INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

A COMPARISON OF THE ACCURACY OF AD HOC AND BAYESIAN ESTIMATES OF THE POPULATION SQUARED MULTIPLE CORRELATION COMPUTED FROM INCOMPLETE DATA.

BY

JOSEPH ONUBOGU

A dissertation submitted to the Graduate Faculty
in Educational Psychology in partial fulfillment
of the requirements for the degree of doctor of
Philosophy, The City University of New York

1999

UMI Number: 9924834

UMI
300 North Zeeb Road
Ann Arbor, MI 48103

APPROVAL

This manuscript has been read and accepted for the Graduate Faculty in Educational Psychology in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

4/28/99
Date

Chair of Examining Committee
Alan L. Gross

4/28/99
Date

Executive Officer
Alan L. Gross

Professor David Rindskopf
Professor Carol Kehr Tittle
Supervisory Committee.

The City University of New York

Abstract

A COMPARISON OF THE ACCURACY OF AD HOC AND BAYESIAN ESTIMATES OF THE POPULATION SQUARED MULTIPLE CORRELATION COMPUTED FROM INCOMPLETE DATA.

BY

JOSEPH ONUBOGU

Adviser: Professor Alan Gross

The accuracy of interval estimates for the population squared multiple correlation (ρ^2) computed from incomplete data was assessed using three ad hoc methods and the Bayesian procedure based on the Gibbs sampler. Listwise deletion, pairwise deletion, and unconditional fill-in were the three ad hoc methods used.

A multivariate normal distribution for $q > 1$ predictor variables and one dependent variable, with a variance-covariance matrix Σ for the distribution, and a value for population squared multiple correlation ρ^2 was specified. A random sample of size n was then drawn from this distribution and data deleted on both the dependent variable and independent variables using a MCAR and MAR processes. The incomplete data set was analyzed using the four methods, and a sample interval estimate of ρ^2 for each method was computed. This analysis was repeated by varying the sample size

(n = 50, 100, 250) the number of predictors (2, 5, 7) the value of the squared multiple correlation ρ^2 (.1, .25, .50) and the amount of the missing data (.10, .40). 95% confidence intervals (CI) for ρ^2 estimates from the three ad hoc methods were computed and 95% central posterior intervals were calculated for the Bayesian approach based on the Gibbs sampler.

The results indicate that listwise deletion has the most accurate interval estimates for ρ^2 among the three ad hoc methods and fill-in has the least. Our overall result indicated that the Bayesian interval estimates for ρ^2 covered the parameter values fewer times than interval estimates from the ad hoc methods for sample sizes less than or equal to 100 and ρ^2 less than .5. The prior for ρ^2 was the main reason for the Bayesian relative poor performance.

ACKNOWLEDGEMENTS

This has been a long and interesting academic pursuit. I have enjoyed everyday of this period. I wish to thank all the members of the educational psychology department staff who had taught and helped me during this period to successfully complete my PHD program.

I wish to express a special appreciation to professors Tittle and Rindskopf for their supportive and advisory roles as members of my dissertation committee. Your inputs made a big difference in this dissertation research effort.

To my adviser, Professor Alan Gross, I wish to say thank you for everything you have done for me to make this dissertation project a success. You were very supportive, understanding, and accommodating. I am indeed fortunate to have you as my adviser.

I wish to express special thanks to my parents Ogbuefi and Mrs Onubogu for teaching me to be diligent. My thanks go to all my eight siblings for their supportive roles. Anthony Onubogu and his wife Ann were always there when I needed emotional support. I wish also to thank Professor Wilson Obi and Dolores Abernathy for all their encouragement and advice during this period of my life.

Finally, I wish to thank my wife Chizoba for the encouragement and the tremendous support she had given me during this period. The birth of our son, Joseph Jr., on 11/12/98, was a source of happiness and joy for Chizoba and I. We are enjoying every minute of his growing up.

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF FIGURES

Figure

Chapter I

INTRODUCTION

Educational researchers are often concerned with the relationship between a dependent variable (y) and a set of continuous independent variables (**x**) within a population of students. They typically use multiple regression analyses computed on a sample to provide information on the population squared multiple correlation coefficient, ρ^2, (coefficient of multiple determination). This parameter is a measure of the linear relationship of the dependent variable to the set of independent variables. The sample squared multiple squared, R^2 , is used to estimate ρ^2 (Dillon & Goldstein, 1984, p. 221). Although it is a relatively easy process to compute the sample multiple correlation, R^2 if the data are complete, when missing values are encountered, the problem of computing an estimate of ρ^2 can become complex. In the present study the problem of estimating the population ρ^2 values when data are missing on both the dependent variable and independent variables was investigated. Rather than considering only point estimates, interval estimates for the population ρ^2 values were considered. An interval estimate of a parameter allows one to establish a confidence bound around the parameter estimated.

Missing values occur in different forms when one performs a multiple regression analysis. In some cases, the missingness may be unrelated to either y or $\underline{x}$, that is, the missing values may be missing completely at random (MCAR). As an example, consider a situation where a college administers a test battery $\underline{x}$ to all incoming freshmen students. The college administration would like to know the relationship between students' test scores and their subsequent first year GPA (y) in order to validate their freshmen test-battery. The investigator randomly selects a sample of students that originally took the test and examines their GPAs. The unincluded data are MCAR because the probability that GPA or $\underline{x}$ is examined is the same for all freshmen students regardless of their $\underline{x}$ scores or GPAs.

In other cases, the missing data process may depend on the observed data. This type of missingness is referred to as missing at random (MAR). For example, suppose the freshmen students in the above example were admitted on the basis of a cutoff score. As an illustration, suppose $\underline{x}$ consists of $x_1 = SAT_V$, $x_2 = SAT_Q$. Suppose only students for whom the total SAT scores exceeded or are equal to 800 points (i.e., $x_1 + x_2 \geq 800$) were admitted. Scores for the SAT will be observed for everybody who applied to the college but subsequent GPAs will be observed only for those whose total SAT score is 800 or more. The data will be MAR since the probability that the GPAs are missing is a function of $\underline{\mathbf{x}}$, which is fully observed.

