1 + \widehat{\theta}_2 - z_{\alpha/2}\sqrt{(\widehat{\theta}_1 - l_1)^2/z^2_{\alpha/2} + (\widehat{\theta}_2 - l_2)^2/z^2_{\alpha/2} + 2\widehat{\rho}(\widehat{\theta}_1 - l_1)(\widehat{\theta}_2 - l_2)/z^2_{\alpha/2}} \\ &= \widehat{\theta}_1 + \widehat{\theta}_2 - \sqrt{(\widehat{\theta}_1 - l_1)^2 + (\widehat{\theta}_2 - l_2)^2 + 2\widehat{\rho}(\widehat{\theta}_1 - l_1)(\widehat{\theta}_2 - l_2)} \end{aligned} \tag{3.3}$$

and similarly,

$$U = \widehat{\theta}_1 + \widehat{\theta}_2 + \sqrt{(u_1 - \widehat{\theta}_1)^2 + (u_2 - \widehat{\theta}_2)^2 + 2\widehat{\rho}(u_1 - \widehat{\theta}_1)(u_2 - \widehat{\theta}_2)} \tag{3.4}$$

In fact, the Wald method being applied for the sum of two parameters is a special case of the MOVER when the confidence limits of each parameter are obtained by using the

Wald method. This also highlights the superiority of the MOVER: It acknowledges the asymmetric nature of the sampling distributions for a single parameter, whereas the Wald method ignores this fact.

This method can be applied for constructing the confidence interval for a difference between two parameters (Zou, 2007). The difference $\theta_1 - \theta_2$ can be rewritten as $\theta_1 + (-\theta_2)$, while the confidence interval for $-\theta_2$ is $(-u_2, -l_2)$, through replacing θ_2, l_2 and u_2 in equations (3.3) and (3.4) respectively by $-\theta_2$, $-u_2$ and $-l_2$, then the confidence limits for $\theta_1 - \theta_2$ are:

$$L = \widehat{\theta_1} - \widehat{\theta_2} - \sqrt{(\widehat{\theta_1} - l_1)^2 + (u_2 - \widehat{\theta_2})^2 - 2\widehat{\rho}(\widehat{\theta_1} - l_1)(u_2 - \widehat{\theta_2})} \quad (3.5)$$

$$U = \widehat{\theta_1} - \widehat{\theta_2} + \sqrt{(u_1 - \widehat{\theta_1})^2 + (\widehat{\theta_2} - l_2)^2 - 2\widehat{\rho}(u_1 - \widehat{\theta_1})(\widehat{\theta_2} - l_2)} \quad (3.6)$$

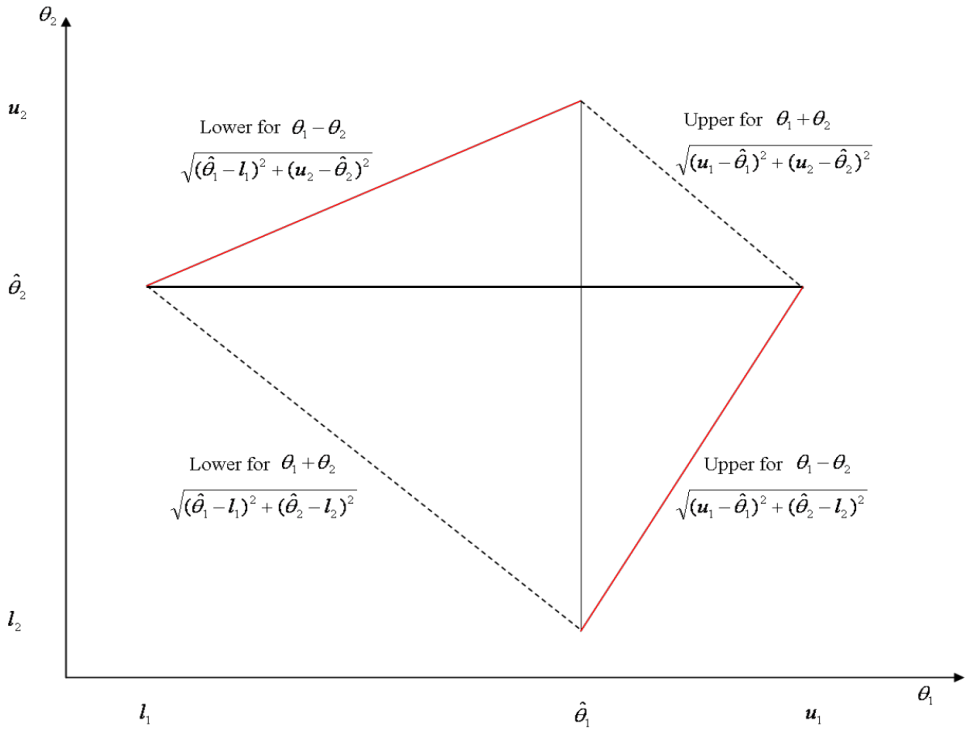


Figure 3.2: Geometric illustration of margins of errors as obtained using the MOVER for $\theta_1 - \theta_2$ and $\theta_1 + \theta_2$ which are identical to that by the Pythagorean theorem.

When $\widehat{\theta}_1$ and $\widehat{\theta}_2$ are independent, i.e., $\rho = 0$, the MOVER can be geometrically illustrated in Figure 3.2. Note that the lower and upper confidence limits of a parameter are written as [*point estimate-lower margin of error, point estimate+upper margin of error*]. Thus, the margins of errors for the difference $\theta_1 - \theta_2$ and the sum $\theta_1 + \theta_2$ obtained by using the MOVER are identical to that by using the Pythagorean theorem. In general, these confidence limits are asymmetric, unless confidence limits for θ_i, $i = 1, 2$, are symmetric. It also reflects that the MOVER does not enforce symmetry for confidence interval construction for a parameter of interest.

The above procedures can also be generalized to obtain the confidence interval for any a linear combination of K parameters $\sum_{i=1}^{K} c_i\theta_i$ (Zou, 2008), when the sample estimators of $\theta_1, \ldots, \theta_K$ are independent, given by

$$\begin{aligned} L &= \sum_{i=1}^{K} c_i\widehat{\theta}_i - \sqrt{\sum_{i=1}^{K}\left[c_i\widehat{\theta}_i - \min(c_i l_i, c_i u_i)\right]^2} \\ U &= \sum_{i=1}^{K} c_i\widehat{\theta}_i + \sqrt{\sum_{i=1}^{K}\left[c_i\widehat{\theta}_i - \max(c_i l_i, c_i u_i)\right]^2} \end{aligned}$$

For a ratio of two parameters θ_1/θ_2, there are two approaches to constructing its confidence interval. One way is only applicable to positive parameter components and simply use a logarithmic transformation (Zou and Donner, 2008). That is, constructing the confidence interval for $\log(\theta_1) - \log(\theta_2)$ is done by applying the MOVER then exponentiating the limits. A more general way is to apply a generalization of Fieller's theorem proposed by Zou and Donner (2010), that is, a confidence interval for $R = \theta_1/\theta_2$ can be constructed by setting the resulting confidence limits for $\theta_1 - R\theta_2$ obtained by using the MOVER to be zero, because R must satisfy $\theta_1 - R\theta_2 = 0$ to be a ratio of θ_1/θ_2.

The MOVER has been widely applied in confidence interval construction for parameters in many fields. It has been applied for a difference between two parameters including kappa statistics (Donner and Zou, 2002), product-moment correlations and independent R^2s (Zou, 2007), dependent intraclass correlation coefficients (Ramasundarahettige et al., 2009), normal means (Wang and Chow, 2002) and lognormal means (Zou, Huo and Tale-

ban, 2009).

It also has been applied to simple functions of parameters, including effect measures such as the relative risk (Zou and Donner, 2008) and that obtained by using counterfactuals (Zou, 2010), additive interaction (Zou, 2008), lognormal means (Zou and Donner, 2008; Zou, Huo and Taleban, 2009) and other lognormal data (Zou, Taleban and Huo, 2009), linear functions of binomial proportions (Zou, Huang and Zhang, 2009) or negative binomial proportions under inverse sampling (Zou, 2010), functions of normal means and standard deviations (including the area under the receiver operating characteristic curve for quantifying the ability of a biomarker to correctly classify individuals into two groups) (Li et al., 2010), the Bland-Altman limits of agreement with multiple observations per individual as a standard for assessing agreement between different methods measuring the same quantity (Zou, 2011), the normal distribution percentiles, the coefficient of variation and Cohen's effect size (Donner and Zou, 2010). It has also been extended to conducting simultaneous confidence intervals for multiple contrasts of proportions (Donner and Zou, 2011).

3.3 Application of MOVER to differences between two ρ^2s from non-nested models

The purpose of this study is to construct a confidence interval for differences between two R^2s from non-nested models by using the MOVER. However, we first have to obtain confidence limits for each R^2 and the correlation coefficient between the two R^2s.

As mentioned in last chapter, an asymptotic confidence interval for a squared multiple correlation coefficient ρ^2 can be constructed by using a scaled central F distribution (Gurland, 1968; Helland, 1987) or a scaled noncentral F distribution (Lee, 1971) approximated to the density of $R^2/(1-R^2)$.

For a multiple regression model with n samples and p predictors, if the variable $\widetilde{R}^2 = R^2/(1-R^2)$ is assumed to be approximated to be scaled central F distributed, we have

$$\widetilde{R}^2 \approx \frac{(n-p-1)\rho^2+p}{(n-p-1)(1-\rho^2)}F_{v,n-p-1}.$$

Thus an asymptotic confidence interval for ρ^2, $\left[\rho_L^2, \rho_U^2\right]$, is given by

$$\rho_L^2 = \frac{(n-p-1)R^2 - \left(1-R^2\right) pF_{1-\alpha/2;v,(n-p-1)}}{(n-p-1)\left[R^2 + (1-R^2)F_{1-\alpha/2;v,(n-p-1)}\right]} \tag{3.7}$$

$$\rho_U^2 = \frac{(n-p-1)R^2 - \left(1-R^2\right) pF_{\alpha/2;v,(n-p-1)}}{(n-p-1)\left[R^2 + (1-R^2)F_{\alpha/2;v,(n-p-1)}\right]}, \tag{3.8}$$

where

$$v = \frac{\left[(n-1)\widetilde{\rho}^2 + p\right]^2}{(n-1)\widetilde{\rho}^2(\widetilde{\rho}^2+2)+p} \tag{3.9}$$

and $F_{\alpha/2;v,(n-p-1)}$ and $F_{1-\alpha/2;v,(n-p-1)}$ are respectively the $100\cdot\alpha/2$ and $100\cdot(1-\alpha/2)$ percentile points of the central F distribution having v and $n-p-1$ degrees of freedom. Since the parameter v is still a function of ρ^2, the inference procedure requires an iterative process between equations 3.7 (or 3.8) and 3.9. Given a series of observations on the outcome and predictors, the whole process is easily implemented by SAS PROC CANCORR with the option SMC.

The variable $\widetilde{R}^2$ can also be assumed to be approximately scaled noncentral F distributed, given by

$$\widetilde{R}^2 \approx \frac{gw}{n-p-1} F_{w,n-p-1}(\lambda),$$

where

$$\begin{aligned}
g &= \left[\phi_2 - \sqrt{{\phi_2}^2 - \phi_1\phi_3}\right]/\phi_1, \\
w &= \left[\phi_2 - 2\widetilde{\rho}^2\gamma\sqrt{(n-1)(n-p-1)}\right]/g^2, \\
\lambda &= \widetilde{\rho}^2\gamma\sqrt{(n-1)(n-p-1)}/g^2, \\
\widetilde{\rho}^2 &= \rho^2/(1-\rho^2) \\
\gamma &= 1/(1-\rho^2) \\
\phi_j &= (n-1)\left(\gamma^{2j}-1\right)+p, \quad j=1,2,3.
\end{aligned}$$

The above steps required to obtain a confidence interval for ρ^2 has been implemented by Zou (2007) in a SAS macro and an SPSS syntax publicly available at http://dx.

doi.org/10.1037/1082-989x.12.4.399.supp . The programmes only need to input the sample estimate R^2, the sample size and the number of predictors.

Assume that we are interested in inference for the difference between two squared multiple correlation coefficients between the outcome y and different sets of predictors from non-nested models, ${\rho_{\alpha,\beta}}^2 - {\rho_{\alpha,\gamma}}^2$, in which α, β and γ are three sets of subscripts of predictors.

According to the definition, the point estimates of ${\rho_{\alpha,\beta}}^2$ and ${\rho_{\alpha,\gamma}}^2$, denoted as ${R_{\alpha,\beta}}^2$ and ${R_{\alpha,\gamma}}^2$ can be calculated by

$$
\begin{aligned}
{R_{\alpha,\beta}}^2 &= 1 - \frac{\left|\widehat{\Sigma}(0,\alpha,\beta)\right|}{\left|\widehat{\Sigma}(\alpha,\beta)\right|}, && (3.10)\\
{R_{\alpha,\gamma}}^2 &= 1 - \frac{\left|\widehat{\Sigma}(0,\alpha,\gamma)\right|}{\left|\widehat{\Sigma}(\alpha,\gamma)\right|}, && (3.11)
\end{aligned}
$$

where $\widehat{\Sigma}$ indicates the sample estimate of population correlation matrices Σ. For example, $|\widehat{\Sigma}(0,\alpha,\beta)|$ is the determinant of the matrix of sample correlations among the response variable y and predictors with subscripts α and β, and $|\widehat{\Sigma}(\alpha,\beta)|$ is the determinant of the matrix of sample correlations among predictors with subscripts α and β.

By using the delta method, the covariance between these two sample squared multiple correlation coefficients from non-nested models is given by (Hedges and Olkin, 1983)

$$
\begin{aligned}
\operatorname{cov}\left({R_{\alpha,\beta}}^2, {R_{\alpha,\gamma}}^2\right) = {} & \frac{\operatorname{cov}\left(|\widehat{\Sigma}(0,\alpha,\beta)|, |\widehat{\Sigma}(0,\alpha,\gamma)|\right)}{|\Sigma(\alpha,\beta)||\Sigma(\alpha,\gamma)|} \\
& + \frac{|\Sigma(0,\alpha,\beta)||\Sigma(0,\alpha,\gamma)|}{|\Sigma(\alpha,\beta)|^2|\Sigma(\alpha,\gamma)|^2}\operatorname{cov}\left(|\widehat{\Sigma}(\alpha,\beta)|, |\widehat{\Sigma}(\alpha,\gamma)|\right) \\
& - \frac{|\Sigma(0,\alpha,\beta)|}{|\Sigma(\alpha,\beta)|^2|\Sigma(\alpha,\gamma)|}\operatorname{cov}\left(|\widehat{\Sigma}(\alpha,\beta)|, |\widehat{\Sigma}(0,\alpha,\gamma)|\right) \\
& - \frac{|\Sigma(0,\alpha,\gamma)|}{|\Sigma(\alpha,\beta)||\Sigma(\alpha,\gamma)|^2}\operatorname{cov}\left(|\widehat{\Sigma}(0,\alpha,\beta)|, |\widehat{\Sigma}(\alpha,\gamma)|\right) \qquad (3.12)
\end{aligned}
$$

Furthermore, it is easy to show that any matrix of correlations among variables with sub-

scripts ξ, $\Sigma(\xi)$, satisfies

$$\begin{aligned}\frac{\partial|\Sigma(\xi)|}{\partial\rho_{ij}} &= |\Sigma(\xi)|\rho^{ij}, \quad \text{if} \quad i,j\in\xi,\\ \frac{\partial|\Sigma(\xi)|}{\partial\rho_{ij}} &= 0, \qquad \text{if} \quad i,j\notin\xi,\end{aligned}$$

where ρ^{ij} is the (i,j)th element of the inverse of the correlation matrix $\Sigma(\xi)=(\rho_{ij})$, and the asymptotic covariance between two sample simple correlations (Olkin and Siotani, 1976) is given by

$$\begin{aligned}\operatorname{cov}(r_{ij},r_{kl}) &= \frac{1}{n}\Big[\frac{1}{2}\rho_{ij}\rho_{kl}\left({\rho_{ik}}^2+{\rho_{il}}^2+{\rho_{jk}}^2+{\rho_{jl}}^2\right)+\rho_{ik}\rho_{jl}+\rho_{il}\rho_{jk}\\ &\quad -\left(\rho_{ij}\rho_{ik}\rho_{il}+\rho_{ji}\rho_{jk}\rho_{jl}+\rho_{ki}\rho_{kj}\rho_{kl}+\rho_{li}\rho_{lj}\rho_{lk}\right)\Big].\end{aligned}$$

Thus the covariance terms between two determinants of sample correlation matrices can be estimated as (Hedges and Olkin, 1981)

$$\begin{aligned}\operatorname{cov}\left(|\widehat{\Sigma}(\xi)|,|\widehat{\Sigma}(\eta)|\right) &= \sum_{i,j\in\xi}\sum_{k,l\in\eta}\frac{\partial|\Sigma(\xi)|}{\partial\rho_{ij}}\frac{\partial|\Sigma(\eta)|}{\partial\rho_{kl}}\operatorname{cov}(r_{ij},r_{kl})\\ &= \frac{1}{n}|\Sigma(\xi)||\Sigma(\eta)|\sum_{i,j\in\xi}\sum_{k,l\in\eta}\rho^{ij}\rho^{kl}\\ &\quad \Big\{\frac{1}{2}\rho_{ij}\rho_{kl}({\rho_{ik}}^2+{\rho_{il}}^2+{\rho_{jk}}^2+{\rho_{jl}}^2)+\rho_{ik}\rho_{jl}+\rho_{il}\rho_{jk}\\ &\quad -\rho_{ij}\rho_{ik}\rho_{il}-\rho_{ji}\rho_{jk}\rho_{jl}-\rho_{ki}\rho_{kj}\rho_{kl}-\rho_{li}\rho_{lj}\rho_{lk}\Big\}\\ &= \frac{2}{n}|\Sigma(\xi)||\Sigma(\eta)|\operatorname{tr}\left\{\left(\Sigma(\xi)^{-1}-I\right)\Psi(\xi,\eta)\right.\\ &\quad \left.\left(\Sigma(\eta)^{-1}-I\right)\Psi'(\xi,\eta)\right\},\end{aligned}\tag{3.13}$$

where ρ^{ij} and ρ^{kl} are respectively the elements of the inverse of two population correlation matrices $\Sigma(\xi)=(r_{ij})_{i,j\in\xi}$ and $\Sigma(\eta)=(r_{kl})_{k,l\in\eta}$, $\Psi(\xi,\eta)=(r_{st})_{s\in\xi,t\in\eta}$, 'tr' denotes the trace and I is an identity matrix.

