AN APPROXIMATE F STATISTIC FOR TESTING POPULATION EFFECTS IN LONGITUDINAL STUDIES VIA MIXED MODELS

by

Dennis Dale Wallace and Ronald W. Helms

Department of Biostatistics, University of North Carolina at Chapel Hill, NC.

Institute of Statistics Mimeo Series No. 2115T

April 1993
AN APPROXIMATE F STATISTIC FOR TESTING POPULATION EFFECTS IN LONGITUDINAL STUDIES VIA MIXED MODELS

by

Dennis Dale Wallace

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 in the School of Public Health

Chapel Hill

1993

Approved by:

[Signatures]

Ronald Helm
Advisor

Jeffrey Shaw
Reader

Lawrence S. Rappaport
Reader
DENNIS DALE WALLACE. An Approximate F Statistic for Testing Population Effects in Longitudinal Studies Via Mixed Models. (Under the direction of Ronald W. Helms.)

ABSTRACT

This work focuses on developing and characterizing a statistic for testing contrasts among population effects and developing confidence regions for those effects using data from longitudinal studies. Historically, likelihood ratio or Wald-type statistics were used for such analyses, but those statistics produce very optimistic Type I error rates. McCarroll and Helms (1987) introduced an ad hoc F statistic ($F_H$) with reasonable Type I error rates, but no information was available on distributional properties of that statistic. This research substantially extends McCarroll and Helms' results by characterizing the distributional and numerical properties of an alternative form of $F_H$.

Longitudinal studies, which play a key role in medical, epidemiological, and environmental research, are designed to characterize patterns in experimental units' response over some longitudinal metameter and to investigate the effects of covariates on responses. Characteristics of longitudinal data from such research limit traditional univariate and multivariate methods. Often data are mistimed or irregularly timed, and missing observations result in incomplete data. This work establishes procedures that provide hypothesis tests and confidence intervals for such data using the Linear Mixed Model (MixMod).

An alternative parameterization of MixMod is developed, and a modified form of the $F_H$ statistic is derived. The Helms-McCarroll procedure is used to derive restricted maximum likelihood estimates for the covariance parameters using the concentrated likelihood (Callanan and Harville, 1988) to concentrate the error variance out of the estimating equations. Conditional on these estimates, the
distribution of $F_H$ is established for both the central and noncentral cases. As the covariance parameter estimates converge to the true covariance parameters, the value of $F_H$ is shown to converge to the value of $F_{WLS}$, the statistic that would be used if the covariance parameters were known.

Finally, the numerical behavior of $F_H$ is examined via computer-generated data sets. It is shown to provide reasonable inferential results as covariance parameter estimates are perturbed over their expected error range. Also, the statistics $F_H$ and $F_{REML}$, which can be obtained through classic REML procedures using standard statistical packages are shown to produce comparable numerical results.
ACKNOWLEDGEMENTS

First, I wish to express my sincere gratitude to my advisor, Dr. Ron Helms for his encouragement and guidance throughout this dissertation research. He provided intellectual stimulation and emotional support for this research, but just as importantly, he provided insight into what it means to be a professional biostatistician. I also want to thank my committee, Dr. Lloyd Edwards, Dr. Larry Kupper, Dr. P. K. Sen, and Dr. Carl Shy for their support and constructive comments on this dissertation.

On a personal note, I want to thank my family for their support and encouragement throughout my graduate career; I couldn’t have persevered without them. I am particularly indebted to my parents Ken and Ellen Wallace; their personal belief in the value of knowledge and education that they instilled in my siblings and me early in life has been a continuing force in my graduate program. Finally, I want to thank my wife, Carol Athey for her love, support, and unwavering patience throughout my dissertation.
TABLE OF CONTENTS

LIST OF TABLES ........................................ viii
LIST OF FIGURES ........................................ ix
LIST OF ABBREVIATIONS ................................. xiii

Chapter

I. INTRODUCTION AND LITERATURE REVIEW .............. 1
  1.1 INTRODUCTION ..................................... 1
  1.2 NOTATION AND MODEL STRUCTURE ..................... 6
  1.3 LIKELIHOOD-BASED PARAMETER ESTIMATION IN MixMod ... 16
    1.3.1 Fixed and Random Effects Estimation--
         Known Covariance Parameters .................. 18
    1.3.2 Estimation of Covariance Parameters .......... 20
    1.3.3 Properties of ML and REML Estimates .......... 26
  1.4 ALGORITHMS FOR LIKELIHOOD-BASED PARAMETER
       ESTIMATES ....................................... 31
    1.4.1 The Newton-Raphson Algorithm ................ 33
    1.4.2 The Scoring Algorithm ........................ 36
    1.4.3 The Method of Successive Approximations ...... 37
    1.4.4 The EM Algorithm ................................ 38
    1.4.5 Relative Comparison of the Algorithms .......... 40
  1.5 EXACT AND APPROXIMATE TEST STATISTICS .......... 41
    1.5.1 Hypothesis Test Statistics in GLUM ............ 42
    1.5.2 Hypothesis Test Statistics in GLMM ............ 45
    1.5.3 Hypothesis Test Statistics in MixMod .......... 51

II. DEVELOPMENT OF MixMod TEST STATISTIC ............. 54
  2.1 ALTERNATIVE PARAMETERIZATION OF MixMod .......... 56
  2.2 DERIVATION OF COVARIANCE PARAMETER ESTIMATES .... 57
LIST OF TABLES

Table 3.1: Summary of Distributional Terminology ... 79

Table 4.1: Population Parameters for Fairclough-Based Data ... 118

Table 4.2: Fixed Effect Parameters for the FVC-Mod2 Data ... 119

Table 4.3: Variance Parameters for the FVC-Mod2 Data ... 119

Table 4.4: Summary of the Parameter Estimates for the Fairclough-Based Data ... 121

Table 4.5: Summary of the Parameter Estimates for the FVC-Mod2 Data ... 124

Table A.1: Summary of the Fixed Effect Intercept Parameter Estimates ... 210

Table A.2: Summary of the Fixed Effect Slope Parameter Estimates ... 213

Table A.3: Summary of the Covariance Parameter Estimates ... 216

Table A.4: Summary of Estimated Test Statistics ... 217
| Figure 4.1: | Illustration of the FVC Linear Model | 148 |
| Figure 4.2: | Comparison of Intercept Variance Estimates Fairclough-Based Data | 149 |
| Figure 4.3: | Comparison of Intercept/Slope Covariance Estimates Fairclough-Based Data | 150 |
| Figure 4.4: | Comparison of Slope Variance Estimates Fairclough-Based Data | 151 |
| Figure 4.5: | Comparison of Intercept Variance Estimates Modified FVC Data | 152 |
| Figure 4.6: | Comparison of Intercept/Slope Covariance Estimates Modified FVC Data | 153 |
| Figure 4.7: | Comparison of Slope Variance Estimates Modified FVC Data | 154 |
| Figure 4.8: | F-Statistic Comparison HM vs WLS Fairclough-Based Data | 155 |
| Figure 4.9: | F-Statistic Comparison Classic REML vs WLS Fairclough-Based Data | 156 |
| Figure 4.10: | F-Statistic Comparison HM vs WLS Modified FVC Data | 157 |
| Figure 4.11: | F-Statistic Comparison Classic REML vs WLS Modified FVC Data | 158 |
| Figure 4.12: | P-Value Comparison HM vs WLS Fairclough-Based Data | 159 |
| Figure 4.13: | P-Value Comparison Classic REML vs WLS Fairclough-Based Data | 160 |
| Figure 4.14: | P-Value Comparison HM vs WLS Modified FVC Data (Full Scale) | 161 |
| Figure 4.15: | P-Value Comparison Classic REML vs WLS Modified FVC Data (Full Scale) | 162 |
| Figure 4.16: | P-Value Comparison HM vs WLS Modified FVC Data (Reduced Scale) | 163 |
| Figure 4.17: | P-Value Comparison Classic REML vs WLS Modified FVC Data (Reduced Scale) | 164 |
| Figure 4.18: | EDF Plot for AR Error in F, Fairclough-Based Data, Perturbation Set 1 | 165 |
Figure 4.47: P-Value Relative Error vs Mahalanobis Distance, Modified FVC Data, Perturbation Set 3  .  194

Figure 4.48: P-Value Actual Error vs Mahalanobis Distance, Modified FVG Data, Perturbation Set 3  .  195

Figure 4.49: Eigenvalue Error vs Mahalanobis Distance, Modified FVC Data, Perturbation Set 3  .  196

Figure 4.50: Comparison of HM and WLS F-Statistics, Modified FVC Data, Perturbation Sets 1 and 2  .  197

Figure 4.51: Comparison of HM and WLS p-Values, Modified FVC Data, Perturbation Sets 1 and 2  .  198

Figure 4.52: Comparison of HM and WLS F-Statistics, Modified FVC Data, Perturbation Set 3  .  199

Figure 4.53: Comparison of HM and WLS p-Values, Modified FVC Data, Perturbation Set 3  .  200

Figure 4.54: Comparison of Smaller HM and WLS p-Values, Modified FVC Data, Perturbation Sets 1 and 2  .  201

Figure 4.55: Comparison of Smaller HM and WLS p-Values, Modified FVC Data, Perturbation Set 3  .  202
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AE</td>
<td>Absolute Error—the difference between an estimated value and a reference value</td>
</tr>
<tr>
<td>ANOVA</td>
<td>Analysis of variance</td>
</tr>
<tr>
<td>AR Error</td>
<td>Absolute value of the relative error</td>
</tr>
<tr>
<td>ARMA</td>
<td>Autoregressive, moving average model</td>
</tr>
<tr>
<td>b.l.u.e.</td>
<td>Best linear unbiased estimator</td>
</tr>
<tr>
<td>BLUP</td>
<td>Best linear unbiased predictor</td>
</tr>
<tr>
<td>D_M</td>
<td>Mahalanobis distance</td>
</tr>
<tr>
<td>EDF</td>
<td>Empirical distribution function</td>
</tr>
<tr>
<td>EM</td>
<td>Expectation/maximization algorithm</td>
</tr>
<tr>
<td>FEV</td>
<td>Forced expiratory volume</td>
</tr>
<tr>
<td>FR</td>
<td>Full rank (matrix)</td>
</tr>
<tr>
<td>GLMM</td>
<td>General linear multivariate model</td>
</tr>
<tr>
<td>GLS</td>
<td>General least squares</td>
</tr>
<tr>
<td>GLUM</td>
<td>General linear univariate model</td>
</tr>
<tr>
<td>i.i.d</td>
<td>Independent and identically distributed</td>
</tr>
<tr>
<td>IML</td>
<td>Interactive matrix language</td>
</tr>
<tr>
<td>LTFR</td>
<td>Less than full rank (matrix)</td>
</tr>
<tr>
<td>LR</td>
<td>Likelihood ratio (hypothesis test)</td>
</tr>
<tr>
<td>MAR</td>
<td>Missing at random (missing data)</td>
</tr>
<tr>
<td>MCAR</td>
<td>Missing completely at random (missing data)</td>
</tr>
<tr>
<td>MixMod</td>
<td>General linear mixed model</td>
</tr>
<tr>
<td>ML</td>
<td>Maximum likelihood</td>
</tr>
<tr>
<td>MSA</td>
<td>Method of successive approximations</td>
</tr>
<tr>
<td>MVN</td>
<td>Multivariate normal (random vector)</td>
</tr>
<tr>
<td>NID</td>
<td>Normally and independently distributed</td>
</tr>
<tr>
<td>NR</td>
<td>Newton-Raphson</td>
</tr>
<tr>
<td>Acronym</td>
<td>Definition</td>
</tr>
<tr>
<td>---------</td>
<td>------------</td>
</tr>
<tr>
<td>OAR</td>
<td>Observed at random (missing data)</td>
</tr>
<tr>
<td>RE</td>
<td>Relative error--absolute error divided by the reference value</td>
</tr>
<tr>
<td>REML</td>
<td>Restricted (or residual) maximum likelihood</td>
</tr>
<tr>
<td>WLS</td>
<td>Weighted least squares</td>
</tr>
<tr>
<td>UMP</td>
<td>Uniformly most powerful (hypothesis test)</td>
</tr>
</tbody>
</table>
CHAPTER 1
INTRODUCTION AND LITERATURE REVIEW

1.1 INTRODUCTION

Longitudinal studies are broadly defined as studies for which a response variable, which may be either a vector or a scalar quantity, is observed on at least two occasions for each subject or experimental unit. Typically, these studies generate serial measurements indexed by time. However, the occasions of measurement may be indexed by some other naturally scaled metamer such as age, time on study, height of subject, or cumulative exposure to some risk factor. To simplify the presentation, the remainder of the document will use the term subject to denote the experimental units on which observations are made and time as the metamer on which the observations are indexed, even though these studies encompass much wider scenarios. In broad terms, the objectives of longitudinal studies are to characterize patterns of the subjects' response over time and to investigate the effects of covariates on these patterns (Ware, 1985). This research is concerned with the analyses of longitudinal data generated from designed experimental or observational studies and with the tools needed to develop designs for such studies. Particular attention is given to the evaluation of approximate test statistics that can be used to perform hypothesis tests and power calculations for longitudinal studies. The focus is on continuous response variables from designed experimental and observational studies. Neither longitudinal data generated from survey samples nor categorical response variables are considered in this research.
Longitudinal studies play an important role in medical research, epidemiology, and environmental research, as evidenced by the examples below. Fairclough and Helms (1984) evaluated the changes in pulmonary function in growing children. Pulmonary function was evaluated by trends in forced expiratory volume (FEV) and forced vital capacity (FVC) as children grew with the child's height rather than time as the longitudinal metameter. The study examined FEV and FVC behavior as a function of gender and race. Laird and Ware (1982) also used measurements of FEV in children over time to analyze the effects of air pollution on the pulmonary function development in children. Again, the child’s height, not time, was the longitudinal metameter. In another environmental epidemiologic study, Waternaux, Laird, and Ware (1989) examined the effect of prenatal lead concentrations on the IQ development in children. The children’s blood lead levels and adjusted mental development index was examined at 6, 12, and 18 months as a function of prenatal lead exposure. Here, the longitudinal metameter was time. Currently, Shy, et. al. (1993a,b) are examining the effects of air pollution from incineration processes on the respiratory function in children and adults. The response variable in this study is peak expiratory flow (PEF), and the longitudinal metameter is time. The primary objective of this ongoing study is to investigate the effects of the location of subjects relative to waste incineration systems and other specific measures of ambient air quality and air pollution measures on PEF over a 35-day time period. These examples illustrate the types of questions that can be explored with longitudinal studies.

At this point, basic terminology on longitudinal studies is introduced to simplify the subsequent discussion. A longitudinal study has a regularly timed schedule if measurements are scheduled at uniform time intervals on each subject, and it has regularly timed data if measurements are actually obtained at uniform time intervals on each subject.
(Helms 1992). Note that the uniform intervals can be unique for each subject and that time is used here to denote a general longitudinal metameter. Also, note that the longitudinal metameter is typically measured from some baseline level as a difference rather than on an absolute scale, irrespective of whether the measure is time or some other natural metameter. A longitudinal study has a consistently timed schedule if each subject has the same schedule (i.e., if each subject is scheduled to be measured at the same times, whether or not they are regularly timed) and has consistently timed data if all subjects are evaluated at the same times (Helms, 1992).

The design of studies with a regular and consistent schedule and the analyses for studies with regularly and consistently timed data can often be accomplished relatively easily with General Linear Multivariate Model (GLMM) or General Linear Univariate Model (GLUM) methods. However, in medical and environmental research, many studies are, at least in part, observational in nature. The timing of observations in such studies is not generally within the complete control of the investigator. Consequently, the study may generate mistimed or irregularly timed data, or missing observations may result in incomplete (or missing) data. In such cases, standard GLMM and GLUM techniques are inappropriate analytical tools, and alternative techniques must be used for the design of such studies and the analyses of data from them.

Of particular concern in such studies are the procedures for handling incomplete data generated when observations are missing. In essence, these scenarios generate inconsistently and irregularly timed data. The appropriate procedures for handling such data depend upon the mechanisms that lead to the missing data. Knowledge, or the absence thereof, of the mechanisms that led to certain values being missing is a key to choosing an appropriate analytical tool and to interpreting the results under such
conditions (Little and Rubin, 1987). At this point, some additional terminology on missing data is helpful. Let Y be a univariate response variable with a set of cofactors X, and allow Y to have missing values with X assumed to be complete. If the missing data mechanism is independent of both Y and X, then the missing data are said to be missing at random (MAR); furthermore, the observed data are observed at random (OAR), and more compactly, the missing data are classified as being missing completely at random (MCAR) (Little and Rubin, 1987). If the missing data mechanism is dependent on X but independent of Y the missing data are said to be MAR, and if the missing data are dependent on Y the data are neither MAR or OAR. A final class of missing data includes those data that are specifically excluded by the study design; these data are said to be missing by design. A common example of data that are missing by design are the cells in a factorial design that are excluded in a Latin Squares type design.

Little and Rubin (1987) reviewed ad hoc methods that have been used historically to evaluate data with missing values for multivariate data problems like those encountered with longitudinal studies. The three methods considered were complete-case analysis, available-case analysis, and simple forms of imputation. The complete-case (also called casewise deletion) analyses confine attention to those subjects that have no missing observations. The method has the advantage of being simple and can be performed with readily available computer packages, but it can discard a substantial amount of information. Available-case methods are generally limited to regularly and consistently timed data. For observations collected at a particular time, these methods use all available data to estimate mean values and variances, and use pairwise available data to estimate covariances between two observation periods. These methods recover more information than complete-case methods, but have deficiencies that limit their usefulness in practice.
In particular, available-case methods can generate covariance matrix estimates that are not positive definite. Finally, a variety of techniques have been used to impute or fill in the missing response variables. Little and Rubin (1987) found these methods to be of limited value because (a) they all require the missing data to be MCAR; (b) their performance is unreliable; (c) frequently they require ad hoc adjustments to yield acceptable estimates; and (d) situations when the methods work and when they fail are not clearly distinguishable.

Given the limitations of these ad hoc methods for dealing with missing data, likelihood-based approaches provide an attractive alternative for handling such problems. Conceptually, the vector of response variables for a study with missing data can be viewed as a complete data vector that has been partitioned into two components—the observed responses and the missing responses. If the conceptual complete data vector has the property that the marginal likelihood of the observed data is equal to the likelihood of the observed data conditioned on the missing data, then the missing data are said to be ignorably missing. Note that if data are MAR or if the data are missing by design, they are ignorably missing. This research will focus on the use of likelihood-based methods in the context of longitudinal studies with ignorably missing data.

While a substantial body of literature has appeared over the last 10 years on the analyses of incomplete data from longitudinal studies, little attention has been given to missing data in the design of longitudinal experiments. Missing data can affect experimental design in two ways. First, failure to consider the effect of unintentionally missing data on subsequent data analyses when historical information suggests that comparable studies typically generate incomplete data, may lead to overly optimistic power estimates. Second, experimental designs with
intentionally missing data may have advantages over traditional complete designs. Such designs are called intentionally incomplete designs. Helms (1992) has shown that in some cases such intentionally incomplete designs have benefits in terms of both power and cost in comparison to the standard complete design of the same "size." This research provides further examination of appropriate test statistics that can be used to develop and evaluate such designs.

1.2 NOTATION AND MODEL STRUCTURE

The primary likelihood-based methods for evaluating data from longitudinal studies in which the response of interest is a continuous random variable to be considered in this research are based on the General Linear Mixed Model, hereafter denoted as "MixMod." This subsection first introduces notation for a very general linear model form for the situation in which multiple observations are made on the same subject. Then the MixMod notation is introduced as a natural extension of this general form. While this research is concerned primarily with the use of MixMod in the evaluation of longitudinal data, MixMod techniques typically must be compared with techniques based on GLMM and GLUM. Consequently, standard terminology for these two models will also be introduced. The first subsection below introduces the general linear model notation and then specifies notation for MixMod, GLMM, and GLUM. The second subsection defines terminology that will be used to specify hypotheses and presents the typical MixMod hypothesis structure. The final subsection presents some alternative models for MixMod variance component structures.

1.2.1 General Linear Model Notation

Consider the following linear model for the observations made on a single subject in a longitudinal study:
\[ y_k = x_k \beta + e_k \]  \hspace{1cm} (1.2.1)

where

- \( y_k \) is an \( N_k \times 1 \) vector of observations on the \( k \)th subject \( k = 1, 2, \ldots, K \),
- \( \beta \) is a \( q \times 1 \) vector of unknown parameters,
- \( x_k \) is an \( N_k \times q \) known, constant design matrix, corresponding to the fixed effects \( \beta \),
- \( e_k \) is an \( N_k \times 1 \) vector of random error terms.

The following assumptions are made about the model:

1. \( e_k \sim \text{NID}(0, \Sigma_k) \) so \( e_i \) is independent of \( e_j \) for \( i \neq j \).
2. \( \Sigma_k \) is a positive-definite symmetric matrix whose elements are twice differentiable functions of a finite number of constant unknown parameters \( \tau_1, \tau_2, \ldots, \tau_m \), i.e., \( \Sigma_k = \Sigma_k(\tau) \), where \( \tau \in \mathbb{T} \), the set of parameter values that make \( \Sigma_k \) positive definite.
3. Parameters in \( \beta \) are functionally independent of those in \( \tau \).

This very general model encompasses most of the models that are used in analyses of longitudinal studies. For this model, \( E[y_k] = x_k \beta \) and \( \text{Var}[y_k] = \Sigma_k \). The three models described below (MixMod, GLMM, and GLUM) can be viewed as an extension of this very general form.

The model equations for MixMod are:

\[ y_k = x_k \beta + z_k d_k + e_k, \quad k = 1, 2, \ldots, K \]  \hspace{1cm} (1.2.2)

and

\[ y = x \beta + zd + e \]  \hspace{1cm} (1.2.3)

where

- \( y_k \) is an \( N_k \times 1 \) vector of observations of a response variable from the \( k \)-th subject, \( k = 1, 2, \ldots, K \);
- \( Y = [y_1//y_2//\ldots//y_K] \) is the \( N \times 1 \) vector of response variable measures for all \( K \) subjects for which the
total vector size is defined by \( N = \sum_{k=1}^{K} N_k \);

\( X_k \) denotes the \( N_k \times q \) design matrix for the fixed effects for the \( k \)-th subject with the elements of \( X_k \) assumed to be known constants that are measured without appreciable error;

\( X = [X_1//X_2//...//X_K] \) is the \( N \times q \) design matrix of fixed effects for the full model where \( N > q \).
(\emph{Note that if \( \text{Rank}(X) = q \), then the fixed effects component of the model is said to be \textit{full rank} (FR); otherwise the fixed effects component of the model is said to be \textit{less than full rank} (LTFR).});

\( Z_k \) is an \( N_k \times r \) design matrix for the random effects for the \( k \)-th subject with the elements of \( Z_k \) assumed to be known constants that are measured without appreciable error;

\( Z = \text{Diag}(Z_1, Z_2, ..., Z_K) \) is the \( N \times Kr \) random effects design matrix for the model. (\emph{Note that this matrix has a block diagonal form that provides some advantages for computations used in the analyses described in subsequent sections. As with the fixed effects components of the model, the random effects component is said to be FR if \( \text{Rank}(Z) = Kr \), and LTFR otherwise});

\( \beta \) is the \( q \times 1 \) vector of unknown, constant \textit{fixed effect primary parameters};

\( d_k \) is an \( r \times 1 \) vector on unknown, unobservable \textit{random effects}. (\emph{Note that these random effects are subject specific, but that the vector size is consistent from subject to subject.});

\( d = [d_1//d_2//...//d_K] \) is the \( Kr \times 1 \) vector of unknown, unobservable \textit{subject-specific random effects} for the entire model;

\( e_k \) is an \( N_k \times 1 \) vector of \textit{within-subject random errors}; and
\[ e = [e_1/e_2/\ldots/e_k] \text{ is the vector of random errors for the complete model.} \]

The distributional assumptions for the model are:

\[ d_k \sim \text{NID}_r(0, \Delta) \quad (1.2.4) \]

independently of

\[ e \sim N_n(0, \sigma^2 \mathbf{V}) \quad (1.2.5) \]

where

\[ \Delta = \text{Var}(d_k) \text{ is the covariance matrix of the random effects;} \]
\[ \sigma^2 \text{ is an unknown, scalar within-subject error variance parameter;} \]
\[ \mathbf{V}_k = \text{Var}(e_k)/\sigma^2 \text{ is a positive definite, symmetric matrix with parameters that may be either known or unknown; and} \]
\[ \mathbf{V} = \text{Diag}(\mathbf{V}_1, \mathbf{V}_2, \ldots, \mathbf{V}_k). \]

Relative to the above specifications, first note that NID means "normally and independently distributed." Also note the relatively strong assumption of between subject homoscedasticity for the random effects. Generally, both \( \mathbf{V} \) and \( \Delta \) will be assumed to be functions of a vector of parameters \( \tau \), where \( \tau_i, i=1, 2, \ldots, m \) are not functionally dependent on \( k \). Frequently, \( \mathbf{V}_k \) will have the simple form \( \sigma^2 I_{N_k} \), where \( I_{N_k} \) is the identity matrix of order \( N_k \). Under the general assumptions outlined above, we have that for any specific pair of subjects indexed by \( k \) and \( k' \):

\[ \text{Cov}[d_{k'}, d_k] = 0 \]
\[ \text{Cov}[d_{k'}, e_k] = 0 \]
\[ \text{Cov}[e_{k'}, e_k] = 0 \quad (1.2.6) \]

and that the expectation and variance for the fixed effects and random effects are:
\[ E \begin{bmatrix} \mathbf{d}_k \\ \mathbf{e}_k \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \]

and

\[ \text{Var} \begin{bmatrix} \mathbf{d}_k \\ \mathbf{e}_k \end{bmatrix} = \begin{bmatrix} \Delta & 0 \\ 0 & \sigma^2 \mathbf{v}_k \end{bmatrix} \]

Consequently, the expectation and variance for \( \mathbf{Y}_k \) and \( \mathbf{Y} \) are:

\[ E(\mathbf{Y}_k) = \mathbf{X}_k \mathbf{\beta} \]

with

\[ \text{Var}(\mathbf{Y}_k) = \mathbf{Z}_k \Delta \mathbf{Z}_k' + \sigma^2 \mathbf{v}_k \]

and

\[ E(\mathbf{Y}) = \mathbf{X} \mathbf{\beta} \]

with

\[ \text{Var}(\mathbf{Y}) = \mathbf{\Sigma} = \text{Diag}(\Sigma_1, \Sigma_2, \ldots, \Sigma_k) \]

The model equations for GLMM are:

\[ \mathbf{Y}_k' = \mathbf{X}_k' \mathbf{\beta} + \mathbf{e}_k', \; k=1,2,\ldots,K \]

and

\[ \mathbf{Y} = \mathbf{X} \mathbf{\beta} + \mathbf{e} \]

where

\( \mathbf{Y}_k \) is an \( N_k \times 1 \) vector of observations a response variable from the the \( k \)-th subject with the additional assumption that \( N_1 = N_2 = \ldots = N_k \);

\( \mathbf{Y} = [\mathbf{Y}_1'//\mathbf{Y}_2'//\ldots//\mathbf{Y}_K'] \) is the \( K \times n \) matrix of response variable measures for all subjects where \( n = N_k \) is the number of longitudinal observations on each subject;

\( \mathbf{X}_k \) is the \( q \times 1 \) design vector for the \( k \)-th subject;

\( \mathbf{X} = [\mathbf{X}_1||\mathbf{X}_2||\ldots||\mathbf{X}_k]' \) is the \( K \times q \) design matrix for all subjects combined;

\( \mathbf{\beta} \) is the \( q \times n \) matrix of unknown, constant parameters for GLMM;
\( e_k \) is an \( n \times 1 \) vector of within-subject random errors; and
\[ e = [e_1 \parallel e_2 \parallel \ldots \parallel e_k]' \] is the matrix of random errors for the \( K \) subjects combined.

The distributional assumptions of the model can be summarized as:
\[ e_k \sim \text{NID}_n(0, \Sigma) \] (1.2.12)

where \( \Sigma \) is the \( n \times n \) covariance matrix for the within-subject error terms. Note the strong assumptions of between subject independence and homoscedasticity. Now, let \( Y' = \text{Vec}(Y') = [Y_1' \parallel Y_2' \parallel \ldots \parallel Y_k'] \). Under the above assumptions, the expectation and variance of \( Y' \) are:

\[ E(Y') = \text{Vec}((X\beta)') \]
\[ = (I_n \otimes X_1' \parallel \ldots \parallel I_n \otimes X_k') \ast \text{Vec}(\beta) \] (1.2.13)

and

\[ \text{Var}(Y') = I_n \otimes \Sigma \]

for this expanded form of GLMM.

The model equations for GLUM are:
\[ Y_k = X_k\beta + e_k \]
and
\[ Y = X\beta + e \] (1.2.14)

where the symbols are defined identically to the definitions associated with (1.2.1) and (1.2.3). The GLUM distributional assumptions can be summarized by:

\[ e_k \sim N(0, \sigma^2 I_{k_k}) \]
and
\[ e \sim N(0, \sigma^2 I_n) \] (1.2.15)

Imbedded in these formulae are the very strong assumptions of within-subject and between-subject homoscedasticity. The model also assumes that all subjects have the same fixed
effect parameters.

1.2.2 Secondary Parameters and Hypotheses in MixMod

Most analyses of longitudinal data in the MixMod framework are concerned with studying how the mean response of subjects changes over time and, in many cases with how treatment variables, exposure variables, or risk factors affect such changes (Helms, 1991a). These analyses are often based on secondary parameters derived from the fixed effects. In some cases, tests of hypotheses concerning a particular realization of the random effects also may be of interest. Hence, the framework for hypothesis testing outlined here is in a general form that encompasses analyses of fixed and random effects individually or in combination.

Let δ and δ_k be the single realization analog of d and d_k, respectively. The analyses addressed in this document will focus on sets of linear contrasts of the fixed effect parameters, but for the moment, consider more general secondary parameters of the form:

$$\theta = C_1 \beta + C_2 \delta - \theta_0$$  (1.2.16)

where C_1, C_2, and \theta_0 are known, fixed constants, and both C_1 and C_2 are defined to make \theta estimable. Note that \theta, a scalar parameter is estimable if and only if there is a vector a \in \mathbb{R}^n such that a'Y is an unbiased estimate of \theta (Arnold, 1981). Furthermore, the parameter C_1 \beta is estimable only if the rows of C_1 are linearly independent elements of the row space of X, and the parameter \theta is estimable only if C_1 \beta is estimable (McClean, Sanders, and Stroup, 1991). Hypotheses about \theta are testable only if \theta is estimable. (Note that although these definitions hold in the general case described above, the focus of this research is on estimable parameters of the form C_1 \beta.) Under these assumptions the hypotheses of interest are Boolean functions of \theta of the form (Helms, 1988b):
\[ H_0(\theta) = B(\theta = 0) = \begin{cases} \text{TRUE iff } \theta = 0 \\ \text{FALSE iff } \theta \neq 0 \end{cases} \] (1.2.17)

versus

\[ H_a(\theta) = -H_0(\theta) = B(\theta \neq 0) = \begin{cases} \text{TRUE iff } \theta \neq 0 \\ \text{FALSE iff } \theta = 0 \end{cases} \] (1.2.18)

As appropriate, one-tailed hypotheses can be defined with similar terminology. Note that this general formulation can be used for addressing only fixed effects by allowing \( C_2 \) to be an identically zero matrix of appropriate size.

1.2.3 Alternative MixMod Covariance Structure Models

As outlined in Section 1.3.1, the estimation of fixed and random effects in MixMod is reasonably straightforward if the covariance matrix \( \Sigma \) is known. However, because \( \Sigma \) is typically not known, it must be estimated before estimates for the fixed effects and random effect realizations can be developed. Also, in some cases the covariance structure is of interest in its own right. In order to estimate \( \Sigma \), some assumption must be made about its structure. Three common forms of the covariance structure that have been considered in the analyses of longitudinal data are the unstructured covariance matrix, a covariance matrix with linear structure, and a covariance matrix with some type of nonlinear autocovariance structure. Each of these general forms is described briefly in the paragraphs below.

For the unstructured covariance matrix, no assumptions are made about its elements other than that \( \Sigma \) must be a positive definite, symmetric matrix. Hence, in the most general form of the MixMod, an unstructured covariance matrix, \( \text{Var}(Y) \), could theoretically have as many as \( N(N-1)/2 \) unique elements. In the general scenario, this approach to modelling the covariance structure is intractable in that the large number of covariance parameters that must be estimated leads to imprecise and inefficient estimates of the fixed and random effects. However, for the special case in which the longitudinal study yields complete and
consistently timed data and between subject homoscedasticity can be assumed, reasonable estimates of $\Sigma$ ($p \times p$) can be obtained within the GLMM framework. These procedures are well known, and estimation procedures are described in McC Carroll and Helms (1987).

Because the unstructured covariance matrix is intractable in many situations, one alternative that reduces the number of variance components to be estimated is the linear covariance structure. This structure, which is described in detail in Andrade and Helms (1986a,b) and Hocking (1985), requires $\Sigma$ to be of the following form:

$$
\Sigma = \sum_{m=1}^{M} \tau_m G_m
$$

(1.2.19)

where

- $\tau_m$ denote fixed, unknown variance components that are defined such that $\Sigma$ is positive definite and symmetric,
- $G_m$ are known, constant symmetric matrices

This formulation has the advantage of encompassing a wide variety of scenarios while substantially reducing the number of covariance parameters that must be estimated.

While the linear covariance structure is widely applicable, this class of models does not include autoregressive models that are often used to characterize the covariance of serial data in time series settings. A general autoregressive, moving average (ARMA) structure for the error covariance matrix was considered by Rochon and Helms (1985 and 1987). In its most general form, this model assumes an ARMA ($p,q$) form for the covariance structure. Hence the error terms for the $i$-th subject have the form:

$$
\phi(B) e_{i,t} = \eta(B) \epsilon_{i,t}
$$

(1.2.20)

using a backshift operator, $B$, defined by Box and Jenkins (1970) and two functions $\phi$ and $\eta$ to represent the autoregressive and moving average components of the model,
respectively. The operator and functions are defined as:

\[ B^k(e_t) = e_{t-k} \]  \hspace{1cm} (1.2.21)

and

\[ \phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p \]
\[ \eta(B) = 1 - \eta_1 B - \cdots - \eta_q B^q \]  \hspace{1cm} (1.2.22)

In discussing this nonlinear covariance structure, Murray and Helms (1990) indicate that this general model represents an error term as a function of the p previous error terms for that subject (the autoregression component of the model) in conjunction with a moving average of q random shocks. This model, which does not contain any random effects, allows missing values, but it does not accommodate mistimed data. However, Murray and Helms (1990) examined a special case of this model, the AR(1) model, which can accommodate both mistimed and randomly missing data. For this AR(1) model \( Y_k \) has a covariance structure of the form:

\[
\text{Var}(Y_k) = \sigma^2 \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 & \cdots \\ \rho & 1 & \rho & \rho^2 & \cdots \\ \rho^2 & \rho & 1 & \rho & \cdots \\ \rho^3 & \rho^2 & \rho & 1 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix} \]  \hspace{1cm} (1.2.23)

Note that this model has two relatively strong assumptions—
(a) that the observations are homoscedastic; and (b) that the correlations depend upon a single parameter \( \rho \), the correlation between any two successive observations (Helms, 1990). This structure is inherently nonlinear and cannot be modelled using a MixMod formulation with linear covariance structure that is the focus of this research. However, this model does arise in many situations that involve serial data, particularly those that conform to a traditional time series pattern. These AR(1) covariance structures will not be considered in detail in this research, but such models can arise in longitudinal studies of the type considered here. As such, possible test statistics for longitudinal
data from studies with small to moderate sample sizes that conform to AR(1) structures provide a possible avenue for further study.

1.3 LIKELIHOOD-BASED PARAMETER ESTIMATION IN MIXMOD

While the ad hoc methods described in Section 1.1.1 have been used successfully to evaluate the results from studies with incomplete data, they do have inherent limitations. Any specific method is generally associated with a particular type of problem and may not be readily generalizable. Also, these methods are of limited usefulness at the design stage. A general likelihood-based approach can be used to address a broad variety of problems, and this approach lends itself to consideration of incomplete data at the design stage of a study. However one constraint of these procedures is that detailed assumptions about the underlying probability structure of the data are required.

This research focuses on data from longitudinal studies that can be modeled with MixMod, which means that the data can be represented with a model like (1.2.3). Also, the focus of this research is on situations for which a linear covariance structure as described in (1.2.19) can be assumed. Within such a framework, likelihood-based methods can be used to obtain parameter estimates for the fixed and random effects, as well as covariance parameters, as needed, with incomplete as well as complete data. Under the assumption of normality for both the random effects and the error terms the likelihood function for a MixMod (1.2.3) for all subjects combined has the form (Hocking, 1985):
\[ l = l(y) = \log[L(y)] \]
\[ = -\frac{N}{2} \log(2\pi) - \frac{1}{2} \log |\Sigma| \]
\[ - \frac{1}{2} (y - x\beta)' \Sigma^{-1} (y - x\beta) \]
\[ = c - \frac{1}{2} \log |\Sigma| - \frac{1}{2} (y - x\beta)' \Sigma^{-1} (y - x\beta) \]

in which the variables in the model are defined as in (1.2.2) through (1.2.9). For a particular realization of the random effects (i.e., for a particular realization \( \delta \) of \( \delta \)), the likelihood function has the form (Hocking, 1985 and Harville, 1976):

\[ l^* = l(y) = \log[L(y)] \]
\[ = c - \frac{1}{2} \sigma^2 |v| \]
\[ - \frac{1}{2\sigma^2} (y - x\beta - z\delta)' v^{-1} (y - x\beta - z\delta) \]

These likelihood equations are used to obtain maximum likelihood (ML) estimates of the fixed effects and of a particular realization of the random effects. They also are used to derive both ML and restricted maximum likelihood (REML) estimates of the covariance parameters. The subsections below discuss the estimation of the fixed effects and a specific realization of the random effects when \( \Sigma \) is assumed to be known (Section 1.3.1), describe ML and REML procedures for estimating parameters when \( \Sigma \) is unknown (Section 1.3.2), and examine the properties of the fixed and random effects parameter estimates when ML and REML procedures are used (Section 1.3.3). Note that the estimating equations presented in the following subsections assume that the covariance matrix, \( \Sigma \), is positive definite and that the design matrices, \( X \) and \( Z \), are FR; consequently, appropriate matrix inverses exist. In general, however, if these design matrices are LTFR, comparable procedures that use generalized inverse matrices can be used to obtain
1.3.1 Fixed and Random Effects Estimation—Known Covariance Parameters

Typically, a convenient place to begin examining the estimation of model parameters is to consider the case for which the vector $\tau$ of covariance parameters is known. Although this scenario is not typically encountered in practice, the results from this case frequently provide insight into reasonable estimation procedures for cases in which these parameters are unknown.

Suppose that the vector of covariance parameters $\tau$ is known. The likelihood equation (1.3.1) is maximized by any value $\hat{\beta}$, which satisfies the normal equations:

$$ (X'\Sigma^{-1}X)\hat{\beta} = X'\Sigma^{-1}y $$

(1.3.3)

Furthermore, suppose $C_i\beta$ is estimable (i.e., if $C_i' = X'\zeta_i$ for $\zeta_i$ an arbitrary $N \times c$ vector where $c = \text{Rank}(C_i) =$ the number of rows of $C_i$). If $\hat{\beta}$ is a particular solution of normal equations (1.3.3), then an extended version of the Gauss-Markov theorem says that the system of equations in (1.3.3) is consistent and that $C_i\hat{\beta}$ is an essentially-unique best linear unbiased estimator (b.l.u.e.) of $C_i\beta$ in the sense that if $c + r'y$ is a linear unbiased estimator of $C_i\beta$, then $\text{Var}(C_i\hat{\beta}) \leq \text{Var}(c + r'y)$ with equality holding if and only if $c + r'y = C_i\hat{\beta}$ with probability 1 (Harville, 1976).

While many analyses with MixMod are concerned primarily with evaluation of the fixed effects, cases do arise for which inferences about a realization of the random effects are of interest. (For example, in medical research, one such problem is related to how well children "track" with respect to some response variable such as height or weight.) If $\beta$ were known, a reasonable estimator (or alternatively, a reasonable predictor [Harville, 1985 and 1988]) of $C_2\delta$ is $C_2\tilde{\delta}$, where:
\[ \hat{\delta} = \Delta \Sigma^{-1} (Y - X\hat{\beta}) \] (1.3.4)

This estimator has intuitive appeal in that when \( Y \) is normally distributed per the MixMod assumptions, \( \hat{\delta} \) is the conditional expectation of \( \mathbf{d} \) given \( Y \) (Harville, 1976). Harville (1976) went on to demonstrate that even when \( \beta \) is not known, the essentially-unique b.l.u.e. for \( C_1\beta + C_2\delta \), when the parameter is estimable (i.e., when \( C_1\beta \) is estimable), is \( C_1\hat{\beta} + C_2\hat{\delta} \) where \( \hat{\beta} \) is a solution of (1.3.3) and where:

\[ \hat{\delta} = \Delta \Sigma^{-1} (Y - X\hat{\beta}) \] (1.3.5)

This estimator is also known as the best linear unbiased predictor (BLUP) of \( C_1\beta + C_2\mathbf{d} \) (Henderson, 1984 and McClean, Sanders, and Stroup, 1991).

The structure of MixMod allows further simplification of the calculations required to solve (1.3.3) and (1.3.5). First, the block diagonal structure of the covariance matrix allows the solution to be formulated in terms of the subject-specific design and covariance matrices as follows (Laird and Ware, 1982):

\[ \hat{\beta} = \left( \sum_{k=1}^{K} X'_k \Sigma_k^{-1} X_k \right)^{-1} \sum_{k=1}^{K} X'_k \Sigma_k^{-1} Y_k \] (1.3.6)

and

\[ \hat{\delta}_k = \Delta \Sigma_k^{-1} (Y_k - X_k\hat{\beta}) \] (1.3.7)

Also, Harville (1976) developed results for very general conditions under which the solutions to (1.3.3) and (1.3.5) were equivalent to a system of linear equations that could be solved relatively easily if \( \Delta \) and \( \mathbf{V} \) have simple forms, as frequently encountered. Subsequently, Harville (1977) presented these equations in a more simplified form that is applicable to the MixMod system considered here. The system
of equations can be represented as:

\[
\begin{bmatrix}
  x'(\sigma^2\nu)^{-1} & x'(\sigma^2\nu)^{-1}z \Delta \\
  z'(\sigma^2\nu)^{-1} & I + z'(\sigma^2\nu)^{-1}z \Delta
\end{bmatrix}
\begin{bmatrix}
  \hat{\beta} \\
  \hat{\nu}
\end{bmatrix}
= \begin{bmatrix}
  x'(\sigma^2\nu)^{-1}y \\
  z'(\sigma^2\nu)^{-1}y
\end{bmatrix}
\tag{1.3.8}
\]

where

\[
\delta = \Delta \hat{\nu} \quad \text{and} \quad \hat{\nu} = z'\Sigma^{-1}(y - x\hat{\beta})
\tag{1.3.9}
\]

These equations have computational advantages over the normal equations in that they do not require the inversion of either \(\Sigma\) or \(\Delta\). Also, \(\nu\) is frequently assumed to be an identity matrix, and under such an assumption, no matrix inversion is required to solve the alternative system of equations.