Finally, the data can be missing as a function of both the observed and missing data. This case is referred to as data not missing at random (NMAR) or nonignorable non-response. As an example of this type of missingness, suppose some of the students who were accepted for this college took offers from other schools with higher selection standards. These students are likely to be high scorers on unmeasured variables which may be related to their GPAs. Consequently, the probability that the GPA is observed may depend not only on SAT scores but also on the potential value of the GPA.

In this dissertation, I describe, evaluate and compare two different methods for obtaining interval estimates of the squared multiple correlation when data are missing on both the dependent variable (DV) and the predictor variables, and the missing data processes are MCAR and MAR.

First, ad hoc methods for computing the interval estimates for ρ^2 are considered. Three common ad hoc methods for dealing with missing data are listwise deletion, pairwise deletion, and unconditional mean fill-in methods. Listwise deletion excludes all the missing values from analysis. Pairwise deletion uses all cases where the variable(s) under investigation are observed. Unconditional fill-in substitutes each missing datum with the mean of the variable for the observed data.

These procedures can be viewed as methods that produce a pseudo-complete data set. One can then derive confidence interval for ρ^2 based on the sampling distribution of R^2, where R^2 is computed from the pseudo-complete data set. The lower(ρ_l^2) and upper (ρ_u^2) bounds of the 1-α interval are found as solutions to the equations:

$$P(R^2 \geq R_o^2 \mid \rho_l^2) = \alpha/2 \qquad \text{and,}$$

$$P(R^2 \leq R_o^2 \mid \rho_u^2) = \alpha/2 \qquad (1)$$

where;

R_o^2 = the observed sample squared multiple correlation value computed from the complete data,

ρ_l^2, ρ_u^2 = the lower and upper limits of confidence interval (CI),

α = statistical significance level.

The concept and logic of the equations in (1) can be found in Mood, Graybill and Boes (1974). In theory, interval estimates based on ad hoc methods can yield correct intervals when the missing data are MCAR, but not for the MAR case.

The second type of method considered is a Bayesian procedure, which approximates the posterior distribution of ρ^2, given the incomplete data. In theory, given this distribution, one can compute an interval estimate of ρ^2 in terms of a highest density region. Since the computation of the posterior distribution is complex when data are incomplete, one can use an iterative Monte Carlo procedure to obtain interval estimates of ρ^2 in terms of central posterior intervals.

Interval estimates obtained from ad hoc methods and the Bayesian procedure were compared using the following method: A hypothetical multivariate normal distribution having a certain number of predictor variables, variance-covariance matrix and value for ρ^2 was specified. A sample of size (n) was drawn from this distribution and data deleted on both the dependent variable and independent variables using MCAR and MAR processes. The incomplete data set was then analyzed using the two types of methods. The accuracy of the interval estimates was assessed by seeing if they included the value ρ^2. This analysis was repeated by varying the sample size, the number of predictors, the value of the squared multiple correlation ρ^2, and the amount of the missing data.

In section two, a literature survey dealing with the missing data problem and the estimation of ρ^2 from incomplete data sets is presented. Section three describes the sampling methods and the procedures for deleting data. Sections four and five contain the results and conclusions respectively.

Chapter II

LITERATURE SURVEY

2.1 The Problem of Missing Data

There is an abundant literature on the problem of estimating parameters with missing data. However, until approximately two decades ago, much of the literature dealing with the missing data problem was devoted to ad hoc methods (e.g., listwise deletion method). The discourse on these ad hoc methods for dealing with missing data is very common in the literature (Lord & Novick, 1968, Little & Rubin, 1987, Rubin, 1996). Further most of the available statistical packages (e.g., SAS & SPSS) can implement these ad hoc procedures. However, until recently there has been no standard and explicit guide for the applied researcher on how to deal effectively with missing data problems in any practical sense. Following Rubin's (1976) paper describing maximum likelihood estimation, researchers began to consider statistical procedures for dealing with missing data that are based on sound statistical principles and can be objectively evaluated. The extension of maximum likelihood methods to missing data problems has been facilitated by the development of the so called expectation-maximization (EM) algorithm for maximizing the non-standard missing data likelihood functions (Dempster, Laird & Rubin, 1977).

In addition, Bayesian approaches have been generalized to missing data problems. Although the direct calculation of the relevant posterior distribution can often be intractable, the introduction of iterative Monte Carlo (MC) or Monte Carlo Markov Chain (MCMC) procedures, for example, the Gibbs Sampler (Geman & Geman, 1984) have enabled researchers to obtain interval estimates given incomplete data sets. Furthermore, Tanner and Wong (1987) have also described a sample-based technique, data augmentation (DA), that can be used to approximate the actual incomplete posterior distribution of the parameter vector by a mixture of complete data posterior distributions.

In this review, the following topics are discussed:

1) Ad Hoc Methods and their Limitations
2) Maximum likelihood estimation (MLE) and its statistical properties
3) Bayesian Techniques
4) The Gibbs Sampler

2.2 Ad Hoc Methods and their Limitations

Until the early 1970's, efforts to draw statistically valid inference from data with missing values usually involved only ad hoc methods (Rubin, 1976, Kromrey & Hines, 1994). Two ad hoc approaches are commonly used in applied research (deletion and imputation procedures). The first approach involves simply excluding missing data from the statistical analysis. The most common examples of this approach are listwise deletion (complete-case analysis) and available-case analysis (pairwise deletion). In listwise method, all cases with any missing values are excluded from analysis. In deleting cases, there is the danger of increased standard errors due to reduced sample size. Pairwise solutions to missing data problems allow one to use all cases where the variable(s) under investigation are observed. For example, the mean for a variable is estimated using the cases observed on that variable. Statistics involving pairs of measures, (e.g., a correlation) are estimated using cases observed on the pair of variables. It is important to note that pairwise method can produce inconsistent results (e.g., a multiple correlation greater than one) due to changes in sample size from variable to variable according to the missing data patterns.