By choosing ξ and η from four sets of subscripts $\{0,\alpha,\beta\}$, $\{\alpha,\beta\}$, $\{0,\alpha,\gamma\}$ and $\{\alpha,\gamma\}$, and plugging equation (3.13) into equation (3.12), the covariance $\operatorname{cov}\left({R_{\alpha,\beta}}^2,{R_{\alpha,\gamma}}^2\right)$ can be written as a function of simple correlation coefficients among the outcome and predictors. Its sample estimate $\widehat{\operatorname{cov}}\left({R_{\alpha,\beta}}^2,{R_{\alpha,\gamma}}^2\right)$ also can be obtained with all population correlations (ρ) replaced by sample correlations (r).

As an illustration, considering a difference between two population squared multiple correlations from non-nested models, ${\rho_1}^2 - {\rho_2}^2$, in which ${\rho_1}^2$ is from a model with outcome y and two predictors x_1 and x_2, and ${\rho_2}^2$ is from a model with the identical outcome y and predictor x_1 but a different additional predictor x_3. Given the sample correlation matrix for a vector (x_1, x_2, x_3, y) denoted as $\widehat{\Sigma} = (r_{ij})_{4\times 4}$. According to equations (3.10) and (3.11), ${R_1}^2$ and ${R_2}^2$ are given by

$$\begin{aligned} {R_1}^2 &= 1 - \frac{|\widehat{\Sigma}(1,2,4)|}{|\widehat{\Sigma}(1,2)|} \\ &= \frac{{r_{14}}^2 + {r_{24}}^2 - 2r_{14}r_{24}r_{12}}{1 - {r_{12}}^2}, \qquad (3.14) \\ {R_2}^2 &= 1 - \frac{|\widehat{\Sigma}(1,3,4)|}{|\widehat{\Sigma}(1,3)|} \\ &= \frac{{r_{14}}^2 + {r_{34}}^2 - 2r_{14}r_{34}r_{13}}{1 - {r_{13}}^2}. \qquad (3.15) \end{aligned}$$

Applying equations (3.12) and (3.13) results in

$$\begin{aligned} \widehat{\mathrm{cov}}\left({R_1}^2, {R_2}^2\right) &= \frac{\widehat{\mathrm{cov}}\left(|\widehat{\Sigma}(1,2,4)|, |\widehat{\Sigma}(1,3,4)|\right)}{|\widehat{\Sigma}(1,2)||\widehat{\Sigma}(1,3)|} \\ &\quad + \frac{|\widehat{\Sigma}(1,2,4)||\widehat{\Sigma}(1,3,4)|}{|\widehat{\Sigma}(1,2)|^2|\widehat{\Sigma}(1,3)|^2}\widehat{\mathrm{cov}}\left(|\widehat{\Sigma}(1,2)|, |\widehat{\Sigma}(1,3)|\right) \\ &\quad - \frac{|\widehat{\Sigma}(1,2,4)|}{|\widehat{\Sigma}(1,2)|^2|\widehat{\Sigma}(1,3)|}\widehat{\mathrm{cov}}\left(|\widehat{\Sigma}(1,2)|, |\widehat{\Sigma}(1,3,4)|\right) \\ &\quad - \frac{|\widehat{\Sigma}(1,3,4)|}{|\widehat{\Sigma}(1,2)||\widehat{\Sigma}(1,3)|^2}\widehat{\mathrm{cov}}\left(|\widehat{\Sigma}(1,2,4)|, |\widehat{\Sigma}(1,3)|\right) \\ &= \frac{(1 + 2r_{14}r_{24}r_{12} - {r_{12}}^2 - {r_{14}}^2 - {r_{24}}^2)(1 + 2r_{14}r_{24}r_{13} - {r_{13}}^2 - {r_{14}}^2 - {r_{34}}^2)}{n(1 - {r_{12}}^2)(1 - {r_{13}}^2)} \\ &\quad \left(\sum_{i,j\in\{1,2,4\}} - \sum_{i,j\in\{1,2\}}\right)\left(\sum_{k,l\in\{1,3,4\}} - \sum_{k,l\in\{1,3\}}\right) r^{ij} r^{kl} \\ &\quad \left\{\frac{1}{2} r_{ij}r_{kl}({r_{ik}}^2 + {r_{il}}^2 + {r_{jk}}^2 + {r_{jl}}^2) + r_{ik}r_{jl} + r_{il}r_{jk}\right. \\ &\quad \left. - r_{ij}r_{ik}r_{il} - r_{ji}r_{jk}r_{jl} - r_{ki}r_{kj}r_{kl} - r_{li}r_{lj}r_{lk}\right\} \end{aligned}$$

$$
\begin{aligned}
= \quad & \frac{2(1+2r_{14}r_{24}r_{12}-{r_{12}}^2-{r_{14}}^2-{r_{24}}^2)(1+2r_{14}r_{24}r_{13}-{r_{13}}^2-{r_{14}}^2-{r_{34}}^2)}{n(1-{r_{12}}^2)(1-{r_{13}}^2)} \\
& \left\{\text{tr}\left[\left(\Sigma(1,2,4)^{-1}-I_{3\times 3}\right)\Psi(\{1,2,4\},\{1,3,4\})\right.\right. \\
& \left.\left(\Sigma(1,3,4)^{-1}-I_{3\times 3}\right)\Psi'(\{1,2,4\},\{1,3,4\})\right]+ \\
& \text{tr}\left[\left(\Sigma(1,2)^{-1}-I_{2\times 2}\right)\Psi(\{1,2\},\{1,3\})\right. \\
& \left.\left(\Sigma(1,3)^{-1}-I_{2\times 2}\right)\Psi'(\{1,2\},\{1,3\})\right]- \\
& \text{tr}\left[\left(\Sigma(1,2,4)^{-1}-I_{3\times 3}\right)\Psi(\{1,2,4\},\{1,3\})\right. \\
& \left.\left(\Sigma(1,3)^{-1}-I_{2\times 2}\right)\Psi'(\{1,2,4\},\{1,3\})\right]- \\
& \text{tr}\left[\left(\Sigma(1,2)^{-1}-I_{2\times 2}\right)\Psi(\{1,2\},\{1,3,4\})\right. \\
& \left.\left.\left(\Sigma(1,3,4)^{-1}-I_{3\times 3}\right)\Psi'(\{1,2\},\{1,3,4\})\right]\right\}.
\end{aligned}
$$

Since the asymptotic variance of R^2 is given by

$$
\text{var}\left(R^2\right) \approx \frac{4\rho^2\left(1-\rho^2\right)^2}{n},
$$

then the correlation between ${R_1}^2$ and ${R_2}^2$ can be estimated as

$$
\begin{aligned}
\widehat{\text{corr}}\left({R_1}^2,{R_2}^2\right) &= \frac{\widehat{\text{cov}}\left({R_1}^2,{R_2}^2\right)}{\sqrt{\text{var}\left({R_1}^2\right)\text{var}\left({R_2}^2\right)}} \\
&= \frac{\sqrt{(1-{r_{12}}^2)(1-{r_{13}}^2)}}{2\sqrt{({r_{14}}^2+{r_{34}}^2-2r_{14}r_{34}r_{13})({r_{14}}^2+{r_{24}}^2-2r_{14}r_{24}r_{12})}} \\
& \quad \left\{\text{tr}\left[\left(\Sigma(1,2,4)^{-1}-I_{3\times 3}\right)\Psi(\{1,2,4\},\{1,3,4\})\right.\right. \\
& \quad \left.\left(\Sigma(1,3,4)^{-1}-I_{3\times 3}\right)\Psi'(\{1,2,4\},\{1,3,4\})\right]+ \\
& \quad \text{tr}\left[\left(\Sigma(1,2)^{-1}-I_{2\times 2}\right)\Psi(\{1,2\},\{1,3\})\right. \\
& \quad \left.\left(\Sigma(1,3)^{-1}-I_{2\times 2}\right)\Psi'(\{1,2\},\{1,3\})\right]- \\
& \quad \text{tr}\left[\left(\Sigma(1,2,4)^{-1}-I_{3\times 3}\right)\Psi(\{1,2,4\},\{1,3\})\right. \\
& \quad \left.\left(\Sigma(1,3)^{-1}-I_{2\times 2}\right)\Psi'(\{1,2,4\},\{1,3\})\right]- \\
& \quad \text{tr}\left[\left(\Sigma(1,2)^{-1}-I_{2\times 2}\right)\Psi(\{1,2\},\{1,3,4\})\right. \\
& \quad \left.\left.\left(\Sigma(1,3,4)^{-1}-I_{3\times 3}\right)\Psi'(\{1,2\},\{1,3,4\})\right]\right\}. \qquad (3.16)
\end{aligned}
$$

Now, ${R_1}^2$ and ${R_2}^2$ can be calculated by formula (3.14) and (3.15). Based on a scaled central or non-central F approximation, we can obtain the asymptotic confidence limits for each ${\rho_i}^2$ denoted as (l_i, u_i) , $i = 1, 2$. Moreover, the correlation between ${R_1}^2$ and ${R_2}^2$ also can be estimated by equation (3.16). The next step is to generate a confidence interval for $\Delta\rho^2 = {\rho_1}^2 - {\rho_2}^2$ using the MOVER. Applying the formula (3.5), we have the confidence limits for ${\rho_1}^2 - {\rho_2}^2$ given by

$$
\begin{aligned}
L &= {R_1}^2 - {R_2}^2 - \sqrt{({R_1}^2 - l_1)^2 + (u_2 - {R_2}^2)^2 - 2\widehat{\text{corr}}\left({R_1}^2, {R_2}^2\right)({R_1}^2 - l_1)(u_2 - {R_2}^2)}, \\
U &= {R_1}^2 - {R_2}^2 + \sqrt{(u_1 - {R_1}^2)^2 + ({R_2}^2 - l_2)^2 - 2\widehat{\text{corr}}\left({R_1}^2, {R_2}^2\right)(u_1 - {R_1}^2)({R_2}^2 - l_2)}.
\end{aligned}
$$

3.4 Summary

In this chapter, we proposed a new confidence interval constructed for a difference between two population squared multiple correlation from non-nested models by using the MOVER. The proposed confidence interval accounts for the potential asymmetry of the sampling distribution of the difference between two correlated R^2s using the MOVER. Since the MOVER in principle is derived by applying the central limit theorem, the procedure for differences between two R^2s is asymptotic in nature. Thus, properties in practical sample sizes must be evaluated.

Chapter 4

SIMULATION STUDY

4.1 Introduction

We have presented three versions of approximate confidence intervals (CIs) for differences between two squared multiple correlation coefficients arising from non-nested models in the previous chapter. These include the Wald method and two procedures based on a scaled central or noncentral F approximation and the MOVER. The validity of all procedures rely on the central limit theorem. Therefore, we need to evaluate their performance before they are applied in practice.

The lower and upper $100 \cdot (1-\alpha)\%$ confidence limits for parameter θ, l_θ and u_θ, are defined by $P(l_\theta < \theta < u_\theta) = 1-\alpha$. Accordingly, we assess the performance of the confidence interval procedures in terms of coverage rate, balance of tail error rates and average interval width. The empirical coverage rate (CV) was estimated by the proportion of confidence intervals that cover the true value of the parameter among a large number of replications. Ideally, the empirical coverage rates should be matched with the nominal confidence level of $100(1-\alpha)\%$. In practice, three criteria for acceptable deviation from the nominal confidence level commonly seen in literature (Bradley, 1978; Robey and Barcikowski, 1992) were adopted: strict criterion, $94.5\% - 95.5\%$; moderate criterion, $93.75\% - 96.25\%$; and liberal criterion, $92.5\% - 97.5\%$. Furthermore, an often neglected but important measure for evaluating confidence intervals is that the lower and upper tail error rates from two-sided confidence intervals are equal. Tail error rates were estimated by the proportion of confidence intervals lying completely to the left of the true value of parameter (missing from left, ML) and those lying completely to the right of the true value of parameter (missing

from right, MR). The rates of ML and MR should be asymptotically balanced and equal to $\alpha/2$. An additional criterion is that the such that the more narrow the average interval width (WD), the better.

4.2 Study design

Without loss of generality, we considered the case where it is of interest in determining whether a predictor X_2 or X_3 adds more to a model that already contains a predictor X_1 and the outcome Y. Explicitly, we evaluated the confidence intervals constructed for a difference between two squared multiple correlation coefficients from non-nested models denoted as $\Delta\rho^2 = {\rho_1}^2 - {\rho_2}^2$, where ${\rho_1}^2$ comes from the model predicting the outcome Y from predictors $\{X_1, X_2\}$ and ${\rho_2}^2$ comes from the model predicting the identical response variable Y from predictors $\{X_1, X_3\}$.

Based on discussions presented in the previous chapter, it was found that all three versions of asymptotic confidence intervals (CIs) constructed for a difference between two correlated R^2s highly depend on the correlations among the involved variables. Therefore, depending on the correlation matrix of a vector of involved predictors and the outcome, the same difference between ρ^2s could lead to very different confidence intervals. Hence, to investigate the performance of inference procedures for a wide variety of possible cases, we conducted a simulation study by using various population correlation matrices, as well as different values of $\Delta\rho^2$.

Denote Σ as the population correlation matrix for a vector of the response variable and all three predictors $[X_1, X_2, X_3, Y]$. In this simulation study, we considered a total of 50 typical parameter combinations ($10\Sigma \times 5n$), where ten population correlation matrices (Σ =A, B, ..., J) are shown in Table 4.1 and sample size n =50, 75, 100, 150 and 300. These ten correlation matrices were selected to represent a variety of common values of $\Delta\rho^2$ ranging from 0 to 0.30 and moderate values of ${\rho_{Y \cdot X_1 X_2 X_3}}^2$ for the full model predicting Y from predictors $\{X_1, X_2, X_3\}$. Among 10 matrices, to better represent generality of parameter se-

lection, matrices with the same value of $\Delta\rho^2$ (e.g., 0.10, 0.15 or 0.20) but different values of ${\rho_1}^2$ and ${\rho_2}^2$ were involved. For example, both matrix C and matrix D result in the same value of $\Delta\rho^2 = 0.10$, but the values of ${\rho_1}^2$ and ${\rho_2}^2$ calculated from matrix D are almost twice than those from matrix C.

We assume that the vector $[X_1, X_2, X_3, Y]$ follows a multivariate normal distribution, for each of ten population correlation matrices, we generated multivariate normal data with sample size n, then calculated the sample estimator of $\Delta\rho^2$ and constructed its confidence intervals by separately applying three versions of procedures. Finally, the performance of these three procedures were evaluated through simulating 1000 data sets.

Since our general idea is to obtain a CI for a difference using confidence limits for single parameters, we also evaluated the performance of procedures for constructing CIs for a single squared multiple correlation coefficient. For each of ten correlation matrices, the resulting values of ${\rho_1}^2$ and ${\rho_2}^2$ were calculated as shown in Table 4.1. For example, the ten correlation matrices led to the values of ${\rho_1}^2$ ranging from 0.31 to 0.68, which are common in practice. For each of the resulting values of ${\rho_1}^2$, sample size and replication, the Wald CI and the two CIs based on a scaled central/noncentral F approximation for ${\rho_1}^2$ were constructed.