1.3.2 Estimation of Covariance Parameters

The methods described in Section 1.3.1 provide some insight into the estimation of fixed and random effects, but typically they are not directly applicable because the vector of variance components, \(\tau\), is not known. A standard technique in such situations is to use a consistent estimate of these variance components and proceed to use the methods outlined above to estimate fixed and random effects. This section describes two likelihood based approaches for generating consistent estimates of the variance components, ML and REML, that have been suggested as reasonable candidates that provide a unified approach for estimating \(\beta\), \(\delta\) and \(\tau\) (Laird and Ware, 1982).

Using the general form of the likelihood equation (1.3.1), Andrade and Helms (1984a,b) demonstrated that the ML estimates for the \(\tau_m\), \(m=1,2,\ldots,M\) are the solutions to the system of \(M\) equations of the form:

\[
(y - x\beta)\Sigma^{-1}\frac{\partial \Sigma}{\partial \tau_m} \Sigma^{-1}(y - x\beta) = \text{tr} \left[\Sigma^{-1} \frac{\partial \Sigma}{\partial \tau_m}\right]
\tag{1.3.10}
\]

where the elements of \(\Sigma = \Sigma(\tau)\) comprise twice differentiable functions of elements of \(\tau\).

For the special case in which \(\Delta\) has a linear covariance structure, Andrade and Helms (1984a) demonstrated that the
ML estimates for $\tau$ are can be found from the system of equations:

$$\hat{\tau} = \left[\sum_{k=1}^{K} \text{tr}\left[\hat{\Sigma}_k^{-1}G_{ks}\hat{\Sigma}_k^{-1}G_{kt}\right]\right]^{-1} \cdot \left[\sum_{k=1}^{K} \text{tr}\left[\hat{\Sigma}_k^{-1}G_{ss}\hat{\Sigma}_k^{-1}(Y_k - X_k\hat{\beta})\right]\right]_{st}$$  \hspace{1cm} (1.3.11)

where $\langle f(s,t) \rangle_{st}$ represents a matrix with $f(s,t)$ in the $s$-th row and $t$-th column, and $\langle f(s) \rangle_s$ represents a vector with the $s$-th element being $f(s)$. As will be described in Section 1.3.3, these ML estimators have the reasonable asymptotic properties. However, the estimators are biased for finite samples, in part because they do not account for the degrees of freedom lost in estimating the fixed effects.

Patterson and Thompson (1971) proposed an alternative method for estimating covariance parameters that accounts for this loss in degrees of freedom—the restricted maximum likelihood method. In essence this procedure relies on using the error space of the fixed effects design matrix to estimate the covariance parameters. The procedure can be justified from both classical sample theoretic and Bayesian perspectives. The paragraphs below first outline the procedure within a sample theoretic framework, as outlined by Hocking (1985) and then present a Bayesian justification for the procedure as described by Harville (1974) and Laird and Ware (1982).

The sample theoretic justification for the REML first suggested by Patterson and Thompson (1971) relies on modifying the ML procedure through a factorization of the likelihood function. Note that this procedure as developed by Patterson and Thompson (1971) and outlined by Hocking (1985) addressed fixed effects models such as those found in GLUM or GLMM. However, the techniques can be applied in a relatively straightforward fashion to obtain estimates for parameters of MixMod. Because the problem can be addressed
in a more intuitive manner in the framework of a univariate model, the initial discussion presents REML estimation procedures for a univariate model with an arbitrary covariance structure. The discussion is then extended to MixMod.

Consider a univariate model of the form shown in (1.2.1) with the stipulation that elements of $\Sigma_k$ are twice differentiable functions of a set of parameters $\tau$. An error contrast is defined as a linear combination $\lambda'y$ of the observations such that $E(\lambda'y) = 0$. Note that this condition is equivalent to the requirement that $\lambda$ is in the null space of $X'$, i.e., that $X'\lambda = 0$ in this model. (Also, note that an error contrast has the same interpretation in MixMod when a specific realization of the random effects, $d$, is not being considered.) The REML estimate of $\tau$ is based on maximizing the likelihood function of a full rank set of error contrasts $T'y$ with respect to $\tau$, rather than maximizing the likelihood function of the complete observation vector $y$. Mathematically, this process can be accomplished by letting:

$$y^* = Py$$

(1.3.12)

where $P$ and $y^*$ are partitioned as:

$$P = \begin{bmatrix} P_1 \\ P_2 \end{bmatrix} \quad \text{and} \quad y^* = \begin{bmatrix} y_1^* \\ y_2^* \end{bmatrix}$$

(1.3.13)

and where

$$P_1 = I - X(X'X)^{-1}X' \quad \text{and} \quad P_2 = X\Sigma^{-1}$$

(1.3.14)

Hocking (1985) goes on to demonstrate that (a) $y_1^*$ and $y_2^*$ are independent; (b) the distribution of $y_1^*$ does not depend functionally on $\beta$; and (c) that even though $y_1^*$ has a singular normal distribution of rank $N-p$, the degeneracy can be addressed by looking at any set of $N-p$ independent rows of $P_1$, say $P_{11}$, to generate the vector $y_{11}^* = P_{11}y$. The
estimates of \( \tau \) are then based on maximizing the likelihood function of \( y_{11}^* \). Note that this procedure is analogous to that of estimating \( \sigma^2 \) using the error sums of squares in standard ANOVA.

While the partitioning of the likelihood as described above is generally used to explain REML estimates, Harville (1974) demonstrated that the procedure has a natural Bayesian interpretation. The likelihood function for \( y_{11}^* \) developed above was shown to be proportional to the posterior density for \( \tau \) when \( \beta \) and \( \tau \) are assumed to be independent and the prior density for \( \beta \) is such that the components are distributed independently and uniformly over the real line (i.e., a vague prior is assumed). Hence the REML estimate for \( \tau \) is equivalent to the mode of its marginal posterior density ignoring any prior information about \( \beta \) (Harville, 1974). Harville (1974) also demonstrated that a convenient form of a likelihood function that can be used to obtain REML estimates is:

\[
1_r = c - \frac{1}{2} \log |X' \Sigma^{-1} X| - \frac{1}{2} \log |\Sigma| - \frac{1}{2} (y - XB)' \Sigma^{-1} (y - XB)
\]  

Estimates for \( \tau \) can be obtained by maximizing this function to obtain a set of likelihood equations analogous to those shown in (1.3.10).

While REML estimates can be developed for MixMod, the procedures are less straightforward than they were for the univariate model. The added difficulty arises because the random effects contribute to both the estimation (or first moment) space and the error (or second moment) space. Hence, we no longer have strict orthogonality between the spaces. McCarrroll and Helms (1987) and Laird and Ware (1982) present two approaches to REML estimation in MixMod which generate somewhat different forms of the estimating equations. Each of these approaches and the estimation equations that resulted from them are summarized briefly.
below. Both approaches are based on a general form of MixMod comparable to that shown in (1.2.2) through (1.2.9). McC Carroll and Helms (1987) approached REML estimation in MixMod via an algebraic approach in the case of a covariance matrix with linear structure. Using standard MixMod notation, let \( R = [X \mid Z] \) with rank(\( X \)) = \( a \leq \) rank(\( R \)) = \( b \). An orthogonal transformation of \( Y \) can then be partitioned into three components as shown below (McC Carroll and Helms, 1987):

\[
Y^* = T'Y = \begin{bmatrix}
T_1' & Y_1^* \\
T_2' & Y_2^* \\
T_3' & Y_3^*
\end{bmatrix}
\]

where \( T_1 = [t_1, t_2, \ldots, t_a] \) is an orthonormal basis for the column space of \( X \) [denoted as \( M(X) \)], \( T_2 = [t_{a+1}, t_{a+2}, \ldots, t_b] \) is an orthonormal basis for \( M(R) \mid M(X) \) (see Arnold, 1981, p39 for a discussion of orthogonal subspaces and the bar notation), and \( T_3 = [t_{b+1}, t_{b+2}, \ldots, t_n] \) is an orthonormal basis for the orthogonal complement \( M(R) \) in \( \mathbb{R}^n \) [denoted by \( M(R)^\perp \)]. Note that the expectation and variance of this transformed variable are:

\[
E(Y^*) = \begin{bmatrix}
T'X\beta \\
0 \\
0
\end{bmatrix}
\]

and

\[
\text{Var}(Y^*) = T' \text{Var}(Y) T
\]

\[
= \begin{bmatrix}
Z_1'\Delta^*Z_1' - \sigma^2 I_a & Z_1'\Delta^*Z_2' & 0 \\
Z_2'\Delta^*Z_1' & Z_2'\Delta^*Z_2' + \sigma^2 I_{b-a} & 0 \\
0 & 0 & \sigma^2 I_{n-b}
\end{bmatrix}
\]

where \( Z_i = T_i'Z \) and \( \Delta^* = I_x \otimes \Delta \). Note that \( Y_3^* \) is independent of \( Y_2^* \) and \( Y_1^* \) and that the distribution of \( Y_3^* \) depends only on the parameter \( \sigma^2 \). Consequently, McC Carroll and Helms proposed a REML estimate of \( \sigma^2 \) based on the pure error space (i.e., the
realization of $Y_3^* = Y_3^*$) and an ML estimate of the
covariance parameters $\tau_\theta=(\tau_1, \ldots, \tau_m)'$ based on the column
space of $R$. The resulting estimates are (McCarroll and
Helms, 1987):

$$\hat{\sigma}_{\text{REML}}^2 = \frac{(Y_3^*, Y_3^*)}{(N-b)} \quad (1.3.19)$$

and

$$\hat{\tau}_1 = \left[ \left( \text{tr}(\hat{\Sigma}_1 G_1 \hat{\Sigma}_1^{-1} G_1^*) \right) \right]^{-1} \left[ \left( \text{tr}(\hat{\Sigma}_1^{-1} G_1^* \hat{\Sigma}_1^{-1} UU^*) \right) \right] \quad (1.3.20)$$

where

$$\Sigma_1 = \left[ \sum_{m=1}^{m-1} \tau_m G_m^* \right] + \sigma^2 I_b$$

and

$$U = \begin{bmatrix} Y_1^* - T_1 X \beta \\ Y_2^* \end{bmatrix}$$

with

$$G_m^* = \begin{bmatrix} Z_1 (I_x \otimes G_m) Z_1' & Z_1 (I_x \otimes G_m) Z_2' \\ Z_2 (I_x \otimes G_m) Z_1' & Z_2 (I_x \otimes G_m) Z_2' \end{bmatrix}$$

Note that the terminology used here is equivalent to that
defined for the ML estimates in (1.3.10) and (1.3.11).

In contrast to McCarroll and Helms (1987), Laird and Ware (1982) and Laird, Lange, and Stram (1987) used a
Bayesian approach to develop REML estimates that follows
directly from the REML procedures developed by Harville
(1976 and 1977). They produced results for MixMod with an
arbitrary structure for $\Lambda$; consequently, $\tau$ has $r(r+1)/2$
elements. In developing REML estimates, Laird and Ware
(1982) treat MixMod conceptually as a two stage model. The
general form of the model for each individual is equivalent
to the one shown in (1.2.2). At the first stage, $e_k$ is
defined as before but both $\beta$ and $d_k$ are assumed to be fixed.
The second stage then treats the model as defined before.

From the Bayesian perspective, the second stage of the model
is modified to let $\beta \sim N(0, \Sigma)$. The empirical Bayes strategy is then to estimate $\tau$ by integrating the likelihood over $\beta$ and $d$, letting $\Sigma^{-1} = 0$, and maximizing the marginal likelihood of $\tau$ given $y$ (Laird and Ware, 1982). The resulting estimating equations become (Laird and Ware, 1982):

$$\delta^2 = \hat{\tau}_1 / N$$

and

$$\Delta = \hat{\tau}_2 / K$$

where

$$\hat{\tau}_1 = E \left[ \sum_{k=1}^{K} e_k^{'} e_k | y_k, \hat{\tau} \right]$$

$$= \sum_{k=1}^{K} e_k (\hat{\tau})^{'} e_k (\hat{\tau}) + \text{tr} \left[ \text{var} (e_k | y_k, \hat{\tau}) \right]$$

and

$$\hat{\tau}_2 = E \left[ \sum_{k=1}^{K} d_k d_k^{'} | y_k, \hat{\tau} \right]$$

$$= \sum_{k=1}^{K} \hat{d}_k (\hat{\tau}) \hat{d} (\hat{\tau})^{'} + \text{var} (b_k | y_k, \hat{\tau})$$

Note that for the ML and REML estimation systems described above, the sets of equations typically have no explicit solution; the solution for $\beta$ and $\delta$ requires the solution for $\tau$, and vice versa. Consequently, iterative procedures are needed to solve these systems. Alternative procedures that can be used are explored in Section 1.4.

1.3.3 Properties of ML and REML Estimates

Information about the distributional properties of the MixMod estimators and about test statistics that are derived from these estimators are needed for both the design of longitudinal studies and the analyses of data from these studies. While information is needed on both the asymptotic and finite sample properties of these estimators, this
subsection addresses the asymptotic properties because most
information in the literature focuses on these properties.
Also, results discussed here typically relate to the general
linear model of the form:

$$y = X\beta + e$$

(1.3.22)
because of the prevalence of such results in the literature.
However, these results can be extended to the MixMod
framework with appropriate regularity conditions as noted by
Andrade and Helms (1984) and in recent work by Cressie and

Typically, examination of the asymptotic properties of
the general linear model estimators has taken one of two
approaches. One approach assumes normality of the error
terms and uses the resultant multivariate normality of the
observation vector to develop the formulations of the ML
estimators and to derive their asymptotic properties. This
approach is used by Andrade and Helms (1984a,b), Magnus
(1978), and Don and Magnus (1980). The alternative is to
assume the errors are independently and identically
distributed and use least squares procedures for estimation.
(Note that under normality these two estimation procedures
are equivalent.) Then general Central Limit Theorem
techniques are used to derive the asymptotic properties of
the estimators. This approach is exemplified by Arnold
(1981). The paragraphs below describe the results obtained
under these alternative approaches.

Asymptotic properties of ML estimators for the general
linear model under the assumption of normality of errors
have been studied extensively. Of particular interest for
this research are results related to models with an assumed
linear covariance structure. Andrade and Helms (1984a)
provided a comprehensive summary of the properties of ML
estimators for such models. They demonstrated that under
mild regularity conditions for a fixed effects model with
complete data, the joint ML estimates of $\beta$ and $\tau$ are known
to be asymptotically normal with the following mean and covariance structure:

$$\sqrt{N} \begin{bmatrix} \beta - \beta_0 \\ \hat{\tau} - \tau \end{bmatrix} \xrightarrow{D} N\left(0, \begin{bmatrix} \mathbf{v}_1 & 0 \\ 0 & \mathbf{v}_3 \end{bmatrix} \right)$$  \hspace{1cm} (1.3.23)

where

$$\mathbf{v}_1 = \left(\mathbf{x}' \mathbf{\Sigma}^{-1} \mathbf{x}\right)^{-1}$$  \hspace{1cm} (1.3.24)

$$\mathbf{v}_3 = \left(\text{tr}\left(\mathbf{\Sigma}^{-1} \mathbf{G}_s \mathbf{\Sigma}^{-1} \mathbf{G}_t\right)\right)_{st}$$

Andrade and Helms (1984a) also demonstrated that for incomplete data, the data vector could be partitioned into separate populations, each with a characteristic pattern of missing data. Each population is characterized by a matrix \( \mathbf{M}_i \), that comprises a subset of rows of an identity matrix that "picks off" the elements of \( \mathbf{y} \) that are not missing. Andrade and Helms (1984a) go on to show that the observed populations of incomplete data can be represented as \( \mathbf{y}_i = \mathbf{M}_i \mathbf{y} \), \( i = 1, \ldots, L \), where \( \mathbf{y} \) represents a complete vector of observations for a subject. In such cases the known covariance matrices have form \( \mathbf{M}_i \mathbf{G}_s \mathbf{M}_i' \), and the estimates of parameters for the separate populations have asymptotic distributions analogous to those shown in (1.3.17) and (1.3.18). Andrade and Helms (1984a) develop asymptotic distributions for the nonnull cases for both complete and incomplete data scenarios. While these results were developed for ML estimates of \( \beta \) and \( \tau \) in GLUM, the results can be demonstrated to hold within the MixMod framework when the fixed and random effects estimates are developed by ML techniques as described earlier (McCarroll and Helms, 1987). Historically, it has been assumed that similar asymptotic results also hold when the covariance parameters are developed with REML techniques. Recent results by Cressie and Lahiri (1991) suggest that asymptotic results comparable to those that are available for ML estimators do hold for REML estimators but that the REML results require additional
the regularity conditions beyond those needed to establish the ML results. Cressie and Lahiri (1991) demonstrate that if certain regularity conditions related to the smoothness of the covariance parameter functions and the characteristics of its eigenvalues are satisfied, then for \( R_n(\tau) \) defined as the expected information matrix of \( \tau \) (Cressie and Lahiri 1991):

\[
[R_n(\hat{\tau})]^{1/2} (\hat{\tau}_n - \tau) \sim N_n(0, I)
\]  

(1.3.25)

Magnus (1978) addressed a slightly more general problem than the problem treated by Andrade and Helms (1984a,b). Starting with the general linear model as given in (1.3.16) and the additional assumption that the covariance matrix is dependent upon a finite number of parameters, say \( \tau_1, \ldots, \tau_m \), with no additional structure assumed, Magnus examined the asymptotic properties of the ML estimates of the regression coefficients and the covariance parameters. Magnus (1978) demonstrated that under the assumptions that (1) \( e \) is normally distributed with \( E(e) = 0 \) and \( E(ee') = \Omega \) where \( \Omega \) is a positive definite matrix with elements that are twice differentiable functions of the \( \tau_m \); (2) \( X \) is fixed and FR; (3) the parameters in \( \beta \) are independent of those in \( \tau \); and (4) the \( m \) vectors vec(\( \Omega^{-1}/\partial \tau_1 \)), \ldots, vec(\( \Omega^{-1}/\partial \tau_m \)) are linearly independent, then the model has the ML equations equivalent to those shown in (1.3.3) and (1.3.10). Furthermore, the information matrix for the likelihood function is:

\[
\Psi = \begin{bmatrix}
X'\Omega^{-1}X & 0 \\
0 & \frac{1}{2} \Psi_0
\end{bmatrix}
\]  

(1.3.26)

where \( \Psi_0 \) is a symmetric, \( m \times m \), nonsingular matrix of the form:
\begin{equation}
\Psi_0 = \left( \text{tr} \left( \frac{\partial \Omega^{-1}}{\partial \tau_s} \Omega \frac{\partial \Omega^{-1}}{\partial \tau_t} \Omega \right) \right)_{st} \tag{1.3.27}
\end{equation}

An iterative procedure is available for calculating the ML estimators and conditions under which the estimators obtained under this procedure are consistent, asymptotically normal and asymptotically efficient have been established (Magnus, 1978). However, in practice these conditions are difficult to demonstrate, even for relatively common problems (Murray and Helms, 1990). In a similar fashion, Don and Magnus (1980), under the same set of four assumptions listed above, examined the properties of estimators produced by the following iterative scheme:

\begin{align*}
\Omega_j^{-1} &= \Omega^{-1}(\tau_j), \\
b_j &= (x'\Omega_j^{-1}x)^{-1}x'\Omega_j^{-1}y, \tag{1.3.28} \\
e_j &= y - xb_j, \\
\tau_{j+1} &= \tau(e_j)
\end{align*}

using some starting value \( \tau_0 \) is used to start the iteration. Under a reasonably strong assumption about the behavior of the likelihood function and the covariance matrix, which appears to be difficult to demonstrate in practice, the estimates \( b_j \) are demonstrated to be unbiased estimates of \( \beta \), and the iterative procedure outlined above is demonstrated to lead to the ML estimates of \( \beta \), which also are demonstrated to be unbiased.

An alternative approach to evaluating the asymptotic properties of estimators for the general linear model outlined in (1.3.15) was presented by Arnold (1981). This approach partitions the data space into orthogonal components in a fashion comparable to the procedure described for REML estimation. Assume that \( y_n \) is an \( n \)-dimensional vector with \( p \), the number of elements in \( \beta \), fixed and that \( e_n = (e_1, \ldots, e_n) \) with the \( e_i \) independently and
identically distributed with mean 0 and finite variance \( \sigma^2 \). Define \( P_x \) as the projection matrix from \( \mathbb{R}^n \) onto the estimation space generated by \( \mathbf{X} \) (i.e., onto \( M(\mathbf{X}) \) the column space of \( \mathbf{X} \)), and assume that the largest diagonal element of \( P_x \) goes to 0 as \( n \to \infty \) ("Huber’s Condition"). (Note that for this particular model, \( P_x \) is \( \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \), and Huber’s Condition is satisfied if the maximum element in this matrix goes to 0 as \( n \) becomes large.) Arnold makes use of the Cramer-Wold Theorem and the Lindeberg-Feller Theorem to demonstrate that:

\[
\frac{\|P_x e_n\|^2}{\sigma^2} \xrightarrow{d} \chi^2_p(0) \tag{1.3.29}
\]

This result depends only on the error terms in the model being i.i.d. with 0 mean and finite variance and does not require normality (Arnold, 1981, Chapter 10). Arnold goes on to demonstrate that the standard F-test is asymptotically size \( \alpha \) and that the power of the test is unaffected by nonnormality. However, he does suggest that the rate of convergence is affected by the degree of departure from normality (Arnold, 1981, Chapter 10).

Arnold (1980) also examines the asymptotic properties of tests and confidence intervals used in the framework of GLMM. Using similar arguments, he shows that if \( P_x \) represents the projection matrix for the standard GLMM framework and \( m(P_x) \) is the maximum element in absolute value in this projection matrix and if \( m(P_x) \to 0 \), then (1) the size of any invariant test of a multivariate linear hypothesis is asymptotically unaffected by nonnormality; and (2) any simultaneous confidence intervals for the parameter estimates are similarly unaffected (Arnold, 1981, Chapter 19).

1.4 ALGORITHMS FOR LIKELIHOOD-BASED PARAMETER ESTIMATES

In the MixMod problem with one or more unknown
covariance parameters, neither the ML or REML likelihood equations have closed form explicit solutions for estimating \( \tau \). Also, as noted in Section 1.3.1, the least squares equations for the fixed effect parameter estimates depend upon the covariance parameters, and the ML and REML equations for the covariance parameters depend upon the fixed effects. Consequently, iterative techniques are typically needed to estimate both \( \tau \) and \( \beta \). Four general iterative methods that are commonly used by practitioners to obtain ML and REML estimates for the covariance parameter estimates are the Newton-Raphson (NR) method, Fisher's scoring algorithm, the EM algorithm, and ad hoc techniques that fall under the umbrella of the method of successive approximations (MSA) (Callanan and Harville, 1988, and Jennrich and Schluchter, 1986). Typically, these covariance parameter estimation procedures are combined alternatively with the GLS equations for obtaining estimates of the fixed effects to obtain overall solutions to the likelihood equations.

Each of the algorithms described below requires calculation of the score, and most also require calculation of the Hessian of the likelihood. Consequently, the first and second partial derivatives of the likelihood with respect to the elements of \( \beta \) and \( \tau \) are needed. These derivatives have been calculated for the complete likelihood function, the REML likelihood, and the profile likelihood for the MixMod [the likelihood for the model defined in (1.2.2) and (1.2.3)] with \( \sigma^2 \) factored out). Formulae for these derivatives appear widely in the literature and will not be repeated here. (See for example Jennrich and Schluchter (1986), Lindstrom and Bates (1988), and Laird, Lange, and Stram (1987).)

The first four subsections below briefly describe the four different techniques. Included in each of these sections is a description of the algorithm, a discussion of modifications to the techniques that have been proposed in
the literature to improve the convergence properties of the algorithms, and a brief review of some strengths and limitations of the different algorithms. A final subsection then summarizes the relative merits of the different algorithms.

1.4.1 The Newton-Raphson Algorithm

The NR method is a well-established method for determining the stationary points of an arbitrary function and can be viewed as an application of Newton's method for finding roots of a system of nonlinear equations (Callanan and Harville, 1988). For calculating ML and REML estimates in MixMod, the method is applied to find the roots of the system of equations:

\[
\frac{\partial l}{\partial \tau} = 0 \text{ or } \frac{\partial l}{\tau} = 0 , \tag{1.4.1}
\]

the likelihood equations for ML and REML estimates respectively. The (r+1)st iterate of the NR algorithm is (Bard, 1974):

\[
\tau^{(r+1)} = \tau^{(r)} + \left[ \frac{\partial^2 l}{\partial \tau \partial \tau} \right]^{-1} \left[ \frac{\partial l}{\partial \tau} \right] \tag{1.4.2}
\]

where \( <f>_{st} \) represents a matrix with scalar \( f \) in the s-th row and t-th column and vector function \( \frac{\partial l}{\partial \tau} \) is the vector of partial derivatives of the likelihood function with respect to the elements of \( \tau \) using the notation of Searle (1982).

The NR method has been widely used to obtain likelihood-based estimates for the MixMod problem. In comparison to the alternative algorithms, the NR algorithm is reported to converge in relatively fewer iterations, but the computing time for each iteration can be relatively longer (Jennrich and Schluchter, 1986). In addition to the relative cost per iteration, other concerns that have been raised about the NR algorithm are that convergence is not guaranteed, the
estimate of the covariance matrix that is obtained from the algorithm at each step is not necessarily nonnegative definite, and the convergence sequence of the algorithm is dependent upon the particular parameterization of the algorithm. A variety of modifications to the NR algorithm have been reported to enhance the performance of the algorithm and to address these concerns, at least in part.

For the MixMod formulation given in (1.2.2) and (1.2.3), the likelihood function can be factored in such a way that the ML and REML estimates of \( \sigma^2 \) can be obtained as functions of \( \beta \) and \( \tau \) from (Lindstrom and Bates, 1988):

\[
\hat{\sigma}_{\text{ML}}^2 = \frac{1}{N-K} \sum_{k=1}^{K} r'_k \Sigma_k^{-1}(\tau) r_k
\]

\[
\hat{\sigma}_{\text{REML}}^2 = \frac{1}{N-K} \sum_{k=1}^{K} r'_k \Sigma_k^{-1}(\tau) r_k
\]

where,

\[
r_k = y_k - X_k \beta
\]

Callanan and Harville (1988) used a comparable procedure to form what they called the "concentrated likelihood" for different model parameterizations, which they then maximized. These procedures reduce the computation time of the NR algorithm by reducing the dimension of the parameter to be estimated by one.

Lindstrom and Bates (1988) suggest three additional modifications for improving the performance of the NR algorithm. First, the value of \( \beta \) at the \( i \)-th iteration \((\beta^{(i)})\) is replaced with the GLS estimate \( \hat{\beta}(\tau^{(i)}) \), which is reported to speed convergence slightly and facilitate computation of the derivatives at the next iteration. Second, the QR decomposition of the concatenated design matrix was used to calculate the necessary derivatives. Relative to the actual matrix, the decompositions are less susceptible to round-off error, use storage more efficiently, and speed execution by reducing computations.
(Lindstrom and Bates, 1988). Finally, the log-likelihood was maximized with respect to the Cholesky factor of $A$ rather than $A$, itself. This step ensures positive definite estimates of $A$ and dramatically improves the convergence properties of the algorithm (Lindstrom and Bates, 1988). In contrast to the QR decomposition used by Lindstrom and Bates (1988), Wolfinger, Tobias, and Sall (1991) use a sweep operator on an appropriate cross products matrix to calculate necessary derivatives. They chose this procedure because although it is numerically less stable than the QR method, it requires much less computing time (Wolfinger, Tobias, and Sall, 1991).

Callanan and Harville (1991) proposed a procedure that is conceptually much different from those described above for improving the convergence of the NR algorithm. It is well known that the rate of convergence of Newton’s procedure, of which the NR algorithm is a special case, is related to the departure of the equations from linearity. Hence, the concept is to devise modified versions of the likelihood equations that are "more nearly linear" than the original versions and apply Newton’s method to these modified equations (Callanan and Harville, 1991). Callanan and Harville (1991) describe a linearizing procedure that for "balanced data" in the MixMod framework will produce single iteration convergence of the NR algorithm. This modification is then applied in a straightforward manner to the general likelihood equations in an attempt to generate more nearly linear equations. Callanan and Harville (1991) found that this linearized version of the NR algorithm converged in substantially fewer iterations for two sets of animal birth-weight data than did the competing algorithms, which included the standard NR algorithm as well as EM, Scoring and MSA algorithms. They did note that the NR, Scoring, and linearized NR algorithms required about 60 percent more CPU time per iteration than did the EM and MSA algorithms. However, such comparisons should be viewed
cautiously because computation times depend strongly on how the different algorithms are coded (Jennrich and Schluchter, 1986).

1.4.2 The Scoring Algorithm

The Scoring algorithm is a gradient algorithm that is similar to the NR algorithm. The general form of the i-th iteration of the Scoring algorithm is identical to that for the NR algorithm shown in (1.4.1) and (1.4.2) except that the Hessian matrix of the second-order partial derivatives is replaced with its expectation. The primary advantage of the Scoring algorithm over the NR algorithm is that the expectation can often have a simplified structure such as a block diagonal structure, making matrix inversion easier (Murray and Helms, 1990). However, the Scoring algorithm suffers many of the same deficiencies as the NR algorithm. In particular, convergence is not assured and the algorithm can converge to a well-defined ML or REML estimate for $\Sigma$ that is not nonnegative definite (Jennrich and Schluchter, 1986 and Harville, 1977). Harville (1977) does indicate that three techniques—the penalty technique, the gradient projection technique, and the transformation technique—can be used to modify the algorithm to take into account the constraints on the covariance parameters. Jennrich and Schluchter (1986) also describe a procedure for enhancing the diagonal of the expectation of the Hessian at each step to assure that the its inverse is positive definite. In comparing the performance of the Scoring algorithm to the NR algorithm, Jennrich and Schluchter (1986) found the Scoring algorithm to have a lower cost per iteration but found that it sometimes required a substantially greater number of iterations. However, the Scoring algorithm is generally more robust to poor starting values than is the NR algorithm. Consequently, they suggest a compromise of starting with a Scoring algorithm and then switching to a NR algorithm after several steps (Jennrich and Schluchter, 1986).
1.4.3 The Method of Successive Approximations

The two procedures described above are second-order procedures in that they require the calculation of the Hessian matrix or its expectation. The MSA algorithms are first-order procedures in that they require calculation of only the first-order partial derivatives of the log likelihood. In essence the different MSA algorithms are based on manipulation the likelihood equations for the covariance parameters into the form:

$$\tau = g(\tau; y)$$

(1.4.5)

for some $K \times 1$ vector $g$ of functions, typically nonlinear, of $\tau$ (Harville, 1977). The MSA algorithm consists of the iterative procedure (Harville, 1977):

$$\hat{\tau}^{(i+1)} = g(\hat{\tau}^{(i)}; y)$$

(1.4.6)

For a special case of MixMod in which $\text{Var}(d_k)$ is equal to $\sigma^2_k I$ and $\text{Var} \, e_k = \sigma^2_{k-1} I$, Callanan and Harville (1988) identified three alternative MSA algorithms:

$$\hat{\sigma}^{2(i+1)} = [h_1(\hat{\sigma}^{2(i)}), \ldots, h_{K+1}(\hat{\sigma}^{2(i)})]'$$

(1.4.7)

$$\hat{\sigma}^{2(i+1)} = [\epsilon_1(\hat{\sigma}^{2(i)}), \ldots, \epsilon_k(\hat{\sigma}^{2(i)}), h_{K+1}^*(\hat{\sigma}^{2(i)})]'$$

$$\hat{\sigma}^{2(i+1)} = [h_1(\hat{\sigma}^{2(i)}), \ldots, h_k(\hat{\sigma}^{2(i)})', h_{K+1}^*(\hat{\sigma}^{2(i)})]'$$

where the $h_k$ and $\epsilon_k$ are functions of the random effect sums of squares, $k=1, \ldots, K$; $h_{K+1}^*$ is a function of the sums of squares of the residuals; $h_{K+1}^*$ is a function of the inner product of the data and the residuals. These MSA algorithms can be linearized with procedures analogous to those to that were were used to linearize the NR algorithm, thereby increasing their rate of convergence (Callanan and Harville, 1991). In evaluating the linearized NR and MSA algorithms, Callanan and Harville (1991) found that both algorithms behaved better than the original versions and than the EM and Scoring algorithms. Typically, the linearized MSA algorithms required more iterations to converge but less
computing time per iteration than did the linearized NR algorithms. Also, the linearized versions of both algorithms appeared to be less sensitive to starting points and to enhance determining when convergence was achieved in comparison to their nonlinearized counterparts (Callanan and Harville, 1991).

1.4.4 The EM Algorithm

The EM algorithm is an iterative method for obtaining ML estimates in the face of incomplete data that was developed by Dempster, Laird, and Rubin (1977) as a generalization of the method proposed earlier by Orchard and Woodbury (1972) and Sundberg (1976). The primary results of the original paper were that (1) when the likelihood is a member of the exponential family, the value of the likelihood is nondecreasing at each iteration of the algorithm and (2) under a fairly nonrestrictive set of regularity conditions, the algorithm converges to a point of local maximum in the closure of the parameter space. Subsequently, Wu (1983) and Boyles (1983) noted an error in the proof of the first convergence theorem presented by Dempster, Laird, and Rubin (1977). Wu (1983) went on to demonstrate that an EM sequence does increase the value of the likelihood function at each iteration and that if the sequence is bounded from above, then in the case of regular or curved exponential families, the sequence does converge to a stationary value of the likelihood function. However, he was unable to demonstrate that the likelihood function achieved its unique maximum over the parameter space at the point at which the sequence converged. Boyles (1983) demonstrated conditions under which the EM sequence converges to a compact connected set (not necessarily a point) of local maxima.

The EM algorithm is a general-purpose, iterative procedure that can be used to obtain likelihood-based parameter estimates for a wide variety of incomplete data problems. The algorithm requires iteration between two
steps--an expectation (E) step and a maximization (M) step. Conceptually, the algorithm presumes the existence of two sample spaces, a complete data space say $X$ and an incomplete data space say $Y$, and a many to one mapping from $X$ to $Y$ (Dempster, Laird, and Rubin, 1977). At each iteration of the algorithm, the E step estimates sufficient statistics for the complete data conditional on the observed data and the current estimate of the unknown parameters. The M step then produces ML (or REML) estimates of the parameters using the sufficient statistics generated by the E step. The algorithm continues to iterate until convergence is achieved (Laird, Lange, and Stram, 1988). Relative to the type of MixMod problems considered here, Laird, Lange, and Stram (1988) note that application of the EM algorithm arises naturally when mixed models are used to analyze longitudinal data. In that setting, the incomplete, or observed data, are the measurements collected on each subject, while the complete data include those observed data plus the unobserved random parameters and error terms specified in the mixed model (See for example Andrade and Helms, 1984 and Laird and Ware, 1982). This conceptual model for incomplete data analysis encompasses the standard problem of "missing data" (Laird, Lange, and Stram, 1988).

The primary advantages of the EM algorithm over the alternatives described above are that the E and M steps are often computationally attractive because of the relatively nice forms of the complete-data likelihoods (Wu, 1983); the algorithm is guaranteed to increase (or at least not to decrease) the value of the likelihood function at each iteration and when the sequence does converge, the resultant parameters are guaranteed to be in the parameter space (Dempster, Laird, and Rubin, 1977, and Wu, 1983); solutions to the M step often exist in closed form (Murray and Helms, 1990); and the EM algorithm often requires less computation time per iteration than the alternatives (Fairclough and Helms, 1984, Jennrich and Schluchter, 1985, and Callanan and
Harville, 1988). While the EM algorithm does have some advantages, it also has limitations. First, because it is a first-order iterative procedure, it has a much slower rate of convergence than do the gradient algorithms described earlier. It has been found to require a substantially greater number of iterations to converge than do those gradient algorithms (Fairclough and Helms, 1984, Jennrich and Schluchter, 1986, and Callanan and Harville, 1991). Also, if the likelihood equations used in the M step do not have a closed form solution, an iterative procedure will be needed at each M step to obtain the ML estimates (Laird, Lange, and Stram, 1988).

Modifications to the EM algorithm have been introduced to address the concerns outlined above. The rate of convergence of the algorithm can be increased via the Aitken acceleration method, and algorithms for implementing such modified procedures have been developed (Laird, Lange, and Stram, 1988). Note however, that while this modification can increase convergence rates, the modified algorithm does not have the "guaranteed convergence" property of the original algorithm. Jennrich and Schluchter (1986) proposed a hybrid of the EM and Scoring algorithm that can be used to find ML and REML estimates for MixMod when the M step does not produce a closed form solution. The algorithm uses one iterate of the EM algorithm to estimate $\beta$ and $\tau$, and then uses a Scoring step to estimate $\Sigma$. They found this algorithm to be preferable to the gradient algorithms only when the number of covariance parameters that had to be estimated was large (Jennrich and Schluchter, 1986). Also, when either of these methods is used, care must be taken to assure that each iteration meets the GEM criteria of not decreasing the likelihood (Laird, Lange, and Stram, 1988, and Jennrich and Schluchter, 1986).

1.4.5 Relative Comparison of the Algorithms

Lindstrom and Bates (1988) note that the properties of a good optimization algorithm are iterative computations
that are not prohibitively costly, quick and consistent convergence, and a good method of ascertaining when convergence had been achieved. In comparing EM and NR algorithms, they found little difference in computation times in the examples that they used (Lindstrom and Bates, 1988). Relative to the first criterion, findings in the literature are mixed and generally suggest that the relative computation times depend upon the particular problem. Typically, the gradient algorithms converge in the fewest iterations but have the highest cost per iteration, while the EM algorithm requires the most iterations but has the lowest cost per iteration with the MSA algorithms falling between the other two (Jennrich and Schluchter, 1988 and Callanan and Harville, 1991). Relative to the second criterion, Lindstrom and Bates (1988) found that both the EM and NR algorithm performed equivalently. However, they found a major deficiency of the EM algorithm to be the lack of a good measure of attainment of convergence.

1.5 EXACT AND APPROXIMATE TEST STATISTICS

A primary goal of this research is to examine test statistics that have distributions that are known exactly, or can be closely approximated, in order to test hypotheses, perform power calculations, and develop confidence intervals for secondary parameters of the fixed effects in MixMod. As outlined in Section 1.2.2, these hypothesis are generally of the form $H_0 = B(\theta = 0)$ versus $H_a = B(\theta \neq 0)$, where $\theta = c_1 \beta + c_2 \delta - \theta_0$. Attention will be focused on those situations for which $c_0 = 0$ and on test statistics that are developed for finite sample scenarios. Also, for ease of presentation, $\theta_0$ will typically be assumed to be 0. These test statistics are generally ratios of quadratic forms of the data (or in some limiting cases, simply quadratic forms in the data). Consequently, exact and approximate F-tests and asymptotic Chi-squared tests will be considered.
The remainder of this section reviews both exact and approximate F-tests that have been used for such hypotheses for different models. While the primary interest here is with MixMod, test procedures that are used in the GLUM and GLMM setting can provide helpful insights to the MixMod problem. Before discussing tests for fixed effects in the MixMod framework, results for GLUM and GLMM that have relevance to developing an understanding of possible MixMod test statistics are described.

1.5.1 Hypothesis Test Statistics in GLUM

As noted in Section 1.2.1, the standard model formulation for GLUM is:

\[ Y = X\beta + e, \text{ where } e \sim N(0, \sigma^2 I_n) \]  \hspace{1cm} (1.5.1)

with \( Y \) an \( N \times 1 \) vector of observations, \( \beta \) a \( p \times 1 \) vector of unknown parameters, and \( X \) a known \( N \times p \) design matrix. For this model, the likelihood ratio test statistic for the hypothesis \( H_0: B(\theta = 0) \) versus the general alternative when \( \theta = C\beta \) is (Scheffe, 1959 and Arnold, 1981):

\[ F = \frac{\hat{\theta}'(C(X'X)^{-1}C')^{-1}\hat{\theta}}{s^2} \]  \hspace{1cm} (1.5.2)

where,

\[ \theta = C\hat{\beta} \text{ and } \hat{\beta} = (XX')^{-1}X'y \]  \hspace{1cm} (1.5.3)

Note that \( \text{rank}(C) = a, \text{rank}(X) = r \), and \( s^2 \) is the standard REML estimator for \( \sigma^2 \). Under the null hypothesis \( F \) is distributed as \( F(a, N-r) \), and under the alternative hypothesis, \( F \) is distributed as \( F(a, N-r, \delta) \), where (Arnold, 1981):

\[ \delta = \frac{\theta'(C(X'X)^{-1}C')^{-1}\theta}{\sigma^2} \]  \hspace{1cm} (1.5.4)

When the model assumptions described above are satisfied, this test is uniformly most powerful ("UMP") size \( \alpha \) unbiased (Arnold, 1981).

Examination of the canonical form of the likelihood
ratio test provides an intuitive interpretation of the distributional results for $F$. (See, for example, Scheffe, Section 2.6, and Arnold, Chapters 2 and 7.) Let $\mu$ be the expected value of the data vector $Y$, and note that in the most general case, $\mu \in \mathbb{R}^n$. Now, let $V$ be the subspace of $\mathbb{R}^n$ spanned by the columns of $X$ [i.e., $V=M(X)$], and let $W$ be the subspace of $V$ such that $\theta=0$. Note that $\dim(V)=r$ and $\dim(W)=a$. It is possible to choose an orthonormal basis ("o.n.b.") of $r-a$ vectors, say $(v_{a+1}, v_{a+2}, \ldots, v_r)$, for the subspace $V \setminus W$ (Arnold, Chapter 2). This o.n.b. can be extended to an o.n.b. of $r$ vectors of the form $(v_1, \ldots, v_a, v_{a+1}, \ldots, v_r)$ for $V$, and to a set an o.n.b. of $N$ vectors, $(v_1, \ldots, v_a, v_{a+1}, \ldots, v_r, v_{r+1}, \ldots, v_n)$ for $\mathbb{R}^n$. Let $V=[v_1, v_2, \ldots, v_n]$, and $z_i$ be the coordinates of $Y$ relative to the o.n.b. basis $(v_i)$. Then $z_i=v_i'Y$, and $Z=V'Y$. The test statistic $F$ can then be written in terms of the variables $z_i$ as (Scheffe, 1959):

$$F = \frac{\left(\sum_{i=1}^{a} z_i^2\right)/a}{\left(\sum_{i=a+1}^{n} z_i^2\right)/(N-r)}$$

Under the procedure described above, $Z=V'Y$ is distributed $N(V'X\beta, \sigma^2 I)$. Hence the $z_i$ are independent normal variables with equal variances, and for $i=r+1, r+2, \ldots, N$, $\text{E}(z_i)=0$ (Scheffe, 1959). The distributional result for $F$ follows in a straightforward fashion by dividing both the numerator and denominator by $\sigma^2$.

While the GLUM results presented above, particularly those related to the canonical form, provide useful insights for testing sets of fixed effects contrasts, their applicability to longitudinal data of the form typically evaluated with MixMod is limited because of the strong assumptions of independence and homoscedasticity. However, Robbins and Pittman (1949) and Box (1954a and b) developed both exact and approximate distributional results that could
be used to address heteroscedasticity and correlation among observations in the univariate model. These results for the univariate model have subsequently been applied to repeated measures problems comparable to those addressed with MixMod. The primary distributional results presented by Box (1954a) are:

1. If \( z \) is distributed \( N_p(0, \Sigma) \), \( M \) is a positive semidefinite ("p.s.d.") constant matrix, and \( Q = z' M z \) is a real quadratic form with rank \( r \leq p \), then \( Q \) is distributed like a variable \( X \) which has the form:

\[
X = \sum_{j=1}^{r} \lambda_j \chi^2(1) \tag{1.5.6}
\]

where the chi-square variates are distributed independently and the \( \lambda_j \)'s are the nonzero latent roots of \( U = \Sigma M \).