Imputation procedure involves estimating each missing datum and using the pseudo-complete data set in a standard statistical analysis (Kromrey & Hines, 1994, Little & Rubin, 1987). The estimated missing value may be the mean of the variable for the total set of data (the unconditional mean substitution approach), the value of the variable occurring on a similar data record (the hot-deck approach, Little & Rubin, 1987) or a predicted value based on the relationships among variables in the data (conditional mean approach, Little & Rubin, 1987).

However, none of the ad hoc methods proves entirely adequate for an analysis of an incomplete data set. Ad hoc methods yield biased estimates of parameters from data that are MAR. They may not be appropriate even when one can assume that the data are missing completely at random (MCAR), that is, the missingness is in no way related to the variables of interest (Little & Rubin, 1987). In other words, they may produce biased estimates of parameters of interest even when the missing data structure can be regarded as a sub-sample of the complete data.

For a simple example, suppose one wants to analyze a sample data of size (n=5) in which two data points are MCAR (2, 3, 4, ?, ?). The mean of the observed data is three and the variance is one. Under MCAR, these are consistent estimates. If this data set is now analyzed using the unconditional mean fill-in method, the mean is still the same whereas the variance is now one half. The variance is clearly underestimated using the unconditional fill-in method.

Ad hoc methods will in general yield even worse methods of analysis for incomplete data when the missing data are missing as a function of the observed data, that is (the MAR case). The difficulties associated with ad hoc methods can most easily be explained in terms of a simple analysis. Consider the following hypothetical bivariate data with sample size n = 20. The complete data are presented in Table 1:

Table 1

A complete data set

x	y
2	1
2	2
2	3
2	8
3	3
3	4
3	5
3	10
4	5
4	7
4	10
4	12
5	2
5	4
5	7
5	9
6	3
6	7
6	9
6	10

Table 1 shows that the variance of $\underline{y}$ given $\underline{x}$, Var(y|x), of data is homogeneous across the values of $\underline{x}$, and the regression of $\underline{y}$ on $\underline{x}$ is linear.

Suppose missing data values are created by removing any y-value whose observed x is equal to six. Thus the missing y scores are MAR. One can then compute estimates of the means, standard deviations, regression coefficients, the correlation coefficient and estimated standard errors. Using ad hoc methods, the estimates are then compared with the estimates obtained from the complete data set. The estimated parameter values are presented in table 2.

Table 2

Comparison of parameter estimates

Method	$\hat{\rho}_{xy}$	$\hat{\mu}_x$	$\hat{\mu}_y$	$\hat{\sigma}_x$	$\hat{\sigma}_y$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\frac{\hat{\sigma}_y}{\sqrt{n}}$
Listwise	.24	3.5	3.2	1.4	1.3	2.4	.22	.42
Pairwise	.24	3.5	3.2	1.4	1.3	2.4	.22	.42
*Fill-in	.16	4.0	3.2	1.5	.91	2.8	.10	.20
Complete data	.33	4.0	6.1	1.5	3.3	3.1	.75	.73

*- fill-in with unconditional mean

A careful examination of Table 2 shows that ad hoc methods are not statistically sound techniques for dealing with missing data problem. They certainly yield biased estimates. The ad hoc methods yield correlation coefficients ($\hat{\rho}_{listwise}$=.24, $\hat{\rho}_{pairwise}$= .24 ,and $\hat{\rho}_{fill-in}$ =.16) that are underestimates of the complete-case method ($\hat{\rho}_{xy}$= .33). Similarly, these methods produce underestimates of the mean ($\hat{\mu}_y$), the standard deviation ($\hat{\sigma}_y$), and the slope coefficient ($\hat{\beta}_1$).

One can also use the estimators in Table 2 to compute confidence intervals. For instance, to find an approximate 95% confidence interval (CI) for $\hat{\mu}_y$, one would use the formula:

$$.95(CI) = \hat{\mu}_y \pm t^{*}(\hat{\sigma}_y/\sqrt{n}) \qquad (2.2.1)$$

Where;

$\hat{\mu}_y$ = the estimate for the population mean,

t^{*} = the .95 t-value with n-1 degrees of freedom,

$\hat{\sigma}_y/\sqrt{n}$ = the estimated standard error of the sample mean.

For the complete data case, the estimate will be 6.1 ± 1.73 (.73) or (4.79, 7.31). However, using the pairwise analysis, for example, the interval estimate is quite different, 3.2 ± 1.81(.42) or (2.44, 3.96).

2.3 Maximum Likelihood Estimation

All of the previously described ad hoc methods are not recommended for the analysis of missing data (even when MCAR is assumed) because they often produce biased estimates. Maximum likelihood estimation techniques, as will be seen below, are superior alternatives for the analysis of missing data structures since even under the stronger MAR assumption consistent estimates can be obtained.

For a complete data set, a maximum likelihood estimate of a parameter, θ, is a value for θ that maximizes the likelihood function L (θ|O) given the observed sample data, (O). One chooses as an estimate that value of the parameter that makes the sample data most likely. It is often more convenient to maximize the logarithm of the likelihood function. In a missing data problem, there are both observed data (O) and missing data (M). The likelihood function given an incomplete data set can be constructed using only the observed data set as long as the missing data process can be ignored. According to Little and Rubin (1987), the missing-data mechanism leading to non-response is ignorable for likelihood based inference when the missing data process is MCAR or MAR and the parameters of the sample data and response models are distinct.

Maximum likelihood estimates have many desirable properties (Mood, Graybill & Boes, 1974, Little & Rubin, 1987):

1) They depend on the data only through sufficient statistics when they exist.