For each of ten population correlation matrix, the performance of three asymptotic inference procedures for a single ρ^2 and a difference between two ρ^2s from non-nested models was evaluated by the following process.

1. For each selected correlation matrix Σ, three population squared multiple correlation coefficients ${\rho_1}^2$ for predicting Y from X_1 and X_2, ${\rho_2}^2$ for predicting Y from X_1 and X_3 and ${\rho_{Y \cdot X_1X_2X_3}}^2$ for predicting Y from X_1, X_2 and X_3, as well as the difference $\Delta\rho^2 = {\rho_1}^2 - {\rho_2}^2$, were calculated according to their definitions and shown in Table 4.1. For example, ${\rho_{Y \cdot X_1X_2}}^2 = 1 - |\Sigma(1,2,4)|/|\Sigma(1,2)|$, where $|\Sigma(1,2,4)|$ and $|\Sigma(1,2)|$ are respectively the principle minors of the determinant of the specified correlation matrix for vector $[X_1, X_2, Y]$ and $[X_1, X_2]$.

2. Drawing a random sample of n observations of vector $[X_1, X_2, X_3, Y]$ from a multivariate normal distribution with a specified population correlation matrix Σ, where the sample size n was varied from 50, 75, 100, 150, 300. For a given population correlation matrix Σ and sample size n, a $n \times 4$ matrix A was first generated as n samples of 4 independent standard normal variables. The desired data sets were then obtained as AU, in which U is the upper triangular matrix which satisfies the Cholesky decomposition $U^T U = \Sigma$, where the superscript T denotes the transpose of a matrix.

3. For each simulated data set, computing the sample correlation matrix of vector $[X_1, X_2, X_3, Y]$ denoted as $\widehat{\Sigma}$ by using SAS PROC CORR. This matrix $\widehat{\Sigma}$ was then used to compute all sample squared multiple correlation coefficients by using SAS PROC RSQUARE, which include the sample estimates of ${\rho_1}^2$ and ${\rho_2}^2$ denoted by ${R_1}^2$ and ${R_2}^2$.

4. Estimating the asymptotic variances of ${R_1}^2$ and ${R_2}^2$, as well as their covariance by using the formula presented in equations (3.14), (3.15) and (3.16) in last chapter, which are denoted by $\widehat{\text{var}}({R_1}^2)$, $\widehat{\text{var}}({R_2}^2)$ and $\widehat{\text{cov}}({R_1}^2, {R_2}^2)$. Then, from the formula $\text{corr}({R_1}^2, {R_2}^2) = \text{cov}({R_1}^2, {R_2}^2) / \sqrt{\text{var}({R_1}^2)\text{var}({R_2}^2)}$, the sample correlation coefficient between ${R_1}^2$ and ${R_2}^2$ denoted by $\widehat{\text{corr}}({R_1}^2, {R_2}^2)$ was also calculated.

5. Constructing three approximate confidence intervals for two single squared multiple correlation coefficients including ${\rho_1}^2$ and ${\rho_2}^2$, respectively by applying the Wald method, a scaled central or noncentral F approximation.

   - The two-sided $100(1-\alpha)\%$ Wald confidence intervals for ${\rho_i}^2$ denoted by (l_{iW}, u_{iW}), $i = 1, 2$, were constructed by $(l_{iW}, u_{iW}) = {R_i}^2 \mp z_{\alpha/2} \sqrt{\widehat{\text{var}}({R_i}^2)}$;
   - Given a simulated data set, the approximate confidence intervals for ${\rho_1}^2$ and ${\rho_2}^2$ based on a scaled central F approximation were constructed by using SAS

PROC CANCORR with the option SMC, which were denoted by (l_{1F}, u_{1F}) and (l_{2F}, u_{2F});

- By inputting the values of ${R_1}^2$ and ${R_2}^2$, we obtained the approximate confidence intervals for ${\rho_1}^2$ and ${\rho_2}^2$ based on a scaled noncentral F approximation implemented by Zou (2007), which were denoted by (l_{1NF}, u_{1NF}) and (l_{2NF}, u_{2NF}).

6. Calculating the sample estimate of the difference $\Delta\rho^2 = {\rho_1}^2 - {\rho_2}^2$ by using $\Delta R^2 = {R_1}^2 - {R_2}^2$.

7. Three corresponding confidence intervals for the difference can then be obtained by applying the MOVER formula given by

$$\begin{aligned} L &= {R_1}^2 - {R_2}^2 - \sqrt{({R_1}^2 - l_1)^2 + (u_2 - {R_2}^2)^2 - 2\widehat{\text{corr}}\left({R_1}^2, {R_2}^2\right)({R_1}^2 - l_1)(u_2 - {R_2}^2)}, \\ U &= {R_1}^2 - {R_2}^2 + \sqrt{(u_1 - {R_1}^2)^2 + ({R_2}^2 - l_2)^2 - 2\widehat{\text{corr}}\left({R_1}^2, {R_2}^2\right)(u_1 - {R_1}^2)({R_2}^2 - l_2)}, \end{aligned}$$

where the confidence interval for ${\rho_i}^2$, (l_i, u_i), is equal to (l_{iW}, u_{iW}), (l_{iF}, u_{iF}) or (l_{iNF}, u_{iNF}), $i = 1, 2$ for each procedure respectively.

These above steps were repeated 1000 times for each parameter combination. The following results were recorded: (a) the empirical expected value of $\Delta\rho^2$ or the average value of ΔR^2 over 1000 replications; (b) the *coverage rate*(CV) of confidence intervals constructed respectively for ${\rho_1}^2$, ${\rho_2}^2$ and $\Delta\rho^2$; c) tail error rates including *missing from left* (ML) and *missing from right* (MR) of confidence intervals constructed for ${\rho_1}^2$, ${\rho_2}^2$ and $\Delta\rho^2$, respectively; d) the *average interval widths* (WD) respectively for ${\rho_1}^2$, ${\rho_2}^2$ and $\Delta\rho^2$ over 1000 replications; e) the *power* (P), defined as the proportion of confidence intervals for $\Delta\rho^2$ that did not contain 0 for non-null cases ($\Delta\rho^2 \neq 0$).

Since α was set at 5%, the coverage rates were expected to be 95%. The rates of missing from left and missing from right of confidence intervals were expected to be balanced and equal to 2.5%. For a procedure maintaining a given α level, the higher power the better.

4.3 Results

4.3.1 Results for Point Estimation

Table 4.2 presents sample estimates of the parameter ($\Delta\rho^2$). The results show that, in general, as sample size increases the discrepancy between the estimate and the true parameter value decreases and the estimates are reasonably close to the true parameter value.

4.3.2 Confidence intervals for a single ρ^2

The performance of three procedures for constructing two-sided 95% confidence intervals for ${\rho_1}^2$ is presented in Table 4.3. Here, the performance of procedures for ${\rho_2}^2$ is not shown in the table, since it showed similar trends to that for ${\rho_1}^2$.

From Table 4.3, it can be seen that the Wald method resulted in a coverage rate within the range of $94.5\% - 95.5\%$, specified by Bradley's strict criterion (Bradley, 1978), only for 9 parameter combinations. Among those outside the range, all fell below 94.5%, even when the sample size was as large as 150. In light of the moderate criterion ($93.75\% - 96.25\%$), the Wald method still provided adequate coverage only in 20 of 50 parameter combinations. The two procedures based on a central/noncentral F approximation provided coverage within the range of the strict criterion in 42 of 50 parameter combinations. Among those outside this range, all but one case showed coverage within the range of the moderate criterion. Moreover, two procedures based on a central/noncentral F approximation led to well balanced tail errors, while the Wald method resulted in the upper tail errors more than twice than the corresponding lower tail errors. In addition, all three procedures had very similar interval widthes for all cases. Hence, the poor performance of the Wald method is consistent with previous evaluations (Algina, 1999).

Table 4.3 also shows the close performance of two procedures based on a central or noncentral F approximation. For the eight moderate values of ρ^2, which are 0.31, 0.38, 0.45, 0.50, 0.54, 0.58, 0.65 and 0.68, both provided similar coverage rates and interval width, but the noncentral F approximation having relatively larger lower tail error rates

resulted in less balanced tail errors than the central F approximation did. This suggests that the procedure based on a two-moment scaled central F approximation for a single ρ^2 performs adequately, despite Lee (1971)'s argument that a two-moment scaled central F approximation 'seems reasonable though rather crude'. Furthermore, the theory and usage of a two-moment scaled central F approximation is slightly simpler than that of a three-moment scaled noncentral F approximation. Hence, for a single squared multiple correlation coefficient, we recommend the procedure based on a two-moment scaled central F approximation, as implemented in SAS PROC CANCORR.

4.3.3 Differences between two ρ^2s from non-nested models

In this simulation study, three procedures were separately employed to construct two-sided 95% confidence intervals for $\Delta\rho^2$, including those based on the Wald-type procedure, a scaled central F approximation and the MOVER (referred to as the 'FM' procedure for short), or a scaled noncentral F approximation and the MOVER (referred to as the 'NFM' procedure for short).

For the null case (Matrix A), the proportion of confidence intervals that did not contain 0 represents Type I error rate, which was set at 5% here. The performance of three procedures in terms of coverage rate, tail errors, average interval width and Type I error rate are shown in Table 4.4 for samples of size 50, 75, 100, 150, and 300. Using the range of Bradley's strict criterion 94.5% to 95.5% as acceptable for empirical results, the Wald-type procedure appears to control Type I error rate reasonably well and slightly better than other two procedures, even for small sample sizes. However, the discrepancy among three procedures was so small that all three procedures became acceptable when adopting Bradley's liberal criterion $92.5-97.5\%$. Note that the coverage rate in the null case is simply one minus the Type I error rate. In addition, the three procedures provided balanced tail errors and similar interval width for the null case.

For all nine non-null cases (matrix B-J), the performance of three procedures in terms of coverage rate, tail errors, average interval width and power rate over 1000 replications

was presented in Table 4.5. A visual representation of coverage rates and power rates for three procedures is provided in Figures 4.1-4.6, respectively.

In the first three figures of describing coverage rates, the solid line represents the expected coverage rate (95%), the dotted lines represent the limits of Bradley's strict criterion ($94.5-95.5\%$), and the dot-dashed lines represent the limits of Bradley's liberal criterion ($92.5-97.5\%$).

Figure 4.1 suggests that the Wald-type procedure resulted in a coverage rate within the range of Bradley's strict criterion, only for 7 parameter combinations. Among those outside the range, all fell below 94.5%, even with sample size as large as 300. By the identical strict criterion, as shown in Figures 4.2 and 4.3, the 'FM' and 'NFM' procedures resulted in a similar coverage rate within the range respectively in 18 of 45 and 19 of 45 parameter combinations, even for some cases with small sample size. For example, for matrix B (i.e., $\Delta\rho^2 = 0.05$ and ${\rho_1}^2 = 0.54$) and sample size $n = 300$, the Wald-type procedure provided a coverage of only 93.4% less than the strict criterion of 94.5%, while the 'FM' procedure provided a coverage of 95.2% and the 'NFM' procedure provided a coverage of 95.5%. Another example with small sample size $n = 50$, for matrix I (i.e., $\Delta\rho^2 = 0.25$ and ${\rho_1}^2 = 0.50$), the coverage rates from the 'FM' and 'NFM' procedures with the values of 94.6% and 95.3% were very closed to the nominal level, while for the Wald-type procedure, it was only 92.7%, far falling short of 94.5%. As the sample size increases, the coverage rates from the Wald-type procedure tend to reach the nominal level from below, rather than hovering around the expected coverage rate as those from the 'FM' and 'NFM' procedures. In summary, according to Bradley's strict criterion ($94.5-95.5\%$), the Wald-type procedure did not provide adequate coverage percentage even with a sample sizes as large as 300, while both the 'FM' and 'NFM' procedures provided adequate coverage even for small sample size ($n = 50$).

Even with the liberal criterion of 92.5% to 97.5%, the Wald-type procedure failed in 11 of 45 parameter combinations, all but one occurred at values of ${\rho_1}^2$ above 0.45 (such as matrix B, F, H, J) and small sample size $n = 50$ and 75. In contrast, the 'FM' and 'NFM'

procedure provided adequate coverage in 41 of 45 parameter combinations. Almost all cases outside the range of liberal criterion occurred at small values of $\Delta\rho^2$ (0.05) with small sample sizes of 50 and 75. For these cases, the coverage rates from the Wald-type procedure even went further away from the expected coverage rate (95%). Hence, according to Bradley's liberal criterion, we can conclude that the 'FM' and 'NFM' procedures provide better coverage than the Wald-type procedure for all non-null cases ($\Delta\rho^2 \neq 0$).

Table 4.5 also shows tail error rates of non-null cases. These results show that the Wald-type procedure provided less balanced tail error rates than the 'FM' and 'NFM' procedures, and the 'NFM' procedure provided slightly more balanced tail error rates than the 'FM' procedure, especially when the difference between $\Delta\rho^2$ and ${\rho_1}^2$ was above 0.50. For example, for matrix D with $\Delta\rho^2 = 0.10$ and ${\rho_1}^2 = 0.68$, the mean difference between lower and upper tail error rates for all sample sizes considered was equal to 2.34 for the Wald-type procedure, 0.38 for the 'FM' procedure and 0.1 for the 'NFM' procedure. Furthermore, these three procedures had very similar average interval width for all correlation matrices and sample sizes considered.

Based on the above results for coverage rate, tail error rates and average interval width, the Wald-type procedure performed much worse than the 'FM' and 'NFM' procedures. The 'NFM' procedure was shown to perform slightly better than the 'FM' procedure in terms of balanced tail error rates.

The observed (empirical) power rates from three procedures for non-null cases were also compared as shown in column ('*P%*') of Table 4.5 and Figures 4.4, 4.5 and 4.6. For each of three procedures, power increases with sample size as expected for all parameter combinations considered. The powers from the Wald-type procedure were larger than other two procedures at all cases. In practice, due to the poor performance of the Wald-type procedure in terms of coverage probability and balanced tail error rates, the resulting power values were not valid. Furthermore, the powers of the 'FM' procedure were slightly higher than those of the 'NFM' procedure at all cases considered. It is well known that the power in non-null cases is simply one minus Type II error. To control Type I and II errors in the

hypothesis test $H_0 : \Delta\rho^2 = 0$, the 'FM' procedure seems to be the best choice. Hence, for a difference between two squared multiple correlation coefficients from non-nested models, we recommend applying the MOVER with the scaled central F approximation for single R-squares.

4.4 Discussion

The simulation results we conducted led to two general conclusions. First, to construct a confidence interval for a single squared multiple correlation coefficient, the procedure based on a two-moment scaled central F approximation performed best among three procedures. Second, through a series of evaluations of the performance of three procedures for constructing confidence intervals for a difference between two squared multiple correlation coefficients from non-nested models, the Wald-type procedure was shown to be clearly inappropriate. In contrast, the MOVER procedure with the two-moment scaled central F approximation for single R-squares is the best choice. The procedure based on the three-moment scaled noncentral F approximation is a viable alternative.

The purpose of this simulation study was to serve as a preliminary evaluation of the performance of three asymptotic procedures using 50 typical parameter combinations and a relatively simple R^2 comparison. However, the general patterns observed in this study should readily generalize to R^2 comparisons with any a population correlation matrix of a vector of the response and more than three predictor variables.

Table 4.1: Population correlation matrices (Σ) of the vector of three predictors and the outcome $[X_1, X_2, X_3, Y]$ used in simulation studies and the resulting population squared multiple correlation coefficients including ${\rho_1}^2 = {\rho_{Y \cdot X_1 X_2}}^2$, ${\rho_2}^2 = {\rho_{Y \cdot X_1 X_3}}^2$ and ${\rho_{Y \cdot X_1 X_2 X_3}}^2$, as well as the values of $\Delta\rho^2 = {\rho_1}^2 - {\rho_2}^2$.