2. The \( s \)'th cumulant of the distribution of \( Q \), \( K_s(Q) \), is:

\[
K_s(Q) = 2^{s-1}(s-1)! \sum_{j=1}^{r} \lambda_j^s \tag{1.5.7}
\]

3. \( Q \) is approximately distributed as \( g \chi^2(h) \), where:

\[
g = \frac{K_2(Q)}{2K_1(Q)} = \frac{\sum v_j \lambda_j^2}{\sum v_j \lambda_j} \tag{1.5.8}
\]

\[
h = \frac{2[K_1(Q)]^2}{K_2(Q)} = \frac{(\sum v_j \lambda_j)^2}{\sum v_j \lambda_j^2}
\]

4. If \( Q \) and \( Q' \) are independent quadratic forms with approximate distributions \( g \chi^2(h) \) and \( g' \chi^2(h') \) with \( g, g', h, \) and \( h' \) defined as above, then a quantity whose distribution approximates that of \( Q'/Q \) is \( bF(h', h) \), where \( b = K_1(Q')/K_1(Q) \). Box notes that the approximation is not very accurate, but that it may be used to supplement more accurate but less suggestive methods.
As described by Muller and Barton (1989), Box's results suggest that an F-test statistic for an effect involving repeated measures is approximately distributed as an F with modified degrees of freedom of the form $F(\epsilon \cdot df_1, \epsilon \cdot df_2)$. Four tests that use different approximations for $\epsilon$ have been suggested by Huynh and Feldt (1976) and Geisser and Greenhouse (1958, 1959), and the properties of these different tests are discussed by Muller and Barton (1989). Because these approximate tests are frequently suggested as univariate alternatives to analyses of repeated measures data that meet the conditions of GLMM, they will be discussed further with the GLMM presentation in Section 1.5.2.

Although the primary focus of this research is on finite sample situations, asymptotic results in the literature are of interest. First, Andrade and Helms (1984) considered a univariate model with the same mean structure as (1.5.1) and a covariance matrix $\Sigma$ with linear covariance structure as defined in (1.2.13). They considered results for both complete and incomplete data structures. For hypotheses involving three types of constraints ($L'\beta=0$, $S'\tau=0$, and those two constraints jointly), the likelihood ratio test statistics were found to be distributed asymptotically chi-squared with 1, s, and 1+s degrees of freedom respectively, where $l=\text{rank}(L)$ and $s=\text{rank}(S)$ (Andrade and Helms, 1984).

Also as noted in Section 1.3.3, the asymptotic results for the distributions of the quadratic forms used in the standard univariate F-test depend only upon the error terms being independent and identically distributed with 0 mean and finite variance. Consequently, if these assumptions hold, the size and power of the univariate F-tests described above are not affected asymptotically by nonnormality of the error terms (Arnold, 1981).

1.5.2 Hypothesis Test Statistics in GLMM

As described in Section 1.2, another alternative model
that can sometimes be used to analyze data from longitudinal studies is the GLMM. As noted in Section 1.2.2, this model has the form:

$$Y = [x_1//x_2//...//x_n] \beta + [e_1//e_2//...//e_n]$$ \hspace{1cm} (1.5.9)

where \(Y\) is an \(N \times q\) matrix of observations, \(X = [x_1//...//x_n]\) is a known \(N \times p\) design matrix, \(\beta\) is a \(p \times q\) matrix of parameters and \(e\) is an \(N \times q\) matrix of error terms with each row of \(e\) distributed \(N_q(0, \Sigma)\). The hypotheses of interest for this model are generally of the form \(H_0 = B(\Theta = 0)\), where \(\Theta = C\beta U\), with \(\text{Rank}(C) = a \leq q\) and \(\text{Rank}(U) = b \leq p\). In contrast to the GLUM tests described above, no UMP test is available among tests that are unbiased or among tests that are both invariant and unbiased for hypotheses of this form in GLMM (Arnold, 1981). Three methods that are used to define sensible tests when no optimal methods are available are the likelihood ratio method, the substitution method, and the union-intersection method (Arnold, 1981). Because no single UMP test is available, four tests that were developed with these methods are widely used. These tests are Wilk's test with the test statistic Wilk's Lambda, which is based on the likelihood ratio method, the Lawley-Hotelling test with the test statistic Hotelling-Lawley trace, which is based on the substitution method, Roy's Largest Root test, which is based on the union-intersection method, and Pillai's test with the test statistic the Pillai-Bartlett trace. Each of these test statistics is based on functions of the eigenvalues of the matrices \(H\) and \(E\), where (Muller and Barton, 1989):

$$H = (\Theta - \Theta_0)' [C(X'X)^{-1}C']^{-1} (\Theta - \Theta_0)$$

and

$$E = U'(Y-X\hat{\beta})'(Y-X\hat{\beta})U$$ \hspace{1cm} (1.5.10)

These tests, which are described in detail in the literature (see for example, Arnold, 1981, Morrison, 1976, and Timm, 1975), are outlined briefly below. In the discussion below,
note that $|A|$ represents the determinant of the matrix $A$.

Let $\Lambda_u$ be Wilk’s lambda, $\Lambda_{HL}$ be the Hotelling-Lawley trace, $\Lambda_R$ be Roy’s largest root, and $\Lambda_p$ be the Pillai-Bartlett trace. Also define $\lambda_l$, $l=1,2,\ldots,L$ be the eigenvalues of $HE^{-1}$, ordered from largest to smallest. Then the four multivariate test statistics identified above are defined as (Muller and Peterson, 1984):

Wilk’s lambda:

$$
\Lambda_u = |E| / |H+E| = \prod_{l=1}^{L} (1+\lambda_l)^{-1}
$$

(1.5.11)

Hotelling-Lawley trace:

$$
\Lambda_{HL} = \text{trace}(HE^{-1}) = \sum_{l=1}^{L} \lambda_l
$$

(1.5.12)

Roy’s largest root:

$$
\Lambda_R = \lambda_{\max}(HE^{-1}) = \lambda_1
$$

(1.5.13)

Pillai-Bartlett trace:

$$
\Lambda_{PB} = \text{trace}(H(H+E)^{-1}) = \sum_{l=1}^{L} \lambda_l/(1+\lambda_l)
$$

(1.5.14)

Each of the four tests described above is admissible, unbiased, and invariant, and there seems to be no reason to prefer any test to the others universally (Arnold, 1981). Unfortunately, exact procedures for computing probabilities are unavailable for any of the four tests except for special cases. If $s=\min(a,b)=1$ (as is the case for all univariate tests), then all four tests are equivalent and exact probabilities are calculable from an F distribution (Muller and Peterson, 1984). Also, when $s=2$, an exact F-test is available for the likelihood ratio test (Wilk’s test). However, in general, closed form, finite series expressions do not exist for the distribution functions of these test
statistics, particularly under the alternative hypothesis, and approximate asymptotic distributions based F- or chi-square distributions must be used (McCarroll and Helms, 1987, and Muller and Peterson, 1984). Muller and Peterson (1984) provide a discussion of approximations for both the null and non-null cases that can be used for computing approximate power for general forms of GLMM.

The characteristics of the four GLMM test statistics described above limit the application of the most general form of GLMM in the design of longitudinal studies and in the analysis of data from such studies. However, results developed by Box (1954 a,b) for univariate ANOVA with correlated data (see Section 1.5.1) have been used to identify special cases for which exact univariate tests are available for repeated measures ANOVA. These special cases are equivalent to special structures for $\Sigma$ in GLMM. For these special structures, exact F distributions are known for both null and nonnull cases. Also, approximate univariate F-tests have been developed for general repeated measures ANOVA designs that are equivalent to GLMM designs without the special structure for $\Sigma$, as described below.

Letting $\Sigma=[\sigma_{ij}]$, Huynh and Feldt (1970) demonstrated for a randomized block design equivalent to GLMM that the ratio $MSR_A$ has an exact F distribution when $\sigma_{ij}$ have a linear structure of the form:

$$\sigma_{ij} = \alpha_i + \alpha_j + \lambda \delta_{ij}; \lambda > 0 \text{ and } \delta_{ij} = \text{Kronecker } \delta \quad (1.5.15)$$

and where:

$$MSR_A = \frac{MS_A}{MS_{\text{SS}}} = (N-1) \frac{SS_A}{SS_{\text{SS}}}$$

$$= \frac{\sum_{j=1}^{q} N(Y_{qj} - \bar{Y}_{..})^2}{\sum_{q=1}^{Q} \sum_{n=1}^{N} (Y_{qn} - Y_{kj} - Y_{..})^2} \quad (1.5.16)$$

A covariance matrix with the structure specified by (1.5.15)
is said to have **Huynh-Feldt linear structure** or to meet the condition termed **sphericity** (Helms and McCarroll, 1987 and Muller and Barton, 1989). Note that the case in which $\Sigma$ has equal correlation structure is a special case of the Huynh-Feldt linear structure, and hence, equal correlation is a sufficient but not a necessary condition for repeated measures-type problems to have an exact univariate F-test. Relative to GLMM, the covariance matrix $\Sigma$ has Huynh-Feldt linear structure if and only if for a matrix $U$, that spans the contrast space (i.e., $U$ is a $p \times (p-1)$ column orthonormal matrix with $U'U=I_{p-1}$ and $U'1_p=0$), $U'\Sigma U=\lambda I_{p-1}$ (McCarroll and Helms, 1987). McCarroll and Helms (1987) demonstrate conditions under which exact tests can be obtained for longitudinal data under GLMM for both complete and incomplete data.

Approximate F-tests have been developed using the results of Box (1954 a,b) to treat two-way analysis of variance and repeated measures problems that involve a covariance matrix that departs from sphericity. First, note that the condition of sphericity is equivalent to the condition that $\epsilon=1$, where $\epsilon$ is the population value of the $h$ defined by Box (1.5.8) (Huynh and Feldt, 1970). Suppose that the matrix $U$ in the GLMM hypothesis is a column orthonormal matrix with $U'1=0$, and define $\Sigma=U'\Sigma U$ and let $\lambda_q$, $q=1,2,\ldots,Q$ be the ordered eigenvalues of $\Sigma$. The population value of $\epsilon$ is defined as (Muller and Barton, 1989):

\[
\epsilon = \frac{tr^2(\Sigma)}{Q tr(\Sigma^2)} = \left(\frac{\sum_{q=1}^{Q} \lambda_q}{\sqrt{\sum_{q=1}^{Q} \lambda_q^2}}\right)^2
\]

(1.5.17)

Because the population covariance matrix is seldom known, the value of $\epsilon$ is generally estimated via one of four methods. The population value of $\epsilon$ is bounded by $1/Q$ and 1.
(Geisser and Greenhouse, 1959). Consequently, the least conservative and most conservative estimates of \( \eta \) are 1 and \( 1/Q \), respectively. Intermediate estimates have been proposed by Geisser and Greenhouse (1958 and 1959) and by Huynh and Feldt (1976). The Geisser-Greenhouse estimate framed relative to GLMM is (Muller and Barton, 1989):

\[
\eta = \frac{\text{tr}^2(\hat{\Sigma}_e)}{Q\text{tr}(\hat{\Sigma}_e^2)} = \frac{\text{tr}^2(\hat{\Sigma})}{Q\text{tr}(\hat{\Sigma}^2)} \tag{1.5.18}
\]

Huynh and Feldt (1976) noted that \( \hat{\eta} \) is a biased estimate of \( \eta \), and suggested the alternative:

\[
\hat{\eta} = \frac{NQ\hat{\eta} - 2}{Q(N - r - Q\hat{\eta})} \tag{1.5.19}
\]

In all cases, the test uses the actual calculated value of the F-statistic with adjustments in degrees of freedom as described in Section 1.5.1. In their analysis of the analysis of the alternative approximations, Muller and Barton (1989) noted that the ordering of the tests with respect to increasing critical value, decreasing size, and decreasing power is the uncorrected test, the Huynh-Feldt test, the Geisser-Greenhouse test, and the conservative Geisser-Greenhouse test. Based on their simulation analyses they suggest that the Geisser-Greenhouse test generally be used in that it provides acceptable Type I error control while maximizing power among these tests (Muller and Barton, 1989).

The exact and approximate univariate procedures discussed above identify certain cases for which GLMM techniques provide a reasonable alternative for evaluating data from longitudinal studies and provide mechanisms of longitudinal study design. However, these procedures still assume homoscedasticity between subjects. They are also limited in their ability to handle incomplete data or inconsistently timed data.
1.5.3 Hypothesis Test Statistics in MixMod

The literature contains a substantial body of information on estimation and hypothesis testing in GLUM and GLMM. A substantial amount of information is also available on both ML and REML estimation in MixMod. However, the amount of information on hypothesis testing in MixMod appears to be quite limited. Kackar and Harville (1984) develop approximate procedures for calculating standard errors of fixed and random effects estimates in MixMod. Jeske and Harville (1988) address the general problem of developing approximate prediction intervals in MixMod. Laird and Ware (1982) suggest using chi-squared approximations to the LR tests comparable to those evaluated by Andrade and Helms (1984a,b). However, little attention appears to have been given to the development of exact and approximate test statistics for general linear hypotheses for fixed effects as described in Section 1.2.2. The paragraphs below briefly summarize the findings of McCarroll and Helms (1987) relative to MixMod test statistics.

Assume a data vector $Y$ with the MixMod framework defined in (1.2.2) through (1.2.9). Then if each covariance matrix $\Sigma_k$ has a Huynh-Feldt linear structure with constant $\lambda$, an exact $F$ test is available for a testable secondary parameter of the form $\theta = C_1\beta$. The test statistic for $B(\theta=0)$ is (McCarroll and Helms, 1987):

$$F = \frac{\theta' [C_1 (X'HH'X)^{-1} C_1]^{-1} \theta/c}{S/[(N-K)-(r-1)]} \quad (1.5.20)$$

where
\[ H = \text{Diag}(H_1, H_2, \ldots, H_k) \]
\[ S = (H'Y - H'X\hat{\beta})'(H'Y - H'X\hat{\beta}) \]
\[ = (Y - X\hat{\beta})'HH'(Y - X\hat{\beta}) \]  
(1.5.21)
\[ \hat{\beta} = [(H'X)'(H'X)]^{-1}(H'X)'H'Y \]
\[ \theta = C_1\hat{\beta} \]

and each \( H_k \) is a Helmert matrix of appropriate size satisfying the Huynh-Feldt conditions. This test statistic has an exact F distribution with \( c = \text{Rank}(C_1) \) and \( (N-K)-(r-1) \) degrees of freedom and noncentrality parameter \( cF_a \), where \( F_a \) is the value of the test statistic in (1.5.20) evaluated at the true population value of the parameter \( \beta \) (McCarroll and Helms, 1987). While this mechanism does provide an exact test, it is difficult to implement in practice because covariances that have the MixMod structure are unlikely to have the Huynh-Feldt linear structure (McCarroll and Helms, 1987).

Because of the difficulty of implementation of the exact test, McCarroll and Helms (1987) developed three approximate test statistics based on the weighted least squares approach to estimation, the LR-test based on a ML approach, and a test statistic based on REML estimation via a canonical form of the model. They then examined the performance of each of these statistics through a series of simulations. The general findings of the analysis were that the REML-based test, hereafter denoted as \( F_{\text{REML}} \), produced the most accurate Type I error rates and had the closest conformance with the hypothesized distribution (McCarroll and Helms, 1987). A convenient form of \( F_{\text{REML}} \) is (Helms, 1991a):
\[ F_{\text{REML}} = \frac{\theta'[C_1(X'\hat{\Sigma}_Y^{-1}X)^{-1}C_1']^{-1}\theta}{c} \]  
(1.5.22)

where the REML estimators of the covariance parameters are used to compute \( \hat{\sigma}^2 \) and \( \hat{\Lambda} \), which in turn are used to compute:
Although this ad hoc statistic performed well in most simulations, little information is available on either its asymptotic distribution or, more importantly, its exact distribution in finite samples. Further examination of this test statistic and of other comparable statistics derived using procedures based on those developed by McC Carroll and Helms (1987) will be the focus of the research presented in the remainder of this document.
CHAPTER 2
DEVELOPMENT OF MixMod TEST STATISTICS

As outlined in Section 1.5.3, several procedures have been considered for testing hypotheses involving sets of linear contrasts of the fixed effects for those longitudinal studies in which MixMod can be used to evaluate the data. These procedures include Chi-squared tests based on the LR-statistics (Laird and Ware, 1982), approximate F-tests based on standard WLS test statistics that are described along with the LR-test for the SAS procedure PROC MIXED (Wolfinger, Tobias, and Sall, 1991a and 1991b), Wald-type statistics based on one of several methods that can be used to obtain standard error estimates for the parameter estimates (Woolson, Leeper, and Clarke, 1978, Leeper and Woolson, 1982, Kackar and Harville, 1984, and Jeske and Harville, 1988), and an approximate F-test based REML-type estimates of the covariance parameters (McCarroll and Helms, 1987). While the LR, WLS, and Wald procedures are known to be valid asymptotically, the results from McCarroll and Helms (1987), as well as the earlier results from Leeper and Woolson (1982) for standard Wald type tests, suggest that these test statistics have limitations in studies with moderate sample sizes. None of the statistics performed particularly well in simulation studies with small to moderate sample sizes. In fact, the McCarroll and Helms (1987) study indicated that the LR test in particular can produce severely inflated Type I error rates for even moderately large sample sizes. In contrast, the Helms-McCarroll procedure performed well in the simulation studies that they reported. However, the mathematical properties of their test statistic are not established. A major objective
of the research reported in this chapter is to extend the results of McC Carroll and Helms (1987) to develop a test statistic for MixMod with well-defined mathematical and numerical properties that performs well for small and moderate sample sizes. Such a statistic is needed to improve statistical inference and generate reasonable interval estimates for longitudinal studies, particularly when such studies generate incomplete and inconsistently timed data. Such a statistic will also enhance the ability to conduct power calculations for longitudinal studies that include the option of intentionally incomplete data. This chapter describes the derivation of a modified form of the test statistic suggested by McC Carroll and Helms (1987) as well as a form of the statistic based on classic REML estimators.

Although the derivation of the test statistic follows the spirit of McC Carroll and Helms (1987), it yields a statistic with a form that is somewhat different from the one shown in (1.5.22). The different formulation, which has a more intuitive structure than does the original $F_n$ test statistic, is derived from a reparameterization of MixMod into a structure like that described by Callanan and Harville (1991). The first section below presents the reparameterized form of the model. The second section derives estimates of the covariance parameters for the revised model under two estimating procedures—the classic REML methods and the Helms-McCarroll estimating procedure. The third section derives a general form of the MixMod test statistic and develops specific formulae for the test statistic for the two REML-type estimators that are derived in Section 2.2. The fourth section describes some characteristics of the alternative statistics, which are shown to be ratio functions of quadratic-type forms of the data vector.
2.1 ALTERNATIVE PARAMETERIZATION OF MIXMOD

Recall from Section 1.2.1 that the general formulation of MixMod is:

\[ Y = X\beta + Zd + e \]  
\[ (2.1.1) \]

where \( d = [d_1//d_2//...//d_k] \), with

\[ d_k - \text{NID}_r(0, \Delta), \quad \Delta = \sum_{m=1}^{n-1} \tau_m G_m, \text{ and} \]

\[ e - \text{N}_n(0, \sigma^2 I_n) \], independently of \( d \)

Under this parameterization of the model, we have:

\[ E(Y) = X\beta \]  
\[ (2.1.2) \]

and

\[ \text{Var}(Y) = \Sigma = Z(I_k \otimes \Delta)Z' + \sigma^2 I_n \]
\[ = \sum_{m=1}^{n-1} \tau_m G_m^* + \sigma^2 I_n \]  
\[ (2.1.3) \]

where

\[ G_m^* = Z(I_k \otimes G_m)Z' = \text{Diag}(Z_kG_mZ_k') \]

\[ = \text{Diag}(G_m^{*}), \quad k = 1, 2, \ldots, K \]

The remaining discussion will refer to the model formulation described above as the Model 1 form of MixMod. Under Model 1, note that \( Y - \text{N}_n(X\beta, \Sigma) \).

In order to develop parameter estimates that lead to test statistics with desirable distributional properties, the model can be reparameterized in a way that allows \( \sigma^2 \) to be concentrated out of the covariance parameter estimating equations via a modification of the form of the likelihood function. For this model, hereafter known as the Model 2.
formulation of MixMod, let

\[ \Gamma = \frac{1}{\sigma^2} \Sigma = \sum_{m=1}^{M-1} \gamma_m G_m^* + I_m \]

\[ = \text{Diag} \left( \sum_{m=1}^{M-1} \gamma_m G_{mk}^* + I_{mk} \right) \tag{2.1.4} \]

where

\[ \gamma_m = \frac{r_m}{\sigma^2}, \text{ and } \gamma = (\gamma_1, \gamma_2, \ldots, \gamma_{M-1})' \]

Under the Model 2 form of MixMod, we have \( Y \sim N_m(\mathbf{X}\beta, \sigma^2 \Gamma) \).

The log of the likelihood equation for Model 2 (1.2) can easily be developed from the log of the likelihood equation for Model 1 that was presented in (1.3.1). The resulting equation is:

\[ l_2 = l_2(\beta, \gamma, \sigma^2; Y) \]

\[ = -\frac{N}{2} \log \sigma^2 - \frac{1}{2} \log |\Gamma| \tag{2.1.5} \]

\[ - \frac{1}{2\sigma^2} (Y - \mathbf{X}\beta)' \Gamma^{-1} (Y - \mathbf{X}\beta) \]

The Model 2 formulation and its associated likelihood will be used in subsequent sections to develop and characterize the MixMod test statistic.

2.2 DERIVATION OF COVARIANCE PARAMETER ESTIMATES

Section 1.3.1 described two conceptually different approaches to REML estimation of the MixMod covariance parameters. McCarron and Helms (1987) use an algebraic approach in which the parameter estimates are obtained from different subspaces of the data space. Harville (1976 and 1977), Laird and Ware (1982), and Laird, Lange, and Stram (1987) approached REML estimation of the covariance parameters from a Bayesian perspective using a modified
version of the likelihood equation. Further examination of these two estimation procedures indicates that they lead to structurally different estimators. The discussion below will explore these structural differences in estimators developed via the two methods for the special case of a covariance matrix with linear covariance structure. In subsequent chapters, the distributional properties of estimators obtained via the Helms-McCarroll procedures will be explored in some detail. Also, numerical differences in the estimators and in the test statistics produced by the two methods will be examined in Chapter 4.

Separate estimating equations are developed via the two approaches for the Model 2 form of MixMod in the two subsections below. However, before those derivations are presented, four matrix derivative results that are used in both derivations are presented. Three second order derivatives, which are helpful in characterizing the components of the various test statistics are also presented. Because these derivatives can be obtained with minor adjustments to matrix derivative results that are presented elsewhere (Searle, 1982, Graybill, 1969, and Lindstrom and Bates, 1988), the results are presented without detailed derivations.

First, define $\Delta_1$ as:

$$\Delta_1 = \frac{1}{\sigma^2} \Delta = \sum_{m=1}^{M-1} \gamma_m \gamma_m$$  \hspace{1cm} (2.2.1)

and note that

$$\Gamma = Z (I_k \otimes \Delta_1) Z' + I_n$$

$$= \text{Diag} (Z_k \Delta_1 Z_k' + I_{nk})$$  \hspace{1cm} (2.2.2)

$$= \text{Diag} (\Gamma_k), \ k=1, 2, \ldots, K$$

The notation shown above can now be used to express the following general derivatives for $\beta$ and for $\gamma_m$ with $m=1, 2, \ldots, M-1$:
Result 1:

\[
\frac{\partial (y - X\beta)' \Gamma^{-1} (y - X\beta)}{\partial \beta} = -2X' \Gamma^{-1} (y - X\beta). \tag{2.2.3}
\]

Result 2:

\[
\frac{\partial \log |\Gamma|}{\partial y_m} = \text{tr} \left( \Gamma^{-1} \frac{\partial \Gamma}{\partial y_m} \right)
\]

\[
= \text{tr} \left( \Gamma^{-1} G_m^* \right)
\]

\[
= \sum_{k=1}^{K} \text{tr} \left[ \Gamma_k^{-1} (Z_k G_m Z_k') \right]
\]

\[
= \sum_{k=1}^{K} \text{tr} \left( \Gamma_k^{-1} G_m^* \right) \tag{2.2.4}
\]

Result 3:

\[
\frac{\partial \log |X' \Gamma^{-1} X|}{\partial y_m} = \text{tr} \left( -\Gamma^{-1} X (X' \Gamma^{-1} X)^{-1} X' \Gamma^{-1} \frac{\partial \Gamma}{\partial y_m} \right)
\]

\[
= -\text{tr} \left[ \Gamma^{-1} X (X' \Gamma^{-1} X)^{-1} X' \Gamma^{-1} G_m^* \right]
\]

\[
= -\sum_{k=1}^{K} \text{tr} \left[ X_k (X_k' \Gamma_k^{-1} X_k)^{-1} X_k' \Gamma_k^{-1} G_m^* \right]
\]

Result 4:

\[
\frac{\partial (y - X\beta)' \Gamma^{-1} (y - X\beta)}{\partial y_m} = -(y - X\beta)' \Gamma^{-1} \frac{\partial \Gamma}{\partial y_m} \Gamma^{-1} (y - X\beta)
\]

\[
= -\text{tr} \left[ (y - X\beta)' \Gamma^{-1} G_m^* \Gamma^{-1} (y - X\beta) \right] \tag{2.2.6}
\]

\[
= -\sum_{k=1}^{K} \text{tr} \left[ (y_k - X_k\beta)' \Gamma_k^{-1} G_m^* \Gamma_k^{-1} (y_k - X_k\beta) \right]
\]

where $G_m^*$ and $G_{mk}^*$ are defined as in (2.1.3), $\Gamma_k$ is defined in (2.2.2), and the function "tr" represents the trace of the square matrices. Although the result is not immediately obvious, it can be demonstrated mathematically that if $\beta$ is replaced with $\hat{\beta}$ on the left hand side in Result 4, a
derivative with an identical form is obtained on the right hand side with \( \beta \) replaced by \( \hat{\beta} \).

The results developed above can be extended to obtain the following forms for second order derivatives.

**Result 5:**

\[
\frac{\partial^2 (y - X\beta)' \Gamma^{-1} (y - X\beta)}{\partial \beta \partial \beta'} = 2X'\Gamma^{-1}X
\]  

(2.2.7)

**Result 6:**

\[
\frac{\partial^2 (y - X\beta)' \Gamma^{-1} (y - X\beta)}{\partial \beta \partial \gamma_m} = 2X'\Gamma^{-1}G_m^*\Gamma^{-1}(y - X\beta)
\]  

(2.2.8)

**Result 7:**

\[
\frac{\partial^2 \log |\Gamma|}{\partial \gamma_m \partial \gamma_m} = -tr[\Gamma^{-1}G_m^*\Gamma^{-1}G_m^*]
\]  

(2.2.9)

**Result 8:**

\[
\frac{\partial^2 (y - X\beta)' \Gamma^{-1} (y - X\beta)}{\partial \gamma_m \partial \gamma_m} = \\
(y - X\beta)' \Gamma^{-1}G_m^*\Gamma^{-1}G_m^*\Gamma^{-1}(y - X\beta)
\]  

(2.2.10)

\[
+ (y - X\beta)' \Gamma^{-1}G_m^*\Gamma^{-1}G_m^*\Gamma^{-1}(y - X\beta)
\]

The forms for the derivatives or slight modifications of these forms are used in the subsections below to derive estimating equations for the covariance parameters for the alternative REML procedures as well as estimators for the fixed effect parameters.

2.2.1 **Classic REML Estimating Equations**

Harville (1974 and 1977) used a Bayesian approach to develop the likelihood equations for REML estimation procedures for mixed linear models that were equivalent to the procedures developed by Patterson and Thompson (1971). This likelihood function, which was given in (1.3.15), can be modified to encompass the Model 2 form of MixMod as shown
below. From (1.3.15), the restricted (or residual) likelihood equation for the Model 1 form of MixMod is:

\[
1_r(\beta, \tau; y) = C - \frac{1}{2} \log |\Sigma| - \frac{1}{2} \log |X'\Sigma^{-1}X| - \frac{1}{2} (y - X\hat{\beta})'/\Sigma^{-1}(y - X\hat{\beta}) \tag{2.2.11}
\]

By setting \( \Sigma = \sigma^2 \Gamma \) in the above equation and assuming that \( X \) is FR, the restricted (or residual) likelihood under Model 2 becomes:

\[
l_x(\beta, \sigma^2, \gamma; y) = C - \frac{1}{2} (N-r) \log \sigma^2 - \frac{1}{2} \log |\Gamma| - \frac{1}{2} \log |X'\Gamma^{-1}X| - \frac{1}{2\sigma^2} (y - X\hat{\beta})'/\Gamma^{-1}(y - X\hat{\beta}) \tag{2.2.12}
\]

Note that the likelihood involves the estimator \( \hat{\beta} \) rather than the parameter \( \beta \). The REML estimating equations for the covariance parameters can be obtained by differentiating \( l_x \) with respect to \( \sigma^2 \) and the elements of \( \gamma \) and setting the derivatives equal to 0.

First consider the derivatives with respect to the elements of \( \gamma \). Let

\[
P = \Gamma^{-1} - \Gamma^{-1}X(X'\Gamma^{-1}X)^{-1}X'\Gamma^{-1}
\]

and note that \( P \) is a block diagonal matrix of the form:

\[
P = \text{Diag}(P_k)
\]

where

\[
P_k = \Gamma_k^{-1} - \Gamma_k^{-1}X_k(X_k\Gamma_k^{-1}X_k)^{-1}X_k\Gamma_k^{-1}
\]

The matrix derivatives presented earlier can be used to obtain:
\[
\frac{\partial l_k}{\partial \gamma_m} = -\frac{1}{2} \text{tr}(P G_m^*) \\
+ \frac{1}{2\sigma^2} \text{tr}\left[(y - X\hat{\beta})'(\Gamma_m^{-1} G_m^* \Gamma_m^{-1})(y - X\hat{\beta})\right]
\] (2.2.14)

Recall from Section 1.3.1 that, for the classic REML estimating procedure, the fixed effect parameter estimates are obtained by maximizing the original likelihood (1.3.1) with respect to \( \beta \) under the assumption that \( \Sigma \) is known and then replacing \( \Sigma \) in the estimating equations with a consistent estimator such as the REML estimator. That estimator can be modified to use a consistent estimator of \( \Gamma \) as shown below.

\[
\hat{\beta} = (X' \hat{\Sigma}^{-1} X)^{-1} X' \hat{\Sigma}^{-1} y \\
= (X' (\sigma^2 \hat{\Gamma})^{-1} X)^{-1} X' (\sigma^2 \hat{\Gamma})^{-1} y \\
= (X' \hat{\Gamma}^{-1} X)^{-1} X' \hat{\Gamma}^{-1} y
\] (2.2.15)

Note that the same estimator for \( \beta \) can be obtained by maximizing a modified form of (1.3.1) with \( \Sigma \) replaced by \( \sigma^2 \Gamma \). In a similar fashion, modify the estimator for the realization of the random effects as follows:

\[
\hat{\delta} = \Delta z' \hat{\Sigma}^{-1} (y - X\hat{\beta}) \\
= \sigma^2 \Delta z' (\sigma^2 \hat{\Gamma})^{-1} (y - X\hat{\beta}) \\
= \Delta z' \hat{\Gamma}^{-1} (y - X\hat{\beta})
\] (2.2.16)

Now, returning to the problem of estimating \( \gamma \), In Equation 2.2.14, note that the derivative of the likelihood function with respect to \( \gamma_m \) involves \( \sigma^2 \). This functional relationship presents a complication in that the ultimate goal of this analysis is to estimate \( \gamma \) independently of the estimate of \( \sigma^2 \). As demonstrated by Bard (1974) and subsequently used by Callanan and Harville (1991) to simplify computation of the REML estimates, \( \sigma^2 \) can be concentrated out of the likelihood analytically to generate
a modified likelihood function, which was denoted as the concentrated likelihood function by Bard (1974). To generate this function, note that:

\[
\frac{\partial l_R}{\partial \sigma^2} = -\frac{(N-r)}{2\sigma^2} + \frac{(y - X\hat{\beta})' \Gamma^{-1} (y - X\hat{\beta})}{2(\sigma^2)^2}
\]

\[
= -\frac{(N-r)}{2(\sigma^2)^2} \left[ \sigma^2 - \frac{(y - X\hat{\beta})' \Gamma^{-1} (y - X\hat{\beta})}{N-r} \right]
\]

(2.2.17)

For any fixed value of \( \gamma \), let

\[
\sigma^2(\gamma) = \frac{(y - X\hat{\beta})' \Gamma^{-1} (y - X\hat{\beta})}{N-r}
\]

(2.2.18)

and note that \( l_R \) is maximized at \( \sigma^2 = \sigma^2(\gamma) \). Hence, \( l_R \) can be modified to obtain the concentrated log likelihood function \( l_{RC} \), which contains the \( M-1 \) random effects covariance parameters, as follows:

\[
l_{RC}(\beta, \sigma^2(\gamma), \gamma; y) = C - \frac{1}{2} (N-r) \log \sigma^2(\gamma)
\]

\[
- \frac{1}{2} \log |\Gamma| - \frac{1}{2} \log |X' \Gamma^{-1} X| 
\]

\[
- \frac{1}{2\sigma^2(\gamma)} (y - X\hat{\beta})' \Gamma^{-1} (y - X\hat{\beta})
\]

(2.2.19)

This equation can be simplified algebraically to obtain:

\[
l_{RC}(\beta, \sigma^2(\gamma), \gamma; y) = C - \frac{N-r}{2} \left[ 1 + \log \sigma^2(\gamma) \right]
\]

\[
- \frac{1}{2} \log |\Gamma| - \frac{1}{2} \log |X' \Gamma^{-1} X| 
\]

Hence,

\[
\frac{\partial l_{RC}}{\partial \gamma_m} = -\frac{1}{2} \text{tr}(P_{GM}) - \frac{N-r}{2 \sigma^2(\gamma)} \frac{\partial \sigma^2(\gamma)}{\partial \gamma_m}
\]

(2.2.20)

Using the results on matrix derivatives presented earlier,
this equation can be manipulated algebraically to yield:

\[
\frac{\partial \ell_{rc}}{\partial \gamma_m} = -\frac{1}{2} \frac{\partial^2(\gamma)}{\partial \gamma^2} \left[ \text{tr} \left\{ \hat{G}_m \right\} \right] \]

\[
- \text{tr} \left\{ \left( y - X \hat{\beta} \right)' R^{-1} G_m^* R^{-1} (y - X \hat{\beta}) \right\}
\]

(2.2.21)

Now, define \( R' \) as:

\[
R' = (y - X \hat{\beta}) (y - X \hat{\beta})'
\]

Finally, let \( \gamma_r = (\gamma_{r,1}, \ldots, \gamma_{r,M-1}) \), be the vector of classic REML estimators of \( \gamma \), and define \( \hat{\gamma}_r, \hat{\beta}_r, \hat{\hat{\beta}}_r, \hat{\delta}_r, \) and \( \hat{\lambda}_r \) to be the estimators obtained by evaluating \( \Gamma, \beta, \hat{\beta}, \delta, \) and \( R' \), respectively at \( \gamma = \gamma_r \). Then the estimating equations for \( \hat{\gamma} \) can be derived as follows:

\[
\left. \frac{\partial \ell_{rc}}{\partial \gamma_m} \right|_{\gamma_r, \delta} = 0
\]

\[
- \frac{\partial^2(\gamma)}{\partial \gamma^2} \cdot \text{tr} \left\{ \hat{\gamma}_r G_m^* \right\} - \text{tr} \left\{ \hat{\gamma}_r^{-1} G_m^* \hat{\gamma}_r^{-1} \hat{\lambda}_r \right\} = 0
\]

\[
- \frac{\partial^2(\gamma)}{\partial \gamma^2} \cdot \text{tr} \left\{ \hat{\gamma}_r G_m^* \hat{\gamma}_r^{-1} \left[ \sum_{t=1}^{M-1} \hat{\gamma}_t \hat{G}_t^* + I_m \right] \right\} = \text{tr} \left\{ \hat{\gamma}_r^{-1} G_m^* \hat{\gamma}_r^{-1} \hat{\lambda}_r \right\}
\]

\[
- \text{tr} \left\{ \hat{\gamma}_r G_m^* \hat{\gamma}_r^{-1} \left[ \sum_{t=1}^{M-1} \hat{\gamma}_t \hat{G}_t^* \right] \right\} = \text{tr} \left\{ \hat{\gamma}_r^{-1} G_m^* \hat{\gamma}_r^{-1} \hat{\lambda}_r - \hat{\gamma}_r G_m^* \hat{\gamma}_r^{-1} \right\}
\]

The \( M-1 \) equations obtained by this process can be manipulated algebraically into the matrix format:

\[
\left\langle \text{tr} \left\{ \hat{\gamma}_r G_m^* \hat{\gamma}_r^{-1} \hat{G}_t^* \right\} \right\rangle_{st} \cdot \hat{\gamma}_r = \left\langle \text{tr} \left\{ \hat{\gamma}_r^{-1} G_m^* \hat{\gamma}_r^{-1} \hat{\lambda}_r - \hat{\gamma}_r G_m^* \hat{\gamma}_r^{-1} \right\} \right\rangle_s
\]

where again \( \langle f(s,t) \rangle_{st} \) is a matrix with \( f(s,t) \) as the element in the \( s \)-th row and \( t \)-th column and \( \langle f(s) \rangle_s \) is a
vector with \( f(s) \) in the \( s \)-th position. Using these results, the estimating equations for the elements of \( \gamma \) can be written as:

\[
\hat{\gamma}_r = \left[ \text{tr} \left( \hat{\mathbf{P}}_r \mathbf{G}_r^s \mathbf{F}^{-1}_r \mathbf{F}^s_r \right) \right]^{-1} \cdot \left[ \text{tr} \left( \hat{\mathbf{F}}_r^{-1} \mathbf{G}_s^s \mathbf{F}^{-1}_r \mathbf{F}^s_r \right) \right]_{st} \\
\left[ \text{tr} \left( \frac{\hat{\mathbf{F}}_r^{-1} \mathbf{G}_s^s \mathbf{F}^{-1}_r \mathbf{F}^s_r - \hat{\mathbf{P}}_r \mathbf{G}_s^s \mathbf{F}^{-1}_r \right)}{\delta^2(\gamma)} \right]_s
\]  

(2.2.22)

Next, consider the estimating equation for \( \sigma^2 \). Note from the results presented above that the elements of \( \gamma \) can be estimated with \( \sigma^2 \) concentrated out of the likelihood and that \( \hat{\beta} \) can be expressed in a form that is independent of \( \sigma^2 \) as shown in (2.2.10). Recall that in concentrating \( \sigma^2 \) out of the likelihood, we found the derivative of the likelihood with respect to \( \sigma^2 \) to be:

\[
\frac{\partial \ln L_r}{\partial \sigma^2} = -\frac{1}{2} (N-r) (\sigma^2)^{-1} \\
+ \frac{1}{2} (\sigma^2)^{-2} (\gamma - \mathbf{X}\hat{\beta})' \Gamma^{-1} (\gamma - \mathbf{X}\hat{\beta})
\]  

(2.2.23)

The resultant estimator for \( \sigma^2 \) is calculated as follows:

\[
\frac{\partial \ln L_r}{\partial \sigma^2 \hat{\sigma}^2_r} \bigg|_{\hat{\sigma}^2_r} = 0
\]

\[ \Rightarrow \delta_{\text{REML}} = (N-r)^{-1} (\gamma - \mathbf{X}\hat{\beta})' \hat{\Gamma}^{-1} (\gamma - \mathbf{X}\hat{\beta}) \]

\[ = (N-r)^{-1} \gamma' \hat{\Gamma}^{-1} (\gamma - \mathbf{X}\hat{\beta}) \]

\[ = (N-r) \gamma' (\gamma - \mathbf{X}\hat{\beta} - \mathbf{Z}\hat{\delta}) \]

Note from the above equation that estimates of \( \beta \) and \( \delta \) are needed to estimate \( \sigma^2 \); the classic REML approach uses \( \hat{\gamma}_r \) for these estimates and the final estimator for \( \sigma^2 \) becomes:
\[
\hat{o}_r^2 = (N-r)^{-1} y' (y - X\hat{\beta}_r - Z\hat{\delta}_r) \\
= (N-r)^{-1} (y - X\hat{\beta}_r - Z\hat{\delta}_r)' (y - X\hat{\beta}_r - Z\hat{\delta}_r)
\] (2.2.24)

Note that although \( \hat{\gamma}_r \) does not appear in (2.1.24), \( \hat{o}_r^2 \) is functionally dependent on \( \hat{\gamma}_r \) through its dependence on \( \hat{\beta}_r \) and \( \hat{\delta}_r \).

Although the likelihood function used in the above derivation was developed by both Harville (1974 and 1976) and Laird and Ware (1982) with Bayesian procedures, Harville (1974) demonstrated that the result was equivalent to the results of Patterson and Thompson (1971), which had an algebraic basis. Algebraically then, the estimates of \( \gamma \) and \( \sigma^2 \) are based on transformed normal random variables obtained from mapping \( y \) onto the orthogonal complement of \( M(X) \) in \( \mathbb{R}^n \). When the covariance parameters are known, the classic procedure estimates \( \beta \) from the column space of a linear transformation of \( X \) based on the root of the inverse of \( \text{Var}(Y) \). When the covariance parameters are unknown, the procedure estimates the covariance parameters from the error space, which is equivalent to the orthogonal complement of the column space of \( X \). The elements of \( \beta \) are obtained from the linear transformation of the column space of \( X \) obtained from the resultant estimate of \( \text{Var}(Y) \); consequently the algebraic properties of this estimator are no longer clear.