2) They are consistent and asymptotically normal, that is as the sample size increases, the standard error tends to zero and the sampling distribution of the MLE will tend toward normality with a mean approaching the parameter value.

3) Following (2), MLE is asymptotically unbiased.

4) Among the class of estimators that are consistent and asymptotically normal, the MLE has the smallest sampling variance, that is, it is efficient. It should be noted, however, that MLE is a large sample technique; that is, all of the desirable properties are assured to hold asymptotically. Consequently, its use in practical situations may be limited due to small sample sizes.

Maximum Likelihood estimation methods may not yield closed formulas for the estimates when data are incomplete even for situations where the missing data process is ignorable. Therefore, iterative algorithms, for example, Expectation-Maximization (EM) are used to maximize the likelihood with respect to the unknown parameters (Dempster, Laird & Rubin, 1977, Little & Rubin, 1987).

As a simple example of the application of MLE, consider the estimation of θ where the binary variable y has the probability distribution:

$$P(y|\theta) = \theta^{y}(1-\theta)^{1-y} \tag{2.3.1}$$

Suppose a random sample of n = 5 yields y scores (1, 1, 0, 1,0) and these data are used to estimate θ.
The likelihood is:

$$P(y|\theta) = \theta^{3}(1-\theta)^{2} \tag{2.3.2}$$

The loglikelihood is:

$$\begin{aligned} L(\theta|y) &= \mathrm{Log}P(y|\theta) \\ &= 3\mathrm{Log}(\theta) + 2\mathrm{Log}(1-\theta) \end{aligned} \tag{2.3.3}$$

The MLE can be obtained by maximizing (2.3.3). This can be done by differentiating $L(\theta|y)$ with respect to θ, setting the result to zero and solving for θ. In this case, $\hat{\theta} = 3/5$.

However, suppose y is observed for only n_c subjects and is missing for $n - n_c$ cases. Suppose, the actual data set is presented as follows:

y	R
1	1
1	1
0	1
?	0
?	0

The observed data now consists of both the observed y scores (1, 1, 0) as well as the response variable vector R (1, 1, 1, 0, 0), where R a is binary vector describing the pattern of observed and missing data (1 = observed, 0 = missing). The modified loglikelihood will be:

$$\begin{aligned} L(\theta,\phi|\text{Observed}) &= L(\theta,\phi|\ O,\ R) \\ &= \log P(y_1 = 1|\theta)P(R_1 = 1|\ y_1 = 1,\phi) + \\ &\quad \log P(y_2 = 1|\theta)P(R_2 = 1|\ y_2 = 1,\phi) + \\ &\quad \log P(y_3 = 0\ |\theta)P(R_3 = 1|\ y_3 = 0,\phi) + \\ &\quad \log P(R_4 = 0\ |\theta,\phi) + \log P(R_5 = 0|\theta,\phi) \end{aligned} \tag{2.3.4}$$

where:

$P(R_j \mid y_j, \phi)$ is the missing data model which provides the probability of R conditional on some unknown parameter ϕ, and the observed or missing value of y, and:

$$P(R_j = 0 \mid \theta, \phi) = P(y_j = 0 \mid \theta) P(R_j = 0 \mid y_j = 0, \phi) + P(y_j = 1 \mid \theta)\, P(R_j = 0 \mid y_j = 1, \phi)$$

= the marginal probability that the j^{th} ($j=4,5$) y score is not observed.

The loglikelihood in (2.3.4) is a function of both θ and ϕ and must be maximized with respect to the two parameters simultaneously. This estimation very much depends on the choice of $P(R_i \mid y_i, \phi)$. Suppose the missing data model is specified as follows:

$$P(R = 1 \mid y = k, \phi) = \phi_1^k \phi_2^{1-k}, \qquad k = 0, 1 \qquad (2.3.5)$$

where ϕ_1, ϕ_2 will represent the unknown parameters for the missing data model. This specification implies the following probabilities given observed or missing y-values:

$$P(R = 1 \mid y = 1) = \phi_1$$
$$P(R = 1 \mid y = 0) = \phi_2$$

One can write the loglikelihood of the distribution for our sample as follows:

$$
\begin{aligned}
L(\theta,\phi| \text{ Observed data}) = {} & \log(P(y=1|\theta)\phi_1) + \\
& \text{Log}(P(y=1|\theta)\phi_1) + \\
& \text{Log}(P(y=0|\theta)\phi_2) + \\
& 2\log[(1-\theta)(1-\phi_2) + \theta(1-\phi_1)]
\end{aligned}
\tag{2.3.6}
$$

Clearly, the loglikelihood equation in (2.3.6) depends on θ and ϕ jointly. The implication is that one needs to specify the missing data process underlying the missing data and simultaneously estimate θ and ϕ.

Suppose that the missing data are MCAR, that is, $\phi_1 = \phi_2 = \phi$ and that the parameters θ and ϕ are distinct. Then the likelihood can be broken into two terms, one a function of θ and the second a function of ϕ:

$$
\begin{aligned}
L(\theta,\phi) &= \log(\theta\phi) + \log(\theta\phi) + \log((1-\theta)\phi) + \\
&\quad 2\log\{(1-\theta)(1-\phi) + \theta(1-\phi)\} \\
&= 2\log(\theta\phi) + \log(1-\theta)\phi + 2\log(1-\phi) \\
&= [2\log\theta + \log(1-\theta)] + [3\log\phi + 2\log(1-\phi)] \\
&= f(\theta) + g(\phi)
\end{aligned}
\tag{2.3.7}
$$

One can thus ignore the missing data process and obtain the maximum likelihood estimate of θ simply by maximizing $f(\theta)$, that is, treat the observed data as if it were a complete data set.