Population correlation matrix (Σ)						$\Delta\rho^2$	${\rho_{Y \cdot X_1 X_2}}^2$	${\rho_{Y \cdot X_1 X_3}}^2$	${\rho_{Y \cdot X_1 X_2 X_3}}^2$
A		X_1	X_2	X_3	Y	0	.39	.39	.47
	X_1	1.0							
	X_2	.3	1.0						
	X_3	.3	.3	1.0					
	Y	.5	.5	.5	1.0				
B						.05	.54	.49	.55
	X_1	1.0							
	X_2	.1	1.0						
	X_3	.5	.1	1.0					
	Y	.7	.3	.3	1.0				
C						.10	.38	.28	.52
	X_1	1.0							
	X_2	.3	1.0						
	X_3	.5	.5	1.0					
	Y	.5	.5	.1	1.0				
D						.10	.68	.58	.69
	X_1	1.0							
	X_2	.1	1.0						
	X_3	.3	.5	1.0					

Continued on next page

Table 4.1 – Continued

Population correlation matrix (Σ)					$\Delta\rho^2$	$\rho_{Y \cdot X_1 X_2}{}^2$	$\rho_{Y \cdot X_1 X_3}{}^2$	$\rho_{Y \cdot X_1 X_2 X_3}{}^2$
	Y	.7	.5	.5	1.0				
E						.15	.31	.16	.37
	X_1	1.0							
	X_1	.1	1.0						
	X_2	.1	.1	1.0					
	Y	.3	.5	.3	1.0				
F						.15	.65	.50	.66
	X_1	1.0							
	X_1	.5	1.0						
	X_2	.3	.5	1.0					
	Y	.7	.7	.3	1.0				
G						.20	.31	.11	.34
	X_1	1.0							
	X_2	.1	1.0						
	X_3	.7	.1	1.0					
	Y	.3	.5	.1	1.0				
H						.20	.45	.25	.46
	X_1	1.0							
	X_2	.1	1.0						
	X_3	.3	.1	1.0					
	Y	.5	.5	.1	1.0				
I						.25	.50	.25	.53
	X_1	1.0							

Continued on next page

Table 4.1 – Continued

Population correlation matrix (Σ)					$\Delta\rho^2$	$\rho_{Y \cdot X_1 X_2}{}^2$	$\rho_{Y \cdot X_1 X_3}{}^2$	$\rho_{Y \cdot X_1 X_2 X_3}{}^2$
	X_2	.3	1.0					
	X_3	.1	.5	1.0				
	Y	.1	.7	.5	1.0			
J					.30	.58	.28	.60
	X_1	1.0						
	X_2	.3	1.0					
	X_3	.5	.1	1.0				
	Y	.5	.7	.1	1.0			

Table 4.2: Sample estimates of the difference $\Delta\rho^2$ averaged over 1000 replications by sample size of 50, 75, 100, 150, and 300.

Matrix / Sample size	A **(0.00)**	B **(0.05)**	C **(0.10)**	D **(0.10)**	E **(0.15)**	F **(0.15)**	G **(0.20)**	H **(0.20)**	I **(0.25)**	J **(0.30)**
50	0.00	0.05	0.10	0.10	0.14	0.15	0.19	0.19	0.24	0.29
75	0.00	0.05	0.10	0.10	0.14	0.15	0.19	0.20	0.24	0.29
100	0.00	0.05	0.10	0.10	0.15	0.15	0.20	0.20	0.24	0.30
150	0.00	0.05	0.10	0.10	0.15	0.15	0.20	0.20	0.25	0.30
300	0.00	0.05	0.10	0.10	0.15	0.15	0.20	0.20	0.25	0.30

Note: Numbers in bold and brackets indicate the population parameters.

Table 4.3: Performance of procedures for constructing two-sided 95% confidence intervals for a single squared multiple correlation coefficient (${\rho_1}^2$) over 1000 replications, by sample size.

${\rho_1}^2$	SS	Wald		Central F approximation		Noncentral F approximation	
		CV (ML, MR)%	WD	CV (ML, MR)%	WD	CV (ML, MR)%	WD
0.31	50	91.0 (3.1, 5.9)	0.41	95.2 (2.5, 2.3)	0.40	90.1 (3.4, 6.5)	0.38
	75	92.3 (2.4, 5.3)	0.34	95.0 (2.2, 2.8)	0.33	94.9 (2.8, 2.3)	0.34
	100	93.5 (2.0, 4.5)	0.30	95.1 (2.0, 2.9)	0.29	95.0 (2.4, 2.6)	0.29
	150	94.0 (2.1, 3.9)	0.24	95.1 (2.1, 2.8)	0.24	94.7 (3.0, 2.3)	0.24
	300	94.7 (2.2, 3.1)	0.17	94.9 (2.2, 2.9)	0.17	94.6 (2.6, 2.8)	0.17
0.38	50	91.5 (2.6, 5.9)	0.41	95.2 (2.6, 2.2)	0.41	94.6 (3.5, 1.9)	0.40
	75	92.3 (2.6, 5.1)	0.34	95.1 (2.7, 2.2)	0.34	94.8 (3.5, 1.7)	0.34
	100	94.5 (1.8, 3.7)	0.29	96.0 (2.1, 1.9)	0.29	95.8 (2.4, 1.8)	0.30
	150	93.0 (2.5, 4.5)	0.24	94.4 (2.7, 2.9)	0.24	94.4 (3.1, 2.5)	0.24
	300	95.2 (2.0, 2.8)	0.17	95.5 (2.2, 2.3)	0.17	95.3 (2.8, 1.9)	0.17
0.45	50	91.0 (1.8, 7.2)	0.39	95.4 (2.2, 2.4)	0.40	95.5 (2.8, 1.7)	0.40
	75	92.8 (1.8, 5.4)	0.32	95.5 (2.4, 2.1)	0.33	95.4 (2.8, 1.8)	0.33
	100	94.1 (1.5, 4.4)	0.28	95.4 (2.0, 2.6)	0.28	95.1 (2.5, 2.4)	0.29
	150	94.7 (1.8, 3.5)	0.23	94.8 (2.5, 2.7)	0.23	94.7 (2.8, 2.5)	0.24
	300	94.7 (1.9, 3.4)	0.17	95.0 (2.6, 2.4)	0.17	95.1 (2.7, 2.2)	0.17
0.50	50	92.3 (1.9, 5.8)	0.38	95.3 (2.4, 2.3)	0.38	95.3 (3.1, 1.6)	0.39
	75	92.3 (1.8, 5.9)	0.31	94.3 (3.1, 2.6)	0.32	94.5 (3.3, 2.2)	0.32
	100	92.8 (2.1, 5.1)	0.27	94.5 (2.8, 2.7)	0.27	94.4 (3.1, 2.5)	0.28
	150	93.8 (1.9, 4.3)	0.22	95.5 (2.0, 2.5)	0.23	95.4 (2.6, 2.0)	0.23
	300	94.5 (2.1, 3.4)	0.16	94.5 (3.0, 2.5)	0.16	94.5 (3.2, 2.3)	0.16
0.54	50	91.0 (1.6, 7.4)	0.36	94.9 (2.3, 2.8)	0.37	95.1 (3.0, 1.9)	0.38
	75	93.6 (1.2, 5.2)	0.30	95.7 (2.2, 2.1)	0.30	95.7 (2.5, 1.8)	0.31
	100	93.1 (1.6, 5.3)	0.26	95.1 (2.6, 2.3)	0.26	95.5 (2.8, 1.7)	0.27

Continued on next page

Table 4.3 – Continued

$\rho_1{}^2$	SS	Wald		Central F approximation		Noncentral F approximation	
		CV (ML, MR)%	WD	CV (ML, MR)%	WD	CV (ML, MR)%	WD
	150	94.2 (1.8, 4.0)	0.21	94.9 (2.7, 2.4)	0.22	94.9 (3.1, 2.0)	0.22
	300	94.7 (1.8, 3.5)	0.15	95.8 (2.3, 1.9)	0.15	95.7 (2.5, 1.8)	0.15
0.58	50	92.2 (1.2, 6.6)	0.34	95.3 (2.5, 2.2)	0.36	94.9 (3.3, 1.8)	0.36
	75	92.8 (1.3, 5.9)	0.28	95.6 (2.2, 2.2)	0.29	95.4 (2.7, 1.9)	0.29
	100	94.2 (0.8, 5.0)	0.25	95.7 (1.9, 2.4)	0.25	95.8 (2.3, 1.9)	0.25
	150	93.9 (1.7, 4.4)	0.20	95.2 (2.6, 2.2)	0.21	95.1 (2.8, 2.1)	0.21
	300	95.0 (1.7, 3.3)	0.14	95.4 (2.3, 2.3)	0.14	95.4 (2.4, 2.2)	0.15
0.65	50	92.1 (0.6, 7.3)	0.30	95.4 (2.5, 2.1)	0.32	95.2 (3.1, 1.7)	0.33
	75	93.6 (0.8, 5.6)	0.25	95.7 (2.2, 2.1)	0.26	95.8 (2.5, 1.7)	0.26
	100	93.9 (0.9, 5.2)	0.22	95.2 (2.4, 2.4)	0.22	95.5 (2.6, 1.9)	0.22
	150	94.2 (1.4, 4.4)	0.18	94.8 (2.9, 2.3)	0.18	95.1 (3.0, 1.9)	0.18
	300	94.4 (1.6, 4.0)	0.13	95.6 (2.5, 1.9)	0.13	95.3 (2.9, 1.8)	0.13
0.68	50	91.8 (0.8, 7.4)	0.29	95.3 (2.7, 2.0)	0.31	94.7 (3.4, 1.9)	0.31
	75	93.6 (0.5, 5.9)	0.24	96.2 (1.7, 2.1)	0.25	95.9 (2.3, 1.8)	0.25
	100	93.8 (1.0, 5.2)	0.21	94.7 (2.8, 2.5)	0.21	95.0 (2.9, 2.1)	0.21
	150	93.9 (1.4, 4.7)	0.17	95.1 (2.5, 2.4)	0.17	95.3 (2.7, 2.0)	0.17
	300	95.1 (1.1, 3.8)	0.12	94.8 (2.7, 2.5)	0.12	94.6 (3.2, 2.2)	0.12 [1]

[1]Note: $\rho_1{}^2 = \rho_{Y \cdot X_1 X_2}{}^2$ denotes the selected population squared multiple correlation coefficient for predicting Y from X_1 and X_2; 'SS' denotes sample size; 'CV' denotes the coverage rate; 'ML' and 'MR' denote that the confidence interval misses the true value of parameter from the left and the right, respectively; 'WD' denotes the average interval width over 1000 replications.

Table 4.4: Null case: the performance of 3 procedures for constructing two-sided 95% confidence intervals for a difference between two ρ^2s from non-nested models ($\Delta\rho^2 = {\rho_1}^2 - {\rho_2}^2$) over 1000 replications, by sample size.

Σ	SS	Wald		MOVER and central F		MOVER and noncentral F	
		CV (ML, MR)% WD	E%	CV (ML, MR)% WD	E%	CV (ML, MR)% WD	E%
A	50	95.0 (2.4, 2.6) 0.36	5.0	96.7 (1.7, 1.6) 0.37	3.3	96.9 (1.5, 1.6) 0.39	3.1
($\Delta\rho^2 = 0$)	75	94.8 (2.7, 2.5) 0.30	5.2	96.3 (1.9, 1.8) 0.31	3.7	96.6 (1.7, 1.7) 0.32	3.4
	100	95.4 (1.9, 2.7) 0.26	4.6	95.9 (1.7, 2.4) 0.27	4.1	96.3 (1.6, 2.1) 0.27	3.7
	150	95.2 (2.0, 2.8) 0.22	4.8	95.8 (1.8, 2.4) 0.22	4.2	95.9 (1.8, 2.3) 0.22	4.1
	300	94.8 (2.6, 2.6) 0.15	5.2	95.0 (2.5, 2.5) 0.15	5.0	95.2 (2.3, 2.5) 0.15	4.8

Note: Σ denotes the specified population correlation matrix of a vector of 3 predictors and the outcome presented in Table 4.1; 'SS' denotes sample size; 'CV' denotes the coverage rate; 'ML' and 'MR' denote that the interval misses the true value of parameter from the left and the right, respectively; 'WD' denotes the average interval width over 1000 replications; 'E' denotes the Type I error rate.

Table 4.5: Non-null cases: the performance of procedures for constructing two-sided 95% confidence intervals for a difference between two ρ^2s from non-nested models ($\Delta\rho^2 = {\rho_1}^2 - {\rho_2}^2$) over 1000 replications, by population correlation matrix and sample size.

Σ	SS	Wald		MOVER and central F		MOVER and noncentral F	
		CV (ML, MR)% WD	P%	CV (ML, MR)% WD	P%	CV (ML, MR)% WD	P%
B	50	90.0 (9.2, 0.8) 0.19	5.9	99.1 (0.4, 0.5) 0.24	4.1	99.3 (0.3, 0.4) 0.27	2.6
	75	90.3 (8.9, 0.8) 0.15	11.0	98.1 (1.3, 0.6) 0.18	8.0	99.0 (0.4, 0.6) 0.20	6.9
	100	91.5 (8.1, 0.4) 0.13	21.0	97.2 (2.4, 0.4) 0.15	16.2	98.3 (1.3, 0.4) 0.16	13.6
	150	92.8 (6.8, 0.4) 0.11	40.8	96.3 (3.3, 0.4) 0.12	35.5	97.2 (2.4, 0.4) 0.12	31.3
	300	93.4 (5.5, 1.1) 0.07	80.1	95.2 (3.7, 1.1) 0.08	76.6	95.5 (3.4, 1.1) 0.08	74.7
C	50	93.8 (4.0, 2.2) 0.38	17.9	95.1 (3.4, 1.5) 0.38	14.9	95.7 (3.1, 1.2) 0.41	13.3
	75	93.6 (3.7, 2.7) 0.31	26.4	94.7 (3.3, 2.0) 0.31	24.0	95.1 (3.0, 1.9) 0.32	22.6
	100	94.0 (3.7, 2.3) 0.27	34.2	94.8 (3.6, 1.6) 0.27	31.7	95.0 (3.5, 1.5) 0.28	30.4
	150	93.1 (3.6, 3.3) 0.22	46.1	94.0 (3.5, 2.5) 0.23	43.9	94.2 (3.4, 2.4) 0.23	43.1
	300	94.2 (3.2, 2.6) 0.16	71.0	94.5 (3.2, 2.3) 0.16	70.5	94.6 (3.2, 2.2) 0.16	70.1
D	50	93.5 (4.8, 1.8) 0.24	27.2	96.8 (1.4, 1.8) 0.29	24.0	97.5 (0.9, 1.6) 0.31	21.4
	75	94.3 (4.6, 1.1) 0.20	43.0	96.5 (1.9, 1.6) 0.22	40.1	96.7 (1.7, 1.6) 0.23	36.9
	100	93.4 (5.0, 1.6) 0.17	56.1	96.1 (2.2, 1.7) 0.19	53.6	96.5 (1.8, 1.7) 0.20	51.7
	150	95.2 (3.5, 1.3) 0.14	76.4	95.4 (2.5, 2.1) 0.15	74.5	95.8 (2.1, 2.1) 0.15	72.7
	300	94.5 (3.6, 1.9) 0.10	97.2	95.1 (3.0, 1.9) 0.10	96.7	95.2 (2.8, 2.0) 0.10	96.5
E	50	93.4 (4.0, 2.6) 0.44	25.5	94.5 (4.3, 1.2) 0.42	21.4	95.8 (3.2, 1.0) 0.47	20.1
	75	93.6 (4.1, 2.3) 0.36	34.5	94.5 (4.2, 1.3) 0.36	31.4	94.9 (3.8, 1.3) 0.37	30.5
	100	95.0 (2.7, 2.3) 0.32	43.7	95.2 (3.0, 1.8) 0.31	40.3	95.4 (2.8, 1.8) 0.32	39.9
	150	94.0 (3.3, 2.7) 0.26	58.5	94.5 (3.7, 2.1) 0.26	57.7	94.3 (3.6, 2.1) 0.26	56.8
	300	94.3 (3.0, 2.7) 0.19	86.1	94.6 (3.2, 2.2) 0.18	85.8	94.7 (3.2, 2.1) 0.19	85.7
F	50	91.1 (8.1, 0.8) 0.25	70.8	96.5 (2.6, 0.9) 0.29	62.3	97.7 (1.5, 0.8) 0.31	56.4
	75	91.0 (8.5, 0.5) 0.20	92.4	96.7 (2.8, 0.5) 0.23	88.5	97.2 (2.2, 0.6) 0.24	84.8
	100	91.5 (7.5, 1.0) 0.18	98.5	95.4 (3.5, 1.1) 0.19	97.5	96.3 (2.6, 1.1) 0.20	96.8

