PROPERTIES OF AN R^2 STATISTIC FOR FIXED EFFECTS IN THE LINEAR MIXED MODEL FOR LONGITUDINAL DATA

Jeanine M. Matuszewski

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biostatistics, Gillings School of Global Public Health.

Chapel Hill
2011

Approved by:

Lloyd J. Edwards

Pranab K. Sen

Bahjat F. Qaqish

John Preisser

James Michael Bowling

ABSTRACT

JEANINE M. MATUSZEWSKI. Properties of an R^2 Statistic for Fixed Effects in the Linear Mixed Model for Longitudinal Data.
(Under the direction of Dr. Lloyd J. Edwards)

The R^2 statistic has become a widely used tool when analyzing data in the linear univariate setting. Many R^2 statistics for the linear mixed model exist but their properties are not well established. The purpose of this dissertation is to examine the properties and performance of R^2_β for fixed effects in the linear mixed model.

Two approaches are considered in deriving approximations for the mean and variance of R^2_β under the null and alternative hypotheses which include using the Beta distribution and a Taylor series approximation. Test statistics based on these two approximations of the mean and variance are proposed and compared to the overall F test for fixed effects in the linear mixed model. Using simulations, the Type I error rate of the proposed R^2_β test statistics derived from the Beta distribution was equivalent to the Type I error rate for the overall F test. The Type I error rates for the test statistic based on the Taylor series approximation moments were slightly inflated.

The impact of covariance structure misspecification, estimation technique, and denominator degrees of freedom method on the asymptotic properties of R^2_β are explored. For the simulation studies examined, the estimation technique does not impact the values of R^2_β. The values and asymptotic properties of R^2_β using Kenward-Roger, containment and Satterthwaite methods are greatly impacted by covariance structure misspecification whereas R^2_β using the residual method is not. Simulations illustrate the impact of underspecification of the covariance structure as compound symmetric when the true structure is more complex.

The asymptotic R_β^2's for the underspecified models using Kenward-Roger degrees of freedom are smaller than the true asymptotic R_β^2's. Conversely, the asymptotic R_β^2's for the underpecified models using residual methods are larger than the true asymptotic R_β^2.

The semi-parital R_β^2 for the four denominator degrees of freedom are computed and compared to the corresponding model R_β^2 in both a real world example and simulation study. The semi-partial R_β^2 using residual degrees of freedom never exceeded the model R_β^2, but the semi-partial R_β^2 using the other three methods sometimes exceeded the model R_β^2. R_β^2 is also evaluated as a fixed effects model selection tool. The performance of R_β^2 is poor; so an adjusted R_β^2 is created for purposes of fixed effects model selection. The adjusted R_β^2 using residual degrees of freedom outperformed the adjusted R_β^2 defined using the other methods.

ACKNOWLEDGEMENTS

There are many people to whom I must show my appreciation. First, to my dissertation advisor, Lloyd Edwards. Thank you for your guidance and advice throughout this process. It has been invaluable. I must also thank my committee members, Mike Bowling, P.K. Sen, Bahjat Qaqish, and John Preisser for their time with regards to my research. To my family, thank you for your constant and unwavering support and encouragement. Your unconditional love has helped me immensely on this difficult journey. Finally, to my husband, you have been with me through this entire adventure, the good times and the bad. I would not have been able to complete this dissertation without your love and support. Finally, we can take on new adventures!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
INTRODUCTION AND LITERATURE REVIEW

1.1 Introduction

The R^2 statistic used in linear regression is well known. Its popularity has lead to the development of R^2 statistics for other types of models, including logistic regression, proportional hazards regression, and the linear mixed model. There is a growing interest and need for R^2 statistics in the linear mixed model because the linear mixed model is an important tool used to analyze continuous longitudinal data. It is an extension of the linear univariate model which accounts for the correlated measurements of a particular unit. While the R^2 statistic and its properties are well developed for linear regression (in the univariate model), there are many different R^2 statistics for the linear mixed model with varying properties. Throughout this chapter, the development and properties of R^2 statistics for the univariate, multivariate, and linear mixed models will be highlighted.

1.2 The R^2 Statistic in Linear Univariate Model

The R^2 statistic for the linear univariate model is well developed and utilized extensively in biomedical research. Researchers in other disciplines find the statistic appealing because it provides an easy to understand way of explaining how the model fits the data. Another appealing feature is that it has several equivalent definitions. It is defined as a goodness of fit measure, squared multiple correlation coefficient and the coefficient of determination.

1.2.1 Model Notation

The linear univariate model for N subjects and p covariates is represented as follows (Muller and Stewart, 2006, p. 40-41) :

$$\underset{(N\times 1)}{\boldsymbol{y}} = \underset{(N\times p)}{\boldsymbol{X}} \underset{(p\times 1)}{\boldsymbol{\beta}} + \underset{(N\times 1)}{\boldsymbol{e}} \tag{1.1}$$

$$\boldsymbol{e} \sim N_N\left(\mathbf{0}, \sigma^2 \boldsymbol{I}_N\right)$$

In model 1.1, $\boldsymbol{y}$ is a vector of independent responses. $\boldsymbol{X}$ is a known constant design matrix of covariates, and $\boldsymbol{\beta}$ is an unknown vector of population parameters.

1.2.2 Formulae and Interpretations of R^2 in Linear Regression

There are many formulae for R^2 in the linear univariate model. Kvalseth (1985) gives several different expressions for R^2 statistics that appear throughout the literature. Two of the R^2's (R_5^2 and R_6^2) are the squared multiple correlation coefficient between the regressand and the regressors and the squared correlation coefficient between $\boldsymbol{y}$ and $\widehat{\boldsymbol{y}}$. These and the other expressions of R^2 are equivalent when using linear least squares regression for a model that includes an intercept (Draper and Smith, 1998, chaps. 1-2).

The predicted values, $\widehat{\boldsymbol{y}}$, in the formulae of R^2 described in Kvalseth (1985), are computed by fitting model (1.1). Defining R^2 in this way, does not emphasize the fact that R^2 is really a comparison of two models. Anderson-Sprecher (1994) suggest that defining R^2 in terms of a model comparison perspective is simpler and minimizes the potential misinterpretations and incorrect usages of the statistic. Muller and Fettermann (2002, p 226) define R^2 using a model comparison perspective as follows:

$$\begin{aligned} R^2 &= R^2(\boldsymbol{y} | \boldsymbol{x}_1 \ldots \boldsymbol{x}_p) \\ &= \frac{[\text{SSE}(\beta_0)\text{-SSE}(p)]}{\text{SSE}(\beta_0)} \end{aligned}$$

where $\text{SSE}(\beta_0)$ is the error sum of squares of the model only including the intercept and $\text{SSE}(p)$ is the error sum of squares of the model with all p variables. Calculating R^2 in the above equation is equivalent to comparing two models. There is a full model which contains all of the p covariates including the intercept and a null model which contains only an intercept.

Equivalently, Draper and Smith (1998, p 141) describe R^2 as follows,

$$R^2 = \frac{\nu_1 F}{\nu_2 + \nu_1 F} \tag{1.2}$$

where F represents the test of the null hypothesis $H_0: \beta_1 = \beta_2 = ... = \beta_{p-1}$ versus the alternative, $H_A:$ at least one of the β's (excluding β_0) is zero (i.e. The null hypothesis is that all non-intercept parameters are equal to zero.); $\nu_1$is the numerator degrees of freedom of the F statistic which is $p-1$; ν_2 is the denominator degrees of freedom of the F statistic which is $N-p$. Under H_0, since F is distributed as $F(p-1,\ N-p)$, then it is known that R^2 is is distributed as a Beta$\left(\frac{p-1}{2},\ \frac{N-p}{2}\right)$.

R^2 is a one-to-one function of the overall F test. The partial R^2 can be used to test an individual covariate given the other covariates in the model. Muller and Fetterman (2002) describe the full multiple partial correlation as the correlation between two variables with both adjusted for other variables. The partial correlation can be denoted as $\rho[(\boldsymbol{y},\ x_j)|\{x_1,\ x_2, ..., x_{j-1}, ...,\ x_p\}]$.

The formulae presented above give rise to the various interpretations of the R^2 statistic in linear regression. The most common way of interpreting R^2 is as the proportion of variation in the response, $\boldsymbol{y}$, that is described by the covariates, $\boldsymbol{X}$. Alternatively, R^2 is interpreted as the square of the correlation coefficient between the observed outcome and the predicted outcomes. One can also express R^2 as a measure of the overall linear association of one dependent variable with several independent variables.

1.2.3 Properties of the Univariate R^2 statistic

Cameron and Windmeijer (1996) and Kvalseth (1985) define properties of an ideal R^2 statistic. The properties proposed by Kvalseth (1985) are as follows:

1. It is a measure of goodness of fit and provides a reasonable interpretation.
2. It is dimensionless.
3. The endpoints of R^2 correspond to complete lack of fit and perfect fit.
4. It should be general enough to be applied to any model.
5. The model fitting technique should not effect the R^2.
6. Comparisons can be made between R^2 values that are computed using the same dataset.
7. Other goodness of fit measures are relatively comparable to the R^2.
8. The residuals that are positive are weighted the same as the residuals that are negative.

Cameron and Windmeijer (1996) describe four additional properties R^2 statistics should have. Those properties are:

1. R^2 does not decrease as covariates are added.
2. R^2 based on the residual sum of squares coincides with R^2 based on explained sum of squares.
3. The R^2 statistic corresponds to a significance test.
4. The interpretation of R^2 is based on the information content of the data.

Barten (1962) and Montgomery and Morrison (1973) both show that the R^2 for model (1.1) is positively biased estimator for the true coefficient of determination. The null expectation of R^2 in linear regression using least squares estimation is

$$E\left(R^2|H_0 : \rho^2_{\boldsymbol{y}|\boldsymbol{X}} = 0\right) = \frac{p-1}{N-1} > 0.$$

When the total number of observations, N is small, there is potential for large bias. As a result of this positive bias, an adjusted version of R^2 has been proposed which takes into account the number of covariates in the model. The adjusted version of R^2 is defined as

$$R_a^2 = 1 - \frac{(N-1)(1-R^2)}{(N-p-1)}.$$

The adjusted R^2 increases when a covariate improves the model, whereas the unadjusted R^2 always increases when covariates are added to the model (represented by the first property in Cameron and Windmeijer (1996). As a result, the adjusted R^2 can be used to compare nested models to determine which model is a better fit to the data.

Another important property is that R^2 converges to the true coefficient of determination as the sample size increases (Barten, 1962). Helland (1987) states that under weak conditions, as $N \to \infty$,

$$R^2 \to_{\text{a.s.}} \frac{\boldsymbol{\beta}'\mathbf{S}_{\text{x}}\boldsymbol{\beta}}{\boldsymbol{\beta}'\mathbf{S}_{\text{x}}\boldsymbol{\beta} + \sigma^2}$$

where $\mathbf{S}_{\text{x}} = \frac{1}{N-1}(\boldsymbol{X} - \mathbf{1}\overline{\boldsymbol{x}})'(\boldsymbol{X} - \mathbf{1}\overline{\boldsymbol{x}})$; $\overline{\boldsymbol{x}}$ is defined as the $1 \times p$ matrix of means of the p covariates, and $\mathbf{1}$ is an $N \times 1$ vector of ones.

1.3 The General Linear Multivariate Model

Correlated data is fairly common in biomedical and social science research. Multivariate models are one tool used to analyze a set of correlated responses. The correlated responses may be the same measurement taken repeatedly over time (longitudinal data), or alternatively there could be different measurements of multiple correlated responses. It is necessary to take into account the correlation when performing estimation of model parameters and conducting inference using the data.

1.3.1 Model Notation

The general linear multivariate model is (Muller and Stewart, 2006, p 58):

$$\underset{(N \times p)}{\boldsymbol{Y}} = \underset{(N \times q)}{\boldsymbol{X}} \underset{(q \times p)}{\boldsymbol{B}} + \underset{(N \times p)}{\boldsymbol{E}} \tag{1.3}$$

where $\boldsymbol{Y}$ is a matrix of repeated measure outcomes where the N rows of $\boldsymbol{Y}$ are mutually independent; $\boldsymbol{X}$ is a design matrix; $\boldsymbol{B}$ is a matrix of fixed parameters. The random matrix $\boldsymbol{E}$ has the following properties (1) $E[\boldsymbol{E}] = \mathbf{0}$ and (2) $V[\text{vec}(\boldsymbol{E}')] = V[\text{vec}(\boldsymbol{Y})] = \boldsymbol{I}_N \otimes \boldsymbol{\Sigma}$. The assumptions of the multivariate model are that there is neither missing nor mistimed data, the same design matrix applies to all response variables, and the covariance matrix is not based on the data. Assuming a full rank design, the least squares estimate of $\boldsymbol{B}$ is $\widehat{\boldsymbol{B}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$ which is a unique and unbiased estimator of $\boldsymbol{B}$ (Muller and Stewart, 2006, p 65).

1.3.2 General Linear Hypothesis Tests

The general linear hypothesis for the multivariate model (1.3) is defined as

$$H_0 : \underset{(a \times b)}{\boldsymbol{\Theta}_0} = \underset{(a \times q)}{\boldsymbol{C}} \underset{(q \times p)}{\boldsymbol{B}} \underset{(p \times b)}{\boldsymbol{U}}$$

where $\boldsymbol{C}$, $\boldsymbol{U}$ and $\boldsymbol{\Theta}_0$ represent fixed, known constraints on $\boldsymbol{B}$. If analyzing repeated measures data, the $\boldsymbol{C}$ matrix represents the between subjects contrasts and the $\boldsymbol{U}$ matrix represents the within subjects contrasts. In order for the hypothesis to be testable three conditions must be met. With $\boldsymbol{M} = \boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-}\boldsymbol{C}'$, the conditions are

$$\begin{aligned} &(1)\ \text{rank}(\boldsymbol{M}) = a \\ &(2)\ \boldsymbol{C} = \boldsymbol{C}(\boldsymbol{X}'\boldsymbol{X})^{-}(\boldsymbol{X}'\boldsymbol{X}) \\ &(3)\ \text{rank}(\boldsymbol{U}) = b. \end{aligned}$$

For the multivariate model, there are two groups of hypothesis tests commonly used: multivariate approach to repeated measures (MULTIREP) tests and univariate approach to repeated measures (UNIREP) tests. The MULTIREP tests include the Hotelling-Lawley trace, Pillai-Bartlett Trace, Wilks' Lambda and Roy's Largest root. The MULTIREP tests are of particular interest because each test has a corresponding multivariate measure of association. A multivariate measure of association is defined as the proportion of variance controlled by the multivariate hypothesis. For the general linear hypothesis test given in (3.2), the hypothesis sum of squares is $\boldsymbol{S}_h = \left(\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}_0\right)' \boldsymbol{M}^{-1}\left(\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}_0\right)$. The error sum of squares is $\boldsymbol{S}_e = \boldsymbol{U}'\widehat{\boldsymbol{E}}'\widehat{\boldsymbol{E}}\boldsymbol{U}$ where $\widehat{\boldsymbol{E}} = \boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{B}}$. These two quantities are used to compute the MULTIREP test statistics and corresponding measures of association. These tests and their measures of association are given in Table 1.1. In Table 1.1, $s = \min(a, b)$ and g is defined by Muller and Stewart (2006, p 71) as

$$g = \begin{cases} 1 & a^2b^2 \leq 4 \\ \left[(a^2b^2 - 4)/(a^2 + b^2 - 5)\right]^{1/2} & \text{otherwise} \end{cases}.$$

The four MULTIREP tests can be expressed as a one-to-one function of each other under the null hypothesis and when $s = 1$ (Muller and Stewart 2006, pg 71). When $s > 1$, the four multivariate statistics do not have this property. Furthermore, there is not one single multivariate test statistic that satisfies all of the standard optimality criteria for the more complex constrasts ($s > 1$). The exact distributions of these statistics are only known for special cases when $s > 1$. Johnson and Wichern (1992) provide conditions where functions of Wilks' Lambda have an F distribution and exact tests are possible. Anderson (2003, p 330) shows that the Hotelling-Lawley trace criterion converges in distribution to the χ^2 distribution.

It is also important to know the distribution of the test statistics under the alternative hypothesis to calculate the power of each of the tests. Anderson (2003, p 334) describes how the power of each of the MULTIREP tests approaches one since the noncentrality parameters of the tests tends to infinity. As a result, to compare the various MULTIREP tests, it is more informative to consider a sequence of alternatives such that the powers of the tests will vary. Sen and Singer (1993, p 238) define a sequence of local Pitman-type alternatives as

$$H_{A_n}:\ \boldsymbol{\Theta}_n = \boldsymbol{\Theta}_0 + \frac{\boldsymbol{\Delta}}{\sqrt{n}},$$

Defining a sequence of local alternatives in this way shows how the alternative is not held fixed but allowed to get closer and closer to the null hypothesis as the sample size increases.

Power of the four MULTIREP tests can be computed using a sequence of Pitman-type alternatives and quantized limits. Glueck (1997), Glueck and Muller (2003) and Anderson (2003) define quantized limits by first, defining a positive integer m, and let $N(m) = mN$. A quantized limit is then defined such that as $m \to \infty$, then through a sequence of quantized steps of size N, $N(m) \to \infty$. It is important to recognize that N remains fixed as $N(m)$ increases. Muller et al. (2007) provide new power approximations for all four UNIREP tests which eliminate the inaccuracies in existing methods.

The bias of various multivariate measures of association for multivariate analysis of variance model was examined by Kim and Olejnik (2005) and Steyn and Ellis (2009). These authors conclude that all of the multivariate measures of association examined, including Wilks' lambda, Hotelling-Lawley trace and Pillai's trace criterion, are biased with the bias increasing when the sample size is small, and when the number of outcome variables increases. Steyn and Ellis (2009) introduce the multivariate measures of association as effect sizes and additionally show that they are biased when the effect size is small.

1.4 The General Linear Mixed Model

The linear mixed model is another way to analyze correlated response data. Laird and Ware (1982) introduced the general linear mixed model for longitudinal data, based on the work of Harville (1977). This two-stage random effects model easily accommodates unbalanced data whereas multivariate models can not. The linear mixed model also differs from the multvariate model in that it is able to handle mistimed data and allows the structure of the covariance matrix to be specified by the data instead of assuming its structure.

1.4.1 Model Notation

The linear mixed model (LMM) is a powerful statistical tool for analyzing longitudinal data. The linear mixed model for an independent sampling unit i is (Muller and Stewart, 2006, Chapter 5):

$$\underset{(n_i \times 1)}{\boldsymbol{y}_i} = \underset{(n_i \times p)}{\boldsymbol{X}_i} \underset{(p \times 1)}{\boldsymbol{\beta}} + \underset{(n_i \times k)}{\boldsymbol{Z}_i} \underset{(k \times 1)}{\boldsymbol{d}_i} + \underset{(n_i \times 1)}{\boldsymbol{e}_i} \; ; \; i = 1, ..., m \tag{1.4}$$

Here, $\boldsymbol{y}_i$ is a vector of observations for subject i; $\boldsymbol{X}_i$ is a known, constant design matrix for subject i, with full column rank p; $\boldsymbol{\beta}$ is a vector of unknown, constant, population parameters; $\boldsymbol{Z}_i$ is a known, constant design matrix for subject i corresponding to the random effects $\boldsymbol{d}_i$, with rank k; $\boldsymbol{d}_i$ is a vector of unknown, random individual parameters; $\boldsymbol{e}_i$ is an $n_i \times 1$ vector of random errors. Also, $N = \sum_{i=1}^{m} n_i$. Throughout, $\boldsymbol{d}_i$ is Gaussian with mean $\mathbf{0}$ $(k \times 1)$ and covariance $\boldsymbol{\Sigma}_{d_i}(\boldsymbol{\tau}_d)$, independently of Gaussian $\boldsymbol{e}_i$ $(n_i \times 1)$ with mean $\mathbf{0}$ $(n_i \times 1)$ and covariance $\boldsymbol{\Sigma}_{e_i}(\boldsymbol{\tau}_e)$ $(n_i \times n_i)$, so that

$$\boldsymbol{\mathcal{V}}\left(\begin{bmatrix} \boldsymbol{d}_i \\ \boldsymbol{e}_i \end{bmatrix}\right) = \begin{pmatrix} \boldsymbol{\Sigma}_{d_i}(\boldsymbol{\tau}_d) & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{e_i}(\boldsymbol{\tau}_e) \end{pmatrix}$$

Here $\boldsymbol{\mathcal{V}}(\,\cdot\,)$ is the covariance operator, $\boldsymbol{\Sigma}_{d_i}(\boldsymbol{\tau}_d)$ is a $k \times k$ positive-definite, symmetric covariance matrix of the random effects, and $\boldsymbol{\Sigma}_{e_i}(\boldsymbol{\tau}_e)$ is an unknown $n_i \times n_i$, constant positive-definite matrix. Under the assumptions, $\boldsymbol{\mathcal{V}}(\boldsymbol{y}_i)$ can be expressed as $\boldsymbol{\Sigma}_i(\boldsymbol{\tau}) = \boldsymbol{Z}_i \boldsymbol{\Sigma}_{d_i}(\boldsymbol{\tau}_d) \boldsymbol{Z}_i' + \boldsymbol{\Sigma}_{e_i}(\boldsymbol{\tau}_e)$. Generally, it is assumed that the covariance $\boldsymbol{\Sigma}_i$ can be characterized by a finite set of parameters represented by an $r \times 1$ vector $\boldsymbol{\tau}$, which consists of the unique parameters in $\boldsymbol{\Sigma}_{d_i}(\boldsymbol{\tau}_d)$ and $\boldsymbol{\Sigma}_{e_i}(\boldsymbol{\tau}_e)$. Additionally, $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\tau}')'$ will be the $s \times 1$ vector of parameters for model (1.4), where $s = p + r$.

1.4.2 Estimation Techniques

There are primarily two estimation techniques used in the linear mixed model: maximum likelihood (ML) and restricted-maximum likelihood (REML) (Laird and Ware, 1982; Jennrich and Schluchter, 1986). The technique used plays an important role in both inference and estimation. The marginal log-likelihood function for model (1.4) is

$$l_{\mathrm{ML}}(\boldsymbol{\beta}, \boldsymbol{\tau}) = -\frac{N}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{m}\log|\boldsymbol{\Sigma}_i(\boldsymbol{\tau})| - \frac{1}{2}\sum_{i=1}^{m}(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})'\boldsymbol{\Sigma}_i(\boldsymbol{\tau})^{-1}(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})$$

The restricted log-likelihood function for model (1.7) is

$$l_{\mathrm{REML}}(\boldsymbol{\beta}, \boldsymbol{\tau}) = -\frac{N-p}{2}\log(2\pi) + \frac{1}{2}\log\left|\sum_{i=1}^{m}\boldsymbol{X}_i'\boldsymbol{X}_i\right| - \frac{1}{2}\sum_{i=1}^{m}\log|\boldsymbol{\Sigma}_i(\boldsymbol{\tau})| - \frac{1}{2}\log\left|\sum_{i=1}^{m}\boldsymbol{X}_i'\boldsymbol{\Sigma}_i(\boldsymbol{\tau})^{-1}\boldsymbol{X}_i\right| - \frac{1}{2}\sum_{i=1}^{m}(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})'\boldsymbol{\Sigma}_i(\boldsymbol{\tau})^{-1}(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})$$

The expression for the estimate of $\boldsymbol{\beta}$ is given by

$$\widehat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{m}\boldsymbol{X}_i'\widehat{\boldsymbol{\Sigma}}_i(\widehat{\boldsymbol{\tau}})^{-1}\boldsymbol{X}_i\right)^{-1}\sum_{i=1}^{m}\boldsymbol{X}_i'\widehat{\boldsymbol{\Sigma}}_i(\widehat{\boldsymbol{\tau}})^{-1}\boldsymbol{y}_i.$$

The expression for $\widehat{\boldsymbol{\beta}}$ is the same for ML and REML, but the estimates differ based on the estimation of $\boldsymbol{\Sigma}_i(\boldsymbol{\tau})$. The estimator, $\widehat{\boldsymbol{\beta}}$, is unbiased; however, there is no closed form expression for the variance of $\widehat{\boldsymbol{\beta}}$. The common approach is to estimate the approximate variance with

$$\widehat{V}(\widehat{\boldsymbol{\beta}}) = \left(\sum_{i=1}^{m}\boldsymbol{X}_i'\widehat{\boldsymbol{\Sigma}}_i(\widehat{\boldsymbol{\tau}})^{-1}\boldsymbol{X}_i\right)^{-1}. \tag{1.5}$$

Kackar and Harville (1984) have shown that formula (1.5) underestimates the true variance of $\widehat{\boldsymbol{\beta}}$.

1.4.3 Inference for the Linear Mixed Model

Inference in the linear mixed model has advanced considerably since it was first popularized in the seminal works of Harville (1977) and Laird and Ware (1982). There are special cases of balanced data where the exact distribution of the parameter estimates is known (Grizzle and Allen, 1969). Although generally, exact distributions are not known. Asymptotic approximations are used for inference. Laird and Ware (1982) suggest using asymptotic likelihood ratio tests for fixed effect hypothesis tests. The simulation studies in Helms (1992) show an inflated Type I error rate of the asymptotic likelihood ratio test. Welham and Thompson (1997) propose adjusted likelihood ratio tests when using REML estimation for fixed effects hypothesis tests. Another complication to using the likelihood ratio test is presented by Verbeke and Molenberghs (2000). The authors explain that the likelihood ratio test based on REML log-likelihood function, (1.10), should not be used for hypotheses involving the fixed effects.

As an alternative to the likelihood ratio test, Helms (1992) and others have proposed an approximate F test for testing the fixed effects. The approximate F tests are a Wald-type test for the hypothesis,

$$H_0 : \boldsymbol{C}\boldsymbol{\beta} = \mathbf{0} \quad \text{vs.} \quad H_\text{A} : \boldsymbol{C}\boldsymbol{\beta} \neq \mathbf{0}$$

which has the general form,

$$F\left(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\tau}}\right) = \frac{\left(\boldsymbol{C}\widehat{\boldsymbol{\beta}}\right)' \left[\boldsymbol{C}\left(\sum_{i=1}^{m} \boldsymbol{X}_i' \widehat{\boldsymbol{\Sigma}}_i(\widehat{\boldsymbol{\tau}})^{-1} \boldsymbol{X}_i\right)^{-1} \boldsymbol{C}'\right]^{-1} \left(\boldsymbol{C}\widehat{\boldsymbol{\beta}}\right)}{\text{rank}(\boldsymbol{C})}. \tag{1.6}$$

The $F\left(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\tau}}\right)$ statistic is usually approximated by an F distribution to account for the underestimation of the variance of $\widehat{\boldsymbol{\beta}}$.

There are several methods for determining the denominator degrees of freedom, ν, for the F-statistic in equation (1.6). Brief overviews of the methods are found in the MIXED procedure in the SAS system (SAS, 2004). The simplest denominator degrees of freedom are based on the residual degrees of freedom. The residual denominator degrees of freedom are $N - p$. This method is correct for designs where the outcome is independent and identically distributed (ie. ignoring covariance structure of the model.) because it is the same degrees of freedom of the error term as the conventional F test. This method should only be considered when there is a very large sample size and asymptotic distributions are good approximations. Kesselman et al. (1999a) conclude that the degrees of freedom using the residual method are too large when the covariance is not spherical.

The MIXED procedure used in SAS (2004) describes an alternative method of calculating the denominator degrees of freedom called the containment method. This method is based on the degrees of freedom in the balanced split-plot designs. The degrees of freedom are calculated by scanning the random effects terms to determine if they contain the fixed effect which is being tested. If no random effect contains the fixed effect being tested, the denominator degrees of freedom is the same as the residual method.

A more conservative method involves adjusting the degrees of freedom using the procedure developed by Satterthwaite (1946). The Satterthwaite (1946) approximation computes the denominator degrees of freedom using the chi-square distribution to approximate the distribution of the estimated variance of $\boldsymbol{C}\widehat{\boldsymbol{\beta}}$. Specifically, the estimated approximate denominator degrees of freedom for a linear combination of fixed effects estimates ($\boldsymbol{C}$ is an $1 \times p$ matrix.) are,

$$\widehat{\nu} = \frac{2\left[\boldsymbol{C}\left(\sum_{i=1}^{m} \boldsymbol{X}_i' \widehat{\boldsymbol{\Sigma}}_i(\widehat{\boldsymbol{\tau}})^{-1} \boldsymbol{X}_i\right)^{-1} \boldsymbol{C}'\right]^2}{\widehat{\mathrm{Var}}\left[\boldsymbol{C}\left(\sum_{i=1}^{m} \boldsymbol{X}_i' \widehat{\boldsymbol{\Sigma}}_i(\widehat{\boldsymbol{\tau}})^{-1} \boldsymbol{X}_i\right)^{-1} \boldsymbol{C}'\right]},$$

where the denominator is approximated using gradients (Littell, 2002). Keselman et al. (1999b) found that the F tests using Satterthwaite method of estimating denominator degrees of freedom yielded better results when compared to the residual degrees of freedom. The authors also noted that the F tests were more robust when the true covariance structure was known.

Another method of estimating the degrees of freedom which is very similar to the Satterthwaite approach was developed by Kenward and Roger (1997). These authors not only estimate the denominator degrees of freedom for the F test in the linear mixed model but also adjusted the estimates of the covariance matrix of the parameter estimates. A more accurate estimate of $\Sigma_i(\boldsymbol{\tau})$ is computed to account for the variability in $\widehat{\boldsymbol{\tau}}$ and account for the small sample bias. Kenward and Roger (1997) approximate the distribution of the F-statistic by choosing a scale λ and denominator degrees of freedom m such that $\lambda F \sim F(l, m)$ approximately.

There have been various studies done to compare these methods of estimating the denominator degrees of freedom. Alnosaier (2007) has shown special cases computing the degrees of freedom using Satterthwaite method coincides with the Kenward-Roger method. Other studies have examined when the two methods differ. Schaalje et al. (2002) found the two methods Type I error rates were affected by simulation scenarios of different imbalance, sample size, and covariance structure complexity. The Kenward-Roger method outperformed or performed similarly to the Satterthwaite approximation in all simulation scenarios. Most recently, Arnau et al. (2009) compared the Type I error rate of the F tests using Kenward-Roger, the Satterthwaite and containment degrees of freedom. The simulation results showed that the Satterthwaite approach had liberal Type I error rates, and that the Kenward Roger approach provides the best control of the Type I error rates.

A sequence of random variables $\{X_n\}$ is said to converge in distribution to X, denoted as, $X_n \to_{\mathrm{d}} X$ if the distribution functions of F_n and F of X_n and X satisfy:

$$F_n(x) \to F(x) \text{ as } n \to \infty \text{ for each continuity point } x \text{ of } F.$$

Given a random variable X with $X \sim F(\nu_1,\, \nu_2,\, \omega)$ then as $\nu_2 \to \infty$, $X \to_{\mathrm{d}} Y$ where $Y \sim \nu_1^{-1}\chi^2(\nu_1,\, \omega)$. Therefore if we assume $F\left(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\tau}}\right)$ has an F distribution, applying large sample theory, gives $F\left(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\tau}}\right) \rightarrow_{\mathrm{d}} a^{-1}\chi^2(a, \omega)$. Schluchter and Elashoff (1990) and Manor and Zucker (2004) have examined the chi-square approximation for small sample data using both REML and ML estimation. The chi-square approximations had inflated Type I error rates.

1.4.4 Misspecification of the Covariance

In practice, information regarding the covariance structure is often unknown. As a result, researchers must assume a covariance structure for the data. Assumptions about the distribution of the random effects and errors are also made when fitting a linear mixed model. The linear mixed model assumptions as seen in Section 1.4.1 are that the errors and the random effects are independent and normally distributed. Much research has been done examining what happens when these assumptions about the random effects and errors are incorrect, and the covariance is misspecified. Thus, there are two types of covariance misspecification that can occur. First, it is possible that the structure has been misspecified. The other misspecification occurs is if either the distribution of the error term or the distribution of the random effects does not meet the linear mixed model assumptions (i.e. neither is normally distributed.).

Many authors suggest that specifying the covariance structure can lead to more accurate fixed effects inference. One potential disadvantage would be if the covariance structure is misspecified. Ferron et al. (2002) examined the sensitivity of various fit criteria to misspecifications of the covariance structure, and then examined the bias of the fixed effects and random effects parameters when there was misspecification. They simulated data from a first order autoregressive error structure and determined that Akaike's Information Criteria (AIC) identified the correct structure 70% of the time and the Schwartz's Bayesian Criterion identified the correct structure 45% of the time. Additionally, when the error structure was misspecified as a simplier structure, there was not any bias in the fixed effects or the tests of the fixed effects, but there was bias in the estimates of the random effects. Gomez et al. (2005) examined the Kenward-Roger F statistic Type I error rates for tests of fixed effects when simulating data based on 15 different covariance structures when the covariance structure is selected using AIC and BIC. The authors concluded that the Type I error rates of the KR F statistic when covariance is unknown were greater than 0.05 for all of their simulation studies.

Ferron, Dailey and Yi (2002) examined the effect of underspecification of the error structure on the fixed effects estimates and their standard errors under REML estimation. In their simulation study, the true error structure was autoregressive of order 1 and the misspecified error structure is independent and identically distributed. Underspecification of the error structure lead to unbiased estimators of the fixed effects but the variance parameters were biased. Kwok, West and Green (2007) found similar results with respect to underspecification of the error structure. These authors expanded the results by also looking at general misspecification and overspecification. They conclude that general misspecification of the error structure lead to an overestimation of the variances of the random effects which implies overestimation of the standard errors of the fixed effect. Alternatively, overspecification of the error structure lead to smaller random effect variances which implies standard errors of fixed effects were smaller.

Verbeke and Lesaffre (1997) investigated the effect of misspecifying the distribution of the random effects on the inference based on the ML estimates of the model. The authors concluded that misspecifying the distribution of the random effects does not affect the ML estimates of the fixed effects. Actually, the fixed effects estimates are consistent and asymptotically normally distributed regardless of the distribution of the random effects, but the misspecification of the distribution does have an effect on the random components.

Similarly, Fellingham and Raghunathan (1995) found that when the distribution of the random effects was symmetric, the REML estimates of the fixed effects were not affected.

Conversely, REML estimation is poor when the random effects distribution is not symmetric. Manor and Zucker (2004) found similar results when they simulated data where the random terms had a t-distribution (symmetric) and a log-normal distribution. The Type I error rates were closer to the nominal level when the random effects had a t-distribution as compared to the log-normally distributed random effects. When the random effects were simulated from the log-normal distribution, the Type I error rates were larger than when random effects were simulated from the t-distribution. Vallejo, Ato, and Valdes (2008) also researched the consequences of covariance misspecification by examining Type I error rates for tests of fixed effects when choosing a model based on various information criteria. The Type I error rates of the models chosen from the AIC criterion that were generated from symmetric distributions were robust.

1.5 R^2 Statistics for the Linear Mixed Model

1.5.1 Criteria for Assessing R^2 statistics for the Linear Mixed Model

As seen in Section 1.2, the R^2 statistics for the linear univariate model using least squares regression are equivalent. Unfortunately, the various R^2 statistics for other models are different and do not coincide (Kvalseth, 1985). Each of the formulae given in Section 2 are ways in which the R^2 statistic can be adapted in the linear mixed model. Those formulae are not equivalent in the case of the linear mixed model. When applying those formulae very different R^2 statistics arise. At this time, there is not a universally accepted R^2 statistic for fixed or random effects in the linear mixed model because opinions differ as to what R^2 should measure in this setting. This poses a problem when determining from which R^2 statistic for the linear mixed model to choose.

Kramer (2005) further describes the impossibility of defining a single R^2 for fixed effects for mixed models due to the complexity of the model and the variety of questions that could be posed from the model. Researchers may be interested in hypothesis tests of the fixed effects, or of the random effects. Edwards et al. (2008) highlight the three types of model comparisons that occur in the linear mixed model. It is possible to compare models with the same covariance structure but different fixed effects, models with the same fixed effects but different covariance structures, and models with different fixed effects and different covariance structures. This distinction is important because in the linear mixed model, variation can be explained due to the fixed effects or due to the covariance specification.

Another issue when defining an R^2 statistic for fixed effects for the linear mixed model is the choice of the null model. Throughout the literature of R^2 statistics in the linear mixed model, there has mainly been two null models that have been discussed. There is the null model with only an intercept in the fixed effects and the null model with both a fixed and random intercept. The null model for R^2_β, an R^2 statistic for fixed effects defined by Edwards et al. (2008), can be different from those two null models. It has a fixed effect in the intercept and the same covariance structure as the model of interest. The interpretation of the R^2 statistic depends on the choice of the null model (Edwards et al., 2008).

1.5.2 Evaluating the R^2 Statistics for the Linear Mixed Model

Table 1.2 is a summary of the R^2 statistics used in the linear mixed model. A comprehensive review and details of notation is provided in Edwards et al. (2008). Since many of the R^2 statistics for the linear mixed model are new, there is an increased interest in evaluating each of them and how they behave. The evaluation methods used in the recent work focus on the properties that are important for an ideal R^2 statistic.

One property that is important for R^2 statistics is that the statistic increases when important covariates are added to the model. Conversely, it is not desirable for the R^2 statistic to increase when fitting an overfitted model. Orelien and Edwards (2008) evaluated the marginal and conditional versions of r_c, R_1^2, and P_{rand} as well as the R^2 statistics proposed by Xu (2003) ($\widehat{\Omega}^2$, R_2^2 and $\widehat{\rho}^2$) in distinguishing between overfitted, true, and underfitted models. The marginal version of these statistics refers to when the predicted values are computed using only the fixed effects $\left(\widehat{\boldsymbol{y}}_i = \boldsymbol{X}_i\widehat{\boldsymbol{\beta}}\right)$, and the conditional version refers to when the predicted values are computed using the random effects$\left(\widehat{\boldsymbol{y}}_i = \boldsymbol{X}_i\widehat{\boldsymbol{\beta}} + \boldsymbol{Z}_i\widehat{\boldsymbol{d}}_i\right)$. The conditional versions of r_c, R_1^2, and P_{rand} as well as $\widehat{\Omega}^2$, R_2^2 and $\widehat{\rho}^2$ performed poorly. Those statistics could not distinguish when important covariates were missing whereas, the marginal versions of r_c, R_1^2, and P_{rand} could. The authors conclude that the marginal R^2 statistics presented are able to determine the most parmonious model among overfitted, true, and underfitted models.

Another important property of a R^2 statistic is that they are monotone. Liu et al. (2008) show that two of the three R^2 statistics that they proposed are monotone through examining the dimension of the projected subspace of their statistics.

Sun et al. (2010) also emphasize the importance of the monotonic nondecreasing property in an R^2 statistic. The authors primary reason for choosing Magee's (1990) likelihood ratio based statistic was due to that property. They evaluated the performance of R_{LR}^2, P_{rand}, r_c and conditional and marginal versions of R_{W}^2. They were intested in finding which statistic best captured the quantitative locus trait effect in association mapping. Another important property for Sun et al. (2010) was that the R^2 statistic reduces to the usual R^2 statistic for the fixed linear model. The authors point out that r_c does not have this property.

Another evaluation method used is based on the ability of an R^2 statistic to select a model. Wang and Schaalje (2009) conducted a simulation study on 17 model selection statistics to determine the success rate of choosing a fixed effect when the covariance structure was known. The R^2 statistics that the authors looked at were adjusted versions of marginal and conditional R_1^2 and also adjusted versions of the marginal and conditional r_c. All of the statistics were successful in selecting the best linear model when there was a compound symmetric covariance structure. There was not one criterion that was consistently better than the others. Performance of these model selection criteria was found to be based on covariance structure, values of parameters, and sample size.

Kramer (2005) focused on examining R^2 statistics as a goodness of fit measurement. R^2_{LR} and R^2_W statistics were evaluated and the author found that as the model complexity increased, the R^2 statistics increased.