On the other hand, if the data were not MCAR, or θ and $\phi = (\phi_1,\phi_2)$ are not distinct, one must maximize the entire likelihood function with respect to θ and (ϕ_1,ϕ_2). In other words, if it is assumed that the two missing $\underline{y}$ data values were not missing at random (NMAR), one must specify the missing data model. Under this condition, it will no longer be possible to separate equation (2.3.6) into two distinct functions of θ and ϕ because the last expression in equation (2.3.6) depends on (ϕ_1, ϕ_2) in a joint manner. Clearly one can no longer estimate θ in terms of the observed $\underline{y}$ scores only, but must maximize the entire likelihood with respect to both θ and (ϕ_1, ϕ_2). More importantly, it becomes crucial and imperative to specify the missing data model because the parameter of the missing data model ϕ must be estimated.

2.4 Bayesian Techniques

It has been demonstrated earlier that ad-hoc methods do not handle missing data problems satisfactorily even when missing data points are MCAR. It was also pointed out that although maximum likelihood estimation techniques may produce consistent estimates of the parameters of an incomplete data set, they may produce biased estimates of parameters in missing data problems with small sample sizes. This is because MLE desirable properties hold asymptotically (i.e., with large sample sizes).

A Bayesian estimation approach provides a plausible and practical alternative in situations (e. g., small sample cases) where the likelihood is not closely approximated by the normal likelihood (Tanner & Wong, 1987). This is because Bayesian inference does not depend on the asymptotic assumptions of the MLE approach. Generally, in Bayesian techniques, investigators state prior beliefs about the parameter(s) of interest and then modify these prior beliefs in the light of the observed data in order to arrive at posterior beliefs (Lee, 1989, Samaniego & Neath, 1996). More specifically, Bayesian analysis allows one to derive (albeit tediously in some cases) the entire posterior distribution, $P''(\theta|O)$, of the parameter(s) θ using the prior distribution, $P'(\theta)$, and the likelihood function, $P(O|\theta)$, where O is the observed data . One can then compute interval estimates for the parameters or obtain a point estimate of some functions of parameters of the posterior distribution such as the mean or median.

Bayesian methods are suitable for many missing data problems because additional prior information about the data specified in the prior distribution are incorporated into the analysis for estimating the parameters of the posterior distribution thereby making the estimates more precise (Geman & Geman, 1984, Tanner & Wong, 1987). In other words, the prior can supplement the incomplete data set.

The next subsection describes the Bayesian estimation method in the complete data case. Subsequent subsections then explain how the Bayesian technique generalizes to the missing data problems.

2.4.1 The Complete Data Case

Suppose we observed a complete set of data, (O), and wish to execute a Bayesian analysis to estimate a parameter, θ, from our data. First, we need to quantify our prior knowledge concerning θ. In other words, we want to formulate a prior distribution, $P'(\theta)$, representing our knowledge of θ before the data are considered. Using the data, we are able to establish the likelihood function, $P(O|\theta)$. In order to estimate θ given the observed sample data and the prior , we compute the posterior distribution, $P''(\theta)$, using Bayes formula:

$$P''(\theta|O) = \frac{P'(\theta)P(O|\theta)}{P(O)} \propto P'(\theta)P(O|\theta) \qquad (2.4.1.1)$$

where:

$P(O)$ = the marginal distribution of the data,

$P(O|\theta)$ = the data likelihood ,

$P'(\theta)$ = the prior distribution of θ, and

$P''(\theta|O)$ = the posterior distribution of θ.

The posterior distribution summarizes all of the information we have about the parameter θ given the prior and the observed data, and thus we can use the posterior to estimate θ. One way to estimate the parameter θ is to compute a point estimate of the parameter of interest by calculating the posterior mean, median, or mode.

Sometimes we may be interested in the probability that the parameter lies in a particular interval. In this context, we can obtain an interval estimate for θ from the posterior distribution. The highest density region (HDR) usually represents the Bayesian confidence interval of choice. The HDR is that interval which is the shortest possible interval for a given probability level. The density at any point inside the HDR is greater than the density for any point outside the HDR (Box & Tiao, 1973, Lee, 1989). We can also calculate central posterior interval estimates for ρ^2. A 1- α central posterior interval represents the interval between $\alpha/2$ and 1 - $\alpha/2$ percentiles of the posterior distribution.

It is clear from (2.4.1.1) that the posterior distribution is a function of the prior and the data likelihood. Therefore, the prior and the posterior will be nearly identical if the prior distribution strongly dominates so that the data do not add any relevant information (Box & Tiao, 1973, Lee, 1989). This is a rare situation since there would be no need to conduct a study of a parameter if the prior knowledge is so convincing that further studies would not contribute new relevant information about the parameter.

More often, an investigator's prior knowledge of the parameter may be very vague (Lee, 1989). One of the ways to represent this case is to assume a uniform prior distribution (non-informative prior) for some function of the unknown parameter. For example, we can specify a univariate normal prior distribution for the unknown mean θ of a normal density with known variance σ^2 as follows:

$$P'(\theta) = 1 \tag{2.4.1.2}$$

Using this prior, the dominant feature of the posterior is the data likelihood. More specifically the posterior distribution of θ will depend only on the sample mean.

As a simple example of a Bayesian analysis, suppose a school district is interested in the proportion of students with learning disability (L_D) who passed a state mandated examination before entry into the seventh grade for one particular year. The problem is to estimate θ, the proportion of the L_D who passed. A random sample (n = 50) of the L_D population is drawn in which r = 5 passed. Reasonable prior distribution can be expressed as:

$$P' (\theta) \propto \theta^{a-1} (1 - \theta)^{b-1} = \text{Beta}(a,b) \qquad (2.4.1.3)$$

where:

a,b = the parameters of a beta distribution (a, b > 0).