2.2.2 Helms-McCarroll Estimating Equations

Recall from Section 1.3.2 that the random vector of the response variable \( Y \) can be transformed through an orthogonal transformation \( T \) and can then be partitioned as shown below:

\[
Y' = T' Y = \begin{bmatrix} T'_{11} & T'_{12} \\ T'_{21} & T'_{22} \end{bmatrix} \begin{bmatrix} Y'_{11} \\ Y'_{12} \end{bmatrix} = \begin{bmatrix} Y'_{11} \\ Y'_{22} \end{bmatrix}
\] (2.2.25)

where these variables are defined as in (1.3.16) and (1.3.17). Now, define \( W = W_1 // W_2 \), where \( W_1 = Y'_{11} // Y'_{12} \).
and $\mathbf{w}_2 = \mathbf{y}_2^*$. For simplification of terminology, also let $\mathbf{T}_1 = \mathbf{T}_{11} \parallel \mathbf{T}_{12}$. The results presented by McCarroll and Helms (1987) and summarized in Section 1.3.2 demonstrate that $\mathbf{w}_1$ and $\mathbf{w}_2$ are independent multivariate normal random vectors with the following mean and covariance structures:

$$
E(\mathbf{w}_1) = \mu_1 = \mathbf{T}_1' \mathbf{x} \beta
$$

$$
= \begin{bmatrix}
\mathbf{T}_{11}' \mathbf{x} \beta \\
0
\end{bmatrix} = \begin{bmatrix}
\mathbf{T}_{11}' \\
0
\end{bmatrix} \beta
$$

with

$$
\text{Var}(\mathbf{w}_1) = \Sigma_1 = \mathbf{T}_1' \Sigma_1 \mathbf{T}_1
$$

$$
= \begin{bmatrix}
\mathbf{T}_{11}' \mathbf{Z} \Delta Z' \mathbf{T}_{11} + \sigma^2 \mathbf{I}_n & \mathbf{T}_{11}' \mathbf{Z} \Delta Z' \mathbf{T}_{12} \\
\mathbf{T}_{12}' \mathbf{Z} \Delta Z' \mathbf{T}_{11} & \mathbf{T}_{12}' \mathbf{Z} \Delta Z' \mathbf{T}_{12} + \sigma^2 \mathbf{I}_{b-a}
\end{bmatrix}
$$

and

$$
E(\mathbf{w}_2) = 0
$$

with

$$
\text{Var}(\mathbf{w}_2) = \sigma^2 \mathbf{I}_{b-a}
$$

where

$$
\Delta^* = \mathbf{I}_k \otimes \Delta
$$

Finally, note that because $\mathbf{w}_2$ has zero expectation and the columns of $\mathbf{T}_2$ are orthogonal to those of $\mathbf{x}$, $\mathbf{Z}$, and $\mathbf{T}_1$, the covariance of $\mathbf{w}_1$ and $\mathbf{w}_2$ can be determined as follows:

$$
\text{Cov}(\mathbf{w}_1, \mathbf{w}_2) = E(\mathbf{w}_1 \mathbf{w}_2') = E(\mathbf{T}_1' \mathbf{y} \mathbf{y}' \mathbf{T}_2)
$$

$$
= \mathbf{T}_1' \text{E}(\mathbf{y} \mathbf{y}') \mathbf{T}_2
$$

$$
= \mathbf{T}_1' \left[ \text{Var}(\mathbf{y}) + E(\mathbf{y}) E(\mathbf{y}') \right] \mathbf{T}_2
$$

$$
= \mathbf{T}_1' \left[ \mathbf{Z} \left( \mathbf{I}_k \otimes \Delta \right) \mathbf{Z}' + \sigma^2 \mathbf{I}_n + \mathbf{x} \beta \mathbf{x}' \right] \mathbf{T}_2
$$

$$
= 0
$$

Hence, $\mathbf{w}_1$ and $\mathbf{w}_2$ are independent normal random vectors.
If the problem is reparameterized in a fashion equivalent to the parameterization described in Section 2.2.1 with

$$\Delta^*_i = \frac{1}{\sigma^2} Z \Delta^* Z'$$

$$= \sum_{m=1}^{N-1} \gamma_m G_m^*$$

then the covariance matrix for \( \mathbf{w}_1 \) can be rewritten as follows:

$$\text{Var} (\mathbf{w}_1) = \Sigma_1 = \sigma^2 \Gamma_1$$

(2.2.29)

where

$$\Gamma_1 = T_1' \Gamma T_1$$

$$= T_1' \begin{bmatrix} \Delta^*_i + I_1 & \Delta^*_i \\ \Delta^*_i & \Delta^*_i + I_{b-1} \end{bmatrix} T_1$$

$$= \begin{bmatrix} T_{11}' \Gamma T_{11} & T_{11}' \Delta^*_i T_{12} \\ T_{12}' \Delta^*_i T_{11} & T_{12}' \Gamma T_{12} \end{bmatrix}$$

Furthermore, let \( G_m^* = T_1' G_m^* T_1 \) and note that

$$\Gamma_1 = T_1' \Gamma T_1 = T_1' \left( \sum_{m=1}^{N-1} \gamma_m G_m^* + I_N \right) T_1$$

$$= \sum_{m=1}^{N-1} \gamma_m G_m^* + T_1' T_1$$

The Helms–McCarroll strategy uses the "pure error space", \( \mathbf{w}_2 \), to estimate \( \sigma^2 \) and the subspace of the data represented by \( \mathbf{w}_1 \) to estimate \( \gamma \) and \( \beta \). Because the vector \( \mathbf{w}_2 \) is multivariate normal of size \( N-b \) (where \( b \) is the rank of \( T_1 \) as defined in Section 1.3.2) with zero expectation, the maximum likelihood estimator for \( \sigma^2 \) based on \( \mathbf{w}_2 \) is:
\[ \hat{\sigma}^2 = \frac{\tilde{W}_2^T \tilde{W}_2}{N - 5} \]  

(2.2.30)

The estimating equations for \( \gamma \) are obtained by maximizing the likelihood function for \( W_1 \) simultaneously for \( \gamma \) and \( \beta \). The likelihood function for \( W_1 \) in reparameterized form is:

\[
\ell_{W_1} = C - \frac{b}{2} \log \sigma^2 - \frac{1}{2} \log |\Gamma_1| - \frac{1}{2\sigma^2} (W_1 - \mu_1)' \Gamma_1^{-1} (W_1 - \mu_1) \]  

(2.2.31)

In order to maximize the likelihood with respect to \( \gamma \), note that

\[
\frac{\partial \Gamma}{\partial Y_m} = \frac{\partial \Delta^*}{\partial Y_m} = G_m^* = \text{Diag}(G_k^*)
\]

and hence,

\[
\frac{\partial \Gamma_1}{\partial Y_m} = \begin{bmatrix} T_{11}' \\ T_{12}' \end{bmatrix} G_m^* \begin{bmatrix} T_{11} & T_{12} \end{bmatrix} = T_1' G_m^* T_1 = G_m^*
\]

The various matrix derivative forms identified earlier can be used to obtain:

\[
\frac{\partial \ell_{W_1}}{\partial Y_m} = -\frac{1}{2} \text{tr}\left(\Gamma_1^{-1} G_m^*\right) + \frac{1}{2\sigma^2} \text{tr}\left((W_1 - \mu_1)' \Gamma_1^{-1} G_m^* \Gamma_1^{-1} (W_1 - \mu_1)\right)
\]

Note that the form of the derivative shown above has essentially the same form as the derivative used to obtain the classic REML estimating equations for \( \gamma \) and that it also involves \( \sigma^2 \). Hence, an analogous sequence of calculations can be used to obtain the estimating equations for \( \gamma \) for the Helms-McCarroll procedure. First, note that

\[
\frac{\partial \ell_{W_1}}{\partial \sigma^2} = -\frac{b}{2\sigma^2} + \frac{1}{2\sigma^2} (W_1 - \mu_1)' \Gamma_1^{-1} (W_1 - \mu_1)
\]
Hence, for any fixed value of $\gamma$, let
\[
\delta^2(\gamma) = \frac{1}{b} (\mathbf{w}_1 - \mu_1)' \Gamma_1^{-1} (\mathbf{w}_1 - \mu_1) \tag{2.2.32}
\]

Procedures like those described earlier for the classic REML estimating equations can be used to obtain a concentrated likelihood function for the Helms-McCarroll procedure:
\[
l_{w1c} = C - \frac{b}{2} [1 + \log \delta^2(\gamma)] - \frac{1}{2} \log |\Gamma_1| \tag{2.2.33}
\]

Then \( \frac{\partial l_{w1c}}{\partial \gamma_m} \bigg|_{\gamma_m} = 0 \) if and only if
\[
\delta^2(\gamma) \cdot \text{tr} \left( \hat{\Gamma}_1^{-1} \Gamma_m \right) - \text{tr} \left( \hat{\Gamma}_1^{-1} \hat{\Gamma}_1^{-1} \hat{\mathbf{R}} \right) = 0
\]
\[
- \delta^2(\gamma) \cdot \text{tr} \left( \hat{\Gamma}_1^{-1} \Gamma_m \hat{\Gamma}_1 \left( \sum_{m' = 1}^{N-1} \hat{\gamma}_{m'} \Gamma_{m'} + I_b \right) \right) - \text{tr} \left( \hat{\Gamma}_1^{-1} \Gamma_m \hat{\Gamma}_1^{-1} \hat{\mathbf{R}} \right) = 0
\]
\[
= \text{tr} \left( \hat{\Gamma}_1^{-1} \Gamma_m \hat{\Gamma}_1 \left( \sum_{m' = 1}^{N-1} \hat{\gamma}_{m'} \Gamma_{m'} \right) \right) + \text{tr} \left( \frac{\hat{\Gamma}_1^{-1} \Gamma_m \hat{\Gamma}_1^{-1} \hat{\mathbf{R}}}{\delta^2(\gamma)} - \hat{\Gamma}_1^{-1} \Gamma_m \hat{\Gamma}_1^{-1} \right)
\]

These equations can be written in matrix form as:
\[
\hat{\gamma}_m = \left[ \text{tr} \left( \hat{\Gamma}_1^{-1} \hat{\Gamma}_1^{-1} \hat{\mathbf{R}} \right) \right]^{-1} \cdot
\]
\[
\left[ \text{tr} \left( \hat{\Gamma}_1^{-1} \Gamma_m \hat{\Gamma}_1^{-1} \hat{\mathbf{R}} \right) \right]_{st}
\]
\[
\left[ \text{tr} \left( \frac{\hat{\Gamma}_1^{-1} \Gamma_m \hat{\Gamma}_1^{-1} \hat{\mathbf{R}}}{\delta^2(\gamma)} \left( \hat{\Gamma}_1^{-1} \hat{\mathbf{R}}_1 - I_b \right) \right) \right]_{st}
\]

where
\[ \hat{\beta}_n = (w_1 - \mu_1) (w_1 - \mu_1)' \]

\[ = (w_1 - T_1' \hat{X}_n) (w_1 - T_1' \hat{X}_n)' \]

and

\[ \hat{\Gamma}_n^{-1} = \Gamma_1^{-1}_{Y_n} \]

In order to develop the estimate for \( \beta \), note that \( w_1 \) is MVN with mean and covariance structure as defined in (2.2.26) through (2.2.28). Hence, the standard ML estimator for \( \beta \) based on \( w_1 \) is:

\[ \hat{\beta}_n = \left( \begin{bmatrix} T_{11}' \hat{X} \\ 0 \end{bmatrix} \hat{\Gamma}_n^{-1} \begin{bmatrix} T_{11}' \hat{X} \\ 0 \end{bmatrix} \right)^{-1} \begin{bmatrix} T_{11}' \hat{X} \\ 0 \end{bmatrix} \hat{\Gamma}_n^{-1} \begin{bmatrix} T_{11}' \\ T_{12}' \end{bmatrix} y \]

(2.2.35)

\[ = \begin{bmatrix} x' \hat{\Gamma}_n^{-1} T_1' x \\ x' T_1 \hat{\Gamma}_n^{-1} T_1' y \end{bmatrix} \]

Further, if we let:

\[ \hat{\Gamma}_n^{-1} = \begin{bmatrix} U_{11} & U_{12} \\ U_{12}' & U_{22} \end{bmatrix} \]

then the Helms-McCarroll estimate of \( \beta \) becomes:

\[ \hat{\beta}_n = \begin{bmatrix} x' T_{11} U_{11} T_{11}' x \\ x' T_{11} U_{11} T_{11}' x + x' T_{11} U_{12} T_{12}' \end{bmatrix} y \]

(2.2.36)

Note that for this estimation procedure, \( \sigma_n^2 \) is functionally independent of both \( \hat{\gamma}_n \) and \( \hat{\beta}_n \).

2.3 DERIVATION OF THE MixMod F-STATISTIC

The primary objective of this section is to derive a test statistic that can be used to test hypotheses about population effects for longitudinal studies with small to moderate numbers of experimental units. In particular, the research is directed toward hypotheses concerning sets of linear contrasts among the fixed effects in MixMod, although
the results can be extended to hypotheses that involve sets of linear contrasts among both fixed and random effects. Using the terminology outlined in Section 1.2.2, let:

\[ \theta = c_1 \beta - \theta_0 \]  

be a \( c \times 1 \) secondary parameter, and consider the hypothesis:

\[ H_0(\theta) = B(\theta = 0) \] versus \[ H_1(\theta) = B(\theta \neq 0) \]  

(2.3.2)

The motivation for the MixMod statistic comes from considering the test statistic that would be used to test an analogous hypothesis in the GLUM. Suppose we have a model of the form GLUM-FR(\( Y, X\beta, \sigma^2 V \)) of order \( N \), where \( V \) is a known p.d. matrix and \( \text{rank}(X) = r \). The likelihood ratio test statistic for testing a hypothesis analogous to the one presented in (2.3.1) and (2.3.2) is:

\[
F_u = \frac{\theta'[C_1(X'V^{-1}X)C_1']^{-1}\theta/c}{s^2} 
\]  

(2.3.3)

where \( \theta \) is the estimate of \( \theta \) evaluated at \( \hat{\beta} \) and \( s^2 \) is the corresponding weighted error sum of squares divided by \( N-r \). Note that this form of \( s^2 \) is the REML estimator for the error variance in GLUM. Also, note that this statistic can be represented as follows:

\[
F_u = \frac{\theta'[\text{var}(\theta)]^{-1}\theta/c}{s^2} 
\]  

(2.3.4)

Under the null hypothesis, the statistic is known to have an exact \( F \) distribution with \( c \) numerator degrees of freedom and \( N-r \) denominator degrees of freedom. Under the alternative hypothesis, it has a noncentral \( F \) distribution with noncentrality parameter:
\[
nc = \frac{\theta'[c_1(x'v^{-1}x)c_1']^{-1}\theta}{2}
\]

Using the above results as a paradigm, the MixMod test statistic is derived via the following procedure. Suppose \( \gamma \) is known. Then \( Y - N(X\beta, \sigma^2\Gamma) \), where \( \Gamma \) is a known p.d. matrix, and a natural form that can be considered for the test statistic is:

\[
F_1 = \frac{\theta'[c_1(x'(\sigma^2\Gamma)^{-1}x)c_1']^{-1}\theta/c}{\sigma^2_{\text{REML}}/\sigma^2} = \frac{\theta'[c_1(x'\Gamma^{-1}x)c_1']^{-1}\theta/c}{\sigma^2_{\text{REML}}}
\]

While the above statistic provides insight into a potential form for the test statistic, it is not useful in practice because \( \gamma \) typically is not known. In practice, we propose to use \( \hat{\gamma}_{\text{REML}} \), which is a consistent estimator of \( \gamma \), to replace \( \gamma \) in (2.3.5) to form an alternative possible test statistic. Consequently, the proposed form of the MixMod \( F \)-statistic becomes:

\[
F_{\text{REML}} = \frac{\theta'_{\text{REML}}[c_1(x'\hat{\gamma}^{-1}_{\text{REML}} x)c_1']^{-1}\theta_{\text{REML}}/c}{\sigma^2_{\text{REML}}}
\]

where

\[
\theta_{\text{REML}} = c_1\beta|_{\beta = \hat{\beta}_{\text{REML}}} - \theta_0
\]

and

\[
\hat{\gamma}^{-1}_{\text{REML}} = \Gamma^{-1}|_{\gamma = \hat{\gamma}_{\text{REML}}}
\]

for appropriate REML estimators.

As outlined in the discussion in Section 2.2, two alternatives are available for obtaining "REML" estimates of \( \sigma^2, \beta, \) and \( \gamma \)--the classic REML methods as described by
Harville (1974 and 1977) and Laird and Ware (1982) and the Helms-McCarroll method. The estimates derived by the first method will be denoted as \( \hat{\beta}_r, \hat{\sigma}_r^2, \) and \( \hat{\gamma}_r, \) while those derived by the second method will be denoted as \( \hat{\beta}_n, \hat{\sigma}_n^2, \) and \( \hat{\gamma}_n. \) First, consider the corresponding test statistics based on the classic REML estimates, which is denoted by \( F_r. \)

Using the form in (2.3.6) this proposed statistic can be written as:

\[
F_r = \frac{\theta'_r [c_1(x'_r \hat{r}^{-1}_r x)^{-1} c'_1]^{-1} \theta_r / c}{\hat{\sigma}_r^2}
\]  

(3.3.7)

and for convenience define:

\[
\hat{\theta}_r = c_1(x'_r \hat{r}^{-1}_r x)^{-1} c'_1
\]  

(3.3.8)

Note that although \( \hat{\theta}_r \) is not the variance of \( \theta, \) it does have the same form as \( \text{Var}(\theta) \) with \( \gamma \) replaced by \( \hat{\gamma}. \)

In developing the Helms-McCarroll form of the test statistic, note that \( \beta, \) and hence \( \theta, \) are no longer estimated directly from \( y \) and \( x, \) but rather from transformed versions. Consequently the form of the statistic is adjusted accordingly and becomes:

\[
F_n = \frac{\theta'_n [c_1(x'_n T_n \hat{r}^{-1}_n T'_n x)^{-1} c'_1]^{-1} \theta_n / c}{\hat{\sigma}_n^2}
\]  

(2.3.9)

Again, for convenience in later derivations, define:

\[
\hat{\theta}_n = c_1(x'_n T_n \hat{r}^{-1}_n T'_n x)^{-1} c'_1
\]  

(2.3.10)

The remaining section of this chapter will develop a more explicit characterization of the "quadratic forms" in the numerators and denominators of these proposed test statistics.
2.4 FURTHER CHARACTERIZATION OF THE COMPONENTS OF $F_r$ AND $F_H$

Following the notation used in Section 2.3, note that $\hat{\mu}_r$ and $\hat{\mu}_H$ are functions of $\hat{y}_r$ and $\hat{y}_H$, respectively. Because both of the variables are nonlinear functions of $y$, the functions involving $\hat{\mu}_r$ and $\hat{\mu}_H$ in the numerators of the test statistics presented above are not quadratic forms in a strict sense. However, conditional on $\hat{y}_r$ and $\hat{y}_H$, they are quadratic forms in $y$, and for notational convenience, they will be referred to as quadratic forms in the remainder of the document.

In order to simplify the presentation of results concerning alternative forms of the MixMod $F$-statistics that are discussed in subsequent sections, define the quadratic forms shown below. For the classic REML estimates, the quadratic forms are:

$$Q_{1r} = \theta_r' \left[ y' \left( x' \hat{r}_r^{-1} x \right)^{-1} \right] \theta_r$$

$$= y' \left( \hat{r}_r^{-1} x \right)^{-1} y' \left[ \left( x' \hat{r}_r^{-1} x \right)^{-1} \right] \left[ \left( x' \hat{r}_r^{-1} x \right)^{-1} \right] y$$

$$= y' \left( \hat{r}_r^{-1} x \right)^{-1} \left[ \left( x' \hat{r}_r^{-1} x \right)^{-1} \right] y \quad (2.4.1)$$

and

$$Q_{2r} = (y - x \hat{\beta}_r - z \hat{\delta}_r)' (y - x \hat{\beta}_r - z \hat{\delta}_r)$$

$$= y' (y - x \hat{\beta}_r - z \hat{\delta}_r)$$

$$= y' \left( I - \left[ x \left( x' \hat{r}_r^{-1} x \right)^{-1} x' \hat{r}_r^{-1} \right] \right) - \left[ z \Delta_r, z' \hat{r}_r^{-1} \left( I - \left( x \left( x' \hat{r}_r^{-1} x \right)^{-1} x' \hat{r}_r^{-1} \right) \right) \right] y \quad (2.4.2)$$

Note that conditionally on $\hat{y}_r$, the numerator of the classic REML statistic is a quadratic form in both $\theta$ and $y$, while the denominator is a quadratic form in $y$.

For the Helms-McCarroll form of the proposed statistic,
the comparable quadratic forms are:

\[ Q_{2H} = w_2' w_2 = y' T_2 T_2' y \]  \hspace{1cm} (2.4.3)

and

\[
Q_{1H} = \theta_n' \left[ C_1 (x' T_1 \hat{\beta}_n^{-1} T_1' x)^{-1} C_1' \right]^{-1} \theta_n \\
= \theta_n' M_n^{-1} \theta_n \\
= w_1' \left\{ \hat{\beta}_n^{-1} T_1' x \left[ x' T_1 \hat{\beta}_n^{-1} T_1' x \right]^{-1} C_1' \hat{\beta}_n^{-1} C_1 \cdot \\
\left[ x' T_1 \hat{\beta}_n^{-1} T_1' x \right]^{-1} \left[ x' T_1 \hat{\beta}_n^{-1} T_1' x \right] \right\} w_1 \\
= y' \left\{ T_1 (T_1' \hat{\beta}_n T_1)^{-1} T_1' x \left[ x' T_1 (T_1' \hat{\beta}_n T_1)^{-1} T_1' x \right]^{-1} C_1' \hat{\beta}_n^{-1} C_1 \cdot \\
\left[ x' T_1 (T_1' \hat{\beta}_n T_1)^{-1} T_1' x \right]^{-1} x' T_1 (T_1' \hat{\beta}_n T_1)^{-1} T_1' \right\} y
\]  \hspace{1cm} (2.4.4)

Hence, conditional on \( \hat{\gamma}_n \), the numerator of the Helms-McCarroll form of the MixMod test statistic can be expressed as a quadratic form in \( \theta, w_1, \) or \( y \). However, the denominator is unconditionally a quadratic form in \( w_2 \) or \( y \). The specific distributional properties of these quadratic forms and their ratios are explored in more detail in Chapter 3.
CHAPTER 3

THE DISTRIBUTIONAL PROPERTIES OF THE MixMod TEST STATISTIC

In Chapter 2, alternative forms of a possible MixMod test statistics were developed based on parameter estimates obtained from both the classic REML and the Helms-McCarroll estimating procedures. The forms of the statistic were F-like in that they comprised ratios of "quadratic forms" divided by degrees of freedom. In this chapter, an appropriate MixMod test statistic is developed further by exploring the distributional properties of the different components of those statistics. Also, a method for obtaining an approximation of the cumulative distribution function for the final recommended test statistic is derived.

As the distributional properties of the quadratic forms were examined, two conclusions were reached. First, the forms developed from the Helms-McCarroll procedure had potential distributional advantages over the forms developed from the classic REML procedures. Consequently, the development focuses on the estimators developed via the Helms-McCarroll procedure, although the numerical differences in the estimators generated by the two procedures will be discussed further in Chapter 4. Second, direct ratios of the quadratic forms, which generate statistics with a beta-type appearance, appear to have analytical and, possibly, computational advantages over the F-type forms developed in Chapter 2. Consequently, the development of the distributional properties is based on such forms, although those forms can be transformed into F-type forms as illustrated in Chapter 4.

The remainder of the chapter is divided into three sections. Section 3.1 establishes distributional notation
that will be used throughout the remainder of the chapter. Section 3.2 explores the distributional properties of the "quadratic forms", $Q_{1M}$ and $Q_{2M}$ that are generated by the Helms-McCarroll estimation procedure and develops an expansion of the distribution as a mixture. This mixture provides a convenient procedure for approximating the cumulative distribution of the ratio. Section 3.2 also examines the distributional properties of a ratio of those forms when the estimate of the covariance parameters is assumed to be "fixed." However, because $Q_{1M}$ and consequently the MixMod statistic do depend on the covariance parameter estimates, Section 3.3 examines the effect of the covariance parameter estimates on the values of $Q_{1M}$ and on the proposed test statistic, as well as on the expansion that can be used to approximate the cumulative distribution of that proposed statistic.

3.1 DISTRIBUTIONAL NOTATION

The primary distributional forms that are of interest in this chapter are the central and noncentral forms of the chi-squared, $F$, and gamma distributions and the distribution of a random variable that is the ratio of two independent random variables with gamma distributions. Because the notation throughout the chapter is relatively complicated, the paragraphs below provide a comprehensive overview of the notation that will be used. For reference, Table 3.1 provides a summary of the distributional notation.

Throughout the discussion, the symbol $X \sim \chi^2(v)$ is used to denote that the random variable $X$ is distributed as a central Chi-squared distribution with $v$ degrees of freedom. A random variable $X$ with this distribution has the density
TABLE 3.1. Summary of Distributional Terminology

<table>
<thead>
<tr>
<th>Distribution Description</th>
<th>Label for Dist'n</th>
<th>CDF Label</th>
<th>Density Fcn Label</th>
<th>Density Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chi-square--d.f.:v noncent:ξ</td>
<td>$\chi^2(v, \xi)$</td>
<td>$\chi^2_{v, \xi}(x)$</td>
<td>---</td>
<td>$\sum_{j=0}^{\infty} \frac{\xi^j e^{-\xi}}{j!} \cdot \frac{x^{v/2+j-1} e^{-x/2}}{2^{v/2+j} \Gamma\left(\frac{v}{2}+j\right)}$</td>
</tr>
<tr>
<td>Chi-square--d.f.:v noncent:0</td>
<td>$\chi^2(v)$</td>
<td>$\chi^2_v(x)$</td>
<td>---</td>
<td>$\frac{x^{v/2-1} e^{-x/2}}{2^{v/2} \Gamma\left(\frac{v}{2}\right)}$</td>
</tr>
<tr>
<td>Gamma--Shape:α Location:ξ Scale:ψ</td>
<td>$X=\psi Y$ Y~$G(\alpha, \xi/\psi)$</td>
<td>---</td>
<td>---</td>
<td>$\frac{(x-\xi)^{\alpha-1} \exp[-(x-\xi)/\psi]}{\psi^\alpha \Gamma(\alpha)}$</td>
</tr>
<tr>
<td>Gamma--Shape:α Location:ξ Scale:1</td>
<td>$G(\alpha, \xi)$ or $G(\alpha+\nu_t)$</td>
<td>$G_{\alpha, \xi}(x)$</td>
<td>$g_{\alpha, \xi}(x)$</td>
<td>$\sum_{k=0}^{\infty} \frac{e^{-\xi} \xi^k}{k!} g_{\alpha, k}(x), x &gt; \xi$</td>
</tr>
<tr>
<td>Distribution Description</td>
<td>Label for Dist'n</td>
<td>CDF Label</td>
<td>Density Fcn Label</td>
<td>Density Function</td>
</tr>
<tr>
<td>--------------------------</td>
<td>------------------</td>
<td>-----------</td>
<td>------------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>Gamma</td>
<td>$G(\alpha)$</td>
<td>$G_{\alpha}(x)$</td>
<td>$g_{\alpha}(x)$</td>
<td>$\frac{x^{\alpha-1}e^{-x}}{\Gamma(\alpha)}$, $x&gt;0$</td>
</tr>
<tr>
<td>Ratio</td>
<td>$R(\alpha_1, \alpha_2, \xi)$</td>
<td>$R_{\alpha_1,\alpha_2,\xi}(x)$</td>
<td>$r_{\alpha_1,\alpha_2,\xi}(x)$</td>
<td>$\sum_{k=0}^{\infty} \frac{e^{-\xi} \xi^k}{k!} r_{\alpha_1+k,\alpha_2}(x)$</td>
</tr>
<tr>
<td>Ratio</td>
<td>$R(\alpha_1, \alpha_2)$</td>
<td>$R_{\alpha_1,\alpha_2}(x)$</td>
<td>$r_{\alpha_1,\alpha_2}(x)$</td>
<td>$\frac{x^{\alpha_1-1}(1-x)^{-(\alpha_1+\alpha_2)}}{B(\alpha_1, \alpha_2)}$, $x&gt;0$</td>
</tr>
<tr>
<td>F</td>
<td>$F(v_1, v_2, \xi)$</td>
<td>$F_{v_1, v_2, \xi}(x)$</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>F</td>
<td>$F(v_1, v_2)$</td>
<td>$F_{v_1, v_2}(x)$</td>
<td>---</td>
<td>---</td>
</tr>
</tbody>
</table>
function:

\[ f_x(x) = \frac{\chi^{v/2-1}e^{-x/2}}{2^{v/2} \Gamma(\frac{v}{2})} \]  

(3.1.1)

The symbol \( X \sim \chi^2(v, \xi) \) is used to denote that the random variable \( X \) is distributed as a noncentral Chi-squared distribution with \( v \) degrees of freedom and noncentrality parameter \( \xi \). To avoid possible ambiguity, the noncentrality parameter is defined as follows. If \( Y \sim N_v(\mu, I) \), then \( X=Y'Y \sim \chi^2(v, \xi) \) with \( \xi=\mu'\mu/2 \). Note that if \( X=Y'Y \sim \chi^2(v, \xi) \) then the density function of \( X \) can be expressed as (Johnson and Kotz, 1970b):

\[ f_x(x) = \sum_{j=0}^{\infty} \frac{\xi^j e^{-\xi}}{j!} \cdot \frac{x^{v/2+j-1}e^{-x/2}}{2^{v/2+j} \Gamma(\frac{v}{2}+j)} \]  

(3.1.2)

Hence, a random variable that is distributed as a noncentral chi-square can be represented as a mixture of central chi-squared random variables with the mixing variable being a Poisson random variable, and the Poisson mixing variable has the noncentrality parameter of the chi-squared random variable as its mean.

The notation presented in the paragraphs below is developed with some modification using the notation and results presented by Saw (1988). First, note that Saw uses the symbol \( X \sim G(\alpha) \) to indicate that the random variable \( X \) is distributed as a gamma distribution with shape parameter \( \alpha \), location parameter \( 0 \) (i.e., a "central" gamma distribution), and unit scale parameter. A random variable with this distribution has the density function:

\[ f_x(x) = g_\alpha(x) = \frac{x^{\alpha-1}e^{-x}}{\Gamma(\alpha)} , \quad x>0 \]  

(3.1.3)

Next, let the symbol \( X \sim G(\alpha, \xi) \) denote that the random variable \( X \) is distributed as a Gamma distribution with shape parameter \( \alpha \), noncentrality parameter \( \xi \), and unit scale parameter.
parameter. Results presented by Saw (1988) and Johnson and Kotz (1970) are sufficient to demonstrate that the density of the random variable with distribution \( G(\alpha, \xi) \) can be represented in two ways:

\[
f_x(x) = g_{\alpha, \xi}(x) = \frac{(x-\xi)^{\alpha-1}e^{-(x-\xi)}}{\Gamma(\alpha)}, \; x > \xi \quad (3.1.4a)
\]

or

\[
f_x(x) = g_{\alpha, \xi}(x) = \sum_{k=0}^{\infty} \frac{e^{-\xi \xi^k}}{k!} g_{\alpha+k}(x), \; x > \xi \quad (3.1.4b)
\]

Hence, the density for the noncentral gamma distribution can be represented as a mixture that is an infinite sum of densities of central gamma densities with the mixing variable being a Poisson random variable with a mean equal to the noncentrality parameter.

In keeping with the characterization of the noncentral gamma distribution as a mixture of central gamma distributions, Saw (1988) suggested alternative terminology for signifying that a random variable has such a distribution. If \( X \sim G(\alpha, \xi) \), then the alternative representation is \( X \sim G(\alpha + P_\xi) \), where \( Y \sim P_\xi \) indicates that \( Y \) is distributed as a Poisson random variable with mean \( \xi \). Saw (1988) notes that a random variable of this type is equivalent to one generated in two steps. First a random variable \( K \sim P_\xi \) is generated; then, conditional on \( K=k \), a random variable \( X - G(\alpha + k) \) is generated.

A very general form of the gamma distribution is represented by a random variable with a gamma distribution with shape parameter \( \alpha \), location parameter \( \xi \), and scale parameter \( \psi \), which has the density function:
\[ f_X(x) = \frac{(x - \xi)^{\alpha - 1} \exp[-(x - \xi)/\psi]}{\psi^\alpha \Gamma(\alpha)} \] (3.1.5)

Note that if \( X \sim G(\alpha, \xi) \) and \( Y = bX \), then the random variable \( Y \) is distributed as a gamma random variable with shape parameter \( \alpha \), scale parameter \( b \), and location parameter \( b\xi \). In order to simplify notation, the symbol \( X = \psi Y \) where \( Y \sim G(\alpha, \xi/\psi) \) will be used to denote that the random variable \( X \) is distributed as a gamma distribution with shape parameter \( \alpha \), scale parameter \( \psi \), and location parameter \( \xi \). The density of the random variable \( X = \psi Y \) with the \( Y \sim G(\alpha, \xi) \) also can be expressed as a mixture as follows:

\[ f_X(x) = \sum_{k=0}^{\infty} \frac{e^{-\xi} \xi^k}{k!} \cdot \frac{x^{\alpha-k-1} e^{x/\psi}}{\psi^\alpha \Gamma(\alpha+k)} \] (3.1.6)

Furthermore, note that if \( X \sim \chi^2(\nu, \xi) \) then \( X = 2Y \), where \( Y \sim G(\nu/2, \xi) \). (Note that one advantage of defining the chi-squared noncentrality parameter \( \xi \) as \( \mu'\mu/2 \) rather than as \( \mu'\mu \) is this equivalence to the location parameter in the density function of the gamma distribution.)

The next general type of random variable of interest in this chapter arises when a random variable is derived as the ratio of two independent gamma random variables. The density function of the central version of this random variable is generated as follows. Suppose that \( Y \sim G(\alpha_1, \xi) \) independently of \( Z \sim G(\alpha_2) \). First, consider the central form with \( \xi = 0 \). Then if \( X = Y/Z \), \( X \) has a density function of the form (Saw, 1988):

\[ f_X(x) = r_{\alpha_1, \alpha_2}(x) = \frac{x^{\alpha_1-1}(1-x)^{-(\alpha_1+\alpha_2)}}{B(\alpha_1, \alpha_2)} , \quad x > 0 \] (3.1.7)

where
\[ B(\alpha_1, \alpha_2) = \frac{\Gamma(\alpha_1) \Gamma(\alpha_2)}{\Gamma(\alpha_1 + \alpha_2)} \]

If a random variable \( X \) has the density defined above, then we will use the notation \( X \sim R(\alpha_1, \alpha_2) \). The noncentral form of the distribution, denoted by \( X \sim R(\alpha_1, \alpha_2, \xi) \), arises when \( Y \sim G(\alpha_1, \xi) \) independently of \( Z \sim G(\alpha_2) \); \( X = Y/Z \) has the density function:

\[ f_x(x) = r_{\alpha_1, \alpha_2, \xi}(x) = \sum_{k=0}^{\infty} \frac{e^{-\xi} \xi^k}{k!} r_{\alpha_1 + k, \alpha_2}(x) \quad (3.1.8) \]

Obviously, the central version (3.1.7) can be obtained from (3.1.8) when the location (or noncentrality) parameter is zero. In keeping with the alternative notation used for the noncentral gamma distribution following (3.1.4), the notation \( X \sim R(\alpha_1 + \mu, \alpha_2) \) will be used as an alternative to \( X \sim R(\alpha_1, \alpha_2, \xi) \). Here, the notation can be interpreted in a fashion that parallels the interpretation following (3.1.4).

The final general types of random variables considered in this chapter are those with central and noncentral \( F \) distributions. If \( X_1 \sim \chi^2(v_1, \xi) \) independently of \( X_2 \sim \chi^2(v_2) \), then the random variable \( Y = (X_1/v_1)/(X_2/v_2) \) is said to have a noncentral \( F \) distribution with \( v_1 \) numerator degrees of freedom, \( v_2 \) denominator degrees of freedom and noncentrality parameter \( \xi \). The symbol that will be used to represent this distribution is \( Y \sim F(v_1, v_2, \xi) \). The central \( F \) distribution is defined as above with \( \xi = 0 \), and \( Y \sim F(v_1, v_2) \) will denote that the random variable \( Y \) is distributed as a central \( F \) distribution with \( v_1 \) and \( v_2 \) degrees of freedom in the numerator and denominator, respectively.

Throughout the remainder of this chapter, \( f_x(x) \) will be used as the generic symbol for the density function of a random variable \( X \). The functional symbols \( g \) and \( r \) will be reserved for the specific densities defined above.
Furthermore, if $X \sim G(\alpha)$, then the cumulative distribution function of $X$ will be denoted by $G_x(\alpha)$, and the cumulative distribution function for the other gamma and ratio random variables defined above will be denoted accordingly. Table 3.1 provides a succinct summary of these symbols.

3.2 MixMod QUADRATIC FORM DISTRIBUTIONAL PROPERTIES

The results presented in Section 2.4, demonstrated that, conditional upon a "fixed" estimate of $\gamma$, the numerator and denominator of the proposed $F$-statistics developed from both the classic REML and Helms-McCarron procedures contain forms that are quadratic forms in $\gamma$. (Again, note that these forms are not strictly quadratic forms in that the weight matrix contains random variables that are not independent of $\gamma$. However, such forms will be called quadratic forms throughout this discussion.) Further examination of the forms suggests that $Q_{1r}$ (2.4.1) and $Q_{1m}$ (2.4.4) are comparable in structure and are likely to present equivalent difficulties relative to distributional considerations. However, the distributional form of $Q_{2r}$ (2.4.3) appears to be much more tractable than that of $Q_{2m}$ (2.4.2). Furthermore, as demonstrated below, the independence of $Q_{1m}$ and $Q_{2m}$ follows readily from the results in Sections 2.3 and 2.4, while the relationship between $Q_{1r}$ and $Q_{2r}$ is not at all apparent. Consequently, the remainder of this section focuses on the distributional properties of $Q_{1m}$ and $Q_{2m}$. First, two results that follow readily from earlier results, the distribution of $Q_{2m}$ and the independence of $Q_{2m}$ and $Q_{1m}$ are established. Then, results on quadratic forms in normal random variables presented by Johnson and Kotz (1970) are used to characterize the distribution of $Q_{1m}$, given that it is a function of a nonstochastic variable $\tilde{\gamma}$. Finally, results established by Saw (1988) are used to characterize the distribution of the ratio of $Q_{1m}$ to $Q_{2m}$.
3.2.1 Distribution of $Q_{2H}$

First, the fact that $Q_{2H}/\sigma^2$ is distributed as a central chi-square random variable follows directly from its estimation as the quadratic form of independent normal random variables, which are obtained from the orthogonal transformation of the data.

Theorem 3.1 Under the general MixMod distributional assumptions established in (1.2.2) through (1.2.9) and with $Q_{2H}$ defined at (2.4.3), $Q_{2H}/\sigma^2 \sim \chi^2(N-b)$, or equivalently, $Q_{2H}/\sigma^2 = 2Y$, where $Y \sim \mathcal{G}((N-b)/2)$.

Proof: From (2.2.17), $\mathbf{w}_2 \sim \mathcal{N}_{N-b}(0, \sigma^2 I)$. Hence the elements of $\mathbf{w}_2/\sigma$ are independently distributed $\mathcal{N}(0,1)$, and by definition,

$$\frac{Q_{2H}}{\sigma^2} = \frac{\mathbf{w}'_2 \mathbf{w}_2}{\sigma^2} - \chi^2(N-b)$$

3.2.2 Independence of $Q_{1H}$ and $Q_{2H}$

Next, the independence of $Q_{1H}$ and $Q_{2H}$ can be demonstrated statistically based on the independence of $\mathbf{w}_1$ and $\mathbf{w}_2$, or algebraically based on the orthogonality of $\mathbf{T}_1$ and $\mathbf{T}_2$. The first approach is used in the theorem below.

Theorem 3.2 The quadratic form $Q_{2H}$ is distributed independently of $Q_{1H}$.

Proof: By definition (2.4.4), $Q_{2H}$ is a scalar function of the random vector $\mathbf{w}_2$, while $Q_{1H}$ is by definition (2.4.3) a function of the random vectors $\hat{\mathbf{r}}_H$ and $\hat{\mathbf{b}}_H$. Both of these vectors are vector functions of the random vector $\mathbf{w}_1$. From the results presented
in Section 2.2.2 (see (2.2.23) ff), \( W_1 \) and \( W_2 \) are independent multivariate normal random vectors. Consequently, \( Q_{1N} \) and \( Q_{2N} \) are functions of independent sets of random variables and therefore are independent random variables.

3.2.3 Distribution of \( Q_{1N} \)

The results in Section 2.4 demonstrated that if \( \gamma \) is considered to be a fixed value, then \( Q_{1N} \) can be expressed as a quadratic form in \( Y, W_1, \) or \( \theta \). Because examination of the different forms suggests that, at least with respect to notation, the form in \( \theta \) appears to be most tractable, the results below are developed using \( \theta \). However, because \( \gamma \), and consequently \( \tilde{\theta} \), is a very complicated nonlinear function of the random data vector, the unconditional distribution of \( Q_{1N} \) does not appear to be readily derivable. Consequently, the discussion below follows an analytical rather than a stochastic approach to evaluate the properties of \( Q_{1N} \).

First, define \( \tilde{\gamma} = \tilde{\gamma}(y) \) to be a vector-valued nonstochastic variable that is a function of a fixed vector of observations \( y \). Then, define the corresponding set of variables:

\[
\begin{align*}
\tilde{\Gamma}_1 &= \tilde{\Gamma}_1(y) = \Gamma_1(y - \gamma) \\
\tilde{\theta} &= \tilde{\theta}(y) = \tilde{\theta}_N(y - \gamma) \\
\tilde{\mu} &= \tilde{\mu}(y) = \tilde{\mu}_N(y - \gamma) \\
\theta &= \theta(y) = c_i\tilde{\theta} \\
\hat{Q}_{1N} &= \hat{Q}_{1N}(y) = \tilde{\theta}'\tilde{\mu}^{-1}\tilde{\theta} 
\end{align*}
\tag{3.2.1}
\]

In the paragraphs below, the distribution of \( \theta \) for a fixed \( \gamma \) is developed. Then the distribution of the quadratic form \( \hat{Q}_{1N} \) is characterized as a weighted sum of noncentral chi-squared random variables based upon that fixed value of \( \gamma \). Finally, results presented by Saw (1988) are used to characterize \( \hat{Q}_{1N} \) as a mixture of central Gamma
random variables, and that characterization is used to characterize the distribution of ratio $\hat{Q}_{1n}$ to $Q_{2n}$.

**Theorem 3.3:** For any nonstochastic value $\tilde{\gamma}$ in the parameter space of $\gamma$, the random vector $\tilde{\theta}$ is distributed as a MVN random vector with mean $\theta$ and with variance structure as shown below.

$$\text{Var}(\tilde{\theta}) = \sigma^2 C_1 \left[ X' T_1 \tilde{\Gamma}_1^{-1} T_1' X \right]^{-1} X' T_1 \tilde{\Gamma}_1^{-1}.$$  \hspace{1cm} (3.2.2)

$$\quad r_1 \tilde{\Gamma}_1^{-1} T_1' X \left[ X' T_1 \tilde{\Gamma}_1^{-1} T_1' X \right]^{-1} C_1'$$

**Proof:** Note that

$$\tilde{\theta} = C_1 \tilde{\beta}$$

$$= C_1 \left[ X' T_1 \tilde{\Gamma}_1^{-1} T_1' X \right]^{-1} X' T_1 \tilde{\Gamma}_1^{-1} T_1' \gamma$$ \hspace{1cm} (3.2.3)

$$= C_1 \gamma$$

Hence, under the assumption that $\tilde{\gamma}$ is nonstochastic, the only random variables in the above equation are contained in the random vector $\gamma$, and as a linear transformation of a MVN random vector $\tilde{\theta}$ is a MVN random vector with mean and covariance structure as shown below.

$$\text{E}(\tilde{\theta}) = C_1 \left[ X' T_1 \tilde{\Gamma}_1^{-1} T_1' X \right]^{-1} X' T_1 \tilde{\Gamma}_1^{-1} T_1' \text{E}(\gamma)$$

$$= C_1 \left[ X' T_1 \tilde{\Gamma}_1^{-1} T_1' X \right]^{-1} X' T_1 \tilde{\Gamma}_1^{-1} T_1' X \beta$$

$$= C_1 \beta = \theta$$

and

88
\[ \text{Var}(\tilde{\theta}) = C_\beta \text{Var}(y) C_\beta' \]
\[ = C_\beta \Sigma C_\beta' \]
\[ = \sigma^2 C_\beta \Gamma C_\beta' \]

Substitute the appropriate form for \( C_\beta \) as given in (3.2.3) into the above result to obtain the final result shown in (3.2.2).

The results of Theorem 3.3 can be used to demonstrate that \( \tilde{Q}_{1\mathbb{N}} \) is distributed as the weighted sum of noncentral chi-squared random variables. The weights are the eigenvalues of a linear transformation of the product of the variance of \( \tilde{\theta} \) and \( \tilde{\Sigma}^{-1} \). These results are based on transformations of the data that use the decomposition properties of positive definite, symmetric matrices described below.