1.5.3 R^2_β for fixed effects in the Linear Mixed Model

Edwards et al. (2008) expanded on formula (1.2) and proposed an R^2 statistic for the fixed effects in the linear mixed model. The newly proposed R^2_β is as follows,

$$R^2_\beta = \frac{(q-1)\nu^{-1} F\left(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Sigma}}\right)}{1 + (q-1)\nu^{-1} F\left(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Sigma}}\right)}.$$

Edwards et al. (2008) showed that under certain conditions where the linear mixed model coincided with the multivariate linear model, the R^2_β is identical to the Hotelling-Lawley trace association statistic. Another correspondence between a statistic for the linear mixed model and a statistic for the multivariate linear model is shown in Bathke et al. (2009). These authors showed, under special cases, the equivalence of an ANOVA type statistic from the linear mixed model and the Greenhouse-Geisser F adjustment.

The R^2_β statistic Edwards et al. (2008) propose is based on the restricted maximum likelihood (REML) estimation while noting that the formulae do apply to the maximum likelihood estimation computations. Additionally, the authors recommend using the Kenward-Roger F to define R^2_β because small sample inference of the Kenward-Roger F is the most accurate.

One advantage of R^2_β is that to compute it, only one model needs to be fit. It is not necessary to fit a null model. Another advantage of the R^2_β proposed is that it corresponds to a signficance test because it is a 1-1 function of the F test for the linear mixed model using Kenward-Roger's F statistic and REML estimation. The R^2 statistic based on the likelihood ratio test introduced by Magee (1990) also has this property.

Another advantage of the R^2_β proposed is that the same statistic generalizes to define a partial R^2 statistic for marginal (fixed) effects of all sorts. None of the other R^2 statistics reviewed appear to have the same important property.

1.6 Summary and Overview

In the linear univariate model, the R^2 statistic has been extensively researched and is a widely implemented analysis tool. It serves as a goodness of fit tool, a model selection tool, and a measure of the strength of association. There are several formulae for the R^2 statistic and they are all equivalent under the context of linear regression models.

Each expression for the R^2 statistic in the univariate model gives rise to an R^2 statistic in the linear mixed model. Unfortunately, these statistics are not equivalent and there is not one universally accepted R^2 statistic for the linear mixed model.

This dissertation will present some of the properties of R^2_β through both theoretical and applied analysis. In Chapter 2, theoretical results are presented which describe the asymptotic properties of R^2_β defined by the Kenward-Roger method using approximations to the mean and variance of R^2_β. In addition, two tests of hypothesis are developed and evaluated using a large scale simulation study. Chapter 3 examines the impact of covariance structure misspecification, denominator degrees of freedom methods, estimation techniques on the values and asymptotic properties of R^2_β. The finite sample properties of R^2_β are discussed in Chapter 4 which include examining the semi-partial form of R^2_β for different denominator degrees of freedom methods and creating an adjusted version of R^2_β for fixed effects model selection.

Table 1.1 Test Statistics for General Linear Hypothesis
and the Corresponding Multivariate Measures of Association

Hotelling-Lawley Trace	**Pillai-Bartlett Trace**	**Wilks' Lambda**	**Roy's Largest Root**
tr$(\boldsymbol{S}_h\boldsymbol{S}_e^{-1})$	tr$[\boldsymbol{S}_h(\boldsymbol{S}_h+\boldsymbol{S}_e)^{-1}]$	tr$[\boldsymbol{S}_e(\boldsymbol{S}_h+\boldsymbol{S}_e)^{-1}]$	max eigenvalue $\boldsymbol{S}_h(\boldsymbol{S}_h+\boldsymbol{S}_e)^{-1}$
$\eta=\frac{\text{HLT}/s}{1+\text{HLT}/s}$	$\eta=\frac{\text{PBT}}{s}$	$\eta=1-\text{WLK}^{1/g}$	$\frac{\text{tr}(\boldsymbol{S}_h)}{\text{tr}(\boldsymbol{S}_h+\boldsymbol{S}_e)}$

Table 1.2. Summary of R^2 Statistics in the Linear Mixed Model

Source	**Formula**
Vonesh et al. (1996)	$r_c=1-\frac{\sum_{i=1}^{m}(\boldsymbol{y}_i-\widehat{\boldsymbol{y}}_i)'(\boldsymbol{y}_i-\widehat{\boldsymbol{y}}_i)}{\sum_{i=1}^{m}(\boldsymbol{y}_i-\overline{y}\boldsymbol{1}_{n_i})'(\boldsymbol{y}_i-\overline{y}\boldsymbol{1}_{n_i})+\sum_{i=1}^{m}(\widehat{\boldsymbol{y}}_i-\widehat{y}\boldsymbol{1}_{n_i})'(\widehat{\boldsymbol{y}}_i-\widehat{y}\boldsymbol{1}_{n_i})+N(\overline{y}-\widehat{y})^2}$
Vonesh and Chinchilli (1997)	$R_1^2=1-\frac{\sum_{i=1}^{m}(\boldsymbol{y}_i-\widehat{\boldsymbol{y}}_i)'(\boldsymbol{y}_i-\widehat{\boldsymbol{y}}_i)}{\sum_{i=1}^{m}(\boldsymbol{y}_i-\overline{y}\boldsymbol{1}_{n_i})'(\boldsymbol{y}_i-\overline{y}\boldsymbol{1}_{n_i})}$
Zheng (2000)	$P_{\text{rand}}=1-\frac{\sum_{i=1}^{m}d_i(\boldsymbol{y}_i,\widehat{\boldsymbol{y}}_i)/2\widehat{\sigma}+\widehat{\boldsymbol{b}}'(\boldsymbol{G}\otimes\boldsymbol{I_m})\widehat{\boldsymbol{b}}/2}{\sum_{i=1}^{m}d_i(\boldsymbol{y}_i,\overline{y}\boldsymbol{1}_{n_i})/(2\widehat{\sigma})}$
Xu (2003)	1. $\widehat{\Omega}^2=1-\frac{\widehat{\sigma}^2}{\widehat{\sigma}_0^2}$ 2. $R_2^2=1-\frac{RSS}{RSS_0}$ 3. $\widehat{\rho}^2=1-\frac{\widehat{\sigma}^2}{\widehat{\sigma}_0^2}\exp\left(\frac{RSS}{N\widehat{\sigma}^2}-\frac{RSS_0}{N\widehat{\sigma}_0^2}\right)$
Magee (1990)	$R_{\text{LR}}^2=1-\exp\left[-\frac{2}{n}(\log L_M-\log L_0)\right]$
Buse (1973)	$R_W^2=1-\frac{\sum_{i=1}^{m}(\boldsymbol{y}_i-\widehat{\boldsymbol{y}}_i)'\boldsymbol{V}_i^{-1}(\boldsymbol{y}_i-\widehat{\boldsymbol{y}}_i)}{\sum_{i=1}^{m}(\boldsymbol{y}_i-\overline{y}\boldsymbol{1}_{n_i})'\boldsymbol{V}_i^{-1}(\boldsymbol{y}_i-\overline{y}\boldsymbol{1}_{n_i})}$

REFERENCES

Alnosaier, W.S. (2007). *Kenward-Roger approximate F test for fixed effects in mixed linear models*. Unpublished doctoral dissertation, Oregon State University, Portland, OR.

Anderson, T.W. (2003). *An introduction to multivariate statistical analysis*. (3rd ed.). Hoboken, NJ: Wiley.

Anderson-Sprecher, R. (1994). Model comparisons and R^2. *The American Statistician* **48**, 113-117.

Arnau, J., Bono, R. and Vallejo, G. (2009). Analyzing small samples of repeated measures data with mixed model adjusted F test. *Communications in Statistics- Simulation and Computation* **38**, 1083-1103.

Barten, A.P. (1962). Note on unbiased estimation of the squared multiple correlation coefficient. *Statistica Neerlandica* **16**, 151-163.

Bathke, A.C., Schabenberger, O., Tobias, R.D., and Madden, L.V. (2009). Greenhouse-Geisser adjustment and the ANOVA-type Statistic: Cousins or twins? *The American Statistician* **613**, 239-246.

Buse, A. (1973). Goodness of fit in generalized least-squares estimation. *American Statistician,* 27, 106–108.

Cameron, A.C. and Windmeijer, F.A.G. (1996). R-squared measures for count data regression models with applications to health-care utilization. *Journal of Business and Economic Statistics* **14**, 209-220.

Draper, N.R., and Smith, H. (1998). *Applied Regression Analysis* (3rd ed.). New York, NY: Wiley.

Edwards, L.J., Muller, K.E., Wolfinger, R.D., Qaqish, B.F., and Schabenberger, O. (2008). An R^2 statistic for fixed effects in the linear mixed model. *Statistics in Medicine* **27**, 6137-6157.

Fellingham, G.W. and T.E. Raghunathan. (1995). Sensitivity of point and interval estimates to distributional assumptions in longitudinal data analysis of small samples. *Communications in Statistics- Simulation and Computation* **24**, 617-630.

Ferron, J., Dailey, R., and Yi, Q. (2002). Effects of misspecifying the first-level error structure in two-level models of change. *Multivariate Behavioral Research* **37**, 379-403.

Ferron, J., Hogarty, K. Y., Hess, M., Romano, J., Kromrey, J. D., Niles, J. D., Dawkins, G.R., and Sentovich, C. (2002). A macro for examining the consequences of error structure misspecifications. *Proceedings of the Twenty-Seventh Annual SAS Users Group International Conference*, Cary, NC: SAS Institute Inc. 201-227.

Glueck, D.H. (1996). *Power for a generalization of the GLMM with fixed and random predictors*. Unpublished doctoral dissertation, University of North Carolina at Chapel Hill, Chapel Hill, NC.

Glueck, D.H. and Muller, K.E. (2003). Adjusting power for a baseline covariate in linear models. *Statistics in Medicine* **22**, 2535-2551.

Gómez, V.E., Schaalje, G.B., & Fellingham, G.W. (2005). Performance of the Kenward-Roger method when the covariance structure is selected using AIC and BIC. *Communications in Statistics- Simulation and Computation* **34**, 377–392.

Grizzle, J.E., Allen, D.M. (1969). Analysis of growth and dose response curves. *Biometrics* **25**, 357–381.

Harville, D.A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. *Journal of the American Statistical Association* **72**, 320-340.

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

Helms, R.W. (1992). Intentionally incomplete longitudinal designs: I. Methodology and comparison of some full span designs. *Statistics in Medicine* **11**, 1889-1913.

Jennrich, R.I. and Schluchter, M.D. (1986). Unbalanced repeated-measures models with structured covariance matrices. *Biometrics* **42**, 805-820.

Johnson, R.A. and Wichern, D.W. (2002). *Applied Multivariate Statistical Analysis* (5th ed.). Hoboken, NJ: Prentice Hall

Kackar, R.N. and Harville, D.A. (1984). Approximations for standard errors of estimations of fixed and random effects in mixed linear models. *Journal of the American Statistical Association* **79**, 853-862.

Kim, S. and Olejnik, S. (2005). Bias and precision of measures of association for a fixed-effect multivariate analysis of variance model. *Multivariate Behavioral Research* **40**, 401-421.

Kenward, M.G. and Roger, J.H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics* **53**, 983-997.

Keselman, H.J., Algina, J., Kowalchuk, R.K., and Wolfinger, R.D. (1999a). A comparison of recent approaches to the analysis of repeated measurements. *British Journal of Mathematical and Statistical Psychology* **52**, 63-78.

Keselman, H.J., Algina, J., Kowalchuk, R.K., and Wolfinger, R.D. (1999b). The analysis of repeated measurements: A comparison of mixed model Satterthwaite F tests and a nonpooled adjusted degrees of freedom multivariate test. *Communications in Statistics- Theory and Methods* **28**, 2967-2999.

Kramer, M. (2005). R^2 statistics for mixed models. 2005 *Proceedings of the Conference on Applied Statistics in Agriculture*, Manhattan, KS, pp 148–160.

Kvalseth, T.O. (1985). Cautionary note about R^2. *The American Statistician* **39**, 279-285.

Kwok, O., West, S.G. and Green, S.B. (2007). The impact of misspecifying the within-subject covariance structure in multiwave longitudinal multilevel models: A monte carlo study. *Multivariate Behavioral Research* **42**, 557-592.

Laird, N.M. and Ware, J.H. (1982). Random-effects models for longitudinal data. *Biometrics* **38**, 963-974.

Liu, H., Zheng, Y, and Shen, J. (2008). Goodness-of-fit measures of R^2 for repeated measures mixed effect models. *Journal of Applied Statistics* **35**, 1081-1092.

Littell, R.C. (2002). Analysis of unbalanced mixed model data: A case study comparison of ANOVA vs. REML/GLS. *Journal of Agricultural, Biological and Environmental Statistics* **7**, 472-490.

Magee, L. (1990). R^2 measures based on Wald and likelihood ratio joint significance tests. *The American Statistician* **44**, 250-253.

Manor, O. and Zucker, D.M. (2004). Small sample inference for the fixed effects in the mixed linear model. *Computational Statistics and Data Analysis* **46**, 801-817.

Montgomery, D.B. and Morrison, D.G. (1973). A note on adjusting R^2. *The Journal of Finance* **28**, 1009-1013.

Muller, K.E., Edwards, L.J., Simpson, S.L. and Taylor, D.J. (2007). Statistical tests with accurate size and power for balanced linear mixed models. *Statistics in Medicine* **26**, 3639-3660.

Muller, K.E., and Fetterman, B.A. (2002). *Regression and ANOVA: An integrated approach using SAS software*. Cary, NC: SAS Institute.

Muller, K.E., and Stewart, P.W. (2006). *Linear model theory: Univariate, multivariate, and mixed models.* Hoboken, NJ: Wiley.

Orelien, J.G., and Edwards, L.J. (2008). Fixed-effect variable selection in linear mixed models using R^2 statistics. *Computational Statistics and Data Analysis* 52, 1896–1907.

SAS Institute. (2004). *SAS/STAT User's Guide, Version 9.1*. Cary, NC: SAS Institute.

Satterthwaite, F.E. (1946). An approximate distribution of estimates of variance components. *Biometrika Bulletin* **2**, 110-114.

Schaalje, G.B., McBride, J.B. and Fellingham, G.W. (2002). Adequacy of approximations to distributions of test Statistics in complex mixed linear models. *Journal of Agricultural, Biological, and Environmental Statistics* **7**, 512-524.

Schluchter, M.D. and Elashoff, J.D. (1990). Small-sample adjustments to tests with unbalanced repeated measures assuming several covariance structures. *Journal of Statistical Computation and Simulation* **34**, 11-28.

Sen, P.K. and Singer, J.M. (1993). *Large sample methods in statistics: An introduction with applications*. New York, NY: Chapman & Hall.

Steyn Jr., H.S. and Ellis, S.M. (2009). Estimating the effect size in one-way multivariate analysis of variance (MANOVA). *Multivariate Behavioral Research* **44**, 106-129.

Sun, G., Zhu, C., Kramer, M.H., Yang, S.S., Song, W., Piepho, H.P., and Yu, J. (2010). Variation explained in mixed-model association mapping. *Heredity* 1-8.

Vallejo, G., Ato, M. and Valdes, T. (2008). Consequences of misspecifying the error covariance structure in linear mixed models for longitudinal data. *Methodology* **4**, 10-21.

Verbeke, G. and Lesaffre, E. (1997). The effect of misspecifying the random-effects distribution in linear mixed models for longitudinal data. *Computational Statistics and Data Analysis* **23**, 541-556.

Verbeke, G. and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data* New York, NY: Springer-Verlag.

Vonesh, E.F., and Chinchilli, V.M. (1997) *Linear and Nonlinear Models for the Analysis of Repeated Measures* New York: Marcel Dekker.

Vonesh, E.F., Chinchilli, V.M. and Pu, K. (1996) Goodness-of-fit in generalized nonlinear mixed-effects models. *Biometrics*, **52**, 572-587.

Wang, J. and Schaalje, G.B. (2009). Model selection for linear mixed models using predictive criteria. *Communications in Statistics- Simulation and Computation* **38**, 788-801.

Welham, S.J. and Thompson, R. (1997). A likelihood ratio test for fixed model terms using residual maximum likelihood. *Journal of the Royal Statistical Society, Series B* **59**, 701-714.

Xu, R.H. (2003). Measuring explained variation in linear mixed effects models. *Statistics in Medicine*, **22**, 3527-3541.

Zheng, B.Y. (2000). Summarizing the goodness of fit of generalized linear models for longitudinal data. *Statistics in Medicine,* **19**, 1265–1275.

CHAPTER 2

ASYMPTOTIC PROPERTIES OF R^2_β AND TESTS OF HYPOTHESES

2.1 Introduction

Two approaches are considered in deriving approximations for the mean and variance of $R^2_\beta(\nu)$ under the null and alternative hypotheses which include using the Beta distribution and a Taylor series approximation. The former assumes the Wald F statistic has an F distribution and the latter assumes values for only the mean and variance of the F statistic. Test statistics are developed based on these approaches.

2.2 Distributions and Their Properties

2.2.1 Gamma distribution

If a random variable, X, has a Gamma distribution with parameters $\alpha,\ \beta > 0,$ the probability density function is,

$$f(x;\alpha,\ \beta) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}}, \qquad x > 0.$$

The chi-squared distribution is the special case of the Gamma distribution where $\alpha = \frac{\nu}{2}$ and $\beta = 2$ and will be denoted a χ^2_ν.

2.2.2 Central F distribution

The ratio of two independent chi-squared variables over their respective degrees of freedom, ν_1 and ν_2 degrees of freedom, results in an F distribution with ν_1 and ν_2 degrees of freedom. The probability density function of the F distribution is

$$f(x;\, \nu_1,\, \nu_2) = \frac{1}{B(\frac{\nu_1}{2},\, \frac{\nu_2}{2})} \left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}} x^{\frac{\nu_1}{2}-1} \left(1 - \frac{\nu_1}{\nu_2}x\right)^{-\left(\frac{\nu_1+\nu_2}{2}\right)}, \qquad x > 0,$$

where ν_1 and ν_2 are positive integers and $B(a,\, b)$ is the Beta function. Some well known facts concerning the F distribution are as follows:

1. If $X \sim F(\nu_1,\, \nu_2)$ then as $\nu_2 \to \infty$, $\nu_1 X$ converges in distribution to $\chi^2_{\nu_1}$.
2. If $X \sim F(\nu_1,\, \nu_2)$ then $\frac{1}{X} \sim F(\nu_2,\, \nu_1)$.
3. If $X \sim F(\nu_1,\, \nu_2)$ and $Y = \frac{\nu_1 X/\nu_2}{1+\nu_1 X/\nu_2}$ then $Y \sim \text{Beta}(\frac{\nu_1}{2},\, \frac{\nu_2}{2})$ (The Beta distribution will be defined in section 2.2.4.) (Draper and Smith, 1998).

2.2.3 Non-Central F distribution

The distribution of the ratio of a non-central chi-squared and a central chi-squared that are independent is a non-central F distribution. If $X_1 \sim \chi^2(\nu_1,\, \lambda)$ and $X_2 \sim \chi^2(\nu_2)$ with X_1 independent of X_2 then,

$$\frac{X_1/\nu_1}{X_2/\nu_2} \sim F(\nu_1,\, \nu_2,\, \lambda),$$

where $F(\nu_1,\, \nu_2,\, \lambda)$ denotes the non-central F distribution. The mean of the non-central F distribution when $\nu_2 > 2$ is

$$E[F] = \frac{\nu_2(\nu_1 + \lambda)}{\nu_1(\nu_2 - 2)}.$$

If $\nu_2 \leq 2$, then the mean does not exist. Additionally,

1. When $\lambda = 0$, the non-central F distribution becomes the F distribution.

2. If $X \sim F(\nu_1,\, \nu_2,\, \lambda)$ and λ is not dependent upon ν_2 then as $\nu_2 \to \infty$, $\nu_1 X$ converges in distribution to $\chi^2_{\nu_1}$.

3. If $X \sim F(\nu_1,\, \nu_2,\, \lambda)$ then $Y = \frac{\nu_1 \nu_2^{-1} X}{1+\nu_1 \nu_2^{-1} X} \sim$ Beta$(\frac{\nu_1}{2},\, \frac{\nu_2}{2},\, \lambda)$ which will be defined in Section 2.2.5.

2.2.4 Central Beta distribution

The probability density function of the Beta distribution is:

$$f(x;\alpha,\,\gamma) = \frac{\Gamma(\alpha+\gamma)}{\Gamma(\alpha)\Gamma(\gamma)} x^{\alpha-1}(1-x)^{\gamma-1}, \qquad 0 < x < 1,$$

where Γ is the gamma function. Beta$(\alpha,\, \gamma)$ denotes the Beta distribution with parameters α and γ. The mean and variance of a random variable X with a Beta distribution with parameters $\alpha > 0$ and $\gamma > 0$ are:

$$\begin{aligned} E[X] &= \frac{\alpha}{\alpha+\gamma}, \\ V[X] &= \frac{\alpha\gamma}{(\alpha+\gamma)^2(\alpha+\gamma+1)}. \end{aligned}$$

There are several important properties of the Beta distribution.

1. If $X \sim$ Beta$(\alpha,\, \gamma)$, then $1 - X \sim$ Beta$(\gamma,\, \alpha)$.

2. If X has a Beta distribution where both of the parameters are equal to 1 $(\alpha = 1,\, \gamma = 1)$, then X has a Uniform distribution on $(0,\, 1)$.

3. If $X \sim \chi^2(\alpha)$ and independently $Y \sim \chi^2(\gamma)$ then $\frac{X}{X+Y} \sim$ Beta$(\frac{\alpha}{2},\, \frac{\gamma}{2})$ (Johnson, Kotz and Balakrishnan, 1994).

4. If $X \sim$ Beta$(\alpha,\, \gamma)$ then as $\gamma \to \infty$, then $X \to {}_d$Gamma $(\alpha,\, 1)$

2.2.5 Non-central Beta distribution

The non-central Beta distribution is defined as the ratio $Z = \frac{X}{X+Y}$ where $X \sim \chi^2(\nu_1,\, \lambda)$ and $Y \sim \chi^2(\nu_2)$. It is clear to see that for $Y \neq 0$, we can write $Z = \frac{X/Y}{X/Y+1}$. If $X \sim \chi^2(\nu_1,\, \lambda)$ and $Y \sim \chi^2(\nu_2)$ and X is independent of Y then X/Y has a non-central F distribution. Approximations of the mean and variance of the non-central Beta have been implemented in computer programs. Chattamvelli and Shanmugam (1997) give an approximate expression for the mean of the non-central Beta distribution derived using the delta method. The approximation of the mean of a non-central Beta distribution is denoted as Beta $(\alpha,\, \gamma,\, \lambda)$ where $\alpha > 0,\, \gamma > 0,$ and $\lambda > 0$ is,

$$E[X] \simeq 1 - \left(\frac{\gamma}{C}\right)\left(1 + \frac{\lambda}{2C^2}\right) + O\big(n^{-1}\big)$$

where $C = \alpha + \gamma + \frac{\lambda}{2}$. The order of $O(n^{-1})$ was shown in Oehlert (1992).

The approximation of the variance of a non-central Beta distribution is:

$$\begin{aligned} V[X] = V\big[R^2_\beta\big] \simeq \frac{\lambda(\frac{\nu}{2})^2}{2C^4} + \frac{\frac{\nu}{2}}{G}\left[1 + \frac{\lambda}{2}\big(\lambda^2 + 3\lambda + S\big)/G^2\right] \\ - \frac{(\frac{\nu}{2})^2}{F}\big(1 + D/F^2\big) + O\big(n^{-1}\big) \end{aligned}$$

where

$$\begin{aligned}
C &= \tfrac{q-1}{2} + \tfrac{\nu}{2} + \tfrac{\lambda}{2} \\
G &= C(C+1) + \tfrac{\lambda}{2} = \big(\tfrac{q-1}{2} + \tfrac{\nu}{2} + \tfrac{\lambda}{2}\big)\big(\tfrac{q-1}{2} + \tfrac{\nu}{2} + \tfrac{\lambda}{2} + 1\big) + \tfrac{\lambda}{2}, \\
S &= \big[2\big(\tfrac{q-1}{2} + \tfrac{\nu}{2}\big) + 1\big]^2, \\
F &= \big(\tfrac{q-1}{2} + \tfrac{\nu}{2}\big)^2\big(\tfrac{q-1}{2} + \tfrac{\nu}{2} + 1\big) + H\tfrac{\lambda}{2} + \big[3\big(\tfrac{q-1}{2} + \tfrac{\nu}{2}\big) + 4\big]\big(\tfrac{\lambda}{2}\big)^2 + \big(\tfrac{\lambda}{2}\big)^3, \\
H &= 3\big(\tfrac{q-1}{2} + \tfrac{\nu}{2}\big)^2 + 5\big(\tfrac{q-1}{2} + \tfrac{\nu}{2}\big) + 2, \\
D &= \tfrac{\lambda}{2}\Big[H^2 + 2P\tfrac{\lambda}{2} + Q\big(\tfrac{\lambda}{2}\big)^2 + R\big(\tfrac{\lambda}{2}\big)^3 + 9\big(\tfrac{\lambda}{2}\big)^4\Big], \\
P &= \big[3\big(\tfrac{q-1}{2} + \tfrac{\nu}{2}\big) + 1\big]\big[9\big(\tfrac{q-1}{2} + \tfrac{\nu}{2}\big) + 17\big] \\
&\qquad + 2\big(\tfrac{q-1}{2} + \tfrac{\nu}{2}\big)\big[3\big(\tfrac{q-1}{2} + \tfrac{\nu}{2}\big) + 2\big]\big[3\big(\tfrac{q-1}{2} + \tfrac{\nu}{2}\big) + 4\big] + 15, \\
Q &= 54\big(\tfrac{q-1}{2} + \tfrac{\nu}{2}\big)^2 + 162\big(\tfrac{q-1}{2} + \tfrac{\nu}{2}\big) + 130,
\end{aligned}$$

$R = 6\big[6\big(\frac{q-1}{2} + \frac{\nu}{2}\big) + 11\big]$

2.3 Approximating the Mean and Variance of R^2_β using the Beta Distribution

In this section, the distribution of R^2_β is derived by assuming the $F\left(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Sigma}}\right)$ statistic has an $F_{q-1,\nu}(\lambda)$ distribution. The mean and variance of $R^2_\beta(\nu)$ will be approximated under the null and alternative hypotheses. To emphasize the distributional properties of $F\left(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Sigma}}\right)$ and R^2_β, the statistics will be denoted as $F\left(\widehat{\lambda}\right)$ and $R^2_\beta(\nu)$ respectively.

2.3.1 Motivation and Justification for using the Beta distribution

Hotelling's T^2 statistic is a test statistic used to compare groups in the multivariate model given in equation 1.3. It has the Hotelling's T^2 distribution with parameters m and p and is denoted as $T^2(m,\, p)$. Mardia, Kent and Bibby (1992) show in Theorem 3.5.2 on page 74 that the Hotelling's T^2 distribution has a scaled F distribution,

$$T^2(p,\, m) = \left(\frac{mp}{m-p+1}\right) F_{p,\, m-p+1} \tag{2.1}$$

The multivariate model can be stated as a special case of the linear mixed model (Muller and Stewart, 2006). When the general multivariate hypothesis is equivalent to the linear mixed model hypothesis, the Hotelling-Lawley trace measure of association, denoted as η_{HLT}, exactly coincides with the R^2_β calculated from the linear mixed model (Edwards et al, 2008). Therefore the F value corresponding to the multivariate HLT statistic is a one-to-one function of Hotelling's T^2 statistic. One implication of the coincidence of R^2_β with the Hotelling-Lawley trace measure of association is that the Wald F test has an exact F distribution. Thus, the coincidence of these statistics provides justification for assuming the Wald F statistic has an F distribution.

2.3.1.1 Dental Dataset Example

A well-known example from Potthoff and Roy (1964) will be used to demonstrate the correspondence between Hotelling's T^2 statistic and the Wald F statistic in the linear mixed model. Potthoff and Roy (1964) describe how growth curve analysis can be stated as a Generalized Multivariate Analysis of Variance (GMANOVA) model. The data is from an orthodontic study of the distance (mm) from the center of the pituitary to the pterygomaxillary fissure denoted as dental distance measured at ages 8, 10, 12, 14 for 16 boys and 11 girls.

The F statistic is computed analyzing the data using the linear mixed model and using a multivariate model. As expected, the F statistic, denominator degrees of freedom, and corresponding R_β^2 are the same for both the linear mixed model and the linear multivariate model. The F value is 3.63 with degrees of freedom $(4,\ 22)$, and R_β^2 is 0.398.

Additionally, the T^2 statistic is calculated for the dental dataset using a function in the IML procedure. The T^2 statistic is 16.5. Using Mardia, Kent, and Bibby (1992), the T^2 statistic can be calculated using the F statistic. The first thing we have to do is determine m and p in equation 2.1. So, p represents the numerator degrees of freedom of the F statistic and, $p = 4$. The denominator degrees of freedom of the F statistic in the linear mixed model is $m - p + 1 = 22$, and $m = 25$. The T^2 statistic can be computed as:

$$T^2 = \left(\frac{25{*}4}{25 - 4 + 1}\right)3.63 = 16.5$$

2.3.2 Under the Null Hypothesis

Under the null hypothesis $H_0 : \boldsymbol{C\beta} = \mathbf{0}$, the $F\left(\widehat{\lambda}\right)$ statistic is approximated by a central F distribution. Section 2.3.1 provides justification for assuming $F\left(\widehat{\lambda}\right)$ has an exact central F distribution based on the special case multivariate. Under the properties of a central F distribution, described in Section 2.2.2, if $F\left(\widehat{\lambda}\right) \sim F_{q-1,\nu}$, where $F_{q-1,\nu}$ denotes the F distribution with $q-1$ numerator and ν denominator degrees of freedom, then $R^2_\beta = \frac{(q-1)\nu^{-1}F\left(\widehat{\lambda}=0\right)}{1+(q-1)\nu^{-1}F\left(\widehat{\lambda}=0\right)} \sim \text{Beta}\left(\frac{q-1}{2},\ \frac{\nu}{2}\right)$ approximately. Using the moments of the central Beta distribution,

$$
\begin{aligned}
E\left[R^2_\beta(\nu)\middle|\, H_0\right] &= \frac{(q-1)/2}{(q-1)/2+\nu/2} \\
&= \frac{(q-1)}{(q-1)+\nu} \\
&= \frac{q-1}{q+\nu-1},
\end{aligned}
$$

$$
\begin{aligned}
V\left[R^2_\beta(\nu)\middle|\, H_0\right] &= \frac{\left(\frac{q-1}{2}\right)\left(\frac{\nu}{2}\right)}{\left(\frac{q-1}{2}+\frac{\nu}{2}\right)^2\left(\frac{q-1}{2}+\frac{\nu}{2}+1\right)} \\
&= \frac{\frac{(q-1)\nu}{4}}{\left(\frac{q-1+\nu}{2}\right)^2\left(\frac{q-1+\nu+2}{2}\right)} \\
&= \frac{2(q-1)\nu}{(q+\nu-1)^2(q+\nu+1)}.
\end{aligned}
$$

2.3.3 Under the Alternative Hypothesis

Under the alternative hypothesis $H_{\text{A}} : \boldsymbol{C\beta} \neq \mathbf{0}$, the $F\left(\widehat{\lambda}\right)$ statistic is approximated by a non-central F distribution where $\widehat{\lambda}$ is the estimate of the noncentrality parameter λ. Using properties of the non-central F distribution, if $F\left(\widehat{\lambda}\right) \sim F_{q-1,\,\nu}(\lambda)$ then, $R^2_\beta = \frac{(q-1)\nu^{-1}F\left(\widehat{\lambda}\right)}{1+(q-1)\nu^{-1}F\left(\widehat{\lambda}\right)} \sim \text{Beta}\left(\frac{q-1}{2},\ \frac{\nu}{2},\ \lambda\right)$ approximately. The non-central Beta distribution is defined as the ratio $Z = \frac{X}{X+Y}$ where $X \sim \chi^2(\nu_1,\ \lambda)$ and $Y \sim \chi^2(\nu_2)$. For $Y \neq 0$, $Z = \frac{X/Y}{X/Y+1}$. If $X \sim \chi^2(\nu_1,\ \lambda)$ and $Y \sim \chi^2(\nu_2)$ and X is independent of Y then X/Y has a non-central F distribution. Using the approximate moments of the non-central Beta distribution provided by Chattamvelli and Shanmugam (1997)

$$
\begin{aligned}
E\left[R^2_\beta(\nu)\middle|\, H_\mathrm{A}\right] \quad &\simeq 1 - \frac{\frac{\nu}{2}}{\left(\frac{q-1}{2}+\frac{\nu}{2}+\frac{\lambda}{2}\right)}\left(1+\frac{\lambda}{2\left(\frac{q-1}{2}+\frac{\nu}{2}+\frac{\lambda}{2}\right)^2}\right)+O\left(m^{-1}\right)\\
&= 1 - \frac{\nu}{(q-1+\nu+\lambda)}\left(1+\frac{2\lambda}{(q-1+\nu+\lambda)^2}\right)
\end{aligned}
$$

$$
\mathrm{V}\left[R^2_\beta(\nu)\middle|\, H_\mathrm{A}\right] \simeq \frac{\lambda\left(\frac{\nu}{2}\right)^2}{2C^4}+\frac{\frac{\nu}{2}}{G}\left[1+\frac{\lambda}{2}\left(\lambda^2+3\lambda+H\right)/G^2\right]-\frac{\left(\frac{\nu}{2}\right)^2}{F}\left(1+D/F^2\right)+O\left(m^{-1}\right)
$$

where

$$
\begin{aligned}
&C = \tfrac{q-1}{2}+\tfrac{\nu}{2}+\tfrac{\lambda}{2}\\
&G = C(C+1)+\tfrac{\lambda}{2} = \left(\tfrac{q-1}{2}+\tfrac{\nu}{2}+\tfrac{\lambda}{2}\right)\left(\tfrac{q-1}{2}+\tfrac{\nu}{2}+\tfrac{\lambda}{2}+1\right)+\tfrac{\lambda}{2},\\
&S = \left[2\left(\tfrac{q-1}{2}+\tfrac{\nu}{2}\right)+1\right]^2,\\
&F = \left(\tfrac{q-1}{2}+\tfrac{\nu}{2}\right)^2\left(\tfrac{q-1}{2}+\tfrac{\nu}{2}+1\right)+H\tfrac{\lambda}{2}+\left[3\left(\tfrac{q-1}{2}+\tfrac{\nu}{2}\right)+4\right]\left(\tfrac{\lambda}{2}\right)^2+\left(\tfrac{\lambda}{2}\right)^3,\\
&H = 3\left(\tfrac{q-1}{2}+\tfrac{\nu}{2}\right)^2+5\left(\tfrac{q-1}{2}+\tfrac{\nu}{2}\right)+2,\\
&D = \tfrac{\lambda}{2}\left[H^2+2P\tfrac{\lambda}{2}+Q\left(\tfrac{\lambda}{2}\right)^2+R\left(\tfrac{\lambda}{2}\right)^3+9\left(\tfrac{\lambda}{2}\right)^4\right],\\
&P = \left[3\left(\tfrac{q-1}{2}+\tfrac{\nu}{2}\right)+1\right]\left[9\left(\tfrac{q-1}{2}+\tfrac{\nu}{2}\right)+17\right]\\
&\qquad +2\left(\tfrac{q-1}{2}+\tfrac{\nu}{2}\right)\left[3\left(\tfrac{q-1}{2}+\tfrac{\nu}{2}\right)+2\right]\left[3\left(\tfrac{q-1}{2}+\tfrac{\nu}{2}\right)+4\right]+15,\\
&Q = 54\left(\tfrac{q-1}{2}+\tfrac{\nu}{2}\right)^2+162\left(\tfrac{q-1}{2}+\tfrac{\nu}{2}\right)+130,\\
&R = 6\left[6\left(\tfrac{q-1}{2}+\tfrac{\nu}{2}\right)+11\right]
\end{aligned}
$$

2.4 Mean and Variance of R^2_β using the Taylor Series Approximation

Since R^2_β is a function of $F\left(\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\Sigma}}\right)$, the mean and variance of R^2_β can be calculated using a Taylor series approximation. A specific distribution for R^2_β does not need to be assumed, only the mean and variance of $F\left(\widehat{\boldsymbol{\beta}},\widehat{\boldsymbol{\Sigma}}\right)$ are assumed. The Taylor series of a function, $g(X)$ that is infinitely differentiable in the neighborhood of μ_X is,

$$Y = g(X) = \sum_{n=0}^{\infty} \frac{g^{(n)}(\mu_X)}{n!}(X - \mu_X)^n$$

where $g^{(n)}(\mu_X)$ denotes the n^{th} derivative of the function g evaluated at μ_X. Casella and Berger (2002) define the Taylor polynomial of order r about μ_X is,

$$T_r(X) = \sum_{i=0}^{r} \frac{g^{(i)}(\mu_X)}{i!}(X - \mu_X)^i.$$

The authors also state Taylors theorem as: if $g^{(r)}(\mu_X) = \frac{d^r}{dx^r} g(X)\Big|_{x=\mu_X}$ exists then $\lim_{X \to \mu_X} \frac{g(X) - T_r(X)}{(X - \mu_X)^r} = 0$.

The first order approximation to the mean is given by

$$E[Y] \simeq g(\mu_X) + o(|X - \mu_X|),$$

and the second order approximation to the mean is

$$E[Y] \simeq g(\mu_X) + \frac{1}{2}\sigma_X^2 g''(\mu_X) + o\big(|X - \mu_X|^2\big).$$

The first order approximation to the variance is given by

$$\begin{aligned} V[Y] &= V[g(X)] \\ &= V[g(\mu_X) + (X - \mu_X)g'(\mu_X)] \\ &= V[g'(\mu_X)X] \\ &= [g'(\mu_X)]^2 V[X] \\ &= [g'(\mu_X)]^2 \sigma_X. \end{aligned}$$

The second order approximation to the variance will not be computed since further moments assumptions of X would have to be made.

In the case of $R_\beta^2 = \frac{(q-1)\nu^{-1}F\left(\widehat{\lambda}\right)}{1+(q-1)\nu^{-1}F\left(\widehat{\lambda}\right)}$. Let $X = (q-1)\nu^{-1}F\left(\widehat{\lambda}\right)$, then $g(X) = \frac{X}{1+X}$.

Taking the derivative of $g(X)$,

$$
\begin{aligned}
g'(X) &= -X(1+X)^{-2} + (1+X)^{-1} \\
&= \frac{-X}{(1+X)^2} + \frac{1}{(1+X)} \\
&= \frac{-X}{(1+X)^2} + \frac{1+X}{(1+X)^2} \\
&= \frac{1}{(1+X)^2}
\end{aligned}
$$

Additionally, $g''(X) = \frac{-2}{(1+X)^3}$.

2.4.1 Under the Null Hypothesis

Under the null hypothesis, when $F\left(\widehat{\lambda}\right)$ has a central $F_{q-1,\,\nu}$ distribution,

$$E\left[F\left(\widehat{\lambda}\right)\right] = \frac{\nu}{\nu-2} \text{ for } \nu > 2$$

$$V\left[F\left(\widehat{\lambda}\right)\right] = \frac{2\nu^2(q-1+\nu-2)}{(q-1)(\nu-2)^2(\nu-4)} \text{ for } \nu > 4$$

Recall, the noncentrality parameter for the $F\left(\widehat{\lambda}\right)$ is zero under the null hypothesis.

2.4.1.1 First Order Approximation

Under the null hypothesis, with $X = (q-1)\nu^{-1}F\left(\widehat{\lambda}\right)$, we know that

$$
\begin{aligned}
\mu_X &= E[X] \\
&= (q-1)\nu^{-1}\frac{\nu}{\nu-2} \\
&= \frac{q-1}{\nu-2}.
\end{aligned}
$$

Therefore, plugging in μ_X to the first order approximation formula, we see that

$$
\begin{aligned}
E[Y] &= E\left[R_\beta^2(\nu)\middle|\, H_0\right] \\
&\simeq \frac{\frac{q-1}{\nu-2}}{1+\frac{q-1}{\nu-2}} \\
&= \frac{(q-1)}{(\nu-2)+(q-1)} \\
&= \frac{q-1}{q+\nu-3}
\end{aligned}
$$

$E\left[R_\beta^2(\nu)\middle|\, H_0\right]$ using the Beta distribution is not equal to $E\left[R_\beta^2(\nu)\middle|\, H_0\right]$using the Taylor series approximation; however, for large ν they are approximately the same.