One useful feature of the Beta distribution is that by manipulating a and b one can get different shapes of the distribution, or even approximate other known distributions:

a) When a = b = 1, the distribution is uniform
b) If (a,b) > 1, and a = b, the distribution is symmetric about .5.
c) If (a,b) >1 and a > b, the distribution is skewed to the left.
d) If (a,b) >1 and a < b, the distribution is skewed to the right. This condition will be suitable for the specification of our prior distribution because learning disabled students often do not do well in examinations. As an example, we will let a = 1 and b = 10.

The data likelihood is:

$$P(\text{data}|\theta) \propto \theta^5 (1-\theta)^{45}$$

Following (2.4.1.1), the posterior distribution is computed by the following relationship:

$$\begin{aligned} P''(\theta|O) &\propto P'(\theta)\, P(\text{data}|\theta) \\ &\propto \theta^{a+5-1}\,(1-\theta)^{b+45-1} \\ &\propto \text{Beta}(a+5,\ b+45) \end{aligned} \tag{2.4.1.4}$$

To estimate the parameter θ of interest we can use the posterior mean of $P''(\theta|O)$. The posterior mean can be found using the following expression:

$$\begin{aligned} E\{P''(\theta|O)\} &= \frac{a+5}{a+5+b+45} \\ &= \frac{1+5}{1+5+10+45} \\ &= \frac{6}{61} \end{aligned}$$

For many standard estimation problems with no missing data, the form of the resulting posterior distribution given the prior and the likelihood is known

(Lee, 1989). However, serious complications may arise when one wishes to compute the posterior distribution for incomplete data.

2.4.2 The incomplete data case

The Bayesian approach can be modified when only an incomplete data set is available. In addition to the observed sample data(O), there are missing data values (M) which are unobservable. Given that the missing data process is ignorable, and the prior distribution of θ and ϕ can be expressed as $P'(\theta)\ P'(\phi)$, (Rubin, 1976) the posterior distribution of θ can be expressed as:

$$P''(\theta|O) = P'(\theta)\ L(\theta|O) \qquad (2.4.2.1)$$

It is important to note that due to the non-standard form for $L(\theta|O)$, analytical expressions for various marginal posterior distributions may be difficult to obtain in the incomplete case. Thus interval estimates of particular individual parameters may be difficult to obtain.

For example, suppose we consider the following bivariate normal variable vector:

$$\underline{y} = (y_1, y_2) \sim N(\underline{\mu}, \Sigma)$$

where:

$$\underline{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$$

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}$$

Suppose we now observe the following for a sample of data:

Sample size	y_1	y_2
n_1	O	O
n_2	O	M

where;

n_1 = the fully observed cases,

n_2 = Cases with missing values on y_2,

O = observed data, and

M = missing data.

Suppose it assumed that y_2 is MAR, and that $\theta = \{\underline{\mu}, \Sigma\}$ and the parameters underlying this missing data process (ϕ) are a priori independent. The posterior distribution for our parameter θ given the observed data can be written as:

$$P''(\theta|O) = P'(\underline{\mu}, \Sigma) P(y_1, y_2 | \underline{\mu}, \Sigma, n_1) P(y_1 | \mu_1, \sigma_{11}, n_2) \tag{2.4.2.2}$$

If the third factor is not present, the posterior is of a standard form, and the marginal posteriors, that is, $P''(\underline{\mu}|O)$, $P''(\Sigma|O)$ are well known (De Groot, 1970). However, the $P''(y_1|\mu_1, \sigma_{11}, n_2)$ term creates a non-standard likelihood, and the above marginals are difficult to directly obtain.

Analytical solutions for the posterior distributions may be obtainable for simple cases of incomplete data. For example, Gross and Torres-Quevedo (1995) considered the problem of estimating the bivariate correlation coefficient with data missing on one of the variables. This method can be further generalized to bivariate cases where data are missing on both variables. However, for the more complex problem where the multiple correlation is estimated from a data set with missing values on all the variables one can no longer obtain closed form expressions for marginal posterior distribution of interest, that is $P''(\rho^2 \mid O)$. At this point, it becomes crucial and practical to implement Monte Carlo (MC) methods (e.g., the Gibbs Sampler; see Geman & Geman, 1984) to approximate the marginal posterior distributions of interest.

In the subsection that follows, alternative MC procedures that can be implemented to obtain the desired marginal posterior distributions when analytic methods are not feasible are discussed.

2.4.3 Alternative procedures to analytic techniques

The technical difficulties stemming from the computation of the marginal posterior densities used in Bayesian inference have long hindered the wider application of the Bayesian approach to real data until recently (e.g., see Gelfand et al., 1990a, 1990b). In a Bayesian analysis of an incomplete data set, the calculation of a marginal posterior distribution may become intractable due to missing values (e.g., analysis of multivariate data sets with missing values in many variables). In this context, the numerical and analytic methods will often prove ineffective procedures. The Gibbs sampler (Geman &Geman,

1984), data-augmentation (Tanner & Wong, 1987), and importance-sampling (Rubin, 1987, 1988) are three alternative sampling-based algorithms (or Monte Carlo procedures) for the calculation of numerical estimates of marginal probability distribution. Their conceptual simplicity and ease of implementation make them a good choice for calculating marginal distributions in incomplete data cases. In recent years, extensive use of MC procedures in applied research has been noted both for complete and incomplete data structures. Examples of these Bayesian MC research efforts dealing with different problems include:

(a) Bayesian estimation for multivariate linear regression with fully observed predictor variables and possible ignorable missing values from outcome variables (Liu, 1996).

(b) Examination of the covariance structure of the Gibbs sampler with applications to the comparisons of estimators and augmentation schemes (Liu, Wong & Kong, 1994).

(c) The application of Gibbs sampling, data-augmentation algorithm, and the Rubin importance-sampling algorithm to the calculation of numerical estimates of marginal probability distributions (Gelfand & Smith, 1990b, Chib, 1995). These sampling-based approaches are useful in the calculation of marginal posterior densities within a Bayesian inference framework. Consequently, Gelfand et al. (1990a) illustrated how the Gibbs sampler algorithm can be used to obtain numerical Bayesian inferences with a range of normal data models, including variance components, unordered and ordered means, hierarchical growth curves, and missing data models with complex distribution structures.