Let \( \text{Var}(\tilde{\theta}) = \sigma^2 V_{\tilde{\theta}} \). Note that because \( \text{Var}(\tilde{\theta}) \) is a positive definite, symmetric matrix, \( V_{\tilde{\theta}} \) can be factored via the Cholesky decomposition as:

\[ V_{\tilde{\theta}} = LL' \]

where \( L \) is a nonsingular lower triangular matrix. Also, note that as a consequence of its method of generation, \( \tilde{\Sigma} \) is positive definite and symmetric and hence, so is \( \tilde{\Sigma}^{-1} \). Consequently, we obtain the factorization:

\[ L' \tilde{\Sigma}^{-1} L = \Lambda \Omega \Lambda' \]

where \( \Lambda \) is a diagonal matrix comprising the eigenvalues of \( L' \tilde{\Sigma}^{-1} L \), and \( \Omega \) is the orthogonal matrix of the corresponding eigenvectors. These properties are now used to establish the distribution of \( \tilde{Q}_{1\mathbb{N}} \).
Theorem 3.4: Given Ŝ defined as above, then

\[ \frac{\ddot{Q}_{1W}}{\sigma^2} = \sum_{j=1}^{j} \lambda_j y_j \]

where

\[ y_j \sim \chi^2(1, \xi_j), \text{ and } \xi_j = \frac{\zeta^2_j}{2} \]

where \( \lambda_j \) are the eigenvalues of \( L'M^{-1}L \), and consequently of \( V_{\delta}M^{-1} \), and \( \zeta_j \) are obtained from a linear transformation of \( \theta \), as demonstrated in the proof.

Proof: As denoted above, let \( \sigma^2 V_{\delta} = \text{Var}(\delta) \), and note from the results of Theorem 3.3 that

\[ \delta \sim N_c(\theta, \sigma^2 V_{\delta}) \]

where \( c = \text{Rank} (C_1). \)

Now, define:

\[ z = PL^{-1} \delta \]

and

\[ \zeta = PL^{-1} \theta \]

Hence,

\[ z \sim N[PL^{-1} \theta, PL^{-1} \sigma^2 V_{\delta} (L^{-1})'P' \]

\[ = N[\zeta, \sigma^2 I] \]

indicating that the elements of \( z, z_j \), are independent normal random variables with variance \( \sigma^2 \) and means \( \zeta_j \). Using these results, rewrite \( \ddot{Q}_{1W} \) as follows:
\[ \hat{Q}_{1W} = \hat{\delta}' \hat{\Pi}^{-1} \hat{\delta} \]
\[ = z' P L' \hat{\Pi}^{-1} L P' z \]
\[ = z' \Lambda z \]
\[ = \sum_{j=1}^{J} \lambda_j z_j^2 \]

Divide both sides of the equation by \( \sigma^2 \) to obtain:

\[ \frac{\hat{Q}_{1W}}{\sigma^2} = \sum_{j=1}^{J} \lambda_j \left( \frac{z_j}{\sigma} \right)^2 \]
\[ = \sum_{j=1}^{J} \lambda_j \chi_j \]

Hence, \( \hat{Q}_{1W}/\sigma^2 \) is distributed as a weighted sum of independent random variables which are by definition \( \chi^2 \) random variables with 1 degree of freedom with the noncentrality parameters noted.

Although the above results present a partial characterization of the distribution of \( \hat{Q}_{1W} \), they have limitations with respect to characterizing the distribution of the ratio of \( \hat{Q}_{1W} \) to \( \hat{Q}_{2W} \). To address these limitations, we want to develop a more convenient form for the distribution of \( \hat{Q}_{1W} \). This development relies heavily on results developed by Saw (1988) for use in computing the non-null distributions in ANOVA problems.

In developing his results, Saw (1988), summarizing the work of earlier authors concerning the gamma distributions that were noted in (3.1.4), demonstrated that if \( X \sim G(\alpha, \xi) \), then the distribution of \( X \) can be written as a mixture of central gamma distributions, with unit scale parameter, indexed on the shape parameter. The mixing variable is a Poisson random variable with mean \( \xi \). Saw went on to develop a similar mixture result for problems characterized by mixing variables that are called generalized Poisson or
contagious random variables. The paragraphs below summarize those results and apply them to the characterization of \( \tilde{Q}_{1m} \).

Suppose we have a situation in which \( \lambda \) is a given constant, with \( X = \lambda Y_1 \), where \( Y_1 \sim G(\alpha, \xi) \), and \( \eta \) has an arbitrary value with \( 0 < \eta \leq \lambda \). Saw demonstrated that:

\[
X = \lambda Y_1 \quad \text{where} \quad Y_1 \sim G(\alpha, \xi) \\
\Rightarrow \quad X = \eta Y_2 \quad \text{where} \quad Y_2 \sim G(\alpha + \phi, \xi) \tag{3.2.3}
\]

where

(a) \( \phi \) is an integer-valued random variable with a distribution defined by the probability generating function:

\[
\sum_{k=0}^{\infty} p_k s^k = e^{\phi(s)} \tag{3.2.4}
\]

with

\[
\phi(s) = \sum_{j=1}^{\infty} \left[ \xi \frac{\eta}{\lambda} \left(1 - \frac{\eta}{\lambda}\right)^{j-1} + \frac{\alpha (1 - \frac{\eta}{\lambda})^j}{j} \right] s^j
\]

and

\[
\phi = \phi(1) = \xi - \alpha \log \frac{\eta}{\lambda}
\]

and

(b) Following the spirit of the notation introduced after (3.1.4), the symbol \( G(\alpha+\phi) \) denotes a random variable that is a mixture of random variables with central gamma distributions with mixing variable \( Y = \phi \). Such a random variable can be generated in a two step process. First, generate a random variable \( K = \phi \); then conditional on the event \( K=k \), generate a variable distributed as \( G(\alpha+k) \).

Saw went on to demonstrate that the probability mass function of \( \phi \) can be defined constructively through the recursive relationship shown below.
\[ p_0 = \exp \left[ \xi - \alpha \cdot \log_e \left( \frac{\eta}{\lambda} \right) \right] \]

and

\[ p_k = \sum_{j=0}^{k-1} \left[ \frac{k-j}{k} \right] p_j \phi_{k-j}, \quad 1 \leq k \leq \infty \quad (3.2.5) \]

where

\[ \phi_k = \frac{\xi \eta}{\lambda} \left( 1 - \frac{\eta}{\lambda} \right)^{k-1} + \frac{\alpha \left( 1 - \frac{\eta}{\lambda} \right)^k}{k} \]

In the presentation above, note that the probability generating function for the mixing distribution is defined as an exponential function of \( \phi(s) \). Under the assumption that the random variables that lead to the mixing distributions are independent, then \( \phi_{s_1}, \ldots, \phi_{s_n} \) can also be assumed to be independent. Consequently, because the probability generating function of the sum of independent random variables is the product of the probability generating functions, we have \( \phi_{s_1} + \cdots + \phi_{s_n} = \phi_{s_1 + \cdots + s_n} \).

The final tool needed to complete the characterization of \( \bar{q}_{1n} \) is established in the lemma below. This lemma establishes the distribution of the sum of independent random variables with such mixing distributions.

**Lemma 3.5:** Suppose \( X_1 \sim G(\alpha_1 + \phi_{s_1}) \) independently of \( X_2 \sim G(\alpha_2 + \phi_{s_2}) \) where \( \phi_{s_i} \) is a random variable defined on the integers with probability generating function:

\[ \sum_{k \geq 0} p_{ik} s^k = \exp[\phi_{s_i}(s) - \phi_i] \]

defined as in (3.2.4) and (3.2.5), above, and \( Y = X_1 + X_2 \). Then \( Y \sim G(\alpha_1 + \alpha_2 + \phi_{s_1+s_2}) \).

**Proof:** First, let \( s \in [0,1] \) and let \( s \) and \( t \) be related by \( s = (1-t)^{-1} \), where \( t \in (-\infty, 0] \). Note that the

93
density function of $X_i$ is:

$$f_{X_i}(x) = \sum_{k=0}^{\infty} \frac{p_{ik} x^{a_i+k-1} e^{-x}}{\Gamma(\alpha_i+k)}$$  \hspace{1cm} (3.2.6)

Hence, the moment generating function of $X_i$ for $t \in (-\infty, 0]$ is defined as:

$$M_{X_i}(t) = \int_0^\infty e^{tx} f_{X_i}(x) \, dx$$

$$= \int_0^\infty \left[ \sum_{k=0}^{\infty} p_{ik} \frac{x^{a_i+k-1} e^{-x}}{\Gamma(\alpha_i+k)} \right] dx$$

$$= \int_0^\infty \left[ \sum_{k=0}^{\infty} p_{ik} \frac{x^{a_i+k-1} e^{-(1-t)x}}{\Gamma(\alpha_i+k)} \right] dx$$

Assume for the moment that the integration and summation operators can be interchanged. Hence, the series can be integrated term by term using the transformations $s=(1-t)^{-1}$ and $y=(1-t)x$ defined above to obtain:

$$M_{X_i}(t) = \sum_{k=0}^{\infty} \frac{p_{ik}}{\Gamma(\alpha_i+k)} \left[ \int_0^\infty x^{a_i+k-1} e^{-(1-t)x} \, dx \right]$$

$$= \sum_{k=0}^{\infty} \frac{p_{ik}}{\Gamma(\alpha_i+k)} \left[ \int_0^{\infty} (sy)^{a_i+k-1} e^{-y} \, dy \right]$$

$$= \sum_{k=0}^{\infty} \frac{p_{ik} s^{a_i+k}}{\Gamma(\alpha_i+k)} \left[ \int_0^{\infty} y^{a_i+k-1} e^{-y} \, dy \right]$$

Note that

94
\[ \Gamma(n_i + k) = \int_0^\infty y^{n_i+k-1} e^{-y} \, dy \]

and so

\[ M_{x_i}(t) = s^{n_i} \sum_{k=0}^\infty p_{ik} s^k \quad (3.2.7) \]

But, from (3.2.4), the above series can be expressed as an exponential function of \( \psi_i(s) \) to obtain:

\[ M_{x_i}(t) = s^{n_i} e^{\psi_i(s) - n_i} \quad (3.2.8) \]

The interchange of the sum and integral above can be justified as follows. From Kestelman (1970), we know that if \( f_1, f_2, \ldots \) is a sequence of nonnegative functions defined on \([0, \infty)\) such that

\[ f(x) = \sum_{k=0}^\infty f_k(x) \]

and the following conditions hold:

(a) \[ \int_0^x f_k(x) \, dx \text{ exists for } 0 < x < \infty, \quad k \geq 0 \]

(b) \[ \int_0^x f(x) \, dx \text{ exists for } 0 < x < \infty \]

(c) \[ \sum_{k=0}^\infty \int_0^x f_k(x) \, dx \text{ is finite} \]

then
\[ \int f(x) \, dx = \sum_{k=0}^{\infty} \int f_k(x) \, dx \]

Looking at the three conditions that must be demonstrated, first note that
\[ f_k(x) = p_{ik} \frac{x^{a_i+k-1} e^{-(1-t)x}}{\Gamma(\alpha_i+k)}, \quad x \geq 0, \quad t \in (-\infty, 0] \]

Hence,
\[ \int_{0}^{x} f_k(x) \, dx = \frac{p_{ik}}{\Gamma(\alpha_i+k)} \int_{0}^{x} x^{a_i+k-1} e^{-(1-t)x} \, dx \]

But, note that \( x^{a_i+k-1} e^{-(1-t)x} \) is continuous and hence integrable on \([0, X]\) (Bartle and Sherbet, 1982, p.251). Consequently, so is \( f_k(x) \).

In order to demonstrate that the second criterion holds, we will first show that the partial sum converges uniformly on \([0, X]\), and then by interchange of the sum and integral, demonstrate the integrability of \( f \). Define \( S \) and \( S_n \) as the infinite and partial series:

\[ S(x) = \sum_{k=0}^{\infty} p_{ik} \frac{x^{a_i+k-1} e^{-tx}}{\Gamma(\alpha_i+k)}, \quad x \geq 0 \]

and

\[ S_n(x) = \sum_{k=0}^{n} p_{ik} \frac{x^{a_i+k-1} e^{-tx}}{\Gamma(\alpha_i+k)}, \quad x \geq 0 \]

Now, let \( \epsilon > 0 \) and \( X \in (0, \infty) \) be given, and note that because \( f_k(x) \) is nonnegative,
\[ |S(x) - S_n(x)| = S(x) - S_n(x) \]

\[ = \sum_{k=n+1}^\infty p_{ik} \frac{x^{\alpha_i+k-1} e^{-(1-t)x}}{\Gamma(\alpha_i+k)} \]

\[ = \sum_{k=n+1}^\infty p_{ik} (1-t)^{-(\alpha_i+k-1)} \frac{[(1-t)x]^{\alpha_i+k-1} e^{-(1-t)x}}{\Gamma(\alpha_i+k)} \]

Note that \( p_{ik} \leq 1 \) together with \((1-t) \geq 1\) imply that

\[ S(x) - S_n(x) \leq \sum_{k=n+1}^\infty \frac{[(1-t)x]^{\alpha_i+k-1} e^{-(1-t)x}}{\Gamma(\alpha_i+k)} \]

However, note that because \( S(x) - S_n(x) \) is a continuous function, \( \exists x^* \in [0, X] \) such that

\( S(x) - S_n(x) \) achieves a maximum on \([0, X]\) at \( x^* \).

Let \( y = (1-t)x^* \) and note that

\[ S(x) - S_n(x) \leq \sum_{k=n+1}^\infty \frac{y^{\alpha_i+k-1} e^{-y}}{\Gamma(\alpha_i+k)}, \quad x \in [0, X] \]

Because the gamma function increases monotonically in its argument, \( \Gamma(\alpha_i+k) \geq \Gamma(m+k) \), where \( m = \lfloor \alpha_i \rfloor \) is the greatest integer less than or equal to \( \alpha_i \).

Hence, letting \( j = k+m-1 \), we have

\[ S(x) - S_n(x) \leq \sum_{k=n+1}^\infty \frac{y^{\alpha_i+k-1} e^{-y}}{\Gamma(m+k)} \]

\[ = y^{\alpha_i-m} \sum_{k=n+1}^\infty \frac{y^{m+k-1} e^{-y}}{(m+k-1)!} \]

\[ = y^{\alpha_i-m} \sum_{j=m-n}^\infty \frac{y^j e^{-y}}{j!} \]

But the sum in the last expression is simply \( \Pr(X \geq m+n) \) where \( X \sim \text{Poi}(y) \). Because this probability goes to 0 as \( n \to \infty \), \( \exists N^* \) such that
\[ n \geq N^* \text{ and } x \in [0, X] \implies \Pr(X \geq n) < \frac{\varepsilon}{y^{\alpha_i - m}} \]

\[ = S(x) - S_n(x) < \varepsilon \]

Hence, \( S_n \to S \) uniformly on \([0, X]\) and so

\[ \int_0^x S(x) \, dx = \sum_{k=0}^{\infty} \int_0^x f_k(x) \, dx \]

\[ \leq \sum_{k=0}^{\infty} \int_0^x f_k(x) \, dx \]

\[ = s^{\alpha_i} \sum_{k=0}^{\infty} p_k s^k \leq 1, \quad s \in [0, 1] \]

and \( f \) is integrable on \([0, X]\).

The statements above also demonstrate that the third criterion is met, and consequently, the interchange of operations is justified. Hence,

\[ M_Y(t) = M_{X_1}(t) \cdot M_{X_2}(t) \]

\[ = s^{\alpha_1 + \alpha_2} \exp[(\Phi_1(s) - \phi_1) + (\Phi_2(s) - \phi_2)] \]

But, the sequence of equalities used above to establish the moment generating function of \( X_i \) can be used to show that the moment generating function calculated for \( Y \) is equivalent to the moment generating function for a random variable that is generated as a mixture of central gamma random variables with a generalized Poisson mixing variable of the type defined earlier. When the moment generating function exists, it is unique and completely determines the distribution of a random variable (Hogg and Craig, 1978, p 50ff). Consequently, \( Y \) is distributed as a mixture of central gamma random variables with a generalized Poisson mixing variable. Furthermore, that random
variable has shape parameter \( \alpha_1 + \alpha_2 \) and a mixing variable of the form \( \theta_{\alpha_1+\alpha_2} \). Hence,

\[
Y = G(\alpha_1 + \alpha_2 + \theta_{\alpha_1+\alpha_2})
\]

Note that the result can be extended by induction to any finite sum of such random variables.

The results from Theorem 3.4 and Lemma 3.5 can be combined with the results of Saw (1988) to develop a "closed-form" characterization of the distribution of \( \hat{Q}_{1M} \). The resulting form is the product the minimum eigenvalue defined in Theorem 3.4, and a random variable that is a mixture of central gamma random variables with a generalized Poisson mixing variable as shown in the theorem below.

**Theorem 3.6:** Let \( \eta = \min(\lambda_j, j=1,2,\ldots,J) \) where the \( \lambda \) are defined as in Theorem 3.4. Then

\[
\hat{Q}_{1M}/\sigma^2 = 2\eta X
\]

where

\[
X = G(\frac{1}{2} + \theta_{\epsilon_1, \ldots, \epsilon_J})
\]

with

\[
\phi_j(s) = \sum_{k=1}^{\infty} \left[ \frac{\xi_j \eta (1 - \eta)}{\lambda_j} \left( \frac{1 - \eta}{\lambda_j} \right)^{k-1} + \left( \frac{1 - \eta}{\lambda_j} \right)^k \right] s^k
\]

**Proof:** From Theorem 3.4

\[
\hat{Q}_{1M}/\sigma^2 = \sum_{j=1}^{J} 2 \lambda_j X_j
\]

where

\[
X_j = G\left( \frac{1}{2}, \xi_j \right)
\]

Hence, from (3.2.3)
\[ \frac{\hat{Q}_{1n}}{\sigma^2} = \sum_{j=1}^{\eta} Y_j = \eta \sum_{j=1}^{\eta} Y_j \]

where

\[ Y_j \sim G\left(\frac{1}{2}, \varphi_{\theta_j}\right) \]

and where the \( Y_j \) are independent with \( \varphi_{\theta_j} \) defined as in (3.2.3) and (3.2.4). But, from Lemma 3.5

\[ Y = \sum_{j=1}^{\eta} Y_j \]

\[ = Y - G\left(\frac{1}{2}, \varphi_{\theta_1, \ldots, \theta_\eta}\right) \]

Finally, define \( \phi_j \) as in (3.2.4) and the proof is complete.

The results presented by Saw also can be used to characterize the distribution of the ratio of \( \hat{Q}_{1n} \) to \( Q_{2n} \), as described below. First, let \( \hat{R}_n = \frac{\hat{Q}_{1n}}{Q_{2n}} \). Then, as established in Theorem 3.7, \( \hat{R}_n \) is distributed as a ratio variable of the form defined in (3.1.7) and (3.1.8). (As a note, recall from the discussion following (3.1.4) and (3.1.8), a random variable \( X \sim R(\alpha_1 + \varphi_{\phi}, \alpha_2) \) is equivalent to a random variable generated in two steps. First, an integer-valued random variable \( K \sim \varphi_{\phi} \), and then, conditional on \( K = k \), a random variable \( X, R(\alpha_1 + k, \alpha_2) \) is generated.)

**Theorem 3.7:** Let \( \eta \) and \( \phi(s) \) be defined as in Theorem 3.6, and let \( \hat{R}_n = \frac{\hat{Q}_{1n}}{Q_{2n}} \). Then \( \hat{R}_n = \eta X \), where

\[ X \sim R\left(\frac{1}{2} + \varphi_{\theta_1, \ldots, \theta_\eta}, \frac{n-\varphi}{2}\right) \]

and the \( \phi_j \) are defined in Theorem 3.6.
Proof: From Theorem 3.6 \( \hat{\Omega}_{1H} / \sigma^2 = 2\eta X_1 \), where
\[
X_1 = G(\frac{j}{2} + \psi_1, \ldots, \psi_j),
\]
and from Theorem 3.1,
\[
Q_{2H} / \sigma^2 = 2X_2, \quad \text{where} \quad X_2 = G(\frac{n-b}{2}).
\]
Furthermore, because \( \hat{\Omega}_{1H} \) and \( Q_{2H} \) are independent, so are \( X_1 \) and \( X_2 \). Hence,
\[
\hat{R}_H = \frac{\hat{\Omega}_{1H}}{Q_{2H}} = \frac{\hat{\Omega}_{1H} / \sigma^2}{Q_{2H} / \sigma^2} = \frac{2\eta X_1}{2X_2} = \eta \frac{X_1}{X_2}
\]
But because \( X_1 \) and \( X_2 \) are independent, the results from Saw (1988) are sufficient to demonstrate that
\[
Y = \frac{X_1}{X_2} - \hat{R}\left(\frac{j}{2} + \psi_1, \ldots, \psi_j\right),
\]
which completes the proof.

3.3 FURTHER CHARACTERIZATION OF \( \hat{\Omega}_{1H} \) AND \( \hat{R}_H \)

While Theorems 3.6 and 3.7 provide convenient expansions that characterize the distributions of \( \hat{\Omega}_{1H} \) and \( \hat{R}_H \) in terms of a fixed estimate of the covariance parameter vector \( \hat{\psi} \), they have limited value in actual applications. Because \( \Gamma \), typically is not known, \( \text{Var}(\hat{\theta}) \) cannot be calculated, and consequently, the weights \( \lambda_j \) cannot be determined explicitly. This section explores the problem further. First, the distributions are characterized when the estimates are equal to the true covariance parameter estimates, say \( \gamma_0 \). Then the behavior of the test statistics is examined when the estimates are "close" to the true parameters.

First, let \( \gamma_0 \) be the vector of true values of the covariance parameters, and assume \( \hat{\gamma} = \gamma_0 \). Then \( \hat{\Gamma}_1 = \Gamma_1 \), and
\[ \text{Var}(\hat{\theta}) = \sigma^2 C_1 \left[ \mathbf{X}' \Gamma_1^{-1} \mathbf{T}' \mathbf{X} \right]^{-1} \mathbf{X}' \Gamma_1^{-1} \mathbf{T}' \mathbf{X} C_1' \cdot (3.3.1) \]

\[ = \sigma^2 C_1 \left[ \mathbf{X}' \Gamma_1^{-1} \mathbf{T}' \mathbf{X} \right]^{-1} C_1' \]

From (2.3.10) and (3.2.1),

\[ \tilde{\mathbf{N}} = C_1 \left[ \mathbf{X}' \Gamma_1 \tilde{\Gamma}_1^{-1} \mathbf{T}' \mathbf{X} \right]^{-1} C_1' \cdot (3.3.2) \]

Hence, under the assumption that \( \mathbf{y} = \mathbf{y}_0 \),

\[ \text{Var}(\hat{\theta}) \cdot \tilde{\mathbf{N}}^{-1} = \sigma^2 I_j \]

and consequently, \( \lambda_j = 1, j=1,\ldots,J \). Note that the same results hold if \( \Gamma_1 \) is replaced with \( \tilde{\Gamma}_1 \) in (3.3.1).

Under these assumptions, the results of Theorem 3.4 reduce to:

\[ \tilde{Q}_{1H}/\sigma^2 = \sum_{j=1}^{J} Y_j, \quad Y - \chi^2(1, \xi_j) \]

or equivalently

\[ \tilde{Q}_{1H}/\sigma^2 = Y = 2X, \quad Y - \chi^2(j, \xi) \quad \text{and} \quad X \sim G\left(\frac{j}{2}, \xi\right) \]

where

\[ \xi = \sum_{j=1}^{J} \xi_j. \]

Also, the results for Theorem 3.6 reduce to:

\[ \Phi_j(s) = \xi_j s \]

\[ - \sum_{k=0}^{s-1} p_{jk} s^k = e^{\xi_j(s-1)}, \quad (3.3.3) \]

which is the probability generating function for a random variable \( Y \sim P_{\xi_j} \). Consequently, Theorem 3.6 yields the result.
\[ \frac{\hat{Q}^2}{\sigma^2} = 2X, \quad X \sim G\left(\frac{j}{2} + P_x \right), \quad \text{(3.3.4)} \]

which is equivalent to the distributional results shown above. Furthermore, based on these results,

\[ \tilde{R}_n = R\left(\frac{j}{2} + P_x, \frac{w-b}{2}\right). \quad \text{(3.3.5)} \]

These results lead to the observation that when \( \hat{\gamma} = \gamma_0 \), the distributional form of the equivalent alternative form of the test statistic, which was proposed by McCarrroll and Helms (1987) [see (2.3.9)], is the noncentral F distribution that they postulated as a reasonable distribution.

While the above analysis provides an explicit characterization of the MixMod test statistic when the covariance parameters are estimated exactly, its usefulness in practice is limited in that exact estimates are generally not obtained. The paragraphs below examine the behavior of the test statistic when \( \hat{\gamma} \) is "close" to \( \gamma_0 \) in some sense. First, an explicit definition of "closeness" is developed as the limit of a sequence of vectors. Then, that definition is used in concert with results from matrix perturbation theory to characterize the effect of inexact estimates of the covariance parameters on the estimates of \( Q^2_{1n} \). (Note that because the denominator of the \( F_n \) statistic is independent of the covariance parameter estimates, the effect of these estimates on the statistic is related to their effect on \( Q^2_{1n} \).)

In order to define "closeness" more formally, the notion of convergence as presented by Stewart and Sun (1990), Horn and Johnson (1985), and Lancaster and Tismenetsky (1985) is introduced. Suppose that \( \hat{\gamma} \) converges to \( \gamma_0 \) (\( \hat{\gamma} \rightarrow \gamma_0 \)). Here convergence is defined in terms of the limit of a sequence of vectors as follows. Let \( \{x_k\} \) be a sequence of vectors in \( \mathbb{R}^n \) of the form

\[ x_k = (x_1^{(k)}, \ldots, x_n^{(k)})', \quad \text{and let} \quad \xi \text{ be a fixed vector of the} \]

103
form $\xi = (\xi_1, \cdots, \xi_n)'$. Then the sequence of vectors $x_k$ is said to have the limit $\xi$ ($\lim_{k \to \infty} x_k = \xi$), or $x_k$ is said to converge to $\xi$ ($x_k \to \xi$), if $\lim_{k \to \infty} x_i^{(k)} = \xi_i$, $i = 1, \cdots, n$ (Stewart and Sun, 1990). Stewart and Sun (1990) go on to demonstrate that if $\| \cdot \|$ represents an arbitrary vector norm on $\mathbb{R}^n$, then

$$\lim_{k \to \infty} x_k = \xi \iff \lim_{k \to \infty} \| x_k - \xi \| = 0$$

Hence, component-wise convergence of a sequence of vectors is equivalent to convergence of that sequence in any norm.

Now, let $(\hat{\gamma}_k(y)) = (\hat{\gamma}_k)$ be a sequence of nonstochastic estimates of the vector of random effects covariance parameters with $\gamma_0$ the "true" population value for the random effects covariance parameters. Using definitions comparable to those in (3.2.1), let

$$\beta_0 = \beta(y) = \hat{\beta}|_{\gamma = \gamma_0}$$
$$\Gamma_0 = \Gamma_0(y) = \hat{\Gamma}|_{\gamma = \gamma_0}$$
$$\theta_0 = \theta_0(y) = \mathbf{c}_1 \beta_0$$
$$Q_{1H(0)} = Q_{1H(0)}(y) = \theta_0' \Gamma_0^{-1} \theta_0$$

and note that $\Gamma_1$ is the modified form of the covariance matrix as defined in (2.2.24). The theorem below establishes a type of "pointwise continuity" between $y(y)$ and $\hat{Q}_{1H}(y)$, in that $y(y) - y_0 - \hat{Q}_{1H}(y) - Q_{1H(0)}(y)$ for a fixed vector of observations $y$.

The proof of the theorem relies heavily on matrix perturbation results presented by Stewart and Sun (1990). In order to provide a basis for the proof, the vector and
matrix norm notation that will be used is introduced and some basic matrix perturbation results are summarized below.

- Let \( \mathbf{x} \in \mathbb{R}^n \), and let \( \mathbf{A} \) be an \( m \times n \) matrix (with \( n \leq m \)) of real elements. The 2-norm (or Euclidean norm) of \( \mathbf{x} \) is defined as:

\[
\| \mathbf{x} \|_2 = \sqrt{\sum_{i=1}^{n} x_i^2}
\]  

(3.3.7)

The Frobenius norm of \( \mathbf{A} \) is defined as:

\[
\| \mathbf{A} \|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij}^2} = \left[ \text{trace}(\mathbf{A}'\mathbf{A}) \right]^{1/2}
\]

(3.3.8)

Also note that the Frobenius norm and the 2-norm are consistent in that

\[
\| \mathbf{A} \mathbf{x} \|_F = \| \mathbf{A} \|_F \cdot \| \mathbf{x} \|_2
\]

- Matrix perturbations will be represented by the matrix \( \mathbf{E} \) (or by some indexed set of matrices \( \mathbf{E}_i \)), and perturbed matrices will be denoted by using the diacritical \( \tilde{\cdot} \). That is if a matrix \( \mathbf{A} \) is perturbed by a matrix \( \mathbf{E} \), we write \( \tilde{\mathbf{A}} = \mathbf{A} + \mathbf{E} \).

- If \( \tilde{\mathbf{A}} = \mathbf{A} + \mathbf{E} \), then the norm of the perturbation in \( \mathbf{A}^{-1} \) can be bounded by the relationship (Stewart and Sun, 1990):

\[
|\tilde{\mathbf{A}}^{-1} - \mathbf{A}^{-1}| \leq |\mathbf{A}^{-1}\mathbf{E}| \cdot |\mathbf{A}^{-1}|
\]

(3.3.9)

Furthermore,

\[
\tilde{\mathbf{A}}^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{E}\mathbf{A}^{-1}
\]

(3.3.10)

with this first order approximation accurate up to terms of order \( \| \mathbf{E} \|_2 \). That is
\[ |\bar{A}^{-1} - A^{-1} + A^{-1}EA^{-1}| = o(|E|^2) \]  

(3.3.11)

Suppose the elements of the matrix A are functions of the elements of the vector x. Also let \( \bar{A} = A|_{x=x} \) and \( A_0 = A|_{x=x_0} \). Then using concepts comparable to the convergence concept presented by Stewart and Sun (1990) and Lancaster and Tismenetsky (1985), the statement \( \bar{x} - x_0 \rightarrow \bar{A} - A_0 \) holds iff \( \forall \, \varepsilon > 0, \exists \, \delta > 0 \) such that

\[ |\bar{x} - x_0|_2 < \delta - |\bar{A} - A_0|_f < \varepsilon. \]

or equivalently.

\[ |\bar{x} - x_0|_2 < \delta - \bar{A} = A_0 + E, \text{ where } |E| = o(\varepsilon) \]

These concepts can be used to develop the relationship between the covariance parameter estimates and the numerator of the MixMod test statistic shown below. However, a lemma related to matrix products is developed first.

**Lemma 3.8:** Suppose the matrices A and B have elements that are functions of the elements of the vector x and that \( \bar{x} - x_0 \rightarrow \bar{A} - A_0 \) and \( \bar{B} - B_0 \).

Then \( \bar{x} - x_0 \rightarrow \bar{A} - \bar{B} - A_0 - B_0 \).

**Proof:** Let \( \varepsilon > 0 \) be given, and let \( \varepsilon^* = \varepsilon \cdot \frac{1}{k} \), where

\[ k = \max\{ |A_0|, |B_0|, 1 \}. \]

Then, \( \exists \, \delta_1 \) and \( \delta_2 \) such that

(i) \[ |\bar{x} - x_0|_2 < \delta_1 - |\bar{A} - A_0|_f < \varepsilon^*. \]

(ii) \[ |\bar{x} - x_0|_2 < \delta_2 - |\bar{B} - B_0|_f < \varepsilon^*. \]
(iii) \[ \| \tilde{x} - x_0 \|_2 < \delta_1 \rightarrow \tilde{A} = A_0 + E, \| E \|_F = \mathcal{O}(\epsilon^*) \]

Let \( \delta = \max(\delta_1, \delta_2) \). Then

\[
\| \tilde{A} \cdot \tilde{B} - A_0 \cdot B_0 \|_F = \| \tilde{A} \cdot \tilde{B} - \tilde{A} \cdot B_0 + \tilde{A} \cdot B_0 - A_0 \cdot B_0 \|_F
\]

\[
= \| \tilde{A} \cdot (\tilde{B} - B_0) + (\tilde{A} - A_0) \cdot B_0 \|_F
\]

\[
= \| A_0 \cdot (\tilde{B} - B_0) + E \cdot (\tilde{B} - B_0) + (\tilde{A} - A_0) \cdot B_0 \|_F
\]

\[
\leq \| A_0 \|_F \cdot \| \tilde{B} - B_0 \|_F + \| E \|_F \cdot \| \tilde{B} - B_0 \|_F + \| \tilde{A} - A_0 \|_F \cdot \| B_0 \|_F
\]

\[
< \| A_0 \|_F \cdot \epsilon^* + (\epsilon^*)^2 - \epsilon^* \cdot \| B_0 \|_F
\]

\[
< \epsilon
\]

which completes the proof.

**Theorem 3.9:** For a fixed observation vector \( y \) and a nonstochastic estimate \( \tilde{y}(y) \) of \( y \), if \( \tilde{y}(y) = \tilde{y} - y_0 \), then \( \tilde{Q}_1 - Q_{1_{W(y)}} \).

**Proof:** To establish the theorem, we will demonstrate that comparable results hold for \( \tilde{\theta} \) and \( \tilde{\theta}^{-1} \), and then use the above lemma to complete the proof. First, define the following matrix perturbations

- \( E_1 = \tilde{T}_1 - T_1 \)
- \( E_2 = \tilde{T}_1^{-1} - T_1^{-1} \)
- \( E_3 = X' T_1 \tilde{T}_1^{-1} T_1' X - X' T_1 T_1^{-1} T_1' X = X' T_1 E_2 T_1' X \)
- \( E_4 = [X' T_1 \tilde{T}_1^{-1} T_1' X]^{-1} - [X' T_1 T_1^{-1} T_1' X]^{-1} \)

Also define the following constants based upon the norms of fixed matrices and vectors used in the calculation of the fixed effects estimates and test statistics.
• \( k_1 = \| C_1 \|_f = \left[ \text{trace}(C'_1 C_1) \right]^{1/2} \)

• \( k_2 = \left\| \left[ x' T_1 \Gamma_1^{-1} T_1' x \right]^{-1} \right\|_f \)

• \( k_3 = \| x \|_f = \| x' \|_f = \left[ \text{trace}(x' x) \right]^{1/2} \)

• \( k_4 = \| \Gamma_1^{-1} \|_f \)

• \( k_5 = \| y \|_2 \)

• \( k_6 = \left\| \left[ \sum_{n=1}^{N-1} T_n G_n T_1 \right] \right\|_p \)

• \( k_7 = \left\| C_1 \left[ x' T_1 \Gamma_1^{-1} T_1' x \right]^{-1} \right\|_f \)

Also note that because the columns of \( T_1 \) form an orthonormal basis of \( M(x \| z) \), \( \| T_1 \|_f = r^{1/2} \), where \( r = \text{rank} \ (x \| z) \). Furthermore, for any matrix \( A \) that is conformable to \( T_1 \)

\[ \| T_1 A \| = \| A \| \]

Next, using the above results, we characterize the behavior of \( \delta \) and \( \bar{\Gamma}^{-1} \) as \( \bar{y} \rightarrow y_0 \).

Recall that

\[ \delta = C_1 \left[ x' T_1 \bar{\Gamma}_1^{-1} T_1' x \right]^{-1} x' T_1 \bar{\Gamma}_1^{-1} T_1' y \]

By replacing expressions involving \( \bar{\Gamma}_1^{-1} \) with appropriate expressions involving \( \Gamma_1^{-1} \), \( E_2 \), and \( E_4 \), we obtain:
\[ \delta = c_1 \left\{ \left[ x'T_1 \Gamma_1^{-1} T_1'x \right]^{-1} + e_4 \right\} x'T_1 \left\{ \Gamma_1^{-1} + e_2 \right\} T_1'y \\
= c_1 \left[ x'T_1 \Gamma_1^{-1} T_1'x \right]^{-1} x'T_1 \Gamma_1^{-1} T_1'y + c_1 \left[ x'T_1 \Gamma_1^{-1} T_1'x \right]^{-1} x'T_1 \epsilon_2 T_1'y \\
+ c_1 e_4 x'T_1 \Gamma_1^{-1} T_1'y + c_1 e_4 x'T_1 \epsilon_2 T_1'y \]

Because
\[ \theta_0 = c_1 \left[ x'T_1 \Gamma_1^{-1} T_1'x \right]^{-1} x'T_1 \Gamma_1^{-1} T_1'y, \]
we have
\[ \delta - \theta_0 = c_1 \left\{ \left[ x'T_1 \Gamma_1^{-1} T_1'x \right]^{-1} x'T_1 \epsilon_2 T_1' \\
+ e_4 x'T_1 \Gamma_1^{-1} T_1'y + e_4 x'T_1 \epsilon_2 T_1'y \right\} y \]

Because the triangle inequality holds for all matrix norms and because the Frobenius norm and 2-norm are consistent norms, the constants defined above can be used to generate
\[ \left\| \delta - \theta_0 \right\| = k_1 \left[ k_2 k_3 \left\| \epsilon_2 \right\|_f + k_3 k_4 \left\| \epsilon_4 \right\|_f + k_3 \left\| \epsilon_2 \right\|_f \cdot \left\| \epsilon_4 \right\|_f \right] k_5 \]

Recall that \[ E_1 = \sum_{m=1}^{N-1} \delta_m T_1' G_m' T_1, \] where \[ \delta_m = \gamma_m - \gamma_{0m}. \]

Let \[ \delta^{\ast}_1 = \min(\delta_m, \ m=1, \cdots, M-1), \] and note that
\[ \left\| E_1 \right\|_f \leq \delta^{\ast}_1 k_6. \] This bound on the norm of \( E_1 \), along with the matrix perturbation results presented in (3.3.9) through (3.3.11) can be used to obtain the following bounds for the norms of \( E_2 \) and \( E_4 \).
\[ \|E_2\|_f \leq \left\| \Gamma_1^{-1} \right\|_f \cdot \|E_1\|_f \cdot \left\| \tilde{\Gamma}_1^{-1} \right\|_f \]
\[ \leq \left\| \Gamma_1^{-1} \right\|^2_\infty \cdot \|E_1\|_f \cdot \left[ 1 + \left\| \Gamma_1^{-1} \right\|_f \cdot \|E_1\|_f + o(\|E_1\|^2_\infty) \right] \]
\[ = k_2^5 \|E_1\|_f + o(\|E_1\|^2_\infty) + o(\|E_1\|^2_\infty) \]

and similarly,
\[ \|E_4\| \leq k_2^5 \|E_3\|_f + o(\|E_3\|^2_\infty) + o(\|E_3\|^2_\infty) \]
\[ \leq k_2^5 k_3^2 \|E_2\|_f + o(\|E_2\|^2_\infty) + \ldots \]
\[ \leq k_2^5 k_3^2 k_4^2 \|E_1\|_f + o(\|E_1\|^2_\infty) + \ldots \]

Hence,
\[ \|\delta - \theta_0\| \leq k_1 k_2 k_3 k_4^2 k_5 \left[ \|E_1\|_f + k_2 k_3^2 k_4 \|E_1\|_f \right] \]
\[ + k_2 k_3^2 \|E_1\|_f + o(\|E_1\|^2_\infty) \]
\[ = \|E_1\|_f k_1 k_2 k_3 k_4^2 k_5 \left[ 1 + k_2 k_3^2 k_4 + k_2 k_3^2 k_4^2 \right] \]
\[ + k_1 k_2 k_3 k_4^2 k_5 o(\|E_1\|^2_\infty) \]

Let \( \delta^*_1 = \epsilon / k_{(1)} \), where
\[ k_{(1)} = k_1 k_2 k_3 k_4^2 k_5 \left[ 1 + k_2 k_3^2 k_4 + k_2 k_3^2 k_4^2 \right] \]

Hence, \( |\gamma_m - \gamma_0| < \delta^*_1 \rightarrow \|\delta - \theta_0\|_2 < \epsilon \), and so if \( \tilde{\gamma} - \gamma_0 \), then \( \delta = \theta_0 \) and \( \delta' = \theta_0 \).

Next, we consider the behavior of \( \tilde{M}^{-1} \) as \( \tilde{\gamma} - \gamma_0 \). Note that
\[ \tilde{M} = c_1 \left[ X'T_1 \tilde{\Gamma}_1^{-1} T'_1 X \right]^{-1} c_1' \]

Hence, \( \tilde{M} - M_0 = c_1 E_4 c_1' \), and so
\[ \| \tilde{M} - M_0 \|_f \leq k_1 \| E_4 \|_f \leq k_1^2 k_2^2 k_3^2 k_4^2 \| E_1 \|_f + O(\| E_1 \|_f^2) + \cdots \]

Then an argument similar to the one used above can be used to demonstrate that

\[ \| \tilde{M}^{-1} - M_0^{-1} \|_f \leq k_1^2 \| M_0^{-1} \|_f \cdot \| E_4 \|_f + O(\| E_4 \|_f^2) \leq k_1^2 k_2^2 k_3^2 k_4^2 \| E_1 \|_f + O(\| E_1 \|_f^2) + \cdots \]

Now, let \( \delta_2^* = \epsilon / k_{(2)} \), where

\[ k_{(2)} = k_1^2 k_2^2 k_3^2 k_4^2 \]

Then \( |\gamma_m - \gamma_0| < \delta_2^* = \| \tilde{M}^{-1} - M_0^{-1} \|_2 < \epsilon \), and so if \( \theta - \theta_0 \), then \( \tilde{M}^{-1} - M_0^{-1} \). The convergence of \( Q_{1m} \) to \( Q_{1m(0)} \) follows directly from matrix perturbation results on the convergence of products demonstrated in the Lemma above.

The above theorem demonstrates a continuity-type relationship between the covariance parameter estimates and the value of the proposed MixMod test statistic. However, because the bounds developed in the proof are quite rough, the results provide little indication of how "close" the covariance parameter estimates need to be in practice to provide acceptably accurate estimates for the test statistic. Further insight can be gathered by examining the effect of perturbations in the covariance parameter estimates on the coefficients of the expansion of the test statistic discussed earlier. Recall from the earlier discussion that if \( \tilde{\gamma} = \gamma_0 \), or that if the covariance parameter estimates, which are treated as nonstochastic (i.e., if we condition on \( \tilde{\gamma} \)), are used in the expression
for \( \text{Var}(\hat{\Theta}) \), then the test statistic behaves like a weighted sum of gamma random variables with the weights being the probability mass values from a Poisson distribution. The discussion below examines the behavior of the eigenvalues of \( V_{\hat{\Theta}} \cdot \tilde{M}^{-1} \) and then characterizes the behavior of the expansion developed for \( \hat{Q}_{1n}/\sigma^2 \) in Theorem 3.6. First, a result related to the eigenvalues of a matrix product in which one of the matrices is positive definite and symmetric (p.d.s.) is established. The paragraphs that follow discuss the behavior of the eigenvalues of \( V_{\hat{\Theta}} \cdot \tilde{M} \) and show that if \( \hat{\gamma} - \gamma_0 \), then the generating function for the weights obtained from using \( \hat{\gamma} \) converges to the generating function established in (3.3.4) for the case in which \( \gamma_0 \) is used.

**Lemma 3.10:** If \( A \) and \( B \) are symmetric matrices of the same size with \( B \) p.d.s., then the eigenvalues of \( AB \) are the same as those of \( SAS' \), where \( S'S \) is any square root decomposition (see Searle, 1982) of \( B \).