Additionally,

$$
\begin{aligned}
\sigma_X^2 &= V[X] \\
&= V\left[(q-1)\nu^{-1}F\left(\widehat{\lambda}\right)\right] \\
&= \frac{(q-1)^2}{\nu^2}V\left[F\left(\widehat{\lambda}\right)\right] \\
&= \frac{(q-1)^2}{\nu^2}\,\frac{2\nu^2[(q-1)+\nu-2]}{(q-1)(\nu-2)^2(\nu-4)} \\
&= \frac{2(q-1)(q+\nu-3)}{(\nu-2)^2(\nu-4)},
\end{aligned}
$$

$$
\begin{aligned}
g'(\mu_X) &= \frac{1}{(1+\mu_X)^2} \\
&= \frac{1}{\left(1+\frac{q-1}{\nu-2}\right)^2} \\
&= \frac{(\nu-2)^2}{(\nu-2+q-1)^2} \\
&= \frac{(\nu-2)^2}{(\nu+q-3)^2}.
\end{aligned}
$$

Under the null hypothesis, the variance of $R_\beta^2(\nu)$ can be approximated by

$$
\begin{aligned}
V[Y] \quad &= V\left[R^2_\beta(\nu)\middle|\, H_0\right] \\
&\simeq \left[\frac{(\nu-2)^2}{(\nu+q-3)^2}\right]^2 \frac{2(q-1)(q+\nu-3)}{(\nu-2)^2(\nu-4)} \\
&= \frac{2(q-1)(\nu-2)^2}{(\nu-4)(\nu+q-3)^3}
\end{aligned}
$$

2.4.1.2 *Second Order Approximation*

The second order approximation involves $g''(\mu_X)$. Since $g''(X) = \frac{-2}{(1+X)^3}$,

$$
\begin{aligned}
g''(\mu_X) \quad &= \frac{-2}{(1+\frac{q-1}{\nu-2})^3} \\
&= \frac{-2}{(\frac{\nu-2+q-1}{\nu-2})^3} \\
&= \frac{-2(\nu-2)^3}{(\nu+q-3)^3}.
\end{aligned}
$$

Therefore, using the second order approximation to the mean,

$$
\begin{aligned}
E[Y] \quad &= E\left[R^2_\beta(\nu)\middle|\, H_0\right] \\
&\simeq g(\mu_X) + \frac{1}{2}\sigma^2_X g''(\mu_X) \\
&= \frac{q-1}{q+\nu-3} + \frac{1}{2}\frac{2(q-1)[q+\nu-3]}{(\nu-2)^2(\nu-4)}\frac{(-2)(\nu-2)^3}{(\nu+q-3)^3} \\
&= \frac{q-1}{q+\nu-3} - \frac{2(q-1)}{(\nu-4)}\frac{(\nu-2)}{(\nu+q-3)^2}.
\end{aligned}
$$

2.4.2 ***Under the Alternative Hypothesis***

Under the alternative hypothesis, the noncentrality parameter is defined as

$$
\lambda = (\boldsymbol{C\beta})'\left[\boldsymbol{C}\left(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{C}'\right]^{-1}(\boldsymbol{C\beta}).
$$

where $\widehat{\lambda} = \left(\boldsymbol{C}\widehat{\boldsymbol{\beta}}\right)' \left[\boldsymbol{C}\left(\boldsymbol{X}'\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{C}'\right]^{-1} \left(\boldsymbol{C}\widehat{\boldsymbol{\beta}}\right)$ is an estimator of λ. The noncentrality parameter is the value of the test statistic computed using the population values of $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ and is dependent upon the subject sample size. When $F\left(\widehat{\lambda}\right)$ is distributed as $F_{q-1,\,\nu}(\lambda)$,

$$E\left[F\left(\widehat{\lambda}\right)\right] = \frac{\nu(q-1+\lambda)}{(\nu-2)(q-1)} \text{ for } \nu > 2,$$

$$V\left[F\left(\widehat{\lambda}\right)\right] = 2\frac{(q-1+\lambda)^2 + (q-1+2\lambda)(\nu-2)}{(\nu-2)^2(\nu-4)}\left(\frac{\nu}{q-1}\right)^2 \text{ for } \nu > 4.$$

2.4.2.1 First Order Approximation

Under the alternative hypothesis, with $X = (q-1)\nu^{-1}F(\widehat{\lambda})$,

$$\begin{aligned}
\mu_X &= E[X] \\
&= (q-1)\nu^{-1}\frac{\nu(q-1+\lambda)}{(\nu-2)(q-1)} \\
&= \frac{(q-1+\lambda)}{\nu-2},
\end{aligned}$$

$$\begin{aligned}
\sigma_X^2 &= V[X] \\
&= V\left[(q-1)\nu^{-1}F(\widehat{\lambda})\right] \\
&= \frac{(q-1)^2}{\nu^2}V\left[F(\widehat{\lambda})\right] \\
&= \frac{(q-1)^2}{\nu^2}2\frac{(q-1+\lambda)^2 + (q-1+2\lambda)(\nu-2)}{(\nu-2)^2(\nu-4)}\left(\frac{\nu}{q-1}\right)^2 \\
&= 2\frac{(q-1+\lambda)^2 + (q-1+2\lambda)(\nu-2)}{(\nu-2)^2(\nu-4)},
\end{aligned}$$

$$
\begin{aligned}
g'(\mu_X) &= \frac{1}{(1+\mu_X)^2} \\
&= \frac{1}{\left(1+\frac{q-1+\lambda}{\nu-2}\right)^2} \\
&= \frac{(\nu-2)^2}{(\nu-2+q-1+\lambda)^2} \\
&= \frac{(\nu-2)^2}{(\nu+q+\lambda-3)^2}.
\end{aligned}
$$

Therefore, plugging in μ_X to the first order approximation formula,

$$
\begin{aligned}
E[Y] &= E\left[R^2_\beta(\nu)\middle|\, H_{\mathrm{A}}\right] \\
&\simeq \frac{\frac{(q-1+\lambda)}{\nu-2}}{1+\frac{(q-1+\lambda)}{\nu-2}} \\
&= \frac{(q-1+\lambda)}{(\nu-2)+(q-1+\lambda)} \\
&= \frac{q-1+\lambda}{q+\nu+\lambda-3}.
\end{aligned}
$$

Under the alternative hypothesis, the variance of $R^2_\beta(\nu)$ can be approximated by

$$
\begin{aligned}
V[Y] &= V\left[R^2_\beta(\nu)\middle|\, H_{\mathrm{A}}\right] \\
&\simeq \left[\frac{(\nu-2)^2}{(\nu+q+\lambda-3)^2}\right]^2 \left[2\frac{(q-1+\lambda)^2+(q-1+2\lambda)(\nu-2)}{(\nu-2)^2(\nu-4)}\right] \\
&= \frac{2(\nu-2)^2\left[(q-1+\lambda)^2+(q-1+2\lambda)(\nu-2)\right]}{(\nu+q+\lambda-3)^4(\nu-4)}
\end{aligned}
$$

2.4.2.2 *Second Order Approximation*

To calculate the second order approximation, we will need to compute σ^2_X and $g''(\mu_X)$. Since $g''(X) = \frac{-2}{(1+X)^3}$,

$$
\begin{aligned}
g''(\mu_X) &= \frac{-2}{(1+\frac{(q-1+\lambda)}{\nu-2})^3} \\
&= \frac{-2}{(\frac{\nu-2+q-1+\lambda}{\nu-2})^3} \\
&= \frac{-2(\nu-2)^3}{(\nu+q+\lambda-3)^3}
\end{aligned}
$$

Therefore, using the second order approximation to the mean,

$$
\begin{aligned}
E[Y] &= E\left[R^2_\beta(\nu)\middle|\, H_{\mathrm{A}}\right] \\
&= g(\mu_X) + \frac{1}{2}\sigma^2_X g''(\mu_X) \\
&= \frac{q-1+\lambda}{q+\nu+\lambda-3} + \frac{1}{2}2\frac{\left[(q-1+\lambda)^2+(q-1+2\lambda)(\nu-2)\right]}{(\nu-2)^2(\nu-4)}\frac{(-2)(\nu-2)^3}{(\nu+q+\lambda-3)^3} \\
&= \frac{q-1+\lambda}{q+\nu+\lambda-3} - 2\frac{\left[(q-1+\lambda)^2+(q-1+2\lambda)(\nu-2)\right]}{(\nu-4)}\frac{(\nu-2)}{(\nu+q+\lambda-3)^3} \\
&= \frac{q-1+\lambda}{q+\nu+\lambda-3} - 2\frac{\left[(q-1+\lambda)^2(\nu-2)+(q-1+2\lambda)(\nu-2)^2\right]}{(\nu-4)(\nu+q+\lambda-3)^3}.
\end{aligned}
$$

2.5 Asymptotic Values of the Mean and Variance of R^2_β

To develop the asymptotic values of the mean of R^2_β, the order of convergence for ν and λ is needed. Under the null hypothesis using both the Beta distribution and the Taylor series approximation, ν is the only component of $E\left[R^2_\beta(\nu)\right]$ and $V\left[R^2_\beta(\nu)\right]$ that depends on m (the number of independent sampling units), the asymptotic properties of ν are the only properties of interest.

Under the alternative hypothesis, in the non-central case, the denominator degrees of freedom, ν and the noncentrality parameter, λ, depend on m. To emphasize the dependence of both the noncentrality parameter, λ, and the denominator degrees of freedom, ν, on m, denote ν as $\nu(m)$ and λ as $\lambda(m)$.

For the asymptotic theory, it is assumed that $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\Sigma}}$ are consistent estimators of the population parameters, $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$**,** and $\boldsymbol{\Sigma}$ is correctly specified**.** In addition, it is assumed that

$m >> n_i$ i.e., the number of independent sampling units dominates the number of observations per unit.

2.5.1 Noncentrality Parameter

Section 2.4.2 defines the noncentrality parameter. The term of $\lambda(m)$ that depends on m is $\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}$. All other components do not depend on m. The essence matrix will be used to write that term as a function of m. Helms (1988) defines the essence matrix as the matrix that has only one copy of each unique row of the design matrix. The essence matrix for the fixed effects is denoted as $\boldsymbol{X}_{\text{Ess}}$. Similarly define the essence matrix for the covariance matrix and denote it as $\boldsymbol{\Sigma}_{\text{Ess}}$.

2.5.1.1 Case 1: $\boldsymbol{X}_i$ is the same for all i and $\boldsymbol{\Sigma}_i$ is the same for all i

For designs, where the design matrix is the same for all i, then $\boldsymbol{X}_i = \boldsymbol{X}_{\text{Ess}}$ for all i and $\boldsymbol{\Sigma}_i = \boldsymbol{\Sigma}_{\text{Ess}}$ for all i.

$$\begin{aligned}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X} &= \sum_{i=1}^{m}\boldsymbol{X}_i'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{X}_i \\ &= m\boldsymbol{X}_{\text{Ess}}'\boldsymbol{\Sigma}_{\text{Ess}}^{-1}\boldsymbol{X}_{\text{Ess}}.\end{aligned}$$

The noncentrality parameter as a function of m is,

$$\begin{aligned}\lambda(m) &= (\boldsymbol{C\beta})'\left[\boldsymbol{C}\left(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{C}'\right]^{-1}(\boldsymbol{C\beta}) \\ &= (\boldsymbol{C\beta})'\left[\boldsymbol{C}\left(\sum_{i=1}^{m}\boldsymbol{X}_i'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{X}_i\right)^{-1}\boldsymbol{C}'\right]^{-1}(\boldsymbol{C\beta}) \\ &= (\boldsymbol{C\beta})'\left[\boldsymbol{C}\left(m\boldsymbol{X}_{\text{Ess}}'\boldsymbol{\Sigma}_{\text{Ess}}^{-1}\boldsymbol{X}_{\text{Ess}}\right)^{-1}\boldsymbol{C}'\right]^{-1}(\boldsymbol{C\beta}) \\ &= m(\boldsymbol{C\beta})'\left[\boldsymbol{C}\left(\boldsymbol{X}_{\text{Ess}}'\boldsymbol{\Sigma}_{\text{Ess}}^{-1}\boldsymbol{X}_{\text{Ess}}\right)^{-1}\boldsymbol{C}'\right]^{-1}(\boldsymbol{C\beta}).\end{aligned}$$

Under this special case,

$$\lim_{m\to\infty}\frac{\widehat{\lambda}(m)}{m}=(\boldsymbol{C\beta})'\left[\boldsymbol{C}\left(\boldsymbol{X}_{\text{Ess}}'\boldsymbol{\Sigma}_{\text{Ess}}^{-1}\boldsymbol{X}_{\text{Ess}}\right)^{-1}\boldsymbol{C}'\right]^{-1}(\boldsymbol{C\beta}).$$

2.5.1.2 Case 2: $\boldsymbol{X}_i=\boldsymbol{X}_{ik}$ for $k=1,...,K$ groups of size $n_K=\frac{m}{K}$ and $\boldsymbol{\Sigma}_i$ is the same for all i

Assume there are K groups of unique design matrices of equal size $\mathrm{n}_K=\frac{m}{K}$ with K being an integer. Under this case,

$$\begin{aligned}\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\quad&=\sum_{i=1}^{m}\boldsymbol{X}_i'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{X}_i\\&=\sum_{k=1}^{K}n_K\boldsymbol{X}_{ik}'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{X}_{ik}\\&=n_K\left(\boldsymbol{X}_{i1}'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{X}_{i1}+\boldsymbol{X}_{i2}'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{X}_{i2}+...+\boldsymbol{X}_{iK}'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{X}_{iK}\right)\\&=\frac{m}{K}\left(\boldsymbol{X}_{i1}'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{X}_{i1}+\boldsymbol{X}_{i2}'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{X}_{i2}+...+\boldsymbol{X}_{iK}'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{X}_{iK}\right).\end{aligned}$$

The noncentrality parameter as a function of m is,

$$\begin{aligned}\lambda(m)\quad&=(\boldsymbol{C\beta})'\left[\boldsymbol{C}\left(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{C}'\right]^{-1}(\boldsymbol{C\beta})\\&=(\boldsymbol{C\beta})'\left\{\boldsymbol{C}\left[\frac{m}{K}\left(\boldsymbol{X}_{i1}'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{X}_{i1}+\boldsymbol{X}_{i2}'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{X}_{i2}+...+\boldsymbol{X}_{iK}'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{X}_{iK}\right)\right]^{-1}\boldsymbol{C}'\right\}^{-1}(\boldsymbol{C\beta})\\&=\frac{m}{K}(\boldsymbol{C\beta})'\left[\boldsymbol{C}\left(\boldsymbol{X}_{i1}'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{X}_{i1}+\boldsymbol{X}_{i2}'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{X}_{i2}+...+\boldsymbol{X}_{iK}'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{X}_{iK}\right)^{-1}\boldsymbol{C}'\right]^{-1}(\boldsymbol{C\beta}).\end{aligned}$$

Under this special case,

$$\lim_{m\to\infty}\frac{\widehat{\lambda}(m)}{m}=\frac{1}{K}(\boldsymbol{C\beta})'\left[\boldsymbol{C}\left(\boldsymbol{X}_{i1}'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{X}_{i1}+\boldsymbol{X}_{i2}'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{X}_{i2}+...+\boldsymbol{X}_{iK}'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{X}_{iK}\right)^{-1}\boldsymbol{C}'\right]^{-1}(\boldsymbol{C\beta}).$$

2.5.1.3 Case 3: $\boldsymbol{X}_i$ is the same for all i and $\boldsymbol{\Sigma}_i$=$\boldsymbol{\Sigma}_{ik}$ for $k=1,...,K$ groups of size $n_K=\frac{m}{K}$

Assume there are K groups of unique covariance matrices of equal size $\mathrm{n}_K=\frac{m}{K}$ with K being an integer. Under this case,

$$
\begin{aligned}
\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X} \quad &= \sum_{i=1}^{m} \boldsymbol{X}_i'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{X}_i \\
&= \sum_{k=1}^{K} n_K \boldsymbol{X}_i'\boldsymbol{\Sigma}_{ik}^{-1}\boldsymbol{X}_i \\
&= n_K\big(\boldsymbol{X}_i'\boldsymbol{\Sigma}_{i1}^{-1}\boldsymbol{X}_i + \boldsymbol{X}_i'\boldsymbol{\Sigma}_{i2}^{-1}\boldsymbol{X}_i + ... + \boldsymbol{X}_i'\boldsymbol{\Sigma}_{iK}^{-1}\boldsymbol{X}_i\big) \\
&= \frac{m}{K}\big(\boldsymbol{X}_i'\boldsymbol{\Sigma}_{i1}^{-1}\boldsymbol{X}_i + \boldsymbol{X}_i'\boldsymbol{\Sigma}_{i2}^{-1}\boldsymbol{X}_i + ... + \boldsymbol{X}_i'\boldsymbol{\Sigma}_{iK}^{-1}\boldsymbol{X}\big).
\end{aligned}
$$

The noncentrality parameter as a function of m is,

$$
\begin{aligned}
\lambda(m) \quad &= (\boldsymbol{C\beta})'\Big[\boldsymbol{C}\big(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\big)^{-1}\boldsymbol{C}'\Big]^{-1}(\boldsymbol{C\beta}) \\
&= (\boldsymbol{C\beta})'\Big\{\boldsymbol{C}\Big[\frac{m}{K}\big(\boldsymbol{X}_i'\boldsymbol{\Sigma}_{i1}^{-1}\boldsymbol{X}_i + \boldsymbol{X}_i'\boldsymbol{\Sigma}_{i2}^{-1}\boldsymbol{X}_i + ... + \boldsymbol{X}_i'\boldsymbol{\Sigma}_{iK}^{-1}\boldsymbol{X}\big)\Big]^{-1}\boldsymbol{C}'\Big\}^{-1}(\boldsymbol{C\beta}) \\
&= \frac{m}{K}(\boldsymbol{C\beta})'\Big[\boldsymbol{C}\big(\boldsymbol{X}_i'\boldsymbol{\Sigma}_{i1}^{-1}\boldsymbol{X}_i + \boldsymbol{X}_i'\boldsymbol{\Sigma}_{i2}^{-1}\boldsymbol{X}_i + ... + \boldsymbol{X}_i'\boldsymbol{\Sigma}_{iK}^{-1}\boldsymbol{X}\big)^{-1}\boldsymbol{C}'\Big]^{-1}(\boldsymbol{C\beta}).
\end{aligned}
$$

Under this special case,

$$
\lim_{m\to\infty}\frac{\lambda(m)}{m} = \frac{1}{K}(\boldsymbol{C\beta})'\Big[\boldsymbol{C}\big(\boldsymbol{X}_i'\boldsymbol{\Sigma}_{i1}^{-1}\boldsymbol{X}_i + \boldsymbol{X}_i'\boldsymbol{\Sigma}_{i2}^{-1}\boldsymbol{X}_i + ... + \boldsymbol{X}_i'\boldsymbol{\Sigma}_{iK}^{-1}\boldsymbol{X}\big)^{-1}\boldsymbol{C}'\Big]^{-1}(\boldsymbol{C\beta}).
$$

2.5.1.4 Case 4: $\boldsymbol{X}_i = \boldsymbol{X}_{ik}$ and $\boldsymbol{\Sigma}_i$=$\boldsymbol{\Sigma}_{ik}$ for $k = 1, ..., K$ groups of size $n_K = \frac{m}{K}$

Assume there are K groups of unique covariance matrices of equal size $n_K = \frac{m}{K}$ with K being an integer. Under this case,

$$
\begin{aligned}
\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X} \quad &= \sum_{i=1}^{m} \boldsymbol{X}_i'\boldsymbol{\Sigma}_i^{-1}\boldsymbol{X}_i \\
&= \sum_{k=1}^{K} n_K \boldsymbol{X}_{ik}'\boldsymbol{\Sigma}_{ik}^{-1}\boldsymbol{X}_{ik} \\
&= n_K\big(\boldsymbol{X}_{i1}'\boldsymbol{\Sigma}_{i1}^{-1}\boldsymbol{X}_{i1} + \boldsymbol{X}_{i2}'\boldsymbol{\Sigma}_{i2}^{-1}\boldsymbol{X}_{i2} + ... + \boldsymbol{X}_{iK}'\boldsymbol{\Sigma}_{iK}^{-1}\boldsymbol{X}_{iK}\big) \\
&= \frac{m}{K}\big(\boldsymbol{X}_{i1}'\boldsymbol{\Sigma}_{i1}^{-1}\boldsymbol{X}_{i1} + \boldsymbol{X}_{i2}'\boldsymbol{\Sigma}_{i2}^{-1}\boldsymbol{X}_{i2} + ... + \boldsymbol{X}_{iK}'\boldsymbol{\Sigma}_{iK}^{-1}\boldsymbol{X}_{iK}\big).
\end{aligned}
$$

The noncentrality parameter as a function of m is,

$$\begin{aligned}\lambda(m) &= (\boldsymbol{C\beta})'\left[\boldsymbol{C}\left(\boldsymbol{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}\right)^{-1}\boldsymbol{C}'\right]^{-1}(\boldsymbol{C\beta})\\ &= (\boldsymbol{C\beta})'\left\{\boldsymbol{C}\left[\frac{m}{K}\left(\boldsymbol{X}_{i1}'\boldsymbol{\Sigma}_{i1}^{-1}\boldsymbol{X}_{i1}+\boldsymbol{X}_{i2}'\boldsymbol{\Sigma}_{i2}^{-1}\boldsymbol{X}_{i2}+...+\boldsymbol{X}_{iK}'\boldsymbol{\Sigma}_{iK}^{-1}\boldsymbol{X}_{iK}\right)\right]^{-1}\boldsymbol{C}'\right\}^{-1}(\boldsymbol{C\beta})\\ &= \frac{m}{K}(\boldsymbol{C\beta})'\left[\boldsymbol{C}\left(\boldsymbol{X}_{i1}'\boldsymbol{\Sigma}_{i1}^{-1}\boldsymbol{X}_{i1}+\boldsymbol{X}_{i2}'\boldsymbol{\Sigma}_{i2}^{-1}\boldsymbol{X}_{i2}+...+\boldsymbol{X}_{iK}'\boldsymbol{\Sigma}_{iK}^{-1}\boldsymbol{X}_{iK}\right)^{-1}\boldsymbol{C}'\right]^{-1}(\boldsymbol{C\beta}).\end{aligned}$$

Under this special case,

$$\lim_{m\to\infty}\frac{\widehat{\lambda}(m)}{m}=\frac{1}{K}(\boldsymbol{C\beta})'\left[\boldsymbol{C}\left(\boldsymbol{X}_{i1}'\boldsymbol{\Sigma}_{i1}^{-1}\boldsymbol{X}_{i1}+\boldsymbol{X}_{i2}'\boldsymbol{\Sigma}_{i2}^{-1}\boldsymbol{X}_{i2}+...+\boldsymbol{X}_{iK}'\boldsymbol{\Sigma}_{iK}^{-1}\boldsymbol{X}_{iK}\right)^{-1}\boldsymbol{C}'\right]^{-1}(\boldsymbol{C\beta}).$$

2.5.2 Denominator Degrees of Freedom

The denominator degrees of freedom also depend on m and convergence properties of each denominator degrees of freedom method vary greatly. The asymptotic properties for specific mean models and covariance structures of the Kenward-Roger (1997) method is examined in the simulation study presented in Section 2.6. Convergence properties for specific mean models and covariance structuresof the Satterthwaite, containment, and residual method are examined in Chapter 3.

2.5.3 Asymptotic Properties of $\boldsymbol{R_\beta^2}$ using the Beta Distribution Approach Moments under the Null Hypothesis

Under the null hypothesis $H_0 : \boldsymbol{C\beta} = \boldsymbol{0}$,

$$\lim_{\nu\to\infty}E\left[R_\beta^2(\nu)\middle|\,H_0\right]=\lim_{\nu\to\infty}\frac{q-1}{q+\nu-1}=0,$$

$$\lim_{\nu\to\infty}V\left[R_\beta^2(\nu)\middle|\,H_0\right]=\lim_{\nu\to\infty}\frac{2(q-1)\nu}{(q+\nu-1)^2(q+\nu+1)}=0.$$

Theorem 1: Using the Beta distribution approach moments, $R_\beta^2(\nu)$ is mean square consistent for 0 under the null hypothesis $H_0 : \boldsymbol{C\beta} = \boldsymbol{0}$.

Proof: From Serfling (1980, Section 1.15.2), if $E\left[R_\beta^2(\nu)\middle|\, H_0\right]$ is asymptotically unbiased and $V\left[R_\beta^2(\nu)\middle|\, H_0\right]$ converges to 0, then R_β^2 is mean square consistent for 0.

The asymptotic null results are intuitive because under the null hypothesis, it is assumed that none of the fixed effects are associated with the outcome. $R_\beta^2(\nu)$ values of zero correspond to no multivariate association between the outcome and the predictors.

2.5.4 *Asymptotic Properties of $\boldsymbol{R_\beta^2}$ using Beta Distribution Approach Moments under the Alternative Hypothesis*

Under the alternative hypothesis $H_{\mathrm{A}} : \boldsymbol{C\beta} \neq \mathbf{0}$,

$$\begin{aligned} E\left[R_\beta^2(\nu)\middle|\, H_{\mathrm{A}}\right] &\simeq 1-\frac{\frac{\nu}{2}}{\left(\frac{q-1}{2}+\frac{\nu}{2}+\frac{\lambda}{2}\right)}\left(1+\frac{\lambda}{2\left(\frac{q-1}{2}+\frac{\nu}{2}+\frac{\lambda}{2}\right)^2}\right) \\ &= 1-\frac{\nu}{(q-1+\nu+\lambda)}\left(1+\frac{2\lambda}{(q-1+\nu+\lambda)^2}\right). \end{aligned}$$

Notice that

$$\begin{aligned} \mu_{R_\beta^2} = \lim_{m\to\infty} E\left[R_\beta^2(\nu)\middle|\, H_{\mathrm{A}}\right] &= \lim_{m\to\infty}\left(1-\frac{\nu(m)}{\left(q-1+\nu(m)+\widehat{\lambda}(m)\right)}\left\{1+\frac{2\widehat{\lambda}(m)}{\left[q-1+\nu(m)+\widehat{\lambda}(m)\right]^2}\right\}\right) \\ &= \lim_{m\to\infty}\left(1-\frac{\nu(m)}{\left(q-1+\nu(m)+\widehat{\lambda}(m)\right)}-\frac{2\widehat{\lambda}(m)\nu(m)}{\left[q-1+\nu(m)+\widehat{\lambda}(m)\right]^3}\right) \\ &= 1-\lim_{m\to\infty}\frac{\nu(m)}{\left(q-1+\nu(m)+\widehat{\lambda}(m)\right)}-\lim_{m\to\infty}\frac{2\widehat{\lambda}(m)\nu(m)}{\left[q-1+\nu(m)+\widehat{\lambda}(m)\right]^3} \\ &= 1-\lim_{m\to\infty}\frac{\nu(m)}{\left(q-1+\nu(m)+\widehat{\lambda}(m)\right)} \\ &= 1-\lim_{m\to\infty}\frac{\frac{\nu(m)}{m}}{\left(\frac{q-1}{m}+\frac{\nu(m)}{m}+\frac{\widehat{\lambda}(m)}{m}\right)} \\ &= \lim_{m\to\infty}\frac{\frac{q-1}{m}+\frac{\widehat{\lambda}(m)}{m}}{\left(\frac{q-1}{m}+\frac{\nu(m)}{m}+\frac{\widehat{\lambda}(m)}{m}\right)} \\ &= \frac{\lim_{m\to\infty}\frac{\widehat{\lambda}(m)}{m}}{\lim_{m\to\infty}\frac{\widehat{\lambda}(m)}{m}+\lim_{m\to\infty}\frac{\nu(m)}{m}}. \end{aligned}$$

The orders of each of the components will be calculated to evaluate the asymptotic properties of $V\left[R^2_\beta(\nu)|\, H_\text{A}\right]$ using the Beta distribution. Since both λ and ν are linear functions of m, the notation $O(.)$ will denote the highest growth rate of m for each term,

$C = O(m)$

$G = O(m^2)$

$S = O(m^2)$

$F = O(m^3)$

$H = O(m^2)$

$D = O(m^4)$

$P = O(m^3)$

$Q = O(m^2)$

$R = O(m)$

Therefore, in writing $V\left[R^2_\beta(\nu)|\, H_\text{A}\right]$ using this notation,

$$V\left[R^2_\beta(\nu)|\, H_\text{A}\right] \simeq \frac{O(m^3)}{[O(m)]^4} + \frac{O(m)}{O(m^2)}\left[1 + \frac{O(m)O(m^2)}{[O(m^2)]^2}\right] - \frac{O(m^2)}{O(m^3)}\left[1 + \frac{O(m^4)}{[O(m^3)]^2}\right]$$

resulting in

$$\lim_{m\to\infty} V\left[R^2_\beta(\nu)|\, H_\text{A}\right] = 0.$$

Theorem 2: Using the Beta distribution approach moments, if $\mu_{R^2_\beta}$ exists and is finite, then R^2_β is mean square consistent for $\mu_{R^2_\beta}$ under the alternative hypothesis $H_\text{A} : \boldsymbol{C\beta} \neq \mathbf{0}$.

Proof: The proof of Theorem 2 is the same as Theorem 1.

2.5.5 Asymptotic Properties of $\boldsymbol{R_\beta^2}$ using the Taylor Series Approximation Moments under the Null Hypothesis

Under the null hypothesis $H_0 : \boldsymbol{C\beta} = \mathbf{0}$,

$$\lim_{\nu\to\infty} E\left[R_\beta^2(\nu)\middle|\, H_0\right] = \lim_{\nu\to\infty} \frac{q-1}{q+\nu-3} = 0,$$

$$\lim_{\nu\to\infty} V\left[R_\beta^2(\nu)\middle|\, H_0\right] = \lim_{\nu\to\infty} \frac{2(\nu-2)^2\left[(q-1+\lambda)^2 + (q-1+2\lambda)(\nu-2)\right]}{(\nu+q+\lambda-3)^4(\nu-4)} = 0.$$

Theorem 3: Using the Taylor series approximation moments, R_β^2 is mean square consistent for 0 under the null hypothesis.

Proof: The proof of Theorem 3 is the same as Theorem 1.

2.5.6 Asymptotic Properties of $\boldsymbol{R_\beta^2}$ using the Taylor Series Approximation Moments under the Alternative Hypothesis

Under the alternative hypothesis $H_\mathrm{A} : \boldsymbol{C\beta} \neq \mathbf{0}$,

$$\begin{aligned}
\mu_{R_\beta^2} = \lim_{m\to\infty} E\left[R_\beta^2(\nu)\middle|\, H_\mathrm{A}\right] &= \lim_{m\to\infty} \frac{q-1+\widehat{\lambda}}{q+\nu+\lambda-3} \\
&= \lim_{m\to\infty} \frac{\frac{q-1}{m} + \frac{\widehat{\lambda}}{m}}{\frac{q-3}{m} + \frac{\nu}{m} + \frac{\widehat{\lambda}}{m}} \\
&= \frac{\lim\limits_{m\to\infty} \frac{\widehat{\lambda}}{m}}{\lim\limits_{m\to\infty} \frac{\widehat{\lambda}}{m} + \lim\limits_{m\to\infty} \frac{\nu}{m}}
\end{aligned}$$

$$\lim_{m\to\infty} V\left[R_\beta^2(\nu)\middle|\, H_\mathrm{A}\right] = 0$$

Theorem 4: Using the Taylor series approximation moments, if $\mu_{R_\beta^2}$ exists and is finite, then $R_\beta^2(\nu)$ is mean square consistent for $\mu_{R_\beta^2}$ under the alternative hypothesis $H_\mathrm{A} : \boldsymbol{C\beta} \neq \mathbf{0}$.

Proof: The proof of Theorem 4 is the same as Theorem 1.

2.5.7 Summary of Asymptotic Results

Sections 2.5 and 2.6 both showed that under the null hypothesis using both approaches,

$$\lim_{m\to\infty} E\left[R_\beta^2(\nu)\middle|\, H_0\right] = 0 \text{ and} \lim_{m\to\infty} V\left[R_\beta^2(\nu)\middle|\, H_0\right] = 0.$$

Additionally, under the alternative hypothesis using both approaches,

$$\lim_{m\to\infty} E\left[R_\beta^2(\nu)\middle|\, H_\mathrm{A}\right] = \frac{\lim\limits_{m\to\infty} \frac{\lambda(m)}{m}}{\lim\limits_{m\to\infty} \frac{\lambda(m)}{m} + \lim\limits_{m\to\infty} \frac{\nu(m)}{m}},$$

and

$$\lim_{m\to\infty} V\left[R_\beta^2(\nu)\right] = 0.$$

2.6 Tests of Hypotheses

The goal of this section is to use the two approaches shown in the previous sections to develop a test for the null hypothesis, $H_0 : \rho^2_{\boldsymbol{y}|\boldsymbol{X}} = 0$, that is equivalent to $H_0 : \boldsymbol{C\beta} = \mathbf{0}$. The $\rho^2_{\boldsymbol{y}|\boldsymbol{X}}$ represents a population parameter of the measure of multivariate association between the outcome $\boldsymbol{y}$ and the nonintercept covariates of interest represented by $\boldsymbol{X}$.

The Type I error of a test statistic is defined as the probability of rejecting the null hypothesis given the the null hypothesis is true. The Type I error rates will be estimated for each test statistic developed in the following sections.

The power of a test statistic is $1 - P(\text{Type II error})$. The probability of a Type II error is the probability of failing to reject the null hypothesis when the alternative hypothesis is true. When the alternative is true, our resulting tests statistics must be compared to the alternative hypothesized distributions.

2.6.1 Beta Distribution Test Statistic

Under the null hypothesis $H_0 : \boldsymbol{C\beta} = \mathbf{0}$, $R^2_\beta(\nu) \sim \text{Beta}\left(\frac{q-1}{2},\ \frac{\nu}{2}\right)$. To test, $H_0 : \rho^2_{\boldsymbol{y}|\boldsymbol{X}} = 0$, compare R^2_β to the central Beta distribution with $\frac{q-1}{2}$ and $\frac{\nu}{2}$ degrees of freedom.

The power of the test statistic will be computed by comparing R^2_β to the non-central Beta distribution with $\frac{q-1}{2}$ and $\frac{\nu}{2}$ degrees of freedom and noncentrality parameter λ. Define $F_{\text{Beta}}\left[\cdot \,\middle|\, \frac{q-1}{2},\ \frac{\nu}{2},\ \lambda\right]$ as the cumulative distribution function of the noncentral Beta distribution with $\frac{q-1}{2}$ and $\frac{\nu}{2}$ degrees of freedom and λ as the noncentrality parameter with the corresponding probability density function, $f_{\text{Beta}}\left[\cdot \mid \frac{q-1}{2},\ \frac{\nu}{2},\ \lambda\right]$. The power of the test statistic is

$$P = 1 - F_{\text{Beta}}\left[f_{\text{crit}}(1-\alpha) \middle| \ \frac{q-1}{2},\ \frac{\nu}{2},\ \lambda \right]$$

where $f_{\text{crit}}(1-\alpha)$ is the $100(1-\alpha)$ percentile from the central Beta distribution with $\frac{q-1}{2}$ and $\frac{\nu}{2}$ degrees of freedom.

2.6.2 Taylor Series Approximation Test Statistic

Serfling (1980) states that it is necessary to determine normalizing constants $a_{\nu|H_0}$ and $b_{\nu|H_0}$ such that $\frac{R^2_\beta - a_{\nu|H_0}}{b_{\nu|H_0}}$ converges in distribution to a random variable having a nondegenerate distribution. Define $a_{\nu|H_0} = E\left[R^2_\beta(\nu)\middle|\, H_0\right]$ and $b_{\nu|H_0} = \left\{V\left[R^2_\beta(\nu)\middle|\, H_0\right]\right\}^{1/2}$ as computed from Section 2.4.1. Therefore,

$$a_{\nu|H_0} = \frac{q-1}{q+\nu-3},$$

$$b_{\nu|H_0} = \frac{2(q-1)(\nu-2)^2}{(\nu-4)(\nu+q-3)^3}.$$

The test statistic is

$$T_\nu = \frac{R^2_\beta(\nu) - \frac{q-1}{q+\nu-3}}{\left[\frac{2(q-1)(\nu-2)^2}{(\nu-4)(\nu+q-3)^3}\right]^{\frac{1}{2}}}.$$

It will be compared to the standard normal distribution. Since the Taylor series approximation test statistic is based on the standard normal distribution, the one-sided test of the hypothesis, $H_0 : \boldsymbol{C\beta} = \mathbf{0}$ vs. $H_\mathrm{A} : \boldsymbol{C\beta} \neq \mathbf{0}$ rejects the null hypothesis if $T_\nu > z_{1-\alpha}$, where $z_{1-\alpha}$ is the one-sided critical value of the standard normal distribution with $\alpha = 0.05$.

Define

$$a_{\nu|H_\mathrm{A}} = \frac{q-1+\lambda}{q+\nu+\lambda-3},$$

$$b_{\nu|H_\mathrm{A}} = \frac{2(\nu-2)^2\left[(q-1+\lambda)^2 + (q-1+2\lambda)(\nu-2)\right]}{(\nu+q+\lambda-3)^4(\nu-4)}.$$

Suppose the alternative hypothesis is true then, $P(T_\nu > z_{1-\alpha}|H_\mathrm{A} : \boldsymbol{C\beta} \neq \mathbf{0}) \neq 1-\alpha$. The power is

$$\begin{aligned} P\left(T_\nu > z_{1-\alpha}\middle| H_\mathrm{A} : \boldsymbol{C\beta} \neq \mathbf{0}\right) &= P\left(\frac{R^2_\beta(\nu) - a_{\nu|H_0}}{b^{\frac{1}{2}}_{\nu|H_0}} > z_{1-\alpha}\middle| H_\mathrm{A} : \boldsymbol{C\beta} \neq \mathbf{0}\right) \\ &= P\left(\frac{R^2_\beta(\nu) - a_{\nu|H_\mathrm{A}}}{b^{\frac{1}{2}}_{\nu|H_\mathrm{A}}} > \frac{z_{1-\alpha}\left(b^{\frac{1}{2}}_{\nu|H_0}\right) + a_{\nu|H_0} - a_{\nu|H_\mathrm{A}}}{b^{\frac{1}{2}}_{\nu|H_\mathrm{A}}}\middle| H_\mathrm{A} : \boldsymbol{C\beta} \neq \mathbf{0}\right). \end{aligned}$$

Since $\frac{R^2_\beta(\nu) - a_{\nu|H_\mathrm{A}}}{b^{\frac{1}{2}}_{\nu|H_\mathrm{A}}}$ approximately follows a standard normal distribution when the alternative hypothesis is true, the approximate power for T_ν is calculated as,

$$1 - \Phi\left(\frac{z_{1-\alpha}\left(b^{\frac{1}{2}}_{\nu|H_\mathrm{A}}\right) + a_{\nu|H_0} - a_{\nu|H_\mathrm{A}}}{b^{\frac{1}{2}}_{\nu|H_\mathrm{A}}}\right)$$

where $\Phi(\,\cdot\,)$ represents the standard normal distribution function.

One important observation is that $a_{\nu|H_0}$ and $b_{\nu|H_0}^{\frac{1}{2}}$ are both of the order ν^{-1}. Thus, T_ν needs to be appropriately standardized. In addition, it is likely that the true limiting distribution for this test statistic is not the standard normal distribution because the range of the standard normal distribution is from $-\infty$ to ∞. The range of T_ν is between 0 and ∞. Instead, the chi-squared distribution may be more appropriate. Let us examine this further,

$$\begin{aligned} P\{T_\nu \leq t\} &= P\left\{ \frac{R_\beta^2 - \frac{q-1}{q+\nu-3}}{\left[\frac{2(q-1)(\nu-2)^2}{(\nu-4)(\nu+q-3)^3}\right]^{\frac{1}{2}}} \leq t \right\} \\ &= P\left\{ R_\beta^2 \leq \frac{q-1}{q+\nu-3} + t\left[\frac{2(q-1)(\nu-2)^2}{(\nu-4)(\nu+q-3)^3}\right]^{\frac{1}{2}} \right\} \end{aligned}$$

It is necessary to derive the limit distribution of R_β^2. If we assume R_β^2 has a Beta distribution with shape parameters, $\frac{q-1}{2}$ and $\frac{\nu}{2}$, then we know from section 2.2.4, that the limiting distribution of νR_β^2 is distributed as a chi-squared random variable with parameters, $\alpha = \frac{q-1}{2}$ and $\beta = 2$.

$$P\{T_\nu \leq t\} = P\left\{ \nu R_\beta^2 \leq \frac{\nu(q-1)}{q+\nu-3} + \nu t\left[\frac{2(q-1)(\nu-2)^2}{(\nu-4)(\nu+q-3)^3}\right]^{\frac{1}{2}} \right\}.$$

The νR_β^2 value can be compared to a chi-squared distribution with parameters, $\alpha = \frac{q-1}{2}$ and $\beta = 2$. The performance of this new statistic is an area of future research. The simulation studies presented only use the normal distribution for Type I error rates and power calculations.