Continued on next page

Table 4.5 – Continued

Σ	SS	Wald		MOVER and central F		MOVER and noncentral F	
		CV (ML, MR)% WD	P%	CV (ML, MR)% WD	P%	CV (ML, MR)% WD	P%
	150	93.4 (5.7, 0.9) 0.15	100.0	95.2 (3.3, 1.5) 0.15	100.0	95.7 (2.8, 1.5) 0.16	100.0
	300	93.1 (4.8, 2.1) 0.10	100.0	94.3 (3.4, 2.3) 0.11	100.0	94.8 (2.9, 2.3) 0.11	100.0
G	50	92.4 (5.4, 2.2) 0.42	43.1	93.0 (6.0, 1.0) 0.40	37.2	95.1 (3.9, 1.0) 0.47	34.4
	75	92.5 (5.0, 2.5) 0.34	59.5	93.1 (5.4, 1.5) 0.33	57.2	93.9 (4.7, 1.4) 0.37	56.0
	100	93.2 (4.7, 2.1) 0.30	73.2	93.4 (5.3, 1.3) 0.29	71.3	94.0 (4.8, 1.2) 0.31	70.3
	150	92.8 (4.0, 3.2) 0.24	87.6	93.8 (4.9, 1.6) 0.24	87.5	93.4 (5.0, 1.6) 0.24	86.9
	300	95.0 (3.2, 1.8) 0.17	99.3	95.3 (3.5, 1.2) 0.17	99.3	95.3 (3.5, 1.2) 0.17	99.3
H	50	91.0 (7.2, 1.8) 0.35	57.1	92.9 (6.3, 0.8) 0.35	50.7	93.9 (5.3, 0.8) 0.39	46.6
	75	91.6 (6.7, 1.7) 0.29	82.1	93.0 (6.3, 0.7) 0.29	78.7	93.3 (6.0, 0.7) 0.30	76.2
	100	93.2 (5.8, 1.0) 0.25	92.8	93.7 (5.7, 0.6) 0.25	91.5	94.5 (4.9, 0.6) 0.26	90.6
	150	93.0 (5.3, 1.7) 0.20	98.4	93.2 (5.5, 1.3) 0.20	98.4	93.5 (5.2, 1.3) 0.21	98.0
	300	94.1 (4.0, 1.9) 0.14	100.0	94.3 (4.2, 1.5) 0.14	100.0	94.4 (4.1, 1.5) 0.15	100.0
I	50	92.7 (3.9, 3.4) 0.44	56.5	94.6 (3.6, 1.8) 0.44	50.6	95.3 (2.9, 1.8) 0.47	48.2
	75	94.0 (3.3, 2.7) 0.37	73.0	95.1 (3.2, 1.7) 0.37	70.0	95.7 (2.5, 1.8) 0.38	69.1
	100	93.6 (3.2, 3.2) 0.32	83.6	94.6 (3.4, 2.0) 0.32	82.0	94.7 (3.2, 2.1) 0.33	81.8
	150	94.5 (2.9, 2.6) 0.26	95.5	94.9 (3.2, 2.0) 0.26	94.8	95.1 (2.8, 2.1) 0.27	94.6
	300	94.9 (2.7, 2.4) 0.19	100.0	95.1 (2.8, 2.1) 0.19	100.0	95.0 (2.8, 2.2) 0.19	100.0
J	50	92.3 (5.4, 2.3) 0.39	83.2	94.2 (4.7, 1.1) 0.40	79.4	94.5 (4.2, 1.3) 0.43	78.3
	75	91.7 (5.4, 2.9) 0.32	94.9	93.5 (4.8, 1.7) 0.33	94.2	93.7 (4.5, 1.8) 0.34	94.0
	100	93.2 (5.2, 1.6) 0.28	98.6	93.8 (5.0, 1.2) 0.28	98.5	94.1 (4.7, 1.2) 0.29	98.5
	150	93.5 (4.2, 2.3) 0.23	99.8	94.1 (4.0, 1.9) 0.23	99.7	94.5 (3.3, 2.2) 0.24	99.7
	300	94.5 (3.6, 1.9) 0.16	100.0	95.1 (3.4, 1.5) 0.16	100.0	95.0 (3.4, 1.6) 0.17	100.0 [2]

[2]Note: Σ denotes the specified population correlation matrix of a vector of 3 predictors and the outcome presented in Table 4.1; 'SS' denotes sample size; 'CV' denotes the coverage rate; 'ML' and 'MR' denote that the interval misses the true value of parameter from the left and the right, respectively; 'WD' denotes the average interval width over 1000 replications; 'P' denotes the power rate, defined as the proportion of confidence intervals that did not contain 0.

Figure 4.1: Coverage rates of 95% confidence intervals over 1000 replications using the Wald-type procedure, by sample size.

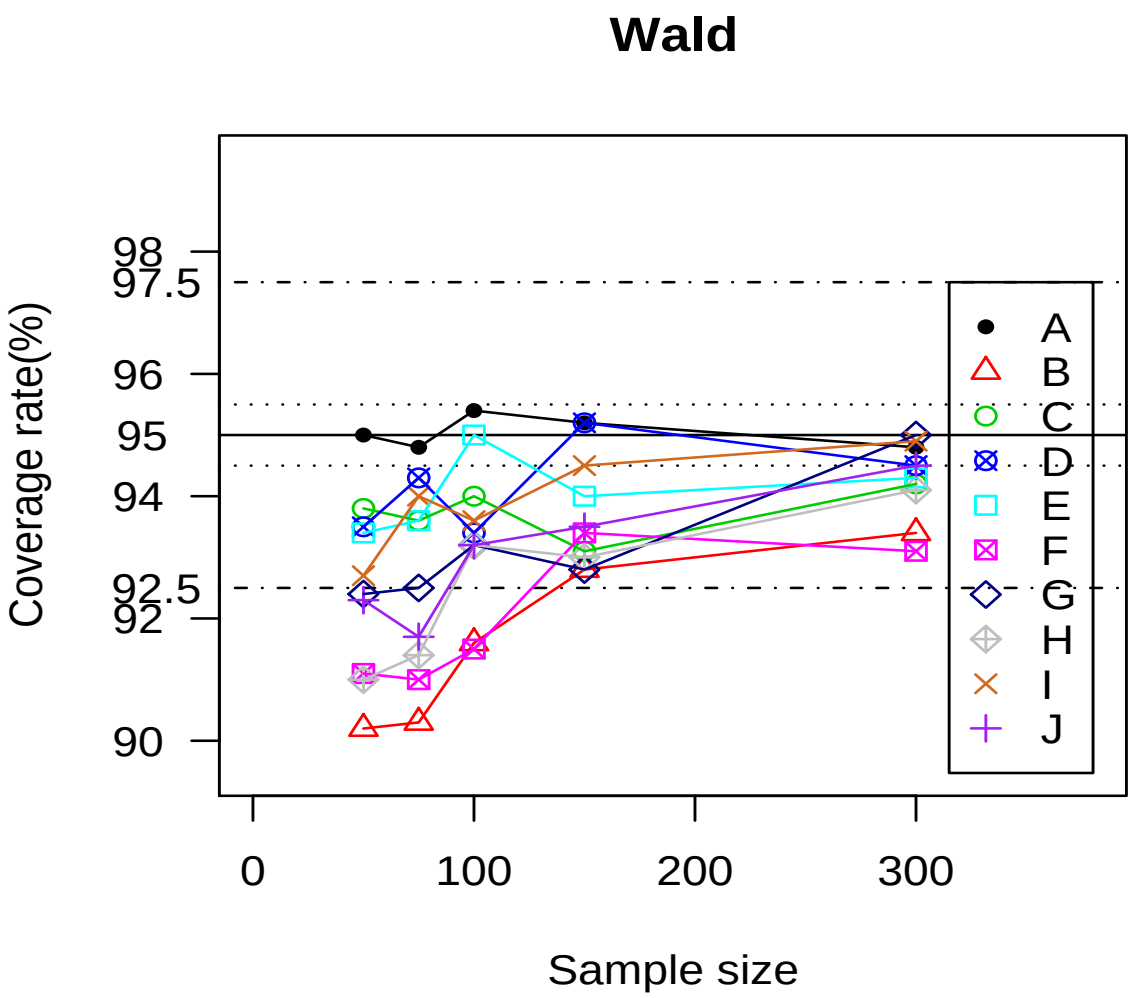


Figure 4.2: Coverage rates of 95% confidence intervals over 1000 replications based on a scaled central F approximation and the MOVER, by sample size.

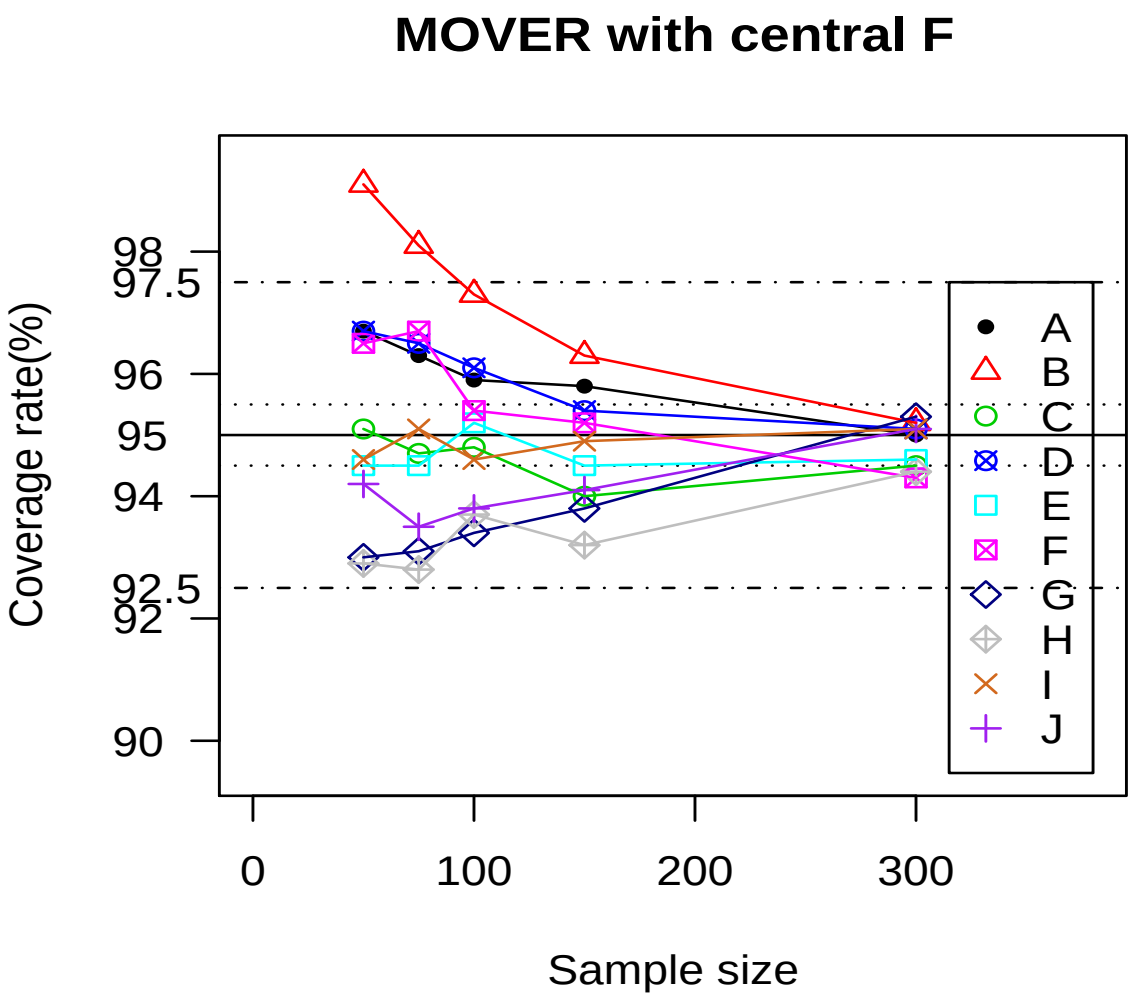


Figure 4.3: Coverage rates of 95% confidence intervals over 1000 replications based on a scaled noncentral F approximation and the MOVER, by sample size.

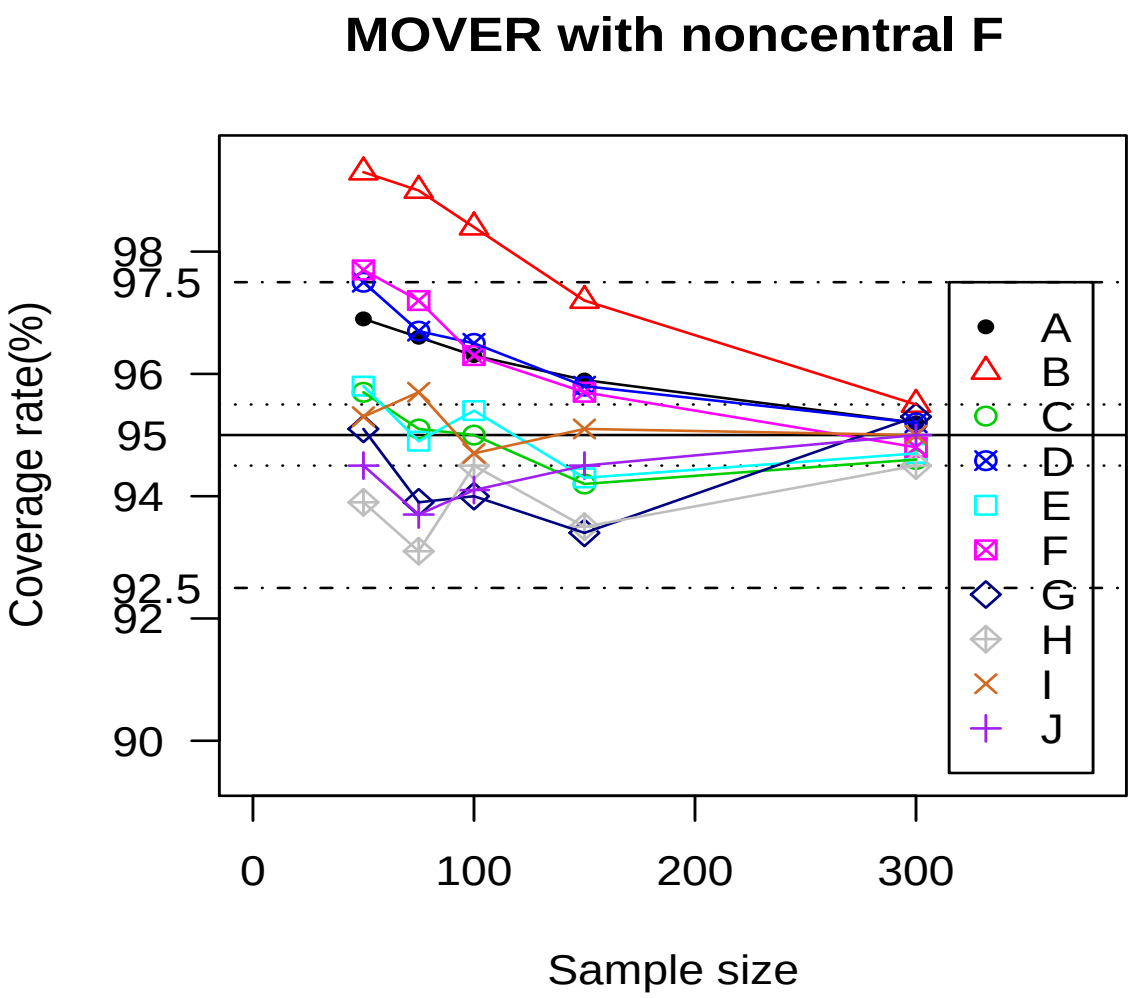


Figure 4.4: Power (null hypothesis rejection) rates over 1000 replications using the Wald-type procedure, by sample size.

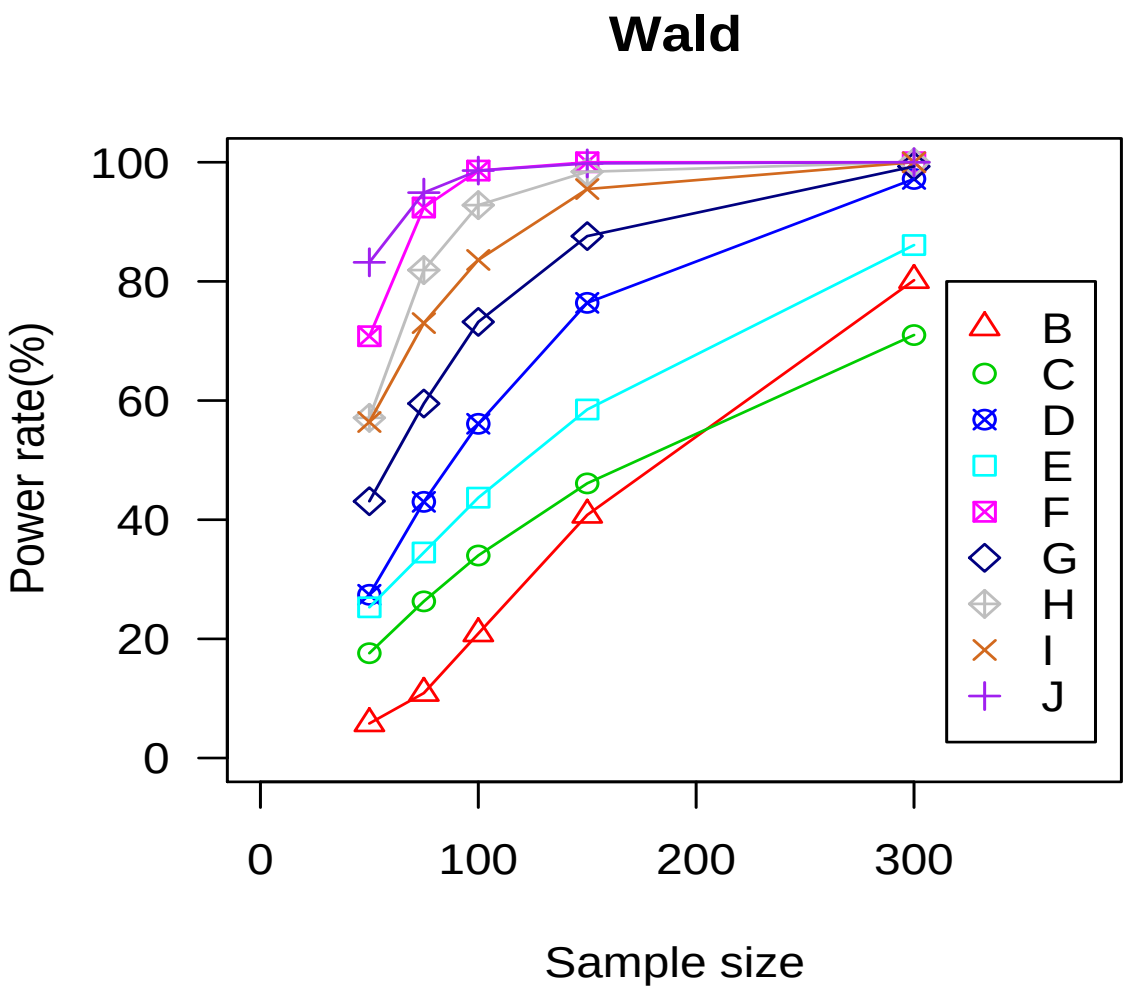


Figure 4.5: Power (null hypothesis rejection) rates over 1000 replications based on a scaled central F approximation and the MOVER, by sample size.