**Proof:** Suppose \( \lambda^* \) is an eigenvalue of \( AB \). Then \( \lambda^* \) is a solution of the characteristic equation 
\[
|AB - \lambda I| = 0.
\]
But,
\[
|AB - \lambda I| = 0 \quad - |S^{-1}SAS'S - \lambda S^{-1}S| = 0
\]
\[
- |S^{-1}(SAS' - \lambda I)S| = 0
\]
\[
- |S^{-1}|SAS^{-1} - \lambda I||S| = 0
\]
However, because \( B \) is p.d.s., \( S \) has a nonzero determinant. Hence, the roots of the characteristic equation of \( AB \) are the same as those of the characteristic equation of \( SAS' \), and they therefore must have the same eigenvalues.
In order to examine the behavior of the expansion developed in Theorem 3.6 as \( \gamma - \gamma_0 \), recall from Theorem 3.3 that for nonstochastic \( \gamma \) and \( \tilde{\Gamma} \),

\[
\text{Var}(\tilde{\theta}) = \sigma^2 C_1 \left[ X' T_1 \tilde{\Gamma}^{-1}_1 T_1' X \right]^{-1} X' T_1 \tilde{\Gamma}^{-1}_1 \Gamma_1 \tilde{\Gamma}^{-1}_1 T_1' X \left[ X' T_1 \tilde{\Gamma}^{-1}_1 T_1' X \right]^{-1} C_1',
\]

and hence,

\[
\text{Var}(\tilde{\theta}/\sigma^2) = \nu_{\tilde{\theta}} \text{Var}(\tilde{\theta}) = C_1 \left[ X' T_1 \tilde{\Gamma}^{-1}_1 T_1' X \right]^{-1} X' T_1 \tilde{\Gamma}^{-1}_1 \Gamma_1 \tilde{\Gamma}^{-1}_1 T_1' X \left[ X' T_1 \tilde{\Gamma}^{-1}_1 T_1' X \right]^{-1} C_1'.
\]

Furthermore, note that as \( \gamma - \gamma_0 \), \( \tilde{\Gamma} \) can be represented as

\[ \tilde{\Gamma}_1 = \Gamma_1 - E, \]

where

\[
E = \sum_{m=1}^{K-1} (\gamma_m - \gamma_m) T_1' Z (I_k \otimes G_m) Z' T_1
\]

(3.3.14)

Note that except for the values of \( \gamma \) and \( \gamma_0 \), the matrices that define \( E \) are fixed and known.

Next, to examine the behavior of the eigenvalues of \( \nu_{\tilde{\theta}} \cdot \tilde{\Gamma}^{-1} \), replace \( \Gamma_1 \) in (3.3.13) with \( \tilde{\Gamma}_1 + E \) and note that

\[
\nu_{\tilde{\theta}} = C_1 \left[ X' T_1 \tilde{\Gamma}^{-1}_1 T_1' X \right]^{-1} C_1' + C_1 \left[ X' T_1 \tilde{\Gamma}^{-1}_1 T_1' X \right]^{-1} X' T_1 \tilde{\Gamma}^{-1}_1 \Gamma_1 \tilde{\Gamma}^{-1}_1 T_1' X \left[ X' T_1 \tilde{\Gamma}^{-1}_1 T_1' X \right]^{-1} C_1'
\]

\[ + E \Gamma_1 \tilde{\Gamma}^{-1}_1 T_1' X \left[ X' T_1 \tilde{\Gamma}^{-1}_1 T_1' X \right]^{-1} C_1' \]

(3.3.15)

Hence,

\[ \nu_{\tilde{\theta}} = C_1 \left[ X' T_1 \tilde{\Gamma}^{-1}_1 T_1' X \right]^{-1} C_1' + E, \]

Hence,
\[ \mathbf{v}_0 \cdot \tilde{M}^{-1} = \mathbf{I} + \mathbf{E}_1 \tilde{M}^{-1} \quad (3.1.16) \]

Because \( \tilde{M} \) is p.d.s., the results from Lemma 3.10 can be used to demonstrate that the eigenvalues of \( \mathbf{v}_0 \cdot \tilde{M}^{-1} \) are equal to the eigenvalues of \( \mathbf{S} \mathbf{v}_0 \mathbf{S}' \), which are equal to the eigenvalues of \( \mathbf{I} + \mathbf{S} \mathbf{E}_1 \mathbf{S}' \), where \( \tilde{M}^{-1} = \mathbf{S}' \mathbf{S} \) is any square root decomposition of \( \tilde{M}^{-1} \).

Now, let \( \tilde{M}^{-1} = \mathbf{U}_n \mathbf{A}_n \mathbf{U}_n' \). Then, based on the results presented above, the eigenvalues of \( \mathbf{v}_0 \cdot \tilde{M}^{-1} \) are the same as the eigenvalues of \( \mathbf{I} + \mathbf{U}_n' \mathbf{A}_n^{1/2} \mathbf{E}_1 \mathbf{A}_n^{1/2} \mathbf{U}_n \). Note that the manner in which they are constructed ensures that both \( \mathbf{E} \) and \( \mathbf{E}_1 \) are symmetric, and consequently, so is \( \mathbf{U}_n' \mathbf{A}_n^{1/2} \mathbf{E}_1 \mathbf{A}_n^{1/2} \mathbf{U}_n \). Because the eigenvalues of \( \mathbf{v}_0 \cdot \tilde{M}^{-1} \) are the same as those of a matrix that is a symmetric perturbation of the identity matrix, results presented by Stewart and Sun (1990) and Wilkinson (1965) are sufficient to demonstrate that if \( \lambda_j \) is an eigenvalue of \( \mathbf{v}_0 \cdot \tilde{M}^{-1} \), then \( \lambda_j \epsilon [1 + \epsilon_1, 1 + \epsilon_j] \), where \( \epsilon_1 \leq \epsilon_2 \leq \ldots \leq \epsilon_j \) are the eigenvalues of \( \mathbf{E}_1 \tilde{M}^{-1} \), and hence of \( \mathbf{U}_n' \mathbf{A}_n^{1/2} \mathbf{E}_1 \mathbf{A}_n^{1/2} \mathbf{U}_n \).

Let \( \epsilon = \max(\epsilon_j, \epsilon_j - \epsilon_1) \), and recall from Theorem 3.6 that

\[ \Phi_j(s) = \sum_{k=1}^{\epsilon_j} \left[ \frac{\epsilon_j \eta}{\lambda_j} (1 - \frac{\eta}{\lambda_j})^{k-1} + \frac{(1 - \frac{\eta}{\lambda_j})^k}{k} \right] s^k \quad (3.1.16) \]

where the \( \lambda_j \) are eigenvalues of \( \mathbf{v}_0 \cdot \tilde{M}^{-1} \) and \( \eta \) is the smallest of these eigenvalues. Note that

114
\[(i) \quad 1 - \frac{\eta}{\lambda_j} = \frac{\lambda_j - \eta}{\lambda_j} \]

\[(ii) \quad 0 \leq 1 - \frac{\eta}{\lambda_j} \leq \epsilon \]

\[(iii) \quad \frac{\eta}{\lambda_j} = 1 - \left(1 - \frac{\eta}{\lambda_j}\right) \]

Hence,

\[
\phi_j(s) = \sum_{k=1}^{\infty} \left[ \frac{\xi_j \eta}{\lambda_j} \left(1 - \frac{\eta}{\lambda_j}\right)^{k-1} + \frac{(1 - \frac{\eta}{\lambda_j})^k}{k} \right] s^k
\]

\[
= \sum_{k=1}^{\infty} \left\{ \xi_j \left[\left(1 - \frac{\eta}{\lambda_j}\right)^{k-1} - \left(1 - \frac{\eta}{\lambda_j}\right)^k \right] + \frac{(1 - \frac{\eta}{\lambda_j})^k}{k} \right\} s^k
\]

\[
= \left\{ \xi_j \left[1 - \mathcal{O}(\epsilon) \right] - \mathcal{O}(\epsilon) \right\} s + \left\{ \xi_j \left[\mathcal{O}(\epsilon) - \mathcal{O}(\epsilon^2) \right] + \mathcal{O}(\epsilon^2) \right\} s^2 + \cdots
\]

Note that as \(\epsilon \to 0\), \(\phi(s) \to \xi_j \) and hence

\[
\sum_{k=0}^{\infty} p_{jk} s^k \to e^{\xi_j(s-1)}
\]

which is the probability generating function for a Poisson random variable with mean \(\xi_j\).

The above results suggest that the departure of the test statistic from the hypothesized approximate distribution is related to the degree to which the covariance parameter estimates depart from the true parameters. While the results do not provide a complete analytical description of how errors in the covariance parameter estimates are reflected to the departure of the test statistic from its hypothesized approximate distribution, they do provide a tool for analyzing the behavior numerically. The results of such a numerical analysis are presented in Chapter 4.
CHAPTER 4
NUMERICAL EVALUATION OF THE MixMod F STATISTIC

The principal objective of this research is to develop and evaluate a test statistic that can be used to test hypotheses about and develop confidence regions for linear contrasts among fixed effect parameters in the general linear mixed model. Such a statistic has practical applications in both study design and data analyses for observational and experimental studies that generate longitudinal data. The distributional properties of a proposed MixMod approximate F-test statistic, $F^*_\nu$, (or more appropriately, a ratio statistic that is a linear transformation of $F^*_\nu$) were developed in Chapter 3. That analysis demonstrated that as the covariance parameter estimates converge to the true covariance parameters, the approximate statistic converges to a variable with the hypothesized F distribution. However, the information presented there did not establish the convergence rates, nor did it characterize the relationship between the magnitude of the test statistic "error" and the errors in the covariance parameter estimates. Consequently, questions still remain about how well the statistic will perform in practical applications to longitudinal studies. Questions also still remain about the practical differences between the statistics generated via the classic REML and Helms-McCarroll procedures.

This chapter describes numerical studies that address these questions to a limited degree. Multiple computer-generated data sets that were constructed using population parameters based on studies of pulmonary function in children reported by Fairclough (1984) were used for the
analyses. These data sets are described in Section 4.1

Section 4.2 presents a comparison of the fixed and random effect parameter estimates and MixMod test statistics that were generated from these data using the classic REML and Helms-McCarroll estimation procedures. Also in Section 4.2, the values of test statistics generated by the classic REML and Helms-McCarroll procedures are compared to the weighted least squares (WLS) test statistic obtained using the "true" random effects covariance parameters. This statistic was selected because results from Section 3.3 are sufficient to show that the WLS statistic has an exact F distribution for both central and noncentral cases.

Section 4.3 addresses questions about how differences between the covariance parameter estimates and the true parameters affect the test statistic and its associated p-values. Covariance parameter "estimates" were obtained by perturbing the "true" covariance parameters for the data sets described in Section 4.1. The values of $F_n$ obtained with these perturbed estimates are compared to the weighted least squares F statistic ($F_{WLS}$).

4.1 DATA USED TO EVALUATE THE TEST STATISTIC

Fairclough (1984) discussed the analyses of respiratory data from a study of pulmonary function in young children. Multiple measurements of forced vital capacity (FVC) were obtained over time from children grouped by race (black and white). These analyses found FVC to linearly related to height for both study groups with both the slope and the intercept of this linear relationship affected by race. The relationships found by Fairclough are shown graphically in Figure 4.1. The random effects covariance parameters were found to be comparable for the two groups. Fairclough’s results were used to develop two groups of data sets to conduct the analyses discussed in this chapter.

The first group of data, denoted as the Fairclough-
based data, was developed using the parameter estimates obtained by Fairclough as "true" population parameters. Twenty independent data sets were generated. (The observations from the first data set were included as an example in the illustration of the model in Figure 4.1.) Each data set contains observations for 40 subjects, 20 from each group. For each subject, 5 observations were generated at different heights. The five observations were obtained at about 10 cm intervals of height centered at 115 cm, which is also the location of the model intercept. Subjects' heights were varied by randomly adding the integer component of a random variable with a N(0,4) distribution to the 5 base levels (0, ±10, ±20). The parameters used in the model are presented in Table 4.1. The data were generated using the SAS interactive matrix language (IML).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Symbol</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept-black</td>
<td>$\beta_1$</td>
<td>1.11</td>
</tr>
<tr>
<td>Slope-black</td>
<td>$\beta_2$</td>
<td>0.029</td>
</tr>
<tr>
<td>Intercept-white</td>
<td>$\beta_3$</td>
<td>1.31</td>
</tr>
<tr>
<td>Slope-white</td>
<td>$\beta_4$</td>
<td>0.036</td>
</tr>
<tr>
<td>Variance-intercept</td>
<td>$\tau_1$</td>
<td>0.018</td>
</tr>
<tr>
<td>Covariance-slope/intercept</td>
<td>$\tau_2$</td>
<td>0.00056</td>
</tr>
<tr>
<td>Variance-slope</td>
<td>$\tau_3$</td>
<td>0.000042</td>
</tr>
<tr>
<td>Error variance</td>
<td>$\sigma^2$</td>
<td>0.0116</td>
</tr>
<tr>
<td>Var Ratio-intercept</td>
<td>$\gamma_1$</td>
<td>1.55</td>
</tr>
<tr>
<td>Cov Ratio-slope/intercept</td>
<td>$\gamma_2$</td>
<td>0.048</td>
</tr>
<tr>
<td>Var Ratio-slope</td>
<td>$\gamma_3$</td>
<td>0.0036</td>
</tr>
</tbody>
</table>

Subsequent analyses of the Fairclough-based model data indicated that the values of the test statistics for the hypothesis of interest were quite large for all data sets. Consequently, a second group of data sets was generated to
analyze the procedures over more varied conditions. These data sets had a larger underlying error variance and different combinations of fixed effects to provide a wider range of test statistics. This second group, denoted as the Modified FVC or FVC-Mod2 data, contains 30 independently-generated data sets, 3 each with the 10 different combinations of fixed effect parameters listed in Table 4.2.

<table>
<thead>
<tr>
<th>Data Sets</th>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
<th>$\beta_3$</th>
<th>$\beta_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-3</td>
<td>1.30</td>
<td>0.035</td>
<td>1.40</td>
<td>0.038</td>
</tr>
<tr>
<td>4-6</td>
<td>1.25</td>
<td>0.035</td>
<td>1.45</td>
<td>0.038</td>
</tr>
<tr>
<td>7-9</td>
<td>1.30</td>
<td>0.033</td>
<td>1.40</td>
<td>0.040</td>
</tr>
<tr>
<td>10-12</td>
<td>1.25</td>
<td>0.033</td>
<td>1.45</td>
<td>0.040</td>
</tr>
<tr>
<td>13-15</td>
<td>1.20</td>
<td>0.035</td>
<td>1.50</td>
<td>0.038</td>
</tr>
<tr>
<td>16-18</td>
<td>1.30</td>
<td>0.031</td>
<td>1.40</td>
<td>0.042</td>
</tr>
<tr>
<td>19-21</td>
<td>1.20</td>
<td>0.031</td>
<td>1.50</td>
<td>0.042</td>
</tr>
<tr>
<td>22-24</td>
<td>1.10</td>
<td>0.035</td>
<td>1.60</td>
<td>0.038</td>
</tr>
<tr>
<td>25-27</td>
<td>1.30</td>
<td>0.027</td>
<td>1.40</td>
<td>0.046</td>
</tr>
<tr>
<td>28-30</td>
<td>1.10</td>
<td>0.027</td>
<td>1.60</td>
<td>0.046</td>
</tr>
</tbody>
</table>

A single vector of random effects covariance parameters and a single error variance were used as shown in Table 4.3.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>$\tau_1$</th>
<th>$\tau_2$</th>
<th>$\tau_3$</th>
<th>$\sigma^2$</th>
<th>$\gamma_1$</th>
<th>$\gamma_2$</th>
<th>$\gamma_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>0.2</td>
<td>0.0022</td>
<td>0.00001</td>
<td>0.01</td>
<td>20</td>
<td>0.22</td>
<td>0.01</td>
</tr>
</tbody>
</table>

4.2 COMPARISON OF PARAMETER ESTIMATES AND TEST STATISTICS

As outlined in Chapter 3, the distributional properties of the approximate F-statistic generated via the Helms-McCarroll estimating procedure can be addressed more
readily than those generated by the classic REML procedure. However, the classic REML procedure does have some computational advantages. Because of these computational advantages, a comparison of the estimates, particularly the covariance parameter estimates, generated by the two procedures is of interest.

The comparison of the procedures involved two sets of evaluations, both of which were primarily qualitative. First, the fixed effect parameter, random effect covariance parameter, and error variance parameter estimates for the two groups of data were examined. The Helms-McCarroll and classic REML estimates were compared to the "true" population values that were used to generate the data. For the fixed effect parameter estimates, these estimates were also compared to the WLS estimates, which represent "best estimates" for a particular data set. Second, test statistics were developed for the hypothesis that the intercepts (at 115 cm) and slopes for the relationship of FVC to height were equal for the two racial groups. The test statistics and p-values associated with those statistics for both the Helms-McCarroll and classic REML procedures were compared to the test statistics and p-values generated by the WLS procedure.

The parameter estimates and test statistics were generated for the classic REML procedure using the SAS procedure PROC MIXED with the REML option. The Helms-McCarroll and WLS estimates were generated from the estimating equations presented in Chapter 2 using an MSA algorithm. The algorithm was implemented with SAS IML code. 4.2.1 Comparison of the Parameter Estimates

Table 4.4 presents a summary of the estimates generated by the classic REML and Helms-McCarroll procedures, as well as those generated by the WLS procedure using the known underlying population covariance parameters for the Fairclough-based data. (Detailed estimates for each data set are presented in Appendix A.) Because all 20 data sets
<table>
<thead>
<tr>
<th>Parameter</th>
<th>True Value</th>
<th>Summary of Parameter Estimates Based on 20 Data Sets</th>
<th>Heims-McCarroll</th>
<th>Classic REML</th>
<th>WLS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Mean Std. Dev. Mean Std. Dev. Mean Std. Dev.</td>
<td>Mean Std. Dev.</td>
<td>Mean Std. Dev.</td>
<td>Mean Std. Dev.</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>1.11</td>
<td>0.0292 0.0289 0.0018</td>
<td>1.113 0.0265</td>
<td>1.114 0.0278</td>
<td>1.115 0.0018</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>1.31</td>
<td>0.0292 0.0320 0.0019</td>
<td>1.312 0.0286</td>
<td>1.312 0.0293</td>
<td>1.312 0.0023</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>1.55</td>
<td>0.0480 0.0313 0.0019</td>
<td>1.554 0.0499</td>
<td>1.554 0.0209</td>
<td>1.554 0.0014</td>
</tr>
<tr>
<td>$\sigma_2$</td>
<td>0.0036</td>
<td>0.0042 0.0022 0.0018</td>
<td>0.0036 0.0025</td>
<td>0.0042 0.0038</td>
<td>0.0036 0.0025</td>
</tr>
</tbody>
</table>

TABLE 4.4: Summary of Parameter Estimates for the Fairclough-Based Data
had the same underlying parameters, the estimates can be compared directly to the population parameters, and the classic REML and Helms-McCarroll estimates can be compared to the WLS estimates.

The results presented in Table 4-4 indicate that both the Helms-McCarroll and classic REML procedures generate acceptable fixed effect parameter estimates. On average, the estimates obtained from both procedures are nearly equal to the "true" parameter values. The variability in the parameter estimates for both procedures is equivalent to the variability in the WLS estimates. These comparable results suggest that the variability is an inherent attribute of the random nature of the observations rather than a function of the analytical or numerical properties of the REML estimates. Both procedures also produce comparable estimates of the error variance, which on average are equivalent to the "true" parameter value.

The random effects covariance parameter estimates generated by the two procedures appear to be somewhat different, with the average of the Helms-McCarroll estimates for the three parameters consistently larger than the classic REML estimates. The differences are shown in more detail in Figures 4-2 through 4-4, which are comparative plots of the estimates for the three parameters. The plots indicate that for all three parameters the classic REML estimate is greater than the Helms-McCarroll estimate more than half of the time; however, when the Helms-McCarroll estimate is the larger of the two, it is often much larger, particularly for $\gamma_1$. These qualitative observations are substantiated by the calculated differences in the estimates generated by the two procedures. For $\gamma_1$, the difference between the Helms-McCarroll and classic REML estimates ranges from -1.0 to 2.4 with an average of 0.21. For $\gamma_2$, the difference ranges from -0.045 to 0.049 with an average of 0.002, while for $\gamma_3$, the difference ranges from -0.004 to 0.005 with an average of 0.0003. The difference,
particularly in the estimates for $\gamma_1$ may be a function of the limitations of the relatively inefficient computation procedure that was used to develop the Helms-McCarroll estimates. In fact, the procedure was unable to achieve convergence in the parameter space for 4 of the 20 data sets. While these differences may be a computational artifact, they do appear to be worthy of further investigation.

Table 4.5 presents a summary of the results for the parameter estimates for the FVC-Mod2 data. (Detailed estimates are presented for each data set in Appendix A.) Recall that these 30 data sets comprise 3 data sets in each of 10 groups with the data sets within each group generated from 1 of the 10 different combinations of fixed effects. However, all data sets have the same covariance parameters. Because of the differences in the "true" fixed effect parameters, averages of the parameter estimates across data sets are not meaningful. However, comparison of the differences between the Helms-McCarroll or classic REML estimates and the WLS estimates averaged across data sets is meaningful. The information in Table 4.5 for fixed effect parameters is based on these differences, while the information on covariance parameter estimates is based on the actual estimates.

The results for the fixed effects and the error variance estimates for the FVC-Mod2 data are consistent with those for the Fairclough data. The fixed effect parameter estimates for both the Helms-McCarroll procedure and the classic REML procedure agree well with the WLS estimates, and on average, the error variance estimates are essentially identical to the "true" error variance.

The results for the random effect covariance parameter estimates for the FVC-Mod2 data contrast somewhat with those for the Fairclough-based data. On average, the estimates for both procedures are generally less than the "true" population values for the FVC-Mod2 data, while those for the
<table>
<thead>
<tr>
<th>Parameter</th>
<th>True Value</th>
<th>Helms-McCarroll Mean</th>
<th>Std. Dev.</th>
<th>Classic REML Mean</th>
<th>Std. Dev.</th>
<th>MRE</th>
<th>WLS estimate</th>
<th>MRE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta_1^*$</td>
<td>0</td>
<td>0.0002</td>
<td>0.0001</td>
<td>-0.001</td>
<td>0.0029</td>
<td>-7.3E-6</td>
<td>0.0001</td>
<td>-7.7E-6</td>
</tr>
<tr>
<td>$\beta_2^*$</td>
<td>0</td>
<td>-1.5E-5</td>
<td>0.0001</td>
<td>-7.8E-6</td>
<td>0.0003</td>
<td>-9.0E-6</td>
<td>0.0003</td>
<td>-4.2E-5</td>
</tr>
<tr>
<td>$\beta_3^*$</td>
<td>0</td>
<td>0.0002</td>
<td>0.0001</td>
<td>-0.001</td>
<td>0.0029</td>
<td>-7.3E-6</td>
<td>0.0001</td>
<td>-7.7E-6</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>0.01</td>
<td>0.0101</td>
<td>0.0007</td>
<td>0.0102</td>
<td>0.0018</td>
<td>0.0102</td>
<td>0.0018</td>
<td>0.0102</td>
</tr>
<tr>
<td>$\gamma_1$</td>
<td>20.0</td>
<td>18.92</td>
<td>6.514</td>
<td>19.47</td>
<td>3.947</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\gamma_2$</td>
<td>0.22</td>
<td>0.214</td>
<td>0.111</td>
<td>0.218</td>
<td>0.105</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\gamma_3$</td>
<td>0.01</td>
<td>0.0098</td>
<td>0.0037</td>
<td>0.0098</td>
<td>0.0027</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Parameter estimates obtained by subtracting the WLS estimate from the Helms_McCarroll or classic REML estimate.
Fairclough-based data were generally greater than the "true" values. Also, the Helms-McCarrow estimates, particularly for \( \gamma_1 \), are less on average than the classic REML estimate. The differences between the random effect covariance parameter estimates generated by the two procedures for these data are illustrated in Figures 4-5 through 4-7. Note that the classic REML estimates are more frequently the larger of the two estimates. However, the estimates from the two procedures appear to be positively correlated, suggesting that some of the departure of the estimates from the "true" parameters is a function of the data set rather than of the particular characteristics of the procedures.

Taken together, the results from the two groups of data indicate that the Helms-McCarrow and the classic REML procedures generated comparable fixed effect parameter estimates that agree quite well with the "true" parameters. Both procedures also generate comparable error variance estimates. The random effect covariance parameter estimates generated by the two procedures are more variable than are the fixed effect parameter estimates. However, these differences do not appear to be systematic and are likely to be associated with the inherent difficulties in estimating variance components and the limitations of the numerical procedures that were used to obtain the Helms-McCarrow parameter estimates for this study.

4.2.2 Comparison of Test Statistics and p-Values

A second question that can be addressed with these two groups of computer generated data is how the test statistics generated from the Helms-McCarrow and classic REML procedures compare to those generated by the WLS procedures, which are the "correct" tests if the covariance parameters are known. The hypothesis examined was \( H_0(\theta) = B(\theta = 0) \) versus \( H_a(\theta) = -H_0(\theta) = B(\theta = 0) \) where
\[ \theta = \begin{pmatrix} \beta_1 - \beta_3 \\ \beta_2 - \beta_4 \end{pmatrix} = C_1 \beta, \text{ where } C_1 = \begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{pmatrix} \]

Given the linear relationship between FVC and height, this hypothesis tests whether the intercepts and slopes of that relationship are equal for the two racial groups.

Figures 4-8 and 4-9 compare the Helms-McCarroll and classic REML test statistics, respectively, to the WLS statistic for the Fairclough-based data. Figures 4-10 and 4-11 present comparable results for the FVC-Mod2 data. In general, both the Helms-McCarroll and the classic REML test statistics appear to show reasonable agreement with the WLS statistic. However, on balance, the classic REML statistic appears to agree more closely with the WLS statistic than does the Helms-McCarroll statistic.

The data in Figure 4.8 indicate that while the two test statistics generally show good agreement, the Helms-McCarroll test statistic differs from the WLS statistic by more than a factor of 2 for two data sets. As noted on Figure 4.8, the value of the Helms-McCarroll statistic for data set No. 1 is about 40% of the value of the WLS statistic, while for data set No. 8 the value of the Helms-McCarroll statistic is about twice the value of the WLS statistic. These results can be explained to some extent by recalling the form of the test statistic (2.3.9) and noting that the value of the test statistic is in some sense inversely related to the magnitude of the covariance parameter estimates. For data set No. 1, the results presented in Appendix A indicate that the estimates for all three covariance parameters are consistently larger than the true values by a factor of 2.5 to 3.5. On the other hand, the estimates for data set No. 8 are consistently smaller than the true values by a factor of 2.5 to 2.7.
Interestingly, comparison of the same data sets for the classic REML procedures (Figure 4.9) shows the same consistent pattern although the magnitude of the differences in the parameters and test statistics is less. Subsequent review of the iteration history of the Helms-McCarroll estimates indicated that although convergence was relatively slow, the results for these two data sets converged to roughly the same estimates with different starting values. Although these results are anecdotal, they suggest that further examination of the properties of these data sets that might lead to consistent, relatively large departures of both sets of parameter estimates/statistics from the true parameters and WLS statistics is worth considering.

The plots in Figures 4-10 and 4-11 indicate that, with the exception of the Helms-McCarroll statistic for three data sets, both the Helms-McCarroll and the classic REML statistics show close agreement with the WLS statistics for the Modified FVC data. For each of the three data sets (data sets No. 11, 23, and 30), the estimates for all covariance parameters are consistently substantially smaller than the true parameters. However, unlike the results for the Fairclough-based data, the patterns for the classic REML estimates are not similar. Further examination of these data sets indicated that the Helms-McCarroll procedure converged to these values consistently from different starting values, including the case for which the classic REML estimates were used as starting values. Consequently, these results are considered to be an anomaly of the data sets not a characteristic of the procedure.

Plots of the p-values associated with these F-statistics are shown in Figures 4-12 through 4-15. More detailed information on smaller p-values for the FVC-Mod2 data is given in Figures 4-16 and 4-17 for the Helms-McCarroll and classic REML results, respectively. As expected from the results for the test statistics, the Helms-McCarroll and classic REML results generally show good
agreement with the WLS results. Large differences for two of the data sets are worth noting. For the Fairclough-based data set No. 6, the p-value for the Helms-McCarroll procedure was much greater than the WLS p-value (0.069 vs. 0.0028). For that data set, the Helms-McCarroll estimates of the covariance parameters were consistently larger than both the true parameters and the classic REML estimates, and the estimate of the intercept variance (the parameter of largest magnitude) was larger than the true parameter by a factor of 2.7. Again, the detailed examination of the convergence patterns this data set showed the results to be consistent for multiple sets of starting values of the covariance parameters. For the FVC-Mod2 data set No. 21, both the Helms-McCarroll and classic REML procedures generate p-values that were larger than the WLS p-values (0.06 to 0.07 vs 0.02). In both cases, the estimates for the three covariance parameters were consistently larger than the true parameter.

The above information substantiates the importance of the accuracy of the parameter estimates on the reliability of the test statistics and p-values. However, it provides little quantitative information about the relationship. To address this relationship, a more detailed perturbation analysis was undertaken. It is described below.

4.3 COVARIANCE PARAMETER PERTURBATION RESULTS

The theoretical results presented in Chapter 3 demonstrated that as the covariance parameter estimates converge to the true covariance parameters, the value of a linear transformation of $F_\nu$ converges to the value of a comparable linear transformation of $F_{WLS}$. However, neither those results nor the numerical results presented in Section 4.2 are sufficient to assess how "close" the covariance parameter estimates need to be to the true covariance parameters for $F_\nu$ to be "close" to $F_{WLS}$. This section
describes the results of perturbation analyses that were conducted to explore that question.

For the 20 Fairclough-based data sets and 30 modified FVC data sets described earlier, multiple parameter "estimates" were obtained by perturbing the vector of true covariance parameters. These parameter "estimates" were then used to generate $F_n$ estimates that could be compared to the $F_{WLS}$ estimates. The first subsection below describes the procedures that were used to generate these perturbations.

Analyses of the effect of the covariance parameter perturbations on the test statistic values require measures of both the magnitude of the covariance parameter perturbation and the effect on the test statistic. The three measures used for the test statistic effect were the difference between the perturbed $F$ (or the "estimated" $F_n$) and the WLS test statistic (i.e., $F_n - F_{WLS}$), the relative difference between the two statistics (i.e., $\frac{F_n - F_{WLS}}{F_{WLS}}$), and the absolute value of the relative difference (i.e., $\frac{|F_n - F_{WLS}|}{F_{WLS}}$). These same measures for the p-values associated with the test statistics were also examined. Because the covariance parameter "estimate" is a vector quantity, some measure of distance between the "estimate" and the true parameter is needed to quantify the magnitude of the perturbation. The first measure considered was the Euclidean distance. This measure has the advantage of being the metric that was used for the theoretical analyses in Chapter 3. However the disadvantages of the Euclidean distance are that it is dependent on the scale of the covariance parameters, it is dominated by the parameter with the largest magnitude, and we have no intuition about what magnitude of Euclidean errors in estimation that we might expect to see in practice. To address these disadvantages,
the Mahalanobis distance measure for the perturbations was considered. Because the exact distribution of the covariance parameter estimates is not known, the Mahalanobis distance could not be calculated exactly. However, because the Helms-McCarroll estimates are ML estimates, the Fisher information matrix for the model was deemed to be appropriate for developing an approximate Mahalanobis distance. The calculations for this approximate distance measure are presented in the second subsection below. After the calculations were completed, the relationship between the perturbation magnitude and the test statistic effects were examined graphically. These results are presented in the third subsection below.

Note that these results represent a preliminary examination of the numerical characteristics of the test statistic, not a comprehensive numerical analysis. The results are generated for only one model with 4 fixed effect and 3 random effect covariance parameters. Also, the results focus on the noncentral form of the statistics because earlier simulation results from McCarroll (1987) indicated that the central form of the statistic provides reasonable Type I error rates. Finally, the results focus on the Helms-McCarroll procedure because it formed the basis for the analyses presented in Chapter 3. While the evaluation has a limited scope, the results indicate that the proposed approximate Helms-McCarroll F-statistic for MixMod does have promising numerical properties as well as the sound theoretical basis described in Chapter 3.

4.3.1 Covariance Parameter Perturbation Procedures

For each of the 20 Fairclough-based data sets and the 30 modified FVC data sets, three sets of perturbed "covariance" parameter estimates were developed. For all three sets, the true covariance parameters were perturbed randomly by generating normal random vectors with the true parameters as the mean of the distribution. The covariance matrices for the for the first two sets of perturbations
were diagonal matrices, with the diagonal elements based on standard deviations that were fixed multiples of the true covariance parameters, as described below. The covariance matrices for the third set of perturbations were based on the information matrix.

The first perturbation algorithm generated 18 vectors of perturbed covariance parameter "estimates" for each of the 50 base data sets. These 18 "estimates" included 3 independent replicates for 6 different covariance matrices. Each covariance matrix was of the form

$$[k \cdot \text{Diag}(\gamma)]^2,$$

i.e., the perturbations for the covariance parameters were assumed to be independent, and they were assumed to be normally distributed with a standard deviation that was a constant multiple of the true parameter. For the first perturbation set, the constants used were 0.001, 0.005, 0.01, 0.05, 0.1, and 0.5. The perturbations for the second set were generated in an identical fashion, but different constants were used. These constants were 0.1, 0.25, 0.4, 0.6, 0.75, and 0.9. The results for these first two sets were combined for many of the subsequent analyses.

The third perturbation algorithm generated 18 i.i.d. vectors of covariance parameter "estimates" for each of the 50 data sets. The covariance matrices used for these "estimates" were obtained by inverting the approximate expected information matrix. This approximate information matrix using the procedures described below for the Mahalanobis distance calculations. This third set of perturbations was generated to provide some insight as to what types of variation might actually be found in practice.

4.3.2 Basis of the Mahalanobis Distance Calculations

The Mahalanobis distance between a normal random vector $\mathbf{y}$ and a fixed vector $\mathbf{y}_0$ is defined as (Arnold, 1981):

$$D_m(\mathbf{y}, \mathbf{y}_0) = \left[(\mathbf{y} - \mathbf{y}_0)' \mathbf{\Sigma}_y^{-1} (\mathbf{y} - \mathbf{y}_0)\right]^{0.5} \quad (4.3.1)$$

where $\mathbf{\Sigma}_y = \text{Var}(\mathbf{y})$. Because the exact distribution of the
Helms-McCarroll covariance parameter estimates is unknown, an exact Mahalanobis distance cannot be calculated for these perturbation analyses. However, the estimates are ML estimates, so the inverse of the expected Fisher information matrix, which is the asymptotic covariance matrix for the estimates, can be used to approximate the Mahalanobis distance. The paragraphs below describe the calculations that were used to obtain this approximation. As outlined in the discussion, exact expectations for all components of the Hessian could not be obtained. Consequently, these components were evaluated using a second order Taylor series approximation.

Recall from the calculations in Chapter 2, (2.2.17) and following, that the concentrated likelihood is obtained by letting:

$$
\hat{\sigma}^2(\gamma) = \frac{1}{b}(\mathbf{w}_1 - \mu_1)\Gamma_1^{-1}(\mathbf{w}_1 - \mu_1)
$$

(4.3.2)

where $b$ is the rank of the error space, $\Gamma_1 = \mathbf{T}_1'\mathbf{T}_1$, $\mathbf{w}_1 = \mathbf{T}_1'\mathbf{y}$, and $\mu_1 = \mathbf{T}_1'\mathbf{x}\beta$. Under these definitions, the concentrated likelihood equation is:

$$
l_{wc} = C - \frac{b}{2}[1 + \log \hat{\sigma}^2(\gamma)] - \frac{1}{2} \log |\Gamma_1|
$$

(4.3.3)

In order to calculate the Hessian for the covariance parameters, we need the second derivatives for the two components of this concentrated likelihood. First, note that:

$$
\frac{\partial^2 \log |\Gamma_1|}{\partial \gamma_i \partial \gamma_j} = - \text{tr}[\Gamma_1^{-1} \mathbf{T}_1' \mathbf{G}_j \mathbf{T}_1' \Gamma_1^{-1} \mathbf{T}_1' \mathbf{G}_i \mathbf{T}_1']
$$

(4.3.4)

$$
= - \text{tr}[\Gamma_1^{-1} \mathbf{G}_j \Gamma_1^{-1} \mathbf{G}_i]
$$

For notational convenience in the following derivation, define the following: $\mathbf{A}_1 = \Gamma_1^{-1}$, $\mathbf{A}_2 = \Gamma_1^{-1} \mathbf{G}_j \Gamma_1^{-1}$, $\mathbf{A}_3 = \Gamma_1^{-1} \mathbf{G}_i \Gamma_1^{-1}$, and
\[ A_k = \Gamma_i^{-1} \left[ G_j \Gamma_i^{-1} G_i - G_i \Gamma_i^{-1} G_j \right] \Gamma_i^{-1} \]. Also, let \( Z_k = (\mathbf{w}_1 - \mu_1)' A_k (\mathbf{w}_1 - \mu_1) \).

Using these definitions, the second derivative of \( \log \sigma^2(\gamma) \) with respect to the covariance parameters can be expressed as:

\[
\frac{\partial^2 \log \sigma^2(\gamma)}{\partial \gamma_i \partial \gamma_j} = \frac{1}{\partial^2 (\gamma)} \left[ -\frac{1}{b} \operatorname{tr} \left\{ (\mathbf{w}_1 - \mu_1)' \Gamma_i^{-1} G_i \Gamma_i^{-1} (\mathbf{w}_1 - \mu_1) \right\} \right]
+ \frac{1}{\partial^2 (\gamma)} \left[ -\frac{1}{b} \operatorname{tr} \left\{ (\mathbf{w}_1 - \mu_1)' \Gamma_i^{-1} G_j \Gamma_i^{-1} (\mathbf{w}_1 - \mu_1) \right\} \right]

= -\frac{1}{\partial^2 (\gamma)} \left[ -\frac{1}{b} \operatorname{tr} \left\{ (\mathbf{w}_1 - \mu_1)' \Gamma_i^{-1} G_j \Gamma_i^{-1} (\mathbf{w}_1 - \mu_1) \right\} \right]
+ \frac{1}{\partial^2 (\gamma)} \left[ -\frac{1}{b} \operatorname{tr} \left\{ (\mathbf{w}_1 - \mu_1)' \Gamma_i^{-1} G_j \Gamma_i^{-1} (\mathbf{w}_1 - \mu_1) \right\} \right]

= \left[ (\mathbf{w}_1 - \mu_1)' \Gamma_i^{-1} (\mathbf{w}_1 - \mu_1) \right]^2 \cdot \left[ (\mathbf{w}_1 - \mu_1)' \Gamma_i^{-1} G_j \Gamma_i^{-1} (\mathbf{w}_1 - \mu_1) \right]
+ \left[ (\mathbf{w}_1 - \mu_1)' \Gamma_i^{-1} (\mathbf{w}_1 - \mu_1) \right]^{-1}

= Z_1 Z_2 Z_3 + Z_1^2 Z_4

To obtain the information matrix, we need to take the expectation of these second derivatives with respect to \( \mathbf{w}_1 \). Note that the first component has a constant second derivative, so calculating the expectation is trivial. However, the second component yields a second derivative that is a nonlinear function of quadratic forms in \( \mathbf{w}_1 \). Note
also that (a) \( E(\mathbf{W}_1 - \mu_1) = 0 \), (b) \( \text{Var}(\mathbf{W}_1 - \mu_1) = \text{Var}(\mathbf{T}_1, \mathbf{Y}) = \sigma^2 \Gamma_1 \), and (c) in general, \( \sigma^2 \mathbf{A}_i \Gamma_j \mathbf{A}_j \neq 0 \). Consequently, the quadratic forms are not independent, so an analytical solution for the expectation cannot be readily obtained. In order to approximate the expectation, second order Taylor series expansions are used to obtain estimates for the two nonlinear components of the above equation as shown below.