2.7 Simulation Study Results

2.7.1 Data Generation

The dental study data example from Potthoff and Roy (1964) used in Section 2.3.1.1 motivates the simulation study. Simulation study regression and covariance parameters are patterned from a linear mixed model analysis of the data.

Simulations and analysis were conducted using SAS version 9.2. The simulated longitudinal data has $m = 25,\ 50,\ 100,\ 200,\ 500,$ 1,000, 2,000, 5,000, $10{,}000$ subjects, and a constant number of observations per subject ($n_i = n = 4$). The repeated measurements were generated using a linear mixed model with four possible mean models and two different covariance structures. Thus, $9 \times 4 \times 2 = 72$ simulation studies are presented each with 10,000 replications.

The mean structure for Model 0 includes only an intercept. The mean structure for Model I includes an intercept and a continuous time effect. Define $\boldsymbol{X}_{\mathrm{I}} = \begin{bmatrix} 1 & 8 \\ 1 & 10 \\ 1 & 12 \\ 1 & 14 \end{bmatrix}$ as the design matrix.

The mean structure for Model II includes an intercept, continuous time effect and a dichotomous effect that is constant per subject. The dichotomous effect is distributed equally among the subjects. Define $\boldsymbol{X}_{\mathrm{II},0} = \begin{bmatrix} 1 & 8 & 0 \\ 1 & 10 & 0 \\ 1 & 12 & 0 \\ 1 & 14 & 0 \end{bmatrix}$ and $\boldsymbol{X}_{\mathrm{II},1} = \begin{bmatrix} 1 & 8 & 1 \\ 1 & 10 & 1 \\ 1 & 12 & 1 \\ 1 & 14 & 1 \end{bmatrix}$ as the two possible design matrices. $\boldsymbol{X}_{\mathrm{II},0}$ refers to the design matrix for all subjects i with a binary effect of 0. $\boldsymbol{X}_{\mathrm{II},1}$ refers to the design matrix for all subjects i with a binary effect of 1.

The mean structure for Model III includes an intercept, continuous time effect, dichotomous effect that is constant per subject, and a time by dichotomous variable interaction. Define $\boldsymbol{X}_{\mathrm{III},0} = \begin{bmatrix} 1 & 8 & 0 & 0 \\ 1 & 10 & 0 & 0 \\ 1 & 12 & 0 & 0 \\ 1 & 14 & 0 & 0 \end{bmatrix}$ and $\boldsymbol{X}_{\mathrm{III},1} = \begin{bmatrix} 1 & 8 & 1 & 8 \\ 1 & 10 & 1 & 10 \\ 1 & 12 & 1 & 12 \\ 1 & 14 & 1 & 14 \end{bmatrix}$ as the two possible design matrices. $\boldsymbol{X}_{\mathrm{III},0}$ refers to the design matrix for all subjects i with a binary effect of 0. $\boldsymbol{X}_{\text{III},1}$ refers to the design matrix for all subjects i with a binary effect of 1. The corresponding true parameter values for each model are shown in Table 2.1.

Covariance structure 1 includes a random intercept with independent errors. Covariance structure 2 includes a random intercept and a random slope for time with independent errors. The covariance parameter values are presented in Table 2.2. Table 2.3 shows the number of replications which there is a positive definite Hessian matrix for each simulation study. The Hessian matrix is the second order partial derivative of the log likelihood function. Only replications for which there is a positive definite Hessian matrix will be included in analysis.

2.7.2 True $\boldsymbol{R^2}$

The true R^2 is introduced by Christou (2005) for a linear univariate model. The author defines R^2 as the squared correlation coefficient between the outcome and predictors and thus defines the true R^2 (population R^2) for a univariate model with one predictor as,

$$\rho^2 = \frac{\beta_1^2 \sigma_X^2}{\beta_1^2 \sigma_X^2 + \sigma_e^2}$$

Similarly, Helland (1987) has shown that under weak conditions, R^2 converges almost surely to,

$$\frac{\beta' S_X \beta}{\beta' S_X \beta + \sigma_e^2}$$

as the total number of observations tends to infinity where S_X is the sample covariance matrix for the explanatory variables which is assumed to be fixed by design.

Using the true parameter values, the asymptotic true R^2_β for each denominator degrees of freedom methods can be calculated for simulations in which there is a constant denominator degrees of freedom across all replications with a positive definite Hessian matrix. The true population measure of association for the linear mixed model will be denoted as $\rho^2_{\boldsymbol{y}_s|\boldsymbol{X}_s}$.

2.7.3 *Objectives and Methods: Large Sample Mean of R^2_β under H_0*

To determine whether our theoretical results coincide with the real world data, a simulation study was conducted to determine whether R^2_β converges to zero under the null hypothesis. To examine this result, the intercept only mean model was simulated using two different covariance structures. The two covariance structures included a random intercept and with independent errors (compound symmetry), and a random intercept and slope with independent errors. Then, these simulated datasets were analyzed using overspecified mean models. The overspecified mean parameters were then tested and an $R^2_\beta(\nu)$ was calculated. Under these conditions, it is expected that $R^2_\beta(\nu)$ values should be close to zero.

2.7.4 *Results: Large Sample Mean of R^2_β under H_0*

Table 2.4 shows the average $R^2_\beta(\nu)$ for Model 0: Covariance 1 when the data were analyzed under different mean models using covariance 1. The average $R^2_\beta(\nu)$ converges to zero as the subject sample size increases for all denominator degrees of freedom methods and all overspecified mean models. Table 2.5 shows the average R^2_β for Model 0: Covariance 2 when the data were analyzed using several mean models using covariance 2. The average R^2_β converges to zero as the subject sample size increases for all denominator degrees of freedom methods and all overspecified mean models.

2.7.5 *Objectives and Methods: Large Sample Mean of R^2_β under H_A*

Using the true parameter values, the asymptotic true $R^2_\beta(\nu)$ for REML estimation with Kenward-Roger F test and corresponding degrees of freedom can be calculated under

H_A for simulations with a constant denominator degrees of freedom across all replications and a positive definite Hessian matrix. The Kenward-Roger denominator degree of freedom formulae and their convergence values for all of the models under REML estimation are provided in Table 2.6.

2.7.5.1 *Asymptotic Properties of R^2_β: Model I: Covariance 1*

Model I with Covariance 1 refers to a model with a continuous time effect and a random intercept and independent errors (compound symmetric covariance matrix). For Model I with Covariance 1, the design matrix is the same for all i and $\mathbf{\Sigma}_i$ is the same for all i, $\boldsymbol{X}_{i,\mathrm{I}} = \boldsymbol{X}_{\mathrm{Ess,I}}$ and $\mathbf{\Sigma}_{i,1} = \mathbf{\Sigma}_{\mathrm{Ess},1}$. As shown in Section 2.5.1.1,

$$\lim_{m\to\infty} \frac{\lambda(m)}{m} = (\boldsymbol{C\beta})'\left[\boldsymbol{C}\left(\boldsymbol{X}'_{\mathrm{Ess,I}}\mathbf{\Sigma}^{-1}_{\mathrm{Ess},1}\boldsymbol{X}_{\mathrm{Ess,I}}\right)^{-1}\boldsymbol{C}'\right]^{-1}(\boldsymbol{C\beta}).$$

When using REML estimation with Kenward-Roger method,

$$\lim_{m\to\infty} E[R^2_\beta(\nu_{\mathrm{KR}})] = \frac{(\boldsymbol{C\beta})'\left[\boldsymbol{C}\left(\boldsymbol{X}'_{\mathrm{Ess,I}}\mathbf{\Sigma}^{-1}_{\mathrm{Ess},1}\boldsymbol{X}_{\mathrm{Ess,I}}\right)^{-1}\boldsymbol{C}'\right]^{-1}(\boldsymbol{C\beta})}{3 + (\boldsymbol{C\beta})'\left[\boldsymbol{C}\left(\boldsymbol{X}'_{\mathrm{Ess,I}}\mathbf{\Sigma}^{-1}_{\mathrm{Ess},1}\boldsymbol{X}_{\mathrm{Ess,I}}\right)^{-1}\boldsymbol{C}'\right]^{-1}(\boldsymbol{C\beta})}.$$

2.7.5.2 *Asymptotic Properties of R^2_β: Model I: Covariance 2*

Model I with Covariance 2 refers to a model with a continuous time effect and a random intercept and slope with independent errors. For Model I with Covariance 2, the design matrix is the same for all i and $\mathbf{\Sigma}_i$ is the same for all i, $\boldsymbol{X}_{i,\mathrm{I}} = \boldsymbol{X}_{\mathrm{Ess,I}}$ and $\mathbf{\Sigma}_{i,2} = \mathbf{\Sigma}_{\mathrm{Ess},2}$. As shown in Section 2.5.1.1,

$$\lim_{m\to\infty} \frac{\lambda(m)}{m} = (\boldsymbol{C\beta})'\left[\boldsymbol{C}\left(\boldsymbol{X}'_{\mathrm{Ess,I}}\mathbf{\Sigma}^{-1}_{\mathrm{Ess},2}\boldsymbol{X}_{\mathrm{Ess,I}}\right)^{-1}\boldsymbol{C}'\right]^{-1}(\boldsymbol{C\beta}).$$

When using REML estimation with Kenward-Roger method,

$$\lim_{m\to\infty} E[R^2_\beta(\nu_{\mathrm{KR}})] = \frac{(\boldsymbol{C\beta})'\left[\boldsymbol{C}\left(\boldsymbol{X}'_{\text{Ess,I}}\boldsymbol{\Sigma}^{-1}_{\text{Ess,2}}\boldsymbol{X}_{\text{Ess,I}}\right)^{-1}\boldsymbol{C}'\right]^{-1}(\boldsymbol{C\beta})}{1+(\boldsymbol{C\beta})'\left[\boldsymbol{C}\left(\boldsymbol{X}'_{\text{Ess,I}}\boldsymbol{\Sigma}^{-1}_{\text{Ess,2}}\boldsymbol{X}_{\text{Ess,I}}\right)^{-1}\boldsymbol{C}'\right]^{-1}(\boldsymbol{C\beta})}.$$

2.7.5.3 Asymptotic Properties of R^2_β: Model II: Covariance 1

Model II with Covariance 1 refers to a model with a continuous time effect and a binary effect with a random intercept with independent errors (compound symmetric). Using this notation, it is possible to express the noncentrality parameter as a direct function of m using Section 2.5.1.2,

$$\begin{aligned}
\lambda(m) &= (\boldsymbol{C\beta})'\left[\boldsymbol{C}\left(\sum_{i=1}^{m}\boldsymbol{X}'_i\boldsymbol{\Sigma}^{-1}_i\boldsymbol{X}_i\right)^{-1}\boldsymbol{C}'\right]^{-1}\boldsymbol{C\beta}\\
&= (\boldsymbol{C\beta})'\left[\boldsymbol{C}\left(\frac{m}{2}\boldsymbol{X}'_{\text{II},0}\boldsymbol{\Sigma}^{-1}_{\text{Ess},1}\boldsymbol{X}_{\text{II},0}+\frac{m}{2}\boldsymbol{X}'_{\text{II},1}\boldsymbol{\Sigma}^{-1}_{\text{Ess},1}\boldsymbol{X}_{\text{II},1}\right)^{-1}\boldsymbol{C}'\right]^{-1}\boldsymbol{C\beta}\\
&= (\boldsymbol{C\beta})'\left[\boldsymbol{C}\left[\frac{m}{2}\left(\boldsymbol{X}'_{\text{II},0}\boldsymbol{\Sigma}^{-1}_{\text{Ess},1}\boldsymbol{X}_{\text{II},0}+\boldsymbol{X}'_{\text{II},1}\boldsymbol{\Sigma}^{-1}_{\text{Ess},1}\boldsymbol{X}_{\text{II},1}\right)\right]^{-1}\boldsymbol{C}'\right]^{-1}\boldsymbol{C\beta}\\
&= (\boldsymbol{C\beta})'\left[\frac{2}{m}\boldsymbol{C}\left(\boldsymbol{X}'_{\text{II},0}\boldsymbol{\Sigma}^{-1}_{\text{Ess},1}\boldsymbol{X}_{\text{II},0}+\boldsymbol{X}'_{\text{II},1}\boldsymbol{\Sigma}^{-1}_{\text{Ess},1}\boldsymbol{X}_{\text{II},1}\right)^{-1}\boldsymbol{C}'\right]^{-1}\boldsymbol{C\beta}\\
&= \frac{m}{2}(\boldsymbol{C\beta})'\left[\boldsymbol{C}\left(\boldsymbol{X}'_{\text{II},0}\boldsymbol{\Sigma}^{-1}_{\text{Ess},1}\boldsymbol{X}_{\text{II},0}+\boldsymbol{X}'_{\text{II},1}\boldsymbol{\Sigma}^{-1}_{\text{Ess},1}\boldsymbol{X}_{\text{II},1}\right)^{-1}\boldsymbol{C}'\right]^{-1}\boldsymbol{C\beta}
\end{aligned}$$

Using this formulation,

$$\lim_{m\to\infty}\frac{\lambda(m)}{m} = \frac{1}{2}(\boldsymbol{C\beta})'\left[\boldsymbol{C}\left(\boldsymbol{X}'_{\text{II},0}\boldsymbol{\Sigma}^{-1}_{\text{Ess},1}\boldsymbol{X}_{\text{II},0}+\boldsymbol{X}'_{\text{II},1}\boldsymbol{\Sigma}^{-1}_{\text{Ess},1}\boldsymbol{X}_{\text{II},1}\right)^{-1}\boldsymbol{C}'\right]^{-1}\boldsymbol{C\beta}.$$

When using REML estimation with Kenward-Roger denominator degrees of freedom,

$$\lim_{m\to\infty} E[R^2_\beta(\nu_{\mathrm{KR}})] = \frac{\frac{1}{2}(\boldsymbol{C\beta})'\left[\boldsymbol{C}\left(\boldsymbol{X}'_{\text{II},0}\boldsymbol{\Sigma}^{-1}_{\text{Ess},1}\boldsymbol{X}_{\text{II},0}+\boldsymbol{X}'_{\text{II},1}\boldsymbol{\Sigma}^{-1}_{\text{Ess},1}\boldsymbol{X}_{\text{II},1}\right)^{-1}\boldsymbol{C}'\right]^{-1}\boldsymbol{C\beta}}{2+\frac{1}{2}(\boldsymbol{C\beta})'\left[\boldsymbol{C}\left(\boldsymbol{X}'_{\text{II},0}\boldsymbol{\Sigma}^{-1}_{\text{Ess},1}\boldsymbol{X}_{\text{II},0}+\boldsymbol{X}'_{\text{II},1}\boldsymbol{\Sigma}^{-1}_{\text{Ess},1}\boldsymbol{X}_{\text{II},1}\right)^{-1}\boldsymbol{C}'\right]^{-1}\boldsymbol{C\beta}}.$$

2.7.5.4 Asymptotic Properties of R^2_β: Model II: Covariance 2

Model II with Covariance 2 refers to a model with a continuous time effect and a binary effect with a random intercept and slope with independent errors. Using section 2.5.1.2,

$$\lambda(m) = \frac{m}{2}(\boldsymbol{C\beta})'\left[\boldsymbol{C}\left(\boldsymbol{X}'_{\text{II},0}\boldsymbol{\Sigma}^{-1}_{\text{Ess},2}\boldsymbol{X}_{\text{II},0} + \boldsymbol{X}'_{\text{II},1}\boldsymbol{\Sigma}^{-1}_{\text{Ess},2}\boldsymbol{X}_{\text{II},1}\right)^{-1}\boldsymbol{C}'\right]^{-1}\boldsymbol{C\beta}$$

and,

$$\lim_{m\to\infty}\frac{\lambda(m)}{m} = \frac{1}{2}(\boldsymbol{C\beta})'\left[\boldsymbol{C}\left(\boldsymbol{X}'_{\text{II},0}\boldsymbol{\Sigma}^{-1}_{\text{Ess},2}\boldsymbol{X}_{\text{II},0} + \boldsymbol{X}'_{\text{II},1}\boldsymbol{\Sigma}^{-1}_{\text{Ess},2}\boldsymbol{X}_{\text{II},1}\right)^{-1}\boldsymbol{C}'\right]^{-1}\boldsymbol{C\beta}.$$

When using REML estimation with Kenward-Roger denominator degrees of freedom,

$$\lim_{m\to\infty} E[R^2_\beta(\nu_{\text{KR}})] = \frac{\frac{1}{2}(\boldsymbol{C\beta})'\left[\boldsymbol{C}\left(\boldsymbol{X}'_{\text{II},0}\boldsymbol{\Sigma}^{-1}_{\text{Ess},2}\boldsymbol{X}_{\text{II},0} + \boldsymbol{X}'_{\text{II},1}\boldsymbol{\Sigma}^{-1}_{\text{Ess},2}\boldsymbol{X}_{\text{II},1}\right)^{-1}\boldsymbol{C}'\right]^{-1}\boldsymbol{C\beta}}{1.3333 + \frac{1}{2}(\boldsymbol{C\beta})'\left[\boldsymbol{C}\left(\boldsymbol{X}'_{\text{II},0}\boldsymbol{\Sigma}^{-1}_{\text{Ess},2}\boldsymbol{X}_{\text{II},0} + \boldsymbol{X}'_{\text{II},1}\boldsymbol{\Sigma}^{-1}_{\text{Ess},2}\boldsymbol{X}_{\text{II},1}\right)^{-1}\boldsymbol{C}'\right]^{-1}\boldsymbol{C\beta}}.$$

2.7.5.5 Asymptotic Properties of R^2_β: Model III: Covariance 1

Model III with Covariance 1 refers to a model with a continuous time effect, a binary effect and their interaction with a random intercept with independent errors. Using this notation, it is possible to express the noncentrality parameter as a direct function of m using Section 2.5.1.2,

$$\begin{aligned}
\lambda(m) &= (\boldsymbol{C\beta})'\left[\boldsymbol{C}\left(\sum_{i=1}^{m}\boldsymbol{X}'_i\boldsymbol{\Sigma}^{-1}_i\boldsymbol{X}_i\right)^{-1}\boldsymbol{C}'\right]^{-1}\boldsymbol{C\beta} \\
&= (\boldsymbol{C\beta})'\left[\boldsymbol{C}\left(\frac{m}{2}\boldsymbol{X}'_{\text{III},0}\boldsymbol{\Sigma}^{-1}_{\text{Ess},1}\boldsymbol{X}_{\text{III},0} + \frac{m}{2}\boldsymbol{X}'_{\text{III},1}\boldsymbol{\Sigma}^{-1}_{\text{Ess},1}\boldsymbol{X}_{\text{III},1}\right)^{-1}\boldsymbol{C}'\right]^{-1}\boldsymbol{C\beta} \\
&= (\boldsymbol{C\beta})'\left[\boldsymbol{C}\left[\frac{m}{2}\left(\boldsymbol{X}'_{\text{III},0}\boldsymbol{\Sigma}^{-1}_{\text{Ess},1}\boldsymbol{X}_{\text{III},0} + \boldsymbol{X}'_{\text{III},1}\boldsymbol{\Sigma}^{-1}_{\text{Ess},1}\boldsymbol{X}_{\text{III},1}\right)\right]^{-1}\boldsymbol{C}'\right]^{-1}\boldsymbol{C\beta} \\
&= (\boldsymbol{C\beta})'\left[\frac{2}{m}\boldsymbol{C}\left(\boldsymbol{X}'_{\text{III},0}\boldsymbol{\Sigma}^{-1}_{\text{Ess},1}\boldsymbol{X}_{\text{III},0} + \boldsymbol{X}'_{\text{III},1}\boldsymbol{\Sigma}^{-1}_{\text{Ess},1}\boldsymbol{X}_{\text{III},1}\right)^{-1}\boldsymbol{C}'\right]^{-1}\boldsymbol{C\beta} \\
&= \frac{m}{2}(\boldsymbol{C\beta})'\left[\boldsymbol{C}\left(\boldsymbol{X}'_{\text{III},0}\boldsymbol{\Sigma}^{-1}_{\text{Ess},1}\boldsymbol{X}_{\text{III},0} + \boldsymbol{X}'_{\text{III},1}\boldsymbol{\Sigma}^{-1}_{\text{Ess},1}\boldsymbol{X}_{\text{III},1}\right)^{-1}\boldsymbol{C}'\right]^{-1}\boldsymbol{C\beta}
\end{aligned}$$

and,

$$\lim_{m\to\infty}\frac{\lambda(m)}{m}=\frac{1}{2}(\boldsymbol{C\beta})'\Big[\boldsymbol{C}\big(\boldsymbol{X}'_{\text{III},0}\boldsymbol{\Sigma}^{-1}_{\text{Ess},1}\boldsymbol{X}_{\text{III},0}+\boldsymbol{X}'_{\text{III},1}\boldsymbol{\Sigma}^{-1}_{\text{Ess},1}\boldsymbol{X}_{\text{III},1}\big)^{-1}\boldsymbol{C}'\Big]^{-1}\boldsymbol{C\beta}.$$

2.7.5.6 Asymptotic Properties of R^2_β: Model III: Covariance 2

Model III with Covariance 2 refers to a model with a continuous time effect, a binary effect and their interaction with a random intercept and slope with independent errors. Using this notation, it is possible to express the noncentrality parameter as a direct function of m using Case 2 in 2.5.1.2,

$$\begin{aligned}
\lambda(m) &= (\boldsymbol{C\beta})'\left[\boldsymbol{C}\left(\sum_{i=1}^{m}\boldsymbol{X}'_i\boldsymbol{\Sigma}_i^{-1}\boldsymbol{X}_i\right)^{-1}\boldsymbol{C}'\right]^{-1}\boldsymbol{C\beta}\\
&= (\boldsymbol{C\beta})'\Big[\boldsymbol{C}\Big(\frac{m}{2}\boldsymbol{X}'_{\text{III},0}\boldsymbol{\Sigma}^{-1}_{\text{Ess},2}\boldsymbol{X}_{\text{III},0}+\frac{m}{2}\boldsymbol{X}'_{\text{III},1}\boldsymbol{\Sigma}^{-1}_{\text{Ess},2}\boldsymbol{X}_{\text{III},1}\Big)^{-1}\boldsymbol{C}'\Big]^{-1}\boldsymbol{C\beta}\\
&= (\boldsymbol{C\beta})'\Big[\boldsymbol{C}\Big[\frac{m}{2}\big(\boldsymbol{X}'_{\text{III},0}\boldsymbol{\Sigma}^{-1}_{\text{Ess},2}\boldsymbol{X}_{\text{III},0}+\boldsymbol{X}'_{\text{III},1}\boldsymbol{\Sigma}^{-1}_{\text{Ess},2}\boldsymbol{X}_{\text{III},1}\big)\Big]^{-1}\boldsymbol{C}'\Big]^{-1}\boldsymbol{C\beta}\\
&= (\boldsymbol{C\beta})'\Big[\frac{2}{m}\boldsymbol{C}\big(\boldsymbol{X}'_{\text{III},0}\boldsymbol{\Sigma}^{-1}_{\text{Ess},2}\boldsymbol{X}_{\text{III},0}+\boldsymbol{X}'_{\text{III},1}\boldsymbol{\Sigma}^{-1}_{\text{Ess},2}\boldsymbol{X}_{\text{III},1}\big)^{-1}\boldsymbol{C}'\Big]^{-1}\boldsymbol{C\beta}\\
&= \frac{m}{2}(\boldsymbol{C\beta})'\Big[\boldsymbol{C}\big(\boldsymbol{X}'_{\text{III},0}\boldsymbol{\Sigma}^{-1}_{\text{Ess},2}\boldsymbol{X}_{\text{III},0}+\boldsymbol{X}'_{\text{III},1}\boldsymbol{\Sigma}^{-1}_{\text{Ess},2}\boldsymbol{X}_{\text{III},1}\big)^{-1}\boldsymbol{C}'\Big]^{-1}\boldsymbol{C\beta}
\end{aligned}$$

Using this formulation,

$$\lim_{m\to\infty}\frac{\lambda(m)}{m}=\frac{1}{2}(\boldsymbol{C\beta})'\Big[\boldsymbol{C}\big(\boldsymbol{X}'_{\text{III},0}\boldsymbol{\Sigma}^{-1}_{\text{Ess},2}\boldsymbol{X}_{\text{III},0}+\boldsymbol{X}'_{\text{III},1}\boldsymbol{\Sigma}^{-1}_{\text{Ess},2}\boldsymbol{X}_{\text{III},1}\big)^{-1}\boldsymbol{C}'\Big]^{-1}\boldsymbol{C\beta}.$$

When using REML estimation with Kenward-Roger denominator degrees of freedom,

$$\lim_{m\to\infty}E[R^2_\beta(\nu_{\text{KR}})]=\frac{\frac{1}{2}(\boldsymbol{C\beta})'\Big[\boldsymbol{C}\big(\boldsymbol{X}'_{\text{III},0}\boldsymbol{\Sigma}^{-1}_{\text{Ess},2}\boldsymbol{X}_{\text{III},0}+\boldsymbol{X}'_{\text{III},1}\boldsymbol{\Sigma}^{-1}_{\text{Ess},2}\boldsymbol{X}_{\text{III},1}\big)^{-1}\boldsymbol{C}'\Big]^{-1}\boldsymbol{C\beta}}{1.1538+\frac{1}{2}(\boldsymbol{C\beta})'\Big[\boldsymbol{C}\big(\boldsymbol{X}'_{\text{III},0}\boldsymbol{\Sigma}^{-1}_{\text{Ess},2}\boldsymbol{X}_{\text{III},0}+\boldsymbol{X}'_{\text{III},1}\boldsymbol{\Sigma}^{-1}_{\text{Ess},2}\boldsymbol{X}_{\text{III},1}\big)^{-1}\boldsymbol{C}'\Big]^{-1}\boldsymbol{C\beta}}.$$

2.7.6 Results: Asymptotic Mean of R^2_β under H_A

Table 2.6 provides the Kenward-Roger denominator degrees of freedom using REML estimation and the equations used to predict the Kenward-Roger denominator degrees of freedom using the subject sample size. Table 2.7 shows the average simulated and true values of $R^2_\beta(\nu)$ for each model and covariance structure for which the Kenward-Roger denominator degrees of freedom are constant across all 10,000 replications. The simulated values of $R^2_\beta(\nu)$ are converging to the asymptotic true $R^2_\beta(\nu)$ for each denominator degrees of freedom method.

2.7.7 Objectives: Hypothesis Testing

Three test statistics were of interest in this simulation study, the Beta distribution theory statistic, the Taylor series approximation statistic, and the F statistic in the linear mixed model. The simulation study objectives were to calculate and compare the Type I error rates and power for these statistics. In addition, the goal was to compare the results from these three test statistics.

Additional simulations were run to calculate the power. The mean model consists of an intercept and a continuous time effect with a parameter estimate of 0.03. This mean model was simulated for subject sample sizes of $10,\ 20,\ 50,\ 100,\ 200$ using covariance structures and values defined in Table 2.2.

The test statistics were examined under the null hypothesis and different alternative hypothesis for the simulation studies. Type I error rates were calculated by comparing the test statistic to their respective comparison distribution. Specifically, the Type I error rates were calculated for true Model 0 with covariance structures 1 and 2 when analyzing the data with overspecified mean models (Models I, II, and III).

To calculate the Type I error, data were generated under the null hypothesis, $H_0 : \boldsymbol{C\beta} = \mathbf{0}$ for three different covariance structures. The mean model consisted of only an intercept denoted as model 0 for these three covariance structures, and the true parameter value is in Table 2.1. Then, the simulated data were analyzed under three different incorrect null hypotheses

$$
\begin{aligned}
&H_{0_{\mathrm{I}}} : \beta_1 = 0 \\
&H_{0_{\mathrm{II}}} : \beta_1 = \beta_2 = 0 \\
&H_{0_{\mathrm{III}}} : \beta_1 = \beta_2 = \beta_3 = 0
\end{aligned}
\tag{9.4}
$$

where β_1 represents the parameter estimate corresponding to the continuous time effect ; β_2 represents the parameter estimate corresponding to the binary effect; β_3 represents the parameter estimate corresponding to the binary effect and continous time interaction effect.

For each test statistic, the p-value was calculated to determine if the null hypothesis was rejected by comparing the test statistic to the appropriate value from the hypothesized distribution.

2.7.8 Results: Hypothesis Testing

The Type I error rates are shown in Tables 2.8 and 2.9. The simulation study results indicate that comparing $R^2_{\beta}(\nu)$ to the Beta distribution is an equivalent test of the hypothesis that $H_0 : \boldsymbol{C\beta} = \mathbf{0}$. The Type I error rates for the Beta distribution theory test statistic are exactly the same as the Type I error rates for the F statistic in the linear mixed model. Comparing the Taylor series approximation to the standard normal distribution consistently resulted in inflated Type I error rates.

There were 10,000 replications for each power simulation. The approximate half-width of a 95% confidence interval for power is calculated as $1.96\left[\frac{P(1-P)}{10{,}000}\right]^{\frac{1}{2}}$. Theoretical values of power for each simulation were computed by first determining the critical value of the null hypothesized distribution. Then, that critical value was evaluated in the cumulative distribution function of the alternative hypothesized distribution. Approximations of the non-central Beta distribution were computed using version 2.13.0 of R. Table 2.10 compares the empirical power with the theoretical prediction of power. For the Beta distribution test statistic, the difference between the theoretical and empirical power fell with the 95% confidence interval for all of the covariance 1 simulations. The Taylor series approximation test statistic was not very accurate. The difference was never within the 95% confidence interval.

2.8 Conclusions

2.8.1 Summary of the Mean and Variance of R_β^2

Using the Beta distribution, under the null hypothesis,

$$E\left[R_\beta^2(\nu) \middle| H_0\right] = \frac{q-1}{q+\nu-1}$$

$$V\left[R_\beta^2(\nu) \middle| H_0\right] = \frac{2(q-1)\nu}{(q+\nu-1)^2(q+\nu+1)}$$

and under the alternative hypothesis,

$$E\left[R_\beta^2(\nu) \middle| H_\mathrm{A}\right] = 1 - \frac{\nu}{(q-1+\nu+\lambda)}\left(1 + \frac{2\lambda}{(q-1+\nu+\lambda)^2}\right)$$

$$V\left[R_\beta^2(\nu) \middle| H_\mathrm{A}\right] \simeq \frac{\lambda\left(\frac{\nu}{2}\right)^2}{2C^4} + \frac{\frac{\nu}{2}}{G}\left[1 + \frac{\lambda}{2}\left(\lambda^2 + 3\lambda + H\right)/G^2\right] - \frac{\left(\frac{\nu}{2}\right)^2}{F}\left(1 + D/F^2\right)$$

where

$$C = \tfrac{q-1}{2} + \tfrac{\nu}{2} + \tfrac{\lambda}{2}$$

$$G = C(C+1) + \tfrac{\lambda}{2} = \left(\tfrac{q-1}{2} + \tfrac{\nu}{2} + \tfrac{\lambda}{2}\right)\left(\tfrac{q-1}{2} + \tfrac{\nu}{2} + \tfrac{\lambda}{2} + 1\right) + \tfrac{\lambda}{2},$$

$$S = \left[2\left(\tfrac{q-1}{2} + \tfrac{\nu}{2}\right) + 1\right]^2,$$

$$F = \left(\tfrac{q-1}{2} + \tfrac{\nu}{2}\right)^2\left(\tfrac{q-1}{2} + \tfrac{\nu}{2} + 1\right) + H\tfrac{\lambda}{2} + \left[3\left(\tfrac{q-1}{2} + \tfrac{\nu}{2}\right) + 4\right]\left(\tfrac{\lambda}{2}\right)^2 + \left(\tfrac{\lambda}{2}\right)^3,$$

$$H = 3\left(\tfrac{q-1}{2} + \tfrac{\nu}{2}\right)^2 + 5\left(\tfrac{q-1}{2} + \tfrac{\nu}{2}\right) + 2,$$

$$D = \tfrac{\lambda}{2}\left[H^2 + 2P\tfrac{\lambda}{2} + Q\left(\tfrac{\lambda}{2}\right)^2 + R\left(\tfrac{\lambda}{2}\right)^3 + 9\left(\tfrac{\lambda}{2}\right)^4\right],$$

$$P = \left[3\left(\tfrac{q-1}{2} + \tfrac{\nu}{2}\right) + 1\right]\left[9\left(\tfrac{q-1}{2} + \tfrac{\nu}{2}\right) + 17\right]$$

$$+ 2\left(\tfrac{q-1}{2} + \tfrac{\nu}{2}\right)\left[3\left(\tfrac{q-1}{2} + \tfrac{\nu}{2}\right) + 2\right]\left[3\left(\tfrac{q-1}{2} + \tfrac{\nu}{2}\right) + 4\right] + 15,$$

$$Q = 54\left(\tfrac{q-1}{2} + \tfrac{\nu}{2}\right)^2 + 162\left(\tfrac{q-1}{2} + \tfrac{\nu}{2}\right) + 130,$$

$$R = 6\left[6\left(\tfrac{q-1}{2} + \tfrac{\nu}{2}\right) + 11\right]$$

Using the Taylor series approximation approach, using the first order approximation, under the null hypothesis,

$$E\left[R_\beta^2(\nu)\middle|\, H_0\right] = \frac{q-1}{q+\nu-3}$$

$$V\left[R_\beta^2(\nu)\middle|\, H_0\right] = \frac{2(q-1)(\nu-2)^2}{(\nu-4)(\nu+q-3)^3}$$

and under the alternative hypothesis using the first order approximation,

$$E\left[R_\beta^2(\nu)\middle|\, H_\mathrm{A}\right] = \frac{q-1+\lambda}{q+\nu+\lambda-3}$$

$$V\left[R_\beta^2(\nu)\middle|\, H_\mathrm{A}\right] = \frac{2(\nu-2)^2\left[(q-1+\lambda)^2 + (q-1+2\lambda)(\nu-2)\right]}{(\nu+q+\lambda-3)^4(\nu-4)}$$

Using the Taylor series approximation approach, using the second order approximation, under the null hypothesis,

$$E\left[R_\beta^2(\nu)\middle|\, H_0\right] = \frac{q-1}{q+\nu-3} - \frac{2(q-1)}{(\nu-4)}\frac{(\nu-2)}{(\nu+q-3)^2}$$

and under the alternative hypothesis using the second order approximation,

$$E\left[R_\beta^2(\nu)\middle|\, H_\mathrm{A}\right] = \frac{q-1+\lambda}{q+\nu+\lambda-3} - 2\frac{\left[(q-1+\lambda)^2(\nu-2) + (q-1+2\lambda)(\nu-2)^2\right]}{(\nu-4)(\nu+q+\lambda-3)^3}$$

2.8.2 Summary of the Asymptotic Properties of the Mean and Variance of $\boldsymbol{R_\beta^2}$

While the approximate values of $E\left[R_\beta^2(\nu)\right]$ and $V\left[R_\beta^2(\nu)\right]$ are different when using the Beta distribution theory and the Taylor series approximation approaches, the asymptotic properties are the same. Under the null hypothesis and assuming consistent estimators of the population parameters $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$, where $\boldsymbol{\Sigma}$ correctly specified,

$$\lim_{m\to\infty} E[R_\beta^2(\nu)|\, H_0] = 0$$

$$\lim_{\nu\to\infty} V[R_\beta^2(\nu)|\, H_0] = 0$$

Under the alternative hypothesis,

$$\lim_{m\to\infty} E[R_\beta^2(\nu)\big|\, H_{\mathrm{A}}] = \frac{\lim\limits_{m\to\infty} \frac{\lambda(m)}{m}}{\lim\limits_{m\to\infty} \frac{\lambda(m)}{m} + \lim\limits_{m\to\infty} \frac{\nu(m)}{m}}$$

$$\lim_{\nu\to\infty} V[R_\beta^2(\nu)\big|\, H_{\mathrm{A}}] = 0$$

2.8.3 Conclusions

For simulations for which the denominator degrees of freedom could be almost perfectly predicted using subject sample size, the asymptotic mean of R_β^2 could be calculated because the convergence properties of the denominator degrees of freedom method could be ascertained.

The Type I error rate for the Beta distribution test statistic is equivalent to the usual F statistic. The power of the Beta distribution test statistic was fairly accurate. The Taylor series approximation test statistic had inflated Type I error rates as well as inaccurate power estimates.

Further research should investigate the asymptotic properties of $R_\beta^2(\nu)$ under misspecified covariance, the finite sample properties of $R_\beta^2(\nu)$, and examine the impact of varying the denominator degrees of freedom methods and estimation techniques used to define $R_\beta^2(\nu)$. In addition, evaluating $R_\beta^2(\nu)$ as a model selection tool should also be addressed.

REFERENCES

Casella, G. and Berger, R.L. (2002). *Statistical Inference*. 2nd edition. Duxbury.

Chattamvelli, R. and Shanmugam, R. (1997). Algorithm AS 310: Computing the non-central beta distribution function. *Applied Statistics* **46**, 146-156.

Christou, N. (2005). The true R^2 and the truth about R^2. Center for the Teaching of Statistics, UCLA.

Edwards, L.J., Muller, K.E., Wolfinger, R.D., Qaqish, B.F., and Schabenberger, O. (2008). An R^2 statistic for fixed effects in the linear mixed model. *Statistics in Medicine* **27**, 6137-6157.

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

Helms, R. W. (1988). Comparisons of parameter and hypothesis definitions in a general linear model. *Communications in Statistics, Series A* **17**, 2725-2753.

Johnson, N.L., Kotz, S. and Balakrishnan, N. (1994). *Continuous Univariate Distributions, Vol. 2*. Wiley.

Kenward, M.G. and Roger, J.H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics* **53**, 983-997.

Mardia, K.V., Kent, J.T., and Bibby, J.M. (1992). *Multivariate Analysis*. New York: Academic Press Limited.

Muller, K.E., and Stewart, P.W. (2006). *Linear model theory: Univariate, multivariate, and mixed models.* Hoboken, NJ: Wiley.

Oehlert, G.W. (1992). A note on the delta method. *The American Statistician* **46**, 27-29.

Potthoff, R.F. and Roy, S.N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika* **51**, 313-326.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. John Wiley and Sons. New York: Wiley Series in Probability and Mathematical Statistics.