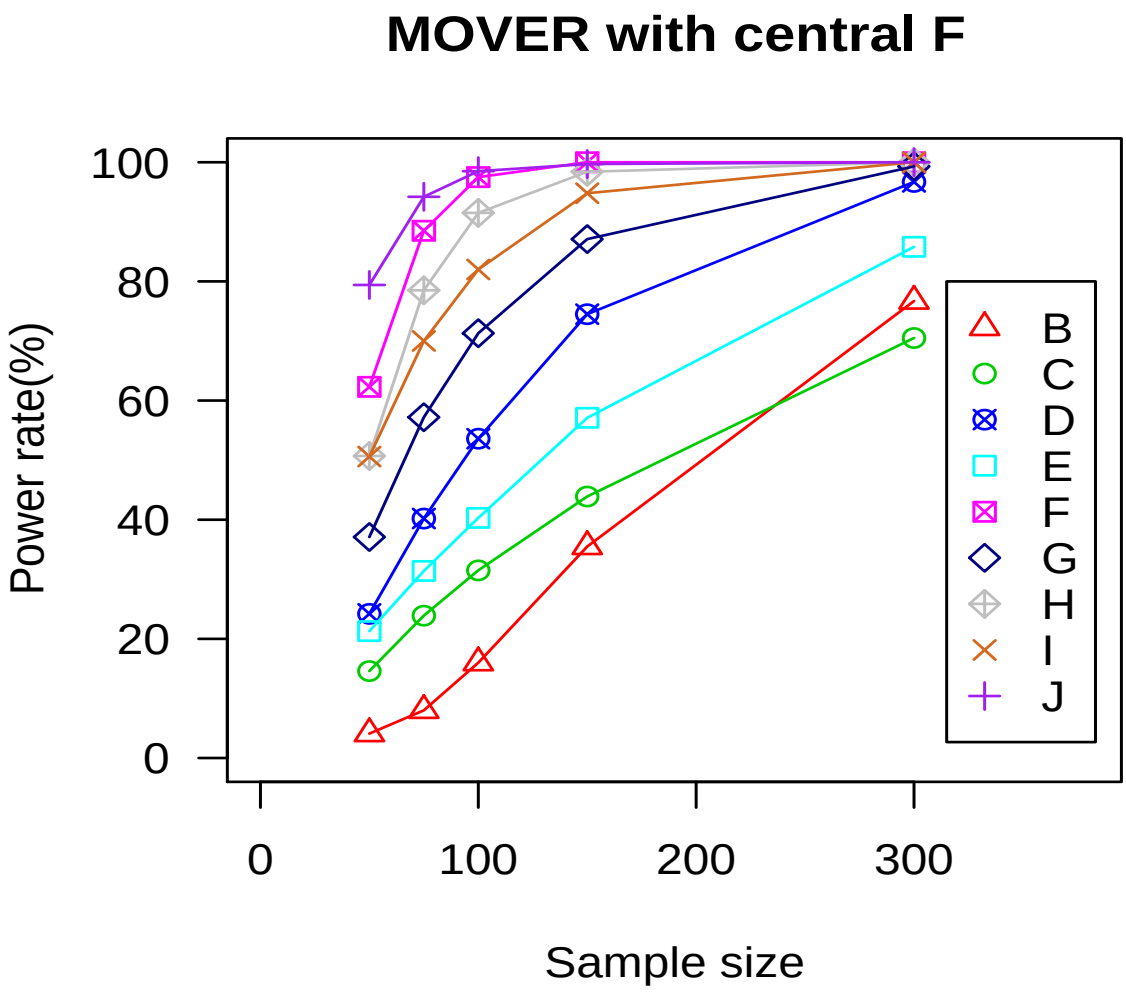
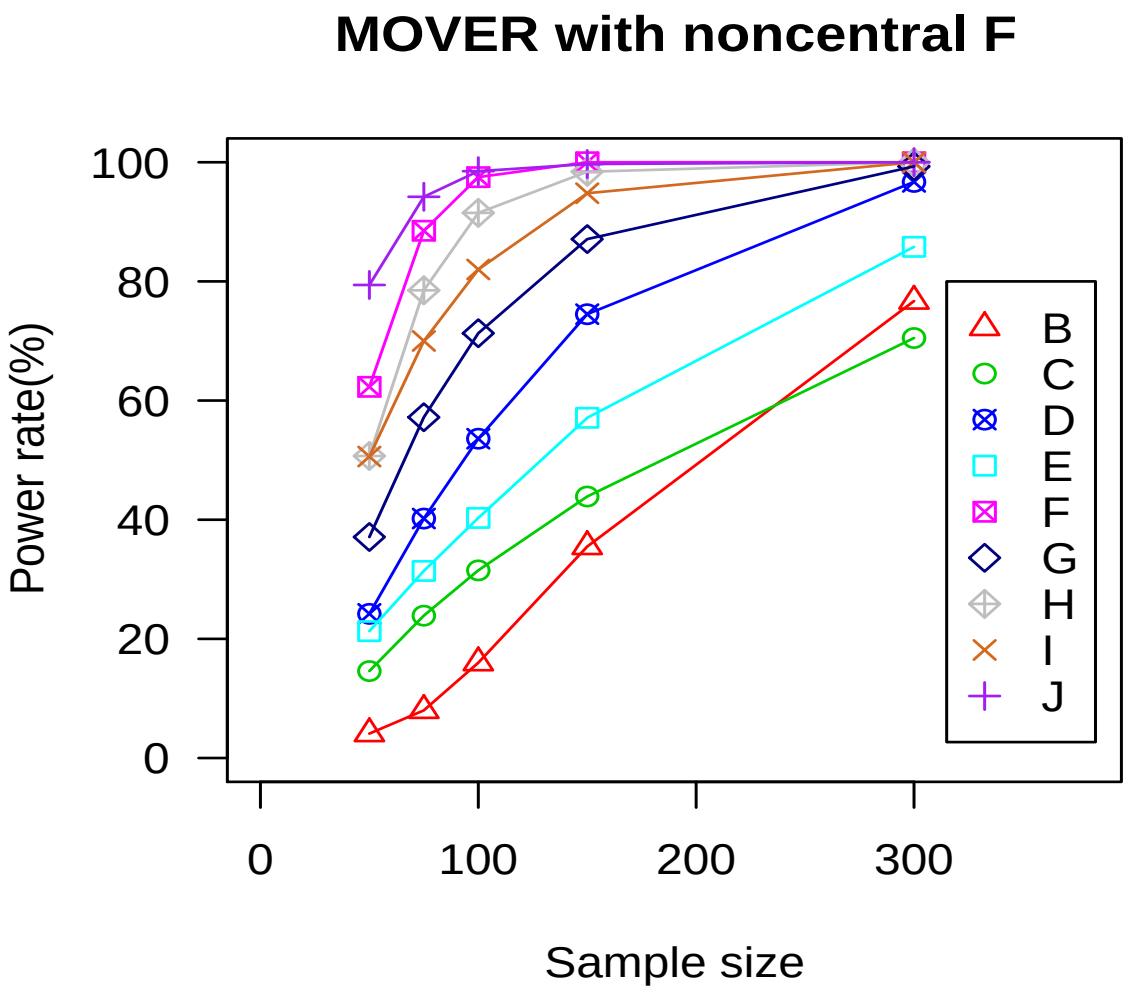


Figure 4.6: Power (null hypothesis rejection) rates over 1000 replications based on a scaled noncentral F approximation and the MOVER, by sample size.



Chapter 5

WORKED EXAMPLES

5.1 Introductory remark

Simulation results presented in the previous chapter have shown that procedures based on the MOVER for confidence intervals for differences between two squared multiple correlations from nonnested models perform better than the Wald-type methods. For illustrative purposes, we present two examples by applying the proposed procedures to determine the relative importance of predictors (Budescu, 1993; Hedges and Olkin, 1981).

The definition of relative importance of predictors in the context of dominance analysis and commonality analysis can be summarized as follows. Consider a multiple regression model with p predictors, dominance is defined as a pairwise relationship examined for all $p(p-1)/2$ pairs of predictors. Consider X_i and X_j, and let X_β stands for any subset of the remaining $p-2$ variables excluding X_i and X_j. Define variable X_i to dominant variable X_j denoted by the notation X_i D X_j if, and only if the squared multiple correlations for predicting the outcome Y from predictors X_i and X_β is larger than those from predictors X_j and X_β, that is,

$$\rho_{Y\cdot X_\beta X_i}{}^2 > \rho_{Y\cdot X_\beta X_j}{}^2,$$

for all possible selections of X_β, including the null set.

Taking $p = 3$ for example, Table 5.1 presents the dominance analysis for three predictors. The first column identifies the predictors included in each of seven submodels including the null model (a model without any predictors). The second column describes the corresponding squared multiple correlation coefficient. The next three columns describe

change in squared multiple correlation as a result of the addition of each predictor. For example, the second row describes the degree to which a model consisting of X_1 is improved by adding to it an additional predictor X_2 or X_3. To determine pairwise dominance relationships, we compare each pair of nonzero contributions of predictors (columns) across all submodels (row). For example, when comparing X_1 and X_2, we need to examine their direct contributions (${\rho_{Y\cdot X_1}}^2$ vs. ${\rho_{Y\cdot X_2}}^2$) and their additional contributions to the model including X_3 (${\rho_{Y\cdot X_1X_3}}^2$ vs. ${\rho_{Y\cdot X_2X_3}}^2$). If there is evidence to support that both differences are positive, a dominance is established (X_1 D X_2).

We considered two examples below. The first example arising from data of breakfast cereals involving 3 predictors and sample size of $n = 76$. The second example contains 4 predictors and a larger sample size, $n = 315$. Our goal is to identify the importance of predictors by applying the proposed procedures.

5.2 Breakfast cereals

The breakfast cereals data is publically available at Statlib Data and Story Library (DASL) at CMU (http://lib.stat.cmu.edu/DASL/Stories/HealthyBreakfast.html) . Data was collected on the nutritional information and consumer rating of 76 brand-name breakfast cereals. For each cereal, observations of 16 variables were collected: cereal name, cereal manufacturer, type (hot or cold), number of calories per serving, grams of protein, grams of fat, milligrams of sodium, grams of fiber, grams of carbohydrates, grams of sugars, milligrams of potassium, typical percentage of the FDA recommended vitamins and minerals, the weight of one serving, the number of cups in one serving, the shelf location, and a rating calculated by consumer reports.

Each of the last 13 numerical variables as a univariate predictor and 'rating' as the response variable was found to be significant with grams of sugars, fiber and fat resulting in largest R-Squared values. A multiple linear model with these three predictors has an overall R-square of 0.868. We now perform a dominance analysis using these three predictors.

Table 5.2 presents the sample correlation matrix among contents of grams of sugars, fiber and fat, and rating. These values show that grams of sugar and fat are negatively correlated with rating, while grams of fiber is positively correlated with rating. Correlations among predictors are relatively smaller.

For three predictors, the resulting coefficients of determination of seven submodels are shown in the second column of Table 5.3. Also shown are the corresponding confidence intervals obtained by using the Wald method, a scaled central/noncentral F approximation, respectively. It is apparent that the confidence intervals based on a scaled central/noncentral F approximation are virtually identical but very different from that by the Wald method, which we have seen in the last chapter to perform poorly.

Table 5.4 summarizes the calculation of asymptotical confidence intervals for all three pairs of predictors including $X_1 - X_2$, $X_1 - X_3$ and $X_2 - X_3$. These results suggest that the dominance relationship among three predictors is that: sugar seems to dominate fiber and fat, and fiber seems to dominate fat. The last three columns show the 95% asymptotic confidence intervals based on the Wald procedure, the MOVER and a scaled central or noncentral F approximation denoted by the 'FM' and 'NFM' procedures. Note that the difference between two independent R^2s is the special case of the difference between two R^2s from non-nested models. All three confidence intervals support the inferiority of fat to sugar. However, the comparisons of grams of sugar vs. fiber, and fiber vs. fat are inconclusive, since both the confidence intervals for ${\rho_{Y\cdot X_1X_2}}^2 - {\rho_{Y\cdot X_1X_3}}^2$ and that for ${\rho_{Y\cdot X_2}}^2 - {\rho_{Y\cdot X_3}}^2$ contain 0. Thus, there is no evidence that grams of sugar is the most important predictor. Similarly, grams of fat is not more important than fiber.

From Table 5.4, we also found that confidence intervals based on the 'FM' and 'NFM' procedures were almost identical, but very different from those based on the Wald procedure. As we concluded at the end of the previous chapter, the Wald confidence interval is not acceptable, especially for small sample size.

5.3 Plasma concentrations of beta-carotene

Plasma concentrations of the micronutrients like retinol, beta-carotene, or other carotenoids vary widely from subject to subject. Observation studies have suggested that low plasma concentrations of beta-carotene might be associated with increased risk of developing certain types of cancer (Nierenberg et al., 1989). The data set for this example arose from a cross-sectional study investigating the effects of personal characteristics and dietary factors on plasma concentrations of beta-carotene (Stukel, 2003). This study enrolled 315 patients who had an elective surgical procedure during a three-year period to biospy or remove a lesion of the lung, colon, breast, skin, ovary or uterus that was found to be non-cancerous. For each patient, the data contains eleven variables, including age, gender, smoking status (1=never, 2=former smoker, 3=current smoker), quetelet (weight/(height^2)), vitamin use (1=Yes, fairly often, 2=Yes, not often, 3=No), number of calories consumed per day, grams of fat consumed per day, grams of fiber consumed per day, number of alcoholic drinks consumed per week, cholesterol consumed (mg per day) and dietary beta-carotene consumed (mcg per day). This raw data set is publically available at Statlib Data and Story Library (DASL) at CMU (`http://lib.stat.cmu.edu/datasets/Plasma_Retinol`) .

A related study of plasma beta-carotene levels found that dietary beta-carotene intake and female sex were positively associated with beta-carotene levels to a statistically significant extent, while smoking status and quetelet index had significantly negative effects (Nierenberg et al., 1989). Furthermore, the sampling distributions of plasma beta-carotene level and dietary beta-carotene intake are positively skewed (Nierenberg et al., 1989). Hence, we analyzed the reduced model considering $Y = \log$(plasma beta-carotene levels) as the outcome and $X_1 = \log$(dietary beta-carotene intake), X_2=gender, X_3=quetelet and X_4=smoking status as the only four predictors. This final reduced model will be used for illustrating complicated dominance analysis.

Table 5.5 presents the sample correlation matrix for the four predictors and the outcome Y. These values show that smoking status and quetelet index are negatively correlated with

log plasma beta-carotene levels, while log dietary beta-carotene intake and female gender are positively correlated with log plasma beta-carotene levels, which is consistent with previous studies (Nierenberg et al., 1989). The correlations among predictors are relatively smaller.

The $2^4 - 1 = 15$ coefficients of determination for all possible submodels are presented in Table 5.6. Also shown are the confidence intervals obtained by using the Wald method, a scaled central/noncentral F approximation, respectively. It is clear that the confidence intervals based on a scaled central/noncentral F approximation are very closed to each other but very different from those based on the Wald method.