First, let \( \mathbf{Z} = (\mathbf{Z}_1, \mathbf{Z}_2, \mathbf{Z}_3)' \), \( \mathbf{Z}^* = (\mathbf{Z}_1, \mathbf{Z}_4)' \), \( f_1(\mathbf{z}) = \mathbf{z}_1^{-2} \mathbf{z}_2 \mathbf{z}_3 \), and \( f_2(\mathbf{z}^*) = \mathbf{z}_1^{-1} \mathbf{z}_4 \). Then, using a multivariate version of Taylor’s expansion for any fixed vector \( \mathbf{a} \), we have:

\[
f_1(\mathbf{z}) \approx f_1(\mathbf{a}) + (\mathbf{z} - \mathbf{a})' f_1^{(1)}(\mathbf{a}) + \frac{1}{2} (\mathbf{z} - \mathbf{a})' f_1^{(2)}(\mathbf{a})(\mathbf{z} - \mathbf{a})
\]

where \( f_1^{(1)}(\mathbf{a}) = \left( \frac{\partial f_1(\mathbf{z})}{\partial \mathbf{z}_s} \right)_{y=\mathbf{a}} \) and \( f_1^{(2)}(\mathbf{a}) = \left( \frac{\partial^2 f_1(\mathbf{z})}{\partial \mathbf{z}_s \partial \mathbf{z}_t} \right)_{y=\mathbf{a}} \). Now, let \( \mathbf{a} = E(\mathbf{y}) \), and note that

\[
E[f_1(\mathbf{y})] \approx E[f_1(\mathbf{a})] + E[(\mathbf{y} - \mathbf{a})' f_1^{(1)}(\mathbf{a})] + \frac{1}{2} E[(\mathbf{y} - \mathbf{a})' f_1^{(2)}(\mathbf{a})(\mathbf{y} - \mathbf{a})]
\]

\[
= f_1(\mathbf{a}) + E[(\mathbf{y} - \mathbf{a})' f_1^{(1)}(\mathbf{a})] + \frac{1}{2} E[tr((\mathbf{y} - \mathbf{a}) f_1^{(2)}(\mathbf{a})(\mathbf{y} - \mathbf{a}))]
\]

\[
= f_1(\mathbf{a}) + \frac{1}{2} E[tr(f_1^{(2)}(\mathbf{a})(\mathbf{y} - \mathbf{a})(\mathbf{y} - \mathbf{a}'))]
\]

\[
= f_1(\mathbf{a}) + \frac{1}{2} tr\{f_1^{(2)}(\mathbf{a}) E[(\mathbf{y} - \mathbf{a})(\mathbf{y} - \mathbf{a})']\}
\]

\[
= f_1(\mathbf{a}) + \frac{1}{2} tr\{f_1^{(2)}(\mathbf{a}) \text{Var} (\mathbf{Z})\}
\]

\[
= f_1(\mathbf{a}) + \frac{1}{2} tr\{f_1^{(2)}(\mathbf{a}) \text{V}_1 (\mathbf{Z})\}
\]

Through a similar sequence of operations, we obtain
\[ E[f_2(z')] \approx f_2(a) + \frac{1}{2} \text{tr}\{f_2^{(2)}(a) \text{Var}(z')\} \]
\[ = f_2(a) + \frac{1}{2} \text{tr}\{f_2^{(2)}(a) \text{V}_2(z')\} \]

The second order partial derivatives needed to generate \( f_1^{(2)}(a) \) and \( f_2^{(2)}(a) \) for the above expression can readily be calculated as follows:

\[
\frac{\partial^2 f_1(z)}{\partial z_1^2} = \frac{6z_2 z_3}{z_1^4}, \quad \frac{\partial^2 f_1(z)}{\partial z_2^2} = 0, \quad \frac{\partial^2 f_1(z)}{\partial z_3^2} = 0
\]

\[
\frac{\partial^2 f_1(z)}{\partial z_1 \partial z_2} = -\frac{2z_3}{z_1^2}, \quad \frac{\partial^2 f_1(z)}{\partial z_1 \partial z_3} = -\frac{2z_2}{z_1^2}, \quad \frac{\partial^2 f_1(z)}{\partial z_2 \partial z_3} = \frac{1}{z_1^2}
\]

\[
\frac{\partial^2 f_2(z^*)}{\partial z_1^2} = \frac{2z_4}{z_1^3}, \quad \frac{\partial^2 f_2(z^*)}{\partial z_2^2} = 0, \quad \frac{\partial^2 f_2(z^*)}{\partial z_3^2} = \frac{-1}{z_1^2}
\]

Recall that for \( i = 1, \ldots, 4 \), \( Z_i \) is a quadratic form in the normal random vector \( \mathbf{w}_1 - \mu_1 \), which has 0 expectation and a variance of \( \sigma^2 \Gamma_1 \). As noted in early papers by Matern (1949) and Aitken (1950) on the independence of quadratic forms, quadratic forms of this type have the following mean and covariance structure:

\[
E(Z_i) = \sigma^2 \text{tr}(\Gamma_1 A_i)
\]

\[
\text{Var}(Z_i) = (\sigma^2)^2 \text{tr}[(\Gamma_1 A_i)^2]
\]

\[
\text{Cov}(Z_i, Z_j) = (\sigma^2)^2 \text{tr}[A_i \Gamma_1 A_j \Gamma_1]
\]

These results can be used to obtain an approximate expression for each component of the information matrix as follows:
\[-E \left[ \frac{\partial^2 \omega_{ij}}{\partial y_i \partial y_j} \right] = \text{tr}[G_i \Gamma_i^{-1} G_j \Gamma_j^{-1}] + \frac{\text{tr}(G_j \Gamma_j^{-1}) \cdot \text{tr}(G_j \Gamma_j^{-1})}{[\text{rank}(T_1)]^2} \]

\[+ \frac{1}{2} \text{tr}[f_1^{(2)}(a) \nu_1(z)] - \frac{2 \text{tr}[G_i \Gamma_i^{-1} G_j \Gamma_j^{-1}]}{\text{rank}(T_1)} \]

\[= \frac{1}{2} \text{tr}[f_2^{(2)}(a) \nu_2(z^*)] \quad (4.3.5) \]

where the matrices that constitute the different components of the expression are defined as follows:

\[f_1^{(2)}(a) = \begin{bmatrix}
-2 \text{tr}(G_j \Gamma_j^{-1}) \cdot \text{tr}(G_i \Gamma_i^{-1}) & -2 \text{tr}(G_j \Gamma_j^{-1}) & -2 \text{tr}(G_j \Gamma_j^{-1}) \\
\frac{[\sigma^2]^2 \cdot [\text{Rank}(T_1)]^6}{[\sigma^2]^2 \cdot [\text{Rank}(T_1)]^6} & \frac{[\sigma^2]^2 \cdot [\text{Rank}(T_1)]^6}{[\sigma^2]^2 \cdot [\text{Rank}(T_1)]^6} & \frac{[\sigma^2]^2 \cdot [\text{Rank}(T_1)]^6}{[\sigma^2]^2 \cdot [\text{Rank}(T_1)]^6} \\
-2 \text{tr}(G_j \Gamma_j^{-1}) & 0 & 1 \\
\frac{[\sigma^2]^2 \cdot [\text{Rank}(T_1)]^6}{[\sigma^2]^2 \cdot [\text{Rank}(T_1)]^6} & \frac{[\sigma^2]^2 \cdot [\text{Rank}(T_1)]^6}{[\sigma^2]^2 \cdot [\text{Rank}(T_1)]^6} & 0
\end{bmatrix} \]

with

\[\nu_1(z) = \frac{[\sigma^2]^2 \cdot [\text{Rank}(T_1)]^6}{[\sigma^2]^2 \cdot [\text{Rank}(T_1)]^6} \begin{bmatrix}
\text{Rank}(T_1) & \text{tr}(\Gamma_1^{-1} G_i) & \text{tr}(\Gamma_1^{-1} G_i) \\
\text{tr}(\Gamma_1^{-1} G_i) & \text{tr}(G_i \Gamma_i^{-1}) & \text{tr}(\Gamma_1^{-1} G_i) \\
\text{tr}(\Gamma_1^{-1} G_i) & \text{tr}(G_i \Gamma_i^{-1}) & \text{tr}(\Gamma_1^{-1} G_i) \\
\text{tr}(\Gamma_1^{-1} G_i) & \text{tr}(G_i \Gamma_i^{-1}) & \text{tr}(\Gamma_1^{-1} G_i)
\end{bmatrix} \]

Similarly,

\[f_2^{(2)}(a) = \begin{bmatrix}
4 \text{tr}(G_j \Gamma_j^{-1} G_i \Gamma_i^{-1}) & -1 \\
\frac{[\sigma^2]^2 \cdot [\text{Rank}(T_1)]^6}{[\sigma^2]^2 \cdot [\text{Rank}(T_1)]^6} & \frac{[\sigma^2]^2 \cdot [\text{Rank}(T_1)]^6}{[\sigma^2]^2 \cdot [\text{Rank}(T_1)]^6} \\
-1 & 0
\end{bmatrix} \]

and

136
\[ \mathbf{v}_2(Z') = \sigma^2 \begin{bmatrix} \text{Rank}(T_i) & 2 \text{tr}(G_i \Gamma_i^{-1} G_j \Gamma_j^{-1}) \\ 2 \text{tr}(G_i \Gamma_i^{-1} T_i G_j) & \text{tr}\left[(G_i \Gamma_i^{-1} G_j \Gamma_j^{-1})^2 + G_i \Gamma_i^{-1} G_j \Gamma_j^{-1}\right] \end{bmatrix} \]

Because the asymptotic covariance matrix is the inverse of the expected information matrix, the matrix calculated using (4.3.5) can be used directly as the weight matrix in the Mahalanobis distance calculations. The equations derived above were implemented via a SAS IML algorithm to generate the Mahalanobis distance estimates used in the next section. As a part of those calculations, the relative contribution of each the 5 components of (4.3.5) was determined. The last 4 components, which are the approximate values, were found to be dominated by the first component, which was determined exactly. Consequently, the part of the approximation associated with the Taylor series expansion did not introduce significant error into the distance calculations.

4.3.3 Covariance Parameter Perturbation Evaluation

The calculations described above generated a substantial amount of information about the effects of "errors" in covariance parameter estimates on the numerical behavior of the Helms-McCarroll approximate F-statistic and p-values associated with that statistic. While these analyses were limited to a single model structure and to two groups of data sets, they provide some intuition about the behavior of the statistic. Because the primary objective of the analyses was to gain some insight into the numerical properties of the statistic not to conduct a comprehensive numerical evaluation of the statistic, the focus of the analysis was on graphical display of the results. Observations about the performance of the test statistic based on these graphical displays are provided in the paragraphs below.

Note that for notational convenience, the Mahalanobis
distance between two vectors, say $\gamma$ and $\gamma'$ will be denoted as $D_W(\gamma, \gamma')$, the difference between $F_H$ and $F_{WLS}$ will be denoted as the absolute error (AE) in $F$, the relative difference will be denoted as the relative error (RE) in $F$, the absolute value of the relative difference will be denoted as the absolute relative error (AR error or RE-A) in $F$. (See Section 4.3 for definitions of these difference terms.) Also, note that all of the analyses are based on censored data in that perturbations that resulted in a random effects covariance matrix that was not positive definite were eliminated from the analyses. Such censoring is deemed reasonable in that the estimation procedures typically used to generate either Helms-McCarroll or classic REML estimates limit parameter estimates to a restricted parameter space that requires the random effects covariance to be positive definite.

The first step in the evaluation was to examine the distribution of the errors in the F-statistic as a function of the "size" of the covariance parameter estimate "error" for the fixed perturbations. Figures 4-18 and 4-19 show the empirical distribution function (EDF) of the AR error in $F$ for the Fairclough-based data for perturbation sets 1 and 2, respectively. A separate EDF is shown for the data associated with each of the multiples that generated the fixed perturbation covariance matrices. Comparable plots are shown for the Modified FVC data in Figures 4-20 and 4-21.

The patterns in the EDF's are remarkably consistent for the two groups of data and for the different perturbation levels. They generally show that the 90th percentile of the EDF of the AR error in $F$ is essentially of the same magnitude as the relative standard deviation of the perturbation of each of the elements of $\gamma$. These results suggest that on average, the relative perturbations in $F$ are on the same order of magnitude as the relative error in the covariance parameter estimates. As such they provide
numerical corroboration of the analytical convergence results. Also, they suggest that further evaluation of the properties of the Helms-McCarroll covariance parameter estimates for small and moderate sample sizes is warranted.

Because the EDF plots described above showed a reasonably strong and consistent relationship between the "size" of the covariance parameter perturbation and the error in F, the relationship was examined further using both the Euclidean distance and Mahalanobis distance measures described earlier. The paragraphs below describe the results of these analyses, first for the Fairclough-based data and then for the Modified FVC data. For these analyses, the fixed perturbation analyses (perturbation sets 1 and 2) were combined, while the perturbations based on the asymptotic covariance matrices for the covariance parameter estimates (perturbation set 3) were addressed separately.

Figure 4.21 shows a plot of the RE in F versus $F_{WLS}$ for the Fairclough-based data and the fixed perturbations. This graph provides an overall indication of the magnitudes of the test statistics involved in the analyses and on the magnitudes of the relative errors. In examining these data, first note that most of the relative errors are near 0. This phenomenon of the data is probably attributable to the number of relatively small perturbation levels in perturbation set 1 and to the normal random vectors used to develop the perturbations. More importantly, however, note that distribution of the relative errors appears to be relatively consistent over the different values of $F_{WLS}$. These results suggest that the actual error increases with increasing size of the WLS statistic, a result that is encouraging in practice in that large errors in F are likely to be associated with large values of the F-statistic. Such a scenario reduce the possibility of the error in F generating a mistaken inference.

All data generated in the analyses with the exception of those eliminated because they did not meet the positive
definiteness criteria described earlier are included in Figure 4.22. However, additional data were eliminated before subsequent graphs were developed in order to provide more informative graphical output. Those observations that had a p-value less than 0.00001 and those observations that had a \( D_w \) less than 0.001 were eliminated from the data base. The observations with very small p-values were eliminated because a review of the data indicated that the possible errors in F at p-values of this level would never have a substantial effect on inference or confidence intervals. The observations with very small \( D_w \) were eliminated because in practice, the chance of obtaining estimates that close to the true parameter value is very small. Consequently, both scenarios were presumed to be uninformative for this graphical presentation.

Figure 4.23 shows the RE in F as a function of the Euclidean distance between the vector of covariance parameter "estimates" and the true covariance parameter vector. As noted earlier, the Euclidean distance measure is scale dependent. For these data sets, the Euclidean norm of the covariance parameter vector is about 2.4, so the distances shown in Figure 4.23 are of the order of the norm of the true parameter and smaller. While the Euclidean distance does have limitations, Figure 4.23 does show the strong relationship between the "error" in the covariance parameter "estimate" and the RE in F. In particular, it provides further numerical corroboration of the analytical convergence results presented in Chapter 3.

To avoid the inherent problems associated with the Euclidean distance measure, the remaining analyses are conducted with the Mahalanobis distance measure. Recall from the definition of the Mahalanobis distance (4.3.1) that the square of \( D_w \) is a quadratic form in the covariance parameter "estimates" centered at the true parameter values. The weight matrix is the inverse of the approximate asymptotic covariance matrix of these estimates. Assuming
that the estimates are unbiased (or at least consistent), this squared distance behaves heuristically like a central chi-squared random variable. This heuristic treatment of the square of \( D_m \) is used in the subsequent discussion to provide an intuitive feel for the approximate magnitude of "error" in covariance parameter estimates that might be expected in practice.

Figures 4-24 and 4-25 show the RE of F and the AE of F, respectively, as a function of the Mahalanobis distance between the covariance parameter "estimates" and the true covariance parameters. Note that on both graphs vertical bars are drawn at \( D_m = 0.5 \) and \( D_m = 3 \). These points are approximately the square roots of the 2.5th and 97.5th percentiles, respectively, of a random variable with a \( \chi^2(3) \) distribution. From Figure 4.24, note that within this the relative error in F ranges from about -0.3 to 0.5, and that most of the observations lie between -0.1 and 0.1.

Furthermore, Figure 4.25 shows that the actual error in F is generally between -1 and 1. In both figures, more observations visually appear to be greater than 0, suggesting that the errors in F may be positively biased in this range. This visual observation was substantiated by univariate analyses of the 169 observations that spanned this range, which indicated that the RE of F had a mean of 0.040, with a standard deviation of 0.14, and a median of 0.023, while the AE of F had a mean of 0.098, a standard deviation of 0.10, and a median of 0.064. This positive bias is not particularly disturbing in that McCarroll (1987) showed a form of the statistic to have reasonable Type I error properties, and positive biases in the noncentral cases considered here will tend to increase the power of the statistic.

Figures 4-26 through 4-30 contain information on the errors in the p-values associated with the F-statistics generated for the Fairclough-based data and perturbation sets 1 and 2. From Figure 4.26, note that these results are
relatively uninformative because all of the WLS p-values are
less than 0.006 and the relative errors are dominated by 4
extreme observations associated with WLS p-values of less
than 0.0001. However, even with these limitations, the
results in Figure 4.28 do indicate that over the Mahalanobis
distance region of greatest interest the actual errors are
small (generally in the range of -0.004 to 0.004). Because
those observations with p-values less than 0.0001 are of
limited interest, they were deleted and revised graphs of
the RE and AE of the p-values were prepared, as shown in
Figures 4-29 and 4-30, respectively. Taken together, these
results indicate that for populations that have relatively
large fixed effect differences compared to the within
subject error variance, like the Fairclough-based model, the
errors in the F-statistic associated with covariance
parameter estimation error are not likely to lead to
inferential mistakes.

Figure 4.31, the final graph associated with the
"fixed" perturbations for the Fairclough-based data, shows
the relationship of the Mahalanobis distance between the
covariance parameter "estimates" and the true covariance
parameters to the Euclidean distance between the eigenvalues
of $V_{\hat{\Theta}} \cdot \tilde{M}^{-1}$ and a vector of 1's. Recall from Chapter 3 that
the departure of these eigenvalues from the vector of 1's
determines how weights for the true underlying mixture
distribution for $F_\mu$ depart from the Poisson weights that are
obtained for the hypothesized F distribution if the
eigenvalues are all 1. Note that the Euclidean distance is
generally relatively small in comparison to 1.4, which is
the Euclidean norm for 1, but to date no information has
been developed on how such distances will affect the
mixture. Additional evaluation of these eigenvalue
differences and on their effect on the mixture distribution
of $F_\mu$ in both the central and noncentral cases appear to be
warranted as a topic for future research.
Figures 4-32 through 4-35 present results for the Fairclough-based data obtained with the Mahalanobis distance-based perturbations (Perturbation Set 3). These results are generally consistent with the results for the "fixed" perturbations described above. Furthermore, these results indicate that in general, the Mahalanobis distances fall between 0.5 and 3, which is consistent with the behavior postulated above.

As shown if Figure 4.32, the relative error in the F-statistic over the range of Mahalanobis distances of interest (0.5 to 3) is between -0.3 and 0.6, a range very comparable to the -0.3 to 0.7 shown for the "fixed" perturbations in Figure 4.24. Similarly, the range in p-values of -0.004 to 0.007 shown in Figure 4.36 is comparable to the range shown in Figure 4.28 for the "fixed" perturbations. Again for this model, these results suggest that although the errors in the covariance parameter estimates do lead to errors in the F-statistics and associated p-values, these errors do not have a substantial effect on the inferences associated with the analyses.

Figure 4.35 shows the relationship of the Mahalanobis distance between the covariance parameter "estimates" and the true covariance parameters to the Euclidean distance between the eigenvalues of $\mathbf{V}_0^{-1/2}$ and the vector of 1's that would have been obtained if the estimates had equalled the true parameters. Again, note the strong relationship between these two factors. Also, note that over the range of Mahalanobis distances of greatest interest, the Euclidean distances are of about the same magnitude of those found for the first two perturbation sets. In almost all cases the Euclidean distance is substantially less than the norm of $1_2$ and is never greater than 1. Again, no quantitative information is available on the effect of errors of this type on the coefficients for the $F_\mathbf{X}$ mixture distribution, and further analysis of these mixtures for different models.
appears to be a viable topic for future research.

Analyses comparable to those described above were also conducted on the perturbation results associated with the modified FVC data. Recall that these data have a structure that is similar to the Fairclough-based data. However, the underlying error variance and covariance parameters are much larger compared to the differences in the fixed effects for the two populations than are those for the Fairclough-based data. Consequently, the range in WLS F-statistics and associated p-values is much larger, making the results from the perturbation analyses somewhat more interesting.

Results that are equivalent to those presented above for the Fairclough-based data are presented graphically in Figures 4-36 through 4-49 and are discussed in the paragraphs below. The major findings from the perturbation analyses of the modified FVC data are comparable to those from the Fairclough-based perturbation analyses. The discussion below focused on differences found for the two groups of data and on additional insights gained from the modified FVC data. Findings that were essentially the same for the two groups are not addressed in detail, but a complete set of graphical representations equivalent to those that were presented for the Fairclough-based data are included for completeness.

The increased variability in the test statistics and associated p-values is reflected in Figures 4-36 and 4-40 which compare the relative error in the F-statistic to $F_{WLS}$ and the relative error in the p-value to the WLS p-value respectively. In particular, Figure 4.40 shows that the WLS p-values for the modified FVC data range from approximately $10^{-5}$ to almost 1, in contrast to the WLS p-values for the Fairclough-based data, which were all less than $10^{-2}$. Further examination of these two figures discloses two patterns worth noting that were not apparent for the Fairclough-based data. First, Figure 4-36 indicates that while most of the errors in the F-statistic fall near 0,
some observations with $F_{WLS}$ values in the range of 1 to 3 have relative errors of 0.5 or greater. Also, for $F_{WLS}$ values in the range of 3 to 5, several observations have relative errors of -0.5 to -0.25. For $F$-statistics in this range errors of such can change inferential decisions. The relative errors in the $p$-values for those observations having WLS $p$-values in the range of 0.02 to 0.1 reflect this pattern. Because these data are all based on a noncentral situation, i.e., the null hypothesis is false, these results suggest that the power of $F_n$ in comparison to $F_{WLS}$ warrants further investigation.

With the exception of Figure 4.45, which shows the relationship of the Mahalanobis distance between the covariance parameter "estimates" and the true covariance parameters to the Euclidean distance between the eigenvalues of $V_0^{-1} \bar{\Sigma}^{-1}$ and a vector of 1's for the modified FVC data and perturbation sets 1 and 2, Figures 4-37 through 4-48 show the relationship between the distance between the covariance parameter "estimates" and the true parameters and different measures of the error in $F$-statistics and $p$-values for the modified FVC data. These comparisons show the same general pattern as those shown for the Fairclough-based data. In particular, the errors in both the $F$-statistics and the $p$-values increase as the distance between the "estimates" and the true parameter increases. However, over the range of Mahalanobis distances of greatest interest (0.5 to 3), both the actual and relative errors appear to be larger for the modified FVC data than they were for the Fairclough-based data. While this increase in magnitude of the errors may be attributable to the increased variability of the test statistics and $p$-values, it warrants further investigation in future research.

Further graphical comparisons of the Helms-McCarroll and WLS test statistics and $p$-values as a function of the Mahalanobis distance between the covariance parameter
"estimates" and true parameters are provided in Figures 4-50 through 4-55. Figures 4-50 and 4-51 compare $F_n$ versus $F_{wls}$ and the Helms-McCarroll versus the WLS p-values for perturbation sets 1 and 2, while Figures 4-52 and 4-53 present similar comparisons for perturbation set 3. With the exception of observations for which the Mahalanobis distance between the "estimates" and the true parameters is greater than 3, the Helms-McCarroll and WLS statistics and p-values show reasonably close agreement as evidenced by their relatively narrow scatter about the 45° line. Although variation about the line appears to be relatively symmetric, there appears to be a slight tendency for the Helms-McCarroll procedure to produce larger F-statistics with associated smaller p-values. If this pattern is real and not an anomaly of this particular data set, these results suggest that $F_n$ might provide slightly greater power than $F_{wls}$. Further investigation of the power of the statistic using the results from the mixture expansion developed in Chapter appears to be an avenue of future research.

A more detailed view over a narrower range of p-values is shown in Figures 4-54 and 4-55 for perturbation sets 1 and 2 and perturbation set 3, respectively. Again, with the exception of those observations for which the Mahalanobis distance is greater that 3, the inferences associated with the Helms-McCarroll and WLS procedures show good agreement at both the $\alpha=0.01$ and $\alpha=0.05$ levels. In observations for which the inferences differed, the Helms-McCarroll procedure rejected the null hypothesis slightly more often than did the WLS procedure, suggesting that it has slightly greater power for this particular model and data set. However, the number of differences in inferential decisions between the two procedures is quite small in comparison to the number of observations for which the decisions were the same, so these results should be considered only suggestive.

Although the numerical evaluations described in this
section were limited in scope, they indicate that the test statistic $F_h$ can reasonably be used to test contrasts among the population effects in longitudinal studies that can be modeled with MixMod. The numerical results corroborate the analytical convergence that was demonstrated in Chapter 3. Furthermore, for the data sets and models considered in the perturbation analyses, the results suggest that $F_h$ has comparable power to the $F_{W1}$ statistic that would be used if the covariance parameters were known, and earlier results obtained by McCarroll (1987) indicate that it has acceptable Type I error properties. Consequently, taken as a whole, the numerical results obtained here and in McCarroll's earlier study support the use of $F_h$ for longitudinal studies.
FIGURE 4.1. Illustration of the FVC Linear Model

Child's Race

- ▲ Black
- ⋯⋯⋯⋯○ White

Forced Vital Capacity (L)

Height (cm)

148
FIGURE 4.2. Comparison of Intercept Variance Estimates
Fairclough-Based Data
FIGURE 4.3. Comparison of Intercept/Slope Covariance Estimates
Fairclough - Based Data
Figure 4.4. Comparison of Slope Variance Estimates
Fairclough-Based Data
FIGURE 4.5. Comparison of Intercept Variance Estimates
Modified FVC Data

[Scatter plot showing the comparison between Classic REML Estimate and Helms-McCarroll Estimate]
FIGURE 4.6. Comparison of Intercept/Slope Covariance Estimates
Modified FVC Data
FIGURE 4.7. Comparison of Slope Variance Estimates
Modified FVC Data
FIGURE 4.8. F–Statistic Comparison HM vs WLS
Fairclough–Based Data
FIGURE 4.9. F–Statistic Comparison Classic REML vs WLS Fairclough–Based Data
FIGURE 4.10. F - Statistic Comparison HM vs WLS
Modified FVC Data
FIGURE 4.11. F–Statistic Comparison Classic REML vs WLS Modified FVC Data
FIGURE 4.12. P-Value Comparison HM vs WLS
Fairclough-Based Data
FIGURE 4.13. P-Value Comparison Classic REML vs WLS
Fairclough-Based Data
FIGURE 4.14. P-Value Comparison HM vs WLS
Modified FVC Data (Full Scale)
FIGURE 4.15. P-Value Comparison Classic REML vs WLS
Modified FVC Data (Full Scale)
FIGURE 4.16. P-Value Comparison HM vs WLS
Modified FVC Data (Reduced Scale)
FIGURE 4.17. P-Value Comparison Classic REML vs WLS
Modified FVC Data (Reduced Scale)
FIGURE 4.18. EDF Plot for AR Error in F
Fairclough-based Data, Perturbation Set 1

Empirical Distribution Function of Absolute Error

Absolute Relative Error in F

Perturbation s.d.  ○ ○ ○ 0.1%  + + + 0.5%  ▲ ▲ ▲ 1.0%
                   ▲ ▲ ▲ 5.0%  ● ● ● 10%  + + + 50%
FIGURE 4.19. EDF Plot for AR Error in F
Fairclough-based Data, Perturbation Set 2

Empirical Distribution Function of Absolute Error

Absolute Relative Error in F

Perturbation s.d.  o  o  10%  * * *  25%  o o o  40%
                   △ △ △  60%  ● ● ●  75%  + + +  90%
FIGURE 4.20. EDF Plot for AR Error in F
Modified FVC Data, Perturbation Set 1

Empirical Distribution Function of Absolute Error

Absolute Relative Error in F

Perturbation

\[ \begin{array}{cccc}
0.1% & 0.5% & 1.0% & 5.0% \\
10% & 50% \\
\end{array} \]
FIGURE 4.21. EDF Plot for AR Error in F
Modified FVC Data, Perturbation Set 2

Empirical Distribution Function of Absolute Error

Absolute Relative Error in F

Perturbation s.d.  ○ ○ ○ 10%  × × × 25%  ○ ○ ○ 40%
               △ △ △ 60%  ● ● ● 75%  --- 90%
FIGURE 4.22. Relative Error in F vs WLS F
Fairclough–based Data, Perturbation Sets 1 and 2
FIGURE 4.23. Relative Error in F vs Euclidean Distance
Fairclough-based Data, Perturbation Sets 1 and 2
FIGURE 4.24. Relative Error in F vs Mahalanobis Distance
Fairclough-based Data, Perturbation Sets 1 and 2
FIGURE 4.25. Actual Error in F vs Mahalanobis Distance
Fairclough-based Data, Perturbation Sets 1 and 2

Error in HM F-Statistic

Mahalanobis Distance
FIGURE 4.26. Relative Error in p-Value vs WLS p-Value
Fairclough-based Data, Perturbation Sets 1 and 2

Relative Error in HM p-Value

Weighted Least Squares p-Value

1.00E-07  1.00E-06  1.00E-05  1.00E-04  1.00E-03  1.00E-02
FIGURE 4.27. Relative Error in p-Value vs Mahalanobis Distance
Fairclough-based Data, Perturbation Sets 1 and 2
FIGURE 4.28. Actual Error in p-Value vs Mahalanobis Distance
Fairclough-based Data, Perturbation Sets 1 and 2
FIGURE 4.29. P-Value Relative Error vs Mahalanobis Distance
Fairclough-based Data (Censored), Perturbation Sets 1 and 2
FIGURE 4.30. P-Value Actual Error vs Mahalanobis Distance
Fairclough-based Data (Censored), Perturbation Sets 1 and 2
FIGURE 4.31. Eigenvalue Error vs Mahalanobis Distance
Fairclough-based Data, Perturbation Sets 1 and 2
FIGURE 4.32. Relative Error in F vs Mahalanobis Distance
Fairclough–based Data, Perturbation Set 3

Mahalanobis Distance
FIGURE 4.33. P-Value Relative Error vs Mahalanobis Distance
Fairclough-based Data, Perturbation Set 3
FIGURE 4.34. P-Value Actual Error vs Mahalanobis Distance
Fairclough-based Data, Perturbation Set 3

Error in HM p-Value

Mahalanobis Distance
FIGURE 4.35. Eigenvalue Error vs Mahalanobis Distance
Fairclough-based Data, Perturbation Set 3
FIGURE 4.36. Relative Error in $F$ vs WLS $F$
Modified FVC Data, Perturbation Sets 1 and 2

Weighted Least Squares $F$ Statistic
FIGURE 4.37. Relative Error in F vs Euclidean Distance
Modified FVC Data, Perturbation Sets 1 and 2
FIGURE 4.38. Relative Error in F vs Mahalanobis Distance
Modified FVC Data, Perturbation Sets 1 and 2
FIGURE 4.39. Actual Error in F vs Mahalanobis Distance
Modified FVC Data, Perturbation Sets 1 and 2
FIGURE 4.40. Relative Error in p-Value vs WLS p-Value
Modified FVC Data, Perturbation Sets 1 and 2

Weighted Least Squares p-Value

Relative Error in HM p-Value

0.000001 0.000010 0.000100 0.001000 0.010000 0.100000 1.000000
FIGURE 4.41. Relative Error in p-Value vs Mahalanobis Distance
Modified FVC Data, Perturbation Sets 1 and 2
FIGURE 4.42. Actual Error in p-Value vs Mahalanobis Distance
Modified FVC Data, Perturbation Sets 1 and 2
FIGURE 4.44. P-Value Actual Error vs Mahalanobis Distance
Modified FVC Data (Censored), Perturbation Sets 1 and 2
FIGURE 4.45. Eigenvalue Error vs Mahalanobis Distance
Modified FVC Data, Perturbation Sets 1 and 2

Eigenvalue Euclidean Norm

Mahalanobis Distance

192
FIGURE 4.46. Relative Error in F vs Mahalanobis Distance
Modified FVC Data, Perturbation Set 3
FIGURE 4.47. P-Value Relative Error vs Mahalanobis Distance
Modified FVC Data, Perturbation Set 3
FIGURE 4.49. Eigenvalue Error vs Mahalanobis Distance
Modified FVC Data, Perturbation Set 3
FIGURE 4.50. Comparison of HM and WLS F−Statistics
Modified FVC Data, Perturbation Sets 1 and 2

<table>
<thead>
<tr>
<th>Mahalanobis Dist</th>
<th>MD</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MD &lt; 0.5</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>0.5 &lt; MD &lt; 1.0</td>
<td>✷</td>
</tr>
<tr>
<td></td>
<td>1.0 &lt; MD &lt; 2.0</td>
<td>▲ ▲</td>
</tr>
<tr>
<td></td>
<td>MD &gt; 3.0</td>
<td>• • •</td>
</tr>
</tbody>
</table>
FIGURE 4.51. Comparison of HM and WLS p-Values
Modified FVC Data, Perturbation Sets 1 and 2

Weighted Least Squares p-Value

Mahalanobis Dist

○ ○ ○ MD < 0.5  • • • 0.5 < MD < 1.0
○ ○ ○ 1.0 < MD < 2.0  △ △ △ 2.0 < MD < 3.0
● ● ● MD > 3.0
FIGURE 4.52. Comparison of HM and WLS F-Statistics
Modified FVC Data, Perturbation Set 3

<table>
<thead>
<tr>
<th>Mahalanobis Dist</th>
<th>Legend</th>
</tr>
</thead>
<tbody>
<tr>
<td>○ ○ ○</td>
<td>MD &lt; 0.5</td>
</tr>
<tr>
<td>○ ○ ○ ○</td>
<td>0.5 &lt; MD &lt; 1.0</td>
</tr>
<tr>
<td>1.0 &lt; MD &lt; 2.0</td>
<td>△ △</td>
</tr>
<tr>
<td>MD &gt; 2.0</td>
<td>△ △ △</td>
</tr>
<tr>
<td>MD &gt; 3.0</td>
<td>• • •</td>
</tr>
</tbody>
</table>

Weighted Least Squares F-Statistic

199
FIGURE 4.54. Comparison of Smaller HM and WLS p-Values
Modified FVC Data, Perturbation Sets 1 and 2

---

Weighted Least Squares p-Value

Mahalanobis Dist

- o o o MD < 0.5
- x x x 0.5 < MD < 1.0
- o o o 1.0 < MD < 2.0
- Δ Δ Δ 2.0 < MD < 3.0
- . . . MD > 3.0

---

201
FIGURE 4.55. Comparison of Smaller HM and WLS p-values
Modified FVC Data, Perturbation Set 3

- ○ ○ ○ MD < 0.5  *=* 0.5 < MD < 1.0
- △ △ △ 1.0 < MD < 2.0 △ △ 2.0 < MD < 3.0
- • • • MD > 3.0
CHAPTER 5
SUMMARY AND SUGGESTIONS FOR FUTURE RESEARCH

The underlying motivation for this study was the need for a test statistic for the linear mixed model (MixMod) that could be used to test population effects or contrasts among those effects using longitudinal data generated from experimental or observational studies. Such a statistic is needed particularly for those studies that generate small to moderate sample sizes. It is also needed to calculate power for alternative longitudinal designs. The initial focus of the study was on the statistic $F_N$ that was proposed by McC Carroll and Helms (1987) and a statistic $F_{REML}$ that has the same functional form as $F_N$ but that is based on covariance parameter estimates derived by the classic REML procedure rather than the Helms-McCarroll procedure. Although simulation results showed that the statistic provided reasonable Type I error rates, information was lacking on the mathematical properties of the statistic. Consequently, the primary objective of this study was to characterize the distribution of $F_N$ or of a comparable test statistic both mathematically and numerically.

The first step in the evaluation of the statistic was a careful examination of $F_N$ as proposed by McC Carroll and Helms (1987). As a result of this evaluation, a modified form of $F_N$ based on a reparameterization of MixMod is proposed as shown below.

$$F_N = \frac{\theta_N'\left[ C_1 (X' T_1 \hat{F}_N^{-1} T_1' X)^{-1} C_1' \right]^{-1} \theta_N / \sigma^2}{\sigma^2_N}$$  \hspace{1cm} (5.1)

Although this statistic is numerically equivalent to the
statistic proposed by McCarroll and Helms (1987), it has several advantages. First, the form of the statistic as a ratio of two "quadratic forms", which are shown to be independent, provides a more intuitive F-statistic. More importantly, this reparameterized form provides a convenient mechanism for modifying the likelihood function in order to develop independent estimates of the covariance parameters and the error variance.

The data vector was partitioned into two orthogonal components based on projections onto the error space and the estimation space using the Helms-McCarroll procedure. Then using the reparameterized form of the model, the error variance was concentrated out of the likelihood for the estimation space component, and the concentrated likelihood was used to derive maximum likelihood estimates of the covariance parameters and fixed effects based on the estimation space. The estimate of the error variance was derived by maximizing the likelihood for the component of the data vector partitioned onto the error space. The orthogonality of the error space and estimation space were used to establish the independence of the random variables in the numerator and denominator of $F_w$.

As a part of the derivation of the modified form of $F_w$, the differences in the estimators developed by the classic REML and Helms-McCarroll procedure were examined. In particular, the effect of the differences on properties of the test statistics were considered. We concluded that further mathematical characterization of the distribution of the $F_{REML}$ statistic generated using classic REML procedures is intractable at this time.

Although the form of $F_w$ given above has the appearance of the ratio of two quadratic forms, it is not strictly so. Because the weight matrix in the numerator is a function of the covariance parameter estimates, it is a random matrix. The distributional properties of the numerator are further complicated by the fact that these covariance parameter
estimates are generated via a numerical rather than an analytical procedure, making their exact distributional properties intractable. In subsequent analyses of the distributional properties of $F_{\hat{y}}$, these complications were addressed by assuming that the covariance parameter estimates are nonstochastic variables. Under this assumption, we demonstrated in Chapter 3 that a linear transformation of $(F_{\hat{y}} | \hat{y}_{\hat{y}})$ ($(R_{\hat{y}} | \hat{y}_{\hat{y}}) = \frac{c}{N-B} \cdot (F_{\hat{y}} | \hat{y}_{\hat{y}})$)

converges analytically to a variable that has the same form as an equivalent linear transformation of $F_{wls}$. Also, $F_{wls}$ was hypothesized by McCarroll and Helms (1987) for $F_{\hat{y}}$. These results provide a theoretical basis for using $F_{\hat{y}}$ as a statistic for testing contrasts among fixed effects in MixMod.

As a part of the convergence demonstration the exact distribution of $(R_{\hat{y}} | \hat{y}_{\hat{y}})$ was demonstrated to be a mixture of ratios of independent random variables with central chi-squared distributions. These results can easily be extended to demonstrate that $(F_{\hat{y}} | \hat{y}_{\hat{y}})$ is distributed as a mixture of random variables with central $F$ distributions. This exact distribution of $(F_{\hat{y}} | \hat{y}_{\hat{y}})$ has little direct applicability for statistical data analyses because the distribution of the mixing variable is a function of the underlying covariance parameters, and if the parameters are replaced by their estimates derived via the Helms-McCarroll procedure the mixture collapses to the hypothesized $F$-distribution. However, while the direct application of the distribution is not viable, the mixture distribution appears to be an important tool for possible future simulation studies of the Type I error rate and power of $F_{\hat{y}}$.

The results in Chapter 3 demonstrate that under the assumption that the covariance parameter estimates are nonstochastic, $(F_{\hat{y}} | \hat{y}_{\hat{y}})$ converges analytically to a function
of \( \gamma \) that has the form of the hypothesized F-distribution. However, they provide no information on the rate of that convergence. In order to develop some information on how close in practice \( F_n \) is to a statistic with the hypothesized F-distribution, numerical evaluations were conducted for a limited set of models. These numerical evaluations also compared \( F_n \) and \( F_{REML} \) for the limited data sets. The results of these numerical analyses are summarized in Chapter 4.

For the two groups of computer generated data sets (one with 20 data sets and the other with 30 data sets), the Helms-McCarroll and classic REML procedures generally showed good agreement. Relative to statistical inference, there appeared to be no apparent biases between the two procedures or between either procedure and the weighted least squares results obtained using the true underlying covariance parameters. These preliminary results suggest that either \( F_n \) or \( F_{REML} \) can be used with some confidence in actual data applications. This result has important practical implications because although \( F_{REML} \) has no substantive theoretical basis, it can be implemented in practice through the SAS procedure PROC MIXED. Hence, it is readily available to the practitioner. On the other hand, although \( F_n \) now has a reasonably sound theoretical basis, it is computationally cumbersome, and the software is not available for implementing it on a wide basis. Consequently, in the near term \( F_{REML} \) will be more readily available for analyzing data from longitudinal studies.

The perturbation results presented in Chapter 4 provide further evidence that \( F_n \) yields reasonable inferential results. In general, the "errors" in \( (F_n | \gamma_n) \) in comparison to \( F_{WLS} \), which was calculated using the known underlying population random effects covariance parameters from the computer-generated data, were of the same order of magnitude as the errors in the covariance parameter estimates. Furthermore, when the Mahalanobis distance between the covariance parameter estimates and the true covariance
parameters was between 0.5 and 3, the range that is expected to predominate in practice for models of the size considered in Chapter 4, errors in both the F-statistic and its associated p-values were reasonably small. For the data sets studied, the number of inferential differences between \( F_H \) and \( F_{WLS} \) was quite small, and the differences showed no bias between the two procedures. Although these results are limited, they suggest that \( F_H \) provides at least as much power as would be obtained from \( F_{WLS} \) if the true covariance parameters were known.

In summary, the earlier results of McCarroll and Helms (1987) indicated that \( F_H \) has acceptable Type I error properties. The results presented here provide a theoretical foundation for using \( F_H \) with the hypothesized F-distribution. The preliminary numerical results suggest that \( F_H \) yields reasonable numerical results and that \( F_H \) appears to provide reasonable power. Furthermore, these results suggest that \( F_{REML} \) can be used as an alternative to \( F_H \) with some confidence.

While the results presented here provide a theoretical basis for \( F_H \) and a preliminary indication that \( F_H \) and \( F_{REML} \) have reasonable numerical properties, they also open avenues for substantial future research. Both theoretical and numerical research are needed to characterize the distributions of \( F_H \) and \( F_{REML} \) more fully. Specific research needs are:

1. The numerical results presented in Chapter 4 are limited to a single MixMod structure that has few underlying covariance parameters. An extensive, carefully structured simulation study is needed to characterize the numerical relationships among \( F_H \), \( F_{REML} \), and \( F_{WLS} \) for a range of model forms and sizes.

2. The theoretical development in Chapter 3 is limited to an error variance of the form \( \sigma^2 I \). Procedures for extending the results within the Helms-McCarroll
framework to the more general form \( \sigma^2 V \) are not readily apparent. However, such covariance structures are of interest, and research to extend the results appears to be warranted.

3. Further research is needed to determine whether the finite sample distribution of \( F_n \) can be determined without the assumption that the covariance parameters are conditioned as nonstochastic variables. As a parallel study, relatively general regularity conditions under which the asymptotic distribution of \( F_n \) can be specified are needed. Both of these problems are likely to be mathematically complicated.