TABLES

Table 2.1 Simulated fixed effect parameter values

Model 0	Model I	Model II	Model III
$\boldsymbol{\beta}_0 = 16.8$	$\boldsymbol{\beta}_{\mathrm{I}} = \begin{bmatrix} 16.8 \\ 0.7 \end{bmatrix}$	$\boldsymbol{\beta}_{\mathrm{II}} = \begin{bmatrix} 15.5 \\ 2.2 \\ 0.7 \end{bmatrix}$	$\boldsymbol{\beta}_{\mathrm{III}} = \begin{bmatrix} 17.4 \\ 0.5 \\ -1.0 \\ 0.3 \end{bmatrix}$

Table 2.2 Simulated covariance parameter values

	Covariance 1	Covariance 2
$\boldsymbol{R}_i$	$\sigma_{\mathrm{e}}^2 = 4.50$	$\sigma_{\mathrm{e}}^2 = 1.72$
$\boldsymbol{D}$	$\sigma_{\mathrm{d}}^2 = 2.02$	$\begin{bmatrix} \sigma_{\mathrm{Int}}^2 & \rho \\ \rho & \sigma_{\mathrm{Slope}}^2 \end{bmatrix} = \begin{bmatrix} 5.79 & -0.29 \\ -0.29 & 0.03 \end{bmatrix}$

Table 2.3 Number of replications where there is a positive definite hessian matrix

	25 subjects	50 subjects	100 subjects	200 subjects	500 subjects	1000 subjects
Model I: Cov 1	9,966	10,000	10,000	10,000	10,000	10,000
Model I: Cov 2	6,863	7,984	8,963	9,692	9,984	10,000
Model I: Cov 3	8,035	9,178	9,812	9,987	10,000	10,000
Model II: Cov 1	9,965	10,000	10,000	10,000	10,000	10,000
Model II: Cov 2	6,901	7,949	8,978	9,721	9,981	10,000
Model II: Cov 3	8,010	9,138	9,806	9,988	10,000	10,000
Model III: Cov 1	9,961	10,000	10,000	10,000	10,000	10,000
Model III: Cov 2	6,801	7,997	9,023	9,708	9,985	10,000
Model III: Cov 3	7,983	9,133	9,816	9,989	10,000	10,000

Table 2.4. Average simulated R_β^2 values for true simulated data from Model 0: Covariance 1 analyzed using various mean models with covariance 1

	Model I Cov 1	Model II Cov 1	Model III Cov 1
25 subjects	0.0134	0.0419	0.0467
50 subjects	0.0067	0.0205	0.0229
100 subjects	0.0034	0.0102	0.0115
200 subjects	0.0017	0.0050	0.0057
500 subjects	0.0007	0.0020	0.0022
1000 subjects	0.0003	0.0010	0.0011

Table 2.5. Average simulated R^2_β values for true simulated data from Model 0: Covariance 2 analyzed using various mean models with covariance 2

	Model I Cov 2	Model II Cov 2	Model III Cov 2
25 subjects	0.0360	0.0590	0.0996
50 subjects	0.0190	0.0298	0.0505
100 subjects	0.0094	0.0147	0.0256
200 subjects	0.0049	0.0074	0.0130
500 subjects	0.0019	0.0029	0.0051
1000 subjects	0.0010	0.0015	0.0026

Table 2.6. Constant Kenward-Roger denominator degrees of freedom using REML estimation for simulated linear mixed models and prediction equations using subject sample size

	Model I Cov 1	Model I Cov 2	Model II Cov 1	Model II Cov 2	Model III Cov 1	Model III Cov 2
25 subjects	74	24	46.09	30.62	61.35	25.29
50 subjects	149	49	96.11	63.96	127.55	54.13
100 subjects	299	99	196.12	130.63	259.92	111.82
200 subjects	599	199	396.13	263.97	524.63	227.20
500 subjects	1499	499	996.13	663.97	1318.75	573.35
1000 subjects	2999	999	1996.13	1330.63	2642.28	1150.28
2000 subjects	5999	1999	3996.13	2663.97	5289.34	2304.12
5000 subjects	14999	4999	9996.13	6663.97	13230.52	5765.66
10000 subjects	29999	9999	19996.13	13330.63	26465.81	11534.89
Prediction Equation:	$\nu = 3m$-1	$\nu = m$-1	$\nu = 2m$-3.9	$\nu = 1.33m$-2.7	$\nu = 2.65m$-4.8	$\nu = 1.15m$-3.6
$\lim\limits_{m \to \infty} \frac{\nu_{\mathrm{KR}}}{m}$	3	1	2	1.33	2.65	1.15

Table 2.7. Average simulated R_β^2 using Kenward-Roger denominator degrees of freedom and REML estimation and the corresponding asymptotic true R_β^2

	Model I Cov 1	Model I Cov 2	Model II Cov 1	Model II Cov 2	Model III Cov 1	Model III Cov 2
25 subjects	0.3960	0.7784	0.5576	0.7401	0.6621	0.8656
50 subjects	0.3943	0.7832	0.5455	0.7436	0.6483	0.8633
100 subjects	0.3937	0.7874	0.5404	0.7463	0.6439	0.8634
200 subjects	0.3934	0.7893	0.5379	0.7471	0.6415	0.8637
500 subjects	0.3924	0.7899	0.5365	0.7472	0.6396	0.8634
1000 subjects	0.3923	0.7899	0.5358	0.7467	0.6392	0.8631
2000 subjects	0.3923	0.7898	0.5356	0.7466	0.6389	0.8630
5000 subjects	0.3923	0.7897	0.5354	0.7466	0.6388	0.8628
10000 subjects	0.3923	0.7897	0.5353	0.7464	0.6387	0.8628
Asymptotic True $R_\beta^2(\nu)$:	**0.3923**	**0.7897**	**0.5353**	**0.7541**	**0.6386**	**0.8718**

Table 2.8. Type I error rates for true Model 0: Covariance 1

	REML KR F test	**REML KR Beta test**	**REML KR Taylor Test**
$H_{0_{\mathrm{I}}} : \beta_1 = 0$ with Covariance 1			
25 subjects	0.0512	0.0512	0.0644
50 subjects	0.0514	0.0514	0.0659
100 subjects	0.0521	0.0521	0.0663
200 subjects	0.0479	0.0479	0.0660
500 subjects	0.0500	0.0500	0.0670
1000 subjects	0.0479	0.0479	0.0661
$H_{0_{\mathrm{II}}} : \beta_1 = \beta_2 = 0$ with Covariance 1			
25 subjects	0.0521	0.0521	0.0618
50 subjects	0.0502	0.0502	0.0664
100 subjects	0.0516	0.0516	0.0689
200 subjects	0.0485	0.0485	0.0672
500 subjects	0.0501	0.0501	0.0707
1000 subjects	0.0504	0.0504	0.0709
$H_{0_{\mathrm{III}}} : \beta_1 = \beta_2 = \beta_3 = 0$ with Covariance 1			
25 subjects	0.0507	0.0507	0.0621
50 subjects	0.0498	0.0498	0.0658
100 subjects	0.0511	0.0511	0.0681
200 subjects	0.0498	0.0498	0.0703
500 subjects	0.0487	0.0487	0.0690
1000 subjects	0.0507	0.0507	0.0704

Table 2.9. Type I error rates for true Model 0: Covariance 2

	REML KR F test	**REML KR Beta test**	**REML KR Taylor Test**
$H_{0_{\mathrm{I}}} : \beta_1 = 0$ with Covariance 2			
25 subjects	0.0383	0.0383	0.0416
50 subjects	0.0450	0.0450	0.0556
100 subjects	0.0449	0.0449	0.0604
200 subjects	0.0452	0.0452	0.0623
500 subjects	0.0479	0.0479	0.0656
1000 subjects	0.0493	0.0493	0.0650
$H_{0_{\mathrm{II}}} : \beta_1 = \beta_2 = 0$ with Covariance 2			
25 subjects	0.0429	0.0429	0.0483
50 subjects	0.0486	0.0486	0.0603
100 subjects	0.0484	0.0484	0.0635
200 subjects	0.0470	0.0470	0.0659
500 subjects	0.0470	0.0470	0.0674
1000 subjects	0.0506	0.0506	0.0703
$H_{0_{\mathrm{III}}} : \beta_1 = \beta_2 = \beta_3 = 0$ with Covariance 2			
25 subjects	0.0377	0.0377	0.0353
50 subjects	0.0449	0.0449	0.0538
100 subjects	0.0491	0.0491	0.0648
200 subjects	0.0478	0.0478	0.0654
500 subjects	0.0483	0.0483	0.0672
1000 subjects	0.0503	0.0503	0.0691

Table 2.10. Theoretical and empirical power estimates for simulated models with $\beta = 0.03$

		Beta Distribution Test Statistic			**Taylor Series Approx Test Statistic**		
Cov	m	Theoretical	Empirical	Half-Width	Theoretical	Empirical	Half-Width
1	10	0.0543	0.0569	0.0044	0.0311	0.0628	0.0034*
1	20	0.0589	0.0598	0.0046	0.0381	0.0740	0.0038*
1	50	0.0729	0.0735	0.0051	0.0615	0.0935	0.0047*
1	100	0.0966	0.0926	0.0058	0.1048	0.1165	0.0060*
1	200	0.1452	0.1465	0.0069	0.1927	0.1790	0.0077*
2	10	0.0580	0.0204	0.0046*	0.0364	0.0071	0.0037*
2	20	0.0671	0.0462	0.0049*	0.0511	0.0450	0.0043*
2	50	0.0946	0.0869	0.0057*	0.1005	0.1026	0.0059*
2	100	0.1416	0.1348	0.0068	0.1857	0.1617	0.0076*
2	200	0.2373	0.2297	0.0083	0.3325	0.2750	0.0092*

CHAPTER 3

IMPACT OF COVARIANCE STRUCTURE MISSPECIFICATION, DENOMINATOR DEGREES OF FREEDOM AND ESTIMATION TECHNIQUE ON R^2_β

3.1 Introduction

There has been considerable interest among researchers regarding an R^2 statistic for the linear mixed model. However, R^2 statistics for the linear mixed model are new statistical tools. Edwards et al (2008) introduced a new R^2 statistic in the linear mixed model, R^2_β for fixed effects with many desirable features. The R^2_β statistic has a semi-partial form as well as a one-to-one correspondence with the Hotelling-Lawley trace multivariate measure of association. While the performance of R^2_β was examined when introduced, further investigation is warranted.

3.1.1 Motivation

Linear mixed models are an important tool used to analyze longitudinal data. In practice, when fitting a linear mixed model, if the model fails to converge, the common practice is to simplify either the mean or the covariance model. It is important to understand what impact that change and potential misspecification has on the statistics being used to analyze the data. This chapter evaluates the impact of covariance structure misspecification on R^2_β.

The denominator degrees of freedom methods and estimation techniques used to define R^2_β are also evaluated. When testing fixed effects in the linear mixed model for longitudinal data, the Kenward-Roger F statistic and corresponding denominator degrees of freedom should be calculated under REML estimation has been shown to have improved Type I error rates (Kenward and Roger, 1997). If inference for the fixed effects is not of interest when using R^2_β, then there is potential that R^2_β could be defined using other denominator degrees of freedom methods and estimation techniques. Therefore, investigation into defining R^2_β for other denominator degrees of freedom methods and estimation techniques is also important.

3.2 R^2_β Notation Discussion

Edwards et al. (2008) proposed an R^2 statistic for the fixed effects in the linear mixed model. The newly proposed R^2_β is as follows,

$$R^2_\beta = \frac{(q-1)\nu^{-1}F\left(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Sigma}}\right)}{1+(q-1)\nu^{-1}F\left(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Sigma}}\right)}.$$

The notation for the R^2_β will be expanded to take into account various linear mixed model conditions and assumptions. Edwards et al. (2008) implicit assumptions for the proposed R^2_β are that the denominator degrees of freedom are known and the covariance structure is correctly specified; therefore, the proposed R^2_β will be denoted as $R^2_\beta(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Sigma}} \big| \boldsymbol{\nu})$. $R^2_\beta(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Sigma}}^* \big| \boldsymbol{\nu})$ will denote the multivariate measure of association under the assumption that the denominator degrees of freedom are known and the true covariance structure is misspecified. While it is true that there are some special cases of the linear mixed model and estimation methods where the denominator degrees of freedom are known, it is more common that the denominator degrees of freedom will have to be estimated from the data. $R^2_\beta(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Sigma}}, \widehat{\boldsymbol{\nu}})$ will denote the multivariate measure of association when the denominator degrees of freedom are estimated from the data and the covariance structure is correctly specified. Similarly, $R^2_\beta(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Sigma}}^*, \widehat{\boldsymbol{\nu}})$ will denote the multivariate measure of association when the denominator degrees of freedom are estimated from the data and the covariance structure is misspecified. The primary focus of this chapter will be on deriving the asymptotic properties of $R^2_\beta(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Sigma}}^* \Big| \boldsymbol{\nu})$ for the four different denominator degrees of freedom methods.

To further clarify notation, the denominator degrees of freedom methods used to define R^2_β will be added. The Kenward-Roger method denominator degrees of freedom will be denoted as $\boldsymbol{\nu}_{\text{KR}}$, the Satterthwaite method as $\boldsymbol{\nu}_{\text{Sat}}$, the residual method as $\boldsymbol{\nu}_{\text{Res}}$, and the containment method as $\boldsymbol{\nu}_{\text{Contain}}$.

3.3 Denominator Degrees of Freedom Methods

There are several methods for determining the denominator degrees of freedom, ν, for the F-statistic which include residual, containment, Satterthwaite, and Kenward-Roger method. Section 1.4.3 provides a detailed review of those methods. An important distinction when comparing each of these methods is that when using the same estimation technique (ML or REML), the residual, Satterthwaite, and containment methods all result in the same F value. These three methods only differ in the distribution to which they compare that statistic. For example, under the residual denominator degrees of freedom method, the F statistic is compared to an F distribution with rank$(\boldsymbol{C})$ numerator degrees of freedom and residual denominator degrees of freedom to compute the p-value. Under most cases, the Kenward-Roger F value differs from the other statistics because under this method, an adjusted version of $[\boldsymbol{X}_s' \boldsymbol{\Sigma}_s^{-1}(\widehat{\boldsymbol{\tau}}) \boldsymbol{X}_s]^{-1}$ is estimated.

3.3.1 Asymptotic Properties of the Denominator Degrees of Freedom Methods

As mentioned in Section 2.5.2. the asymptotic properties of each denominator degrees of freedom method varies. The residual denominator degrees of freedom has a constant formula for all complete and balanced designs. The residual denominator degrees of freedom formula is $\nu_{\text{Res}}(m) = nm - q$. Therefore,

$$\lim_{m\to\infty}\frac{\nu(m)}{m}=\lim_{m\to\infty}\frac{nm-q}{m}=n.$$

The asymptotic properties for specific mean models and covariance structures of other denominator degrees of freedom methods are examined in Section 3.5.

3.4 Covariance Structure Misspecification

One of the simplest covariance structures used in the linear mixed model is the compound symmetric covariance. The compound covariance structure is denoted as $\boldsymbol{\Sigma}_{i,\text{CS}}=\sigma^2\left[\rho\mathbf{1}_{n_i}\mathbf{1}'_{n_i}-(1-\rho)\boldsymbol{I}_{n_i}\right]$ where σ^2 denotes the common variance, ρ denotes the intraclass correlation, and n_i represents the number of repeated measures for subject i. Often researchers will use the compound symmetric covariance structure if there are problems with model convergence when fitting the linear mixed model. It is often overlooked as to the consequences of covariance structure underspecification. Gurka, Edwards, and Muller (2011) describe the bias that arises in inference when underspecifying the covariance structure as compound symmetric.

Suppose the true model is a linear mixed model with a true covariance structure that is not compound symmetric, but the model is fit using a compound symmetric structure. Kistner and Muller (2004) have shown how the estimates of the common variance and intraclass correlation can be derived from the unstructured REML estimates of $\boldsymbol{\Sigma}_i$. Gurka, Edwards and Muller (2011) provide the formulae,

$$\widehat{\sigma}^2=\frac{\text{tr}\left(\widehat{\boldsymbol{\Sigma}}_i\right)}{n},$$

$$\widehat{\rho}=\frac{\left[\mathbf{1}_n\widehat{\boldsymbol{\Sigma}}_i\mathbf{1}'_n-\text{tr}\left(\widehat{\boldsymbol{\Sigma}}_i\right)\right]}{(n-1)\text{tr}\left(\widehat{\boldsymbol{\Sigma}}_i\right)}.$$

Assume $\widehat{\boldsymbol{\Sigma}}_i \rightarrow \boldsymbol{\Sigma}_i$ and n is dominated by m, then $\widehat{\sigma}^2 \rightarrow_p \frac{\text{tr}(\boldsymbol{\Sigma}_i)}{n} = \sigma^2_{\text{Miss}}$ where σ^2_{Miss} denotes the asymptotic misspecified common variance. Also under those assumptions, $\widehat{\rho} \rightarrow_p \frac{[\mathbf{1}_n \boldsymbol{\Sigma}_i \mathbf{1}'_n - \text{tr}(\boldsymbol{\Sigma}_i)]}{(n-1)\text{tr}(\boldsymbol{\Sigma}_i)} = \rho_{\text{Miss}}$ where ρ_{Miss} denotes the asymptotic misspecified intraclass correlation. Given the true covariance parameters, the asymptotic misspecified parameter values can be calculated, and the asymptotic properties of R^2_β under misspecification can be derived.

3.5 Simulation Study

The large simulation study used in Chapter 2 was used to examine the impact of covariance structure misspecification for the various denominator degrees of freedom methods. Simulated data from a new covariance structure was added to this simulation study. Covariance structure 3 is the structure which allows heterogeneity between the dichotomous effect for the error variance, and has a random slope and a random intercept. The covariance parameter values are presented in Table 3.1.

3.5.1 Objectives

The objectives of this simulation study is to examine the impact of covariance structure misspecification, estimation technique, and denominator degrees of freedom method on the large sample properties of R^2_β.

3.5.2 Methods

For each simulation study, eight different values of R^2_β are calculated by varying the estimation technique and the denominator degrees of freedom methods when analyzing the simulated data using the true mean model and covariance structure. The simulated data are also analyzed with incorrect covariance structures for each denominator degrees of freedom method.

The asymptotic true R^2_β were calculated for each denominator degrees of freedom method with a constant value across all replications with a positive definite Hessian matrix. Table 3.2 provides the formula for calculating the denominator degrees of freedom based on the subject sample size. Details of the computation of the asymptotic true R^2_β follow exactly as in Section 2.7.5.

For some of the simulated models, the denominator degrees of freedom for Kenward-Roger, and Satterthwaite are not constant across all 10,000 replications with a positive definite Hessian matrix, but the residual and containment denominator degrees of freedom are constant though.

Figure 3.1-3.4 show the plot of the denominator degrees of freedom by subject sample size each for a different simulation study. For each of the figures, the prediction lines for containment and residual denominator degrees of freedom represent a perfect fit, while, the prediction lines for Kenward-Roger and Satterthwaite denominator degrees of freedom represent a very nearly perfect fit. From the figures, the ordering of the denominator degrees of freedom methods increases from Satterthwaite to Kenward-Roger to containment to residual.

3.5.2.1 Derivation of Asymptotic R^2_β for Misspecified Covariance Structure

The true value of Covariance 2 $\boldsymbol{\Sigma}_i$ is

$$\boldsymbol{\Sigma}_i = \boldsymbol{Z}_i \begin{bmatrix} 5.79 & \text{-0.29} \\ \text{-0.29} & 0.03 \end{bmatrix} \boldsymbol{Z}_i' + (1.72)\, \boldsymbol{I}_4.$$

The estimates of parameter in a misspecified compound symmetric structure defined in Section 3.4 will converge to

$$\sigma^2_{\text{Miss}} = \frac{\text{tr}(\boldsymbol{\Sigma}_i)}{4} = 4.91$$

and

$$\rho_{\text{Miss}} = \frac{[\mathbf{1}_n \boldsymbol{\Sigma}_i \mathbf{1}_n' - \text{tr}(\boldsymbol{\Sigma}_i)]}{3\text{tr}(\boldsymbol{\Sigma}_i)} = 0.61.$$

The asymptotic values of R_β^2 when incorrectly assuming a compound symmetric covariance structure can be calculated for this model using the same principles as outlined in Chapter 2 with

$$\boldsymbol{\Sigma}_i = \begin{bmatrix} 4.91 & 2.99 & 2.99 & 2.99 \\ 2.99 & 4.91 & 2.99 & 2.99 \\ 2.99 & 2.99 & 4.91 & 2.99 \\ 2.99 & 2.99 & 2.99 & 4.91 \end{bmatrix}.$$

Note that $\sigma^2 \rho = 4.91(0.61) = 2.99$. Similar analysis can be done for Covariance 3.

3.5.3 Results

3.5.3.1 ML $R_\beta^2(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Sigma}} | \boldsymbol{\nu})$ vs. REML $R_\beta^2(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Sigma}} | \boldsymbol{\nu})$

The comparison of the average ML R_β^2 and the average REML R_β^2 for a simulated mean model and covariance structure are provided in Tables 3.3, 3.4, 3.5, 3.7, 3.8, 3.10, 3.11, 3.12, 3.14, 3.16, 3.17. The average ML R_β^2's are almost identical to the average REML R_β^2 for all simulated models and covariance structures. Even for those simulation studies with low subject sample size, the REML R_β^2's were very similar to the ML R_β^2's for each denominator degrees of freedom method. The largest difference between the average ML R_β^2 and the average REML R_β^2 was 0.013 in the simulation study of Model III with Covariance 2 with 25 subjects using the residual method.

3.5.3.2 Asymptotic Properties of $R_\beta^2(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Sigma}} | \boldsymbol{\nu})$

The average simulated values of R_β^2 are converging to the asymptotic R_β^2 values for each denominator degrees of freedom method. In addition, the asymptotic true R_β^2 using residual denominator degrees of freedom is significantly smaller than the asymptotic true R_β^2

using the other denominator degrees of freedom methods as a result of the large difference in $\lim_{m\to\infty} \frac{\nu_{\text{Res}}(m)}{m}$ as compared to the other limits.

3.5.3.3 *Asymptotic Properties of* $R^2_\beta(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Sigma}}^* | \boldsymbol{\nu})$

Figures 3.5-3.13 present the average R^2_β for each of the nine simulated models for the four denominator degrees of freedom methods under REML estimation. These figures also show the average R^2_β under covariance structure misspecification for the four denominator degrees of freedom methods under REML estimation. When the true covariance structure is compound symmetric (Figures 3.5, 3.8, 3.11), if the covariance structure is overspecified, the average R^2_β is substantially larger than the true R^2_β when using Kenward-Roger, Satterthwaite, and containment methods. These findings are consistent with Kramer (2005) who found that as the model complexity increased the two R^2 statistics being evaluated increased. In addition, the average R^2_β for the overspecified models are converging to a different value than the average R^2_β for the Kenward-Roger, Satterthwaite, and containment methods. The average R^2_β for the true model using residual method is fairly similar to the average R^2_β using residual method for the overspecified models.

When the simulated covariance structure contains a random intercept and a random slope with hetereogeneous residual errors (Covariance 3 shown in Figures 3.7, 3.10, 3.13), if the covariance structure is underspecified, the average R^2_β for the underspecified models is substantially smaller than the average R^2_β for the true model using Kenward-Roger, Satterthwaite, and containment methods. The average R^2_β for the true model using residual method is fairly similar to the average R^2_β using residual method for the overspecified models. Tables 3.18 and 3.19 show the true asymptotic R^2_β values of both the true covariance structure and the underspecified covariance structure using Kenward-Roger and residual methods. The impact of underspecification of covariance structure varies based on which denominator degree of freedom method is used.

3.6 Conclusions and Discussion

The results from the simulation study are valuable in illustrating the properties and potential pitfalls of defining R^2_β using various denominator degrees of freedom methods and estimation techniques. For the simulation studies examined, the estimation technique does not impact the values of R^2_β even for the smaller sample size simulations. As assumed, the R^2_β using REML estimation converges to the R^2_β using ML estimation.

As suggested in Edwards et al. (2008), the convergence of R^2_β is clearly affected by the choice of the denominator degrees of freedom method. R^2_β using residual degrees of freedom are consistently lower than the R^2_β's defined using other methods.

Covariance structure misspecification greatly impacts the values of R^2_β using Kenward-Roger, containment and Satterthwaite degrees of freedom. Conversely, the values of R^2_β using the residual method are not greatly impacted by covariance structure misspecification. For the case of underspecified covariance structure, the true asymptotic $R^2_\beta(\widehat{\beta}, \widehat{\Sigma}^* | \nu)$ using Kenward-Roger is less than the true asymptotic $R^2_\beta(\widehat{\beta}, \widehat{\Sigma} | \nu)$ using Kenward-Roger. For the case of underspecified covariance structure, the true asymptotic $R^2_\beta(\widehat{\beta}, \widehat{\Sigma}^* | \nu)$ using residual methods is greater than the true asymptotic $R^2_\beta(\widehat{\beta}, \widehat{\Sigma} | \nu)$ using the residual method.

The impact of covariance structure misspecification on R^2_β has important implications. Often, the true model is not known in practice; therefore, it is difficult to know whether R^2_β is measuring the truth or is inflated or deflated as a result of covariance structure misspecification if Kenward-Roger, Satterthwaite, or containment methods are used.

Additionally, R^2 statistics are often provided as a measure of effect size for a fixed effect. If the values of R^2_β are greatly affected by the covariance structure used, then R^2_β is not an accurate measure of the fixed effect size. Since R^2_β using the residual method are not impacted by covariance structure misspecification, when using R^2_β to develop a model or as a measure of effect size, R^2_β should be defined using the residual method.

REFERENCES

Edwards, L.J., Muller, K.E., Wolfinger, R.D., Qaqish, B.F., and Schabenberger, O. (2008). An R^2 statistic for fixed effects in the linear mixed model. *Statistics in Medicine* **27**, 6137-6157.

Gurka, M. J., Edwards, L. J. and Muller, K. E. (2011), Avoiding bias in mixed model inference for fixed effects. Statistics in Medicine. doi: 10.1002/sim.4293

Kenward, M.G. and Roger, J.H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics* **53**, 983-997.

Kistner, E.O., and Muller, K.E. (2004). Exact distributions of intraclass correlation and Cronbach's alpha with Gaussian data and general covariance. *Psychometrika* **69**, 459-474.

TABLES

Table 3.1 Additional simulated covariance parameter values

	Covariance 3
$\boldsymbol{R}_i$	$\sigma_0^2 = 0.44$ $\sigma_1^2 = 2.66$
$\boldsymbol{D}$	$\begin{bmatrix} \sigma^2_{\text{Int}} & \rho \\ \rho & \sigma^2_{\text{Slope}} \end{bmatrix} = \begin{bmatrix} 3.89 & -0.16 \\ -0.16 & 0.03 \end{bmatrix}$

Table 3.2 Formulas to compute the denominator degrees of freedom using subject sample size for each simulation study

	Kenward-Roger	Satterthwaite	Containment	Residual
Model I Cov 1	$\nu = 3m - 1$	$\nu = 3m - 1$	$\nu = 3m - 1$	$\nu = 4m - 2$
Model I Cov 2	$\nu = m - 1$	$\nu = m - 1$	$\nu = m - 1$	$\nu = 4m - 2$
Model I Cov 3	NC*	NC	$\nu = m - 1$	$\nu = 4m - 2$
Model II Cov 1	$\nu = 2m - 3.8897$	$\nu = 1.5001m - 2.931$	$\nu = 3m - 1$	$\nu = 4m - 3$
Model II Cov 2	$\nu = 1.33 - 2.7042$	$\nu = m - 1.5033$	$\nu = 2m$	$\nu = 4m - 3$
Model II Cov 3	NC	NC	$\nu = 2m$	$\nu = 4m - 3$
Model III Cov 1	$\nu = 2.6471m - 4.8019$	NC	$\nu = 3m - 2$	$\nu = 4m - 4$
Model III Cov 2	$\nu = 1.1538 - 3.5643$	NC	$\nu = 2m$	$\nu = 4m - 4$
Model III Cov 3	NC	NC	$\nu = 2m$	$\nu = 4m - 4$

* Indicates that the denominator degrees of freedom are not constant across all replications with a positive definite Hessian matrix.

Table 3.3 True and simulated R_β^2 values for Model I with Covariance 1 under REML and ML estimation when using different denominator degrees of freedom methods

	Model I: Cov 1[1]			
	All other REML	All other ML	REML Residual	ML Residual
25 subjects	0.396	0.396	0.333	0.336
50 subjects	0.394	0.394	0.330	0.331
100 subjects	0.394	0.394	0.328	0.329
200 subjects	0.393	0.393	0.328	0.328
500 subjects	0.392	0.392	0.327	0.327
1000 subjects	0.392	0.392	0.326	0.326
2000 subjects	0.392	0.392	0.326	0.326
5000 subjects	0.392	0.392	0.326	0.326
10000 subjects	0.392	0.392	0.326	0.326
Asymptotic R_β^2	**0.392**	**0.392**	**0.326**	**0.326**

[1] Model I: Covariance 1 is the model which contains an intercept and a continuous time effect with a compound symmetric covariance structure.

Table 3.4 True and simulated R^2_β values for Model I with Covariance 2 under REML and ML estimation when using different denominator degrees of freedom methods

	Model I: Cov 2[2]			
	All other REML	All other ML	REML Residual	ML Residual
25 subjects	0.778	0.776	0.470	0.476
50 subjects	0.783	0.782	0.476	0.480
100 subjects	0.787	0.787	0.482	0.484
200 subjects	0.789	0.789	0.484	0.485
500 subjects	0.790	0.790	0.485	0.485
1000 subjects	0.790	0.790	0.485	0.485
2000 subjects	0.790	0.790	0.484	0.485
5000 subjects	0.790	0.790	0.484	0.484
10000 subjects	0.790	0.790	0.484	0.484
Asymptotic R^2_β	**0.790**	**0.790**	**0.484**	**0.484**

Table 3.5 True and simulated R^2_β values for Model I with Covariance 3 under REML and ML estimation when using different denominator degrees of freedom methods

	Model I: Cov 3[3]			
	REML Contain	ML Contain	REML Residual	ML Residual
25 subjects	0.862	0.867	0.611	0.621
50 subjects	0.863	0.866	0.613	0.618
100 subjects	0.864	0.865	0.614	0.616
200 subjects	0.863	0.864	0.612	0.614
500 subjects	0.862	0.863	0.610	0.611
1000 subjects	0.862	0.862	0.610	0.610
2000 subjects	0.862	0.862	0.609	0.610
5000 subjects	-	-	0.609	0.609
10000 subjects	-	-	0.609	0.609
Asymptotic R^2_β	**0.860**	**0.860**	**0.609**	**0.609**

[2] Model I: Covariance 2 refers to the mean model with an intercept and continous time effect with a random intercept and a random slope and independent errors.

[3] Model I: Covariance 3 refers to a mean model with an intercept and a continuous time effect and a random intercept and a random slope with heterogeneous errors by group.

Table 3.6 Average denominator degrees of freedom for Model II: Covariance 1 under REML estimation from the simulation study results

	Kenward-Roger	**Satterthwaite**	**Containment**
25 subjects	46.09	34.52	74
50 subjects	96.11	72.07	149
100 subjects	196.12	147.10	299
200 subjects	396.13	297.11	599
500 subjects	996.13	747.12	1499
1000 subjects	1996.13	1497.12	2999
2000 subjects	3996.13	2997.12	5999
5000 subjects	9996.13	7497.12	14999
10000 subjects	19996.13	14997.12	39997

Table 3.7 True and average simulated R_β^2 values for Model II: Covariance 1 under REML estimation for the denominator degrees of freedom methods

	Kenward-Roger	**Satterthwaite**	**Containment**	**Residual**
25 subjects	0.558	0.629	0.446	0.382
50 subjects	0.546	0.616	0.439	0.373
100 subjects	0.540	0.611	0.437	0.369
200 subjects	0.538	0.608	0.436	0.367
500 subjects	0.537	0.607	0.435	0.366
1000 subjects	0.536	0.606	0.435	0.366
2000 subjects	0.536	0.606	0.436	0.366
5000 subjects	0.535	0.606	0.434	0.366
10000 subjects	0.535	0.606	-	0.366
Asymptotic R_β^2	**0.535**	**0.606**	**0.434**	**0.366**

Table 3.8 True and average simulated R_β^2 values for Model II: Covariance 1 under ML estimation for the denominator degrees of freedom methods

	Kenward-Roger	**Satterthwaite**	**Containment**	**Residual**
25 subjects	0.548	0.619	0.452	0.388
50 subjects	0.541	0.612	0.442	0.376
100 subjects	0.538	0.609	0.438	0.371
200 subjects	0.537	0.607	0.436	0.368
500 subjects	0.536	0.607	0.435	0.367
1000 subjects	0.536	0.606	0.435	0.366
2000 subjects	0.536	0.606	0.436	0.366
5000 subjects	0.535	0.606	0.434	0.366
10000 subjects	0.535	0.606	-	0.366
Asymptotic R_β^2	**0.535**	**0.606**	**0.434**	**0.366**

Table 3.9 Average denominator degrees of freedom for Model II: Covariance 2 under REML estimation from the simulation study results

	Kenward-Roger	**Satterthwaite**	**Containment**
25 subjects	30.62	23.49	50
50 subjects	63.96	48.50	100
100 subjects	130.63	98.50	200
200 subjects	263.97	198.50	400
500 subjects	663.97	498.50	1000
1000 subjects	1330.63	998.50	2000
2000 subjects	2663.97	1998.50	4000
5000 subjects	6663.97	4998.50	10000
10000 subjects	13330.63	9998.50	20000

Table 3.10 True and average simulated R_β^2 values for Model II: Covariance 2 under REML estimation by the various denominator degrees of freedom methods

	Kenward-Roger	**Satterthwaite**	**Containment**	**Residual**
25 subjects	0.740	0.792	0.644	0.486
50 subjects	0.744	0.794	0.654	0.491
100 subjects	0.746	0.797	0.660	0.495
200 subjects	0.747	0.797	0.662	0.496
500 subjects	0.747	0.797	0.662	0.496
1000 subjects	0.747	0.797	0.662	0.496
2000 subjects	0.747	0.797	0.663	0.496
5000 subjects	0.747	0.797	-	0.495
10000 subjects	0.746	0.797	-	0.495
Asymptotic R_β^2	**0.746**	**0.797**	**0.660**	**0.500**

Table 3.11 True and average simulated R_β^2 values for Model II: Covariance 2 under ML estimation by the various denominator degrees of freedom methods

	Kenward-Roger	**Satterthwaite**	**Containment**	**Residual**
25 subjects	0.740	0.786	0.650	0.493
50 subjects	0.741	0.792	0.657	0.495
100 subjects	0.745	0.796	0.662	0.497
200 subjects	0.747	0.797	0.663	0.497
500 subjects	0.747	0.797	0.663	0.497
1000 subjects	0.747	0.797	0.663	0.496
2000 subjects	0.747	0.797	0.663	0.496
5000 subjects	0.747	0.797	-	0.496
10000 subjects	0.746	0.797	-	0.495
Asymptotic R_β^2	**0.746**	**0.797**	**0.660**	**0.500**

Table 3.12 True and average simulated R_β^2 values for Model II: Covariance 3 for the various denominator degrees of freedom methods

	REML Contain	**ML Contain**	**REML Residual**	**ML Residual**
25 subjects	0.772	0.780	0.639	0.649
50 subjects	0.774	0.778	0.637	0.642
100 subjects	0.776	0.779	0.637	0.640
200 subjects	0.775	0.777	0.635	0.637
500 subjects	0.775	0.776	0.634	0.634
1000 subjects	0.775	0.775	0.633	0.633
2000 subjects	0.775	0.775	0.633	0.633
5000 subjects	-	-	0.632	0.633
10000 subjects	-	-	0.632	0.632
Asymptotic True R_β^2	**0.772**	**0.772**	**0.632**	**0.632**

Table 3.13 Average denominator degrees of freedom for Model III: Covariance 1 under REML estimation from the simulation study results

	Kenward-Roger	**Containment**
25 subjects	61.35	73
50 subjects	127.55	148
100 subjects	259.92	298
200 subjects	524.63	598
500 subjects	1318.75	1498
1000 subjects	2642.28	2998
2000 subjects	5289.34	5998
5000 subjects	13230.52	14998
10000 subjects	26465.81	-

Table 3.14 True and average simulated R_{β}^{2} values for Model III: Covariance 1 for the various denominator degrees of freedom methods

	REML KR	**ML KR**	**REML Contain**	**ML Contain**	**REML Residual**	**ML Residual**
25 subjects	0.662	0.658	0.626	0.636	0.561	0.571
50 subjects	0.648	0.647	0.616	0.621	0.548	0.553
100 subjects	0.644	0.643	0.613	0.615	0.544	0.546
200 subjects	0.642	0.641	0.611	0.613	0.542	0.543
500 subjects	0.640	0.639	0.610	0.610	0.540	0.541
1000 subjects	0.639	0.639	0.610	0.610	0.540	0.540
2000 subjects	0.639	0.639	0.610	0.610	0.539	0.539
5000 subjects	0.639	0.639	0.609	0.610	0.539	0.539
10000 subjects	0.639	0.639	-	-	0.539	0.539
Asymptotic R_{β}^{2}	**0.639**	**0.639**	**0.609**	**0.609**	**0.539**	**0.539**

Table 3.15 Average denominator degrees of freedom for Model III: Covariance 2 under REML estimation from the simulation study results

	Kenward-Roger	**Containment**
25 subjects	25.29	50
50 subjects	54.13	100
100 subjects	111.82	200
200 subjects	227.20	400
500 subjects	573.35	1000
1000 subjects	1150.28	2000
2000 subjects	2304.12	4000
5000 subjects	5765.66	10000
10000 subjects	11534.89	20000

Table 3.16 True and average simulated R_β^2 values for Model III: Covariance 2 for the different denominator degrees of freedom methods

	REML KR	**ML KR**	**REML Contain**	**ML Contain**	**REML Residual**	**ML Residual**
25 subjects	0.866	0.863	0.774	0.783	0.642	0.655
50 subjects	0.863	0.862	0.778	0.783	0.642	0.649
100 subjects	0.863	0.863	0.781	0.784	0.644	0.648
200 subjects	0.864	0.864	0.784	0.785	0.646	0.648
500 subjects	0.863	0.863	0.784	0.785	0.646	0.647
1000 subjects	0.863	0.863	0.784	0.784	0.645	0.646
2000 subjects	0.863	0.863	0.784	0.784	0.645	0.645
5000 subjects	0.863	0.863	-	-	0.645	0.645
10000 subjects	0.863	0.863	-	-	0.645	0.645
Asymptotic R_β^2	**0.863**	**0.863**	**0.784**	**0.784**	**0.645**	**0.645**

Table 3.17 True and average simulated R_β^2 values for Model III: Covariance 3 for the various denominator degrees of freedom methods

	REML Contain	**ML Contain**	**REML Residual**	**ML Residual**
25 subjects	0.864	0.872	0.770	0.782
50 subjects	0.866	0.870	0.768	0.775
100 subjects	0.867	0.869	0.768	0.771
200 subjects	0.867	0.869	0.767	0.769
500 subjects	0.867	0.867	0.765	0.766
1000 subjects	0.867	0.867	0.765	0.765
2000 subjects	0.867	0.867	0.765	0.765
5000 subjects	-	-	0.764	0.765
10000 subjects	-	-	0.764	0.764
Asymptotic True R_β^2	**0.868**	**0.868**	**0.764**	**0.764**

Table 3.18 Average R_β^2 for the true simulated models and misspecified models with their corresponding asymptotic 'true' R_β^2 using Kenward-Roger method

	Model I: Cov 2		**Model II: Cov 2**		**Model III: Cov 2**	
	True	Cov 1	True	Cov 1	True	Cov 1
25 subjects	0.7784	0.6060	0.7401	0.7145	0.8656	0.7842
50 subjects	0.7832	0.6030	0.7436	0.7081	0.8633	0.7756
100 subjects	0.7874	0.6027	0.7463	0.7052	0.8634	0.7719
200 subjects	0.7893	0.6022	0.7471	0.7031	0.8637	0.7703
500 subjects	0.7899	0.6022	0.7472	0.7022	0.8634	0.7694
1000 subjects	0.7899	0.6022	0.7467	0.7017	0.8631	0.7691
Theoretical Asymptotic $R_\beta^2(\nu)$:	**0.7897**	**0.6020**	**0.7541**	**0.7089**	**0.8718**	**0.7788**

Table 3.19 Average R_β^2 for the true simulated models and misspecified models with their corresponding asymptotic 'true' R_β^2 using the Residual method

	Model I: Cov 2		**Model II: Cov 2**		**Model III: Cov 2**	
	True	Cov 1	True	Cov 1	True	Cov 1
25 subjects	0.4695	0.5385	0.4855	0.5494	0.6423	0.7030
50 subjects	0.4760	0.5340	0.4910	0.5451	0.6420	0.6942
100 subjects	0.4817	0.5329	0.4949	0.5432	0.6441	0.6905
200 subjects	0.4841	0.5320	0.4961	0.5414	0.6456	0.6890
500 subjects	0.4847	0.5319	0.4963	0.5408	0.6455	0.6881
1000 subjects	0.4847	0.5318	0.4956	0.5404	0.6451	0.6878
Theoretical Asymptotic $R_\beta^2(\nu)$:	**0.4842**	**0.5315**	**0.5055**	**0.5491**	**0.6624**	**0.6999**