(d) Using a Bayesian methodology to analyze time evolution of earthquake activity (Peruggia & Santner, 1996).

(e) The proposal of an importance sampling technique for converting the output of the Gibbs sampler to a sample from the exact posterior (Ritter & Tanner, 1992).

(f) A semiparametric Bayesian analysis of multiple event time data (e.g., multiple attacks of cardiac arrests) using the Gibbs sampler to sample from the joint posterior distribution of the unknown parameters (Sinha, 1993).

(g) Comparing parameters using Bayes and empirical Bayes procedures (Sobel, 1993).

(h) Comparing Bayesian credible interval for the product of normal means using the reference prior to the standard estimate of the confidence interval (Sun & Ye 1995), implementing generalized linear random effects model in a Bayesian framework via the Gibbs sampler (Zeger & Karim, 1991), and examining the Gibbs sampler convergence criteria (Zelliner & Min, 1995).

(i) Proposal of probabilistic procedure via Gibbs sampling to select good predictors in multiple regression models (George & McCulloch, 1993) and a general algorithm for an approximate Bayesian conditional inference via Gibbs sampler that is more accurate than standard asymptotic approximations (Kolassa & Tanner, 1994) .

(j) Buckle (1995) used Bayesian Gibbs sampling technique to demonstrate that Bayesian inferences can be made for the class of distributions (stable distributions) whose location-scale pairs can be replaced by their means and variances when the distributions attain normality.

(k) The problem of the effect of improper priors on Gibbs sampling in hierarchical linear mixed models (Hobert & Casella,1996).

Our review in the next subsection focuses entirely on the Gibbs sampler, the iterative MC method chosen for this dissertation project.

2.5 The Gibbs Sampler

The nature of the Gibbs sampler

Gibbs sampler is an iterative sampling-based algorithm intended for cases where direct calculation of a marginal distribution is typically not feasible. Many of the examples on the applications of the Gibbs sampler cited in the last section demonstrate the versatility of the Gibbs sampler procedure in practice. The Gibbs iterative algorithm can handle Bayesian missing data problems by approximating the incomplete data marginal posterior distributions (Liu, Wong, & Kong,1994).

The Gibbs procedure operates in the following way:

1) Assume that there exists the joint probability distribution of k variables $P(\underline{x}) = P(X_1, X_2, \ldots, X_k)$. We would like to approximate the marginals $P(X_1)$, $P(X_2), \ldots, P(X_k)$. Suppose it is difficult to either obtain the marginals analytically or to sample directly from the joint distribution.

2) Formulate the univariate conditional distributions

$$P(X_s|X_{r,}\ )$$

where:

s = one particular variable of interest

r = k-1 remaining variables

Then given an arbitrary starting values

$X_1^{(0)} = x_1^{(0)},\ X_2^{(0)} = x_2^{(0)}, \ldots, X_k^{(0)} = x_k^{(0)}$, we can sample

$X_1^{(1)}$ from $P(X_1 | X_2^{(0)} = x_2^{(0)}, \ldots, X_k^{(0)} = x_k^{(0)})$, sample

$X_2^{(1)}$ from $P(X_2 | X_1^{1} = x_1^{(1)}, X_3^{(0)} = x_3^{(0)}, \ldots, X_k^{(0)} = x_k^{(0)})$

and so on up to

$$X_k^{(1)} \text{ from } P(X_k | X_1^{(1)} = x_1^{(1)}, \ldots, X_{k-1}^{(1)} = x_{k-1}^{(1)}) \qquad (2.5.1)$$

to complete one iteration of the algorithm.

After t such iterations, the process yields a Gibbs sequence of random variates $X_1^{(t)}, \ldots, X_1^{(t)}$. It has been shown that for any starting value the distribution $\underline{x}^{(t)}$ approaches the true joint distribution $P(\underline{x})$ as t tends to infinity (Geman & Geman, 1984, Casella & George, 1992). Further the distribution of the s^th^ variable $P(x_s^{(t)})$ approaches the marginal, $P(x_s)$ as t increases.

The Gibbs sampler can be readily applied to the problem of estimating the squared multiple correlation ρ^2 from a multivariate data set with missing data (O,M), and unknown parameters $\underline{\mu}$, Σ where:

O = the observed sample data,

M = the missing data (unobserved data),

$\underline{\mu}$= population mean vector for the set of variables used in the analysis of the data set,

Σ = population variance-covariance matrix for the set of variables used in the analysis of the data set.

The squared multiple correlation is a function of the variance-covariance matrix { i.e., $\rho^2 = f(\Sigma)$, Winer, 1971, p.107}. Thus an estimate of ρ^2 can be computed when the marginal distribution of Σ is approximated using the Gibbs sampler.

Consider the posterior distribution of Σ given the observed data:

$$\begin{aligned} P''(\Sigma|O) &= \int_{\underline{\mu}} P''(\underline{\mu},\Sigma|O) \\ &= \int_{\underline{\mu}}\int_{M} P''(\underline{\mu},\Sigma|O,M)\ P(M|O)dMd\underline{\mu} \\ &= \int_{\underline{\mu}}\int_{M} P''(\underline{\mu},\Sigma,M|O)dMd\underline{\mu} \end{aligned} \tag{2.5.2}$$

Where:

$P''(\underline{\mu},\Sigma|O,M)$ = the posterior distribution of $\underline{\mu}$, Σ conditional on the missing values M, given the observed data, O

$P(M|O)$ = the predictive distribution of the missing data given the observed data, O

$P(\underline{\mu},\Sigma,M|O)$ = the joint distribution of $\underline{\mu}$, Σ, and M given the observed data, O

The expression in (2.5.2) shows that the posterior distribution of Σ, is a marginal distribution obtained from the joint distribution of $\underline{\mu}$, Σ, M. Thus this distribution can be viewed in the context of a Gibbs sampling problem involving the joint distribution of three unknowns $\underline{\mu}$, Σ, M.