4. The series expansion for the distribution \( F_n \) as a mixture of central \( F \) distributions that can be derived from the results in Chapter 3 appears to be a fruitful tool for exploring the Type I error rate and power implications of using the hypothesized \( F \)-distribution in conjunction with \( F_n \). An immediate area of further research is the exploration these relationships via simulated data for a variety of models.
APPENDIX A

TABULAR SUMMARY
OF COMPUTER-GENERATED DATA SET
ANALYSIS RESULTS
<table>
<thead>
<tr>
<th>Data Base</th>
<th>Run</th>
<th>$\beta_{1W}$</th>
<th>$\beta_{1R}$</th>
<th>$\beta_{1H}$</th>
<th>$\beta_{10L}$</th>
<th>$\beta_{3W}$</th>
<th>$\beta_{3R}$</th>
<th>$\beta_{3H}$</th>
<th>$\beta_{30L}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fairclough</td>
<td>1</td>
<td>1.0832</td>
<td>1.08</td>
<td>1.0826</td>
<td>1.0853</td>
<td>1.3316</td>
<td>1.33</td>
<td>1.3317</td>
<td>1.3303</td>
</tr>
<tr>
<td>Fairclough</td>
<td>2</td>
<td>1.1228</td>
<td>1.12</td>
<td>1.1217</td>
<td>1.1253</td>
<td>1.3029</td>
<td>1.3</td>
<td>1.3029</td>
<td>1.3057</td>
</tr>
<tr>
<td>Fairclough</td>
<td>3</td>
<td>1.1338</td>
<td>1.13</td>
<td>1.1343</td>
<td>1.1411</td>
<td>1.2857</td>
<td>1.29</td>
<td>1.2853</td>
<td>1.2841</td>
</tr>
<tr>
<td>Fairclough</td>
<td>4</td>
<td>1.1103</td>
<td>1.11</td>
<td>1.1106</td>
<td>1.1204</td>
<td>1.3073</td>
<td>1.31</td>
<td>1.3071</td>
<td>1.3133</td>
</tr>
<tr>
<td>Fairclough</td>
<td>5</td>
<td>1.1345</td>
<td>1.14</td>
<td>1.1349</td>
<td>1.125</td>
<td>1.2867</td>
<td>1.29</td>
<td>1.2862</td>
<td>1.2792</td>
</tr>
<tr>
<td>Fairclough</td>
<td>6</td>
<td>1.1346</td>
<td>1.13</td>
<td>1.134</td>
<td>1.1356</td>
<td>1.2852</td>
<td>1.28</td>
<td>1.2828</td>
<td>1.3011</td>
</tr>
<tr>
<td>Fairclough</td>
<td>7</td>
<td>1.1066</td>
<td>1.11</td>
<td>.</td>
<td>1.1056</td>
<td>1.3155</td>
<td>1.31</td>
<td>.</td>
<td>1.306</td>
</tr>
<tr>
<td>Fairclough</td>
<td>8</td>
<td>1.1224</td>
<td>1.12</td>
<td>1.1234</td>
<td>1.1205</td>
<td>1.2886</td>
<td>1.29</td>
<td>1.2897</td>
<td>1.2908</td>
</tr>
<tr>
<td>Fairclough</td>
<td>9</td>
<td>1.1435</td>
<td>1.14</td>
<td>.</td>
<td>1.1426</td>
<td>1.3191</td>
<td>1.32</td>
<td>.</td>
<td>1.3174</td>
</tr>
<tr>
<td>Fairclough</td>
<td>10</td>
<td>1.1274</td>
<td>1.13</td>
<td>1.1294</td>
<td>1.119</td>
<td>1.2959</td>
<td>1.3</td>
<td>1.2972</td>
<td>1.2925</td>
</tr>
<tr>
<td>Fairclough</td>
<td>11</td>
<td>1.0753</td>
<td>1.07</td>
<td>1.0736</td>
<td>1.0637</td>
<td>1.2987</td>
<td>1.3</td>
<td>1.2994</td>
<td>1.2989</td>
</tr>
<tr>
<td>Fairclough</td>
<td>12</td>
<td>1.0986</td>
<td>1.1</td>
<td>1.0985</td>
<td>1.0959</td>
<td>1.3488</td>
<td>1.35</td>
<td>1.3484</td>
<td>1.3414</td>
</tr>
<tr>
<td>Fairclough</td>
<td>13</td>
<td>1.1517</td>
<td>1.15</td>
<td>1.1516</td>
<td>1.1532</td>
<td>1.3921</td>
<td>1.39</td>
<td>1.3916</td>
<td>1.3999</td>
</tr>
<tr>
<td>Fairclough</td>
<td>14</td>
<td>1.0453</td>
<td>1.05</td>
<td>1.0451</td>
<td>1.0469</td>
<td>1.352</td>
<td>1.35</td>
<td>1.3525</td>
<td>1.3504</td>
</tr>
<tr>
<td>Fairclough</td>
<td>15</td>
<td>1.1162</td>
<td>1.12</td>
<td>.</td>
<td>1.1163</td>
<td>1.3349</td>
<td>1.33</td>
<td>.</td>
<td>1.3262</td>
</tr>
<tr>
<td>Fairclough</td>
<td>16</td>
<td>1.124</td>
<td>1.12</td>
<td>.</td>
<td>1.1200</td>
<td>1.2913</td>
<td>1.29</td>
<td>.</td>
<td>1.296</td>
</tr>
<tr>
<td>Fairclough</td>
<td>17</td>
<td>1.1486</td>
<td>1.15</td>
<td>1.1487</td>
<td>1.1573</td>
<td>1.3243</td>
<td>1.32</td>
<td>1.324</td>
<td>1.3258</td>
</tr>
<tr>
<td>Data Base</td>
<td>Run</td>
<td>$\beta_{1W5}$</td>
<td>$\beta_{1w}$</td>
<td>$\beta_{1W}$</td>
<td>$\beta_{1ON5}$</td>
<td>$\beta_{3W5}$</td>
<td>$\beta_{3w}$</td>
<td>$\beta_{3W}$</td>
<td>$\beta_{3ON5}$</td>
</tr>
<tr>
<td>------------</td>
<td>-----</td>
<td>--------------</td>
<td>--------------</td>
<td>--------------</td>
<td>----------------</td>
<td>---------------</td>
<td>--------------</td>
<td>--------------</td>
<td>----------------</td>
</tr>
<tr>
<td>Fairclough</td>
<td>18</td>
<td>1.0897</td>
<td>1.09</td>
<td>1.0902</td>
<td>1.0942</td>
<td>1.2842</td>
<td>1.28</td>
<td>1.2854</td>
<td>1.2868</td>
</tr>
<tr>
<td>Fairclough</td>
<td>19</td>
<td>1.1253</td>
<td>1.13</td>
<td>1.1251</td>
<td>1.1146</td>
<td>1.3274</td>
<td>1.33</td>
<td>1.3275</td>
<td>1.3239</td>
</tr>
<tr>
<td>Fairclough</td>
<td>20</td>
<td>1.0971</td>
<td>1.1</td>
<td>1.0983</td>
<td>1.1013</td>
<td>1.2763</td>
<td>1.28</td>
<td>1.2752</td>
<td>1.2745</td>
</tr>
<tr>
<td>Fair_mod2</td>
<td>1</td>
<td>1.2161</td>
<td></td>
<td>1.2158</td>
<td>1.222</td>
<td>1.4715</td>
<td></td>
<td>1.4716</td>
<td>1.4591</td>
</tr>
<tr>
<td>Fair_mod2</td>
<td>2</td>
<td>1.3108</td>
<td>1.31</td>
<td>1.3103</td>
<td>1.3202</td>
<td>1.3951</td>
<td>1.4</td>
<td>1.3955</td>
<td>1.3994</td>
</tr>
<tr>
<td>Fair_mod2</td>
<td>3</td>
<td>1.3673</td>
<td></td>
<td>1.3678</td>
<td>1.3909</td>
<td>1.2972</td>
<td></td>
<td>1.2966</td>
<td>1.2834</td>
</tr>
<tr>
<td>Fair_mod2</td>
<td>4</td>
<td>1.2903</td>
<td>1.29</td>
<td>1.2901</td>
<td>1.3154</td>
<td>1.4219</td>
<td>1.42</td>
<td>1.4215</td>
<td>1.4322</td>
</tr>
<tr>
<td>Fair_mod2</td>
<td>5</td>
<td>1.3438</td>
<td>1.34</td>
<td>1.3445</td>
<td>1.3109</td>
<td>1.3594</td>
<td>1.36</td>
<td>1.359</td>
<td>1.3426</td>
</tr>
<tr>
<td>Fair_mod2</td>
<td>6</td>
<td>1.3427</td>
<td>1.34</td>
<td>1.3426</td>
<td>1.3406</td>
<td>1.304</td>
<td>1.3</td>
<td>1.3038</td>
<td>1.3571</td>
</tr>
<tr>
<td>Fair_mod2</td>
<td>7</td>
<td>1.2505</td>
<td>1.25</td>
<td>1.2504</td>
<td>1.2525</td>
<td>1.4293</td>
<td>1.43</td>
<td>1.4292</td>
<td>1.403</td>
</tr>
<tr>
<td>Fair_mod2</td>
<td>8</td>
<td>1.3058</td>
<td></td>
<td>1.3078</td>
<td>1.298</td>
<td>1.36</td>
<td></td>
<td>1.361</td>
<td>1.3635</td>
</tr>
<tr>
<td>Fair_mod2</td>
<td>9</td>
<td>1.4226</td>
<td>1.42</td>
<td>1.4224</td>
<td>1.4351</td>
<td>1.4239</td>
<td>1.42</td>
<td>1.4239</td>
<td>1.43</td>
</tr>
<tr>
<td>Fair_mod2</td>
<td>10</td>
<td>1.2815</td>
<td>1.28</td>
<td>1.2802</td>
<td>1.2546</td>
<td>1.4342</td>
<td>1.44</td>
<td>1.4356</td>
<td>1.4209</td>
</tr>
<tr>
<td>Fair_mod2</td>
<td>11</td>
<td>1.1258</td>
<td>1.13</td>
<td>1.123</td>
<td>1.0901</td>
<td>1.4567</td>
<td>1.46</td>
<td>1.4572</td>
<td>1.4561</td>
</tr>
<tr>
<td>Fair_mod2</td>
<td>12</td>
<td>1.2276</td>
<td>1.23</td>
<td>1.2275</td>
<td>1.2201</td>
<td>1.5614</td>
<td>1.56</td>
<td>1.5613</td>
<td>1.5349</td>
</tr>
<tr>
<td>Fair_mod2</td>
<td>13</td>
<td>1.3281</td>
<td>1.32</td>
<td>1.3281</td>
<td>1.3385</td>
<td>1.7418</td>
<td>1.74</td>
<td>1.7416</td>
<td>1.7697</td>
</tr>
<tr>
<td>Fair_mod2</td>
<td>14</td>
<td>1.0209</td>
<td>1.02</td>
<td>1.0215</td>
<td>1.0274</td>
<td>1.5813</td>
<td>1.58</td>
<td>1.5812</td>
<td>1.5867</td>
</tr>
<tr>
<td>Data Base</td>
<td>Run</td>
<td>$\beta_{1wS}$</td>
<td>$\beta_{1r}$</td>
<td>$\beta_{1n}$</td>
<td>$\beta_{10LS}$</td>
<td>$\beta_{3wS}$</td>
<td>$\beta_{3r}$</td>
<td>$\beta_{3n}$</td>
<td>$\beta_{30LS}$</td>
</tr>
<tr>
<td>-------------</td>
<td>-----</td>
<td>---------------</td>
<td>--------------</td>
<td>--------------</td>
<td>----------------</td>
<td>----------------</td>
<td>--------------</td>
<td>--------------</td>
<td>----------------</td>
</tr>
<tr>
<td>Fair_mod2</td>
<td>15</td>
<td>1.2243</td>
<td>1.22</td>
<td>1.2242</td>
<td>1.217</td>
<td>1.5717</td>
<td>1.57</td>
<td>1.5717</td>
<td>1.5401</td>
</tr>
<tr>
<td>Fair_mod2</td>
<td>16</td>
<td>1.3289</td>
<td>1.33</td>
<td>1.3304</td>
<td>1.3113</td>
<td>1.3295</td>
<td>1.33</td>
<td>1.3262</td>
<td>1.3493</td>
</tr>
<tr>
<td>Fair_mod2</td>
<td>17</td>
<td>1.4148</td>
<td>1.41</td>
<td>1.4147</td>
<td>1.4436</td>
<td>1.4542</td>
<td>1.45</td>
<td>1.4543</td>
<td>1.4522</td>
</tr>
<tr>
<td>Fair_mod2</td>
<td>18</td>
<td>1.2306</td>
<td>1.23</td>
<td>1.2306</td>
<td>1.243</td>
<td>1.3097</td>
<td>1.31</td>
<td>1.3097</td>
<td>1.3264</td>
</tr>
<tr>
<td>Fair_mod2</td>
<td>19</td>
<td>1.2572</td>
<td>1.26</td>
<td>1.2569</td>
<td>1.2273</td>
<td>1.5812</td>
<td>1.58</td>
<td>1.5813</td>
<td>1.5848</td>
</tr>
<tr>
<td>Fair_mod2</td>
<td>20</td>
<td>1.1485</td>
<td>1.15</td>
<td>1.1487</td>
<td>1.1571</td>
<td>1.3452</td>
<td>1.35</td>
<td>1.3451</td>
<td>1.3425</td>
</tr>
<tr>
<td>Fair_mod2</td>
<td>21</td>
<td>1.2404</td>
<td>1.24</td>
<td>1.2404</td>
<td>1.2352</td>
<td>1.5145</td>
<td>1.51</td>
<td>1.5147</td>
<td>1.5308</td>
</tr>
<tr>
<td>Fair_mod2</td>
<td>22</td>
<td>1.0341</td>
<td>1.03</td>
<td>1.0339</td>
<td>1.0398</td>
<td>1.781</td>
<td>1.78</td>
<td>1.7813</td>
<td>1.7957</td>
</tr>
<tr>
<td>Fair_mod2</td>
<td>23</td>
<td>0.9626</td>
<td>0.962</td>
<td>0.9622</td>
<td>0.9482</td>
<td>1.7618</td>
<td>1.76</td>
<td>1.7627</td>
<td>1.7491</td>
</tr>
<tr>
<td>Fair_mod2</td>
<td>24</td>
<td>1.0897</td>
<td>.</td>
<td>1.0898</td>
<td>1.0921</td>
<td>1.6204</td>
<td>.</td>
<td>1.6206</td>
<td>1.6101</td>
</tr>
<tr>
<td>Fair_mod2</td>
<td>25</td>
<td>1.2418</td>
<td>1.24</td>
<td>1.2415</td>
<td>1.2668</td>
<td>1.4706</td>
<td>1.47</td>
<td>1.4713</td>
<td>1.4564</td>
</tr>
<tr>
<td>Fair_mod2</td>
<td>26</td>
<td>1.1472</td>
<td>.</td>
<td>1.147</td>
<td>1.1512</td>
<td>1.2551</td>
<td>.</td>
<td>1.2551</td>
<td>1.2239</td>
</tr>
<tr>
<td>Fair_mod2</td>
<td>27</td>
<td>1.1923</td>
<td>1.19</td>
<td>1.1925</td>
<td>1.2014</td>
<td>1.6129</td>
<td>1.61</td>
<td>1.6132</td>
<td>1.5972</td>
</tr>
<tr>
<td>Fair_mod2</td>
<td>28</td>
<td>1.1386</td>
<td>1.14</td>
<td>1.1387</td>
<td>1.1072</td>
<td>1.7216</td>
<td>1.72</td>
<td>1.7217</td>
<td>1.7291</td>
</tr>
<tr>
<td>Fair_mod2</td>
<td>29</td>
<td>0.8526</td>
<td>0.854</td>
<td>0.8538</td>
<td>0.8656</td>
<td>1.5158</td>
<td>1.51</td>
<td>1.5165</td>
<td>1.4991</td>
</tr>
<tr>
<td>Fair_mod2</td>
<td>30</td>
<td>1.1167</td>
<td>1.11</td>
<td>1.1231</td>
<td>1.158</td>
<td>1.7009</td>
<td>1.7</td>
<td>1.7008</td>
<td>1.7117</td>
</tr>
</tbody>
</table>
TABLE A.2. Summary of Fixed Effect Slope Parameter Estimates

<table>
<thead>
<tr>
<th>Data Base</th>
<th>Run</th>
<th>$\beta_{2WS}$</th>
<th>$\beta_{2H}$</th>
<th>$\beta_{2B}$</th>
<th>$\beta_{2OLS}$</th>
<th>$\beta_{4WS}$</th>
<th>$\beta_{4H}$</th>
<th>$\beta_{4B}$</th>
<th>$\beta_{4OLS}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fairclough</td>
<td>1</td>
<td>0.0284</td>
<td>0.02837</td>
<td>0.0284</td>
<td>0.0286</td>
<td>0.03636</td>
<td>0.03647</td>
<td>0.0364</td>
<td>0.0361</td>
</tr>
<tr>
<td>Fairclough</td>
<td>2</td>
<td>0.0311</td>
<td>0.03118</td>
<td>0.0311</td>
<td>0.0306</td>
<td>0.0358</td>
<td>0.0358</td>
<td>0.0358</td>
<td>0.0359</td>
</tr>
<tr>
<td>Fairclough</td>
<td>3</td>
<td>0.0286</td>
<td>0.02863</td>
<td>0.0286</td>
<td>0.0289</td>
<td>0.03633</td>
<td>0.0363</td>
<td>0.0364</td>
<td>0.0368</td>
</tr>
<tr>
<td>Fairclough</td>
<td>4</td>
<td>0.0289</td>
<td>0.0289</td>
<td>0.0289</td>
<td>0.0294</td>
<td>0.03662</td>
<td>0.03665</td>
<td>0.0366</td>
<td>0.0364</td>
</tr>
<tr>
<td>Fairclough</td>
<td>5</td>
<td>0.0302</td>
<td>0.03015</td>
<td>0.0302</td>
<td>0.0300</td>
<td>0.03496</td>
<td>0.03494</td>
<td>0.035</td>
<td>0.0344</td>
</tr>
<tr>
<td>Fairclough</td>
<td>6</td>
<td>0.0277</td>
<td>0.02765</td>
<td>0.0277</td>
<td>0.0280</td>
<td>0.03152</td>
<td>0.03175</td>
<td>0.0317</td>
<td>0.0319</td>
</tr>
<tr>
<td>Fairclough</td>
<td>7</td>
<td>0.0281</td>
<td>.</td>
<td>0.0281</td>
<td>0.0282</td>
<td>0.03364</td>
<td>.</td>
<td>0.0337</td>
<td>0.0338</td>
</tr>
<tr>
<td>Fairclough</td>
<td>8</td>
<td>0.0284</td>
<td>0.02836</td>
<td>0.0284</td>
<td>0.0283</td>
<td>0.03801</td>
<td>0.03798</td>
<td>0.38</td>
<td>0.0380</td>
</tr>
<tr>
<td>Fairclough</td>
<td>9</td>
<td>0.0296</td>
<td>.</td>
<td>0.0295</td>
<td>0.0291</td>
<td>0.03372</td>
<td>.</td>
<td>0.0338</td>
<td>0.0331</td>
</tr>
<tr>
<td>Fairclough</td>
<td>10</td>
<td>0.0305</td>
<td>0.0306</td>
<td>0.0307</td>
<td>0.0297</td>
<td>0.0371</td>
<td>0.03694</td>
<td>0.0371</td>
<td>0.0368</td>
</tr>
<tr>
<td>Fairclough</td>
<td>11</td>
<td>0.0269</td>
<td>0.02682</td>
<td>0.0269</td>
<td>0.0263</td>
<td>0.03623</td>
<td>0.03612</td>
<td>0.0362</td>
<td>0.0358</td>
</tr>
<tr>
<td>Fairclough</td>
<td>12</td>
<td>0.0274</td>
<td>0.0274</td>
<td>0.0274</td>
<td>0.0274</td>
<td>0.03694</td>
<td>0.03697</td>
<td>0.037</td>
<td>0.0365</td>
</tr>
<tr>
<td>Fairclough</td>
<td>13</td>
<td>0.0312</td>
<td>0.03117</td>
<td>0.0312</td>
<td>0.0311</td>
<td>0.04184</td>
<td>0.04187</td>
<td>0.0419</td>
<td>0.0414</td>
</tr>
<tr>
<td>Fairclough</td>
<td>14</td>
<td>0.0255</td>
<td>0.02541</td>
<td>0.0253</td>
<td>0.0257</td>
<td>0.03647</td>
<td>0.03655</td>
<td>0.0365</td>
<td>0.0360</td>
</tr>
<tr>
<td>Fairclough</td>
<td>15</td>
<td>0.0317</td>
<td>.</td>
<td>0.0317</td>
<td>0.0320</td>
<td>0.03948</td>
<td>.</td>
<td>0.0394</td>
<td>0.0390</td>
</tr>
<tr>
<td>Fairclough</td>
<td>16</td>
<td>0.0321</td>
<td>.</td>
<td>0.0321</td>
<td>0.0314</td>
<td>0.03731</td>
<td>.</td>
<td>0.0373</td>
<td>0.0379</td>
</tr>
<tr>
<td>Fairclough</td>
<td>17</td>
<td>0.0319</td>
<td>0.03186</td>
<td>0.0319</td>
<td>0.0317</td>
<td>0.0344</td>
<td>0.03443</td>
<td>0.0344</td>
<td>0.0344</td>
</tr>
<tr>
<td>Fairclough</td>
<td>18</td>
<td>0.0294</td>
<td>0.02934</td>
<td>0.0293</td>
<td>0.0296</td>
<td>0.03531</td>
<td>0.03528</td>
<td>0.0353</td>
<td>0.0353</td>
</tr>
<tr>
<td>Fairclough</td>
<td>19</td>
<td>0.0296</td>
<td>0.02966</td>
<td>0.0296</td>
<td>0.0294</td>
<td>0.03613</td>
<td>0.03613</td>
<td>0.0361</td>
<td>0.0362</td>
</tr>
<tr>
<td>Database</td>
<td>Run</td>
<td>$\beta_{2uLS}$</td>
<td>$\beta_{2u}$</td>
<td>$\beta_{2v}$</td>
<td>$\beta_{2OLS}$</td>
<td>$\beta_{4uLS}$</td>
<td>$\beta_{4u}$</td>
<td>$\beta_{4v}$</td>
<td>$\beta_{4OLS}$</td>
</tr>
<tr>
<td>---------------</td>
<td>-----</td>
<td>----------------</td>
<td>--------------</td>
<td>--------------</td>
<td>----------------</td>
<td>----------------</td>
<td>--------------</td>
<td>--------------</td>
<td>----------------</td>
</tr>
<tr>
<td>Fairclough</td>
<td>20</td>
<td>0.0272</td>
<td>0.02717</td>
<td>0.0272</td>
<td>0.0274</td>
<td>0.03318</td>
<td>0.03327</td>
<td>0.0332</td>
<td>0.0337</td>
</tr>
<tr>
<td>FVC_mod2</td>
<td>1</td>
<td>0.0340</td>
<td>0.03395</td>
<td>.</td>
<td>0.0342</td>
<td>0.03748</td>
<td>0.03751</td>
<td>.</td>
<td>0.0370</td>
</tr>
<tr>
<td>FVC_mod2</td>
<td>2</td>
<td>0.0380</td>
<td>0.03792</td>
<td>0.0379</td>
<td>0.0365</td>
<td>0.03751</td>
<td>0.03752</td>
<td>0.0375</td>
<td>0.0381</td>
</tr>
<tr>
<td>FVC_mod2</td>
<td>3</td>
<td>0.0341</td>
<td>0.03412</td>
<td>.</td>
<td>0.0337</td>
<td>0.03946</td>
<td>0.0395</td>
<td>.</td>
<td>0.0414</td>
</tr>
<tr>
<td>FVC_mod2</td>
<td>4</td>
<td>0.0344</td>
<td>0.03442</td>
<td>0.0345</td>
<td>0.0360</td>
<td>0.03877</td>
<td>0.03882</td>
<td>0.0388</td>
<td>0.0389</td>
</tr>
<tr>
<td>FVC_mod2</td>
<td>5</td>
<td>0.0373</td>
<td>0.03725</td>
<td>0.0372</td>
<td>0.0376</td>
<td>0.03704</td>
<td>0.03701</td>
<td>0.037</td>
<td>0.0352</td>
</tr>
<tr>
<td>FVC_mod2</td>
<td>6</td>
<td>0.0316</td>
<td>0.03164</td>
<td>0.0316</td>
<td>0.0326</td>
<td>0.03151</td>
<td>0.03161</td>
<td>0.0316</td>
<td>0.0311</td>
</tr>
<tr>
<td>FVC_mod2</td>
<td>7</td>
<td>0.0307</td>
<td>0.03072</td>
<td>0.0307</td>
<td>0.0298</td>
<td>0.03571</td>
<td>0.03573</td>
<td>0.0357</td>
<td>0.036</td>
</tr>
<tr>
<td>FVC_mod2</td>
<td>8</td>
<td>0.0323</td>
<td>0.03228</td>
<td>.</td>
<td>0.0328</td>
<td>0.04388</td>
<td>0.04385</td>
<td>.</td>
<td>0.0442</td>
</tr>
<tr>
<td>FVC_mod2</td>
<td>9</td>
<td>0.0339</td>
<td>0.03392</td>
<td>0.034</td>
<td>0.0308</td>
<td>0.03715</td>
<td>0.03718</td>
<td>0.0372</td>
<td>0.036</td>
</tr>
<tr>
<td>FVC_mod2</td>
<td>10</td>
<td>0.0353</td>
<td>0.03538</td>
<td>0.0354</td>
<td>0.0345</td>
<td>0.04173</td>
<td>0.04164</td>
<td>0.0417</td>
<td>0.0408</td>
</tr>
<tr>
<td>FVC_mod2</td>
<td>11</td>
<td>0.0291</td>
<td>0.0287</td>
<td>0.0289</td>
<td>0.0278</td>
<td>0.04064</td>
<td>0.04053</td>
<td>0.0406</td>
<td>0.0399</td>
</tr>
<tr>
<td>FVC_mod2</td>
<td>12</td>
<td>0.0310</td>
<td>0.03107</td>
<td>0.0311</td>
<td>0.0320</td>
<td>0.0407</td>
<td>0.04073</td>
<td>0.0407</td>
<td>0.0402</td>
</tr>
<tr>
<td>FVC_mod2</td>
<td>13</td>
<td>0.0385</td>
<td>0.03844</td>
<td>0.0385</td>
<td>0.0385</td>
<td>0.04547</td>
<td>0.04548</td>
<td>0.0455</td>
<td>0.0444</td>
</tr>
<tr>
<td>FVC_mod2</td>
<td>14</td>
<td>0.0304</td>
<td>0.03015</td>
<td>0.0303</td>
<td>0.0295</td>
<td>0.03856</td>
<td>0.03852</td>
<td>0.0385</td>
<td>0.037</td>
</tr>
<tr>
<td>FVC_mod2</td>
<td>15</td>
<td>0.0386</td>
<td>0.03855</td>
<td>0.0386</td>
<td>0.0395</td>
<td>0.04318</td>
<td>0.04319</td>
<td>0.0432</td>
<td>0.0422</td>
</tr>
<tr>
<td>FVC_mod2</td>
<td>16</td>
<td>0.0349</td>
<td>0.03358</td>
<td>0.0349</td>
<td>0.0334</td>
<td>0.04395</td>
<td>0.04431</td>
<td>0.0439</td>
<td>0.0454</td>
</tr>
<tr>
<td>FVC_mod2</td>
<td>17</td>
<td>0.0350</td>
<td>0.03495</td>
<td>0.035</td>
<td>0.0354</td>
<td>0.03988</td>
<td>0.03986</td>
<td>0.0399</td>
<td>0.0407</td>
</tr>
<tr>
<td>FVC_mod2</td>
<td>18</td>
<td>0.0312</td>
<td>0.03115</td>
<td>0.0311</td>
<td>0.0314</td>
<td>0.04168</td>
<td>0.04167</td>
<td>0.0417</td>
<td>0.0410</td>
</tr>
</tbody>
</table>
### TABLE A.3. Summary of Covariance Parameter Estimates

<table>
<thead>
<tr>
<th>Data Base</th>
<th>Run</th>
<th>$\gamma_{1r}$</th>
<th>$\gamma_{2r}$</th>
<th>$\gamma_{1w}$</th>
<th>$\gamma_{1w}$</th>
<th>$\gamma_{2w}$</th>
<th>$\gamma_{3w}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fairclough</td>
<td>1</td>
<td>2.4253</td>
<td>0.085977</td>
<td>0.006437</td>
<td>3.9449</td>
<td>0.135227</td>
<td>0.010651</td>
</tr>
<tr>
<td>Fairclough</td>
<td>2</td>
<td>1.6034</td>
<td>0.011897</td>
<td>0.002069</td>
<td>2.8231</td>
<td>0.020204</td>
<td>0.004068</td>
</tr>
<tr>
<td>Fairclough</td>
<td>3</td>
<td>1.4286</td>
<td>0.048393</td>
<td>0.003661</td>
<td>2.6485</td>
<td>0.082794</td>
<td>0.007161</td>
</tr>
<tr>
<td>Fairclough</td>
<td>4</td>
<td>1.5766</td>
<td>0.039189</td>
<td>0.002973</td>
<td>1.6655</td>
<td>0.041138</td>
<td>0.003165</td>
</tr>
<tr>
<td>Fairclough</td>
<td>5</td>
<td>1.973</td>
<td>0.044054</td>
<td>0.003423</td>
<td>1.6941</td>
<td>0.038422</td>
<td>0.002878</td>
</tr>
<tr>
<td>Fairclough</td>
<td>6</td>
<td>1.7352</td>
<td>0.025677</td>
<td>0.002808</td>
<td>4.1357</td>
<td>0.060767</td>
<td>0.007747</td>
</tr>
<tr>
<td>Fairclough</td>
<td>7</td>
<td>1.4355</td>
<td>0.044194</td>
<td>0.002581</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fairclough</td>
<td>8</td>
<td>1.0085</td>
<td>0.026154</td>
<td>0.002564</td>
<td>0.6236</td>
<td>0.018292</td>
<td>0.001336</td>
</tr>
<tr>
<td>Fairclough</td>
<td>9</td>
<td>0.902</td>
<td>0.033251</td>
<td>0.004963</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fairclough</td>
<td>10</td>
<td>1.9115</td>
<td>0.09646</td>
<td>0.007168</td>
<td>0.8978</td>
<td>0.051313</td>
<td>0.003159</td>
</tr>
<tr>
<td>Fairclough</td>
<td>11</td>
<td>1.2179</td>
<td>0.042244</td>
<td>0.002051</td>
<td>0.9773</td>
<td>0.034947</td>
<td>0.001504</td>
</tr>
<tr>
<td>Fairclough</td>
<td>12</td>
<td>1.9515</td>
<td>0.074369</td>
<td>0.005631</td>
<td>1.5635</td>
<td>0.061191</td>
<td>0.004419</td>
</tr>
<tr>
<td>Fairclough</td>
<td>13</td>
<td>1.7864</td>
<td>0.041942</td>
<td>0.003592</td>
<td>1.5423</td>
<td>0.037321</td>
<td>0.003028</td>
</tr>
<tr>
<td>Fairclough</td>
<td>14</td>
<td>1.8281</td>
<td>0.067891</td>
<td>0.003359</td>
<td>2.4551</td>
<td>0.088957</td>
<td>0.004783</td>
</tr>
<tr>
<td>Fairclough</td>
<td>15</td>
<td>1.2444</td>
<td>0.037111</td>
<td>0.002148</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fairclough</td>
<td>16</td>
<td>1.3542</td>
<td>0.060139</td>
<td>0.004306</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fairclough</td>
<td>17</td>
<td>1.5556</td>
<td>0.049365</td>
<td>0.004127</td>
<td>1.4637</td>
<td>0.046715</td>
<td>0.003841</td>
</tr>
<tr>
<td>Fairclough</td>
<td>18</td>
<td>1.1512</td>
<td>0.042743</td>
<td>0.003103</td>
<td>0.9464</td>
<td>0.036319</td>
<td>0.00243</td>
</tr>
<tr>
<td>Fairclough</td>
<td>19</td>
<td>1.8803</td>
<td>0.066178</td>
<td>0.00464</td>
<td>1.7096</td>
<td>0.06107</td>
<td>0.004112</td>
</tr>
<tr>
<td>Data Base</td>
<td>Run</td>
<td>(Y_{1n})</td>
<td>(Y_{2n})</td>
<td>(Y_{3n})</td>
<td>(Y_{1w})</td>
<td>(Y_{2w})</td>
<td>(Y_{3w})</td>
</tr>
<tr>
<td>------------</td>
<td>-----</td>
<td>-------------</td>
<td>-------------</td>
<td>-------------</td>
<td>-------------</td>
<td>-------------</td>
<td>-------------</td>
</tr>
<tr>
<td>Fairclough</td>
<td>20</td>
<td>1.8718</td>
<td>0.061453</td>
<td>0.004188</td>
<td>1.1362</td>
<td>0.040822</td>
<td>0.002373</td>
</tr>
<tr>
<td>FVC_mod2</td>
<td>1</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>33.4396</td>
<td>0.423617</td>
<td>0.016112</td>
</tr>
<tr>
<td>FVC_mod2</td>
<td>2</td>
<td>15.2475</td>
<td>0.030198</td>
<td>0.006832</td>
<td>17.6778</td>
<td>0.033888</td>
<td>0.008231</td>
</tr>
<tr>
<td>FVC_mod2</td>
<td>3</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>19.9882</td>
<td>0.26863</td>
<td>0.012283</td>
</tr>
<tr>
<td>FVC_mod2</td>
<td>4</td>
<td>21.1458</td>
<td>0.18125</td>
<td>0.008646</td>
<td>31.4553</td>
<td>0.271033</td>
<td>0.013548</td>
</tr>
<tr>
<td>FVC_mod2</td>
<td>5</td>
<td>22.7178</td>
<td>0.181535</td>
<td>0.010166</td>
<td>20.481</td>
<td>0.16425</td>
<td>0.009196</td>
</tr>
<tr>
<td>FVC_mod2</td>
<td>6</td>
<td>18.3391</td>
<td>0.123414</td>
<td>0.009343</td>
<td>18.6224</td>
<td>0.124454</td>
<td>0.009467</td>
</tr>
<tr>
<td>FVC_mod2</td>
<td>7</td>
<td>20.2913</td>
<td>0.200971</td>
<td>0.007961</td>
<td>22.3414</td>
<td>0.220939</td>
<td>0.008966</td>
</tr>
<tr>
<td>FVC_mod2</td>
<td>8</td>
<td>.</td>
<td>.</td>
<td>.</td>
<td>7.41</td>
<td>0.081551</td>
<td>0.004059</td>
</tr>
<tr>
<td>FVC_mod2</td>
<td>9</td>
<td>16.3265</td>
<td>0.115306</td>
<td>0.013848</td>
<td>12.8716</td>
<td>0.090162</td>
<td>0.011367</td>
</tr>
<tr>
<td>FVC_mod2</td>
<td>10</td>
<td>24.4984</td>
<td>0.50792</td>
<td>0.018163</td>
<td>19.8078</td>
<td>0.41494</td>
<td>0.014582</td>
</tr>
<tr>
<td>FVC_mod2</td>
<td>11</td>
<td>16.2044</td>
<td>0.19562</td>
<td>0.006058</td>
<td>4.9304</td>
<td>0.061172</td>
<td>0.001042</td>
</tr>
<tr>
<td>FVC_mod2</td>
<td>12</td>
<td>24.263</td>
<td>0.348073</td>
<td>0.014626</td>
<td>19.9244</td>
<td>0.286363</td>
<td>0.011887</td>
</tr>
<tr>
<td>FVC_mod2</td>
<td>13</td>
<td>23.4163</td>
<td>0.173077</td>
<td>0.00905</td>
<td>21.6117</td>
<td>0.158558</td>
<td>0.008252</td>
</tr>
<tr>
<td>FVC_mod2</td>
<td>14</td>
<td>21.8584</td>
<td>0.262832</td>
<td>0.008319</td>
<td>12.3773</td>
<td>0.15002</td>
<td>0.004075</td>
</tr>
<tr>
<td>FVC_mod2</td>
<td>15</td>
<td>15.7018</td>
<td>0.178947</td>
<td>0.009211</td>
<td>17.9191</td>
<td>0.205469</td>
<td>0.01075</td>
</tr>
<tr>
<td>FVC_mod2</td>
<td>16</td>
<td>20.1653</td>
<td>0.328926</td>
<td>0.010826</td>
<td>2.0038</td>
<td>0.034698</td>
<td>0.000073</td>
</tr>
<tr>
<td>FVC_mod2</td>
<td>17</td>
<td>22.963</td>
<td>0.296296</td>
<td>0.011574</td>
<td>21.4596</td>
<td>0.277232</td>
<td>0.010799</td>
</tr>
<tr>
<td>FVC_mod2</td>
<td>18</td>
<td>13.5514</td>
<td>0.167056</td>
<td>0.008411</td>
<td>17.9298</td>
<td>0.217605</td>
<td>0.011484</td>
</tr>
<tr>
<td>Data Base</td>
<td>Run</td>
<td>$Y_{1R}$</td>
<td>$Y_{2R}$</td>
<td>$Y_{1H}$</td>
<td>$Y_{2H}$</td>
<td>$Y_{3H}$</td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>-----</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td>----------</td>
<td></td>
</tr>
<tr>
<td>FVC_mod2</td>
<td>19</td>
<td>21.0452</td>
<td>0.314972</td>
<td>0.012147</td>
<td>20.5793</td>
<td>0.30842</td>
<td>0.011947</td>
</tr>
<tr>
<td>FVC_mod2</td>
<td>20</td>
<td>25.9259</td>
<td>0.278278</td>
<td>0.01011</td>
<td>23.6158</td>
<td>0.253711</td>
<td>0.009183</td>
</tr>
<tr>
<td>FVC_mod2</td>
<td>21</td>
<td>26.3098</td>
<td>0.310934</td>
<td>0.01344</td>
<td>28.2993</td>
<td>0.334574</td>
<td>0.014742</td>
</tr>
<tr>
<td>FVC_mod2</td>
<td>22</td>
<td>16.7826</td>
<td>0.198261</td>
<td>0.009043</td>
<td>19.08</td>
<td>0.226393</td>
<td>0.010853</td>
</tr>
<tr>
<td>FVC_mod2</td>
<td>23</td>
<td>20.1357</td>
<td>0.247738</td>
<td>0.010068</td>
<td>10.9166</td>
<td>0.130621</td>
<td>0.004761</td>
</tr>
<tr>
<td>FVC_mod2</td>
<td>24</td>
<td></td>
<td></td>
<td></td>
<td>23.2923</td>
<td>0.255777</td>
<td>0.009269</td>
</tr>
<tr>
<td>FVC_mod2</td>
<td>25</td>
<td>16.7173</td>
<td>0.043161</td>
<td>0.007801</td>
<td>17.2952</td>
<td>0.042698</td>
<td>0.008215</td>
</tr>
<tr>
<td>FVC_mod2</td>
<td>26</td>
<td></td>
<td></td>
<td></td>
<td>22.2675</td>
<td>0.294529</td>
<td>0.010639</td>
</tr>
<tr>
<td>FVC_mod2</td>
<td>27</td>
<td>18.5271</td>
<td>0.279845</td>
<td>0.00938</td>
<td>21.6728</td>
<td>0.326189</td>
<td>0.011192</td>
</tr>
<tr>
<td>FVC_mod2</td>
<td>28</td>
<td>13.8614</td>
<td>0.150495</td>
<td>0.009208</td>
<td>16.7489</td>
<td>0.18268</td>
<td>0.011454</td>
</tr>
<tr>
<td>FVC_mod2</td>
<td>29</td>
<td>12.4627</td>
<td>0.24403</td>
<td>0.008955</td>
<td>18.2736</td>
<td>0.35738</td>
<td>0.014059</td>
</tr>
<tr>
<td>FVC_mod2</td>
<td>30</td>
<td>18.3846</td>
<td>0.09</td>
<td>0.008462</td>
<td>6.4474</td>
<td>0.032203</td>
<td>0.002177</td>
</tr>
<tr>
<td>Data Base</td>
<td>Run</td>
<td>$F_{REM}$</td>
<td>$S^2_{W}$</td>
<td>$F_{MS}$</td>
<td>$\sigma^2_{REM}$</td>
<td>$\sigma^2_{W}$</td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>-----</td>
<td>-----------</td>
<td>-----------</td>
<td>----------</td>
<td>-----------------</td>
<td>---------------</td>
<td></td>
</tr>
<tr>
<td>Fairclough</td>
<td>1</td>
<td>26.42</td>
<td>5.5</td>
<td>12.87</td>
<td>0.0087</td>
<td>0.013387</td>
<td></td>
</tr>
<tr>
<td>Fairclough</td>
<td>2</td>
<td>15.95</td>
<td>5.55</td>
<td>7.72</td>
<td>0.0116</td>
<td>0.01177</td>
<td></td>
</tr>
<tr>
<td>Fairclough</td>
<td>3</td>
<td>12.58</td>
<td>1.37</td>
<td>2.33</td>
<td>0.0112</td>
<td>0.03524</td>
<td></td>
</tr>
<tr>
<td>Fairclough</td>
<td>4</td>
<td>19.57</td>
<td>1.14</td>
<td>1.12</td>
<td>0.0111</td>
<td>0.103691</td>
<td></td>
</tr>
<tr>
<td>Fairclough</td>
<td>5</td>
<td>9.29</td>
<td>4.59</td>
<td>4.7</td>
<td>0.0111</td>
<td>0.013518</td>
<td></td>
</tr>
<tr>
<td>Fairclough</td>
<td>6</td>
<td>11.65</td>
<td>2.75</td>
<td>6.27</td>
<td>0.00997</td>
<td>0.010103</td>
<td></td>
</tr>
<tr>
<td>Fairclough</td>
<td>7</td>
<td>20.82</td>
<td>.</td>
<td>2.02</td>
<td>0.0124</td>
<td>0.059878</td>
<td></td>
</tr>
<tr>
<td>Fairclough</td>
<td>8</td>
<td>19.97</td>
<td>20.96</td>
<td>9.83</td>
<td>0.0117</td>
<td>0.01145</td>
<td></td>
</tr>
<tr>
<td>Fairclough</td>
<td>9</td>
<td>32.41</td>
<td>.</td>
<td>10.87</td>
<td>0.00806</td>
<td>0.007943</td>
<td></td>
</tr>
<tr>
<td>Fairclough</td>
<td>10</td>
<td>11.35</td>
<td>3.85</td>
<td>2.59</td>
<td>0.0113</td>
<td>0.032256</td>
<td></td>
</tr>
<tr>
<td>Fairclough</td>
<td>11</td>
<td>22.83</td>
<td>15.18</td>
<td>10.09</td>
<td>0.0156</td>
<td>0.015443</td>
<td></td>
</tr>
<tr>
<td>Fairclough</td>
<td>12</td>
<td>27.76</td>
<td>4.48</td>
<td>4.76</td>
<td>0.0103</td>
<td>0.038899</td>
<td></td>
</tr>
<tr>
<td>Fairclough</td>
<td>13</td>
<td>28.52</td>
<td>19.88</td>
<td>17.89</td>
<td>0.0103</td>
<td>0.01028</td>
<td></td>
</tr>
<tr>
<td>Fairclough</td>
<td>14</td>
<td>35.48</td>
<td>13.5</td>
<td>20.8</td>
<td>0.0128</td>
<td>0.013019</td>
<td></td>
</tr>
<tr>
<td>Fairclough</td>
<td>15</td>
<td>23.86</td>
<td>.</td>
<td>1.11</td>
<td>0.0135</td>
<td>0.123508</td>
<td></td>
</tr>
<tr>
<td>Fairclough</td>
<td>16</td>
<td>11.7</td>
<td>.</td>
<td>2.29</td>
<td>0.0144</td>
<td>0.034075</td>
<td></td>
</tr>
<tr>
<td>Fairclough</td>
<td>17</td>
<td>13.07</td>
<td>7.52</td>
<td>7.23</td>
<td>0.0126</td>
<td>0.012601</td>
<td></td>
</tr>
<tr>
<td>Fairclough</td>
<td>18</td>
<td>27.05</td>
<td>4.07</td>
<td>2.69</td>
<td>0.00999</td>
<td>0.039354</td>
<td></td>
</tr>
<tr>
<td>Fairclough</td>
<td>19</td>
<td>23.25</td>
<td>12.51</td>
<td>13.82</td>
<td>0.00819</td>
<td>0.008259</td>
<td></td>
</tr>
<tr>
<td>Data Base</td>
<td>Run</td>
<td>$F_{REW}$</td>
<td>$F_{MAX}$</td>
<td>$\sigma_{REW}^2$</td>
<td>$\sigma_{MAX}^2$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>-----</td>
<td>-----------</td>
<td>-----------</td>
<td>------------------</td>
<td>------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fairclough</td>
<td>1</td>
<td>13.08</td>
<td>7.85</td>
<td>0.0117</td>
<td>0.00163</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FVC mod2</td>
<td>2</td>
<td>0.46</td>
<td>0.28</td>
<td>0.032451</td>
<td>0.056916</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FVC mod2</td>
<td>3</td>
<td>0.36</td>
<td>0.37</td>
<td>0.010146</td>
<td>0.00117</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FVC mod2</td>
<td>4</td>
<td>0.87</td>
<td>0.04</td>
<td>0.212774</td>
<td>0.00096</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FVC mod2</td>
<td>5</td>
<td>0.01</td>
<td>0.01</td>
<td>0.00964</td>
<td>0.000964</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FVC mod2</td>
<td>6</td>
<td>0.09</td>
<td>0.05</td>
<td>0.00867</td>
<td>0.000964</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FVC mod2</td>
<td>7</td>
<td>1.56</td>
<td>0.14</td>
<td>0.00103</td>
<td>0.000964</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FVC mod2</td>
<td>8</td>
<td>0.64</td>
<td>0.86</td>
<td>0.0017</td>
<td>0.006847</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FVC mod2</td>
<td>9</td>
<td>10.71</td>
<td>6.67</td>
<td>0.00117</td>
<td>0.000964</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FVC mod2</td>
<td>10</td>
<td>0.07</td>
<td>0.04</td>
<td>0.00947</td>
<td>0.006847</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FVC mod2</td>
<td>11</td>
<td>5.11</td>
<td>4.71</td>
<td>0.0137</td>
<td>0.004415</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FVC mod2</td>
<td>12</td>
<td>0.29</td>
<td>0.41</td>
<td>0.00882</td>
<td>0.04415</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FVC mod2</td>
<td>13</td>
<td>8.32</td>
<td>5.15</td>
<td>0.00884</td>
<td>0.013313</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FVC mod2</td>
<td>14</td>
<td>12.6</td>
<td>7.12</td>
<td>0.0113</td>
<td>0.00884</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FVC mod2</td>
<td>15</td>
<td>6.64</td>
<td>0.22</td>
<td>0.0114</td>
<td>0.00884</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FVC mod2</td>
<td>16</td>
<td>0.08</td>
<td>0.76</td>
<td>0.0108</td>
<td>0.00884</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FVC mod2</td>
<td>17</td>
<td>1.07</td>
<td>1.07</td>
<td>0.05988</td>
<td>0.01083</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FVC mod2</td>
<td>18</td>
<td>0.67</td>
<td>0.93</td>
<td>0.05988</td>
<td>0.01083</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
SELECTED BIBLIOGRAPHY


Patterson, H. D. and Thompson, R. (1971). "Recovery of Interblock Information when Block Sizes are Unequal." Biometrika, 58, 545-554.