Table 3.20 Mean (Standard Deviation) of the REML Kenward-Roger denominator degrees of freedom and F statistic for true mean models I

	REML Kenward-Roger Denomatinator Degrees of Freedom			**REML Kenward-Roger *F* statistic**		
	Cov 1	**Cov 2**	**Cov 3**	**Cov 1**	**Cov 2**	**Cov 3**
True Model I: Covariance 1						
25	74 (0)	24 (0)	22.8 (1.87)	50.6 (16.8)	42.1 (14.81)	41.0 (14.57)
50	149 (0)	49 (0)	47.4 (2.39)	99.1 (23.36)	87.3 (21.8)	86.0 (21.68)
100	299 (0)	99 (0)	97.2 (2.5)	196.1 (32.34)	178.4 (31.2)	177.3 (31.33)
200	599 (0)	199 (0)	197.0 (2.87)	390.6 (45.91)	366.7 (46.57)	365.5 (46.5)
500	1499 (0)	499 (0)	496.9 (2.99)	970.1 (71.45)	931.8 (73.55)	930.7 (73.7)
1000	2999 (0)	999 (0)	996.8 (3.11)	1938.5 (101.16)	1883.6 (105.81)	1882.7 (105.69)
True Model I: Covariance 2						
25	74 (0)	24 (0)	22.9 (1.93)	118.7 (31.48)	90.5 (27.7)	88.2 (27.11)
50	149 (0)	49 (0)	47.5 (2.37)	231.3 (44.05)	184.4 (42.54)	182.3 (42.06)
100	299 (0)	99 (0)	97.4 (2.6)	458.3 (60.69)	374.7 (62.14)	372.8 (61.83)
200	599 (0)	199 (0)	197.3 (2.56)	911.5 (85.9)	754.0 (91.11)	752.1 (90.8)
500	1499 (0)	499 (0)	497.3 (2.65)	2274.3 (136.3)	1884.9 (148.58)	1883.2 (148.58)
1000	2999 (0)	999 (0)	997.2 (2.57)	4544.58 (190.23)	3765.5 (208.15)	3763.8 (208.09)
True Model I: Covariance 3						
25	74 (0)	24 (0)	18.6 (3.03)	137.5 (38.43)	102.5 (33.84)	154.2 (51.06)
50	149 (0)	49 (0)	38.1 (4.55)	261.3 (51.14)	205.7 (48.16)	313.8 (77.1)
100	299 (0)	99 (0)	77.3 (6.38)	513.2 (70.13)	417.2 (70.04)	633.4 (113.23)
200	599 (0)	199 (0)	156 (9.24)	1022.6 (100.3)	846.7 (105.55)	1262 (162.26)
500	1499 (0)	499 (0)	392.5 (14.54)	2543.2 (155.41)	2129 (173.89)	3132.4 (249.6)
1000	2999 (0)	999 (0)	786.6 (20.43)	5079.0 (221.75)	4255.8 (255.54)	6247.9 (352.82)

Table 3.21 Mean (Standard Deviation) of the REML Kenward-Roger denominator degrees of freedom and F statistic for true mean models II

	REML Kenward-Roger Denomatinator Degrees of Freedom			**REML Kenward-Roger *F* statistic**		
	Cov 1	**Cov 2**	**Cov 3**	**Cov 1**	**Cov 2**	**Cov 3**
True Model II: Covariance 1						
25	46.1 (0)	30.6 (0)	29.4 (1.61)	30.6 (9.3)	25.9 (8.18)	25.3 (8.09)
50	96.1 (0)	64 (0)	62.4 (1.87)	59.1 (12.6)	52.7 (11.94)	52 (11.81)
100	196.1 (0)	130.6 (0)	128.9 (1.88)	116.7 (17.66)	108 (17.28)	107.4 (17.17)
200	396.1 (0)	264 (0)	262.1 (2.01)	231.9 (24.73)	219.6 (24.92)	218.9 (24.75)
500	996.1 (0)	664 (0)	662.1 (1.99)	578 (38.42)	558.6 (39.63)	558 (39.48)
1000	1996.1 (0)	1330.6 (0)	1328.7 (2.02)	1153.2 (54.76)	1124.5 (56.71)	1124 (56.7)
True Model II: Covariance 2						
25	46.1 (0)	30.6 (0)	29.3 (1.63)	60.4 (15.66)	46.5 (13.81)	45.6 (13.54)
50	96.1 (0)	64 (0)	62.2 (1.92)	119.3 (21.77)	95.9 (20.52)	95 (20.29)
100	196.1 (0)	130.6 (0)	128.2 (2.03)	237.4 (30.77)	195.7 (30.86)	194.9 (30.76)
200	396.1 (0)	264 (0)	260.4 (2.27)	471.8 (42.75)	393.8 (45.72)	392.9 (45.75)
500	996.1 (0)	664 (0)	657.3 (2.75)	1177.2 (67.87)	985.3 (74.52)	984.5 (74.56)
1000	1996.1 (0)	1330.6 (0)	1318.8 (3.34)	2351 (96.84)	1965.1 (106.04)	1964.3 (106.09)
True Model II: Covariance 3						
25	46.1 (0)	30.6 (0)	25.2 (2.98)	70.4 (19.09)	53.3 (16.4)	85.3 (27.07)
50	96.1 (0)	64 (0)	52.3 (4.32)	134.3 (25.86)	106.7 (23.64)	172.5 (40.62)
100	196.1 (0)	130.6 (0)	107.6 (6.12)	265.8 (35.56)	218.6 (35.51)	348.9 (60.27)
200	396.1 (0)	264 (0)	218.1 (8.61)	528.9 (50.05)	442.9 (52.12)	694 (84.52)
500	996.1 (0)	664 (0)	549.5 (13.53)	1318.5 (79.69)	1115.2 (87.27)	1727.1 (132.23)
1000	1996.1 (0)	1330.6 (0)	1102.8 (19.45)	2633.8 (110.61)	2230.8 (125.94)	3446.2 (188.87)

Table 3.22 Mean (Standard Deviation) of the REML Kenward-Roger denominator degrees of freedom and F statistic for true mean models III

	REML Kenward-Roger Denomatinator Degrees of Freedom			**REML Kenward-Roger *F* statistic**		
	Cov 1	**Cov 2**	**Cov 3**	**Cov 1**	**Cov 2**	**Cov 3**
True Model III: Covariance 1						
25	61.3 (0)	25.3 (0)	23 (1.8)	41.6 (9.94)	35.7 (9.15)	35.7 (9.38)
50	127.6 (0)	54.1 (0)	49.4 (1.99)	79.8 (13.14)	71.9 (12.91)	71.9 (13.14)
100	259.9 (0)	111.8 (0)	102.3 (2.2)	158.1 (18.78)	147.5 (18.95)	147.5 (19.29)
200	524.6 (0)	227.2 (0)	208 (2.62)	314.3 (26.16)	298.8 (27.62)	299 (27.85)
500	1318.8 (0)	573.4 (0)	524.6 (3.35)	781.5 (40.58)	757.1 (42.89)	757.4 (43.3)
1000	2642.3 (0)	1150.3 (0)	1051.3 (4.41)	1561.6 (57.38)	1527.4 (61.68)	1527.6 (62.39)
True Model III: Covariance 2						
25	61.3 (0)	25.3 (0)	23.2 (1.63)	77.1 (16.92)	57.7 (15.53)	57.9 (15.96)
50	127.6 (0)	54.1 (0)	50 (1.97)	149.6 (23.09)	117.7 (22.84)	117.9 (23.44)
100	259.9 (0)	111.8 (0)	103.8 (2.38)	295.8 (32.48)	239.9 (34.86)	240.3 (35.24)
200	524.6 (0)	227.2 (0)	211.6 (3.08)	589.1 (45.82)	484.5 (51.66)	485 (52.18)
500	1318.8 (0)	573.4 (0)	535.5 (4.52)	1469 (71.1)	1213.1 (83.11)	1213.7 (83.7)
1000	2642.3 (0)	1150.3 (0)	1075.6 (5.98)	2935.7 (100.51)	2422.6 (116.92)	2423.4 (117.73)
True Model III: Covariance 3						
25	61.3 (0)	25.3 (0)	23.2 (1.63)	88.9 (20.7)	64.8 (18.34)	57.9 (15.96)
50	127.6 (0)	54.1 (0)	50 (1.97)	168.3 (27.12)	130.6 (26.44)	117.9 (23.44)
100	259.9 (0)	111.8 (0)	103.8 (2.38)	332.5 (37.03)	266.9 (39.39)	240.3 (35.34)
200	524.6 (0)	227.2 (0)	211.6 (3.08)	661.8 (53.09)	543.4 (59.36)	485 (52.18)
500	1318.8 (0)	573.4 (0)	535.5 (4.52)	1647.4 (83.59)	1367.8 (99.31)	1213.7 (83.69)
1000	2642.3 (0)	1150.3 (0)	1075.6 (5.98)	3291.7 (118.12)	2738.4 (142.44)	2423.4 (117.73)

Figure 3.1 Plot of Kenward-Roger, Satterthwaite, containment and residual denominator degrees of freedom methods by subject sample size with prediction lines for each method for Model II: Covariance 1

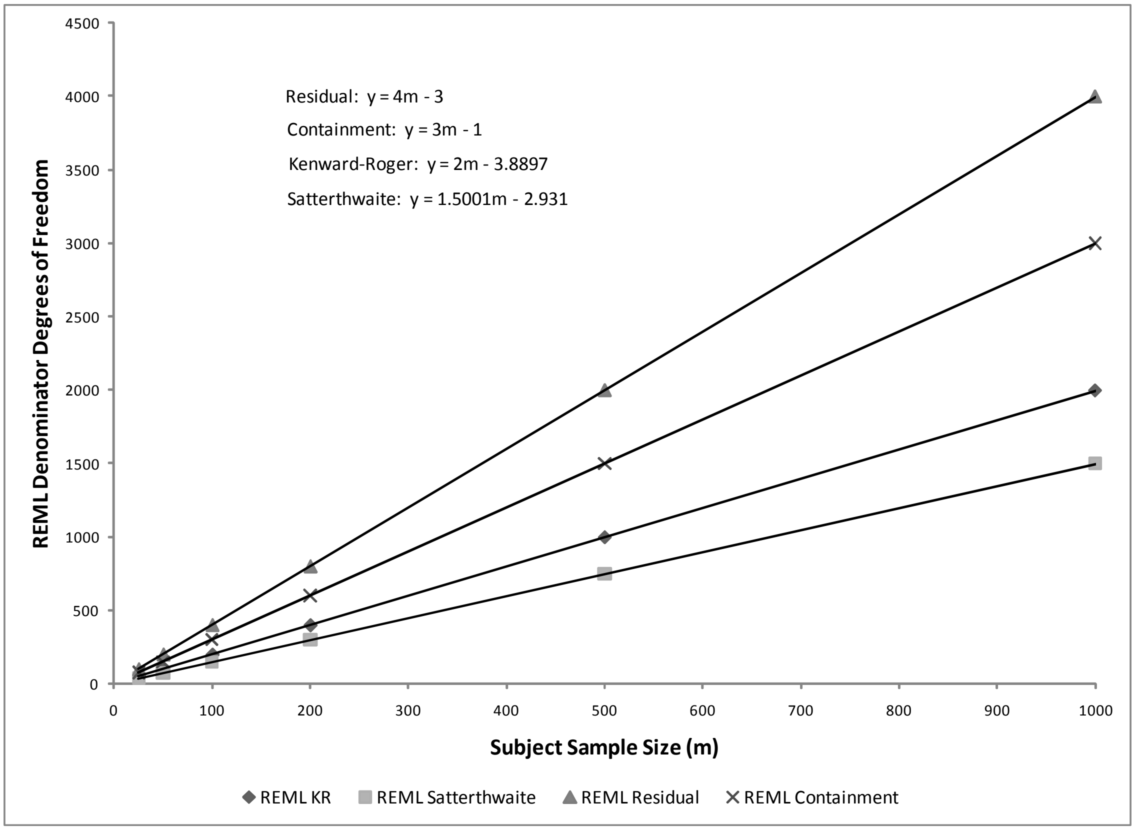


Figure 3.2. Plot of Kenward-Roger, Satterthwaite, containment and residual denominator degrees of freedom methods by subject sample size with prediction lines for each method for Model II: Covariance 2

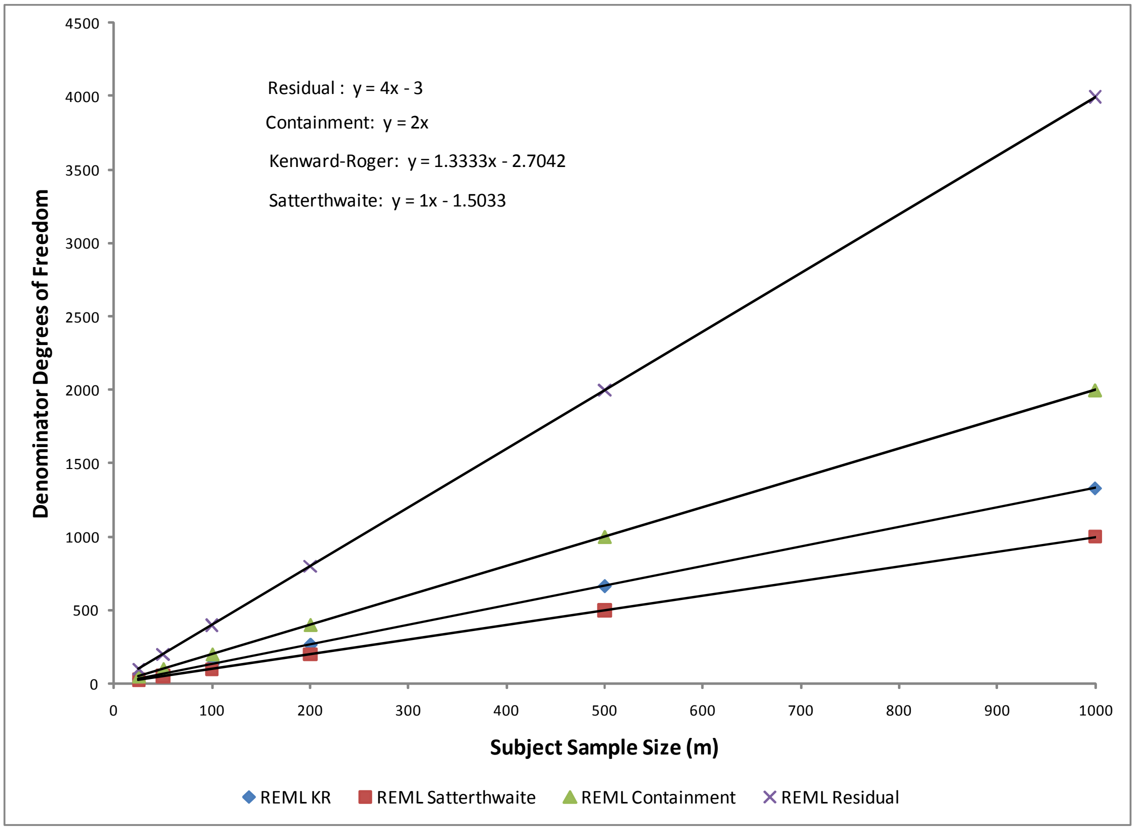


Figure 3.3. Plot of Kenward-Roger, containment and residual denominator degrees of freedom methods by subject sample size with prediction lines for each method for Model III: Covariance 1

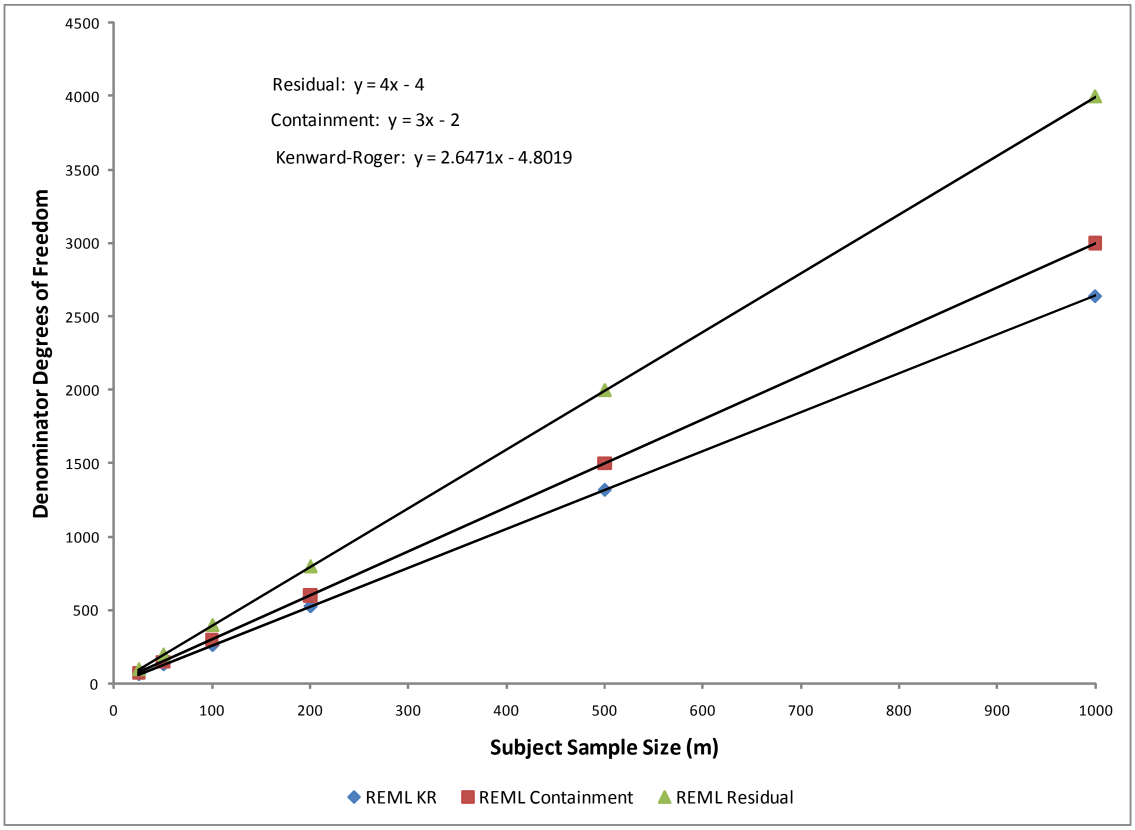


Figure 3.4. Plot of Kenward-Roger, containment and residual denominator degrees of freedom methods by subject sample size with prediction lines for each method for Model III: Covariance 2

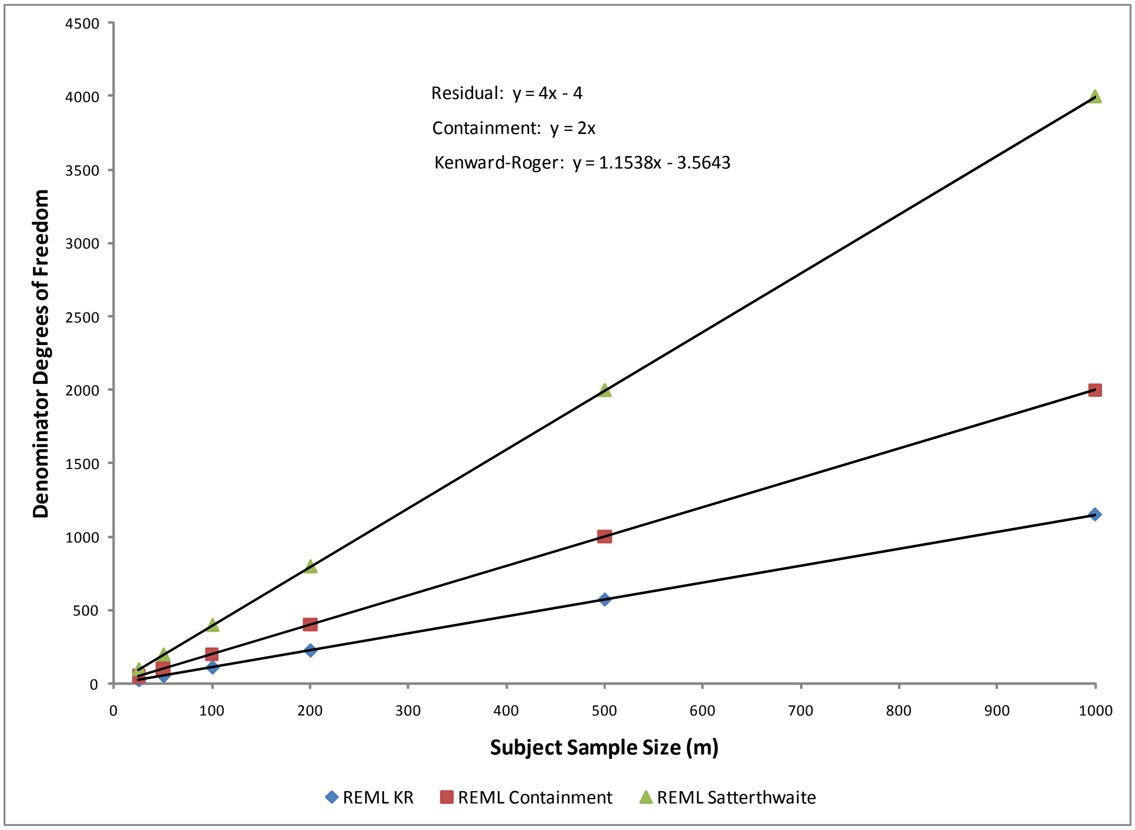


Figure 3.5 Average R_β^2 for Simulated Model - Model I: Covariance 1

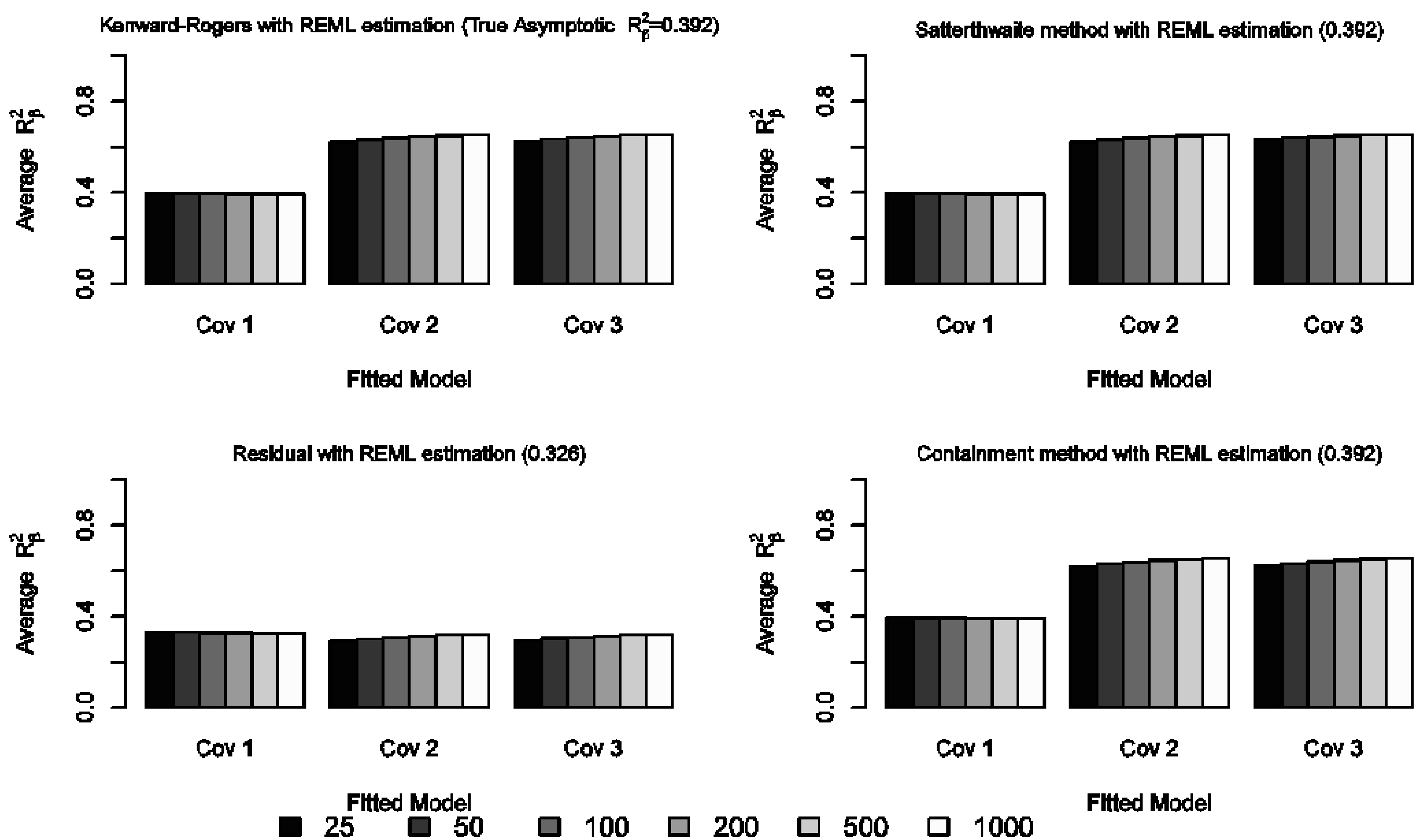


Figure 3.6 Average R_β^2 for Simulated Model - Model I: Covariance 2

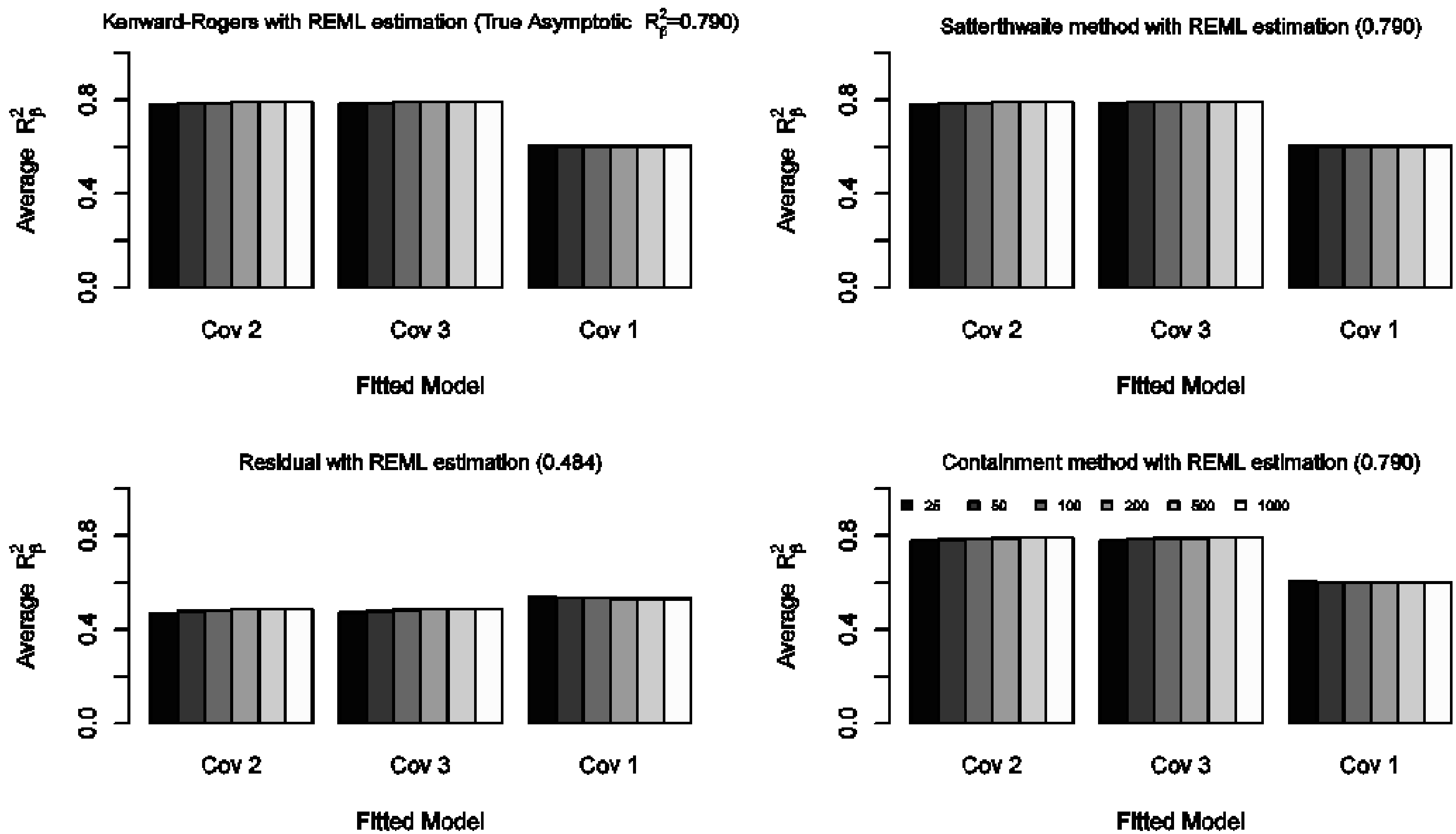


Figure 3.7 Average R_β^2 for Simulated Model - Model I: Covariance 3

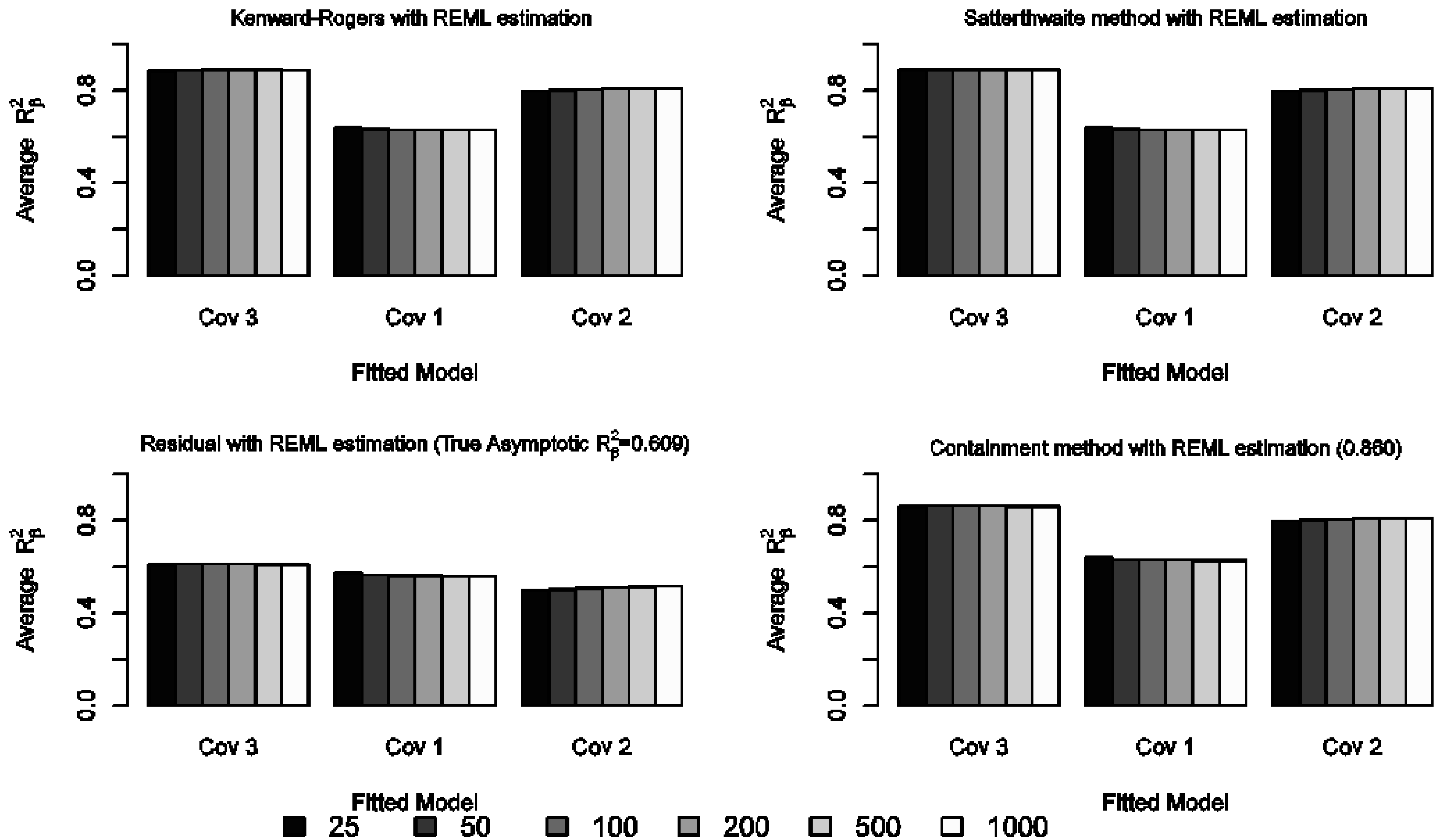
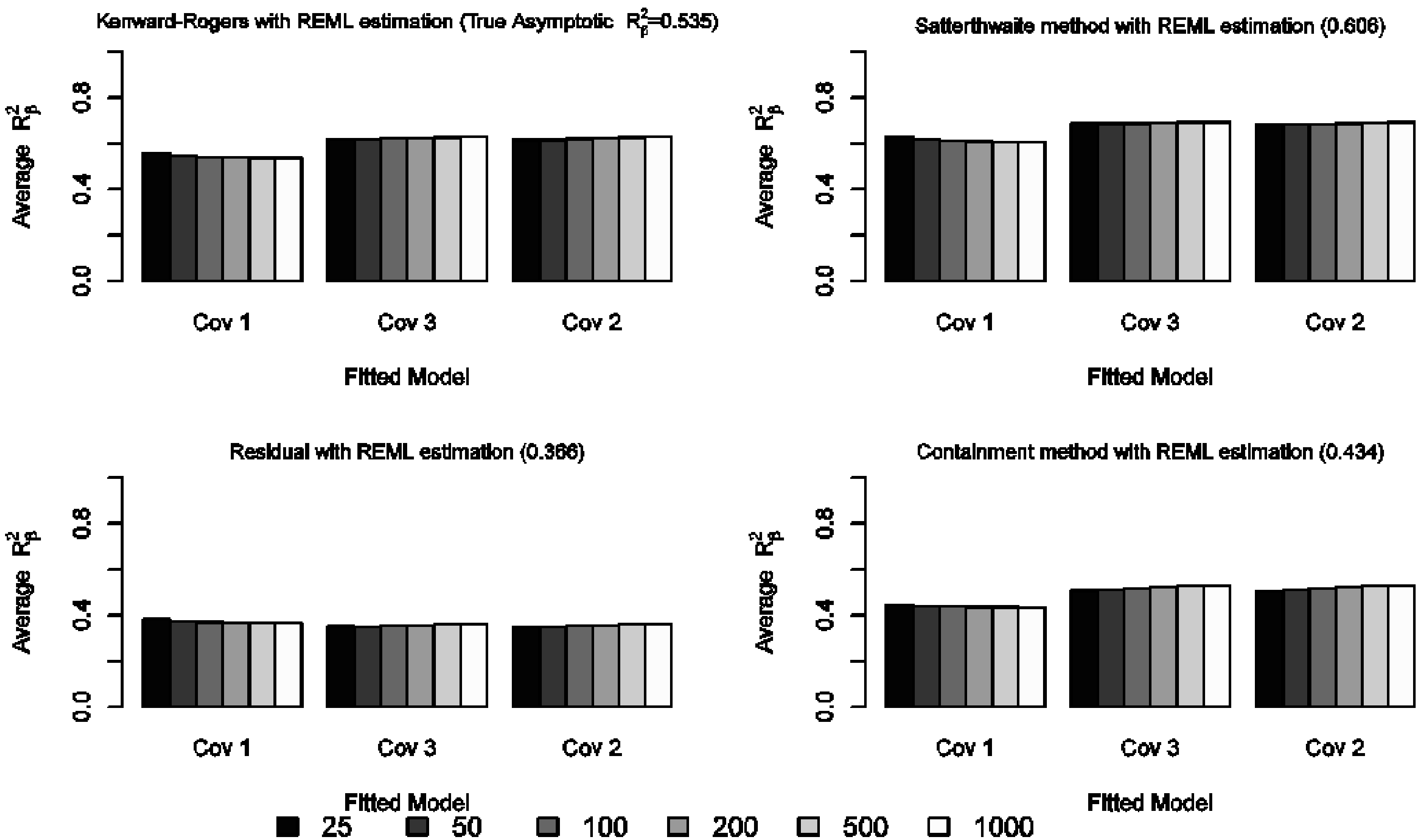