Table 5.7 summarizes the calculation of asymptotical confidence intervals for all six pairs of predictors including $X_1 - X_2$, $X_1 - X_3$, $X_1 - X_4$, $X_2 - X_3$, $X_2 - X_4$ and $X_3 - X_4$. The differences between the relevant squared multiple correlations are calculated and presented in the second column of Table 5.7. According to the signs of those differences, it appears that the dominance relationship among three predictors is that: quetelet index seems to dominate log dietary beta-carotene intake and smoking status, and log dietary beta-carotene intake and smoking status seems to dominate gender. The last three columns show the 95% confidence intervals based on the Wald, 'FM' and 'NFM' procedures, respectively. However, it is inconclusive regarding their dominance relationship. For example, gender is a better predictor than quetelet index in the presence of smoking status or both log dietary beta-carotene intake and smoking status; but we can not determine the relative importance between gender and quetelet index by itself and in the presence of log dietary beta-carotene intake. Except this pair, the confidence intervals for all other pairs presented in the table contain 0. Hence, there is no evidence that any one of those four predictors is better or more important than the other. This suggests that the four predictors cannot be ordered and that their relative importance cannot be determined. Note that the confidence intervals based on the 'FM' and 'NFM' procedures are almost identical, but very different from those based on the Wald procedure.

5.4 Summary

In this chapter, we have presented two examples that illustrate the proposed procedures based on the MOVER with a scaled central/noncentral F approximation for single R-squares being applied for determining the relative importance of predictors. By considering different number of predictors and sample size, it has been shown that the proposed procedures can be readily applied in dominance analysis and commonality analysis.

Table 5.1: Dominance analysis for three predictors.

		Contribution of		
Variable(s)	ρ^2	X_1	X_2	X_3
-	0	${\rho_{Y\cdot X_1}}^2$	${\rho_{Y\cdot X_2}}^2$	${\rho_{Y\cdot X_3}}^2$
X_1	${\rho_{Y\cdot X_1}}^2$	-	${\rho_{Y\cdot X_1X_2}}^2 - {\rho_{Y\cdot X_1}}^2$	${\rho_{Y\cdot X_1X_3}}^2 - {\rho_{Y\cdot X_1}}^2$
X_2	${\rho_{Y\cdot X_2}}^2$	${\rho_{Y\cdot X_1X_2}}^2 - {\rho_{Y\cdot X_2}}^2$	-	${\rho_{Y\cdot X_2X_3}}^2 - {\rho_{Y\cdot X_2}}^2$
X_3	${\rho_{Y\cdot X_3}}^2$	${\rho_{Y\cdot X_1X_3}}^2 - {\rho_{Y\cdot X_3}}^2$	${\rho_{Y\cdot X_2X_3}}^2 - {\rho_{Y\cdot X_3}}^2$	-
X_1, X_2	${\rho_{Y\cdot X_1X_2}}^2$	-	-	${\rho_{Y\cdot X_1X_2X_3}}^2 - {\rho_{Y\cdot X_1X_2}}^2$
X_1, X_3	${\rho_{Y\cdot X_1X_3}}^2$	-	${\rho_{Y\cdot X_1X_2X_3}}^2 - {\rho_{Y\cdot X_1X_3}}^2$	-
X_2, X_3	${\rho_{Y\cdot X_2X_3}}^2$	${\rho_{Y\cdot X_1X_2X_3}}^2 - {\rho_{Y\cdot X_2X_3}}^2$	-	-

Table 5.2: Sample correlation matrix in example of breakfast cereals ($n = 76$).

	Variable			
Variable	sugar (X_1)	fiber (X_2)	fat (X_3)	rating (Y)
sugars (X_1)	1			
fiber (X_2)	-0.1388	1		
fat (X_3)	0.3025	0.0138	1	
rating (Y)	-0.7639	0.5839	-0.4205	1

Note: Three predictors are $X_1 =$ grams of sugars, $X_2 =$ grams of fiber and $X_3 =$ grams of fat. The outcome is $Y =$ a rating of cereal.

Table 5.3: Squared multiple correlation coefficients and their asymptotic 95% confidence intervals in example of breakfast cereal with 3 predictors ($n = 76$).

		Confidence interval based on		
Predictors	R^2	Wald	central F	noncentral F
X_1	0.584	(0.441, 0.727)	(0.419, 0.709)	(0.420, 0.709)
X_2	0.341	(0.168, 0.514)	(0.168, 0.508)	(0.168, 0.508)
X_3	0.177	(0.021, 0.332)	(0.046, 0.345)	(0.046, 0.345)
X_1, X_2	0.816	(0.742, 0.891)	(0.718, 0.877)	(0.718, 0.877)
X_1, X_3	0.623	(0.489, 0.757)	(0.459, 0.736)	(0.460, 0.736)
X_2, X_3	0.525	(0.370, 0.679)	(0.343, 0.659)	(0.344, 0.659)
X_1, X_2, X_3	0.868	(0.813, 0.923)	(0.791, 0.911)	(0.791, 0.911)

Note: Three predictors are X_1 = grams of sugars, X_2 = grams of fiber and X_3 = grams of fat. The outcome is Y = a rating of cereal.

Table 5.4: Asymptotic 95% confidence intervals for all pairwise differences of ρ^2s in example of breakfast cereal with 3 predictors ($n = 76$).

Variable	ΔR^2	Confidence interval based on		
		Wald	'FM'	'NFM'
$X_1 - X_2$	${R_{Y \cdot X_1}}^2 - {R_{Y \cdot X_2}}^2 = 0.243$	(0.018, 0.467)	(0.008, 0.456)	(0.009, 0.456)
	${R_{Y \cdot X_1 X_3}}^2 - {R_{Y \cdot X_2 X_3}}^2 = 0.098$	(-0.089, 0.285)	(-0.095, 0.296)	(-0.095, 0.295)
$X_1 - X_3$	${R_{Y \cdot X_1}}^2 - {R_{Y \cdot X_3}}^2 = 0.407$	(0.195, 0.618)	(0.172, 0.588)	(0.172, 0.588)
	${R_{Y \cdot X_1 X_2}}^2 - {R_{Y \cdot X_2 X_3}}^2 = 0.292$	(0.144, 0.440)	(0.154, 0.463)	(0.154, 0.462)
$X_2 - X_3$	${R_{Y \cdot X_2}}^2 - {R_{Y \cdot X_3}}^2 = 0.164$	(-0.069, 0.397)	(-0.077, 0.376)	(-0.077, 0.376)
	${R_{Y \cdot X_1 X_2}}^2 - {R_{Y \cdot X_1 X_3}}^2 = 0.193$	(0.077, 0.309)	(0.087, 0.337)	(0.087, 0.336)

Note: Three predictors are X_1 = grams of sugars, X_2 = grams of fiber and X_3 = grams of fat. The outcome is Y = a rating of cereal. 'FM' indicates the MOVER with a scaled central F approximation for single R-squares. 'NFM' indicates the MOVER with a scaled noncentral F approximation for single R-squares.

Table 5.5: Sample correlation matrix for five variables in example of plasma beta-carotene levels ($n = 315$).

	Variable				
Variables	logbetadiet	gender	quetelet	smokstat	logbetaplasma
logbetadiet (X_1)	1				
gender(X_2)	-0.047	1			
quetelet(X_3)	0.009	-0.007	1		
smokstat(X_4)	-0.098	-0.121	-0.113	1	
logbetaplasma (X_5)	0.187	0.135	-0.280	-0.191	1

Note: Four predictors are X_1 = log dietary beta-carotene consumed, X_2 = gender, X_3 = quetelet (weight/(height^2)), X_4 = cmoking status. The outcome is Y = log plasma beta-carotene level.

Table 5.6: Squared multiple correlation coefficients and their asymptotic 95% confidence intervals in example of plasma beta-carotene levels with 4 predictors ($n = 315$).

		Confidence interval based on		
Predictors	R^2	Wald	central F	noncentral F
X_1	0.035	(-.005, 0.075)	(0.006, 0.085)	(0.006, 0.085)
X_2	0.018	(-.011, 0.048)	(0.001, 0.058)	(0.000, 0.058)
X_3	0.078	(0.021, 0.135)	(0.030, 0.143)	(0.030, 0.143)
X_4	0.037	(-.004, 0.077)	(0.007, 0.087)	(0.007, 0.087)
X_1, X_2	0.056	(0.007, 0.105)	(0.014, 0.111)	(0.014, 0.111)
X_1, X_3	0.114	(0.048, 0.181)	(0.053, 0.184)	(0.053, 0.184)
X_1, X_4	0.065	(0.013, 0.118)	(0.020, 0.123)	(0.020, 0.123)
X_2, X_3	0.096	(0.034, 0.158)	(0.040, 0.162)	(0.040, 0.162)
X_2, X_4	0.049	(0.003, 0.096)	(0.011, 0.102)	(0.011, 0.102)
X_3, X_4	0.129	(0.060, 0.198)	(0.064, 0.200)	(0.064, 0.201)
X_1, X_2, X_3	0.135	(0.065, 0.205)	(0.066, 0.205)	(0.066, 0.204)
X_1, X_2, X_4	0.081	(0.023, 0.138)	(0.027, 0.139)	(0.027, 0.139)
X_1, X_3, X_4	0.157	(0.083, 0.231)	(0.084, 0.230)	(0.084, 0.230)
X_2, X_3, X_4	0.140	(0.069, 0.211)	(0.070, 0.210)	(0.070, 0.211)
$X_1, X_2, X_3\, X_4$	0.171	(0.095, 0.247)	(0.092, 0.242)	(0.093, 0.243)

Note: Four predictors are X_1 = log dietary beta-carotene consumed, X_2 = gender, X_3 = quetelet (weight/(height^2)), X_4 = smoking status. The outcome is Y = log plasma beta-carotene level.

Table 5.7: Asymptotic 95% confidence intervals for all pairwise differences of ρ^2s in example of plasma beta-carotene levels with four predictors ($n = 315$).

Variables	ΔR^2	Confidence interval based on		
		Wald	'FM'	'NFM'
$X_1 - X_2$	${R_{Y\cdot X_1}}^2 - {R_{Y\cdot X_2}}^2 = 0.017$	(-0.033, 0.066)	(-0.033, 0.070)	(-0.033, 0.070)
	${R_{Y\cdot X_1X_3}}^2 - {R_{Y\cdot X_2X_3}}^2 = 0.018$	(-0.031, 0.067)	(-0.030, 0.068)	(-0.030, 0.068)
	${R_{Y\cdot X_1X_4}}^2 - {R_{Y\cdot X_2X_4}}^2 = 0.016$	(-0.028, 0.060)	(-0.028, 0.062)	(-0.028, 0.062)
	${R_{Y\cdot X_1X_3X_4}}^2 - {R_{Y\cdot X_2X_3X_4}}^2 = 0.017$	(-0.024, 0.058)	(-0.024, 0.058)	(-0.024, 0.058)
$X_1 - X_3$	${R_{Y\cdot X_1}}^2 - {R_{Y\cdot X_3}}^2 = -0.043$	(-0.113, 0.026)	(-0.114, 0.026)	(-0.114, 0.026)
	${R_{Y\cdot X_1X_2}}^2 - {R_{Y\cdot X_2X_3}}^2 = -0.040$	(-0.111, 0.030)	(-0.110, 0.030)	(-0.110, 0.030)
	${R_{Y\cdot X_1X_4}}^2 - {R_{Y\cdot X_3X_4}}^2 = -0.063$	(-0.134, 0.007)	(-0.133, 0.006)	(-0.133, 0.006)
	${R_{Y\cdot X_1X_2X_4}}^2 - {R_{Y\cdot X_2X_3X_4}}^2 = -0.060$	(-0.130, 0.010)	(-0.128, 0.010)	(-0.128, 0.010)
$X_1 - X_4$	${R_{Y\cdot X_1}}^2 - {R_{Y\cdot X_4}}^2 = -0.002$	(-0.059, 0.055)	(-0.060, 0.056)	(-0.060, 0.057)
	${R_{Y\cdot X_1X_2}}^2 - {R_{Y\cdot X_2X_4}}^2 = 0.007$	(-0.046, 0.059)	(-0.046, 0.060)	(-0.046, 0.060)
	${R_{Y\cdot X_1X_3}}^2 - {R_{Y\cdot X_3X_4}}^2 = -0.014$	(-0.072, 0.043)	(-0.071, 0.043)	(-0.071, 0.042)
	${R_{Y\cdot X_1X_2X_3}}^2 - {R_{Y\cdot X_2X_3X_4}}^2 = -0.005$	(-0.061, 0.050)	(-0.060, 0.049)	(-0.060, 0.049)
$X_2 - X_3$	${R_{Y\cdot X_2}}^2 - {R_{Y\cdot X_3}}^2 = -0.060$	(-0.124, 0.004)	(-0.127, 0.002)	(-0.127, 0.002)
	${R_{Y\cdot X_1X_2}}^2 - {R_{Y\cdot X_1X_3}}^2 = -0.059$	(-0.123, 0.006)	(-0.124, 0.005)	(-0.123, 0.005)
	${R_{Y\cdot X_2X_4}}^2 - {R_{Y\cdot X_3X_4}}^2 = -0.079$	(-0.144, -0.015)	(-0.144, -0.016)	(-0.144, -0.016)
	${R_{Y\cdot X_1X_2X_4}}^2 - {R_{Y\cdot X_1X_3X_4}}^2 = -0.077$	(-0.141, -0.013)	(-0.139, -0.012)	(-0.139, -0.013)
$X_2 - X_4$	${R_{Y\cdot X_2}}^2 - {R_{Y\cdot X_4}}^2 = -0.018$	(-0.069, 0.032)	(-0.072, 0.032)	(-0.072, 0.032)
	${R_{Y\cdot X_1X_2}}^2 - {R_{Y\cdot X_1X_4}}^2 = -0.009$	(-0.054, 0.035)	(-0.055, 0.035)	(-0.055, 0.035)
	${R_{Y\cdot X_2X_3}}^2 - {R_{Y\cdot X_3X_4}}^2 = -0.033$	(-0.083, 0.018)	(-0.083, 0.017)	(-0.084, 0.017)
	${R_{Y\cdot X_1X_2X_3}}^2 - {R_{Y\cdot X_1X_3X_4}}^2 = -0.023$	(-0.070, 0.025)	(-0.069, 0.025)	(-0.069, 0.025)
$X_3 - X_4$	${R_{Y\cdot X_3}}^2 - {R_{Y\cdot X_4}}^2 = 0.042$	(-0.028, 0.112)	(-0.028, 0.113)	(-0.028, 0.113)
	${R_{Y\cdot X_1X_3}}^2 - {R_{Y\cdot X_1X_4}}^2 = 0.049$	(-0.022, 0.120)	(-0.021, 0.120)	(-0.021, 0.120)
	${R_{Y\cdot X_2X_3}}^2 - {R_{Y\cdot X_2X_4}}^2 = 0.047$	(-0.025, 0.118)	(-0.024, 0.118)	(-0.024, 0.118)
	${R_{Y\cdot X_1X_2X_3}}^2 - {R_{Y\cdot X_1X_2X_4}}^2 = 0.054$	(-0.014, 0.122)	(-0.014, 0.121)	(-0.014, 0.121)

Note: Four predictors are X_1 = log dietary beta-carotene consumed, X_2 = gender, X_3 = quetelet (weight/(height^2)), X_4 = smoking status. The outcome is Y = log plasma beta-carotene level.

Chapter 6

SUMMARY AND DISCUSSION

This thesis provided a simple and efficient inference procedure for differences between two correlated squared multiple correlations from non-nested models ($\Delta\rho^2$). Specifically, we proposed a closed-form confidence interval for the comparison of the changes in R^2 when each of two sets of predictors is added to a model.

The Wald approach to confidence interval construction for $\Delta\rho^2$ (Olkin and Finn, 1995) ignores the skewness of the distribution of single R-squares. This method is valid only for large sample size (> 200) (Azen and Sass, 2008). As an alternative, we proposed two procedures based on the MOVER (Zou, 2008) with the scaled central or noncentral F approximation for single R-squares.

Comparing these three procedures through simulation studies and the worked examples, we can make two general conclusions. First, the confidence interval for single squared multiple correlations based on the two-moment scaled central F approximation performs better than the Wald method and the method based on the three-moment scaled noncentral F approximation. Second, as for the confidence interval for differences between two squared multiple correlation coefficients from non-nested models, the Wald-type procedure performs poorly in terms of overall coverage and balanced tail errors. In contrast, the MOVER procedure with the scaled central F approximation for single R-squares performs much better, even with sample sizes as small as 50.

Although we considered the simplest case with three predictors and ten correlation matrices in simulation studies, the results can be extended straightforwardly to other general cases as illustrated in the worked examples.

The proposed procedure provides an efficient way for implementing dominance and

commonality analysis (Budescu, 1993; Kerlinger and Pedhazur, 1973). It has been suggested that R-squared can serve as a general measure of determining the relative importance of predictors in multiple regression analysis (Budescu, 1993). Thus, according to the rationale of dominance and commonality analysis, the procedure can be used to rank the predictors according to their contributions in predicting the response variable through a series of pairwise comparisons of squared multiple correlations from all possible submodels. This thesis also provides an general approach to investigate a multivariate regression model by identifying the best subset of predictors.