Figure 3.8 Average R_β^2 for Simulated Model - Model II: Covariance 1



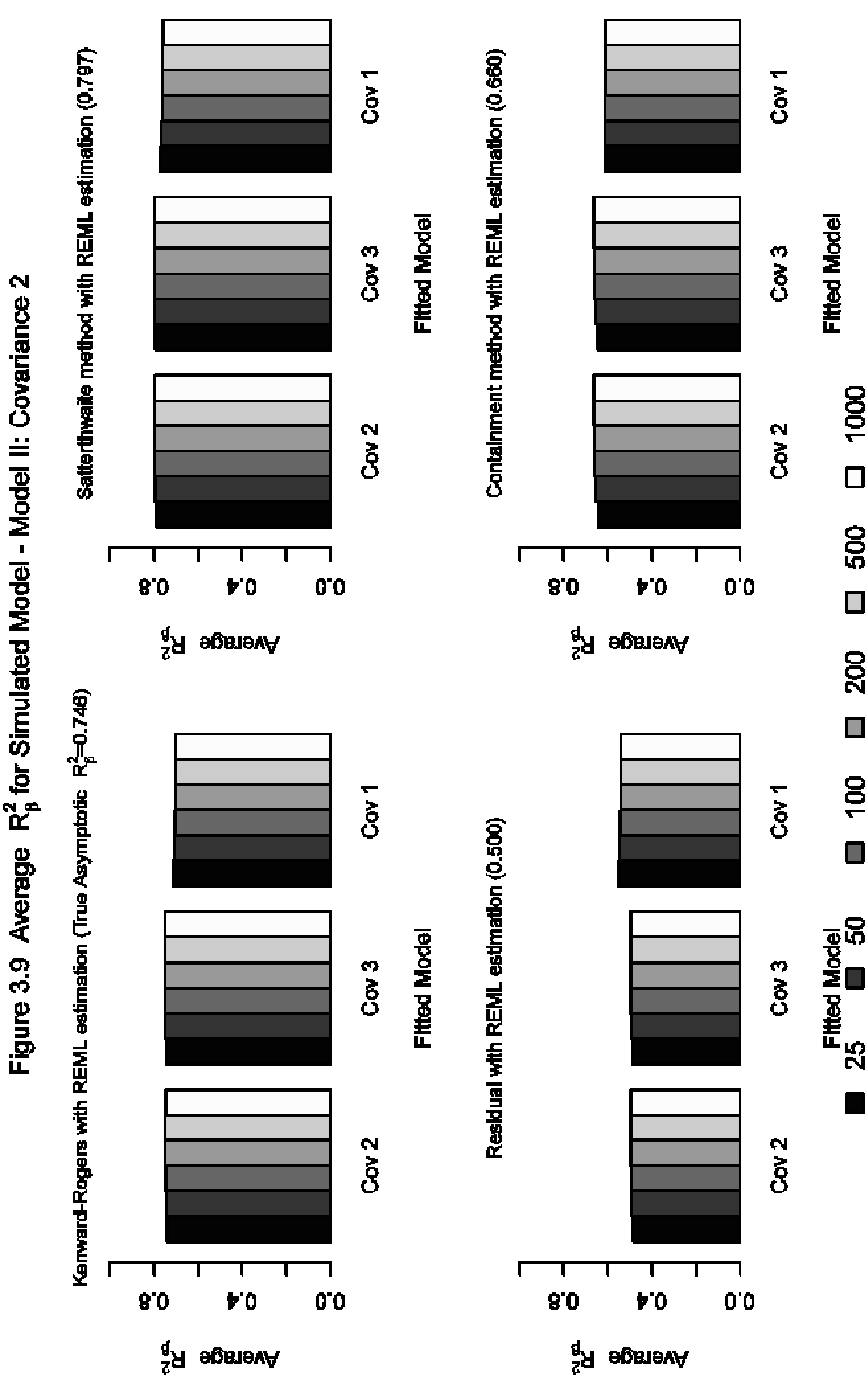


Figure 3.9 Average R^2_β for Simulated Model - Model II: Covariance 2

Figure 3.10 Average R_β^2 for Simulated Model - Model II: Covariance 3

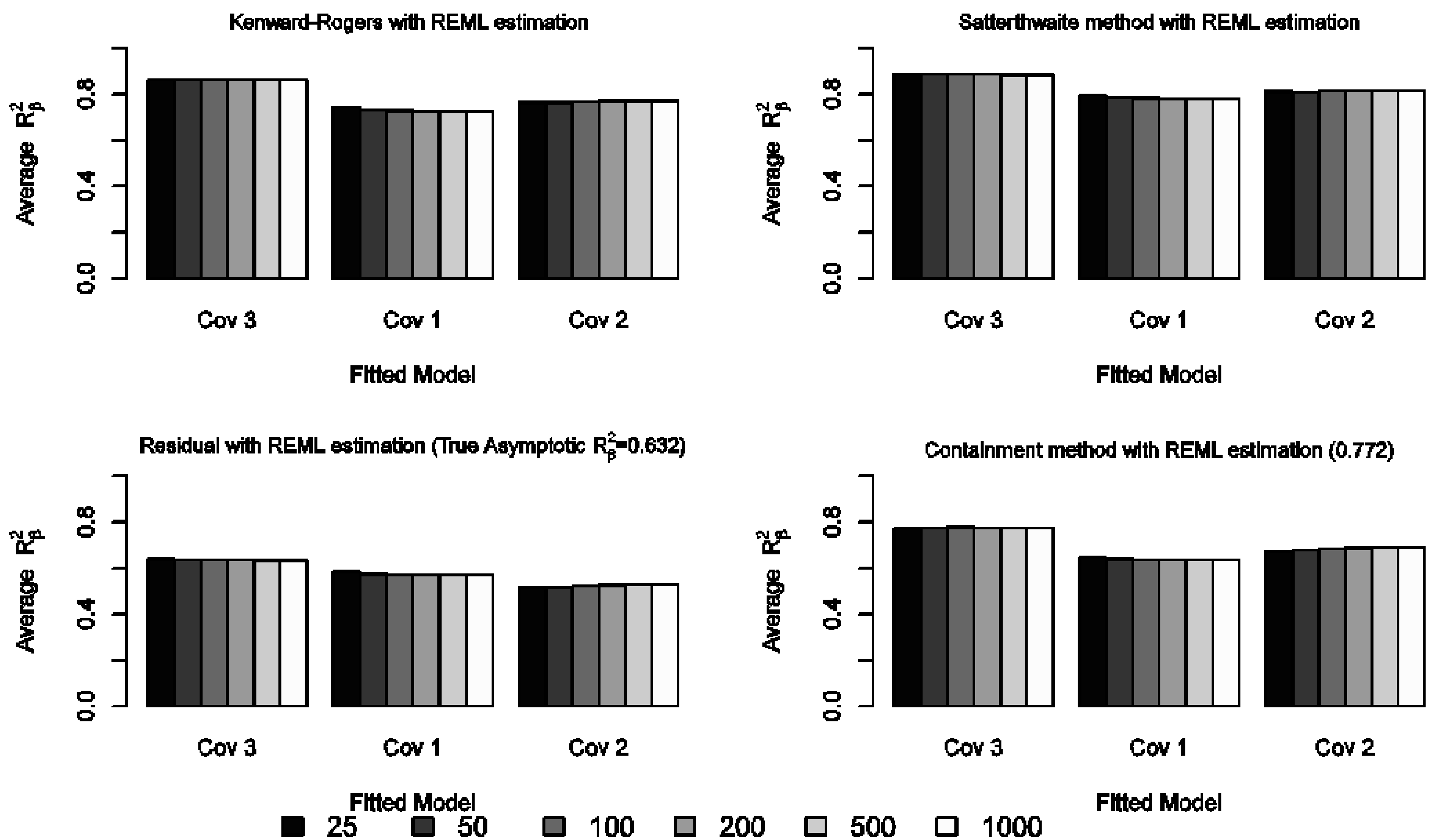


Figure 3.11 Average R_β^2 for Simulated Model - Model III: Covariance 1

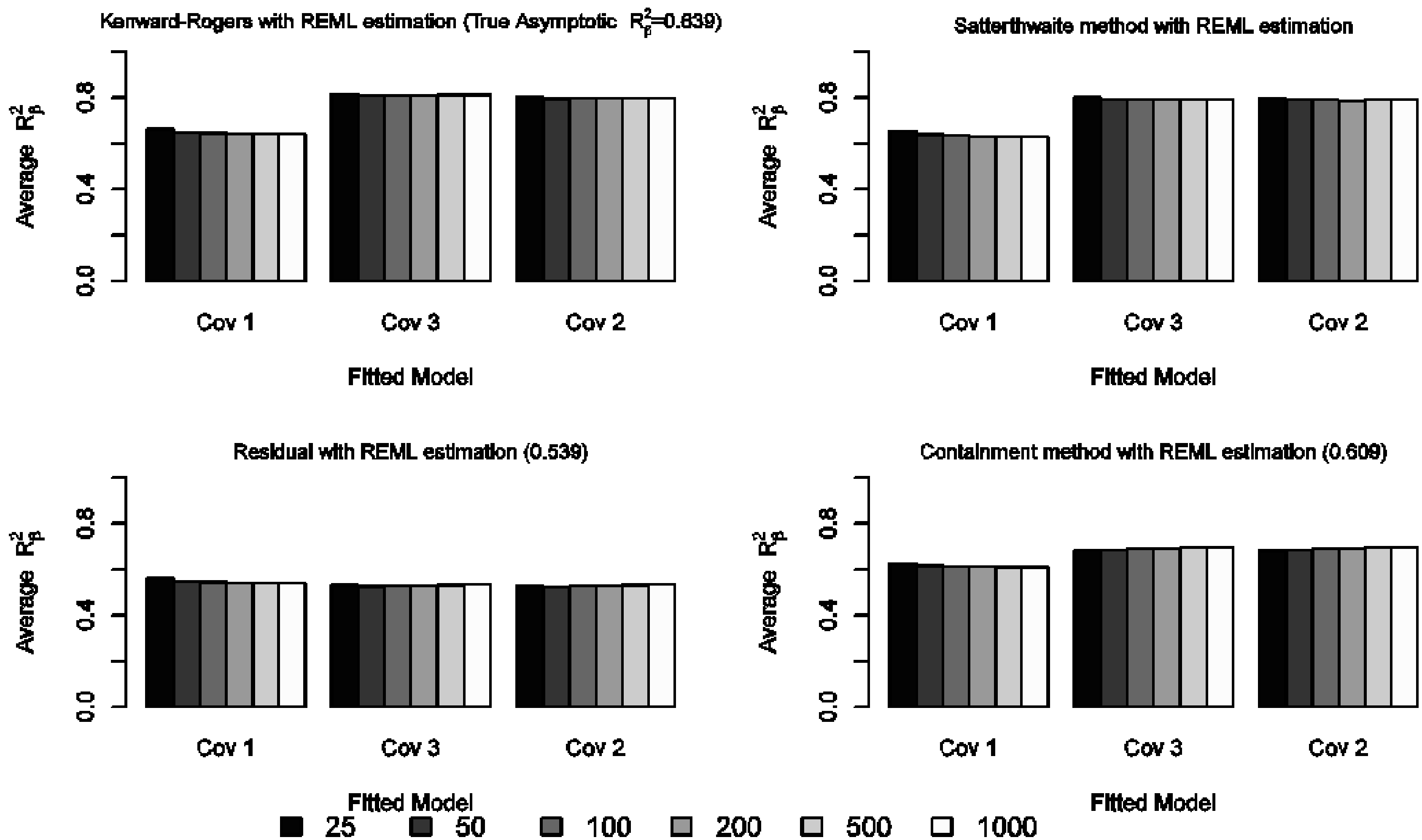


Figure 3.12 Average R_β^2 for Simulated Model - Model III: Covariance 2

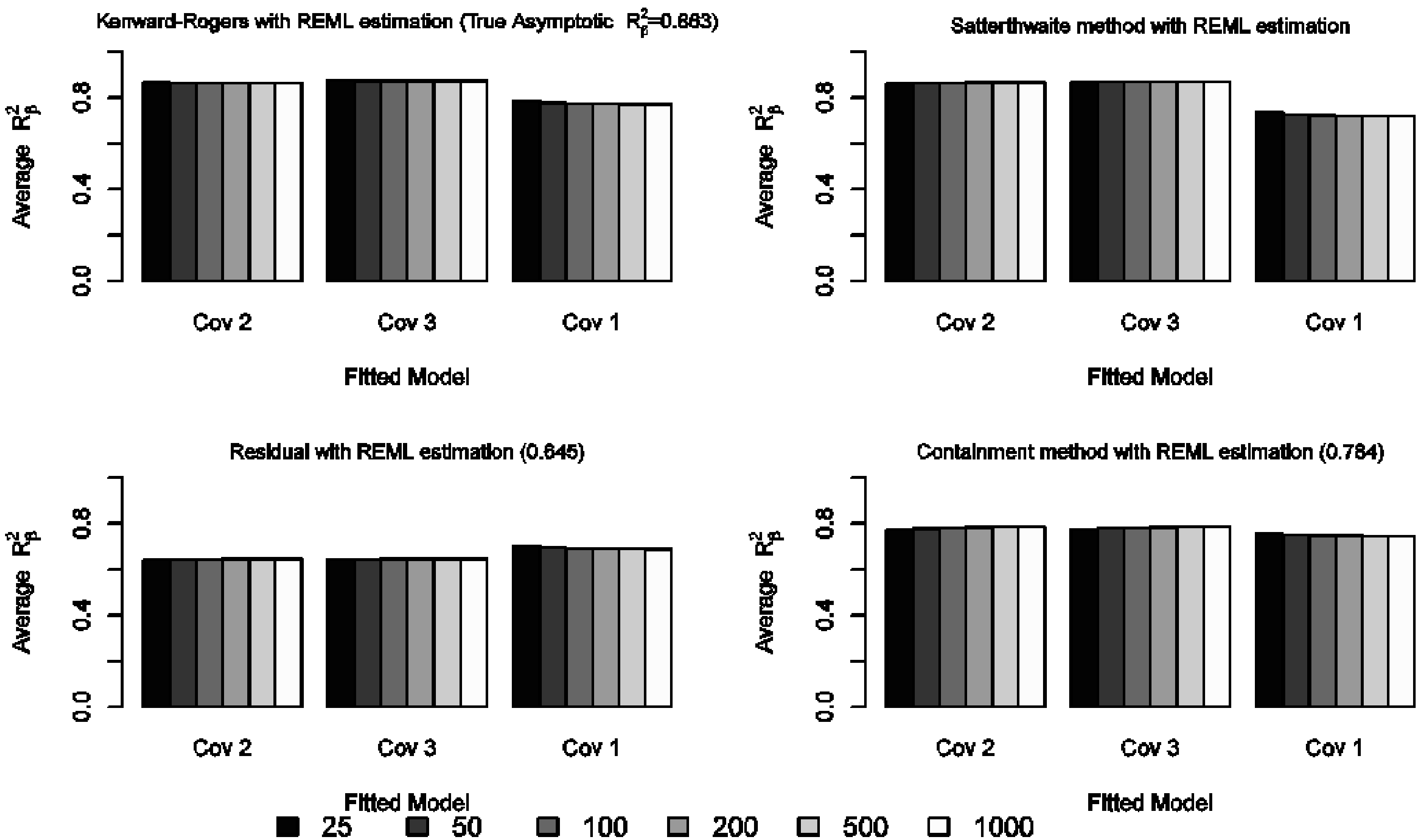
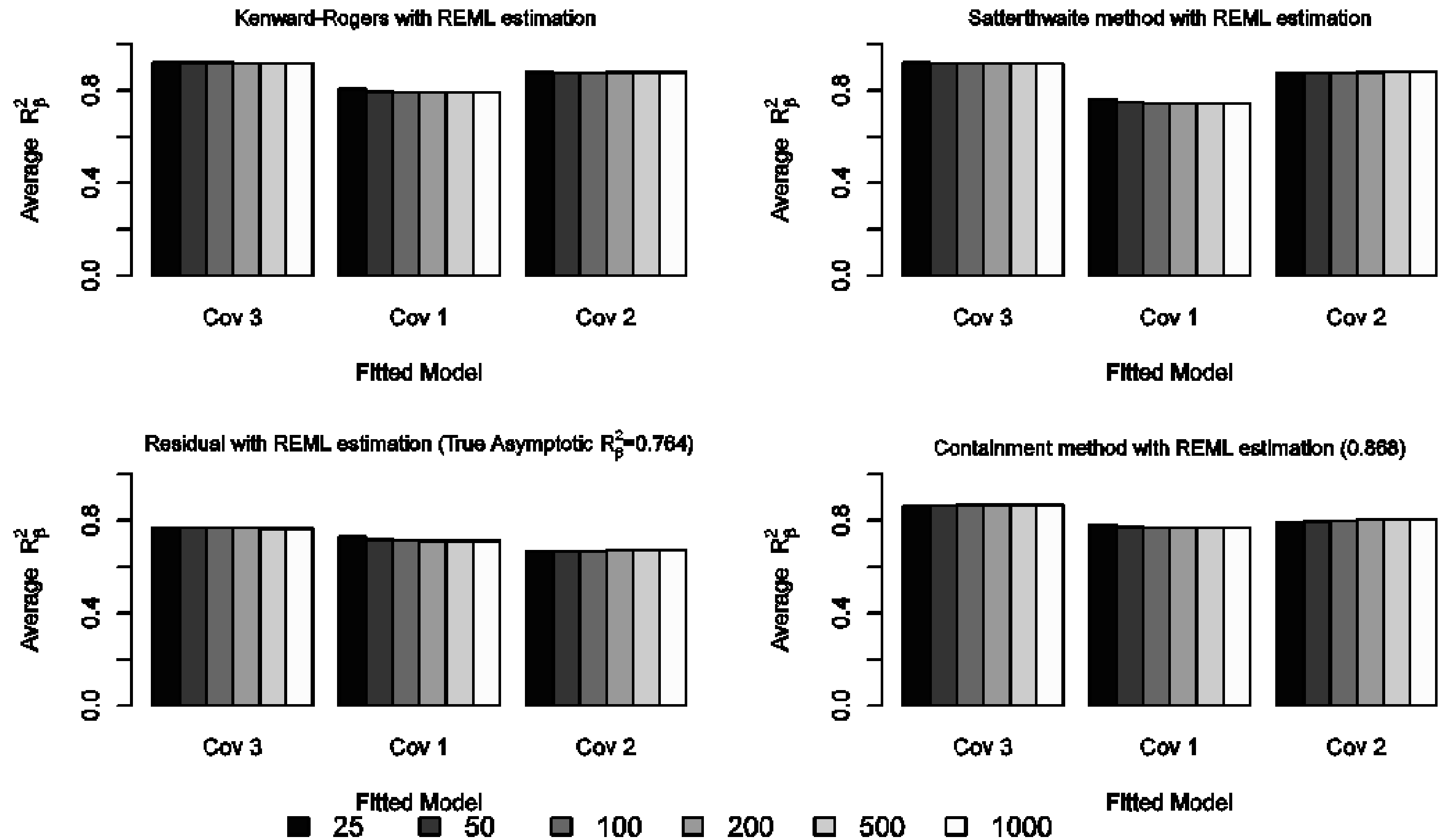


Figure 3.13 Average R_β^2 for Simulated Model - Model III: Covariance 3



CHAPTER 4

SEMI-PARTIAL R_β^2 AND FIXED EFFECTS MODEL SELECTION USING R_β^2

4.1 Introduction

There is an increasing desire for a universal R^2 statistic for the linear mixed model. There are many proposed R^2 statistics available for testing the fixed effects of linear mixed models. Unfortunately, as Kramer (2005) has suggested, a universal R^2 statistic for the linear mixed model may not be possible due to the complexity of the model and its assumptions. In fact, most likely, an R^2 statistic for the linear mixed model will have to be chosen based on the properties for which the investigator is interested. As a result, properties of these R^2 statistics will aid in determining which R^2 statistic to use.

Another consideration when choosing an R^2 statistic for the linear mixed model is to choose the statistic based on the purpose for which it is intended. R^2 statistics serve many functions. They are model selection tools, goodness of fit measures and measures of effect size (express the strength of a relationship between response and predictor). Therefore, it is important that the performance of R^2 statistics in the linear mixed model is evaluated as a model selection tool, goodness of fit measure and effect size.

4.2 Semi-partial R_β^2

4.2.1 Notation

The semi-partial R^2_β is defined for the hypothesis $H_0 : \beta_j = 0$ for $j \in \{1, ..., q-1\}$ where an individual fixed effect is being tested. If F_j is defined as the Wald-type F statistic for the j^{th} fixed effect with corresponding denominator degrees of freedom ν_j, then the semi-partial R^2_β is

$$\frac{\nu_j^{-1} F_j}{1 + \nu_j^{-1} F_j}.$$

More generally, if there is a test of a group of variables, or a contrast, resulting from an F statistic, a corresponding semi-partial R^2_β can be calculated. The semi-partial R^2 allows researchers to assess the relationship between a subset of predictors and the response adjusted for other predictors in the model. It represents the partial multivariate association between the repeated outcomes and one predictor after controlling for the effect of the other predictor.

4.3 Model Selection Methods

Model selection is a fundamental part of statistical analysis because researchers are often trying to build the best model to answer research questions. Muller and Fettermann (2002, pg 224) outline a process for model selection in the linear univariate model. Hosmer and Lemeshow (2000, Chapter 4) provide model selection techniques for logistic regression. Cheng et al (2010) outline model selection in the linear mixed model. Model selection in the linear mixed model is more complex than in the linear univariate model since both the covariance structure and mean model need to be chosen.

Information criteria are often used to choose the covariance structure in the linear mixed model and are becoming an increasingly popular fixed effects model selection tool.

4.3.1 Information Criteria

There are several different fit criteria which include: Akaike's Information Criteria (AIC) (Akaike, 1974), (AICC) (Hurvich and Tsai, 1989), Schwarz Bayesian Criterion (BIC) (Schwarz, 1978), (CAIC) (Bozdogan, 1987), and (HQIC). The criteria will not necessarily choose the best structure nor will they agree on the choice of the covariance structure.

A majority of the research regarding information criteria compares the AIC with the BIC, and the results are varied. Keselman, Algina, Kowalchuk and Wolfinger (1998) have shown that the BIC criterion selects the more parsimonious model as a result of its larger penalty term. Ferron et al. (2002) found AIC outperformed BIC in all of their simulation studies. The performance of these criteria improved with increasing sample size and level of autocorrelation. Ferron et al. (2002) investigated the sensitivity of various information criteria to misspecifications of the covariance structure. They found that data simulated from a first order autoregressive structure, the success rate of identifying the correct covariance structure for AIC was 71% and the success rate for BIC was 45%. Gomez et al (2005) generated data from fifteen covariance structures and found that performance of AIC and BIC depends on the sample size and complexity of the covariance structure. As these studies indicate, the performance of the information criteria is greatly dependent upon the linear mixed model being simulated.

4.4 An Adjusted R^2_β for Fixed Effects Model Selection

Morrison (1990) has shown that the null expectation of R^2 in linear regression using least squares estimation is

$$E\Big(R^2|H_0:\rho^2_{\boldsymbol{y}|\boldsymbol{X}}=0\Big)=\frac{q-1}{N-1}>0.$$

When the total number of observations, N is small, there is potential for large bias. The adjusted R^2 is

$$R_{\mathrm{A}}^2 = 1 - \frac{(1-R^2)(N-1)}{(N-q)},$$

and under the null hypothesis,

$$\begin{aligned} E\big(R_{\mathrm{A}}^2 | H_0\big) &= 1 - \frac{(N-1)E(1-R^2)}{(N-q)} \\ &= 1 - \frac{(N-1)-(N-1)E(R^2)}{(N-q)} \\ &= 1 - \frac{(N-1)-(q-1)}{(N-q)} \\ &= 0. \end{aligned}$$

From Chapter 2, the approximate Beta distribution expectation under the null hypothesis is,

$$E\big[R_\beta^2(\nu)\big|\, H_0\big] = \frac{q-1}{q+\nu-1} > 0.$$

If the denominator degrees of freedom is small, there is potential for large bias. Therefore, using similar process as in the linear univariate model, an adjusted version of R_β^2 is

$$R_{\beta_{\mathrm{adj}}}^2(\nu) = 1 - \frac{\big[1 - R_\beta^2\,(\nu)\big](q-1+\nu)}{\nu},$$

where,

$$\begin{aligned} E\Big[R_{\beta_{\mathrm{adj}}}^2(\nu)\big|\, H_0\Big] &= 1 - \frac{E\big[1 - R_\beta^2\,(\nu)\big](q-1+\nu)}{\nu} \\ &= 1 - \frac{(q-1+\nu)-(q-1+\nu)E\big[R_\beta^2(\nu)\big]}{\nu} \\ &= 1 - \frac{(q-1+\nu)-(q-1)}{\nu} \\ &= 0. \end{aligned}$$

4.5 Real World Examples

4.5.1 Data Background

The data from Potthoff and Roy (1964) described in Section 2.3.1.1 are used to illustrate the performance of the semi-partial R^2_β for various denominator degrees of freedom methods. Linear mixed models were fit with linear age and gender effect for three different covariance structures. The three covariance structures are the same as the simulation study presented in Section 2.7.1. Model R^2_β's and semi-partial R^2_β for both gender and age were calculated for Kenward-Roger, Satterthwaite, and residual denominator degrees of freedom.

4.5.2 Semi-partial R^2_β: Results

Table 4.1-4.3 provide the semi-partial and model R^2_β for the mean model with three different covariance structures. The semi-partial R^2_β for age is greater than the semi-partial R^2_β for gender for all denominator degrees of freedom methods and covariance structures. Both semi-partial R^2_β's are almost equivalent for the Kenward-Roger and Satterthwaite methods whereas, the semi-partial R^2_β's for the residual method are smaller. The semi-partial R^2_β's for gender when using Kenward-Roger and Satterthwaite vary from 0.20 to 0.27 across the three covariance structures used. The semi-partial R^2_β for age when using Kenward-Roger and Satterthwaite vary from 0.59 to 0.85 across the three covariance structures used. The semi-partial R^2_β for age when using the residual method vary from 0.45 to 0.52 across the three covariance structures used.

One of the problematic results is that when using the Kenward-Roger method, the model R^2_β is smaller than the semi-partial R^2_β for age for the two complex covariance structures. The same misalignment is not present for the R^2_β using residual and Satterthwaite denominator degrees of freedom methods.

4.6 Simulation Study for Semi-partial R^2_β

4.6.1 Objectives and Methods

The large scale simulation study summarized in Chapter 2 is used to assess the semi-partial forms of R_β^2 for the four denominator degrees of freedom methods. The simulations for mean Model II with the three different covariance structures will be analyzed according to their true model. Semi-partial R_β^2 for the continuous time effect and the model R_β^2 using the four denominator degrees of freedom methods are computed.

The purpose of this study is to compare the semi-partial R_β^2 for the continuous time effect and the model R_β^2 and examine under which linear mixed model settings and denominator degrees of freedom methods the semi-partial R_β^2 is larger than the model R_β^2.

4.6.2 Results

Table 4.4 provides the differences between the model R_β^2 and the semi-partial R_β^2 for the continuous time effect along with the minimum and maximum values of the difference across all 10,000 replications when the true model was fit to the data. Table 4.5 provides the proportion of times the semi-partial R_β^2 for the continuous time effect exceeds the model R_β^2. For Model II with Covariance 1, the model R_β^2 is always greater than the semi-partial R_β^2 for all denominator degrees of freedom methods and estimation techniques.

For Model II with Covariance 2, for the simulation studies with 100 subjects or more, the model R_β^2 was always less than the semi-partial R_β^2's for the continuous time effect when using Kenward-Roger and containment methods. When using Satterthwaite and residual methods, the model R_β^2 was always greater than the semi-partial R_β^2 for the continuous time effect.

For Model II with Covariance 3 using the Kenward-Roger method, for simulation studies with 500 subjects or more, the model R_β^2 was always less than the semi-partial R_β^2's for the continuous time effect. When using the residual method, the model R_β^2 was always greater than the semi-partial R_β^2's for the continuous time effect. When using the

Satterthwaite method, there was no consistent relationship between the model R_β^2 and the semi-partial R_β^2 for the continuous time effect (The difference was not either positive or negative across all 10,000 replications.) For simulation studies with 50 subjects or more, when using the containment method, the model R_β^2 was always less than the semi-partial R_β^2's for the continuous time effect.

4.6.3 Discussion

The misalignment of the semi-partial R_β^2 in relation to the model is not a desirable property of the statistic. If the semi-partial R_β^2 is greater than the model R_β^2, it suggests that the partial multivariate association between the repeated outcomes and one predictor after controlling for the effect of the other predictor is greater than the multivariate association between the repeated outcomes and both of the predictors.

For the Kenward-Roger, Satterthwaite and containment methods, there were simulations where the model R_β^2 was less than the semi-partial R_β^2 for the continuous time effect. The R_β^2 using the residual method was the only method where the model R_β^2 was never less than the semi-partial R_β^2 for the continuous time effect. Therefore, it is possible that the semi-partial R_β^2 should be defined using the residual method to avoid any misalignment.

Before that recommendation can be made further research is necessary to determine whether the values of the semi-partial R_β^2 using the residual method are true effect sizes. Researchers use the semi-partial R_β^2 as a measure of effect size associated with a statistical test. Therefore, measuring the magnitude of the association is very important. Edwards et al. (2008) has shown that for cases where a multivariate linear hypothesis can be written as a linear mixed model hypothesis, the Hotelling-Lawley trace measure of association is a one-to-one function of R_β^2 defined using Kenward-Roger denominator degrees of freedom.

4.7 Simulation Study assessing R^2_β as a Model Selection Tool

Throughout the model selection process, there are four different mean models and three covariance structures being evaluated. Thus, the maximum model contains four covariates consisting of an intercept, a binary effect, and a continuous time effect and their interaction. The maximum model is denoted as Model III.

4.7.1 Objectives

The overall goal of this simulation study is to determine whether either $R^2_\beta(\nu)$ or $R^2_{\beta_{\text{adj}}}(\nu)$ is a reliable mean model selection tool and evaluate the impact of the denominator degrees of freedom method used to define these statistics. Four denominator degrees of freedom methods will be evaluated including: residual, containment, Satterthwaite and Kenward-Roger methods.

4.7.2 Scenario 1: Model Selection for Known Covariance Structure

Four candidate models were fit for each of the 10,000 replications of each simulation study. These candidate models were fit using REML estimation and four denominator degrees of freedom methods. The candidate models consisted of:

(1) An intercept and slope model

(2) An intercept and group effect model

(3) An intercept, slope, and group effect model

(4) An intercept, slope, group effect and group $\times$ slope interaction.

The mean model corresponding to the largest $R^2_\beta(\nu)$ and $R^2_{\beta_{\text{adj}}}(\nu)$ was selected for each denominator degrees of freedom method.

4.7.2.1 Results

Table 4.6-4.8 shows the accuracy of $R^2_{\beta}(\nu)$ and $R^2_{\beta_{\text{adj}}}(\nu)$ for each denominator degrees of freedom method at selecting the true mean model when the covariance structure is assumed known. Table 4.6 provides the performance when the true model was simulated with only a continuous time effect (Mean Model I) for three covariance structures. As the subject sample size increases, $R^2_{\beta}(\nu)$ performance improved for all denominator degrees of freedom methods. The performance of the $R^2_{\beta_{\text{adj}}}(\nu)$ were relatively constant for the containment and residual methods.

Performance of $R^2_{\beta}(\nu)$ and $R^2_{\beta_{\text{adj}}}(\nu)$ varied greatly based on the true covariance structure. $R^2_{\beta}(\nu)$ for all of the denominator degrees of freedom performed very poorly for the simplest covariance structure (Covariance 1). For the more complex covariance structures, the performance of the unadjusted versions using Kenward-Roger and containment denominator degrees of freedom improved dramatically. For each denominator degrees of freedom method, using $R^2_{\beta_{\text{adj}}}(\nu)$ improved the performance as compared to the corresponding $R^2_{\beta}(\nu)$.

Table 4.7 shows the results of the true mean model with a binary and a continuous time effect (True Mean Model II). For the simplest covariance structure (Covariance 1), the unadjusted versions of Kenward-Roger and Satterthwaite $R^2_{\beta}(\nu)$ performed well in selecting the true model while the unadjusted versions of the containment and residual $R^2_{\beta}(\nu)$ performed poorly. The performance for each denominator degrees of freedom method was improved in the adjusted versions. For the more complex covariance structures (Covariance 2 and 3), $R^2_{\beta}(\nu)$ and $R^2_{\beta_{\text{adj}}}(\nu)$ using Kenward-Roger and containment were outperformed by the $R^2_{\beta}(\nu)$ and $R^2_{\beta_{\text{adj}}}(\nu)$ using Satterthwaite and residual.

Table 4.8 shows the results of the true mean model with a binary effect, continuous time effect and their interaction (True Mean Model III). For the simplest covariance structure, the unadjusted versions of containment or residual performed well in seected the true model while the unadjusted versions of Kenward-Roger and Satterthwaite performed poorly. The performance for each denominator degrees of freedom method was not improved in the adjusted versions. For more complex covariance structures, $R^2_\beta(\nu)$ and $R^2_{\beta_{\text{adj}}}(\nu)$ using residual outperformed the other denominator degrees of freedom methods. For each denominator degrees of freedom method, using $R^2_{\beta_{\text{adj}}}(\nu)$ did not improve the performance as compared to $R^2_\beta(\nu)$.

4.7.3 Scenario 2: Complete Model Selection

Although interest is focused on the fixed effects, the covariance structure is important for accurate inference on the fixed effects (Verbeke and Molenberghs, 2000, p. 62). Thus, the complete model selection process will be examined. The complete model selection process highlighted in Cheng et al (2010) will be conducted for each simulation replication. First, the covariance structure for the maximum mean model will be selected using AIC and BIC separately under REML. Then, $R^2_{\beta_{\text{adj}}}(\nu)$ defined using residual and Kenward-Roger denominator degrees of freedom will be used as a criterion for selecting the mean model.

4.7.3.1 Covariance Structure Selection: Results

The true covariance structure is not known in real world applications of the linear mixed model. In fact, for small samples, there is not a universally accepted method used to identify the best covariance models (Gurka, 2006).

Tables 4.9-4.17 present the covariance structure selected using AIC, AICC, CAIC, BIC, and HQIC for the nine simulated models. For the simulated models with the simplest covariance structure (Covariance 1), the CAIC performed the best out of the five different criteria examined. The BIC performed similarly with the percentage of replications selecting the true model tending to 100%. For simulations with Covariance 2, the information criteria performed poorly for the smaller subject sample sizes but performance greatly improved for the larger subject sample sizes. For the simulated models with the most complex covariance structure (Covariance 3), all of the information criteria accurately selected the true covariance structure.

4.7.3.2 Mean Model Selection: Results

The complete model selection results using AIC and BIC are presented in this chapter. Table 4.18-4.20 present the results for the complete model selection process when using AIC to choose the covariance structure. Table 4.21-4.23 present the results for the complete model selection process when using BIC to choose the covariance structure. The complete model selection results using AIC to first select the covariance structure were very similar to the complete model selection results using BIC.

The performance of REML KR $R^2_{\beta_{\text{adj}}}(\nu)$ in selecting the true mean model when the covariance was chosen with either AIC or BIC was not very consistent. Under certain simulation conditions, the REML KR $R^2_{\beta_{\text{adj}}}(\nu)$ performed very well and others very poorly. The performance of REML KR $R^2_{\beta_{\text{adj}}}(\nu)$ varied based on the true mean model, and the true covariance structure. Conversely, the performance of REML residual $R^2_{\beta_{\text{adj}}}(\nu)$ was fairly consistent. The REML residual $R^2_{\beta_{\text{adj}}}(\nu)$ selected the true mean model approximately 50% of the time when the true model had continuous time effect and was not dependent upon the true covariance structure. The performance of REML residual $R^2_{\beta_{\text{adj}}}(\nu)$ improved as the complexity of the true model increased.

4.8 Conclusions and Discussion

This study investigated the appropriateness of using R^2_β as a model selection tool and an adjusted version of R^2_β was created to aid in model selection. In addition, investigation was conducted as to defining R^2_β and the adjusted version using three other denominator degrees of freedom methods.

Although this simulation study was based on model selection for a maximum model of only three covariates, there is evidence to suggest fixed effects model selection in the linear mixed model should be conducted using the adjusted R^2_β. The unadjusted R^2_β for all denominator degrees of freedom do not perform consistently as a fixed effects model selection tool. The adjusted R^2_β improved performance of its corresponding unadjusted form for most of the simulations. In addition, for purposes of model selection, there is evidence that the adjusted R^2_β should be defined using REML estimation with residual denominator degrees of freedom. The adjusted R^2_β using REML with residual denominator degrees of freedom performed the most consistently for each of the simulation study scenarios.

All of the information criteria performed poorly for the small subject sample size and simpliest covariance structure simulations. As the subject sample size increased, the performance of the information criteria improved. While covariance structure selection is not the primary focus of this paper, it has a direct impact on the mean model selection. Research has indicated that covariance structure greatly impacts the values of R^2_β and thus future work is necessary.

REFERENCES

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transaction on Automatic Control,* **AC-19**, 716-723.

Bozdogan, H. (2000). Akaike information criterion and recent developments in information complexity. *Journal of Mathematical Psychology*, 44, 6291.

Cheng, J., Edwards, L.J., Maldonado-Molina, M.M., Komro, K.A., and Muller, K.E. (2010). Real longitudinal data analysis for real people: Building a good enough mixed model. *Statistics in Medicine*, **29**: 504-520.

Ferron, J., Hogarty, K. Y., Hess, M., Romano, J., Kromrey, J. D., Niles, J. D., Dawkins, G.R., and Sentovich, C. (2002). A macro for examining the consequences of error structure misspecifications. *Proceedings of the Twenty-Seventh Annual SAS Users Group International Conference*, Cary, NC: SAS Institute Inc. 201-227.

Gómez, V.E., Schaalje, G.B., & Fellingham, G.W. (2005). Performance of the Kenward-Roger method when the covariance structure is selected using AIC and BIC. *Communications in Statistics- Simulation and Computation* **34**, 377–392.

Gurka, M. J. (2006). Selecting the Best Linear Mixed Model under REML *The American Statistician*, **60**: 19-26.

Hosmer, D.W. and Lemeshow, S. (2000). *Applied Logistic Regression*. 2nd edition. Wiley.

Hurvich, C. M., and Tsai, C. L. (1989). Regression and Time Series Model Selection in Small Samples *Biometrika*, **76**, 297-307.

Keselman, H.J., Algina, J., Kowalchuk, R.K., and Wolfinger, R.D. (1999a). A comparison of recent approaches to the analysis of repeated measurements. *British Journal of Mathematical and Statistical Psychology* **52**, 63-78.

Kramer, M. (2005). R^2 statistics for mixed models. 2005 *Proceedings of the Conference on Applied Statistics in Agriculture*, Manhattan, KS, pp 148–160.

Morrison, D.F. (1990). *Multivariate statistical methods*. New York: McGraw-Hill.

Muller, K.E., and Fetterman, B.A. (2002). *Regression and ANOVA: An integrated approach using SAS software*. Cary, NC: SAS Institute.

Potthoff, R.F. and Roy, S.N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika*, **51**, 313-326.

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, **6**, 461-464.

Verbeke, G. and Molenberghs, G. (2000). *Linear Mixed Models for Longitudinal Data* New York, NY: Springer-Verlag.

TABLES

Table 4.1 Semi-partial R^2_β and model R^2_β using Kenward-Roger F using REML estimation for the dental data with Model II for three covariance structures

Cov	Variable	F	ν	Semi-Partial R^2_β	Model R^2_β
1	Gender	9.29	25.0	0.27	0.71
	Age	114.84	80.0	0.59	
2	Gender	7.34	25.0	0.23	0.73
	Age	85.85	26.0	0.77	
3	Gender	6.24	25.2	0.20	0.81
	Age	87.38	16.9	0.84	

Table 4.2 Semi-partial R^2_β and model R^2_β using REML estimation with Satterthwaite denominator degrees of freedom for the dental data with model II for three covariance structures

Cov	Variable	F	ν	Semi-Partial R^2_β	Model R^2_β
1	Gender	9.29	25.0	0.27	0.77
	Age	114.84	80.0	0.59	
2	Gender	8.02	25.0	0.24	0.79
	Age	85.85	26.0	0.77	
3	Gender	6.73	25.2	0.21	0.85
	Age	93.74	16.9	0.85	

Table 4.3 Semi-partial R^2_β and model R^2_β using REML estimation with the Residual degrees of freedom for the dental data with model II for three covariance structures

Cov	Variable	F	ν	Semi-Partial R^2_β	Model R^2_β
1	Gender	9.29	105	0.08	0.54
	Age	114.84	105	0.52	
2	Gender	8.02	105	0.07	0.47
	Age	85.85	105	0.45	
3	Gender	6.73	105	0.06	0.51
	Age	93.74	105	0.47	

Table 4.4 Average (Minimum, Maximum) difference between model R_β^2 and semi-partial R_β^2 for the continuous time effect across all 10,000 simulations using REML estimation in Model II

		Kenward-Roger	Satterthwaite	Containment	Residual
Model II: Covariance 1					
	25 subjects	0.16 (0.08, 0.39)	0.23 (0.14, 0.44)	0.05 (0.00, 0.29)	0.05 (0.00, 0.27)
	50 subjects	0.15 (0.10, 0.28)	0.22 (0.16, 0.35)	0.05 (0.00, 0.17)	0.04 (0.00, 0.15)
	100 subjects	0.14 (0.11, 0.23)	0.21 (0.18, 0.29)	0.04 (0.01, 0.12)	0.04 (0.01, 0.11)
	200 subjects	0.14 (0.11, 0.20)	0.22 (0.18, 0.27)	0.04 (0.01, 0.10)	0.04 (0.01, 0.09)
	500 subjects	0.14 (0.12, 0.17)	0.21 (0.19, 0.24)	0.04 (0.02, 0.07)	0.04 (0.02, 0.07)
	1000 subjects	0.14 (0.13, 0.17)	0.21 (0.20, 0.24)	0.04 (0.03, 0.06)	0.04 (0.03, 0.06)
Model II: Covariance 2					
	25 subjects	-0.04 (-0.06, 0.08)	0.02 (0.00, 0.15)	-0.13 (-0.18, -0.02)	0.02 (0.00, 0.13)
	50 subjects	-0.04 (-0.07, 0.03)	0.01 (0.00, 0.09)	-0.13 (-0.17, -0.06)	0.01 (0.00, 0.11)
	100 subjects	-0.04 (-0.06, -0.02)	0.01 (0.00, 0.04)	-0.13 (-0.16, -0.10)	0.01 (0.00, 0.05)
	200 subjects	-0.04 (-0.06, -0.03)	0.01 (0.00, 0.03)	-0.13 (-0.15, -0.10)	0.01 (0.00, 0.04)
	500 subjects	-0.04 (-0.05, -0.03)	0.01 (0.00, 0.02)	-0.13 (-0.14, -0.10)	0.01 (0.00, 0.03)
	1000 subjects	-0.04 (-0.05, -0.04)	0.01 (0.00, 0.01)	-0.13 (-0.14, -0.12)	0.01 (0.01, 0.02)
Model II: Covariance 3					
	25 subjects	-0.02 (-0.07, 0.08)	0.00 (-0.03, 0.11)	-0.09 (-0.17, 0.01)	0.03 (0.00, 0.19)
	50 subjects	-0.02 (-0.06, 0.04)	0.00 (-0.02, 0.06)	-0.08 (-0.14, -0.02)	0.03 (0.00, 0.14)
	100 subjects	-0.02 (-0.05, 0.01)	0.00 (-0.01, 0.04)	-0.08 (-0.13, -0.05)	0.03 (0.00, 0.09)
	200 subjects	-0.02 (-0.04, 0.00)	0.00 (-0.01, 0.02)	-0.08 (-0.11, -0.05)	0.03 (0.01, 0.07)
	500 subjects	-0.02 (-0.03, -0.01)	0.00 (-0.01, 0.01)	-0.08 (-0.10, -0.06)	0.03 (0.02, 0.05)
	1000 subjects	-0.02 (-0.03, -0.01)	0.00 (-0.01, 0.01)	-0.08 (-0.09, -0.07)	0.03 (0.02, 0.05)

Table 4.5 Proportion of times the semi-partial R_β^2 for the continuous time effect exceeds the model R_β^2 across all 10,000 simulations using REML estimation in Model II with Covariance 2 and Covariance 3

		Kenward-Roger	Satterthwaite	Containment	Residual
Model II: Covariance 2					
	25 subjects	98.68	0	100	0
	50 subjects	99.95	0	100	0
	100 subjects	100	0	100	0
	200 subjects	100	0	100	0
	500 subjects	100	0	100	0
	1000 subjects	100	0	100	0
Model II: Covariance 3					
	25 subjects	91.17	44.33	99.99	0
	50 subjects	95.98	42.70	100	0
	100 subjects	99.46	40.87	100	0
	200 subjects	99.98	39.78	100	0
	500 subjects	100	35.52	100	0
	1000 subjects	100	29.96	100	0

Table 4.6 Percentage of the correct mean model selected using different denominator degrees of freedom methods when defining R_β^2 and an adjusted R_β^2 for each of the denominator degrees of freedom methods for mean model I with three different covariance structures

m	REML KR	REML Sat	REML Contain	REML Residual	Adjusted REML KR	Adjusted REML Sat	Adjusted REML Contain	Adjusted REML Residual
Model I: Covariance 1								
25	0.06	0.06	0.01	0	0.22	0.22	63.96	52.45
50	0	0	0	0	0	0	63.89	51.25
100	0	0	0	0	0	0	64.19	51.42
200	0	0	0	0	0	0	63.69	51.49
500	0	0	0	0	0	0	64.33	52.06
1,000	0	0	0	0	0	0	63.85	51.88
Model I: Covariance 2								
25	70.66	2.01	99.99	0.09	77.63	65.03	99.99	47.52
50	86.06	0.86	100	0	87.51	76.69	100	47.93
100	94.81	0.08	100	0	94.9	85.56	100	48.77
200	99.17	0.01	100	0	99.17	89.69	100	47.86
500	99.98	0	100	0	99.98	90.88	100	48.82
1000	10000	0	100	0	100	91.1	100	48.92
Model I: Covariance 3								
25	92.13	89.8	99.99	0.65	92.62	92.16	99.99	48.3
50	98.28	98.02	100	0.11	98.28	98.22	100	47.81
100	99.9	99.9	100	0.03	99.9	99.9	100	49.08
200	100	100	100	0	100	100	100	49.74
500	100	100	100	0	100	100	100	50.01
1,000	100	100	100	0	100	100	100	49.45

Table 4.7 Percentage of the correct mean model selected using different denominator degrees of freedom methods when defining R^2_β and an adjusted R^2_β for each of the denominator degrees of freedom methods for mean model II with three different covariance structures

m	**REML KR**	**REML Sat**	**REML Contain**	**REML Residual**	**Adjusted REML KR**	**Adjusted REML Sat**	**Adjusted REML Contain**	**Adjusted REML Residual**
Model II: Covariance 1								
25	98.67	99.62	0	12.25	98.88	99.79	65.51	66.9
50	99.96	100	0	13.27	99.96	100	68.11	68.17
100	100	100	0	14.41	100	100	69.4	69.4
200	100	100	0	15.49	100	100	67.76	67.76
500	100	100	0	16.24	100	100	67.95	67.95
1,000	100	100	0	15.47	100	100	68.13	68.13
Model II: Covariance 2								
25	1.8	86.13	0	49.73	4.69	65.21	0	58.08
50	0.67	94.37	0	52.92	0.71	85.72	0	66.96
100	0.04	99.12	0	54.56	0.04	98.56	0	68.42
200	0	99.93	0	56.41	0	99.93	0	68.7
500	0	100	0	55.11	0	100	0	68.04
1000	0	100	0	55.66	0	100	0	68.53
Model II: Covariance 3								
25	6	20.89	0	45.84	6.31	13.48	0	58.45
50	1.49	8.23	0	50.23	1.52	5.2	0	65.76
100	0.08	1.63	0	51.24	0.08	0.94	0	68.05
200	0	0.16	0	51.55	0	0.11	0	68.33
500	0	0	0	51.85	0	0	0	68.33
1,000	0	0	0	51.57	0	0	0	68.56

Table 4.8 Percentage of the correct mean model selected using different denominator degrees of freedom methods when defining R^2_β and an adjusted R^2_β for each of the denominator degrees of freedom methods for mean model III with three different covariance structures

m	**REML KR**	**REML Sat**	**REML Contain**	**REML Residual**	**Adjusted REML KR**	**Adjusted REML Sat**	**Adjusted REML Contain**	**Adjusted REML Residual**
Model III: Covariance 1								
25	1.18	0.47	100	97.69	1.01	0.39	72.46	72.46
50	0.07	0.01	100	99.46	0.05	0.01	88.7	88.7
100	0	0	100	99.98	0	0	98.36	98.36
200	0	0	100	100	0	0	99.99	99.99
500	0	0	100	100	0	0	100	100
1,000	0	0	100	100	0	0	100	100
Model III: Covariance 2								
25	74.22	45.8	2.41	93.41	59.38	40.35	2.00	88.8
50	79.67	55.8	0.32	99.05	70.67	51.85	0.26	98.29
100	90.35	70.36	0	99.96	86.9	67.87	0	99.94
200	97.91	81.9	0	100	97.04	80.49	0	100
500	99.99	94.1	0	100	99.97	93.67	0	100
1000	100	98.92	0	100	100	98.8	0	100
Model III: Covariance 3								
25	29.34	11.3	6.12	94.23	21.83	9.67	5.47	90.39
50	22.81	6.34	2.02	99.47	18.12	5.63	1.75	98.87
100	12.01	1.95	0.21	99.99	9.97	1.82	0.19	99.97
200	3.65	0.25	0	100	2.94	0.23	0	100
500	0.15	0	0	100	0.12	0	0	100
1,000	0	0	0	100	0	0	0	100

Table 4.9 Covariance Structure Selection for Model I: Covariance 1 using ML estimation

	Covariance Structure Selected		
	Covariance 1	Covariance 2	Covariance 3
AIC			
25 subjects	0 (0)	7872 (78.72)	2128 (21.28)
50 subjects	7432 (74.32)	1727 (17.27)	841 (8.41)
100 subjects	7762 (77.62)	1487 (14.87)	751 (7.51)
200 subjects	7996 (79.96)	1299 (12.99)	705 (7.05)
500 subjects	8536 (85.36)	1464 (14.64)	0 (0)
1000 subjects	10000 (100)	0 (0)	0 (0)
AICC			
25 subjects	0 (0)	8192 (81.92)	1808 (18.08)
50 subjects	7656 (76.56)	1625 (16.25)	719 (7.19)
100 subjects	7891 (78.91)	1425 (14.25)	684 (6.84)
200 subjects	8064 (80.64)	1270 (12.7)	666 (6.66)
500 subjects	8556 (85.56)	1444 (14.44)	0 (0)
1000 subjects	10000 (100)	0 (0)	0 (0)
CAIC			
25 subjects	0 (0)	8985 (89.85)	1015 (10.15)
50 subjects	9222 (92.22)	719 (7.19)	59 (5.9)
100 subjects	9531 (95.31)	460 (4.6)	9 (0.09)
200 subjects	9728 (97.28)	265 (2.65)	7 (0.07)
500 subjects	9941 (99.41)	59 (0.59)	0 (0)
1000 subjects	10000 (100)	0 (0)	0 (0)
BIC			
25 subjects	0 (0)	8669 (86.69)	1331 (13.31)
50 subjects	8980 (89.8)	875 (8.75)	145 (1.45)
100 subjects	9381 (93.81)	571 (5.71)	48 (0.48)
200 subjects	9639 (96.39)	87 (0.87)	0 (0)
500 subjects	9913 (99.13)	87 (0.87)	0 (0)
1000 subjects	10000 (100)	0 (0)	0 (0)
HQIC			
25 subjects	0 (0)	8160 (81.6)	1840 (18.4)
50 subjects	8313 (83.13)	1274 (12.74)	413 (4.13)
100 subjects	8811 (88.11)	941 (9.41)	248 (2.48)
200 subjects	9164 (91.64)	671 (6.71)	165 (1.65)
500 subjects	9566 (95.66)	434 (4.34)	0 (0)
1000 subjects	10000 (100)	0 (0)	0 (0)

Table 4.10 Covariance Structure Selection for Model I: Covariance 2 using ML estimation

	Covariance Structure Selected		
	Covariance 1	Covariance 2	Covariance 3
AIC			
25 subjects	5299 (52.99)	3367 (33.67)	1334 (13.34)
50 subjects	3597 (35.97)	5028 (50.28)	1375 (13.75)
100 subjects	1451 (14.51)	6967 (69.67)	1582 (15.82)
200 subjects	156 (1.56)	8286 (82.86)	1558 (15.58)
500 subjects	0 (0)	8401 (84.01)	1599 (15.99)
1000 subjects	0 (0)	8445 (84.45)	1555 (15.55)
AICC			
25 subjects	5911 (59.11)	3089 (30.89)	1000 (10.00)
50 subjects	3911 (39.11)	4892 (48.92)	1197 (11.97)
100 subjects	1554 (15.54)	6966 (69.66)	1480 (14.80)
200 subjects	162 (1.62)	8330 (83.30)	1508 (15.08)
500 subjects	0 (0)	8420 (84.20)	1580 (15.80)
1000 subjects	0 (0)	8452 (84.52)	1548 (15.48)
CAIC			
25 subjects	7945 (79.45)	1776 (17.76)	279 (2.79)
50 subjects	7798 (77.98)	2043 (20.43)	159 (1.59)
100 subjects	6026 (60.26)	3846 (38.46)	128 (1.28)
200 subjects	2636 (26.36)	7246 (72.46)	118 (1.18)
500 subjects	57 (0.57)	9870 (98.70)	73 (0.73)
1000 subjects	0 (0)	9940 (99.40)	60 (0.60)
BIC			
25 subjects	7134 (71.34)	2322 (23.22)	544 (5.44)
50 subjects	6766 (67.66)	2907 (29.07)	327 (3.27)
100 subjects	4848 (48.48)	4910 (49.10)	242 (2.42)
200 subjects	1790 (17.90)	8000 (80.00)	210 (2.1)
500 subjects	31 (0.31)	9850 (98.50)	119 (1.19)
1000 subjects	0 (0)	9905 (99.05)	95 (0.95)
HQIC			
25 subjects	5905 (59.05)	3075 (30.75)	1020 (10.2)
50 subjects	4946 (49.46)	4273 (42.73)	781 (7.81)
100 subjects	2774 (27.74)	6480 (64.8)	746 (7.46)
200 subjects	550 (5.50)	8776 (87.76)	674 (6.74)
500 subjects	3 (0.03)	9449 (94.49)	548 (5.48)
1000 subjects	0 (0)	9475 (94.75)	525 (5.25)

Table 4.11 Covariance Structure Selection for Model I: Covariance 3 using ML estimation

	Covariance Structure Selected		
	Covariance 1	Covariance 2	Covariance 3
AIC			
25 subjects	12 (0.12)	11 (0.11)	9977 (99.77)
50 subjects	0 (0)	0 (0)	10000 (100)
100 subjects	0 (0)	0 (0)	10000 (100)
200 subjects	0 (0)	0 (0)	10000 (100)
500 subjects	0 (0)	0 (0)	10000 (100)
1000 subjects	0 (0)	0 (0)	10000 (100)
AICC			
25 subjects	24 (0.24)	13 (0.13)	9963 (99.63)
50 subjects	0 (0)	0 (0)	10000 (100)
100 subjects	0 (0)	0 (0)	10000 (100)
200 subjects	0 (0)	0 (0)	10000 (100)
500 subjects	0 (0)	0 (0)	10000 (100)
1000 subjects	0 (0)	0 (0)	10000 (100)
CAIC			
25 subjects	209 (2.09)	61 (0.61)	9730 (97.3)
50 subjects	0 (0)	0 (0)	10000 (100)
100 subjects	0 (0)	0 (0)	10000 (100)
200 subjects	0 (0)	0 (0)	10000 (100)
500 subjects	0 (0)	0 (0)	10000 (100)
1000 subjects	0 (0)	0 (0)	10000 (100)
BIC			
25 subjects	71 (0.71)	27 (0.27)	9902 (99.02)
50 subjects	0 (0)	0 (0)	10000 (100)
100 subjects	0 (0)	0 (0)	10000 (100)
200 subjects	0 (0)	0 (0)	10000 (100)
500 subjects	0 (0)	0 (0)	10000 (100)
1000 subjects	0 (0)	0 (0)	10000 (100)
HQIC			
25 subjects	24 (0.24)	13 (0.13)	9963 (99.63)
50 subjects	0 (0)	0 (0)	10000 (100)
100 subjects	0 (0)	0 (0)	10000 (100)
200 subjects	0 (0)	0 (0)	10000 (100)
500 subjects	0 (0)	0 (0)	10000 (100)
1000 subjects	0 (0)	0 (0)	10000 (100)

Table 4.12 Covariance Structure Selection for Model II: Covariance 1 using ML estimation

	Covariance Structure Selected		
	Covariance 1	Covariance 2	Covariance 3
AIC			
25 subjects	7138 (71.38)	1890 (18.9)	971 (9.71)
50 subjects	7415 (74.15)	1722 (17.22)	862 (8.62)
100 subjects	7765 (77.65)	1486 (14.86)	749 (7.49)
200 subjects	7951 (79.51)	1266 (12.66)	783 (7.83)
500 subjects	8096 (80.96)	1235 (12.35)	669 (6.69)
1000 subjects	8114 (81.14)	1191 (11.91)	695 (6.95)
AICC			
25 subjects	7591 (75.91)	1705 (17.05)	703 (7.03)
50 subjects	7665 (76.65)	1609 (16.09)	725 (7.25)
100 subjects	7870 (78.7)	1449 (14.49)	681 (6.81)
200 subjects	8003 (80.03)	1243 (12.43)	754 (7.54)
500 subjects	8117 (81.17)	1226 (12.26)	657 (6.57)
1000 subjects	8127 (81.27)	1186 (11.86)	687 (6.87)
CAIC			
25 subjects	8801 (88.01)	1060 (10.6)	138 (1.38)
50 subjects	9216 (92.16)	728 (7.28)	55 (0.55)
100 subjects	9514 (95.14)	463 (4.63)	23 (0.23)
200 subjects	9718 (97.18)	271 (2.71)	11 (0.11)
500 subjects	9930 (99.3)	65 (0.65)	5 (0.05)
1000 subjects	9976 (99.76)	23 (0.23)	1 (0.01)
BIC			
25 subjects	8399 (83.99)	1280 (12.8)	320 (3.2)
50 subjects	8958 (89.58)	905 (9.05)	136 (1.36)
100 subjects	9377 (93.77)	570 (5.7)	53 (0.53)
200 subjects	9654 (96.54)	325 (3.25)	21 (0.21)
500 subjects	9893 (98.93)	96 (0.96)	11 (0.11)
1000 subjects	9948 (99.48)	46 (0.46)	6 (0.06)
HQIC			
25 subjects	7596 (75.96)	1693 (16.93)	710 (7.10)
50 subjects	8299 (82.99)	1290 (12.9)	410 (4.1)
100 subjects	8779 (87.79)	965 (9.65)	256 (2.56)
200 subjects	9191 (91.91)	636 (6.36)	173 (1.73)
500 subjects	9521 (95.21)	377 (3.77)	102 (1.02)
1000 subjects	9586 (95.86)	323 (3.23)	91 (0.91)

Table 4.13 Covariance Structure Selection for Model II: Covariance 2 using ML estimation

	Covariance Structure Selected		
	Covariance 1	Covariance 2	Covariance 3
AIC			
25 subjects	5289 (52.89)	3391 (33.91)	1320 (13.2)
50 subjects	3593 (35.93)	5008 (50.08)	1399 (13.99)
100 subjects	1425 (14.25)	7017 (70.17)	1558 (15.58)
200 subjects	176 (1.76)	8281 (82.81)	1543 (15.43)
500 subjects	0 (0)	8464 (84.64)	1536 (15.36)
1000 subjects	0 (0)	8390 (83.9)	1610 (16.1)
AICC			
25 subjects	5914 (59.14)	3083 (30.83)	1003 (10.03)
50 subjects	3916 (39.16)	4853 (48.53)	1231 (12.31)
100 subjects	1516 (15.16)	7031 (70.31)	1453 (14.53)
200 subjects	183 (1.83)	8323 (83.23)	1494 (14.94)
500 subjects	0 (0)	8473 (84.73)	1527 (15.27)
1000 subjects	0 (0)	8400 (84)	1600 (16)
CAIC			
25 subjects	7935 (79.35)	1766 (17.66)	299 (2.99)
50 subjects	7715 (77.15)	2111 (21.11)	174 (1.74)
100 subjects	6134 (61.34)	3742 (37.42)	115 (1.15)
200 subjects	2583 (25.83)	7311 (73.11)	104 (1.04)
500 subjects	45 (0.45)	9895 (98.95)	60 (0.60)
1000 subjects	0 (0)	9945 (99.45)	55 (0.55)
BIC			
25 subjects	7094 (70.94)	2359 (23.59)	547 (5.47)
50 subjects	6683 (66.83)	2965 (29.65)	352 (3.52)
100 subjects	4950 (49.5)	4801 (48.01)	249 (2.49)
200 subjects	1785 (17.85)	8018 (80.18)	197 (1.97)
500 subjects	15 (0.15)	9872 (98.72)	113 (1.13)
1000 subjects	0 (0)	9909 (99.09)	91 (0.91)
HQIC			
25 subjects	5910 (59.1)	3068 (30.68)	1022 (10.22)
50 subjects	4948 (49.48)	4219 (42.19)	833 (8.33)
100 subjects	2813 (28.13)	6448 (64.48)	739 (7.39)
200 subjects	621 (6.21)	8722 (87.22)	657 (6.57)
500 subjects	0 (0)	9438 (94.38)	562 (5.62)
1000 subjects	0 (0)	9503 (95.03)	497 (4.97)

Table 4.14 Covariance Structure Selection for Model II: Covariance 3 using ML estimation

	Covariance Structure Selected		
	Covariance 1	Covariance 2	Covariance 3
AIC			
25 subjects	5 (0.05)	6 (0.06)	9988 (99.88)
50 subjects	0 (0)	0 (0)	10000 (100)
100 subjects	0 (0)	0 (0)	10000 (100)
200 subjects	0 (0)	0 (0)	10000 (100)
500 subjects	0 (0)	0 (0)	10000 (100)
1000 subjects	0 (0)	0 (0)	10000 (100)
AICC			
25 subjects	20 (0.2)	8 (0.08)	9971 (99.71)
50 subjects	0 (0)	0 (0)	10000 (100)
100 subjects	0 (0)	0 (0)	10000 (100)
200 subjects	0 (0)	0 (0)	10000 (100)
500 subjects	0 (0)	0 (0)	10000 (100)
1000 subjects	0 (0)	0 (0)	10000 (100)
CAIC			
25 subjects	232 (2.32)	42 (0.42)	9725 (97.25)
50 subjects	0 (0)	0 (0)	10000 (100)
100 subjects	0 (0)	0 (0)	10000 (100)
200 subjects	0 (0)	0 (0)	10000 (100)
500 subjects	0 (0)	0 (0)	10000 (100)
1000 subjects	0 (0)	0 (0)	10000 (100)
BIC			
25 subjects	71 (0.71)	19 (0.19)	9909 (99.09)
50 subjects	0 (0)	0 (0)	10000 (100)
100 subjects	0 (0)	0 (0)	10000 (100)
200 subjects	0 (0)	0 (0)	10000 (100)
500 subjects	0 (0)	0 (0)	10000 (100)
1000 subjects	0 (0)	0 (0)	10000 (100)
HQIC			
25 subjects	19 (0.19)	8 (0.08)	9972 (99.72)
50 subjects	0 (0)	0 (0)	10000 (100)
100 subjects	0 (0)	0 (0)	10000 (100)
200 subjects	0 (0)	0 (0)	10000 (100)
500 subjects	0 (0)	0 (0)	10000 (100)
1000 subjects	0 (0)	0 (0)	10000 (100)

Table 4.15 Covariance Structure Selection for Model III: Covariance 1 using ML estimation

	Covariance Structure Selected		
	Covariance 1	Covariance 2	Covariance 3
AIC			
25 subjects	7083 (70.83)	1874 (18.74)	1043 (10.43)
50 subjects	7386 (73.86)	1738 (17.38)	876 (8.76)
100 subjects	7693 (76.93)	1512 (15.12)	795 (7.95)
200 subjects	7990 (79.9)	1292 (12.92)	718 (7.18)
500 subjects	8071 (80.71)	1198 (11.98)	731 (7.31)
1000 subjects	8090 (80.9)	1227 (12.27)	683 (6.83)
AICC			
25 subjects	7536 (75.36)	1706 (17.06)	758 (7.58)
50 subjects	7643 (76.43)	1622 (16.22)	735 (7.35)
100 subjects	7812 (78.12)	1451 (14.51)	737 (7.37)
200 subjects	8041 (80.41)	1264 (12.64)	695 (6.95)
500 subjects	8094 (80.94)	1186 (11.86)	720 (7.2)
1000 subjects	8102 (81.02)	1220 (12.2)	678 (6.78)
CAIC			
25 subjects	8826 (88.26)	1025 (10.25)	149 (1.49)
50 subjects	9218 (92.18)	726 (7.26)	56 (0.56)
100 subjects	9843 (98.43)	492 (4.92)	25 (0.25)
200 subjects	9723 (97.23)	265 (2.65)	12 (0.12)
500 subjects	9938 (99.38)	60 (0.6)	2 (0.02)
1000 subjects	9981 (99.81)	19 (0.19)	0 (0)
BIC			
25 subjects	8346 (83.46)	1306 (13.06)	348 (3.48)
50 subjects	8993 (89.93)	879 (8.79)	128 (1.28)
100 subjects	9334 (93.34)	607 (6.07)	59 (0.59)
200 subjects	9650 (96.5)	321 (3.21)	29 (0.29)
500 subjects	9911 (99.11)	81 (0.81)	8 (0.08)
1000 subjects	9956 (99.56)	39 (0.39)	5 (0.05)
HQIC			
25 subjects	7539 (75.39)	1699 (16.99)	762 (7.62)
50 subjects	8317 (83.17)	1262 (12.62)	421 (4.21)
100 subjects	8776 (87.76)	959 (9.59)	265 (2.65)
200 subjects	9134 (91.34)	678 (6.78)	188 (1.88)
500 subjects	9517 (95.17)	370 (3.7)	113 (1.13)
1000 subjects	9575 (95.75)	348 (3.48)	77 (0.77)

Table 4.16 Covariance Structure Selection for Model III: Covariance 2 using ML estimation

	Covariance Structure Selected		
	Covariance 1	Covariance 2	Covariance 3
AIC			
25 subjects	5342 (53.42)	3346 (33.46)	1312 (13.12)
50 subjects	3649 (36.49)	5015 (50.15)	1336 (13.36)
100 subjects	1391 (13.91)	7027 (70.27)	1581 (15.81)
200 subjects	164 (1.64)	8230 (82.3)	1606 (16.06)
500 subjects	0 (0)	8418 (84.18)	1582 (15.82)
1000 subjects	0 (0)	8373 (83.73)	1627 (16.27)
AICC			
25 subjects	5982 (59.82)	3039 (30.39)	979 (9.79)
50 subjects	3960 (39.6)	4878 (48.78)	1162 (11.62)
100 subjects	1484 (14.84)	7036 (70.36)	1479 (14.79)
200 subjects	173 (1.73)	8269 (82.69)	1558 (15.58)
500 subjects	0 (0)	8443 (84.43)	1557 (15.57)
1000 subjects	0 (0)	8378 (83.78)	1622 (16.22)
CAIC			
25 subjects	7953 (79.53)	1764 (17.64)	283 (2.83)
50 subjects	7788 (77.88)	2052 (20.52)	160 (1.6)
100 subjects	6073 (60.73)	3812 (38.12)	114 (1.14)
200 subjects	2704 (27.04)	7185 (71.85)	111 (1.11)
500 subjects	51 (0.51)	9872 (98.72)	77 (0.77)
1000 subjects	0 (0)	9949 (99.49)	51 (0.51)
BIC			
25 subjects	7210 (72.1)	2259 (22.59)	531 (5.31)
50 subjects	6773 (67.73)	2904 (29.04)	323 (3.23)
100 subjects	4840 (48.4)	4914 (49.14)	245 (2.45)
200 subjects	1834 (18.34)	7952 (79.52)	214 (2.14)
500 subjects	24 (0.24)	9836 (98.36)	140 (1.4)
1000 subjects	0 (0)	9897 (98.97)	103 (1.03)
HQIC			
25 subjects	5975 (59.75)	3019 (30.19)	1006 (10.06)
50 subjects	5020 (50.2)	4229 (42.29)	751 (7.51)
100 subjects	2712 (27.12)	6563 (65.63)	724 (7.24)
200 subjects	598 (5.98)	8692 (86.92)	710 (7.1)
500 subjects	3 (0.03)	9438 (94.38)	559 (5.59)
1000 subjects	0 (0)	9479 (94.79)	521 (5.21)

Table 4.17 Covariance Structure Selection for Model III: Covariance 3 using ML estimation

	Covariance Structure Selected		
	Covariance 1	Covariance 2	Covariance 3
AIC			
25 subjects	12 (0.12)	2 (0.02)	9985 (99.85)
50 subjects	0 (0)	0 (0)	10000 (100)
100 subjects	0 (0)	0 (0)	10000 (100)
200 subjects	0 (0)	0 (0)	10000 (100)
500 subjects	0 (0)	0 (0)	10000 (100)
1000 subjects	0 (0)	0 (0)	10000 (100)
AICC			
25 subjects	17 (0.17)	6 (0.06)	9976 (99.76)
50 subjects	0 (0)	0 (0)	10000 (100)
100 subjects	0 (0)	0 (0)	10000 (100)
200 subjects	0 (0)	0 (0)	10000 (100)
500 subjects	0 (0)	0 (0)	10000 (100)
1000 subjects	0 (0)	0 (0)	10000 (100)
CAIC			
25 subjects	218 (2.18)	67 (0.67)	9714 (97.14)
50 subjects	1 (0.01)	0 (0)	9999 (99.99)
100 subjects	0 (0)	0 (0)	10000 (100)
200 subjects	0 (0)	0 (0)	10000 (100)
500 subjects	0 (0)	0 (0)	10000 (100)
1000 subjects	0 (0)	0 (0)	10000 (100)
BIC			
25 subjects	69 (0.69)	28 (0.28)	9902 (99.02)
50 subjects	0 (0)	0 (0)	10000 (100)
100 subjects	0 (0)	0 (0)	10000 (100)
200 subjects	0 (0)	0 (0)	10000 (100)
500 subjects	0 (0)	0 (0)	10000 (100)
1000 subjects	0 (0)	0 (0)	10000 (100)
HQIC			
25 subjects	17 (0.17)	6 (0.06)	9976 (99.76)
50 subjects	0 (0)	0 (0)	10000 (100)
100 subjects	0 (0)	0 (0)	10000 (100)
200 subjects	0 (0)	0 (0)	10000 (100)
500 subjects	0 (0)	0 (0)	10000 (100)
1000 subjects	0 (0)	0 (0)	10000 (100)

Table 4.18 Model selection results using AIC criterion as the covariance structure selection for Model I with three different covariance structures

m	**AIC selecting true covariance**	**Adjusted REML KR R^2_β selecting true mean**	**Adjusted REML KR R^2_β selecting true mean and covariance**	**Adjusted REML Residual R^2_β selecting true mean**	**Adjusted REML Residual R^2_β selecting true mean and covariance**
Model I: Covariance 1					
25	0	41.81	0	51.88	0
50	74.32	8.36	0	50.95	38.01
100	77.62	7.84	0	51.28	39.91
200	79.96	8.01	0	51.43	41.01
500	85.36	6.38	0	51.94	44.66
1,000	100	0	0	51.88	51.88
Model I: Covariance 2					
25	33.67	37.31	27.72	47.22	15.98
50	50.28	56.7	45.26	48.03	24.76
100	69.67	81.52	66.59	48.72	34.22
200	82.86	97.56	82.18	47.93	39.74
500	84.01	99.97	84	48.82	41.03
1000	84.45	100	84.45	48.92	41.4
Model I: Covariance 3					
25	99.77	92.51	92.44	48.28	48.19
50	100	98.28	98.28	47.81	47.81
100	100	99.9	99.9	49.08	49.08
200	100	100	100	49.74	49.74
500	100	100	100	50.01	50.01
1,000	100	100	100	49.45	49.45

Table 4.19 Model selection results using AIC criterion as the covariance structure selection for Model II with three different covariance structures

m	**AIC selecting true covariance**	**Adjusted REML KR R_β^2 selecting true mean**	**Adjusted REML KR R_β^2 selecting true mean and covariance**	**Adjusted REML Residual R_β^2 selecting true mean**	**Adjusted REML Residual R_β^2 selecting true mean and covariance**
Model II: Covariance 1					
25	71.38	77.08	70.56	67.58	48.03
50	74.16	79.07	74.12	68.63	50.51
100	77.65	80.93	77.65	69.7	53.51
200	79.51	81.35	79.51	67.92	53.64
500	80.96	81.29	80.96	68.1	54.88
1,000	81.14	81.15	81.14	68.26	55.22
Model II: Covariance 2					
25	33.91	55.1	2.11	55.84	19.92
50	50.08	36.63	0.64	65.86	33.58
100	70.17	14.3	0.04	68.14	48.21
200	82.81	1.76	0	68.67	56.78
500	84.64	0	0	68.06	57.61
1000	83.9	0	0	68.52	57.52
Model II: Covariance 3					
25	99.71	6.31	6.31	58.43	58.31
50	100	1.52	1.52	65.76	65.76
100	100	0.08	0.08	68.05	68.05
200	100	0	0	68.33	68.33
500	100	0	0	68.33	68.33
1,000	100	0	0	68.56	68.56

Table 4.20 Model Selection Results using AIC criterion as the Covariance structure selection for Model III with three different covariance structures

m	**AIC selecting true covariance**	**Adjusted REML KR R_β^2 selecting true mean**	**Adjusted REML KR R_β^2 selecting true mean and covariance**	**Adjusted REML Residual R_β^2 selecting true mean**	**Adjusted REML Residual R_β^2 selecting true mean and covariance**
Model III: Covariance 1					
25	70.83	20.38	0.77	71.95	51.2
50	73.86	17.65	0.03	88.46	65.76
100	76.93	14.99	0	98.31	75.59
200	79.9	12.7	0	99.99	79.9
500	80.71	11.07	0	100	80.71
1,000	80.9	9.2	0	100	80.9
Model III: Covariance 2					
25	33.46	27.83	17.9	89.43	29.37
50	50.15	45.38	34.94	98.37	49.23
100	70.27	75.93	61.21	99.93	70.27
200	82.3	95.84	79.96	100	82.3
500	84.18	99.97	84.15	100	84.18
1000	83.73	100	83.73	100	83.73
Model III: Covariance 3					
25	99.85	21.7	21.7	90.38	90.24
50	100	18.12	18.12	98.87	98.87
100	100	9.97	9.97	99.97	99.97
200	100	2.94	2.94	100	100
500	100	0.12	0.12	100	100
1,000	100	0	0	100	100

Table 4.21 Model selection results using BIC criterion as the covariance structure selection for Model I with three different covariance structures

m	**BIC selecting true covariance**	**Adjusted REML KR R^2_β selecting true mean**	**Adjusted REML KR R^2_β selecting true mean and covariance**	**Adjusted REML Residual R^2_β selecting true mean**	**Adjusted REML Residual R^2_β selecting true mean and covariance**
Model I: Covariance 1					
25	0	41.89	0	51.92	0
50	89.8	1.69	0	51.21	45.93
100	93.81	0.78	0	51.33	48.22
200	96.39	0.39	0	51.52	49.66
500	99.13	0.14	0	52.06	51.57
1,000	100	0	0	51.88	51.88
Model I: Covariance 2					
25	23.22	22.29	18.66	46.86	10.62
50	29.07	28.59	26.03	47.78	14.39
100	49.1	49.13	46.9	48.69	24.72
200	80	81.37	79.35	47.7	38.19
500	98.5	99.66	98.48	48.83	48.16
1000	99.05	100	99.05	48.92	48.43
Model I: Covariance 3					
25	99.02	91.96	91.79	48.22	47.83
50	100	98.28	98.28	47.81	47.81
100	100	99.9	99.9	49.08	49.08
200	100	100	100	49.74	49.74
500	100	100	100	50.01	50.01
1,000	100	100	100	49.45	49.45

Table 4.22 Model selection results using BIC criterion as the covariance structure selection for Model II with three different covariance structures

m	**BIC selecting true covariance**	**Adjusted REML KR R_β^2 selecting true mean**	**Adjusted REML KR R_β^2 selecting true mean and covariance**	**Adjusted REML Residual R_β^2 selecting true mean**	**Adjusted REML Residual R_β^2 selecting true mean and covariance**
Model II: Covariance 1					
25	83.99	88.97	82.96	67.17	55.99
50	89.58	94.18	89.54	68.28	61.16
100	93.77	96.72	93.77	69.45	65.13
200	96.54	98.23	96.54	67.77	65.54
500	98.93	99.26	98.93	67.96	67.22
1,000	99.48	99.48	99.48	68.13	67.75
Model II: Covariance 2					
25	23.59	73.13	2.12	54.42	13.84
50	29.65	67.53	0.66	64.31	20.00
100	48.01	49.55	0.04	66.68	33.47
200	80.18	17.85	0	68.15	55.6
500	98.72	0.15	0	68.04	67.19
1000	99.09	0	0	68.53	67.92
Model II: Covariance 3					
25	99.09	7.03	6.31	58.4	57.95
50	100	1.52	1.52	65.76	65.76
100	100	0.08	0.08	68.05	68.05
200	100	0	0	68.33	68.33
500	100	0	0	68.33	68.33
1,000	100	0	0	68.56	68.56

Table 4.23 Model selection results using BIC criterion as the covariance structure selection for Model III with three different covariance structures

m	**BIC selecting true covariance**	**Adjusted REML KR R_β^2 selecting true mean**	**Adjusted REML KR R_β^2 selecting true mean and covariance**	**Adjusted REML Residual R_β^2 selecting true mean**	**Adjusted REML Residual R_β^2 selecting true mean and covariance**
Model III: Covariance 1					
25	83.46	9.70	0.86	72.22	60.55
50	89.93	4.29	0.05	88.65	89.93
100	93.34	2.35	0	98.36	91.82
200	96.5	0.88	0	99.99	96.49
500	99.11	0.21	0	100	99.11
1,000	99.56	0.1	0	100	99.56
Model III: Covariance 2					
25	22.59	16.51	10.68	89.95	19.77
50	29.04	21.87	19.12	98.48	28.48
100	49.14	44.56	42.35	99.93	49.11
200	79.52	79.25	77.13	100	79.52
500	98.36	99.73	98.33	100	98.36
1000	98.97	100	98.97	100	98.97
Model III: Covariance 3					
25	99.02	21.3	21.11	90.38	89.55
50	100	18.12	18.12	98.87	98.87
100	100	9.97	9.97	99.97	99.97
200	100	2.94	2.94	100	100
500	100	0.12	0.12	100	100
1,000	100	0	0	100	100

CHAPTER 5

SUMMARY AND DISCUSSION

5.1 Summary

This dissertation has focused on investigating the properties of R^2_β for fixed effects in the linear mixed model. R^2_β has many desirable features that make it worthwhile to explore. It has a semi-partial form and for special cases, there is a one-to-one correspondence to a multivariate measure of association. .

The first goal of this dissertation research was to examine the asymptotic properties of R^2_β using Kenward-Roger denominator degrees of freedom under the null and alternative hypothesis. The mean and variance of R^2_β are approximated using a Beta distribution and also using a Taylor series expansion. The asymptotic expectation and variance of R^2_β are shown to converge to the same value for both of these approaches. Test statistics based on these two approximations of the mean and variance are derived and compared to the overall F test for fixed effects in the linear mixed model. Using simulations, the Type I error rate of the proposed R^2_β test statistics derived from the Beta distribution was equivalent to the Type I error rate for the overall F test. The Type I error rates for the test statistic based on the Taylor series expansion moments were slightly inflated.

Another goal of this dissertation research was to examine the impact of covariance structure misspecification, estimation technique, and denominator degrees of freedom method on the finite sample properties of R^2_β. For the simulation studies examined, the estimation technique does not impact the values of R^2_β even for the smaller sample size simulations while varying the denominator degrees of freedom has a substantial impact on the values and asymptotic properties of R^2_β. Covariance structure misspecification also greatly impacts the values of R^2_β using Kenward-Roger containment and Satterthwaite degrees of freedom. Conversely, the values of R^2_β using residual degrees of freedom are not impacted by covariance structure misspecification. The great variation in R^2_β values for the misspecified models arises because the covariance structure misspecification impacts the denominator degrees of freedom being used to define R^2_β.

The finite sample properties of R^2_β is also considered which include evaluating R^2_β as a fixed effects model selection tool and evaluating the semi-partial R^2_β. One potentially troublesome feature is that the semi-partial R^2_β is larger than the model R^2_β when using Kenward-Roger denominator degrees of freedom and restricted maximum likelihood; while the semi-partial R^2_β does not exceed the model R^2_β for the residual denominator degrees of freedom. For purposes of fixed effects model selection, an adjusted version of R^2_β was created.

5.2 Conclusions

5.2.1 Denominator Degrees of Freedom Methods

The results from Chapters 2, 3 and 4 show that choosing which denominator degrees of freedom method used in defining R^2_β is critical. The denominator degrees of freedom method affects the values and asymptotic properties of R^2_β as seen in Chapter 2 and 3. In addition, covariance structure misspecification greatly impacts the R^2_β values as a result of changes in the denominator degrees of freedom as seen in Chapter 3. There was great variation seen in the denominator degrees of freedom depending upon the covariance structure specified. Chapter 4 highlights some of properties of R^2_β for the denominator degrees of freedom methods. Some of the denominator degrees of freedom methods used to define R^2_β exhibit problematic results.

The question remains: what denominator degrees of freedom method should be used to define R^2_β? When initially proposed, R^2_β was defined using Kenward-Roger method because of the performance of that method in small sample inference. This dissertation has shown some of the pitfalls of the Kenward-Roger method in defining R^2_β.

5.3 Future Work

While the large simulation study did include many different mean model and covariance settings, it still does have limitations. Future research, should investigate the results of this dissertation for other simulation conditions and settings which include cases of incomplete and/or unbalanced data. Complete and balanced data is not always common in the real world. Oftentimes, longitudinal data are mistimed and not complete.

There has been an increased interest in an R^2 statistic for fixed effects in the linear mixed model and there are many statistics available. These R^2 statistics for fixed effects in the linear mixed model are often being used in data analysis without a detailed examination of their properties. As our research has indicated, the properties of the statistic are essential in understanding how the R^2 statistic will perform. Future work should be done to compare R^2_β and other R^2 statistics and examine for which functions they should be used. For example, further examination of $R^2_{\beta_{\text{adj}}}(\nu_{\text{Res}})$ as a fixed effects model selection tool is necessary. In particular, determining how well the statistic performs in choosing fixed effects for model selection compared to other R^2 statistics and information criteria is important. These comparisons would inform researchers as to which statistic in the linear mixed model to choose as a fixed effects model selection tool.