The results from this research rely on the assumption of the raw data following multivariate normality, which may not be appropriate in practice. Even if the assumption becomes unreasonable, the proposed procedure would still be applicable by replacing suitable confidence limits if available, since the derivation of MOVER itself does not require the normality assumption for the raw data. It suggests that the validity of the proposed procedure depends on that of the confidence limits for single squared multiple correlations.

The discussions in this thesis focus on inference on differences between two correlated ρ^2s from non-nested models. Whereas for differences between two correlated ρ^2s from nested models, the differences have a constrained parameter space, the method proposed here may not work well.

The coefficient of determination (R^2) has been extendedly used in the class of exponential family regression models (Cameron and Windmeijer, 1997). Different from its application to linear regression models, the values of generalized R^2 in nonlinear models have different interpretation, may lie outside the $[0, 1]$ interval and decrease as regressors are added. To assess the relative importance of predictors in such models, future work is needed.

BIBLIOGRAPHY

Akaike, H. (1974), 'A new look at the statistical model identification', *IEEE Transactions on Automatic Control* **19**, 716–723.

Akaike, H. (1977), On entropy maximization principles, *in* P. R.Krishnaiah, ed., 'Applications of Statistics', North Holland, Amsterdam, pp. 27–41.

Aldrich, J. (1995), 'Correlations genuine and spurious in Pearson and Yule', *Statistical Science* **10**, 364–376.

Alf, E. F. and R. G. Graf (1999), 'Asymptotic confidence limits for the difference between two squared multiple correlations: A simplified approach', *Psychological Methods* **4**, 70–75.

Alf, E. F. and R. G. Graf (2002), 'A new maximum likelihood estimator for the population squared multiple correlation', *Journal of Educational and Behavioral Statistics* **27**, 223–235.

Algina, J. (1999), 'A comparison of methods for constructing confidence intervals for the squared multiple correlation coefficient', *Multivariate Behavioral Research* **34**, 493–504.

Algina, J. and H. J. Keselman (1999), 'Comparing squared multiple correlation coefficients: Examination of a confidence interval and a test of significance', *Psychological Methods* **4**, 76–83.

Azen, R. and D. A. Sass (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. and D. V. Budescu (2003), 'The dominance analysis approach for comparing predictors in multiple regression', *Psychological Methods* **8**, 129–148.

Azen, R., D. V. Budescu and B. Reiser (2001), 'Criticality of predictors in multiple regression', *British Journal of Mathematical and Statistical Psychology* **54**, 201–225.

Bradley, J. V. (1978), 'Robustness?', *British Journal of Mathematical and Statistical Psychology* **31**, 144–152.

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.

Burdick, R. K. and F. A. Graybill (1992), *Confidence Intervals on Variance Components*, New York: Dekker.

Cameron, A.C. and F.A.G. Windmeijer (1997), 'An r-squared measure of goodness of fit for some common nonlinear regression models', *Journal of Econometrics* **77**, 329–342.

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.

Claudy, J. G. (1978), 'Multiple regression and validity estimation in one sample', *Applied Psychological Measurement* **2**, 595–607.

Cohen, J., P. Cohen, S. G. West and L. S. Aiken (2003), *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences (Third Edition)*, Mahwah, NJ:,Erlbaum.

Coleman, L S., E. Q. Campbell, C. J. Hobson, J. McPartland, A. M. Mood, F. D. Weinfetd and R. L. York (1966), *Equality of Educational Opportunity*, U.S. Government Printing Office, Washington, D.C.

DiCiccio, T. J. and B. Efron (1996), 'Bootstrap confidence intervals', *Statistical Science.* **11**, 189–228.

Donner, A. and G. Y. Zou (2002), 'Interval estimation for a difference between intraclass kappa statistics', *Biometrics* **58**, 209–214.

Donner, A. and G. Y. Zou (2010), 'Closed-form confidence intervals for functions of the normal mean and standard deviation'.
URL: *http://smm.sagepub.com/content/early/2010/09/07/0962280210383082.full.pdf*

Donner, A. and G. Y. Zou (2011), 'Estimating simultaneous confidence intervals for multiple contrasts of proportions by the method of variance estimates recovery', *Statistics in Biopharmaceutical Research* **3**, 320–335.

Drasgow, Fritz and N. J. Dorans (1982), 'Robustness of estimators of the squared multiple correlation and squared cross-validity coefficient to violations of multivariate normality', *Applied Psychological Measurement* **6**, 185–200.

Dunkin, M. J. (1978), 'Student characteristics, classroom processes, and student achievement', *Journal of Educational Psychology* **10**, 1002–1013.

Efron, B. (1979), 'Bootstrap methods: Another look at the Jackknife', *The Annals of Statistics* **7**, 1–26.

Efron, B. (1981), 'Nonparametric standard errors and confidence intervals', *The Canadian Journal of Statistics* **9**, 139–172.

Efron, B. (1987), 'Better bootstrap confidence intervals', *Journal of the American Statistical Association* **82**, 171C185.

Efron, B. (1988), 'Bootstrap confidence intervals: good or bad?', *Psychological Bulletin* **104**, 293–296.

Efron, B. and R. J. Tibshirani (1993), *An Introduction to the Bootstrap*, Chapman & Hall/CRC Press, Boca Raton, FL.

Exposition, 1993 Statistical Graphics (2005), 'Healthy breakfast'.
URL: *http://lib.stat.cmu.edu/DASL/Stories/HealthyBreakfast.html*

Ezekiel, M. (1929), 'The application of the theory of error to multiple and curvilinear correlation', *Proceedings Supplement, American Statistical Association Journal* **24**, 99–104.

Ezekiel, M. (1930), *Methods of Correlational Analysis*, Wiley, New York.

Fisher, R.A. (1928), 'The general sampling distribution of the multiple correlation coefficient', *Proceedings of the Royal Society of London, Series A* **121**, 654–673.

Gajjar, A. V. (1967), 'Limiting distributions of certain transformations of multiple correlation coefficients', *Metron* **26**, 189–193.

Graf, R.G. and E.F. Alf (1999), 'Correlations redux: asymptotic confidence limits for partial and squared multiple correlations', *Applied psychological measurement* **23**, 116–119.

Gurland, J. (1968), 'A relatively simple form of the distribution of the multiple correlation coefficient', *Journal of the Royal Statistical Society, Series B* **30**, 276–283.

Hagquist, C. and M. Stenbeck (1998), 'Goodness of fit in regression analysis-R^2 and G^2 reconsidered', *Quality & Quantity* **32**, 229–245.

Hall, P. (1992), *The Bootstrap and Edgeworth Expansion*, Wiley, New York: Springer-Verlag.

Hedges, L.V. and I. Olkin (1981), 'The asymptotic distribution of commonality components', *Psychometrika* **46**, 331–336.

Hedges, L.V. and I. Olkin (1983), 'Joint distributions of some indices based on correlation coefficients', *Studies in econometrics, time series, and multivariate analysis* pp. 437–454.

Helland, I. S. (1987), 'On the interpretation and use of R^2 in regression analysis', *Biometrics* **43**, 61–69.

Herzberg, P. A. (1969), 'The parameters of cross-validation', *Psychometrika* **34**, 1–68.

Hotelling, H. (1953), 'New light on the correlation coefficient and its transformation', *J. Roy. Statist. Soc. B* **15**, 193–232.

Howe, W.G. (1974), 'Approximate confidence limits on the mean of $X + Y$ where X and Y are two tabled independent random variables', *Journal of the American Statistical Association* **69**, 789–794.

Huberty, C. J and S. A. Mourad (1980), 'Estimation in multiple correlation/prediction.', *Educational and Psychological Measurement* **40**, 101–112.

Kerlinger, F. N. and E. J. Pedhazur (1973), *Multiple Regression in Behavioral Research*, Holt, Rinehart and Winston, New York, USA.

Khatri, C. G. (1966), 'A note on a large sample distribution of a transformed multiple correlation coefficient', *Annals of the Institute of Statistical Mathematics* **18**, 375–380.

Kramer, K. H. (1963), 'Tables for constructing confidence limits on the multiple correlation coefficients', *Journal of the American Statistical Association* **58**, 1082–1085.

Lee, Y. S. (1971), 'Some results on the sampling distribution of the multiple correlation coefficients', *ournal of the Royal Statistical Society, Series B* **33**, 117–130.

Lee, Y. S. (1972), 'Tables of upper percentage points of the multiple correlation coefficients', *Biometrika* **59**, 175–189.

Li, Y. H., J. J. Koval, A. Donner and G. Y. Zou (2010), 'Interval estimation for the area under the receiver operating characteristic curve when data are subject to error', *Statistics in Medicine* **29**, 2521–2531.

Mardia, K.V. and P.J. Zemroch (1975), 'Algorithm as84: Measures of multivariate skewness and kurtosis', *Applied Statistics* **24**, 262–264.

Mood, A. M. (1971), 'Partitioning variance in multiple regression analyses as a tool for developing learning models', *American Educational Research Journal* **8**, 191–202.

Newcombe, R.G. (1998), 'Interval estimation for the difference between independent proportions: Conparison of eleven methods', *Statistics in Medicine* **17**, 873–890.

Newcombe, R.G. (2011), 'Propagating imprecision: combining confidence intervals from independent sources', *Communications in Statistics - Theory and Methods,* **40**, 3154–3180.

Newton, R. G. and D. J. Spurrell (1969*a*), 'A development of multiple regression for the analysis of routine data', *Applied Statistics* **16**, 51–64.

Newton, R. G. and D. J. Spurrell (1969*b*), 'Examples of the use of elements for clarifying regression analysis', *Applied Statistics* **16**, 165–172.

Nierenberg, D. W., T. A. Stukel, J. A. Baron, B. J. Dain and E. R. Greenberg (1989), 'Determinants of plasma levels of beta-carotene and retinol', *American Journal of Epidemiology* **130**, 511–521.

Ohtani, K. (2000), 'Bootstrapping r^2 and adjusted r^2 in regression analysis', *Economic modelling* **17**, 473–483.

Olkin, I. and J. W. Pratt (1958), 'Unbiased estimation of certain correlation coefficients', *Annals of Mathematical Statistics* **29**, 201–211.

Olkin, I. and J.D. Finn (1995), 'Correlations redux', *Psychological Bulletin* **118**, 155–164.

Olkin, I and M Siotani (1976), Asymptotic distribution of functions of a correlation matrix, *in* 'S. Ikeda (ed). Essays in probability and statistics (pp. 235–251)', Tokyo, Japan: Shinko Tsusho Co, Ltd.

Pearson, K. (1912), 'On the general theory of the influence of selection on correlation and variation', *Biometrika* **8**, 437–443.

Pearson, K. and L.N.G. Filon (1898), 'Mathematical contributions to the theory of evolution. iv. on the probable errors of frequency constants and on the influence of random selection on variation and correlation', *Philosophical Transactions of the Royal Society of London. Series A* **191**, 229–311.

Pedhazur, E. J. (1975), 'Analytic methods in studies of educational effects', *Review of Research in Education* **3**, 243–286.

Pedhazur, E. J. (1982), *Multiple Regression in Behavioral Research: Explanation and Prediction*, 2rd. edn, New York: Holt, Rinehart and Winston, Inc.

Raju, N. S., R. Bilgic, J. E. Edwards and P. F. Fleer (1997), 'Methodology review: Estimation of population validity and cross-validity, and the use of equal weights', *Applied Psychological Measurement* **21**, 291–305.

Ramasundarahettige, C.F., A. Donner and G. Y. Zou (2009), 'Confidence interval construction for a difference between two dependent intraclass correlation coefficients', *Statistics in Medicine* **28**, 1041–1053.

Robey, R. R. and R. S. Barcikowski (1992), 'Type i error and the number of iterations in monte carlo studies of robustness', *British Journal of Mathematical and Statistical Psychology* **45**, 283–288.

SAS (2010), 'Sample 24982: Jackknife and bootstrap analyses'.
URL: *http://support.sas.com/kb/24/982.html*

Scarborough, J. B. (1966), *Numerical Mathematical Analysis. (6th Edition)*, Johns Hopkins Press., Baltimore.

Schenker, N. (1985), 'Qualms about bootstrap confidence intervals', *Journal of the American Statistical Association* **80**, 360–361.

Shao, J., & Tu D. (1995), *The Jackknife and Bootstrap*, Springer-Verlag, New York.

Smithson, M. (2003), *Confidence Intervals. (Sage University Papers Series on Quantitative Applications in the Social Sciences)*, 07-140, Thousand Oaks, CA: Sage.

Steel, R. G. D. and J. H. Torrie (1960), *Principles and Procedures of Statistics*, McGraw-Hill, New York.

Steiger, J. H. (2009), 'Statpower'.
URL: *http://www.statpower.net*

Steiger, J. H. and A. R. Hakstian (1982), 'The asymptotic distribution of elements of a correlation matrix: Theory and application', *British Journal of Mathematical and Statistical Psychology* **35**, 208–215.

Steiger, J. H. and M.W. Browne (1984), 'The comparison of interdependent correlations between optimal linear composites', *Psychometrika* **49**, 11–24.

Steiger, J. H. and R. T Fouladi (1992), 'R2: A computer program for interval estimation, power calculation, and hypothesis testing for the squared multiple correlation', *Behavior Research Methods, Instruments, and Computers* **4**, 581–582.

Stukel, T. (2003), 'Determinants of plasma retinol and beta-carotene levels'.
URL: *http://lib.stat.cmu.edu/DASL/Stories/HealthyBreakfast.html*

Wang, H. and S. C. Chow (2002), 'A practical approach for comparing means of two groups without equal variance assumption', *Statistics in Medicine* **21**, 3137–3151.

Wherry, R. J. (1931), 'A new formula for predicting the shrinkage of the coefficient of multiple correlation', *Annals of Mathematical Statistics* **2**, 440–457.

Wishart, J. (1931), 'The mean and second moment coefficient of the multiple correlation coefficient in samples from a normal population', *Biometrika* **22**, 353–361.

Zou, G. (2007), 'Supplemental material for toward using confidence intervals to compare correlations'.
URL: *http://dx.doi.org/10.1037/1082-989x.12.4.399.supp*

Zou, G. Y. (2007), 'Toward using confidence intervals to compare correlations', *Psychological Methods* **12**, 399–413.

Zou, G. Y. (2008), 'On the estimation of additive interaction by use of the four-by-two table and beyond', *American Journal of Epidemiology* **168**, 212–224.

Zou, G. Y. (2010), 'Confidence interval estimation under inverse sampling', *Computational Statistics & Data Analysis* **54**, 55–64.

Zou, G. Y. (2011), 'Confidence interval estimation for the BlandcAltman limits of agreement with multiple observations per individual'.
URL: *http://www.sagepub.com/content/early/2011/06/17/0962280211402548.full.pdf*

Zou, G. Y. and A. Donner (2008), 'Construction of confidence limits about effect measures: a general approach', *Statistics in Medicine* **27**, 1693–1702.

Zou, G Y and A Donner (2010), A generalization of Fiellers Theorem for ratios of non-normal variables and some practical applications, *in* H.Schuster and W.Metzger, eds, 'Biometrics: Methods, Applications and Analyses', NOVA Publishers, pp. 197–216.

Zou, G. Y., C. Y. Huo and J. Taleban (2009), 'Simple confidence intervals for lognormal means and their differences with environmental applications', *Environmetrics* **20**, 172–180.

Zou, G. Y., J. Taleban and C. Y. Huo (2009), 'Confidence interval estimation for lognormal data with application to health economics', *Computational Statistics & Data Analysis.* **53**, 3755–3764.

Zou, G. Y., W. Huang and X. Zhang (2009), 'A note on confidence interval estimation for a linear function of binomial proportions', *Computational Statistics & Data Analysis* **53**, 1080–1085.

VITA

NAME Li Tan

EDUCATION Department of Epidemiology & Biostatistics
University of Western Ontario
London, Ontario, Canada
2009–2012, M.Sc.

Department of Civil & Environmental Engineering
University of Western Ontario
London, Ontario, Canada
2007–2008, M.Sc.

Division of Probability & Statistics
Department of Mathematics
Beijing Institute of Technology
Beijing, PR China
2003–2006, M.Sc.

Department of Mathematics & Applied Mathematics
Hubei Normal University
PR China
1999–2003, B.Sc.

AWARDS First-Class Scholarship, Hubei Normal University, 2001–2003.

Excellent Student Award, Hubei Normal University, 2003

EXPERIENCE Information Security Engineer
CST Converging Stream Technology Inc.
Beijing, China
2004–2005

PUBLICATIONS

Tan, Li and Hong, Hanping. (2010). Influence of spatial inhomogeneity of tornado occurrence on estimated tornado hazard. *Canadian Journal of Civil Engineering* **37**(2): 279 – 289.

Tan, Li and Tian, Yubin. (2005). Confidence intervals for a binomial proportion and saddlepoint approximations. *9th Probability and Statistics Conference in Beijing & Tianjin Area*, Beijing, China.