Thus the following full conditionals are considered (See appendix for details):

$P(M|\underline{\mu},\Sigma,O)$ = the distribution of the missing data conditional on current parameter values and the observed data,

$P''(\underline{\mu}|\Sigma,M,O)$ = the posterior distribution of $\underline{\mu}$ conditional on the current values for Σ and M, and the O,

$P''(\Sigma|\underline{\mu},M,O)$ = the posterior distribution of Σ conditional the current values for $\underline{\mu}$ and M, and the O.

Using (2.5.1) it is straightforward to sample repeatedly from all three conditional distributions to generate a random sequence of triplets:

$$M^{(0)},\ \underline{\mu}^{(0)},\ \Sigma^{(0)},\ M^{(1)},\ \underline{\mu}^{(1)},\ \Sigma^{(1)},\ \ldots,$$
$$M^{(r)},\ \mu^{(r)},\ \Sigma^{(r)},\ \ldots,\ M^{(p)},\ \mu^{(p)},\ \Sigma^{(p)}.$$

The first r iterations represent the burn in and are usually ignored. The remaining r+1, . . ., p iterations will be used to calculate the marginal distribution of Σ, and thus the estimate for ρ^2.

As an illustration of the Gibbs algorithm, consider the following hypothetical multivariate sample data with missing values:

Subject	y	x1	x2
1	O	O	O
2	O	O	O
3	O	O	O
4	M	O	O
5	O	M	M
6	O	O	O

Step (A) Two missing data patterns are identified

Step (B) Compute initial values for $\underline{\mu}$ and Σ designated as $\underline{\mu}_0$ and Σ_0 from fully observed data, where

$\underline{\mu}$ = a three by one vector of means, and

Σ = a three by three variance-covariance matrix

Step (C) The iterative procedures previously described are used to:

1) Sample a value for y_4 from $P(y|x_1,x_2, \underline{\mu}_0,\Sigma_0)$ for subject 4

2) Sample values for x_{15} and x_{25} from $P(x_1,x_2|y,\underline{\mu}_0,\Sigma_0)$ for subject 5

We call the filled in values M_1. Now we have a pseudo complete data set

3) Sample Σ_1 from $P'(\Sigma|\underline{\mu}_0,O,M_1)$

4) Sample $\underline{\mu}_1$ from $P'(\underline{\mu}|\Sigma_1,O,M_1)$. Now we have M_1, Σ_1, $\underline{\mu}_1$.

5) Compute ρ^2 from Σ_1

6) Repeat steps 1-5, say, b + n times, where b represents a set of burn-in values. We now have b+n values of ρ^2

7) Ignore the first b burn-in values of ρ^2. Plot a frequency distribution for the last n values of ρ^2.

Obtain an interval estimate of ρ^2 from this empirical distribution.

Chapter III

METHOD

In this dissertation the accuracy of interval estimates for the population ρ^2 values computed from data with missing values on both the dependent and q > 1 independent variables was assessed using a Bayesian technique based on the Gibbs sampler and three ad hoc methods. Listwise deletion, pairwise deletion and unconditional fill-in were the three ad hoc methods used.

More specifically, interval estimates from the two types of techniques were compared using the following method: A multivariate normal distribution for q > 1 predictor variables and one dependent variable, a variance-covariance matrix, and a value for ρ^2 was specified. A random sample of size n was drawn from this distribution and data deleted on both the dependent variable and independent variables using a MCAR and MAR processes. The incomplete data set was analyzed using the two types of methods, and a sample interval estimate of ρ^2 for each type of method was computed. This analysis was repeated by varying the sample size, the number of predictors, the value of the squared multiple correlation ρ^2, and the amount of the missing data. 95% confidence intervals (CI) for ρ^2 estimates from the three ad hoc methods were constructed and 95% central posterior intervals were calculated for the Bayesian approach based on the Gibbs sampler. The details of specification, sampling, deletion and analyses are presented below.

The parameters of the hypothetical multivariate normal populations were specified as follows:

1) All variables have zero means, variances of 1.00, and inter-correlations among predictors of .50.
2) The correlation of y with each x was determined such that the squared multiple correlation varied with values of .10, .25, .50.
3) The number of predictors varied with values 2, 5, 7.
4) The sample sizes also varied with values 50, 100, 250.

Given complete samples from the 27 hypothetical populations, described above, data were deleted in four ways:

a) 10% of the data were deleted and b) 40% of the data were deleted such that the missing data are either MCAR or MAR. The methods for deleting data are described in appendix B.

After the deletion process there were 108 incomplete data sets. The four methods were then applied to the incomplete data sets, thus computing the interval estimates for ρ^2. Seeing if these computed interval estimates ($\hat{\rho}^2$) include the ρ^2 values assesses the accuracy of the methods implemented. The next two sections explain the specific steps needed to calculate the interval estimates for ρ^2 for the ad hoc methods and the Gibbs sampler.

3.1 Ad hoc methods

If the available multivariate sample contained no missing data, one can use the following procedure to obtain a 95% CI for ρ^2. Let R_o^2 be the observed sample squared multiple correlation, we obtain the upper (ρ_u^2) and lower (ρ_l^2) confidence interval bounds by solving the following equations:

$$P(R^2 \leq R_o^2 \mid \rho_u^2) = \int_{o}^{R_o^2} f(R^2 \mid \rho_u^2, n, q) = .025$$

$$P(R^2 \geq R_o^2 \mid \rho_l^2) = \int_{R_o^2}^{1} f(R^2 \mid \rho_l^2, n, q) = .025$$

(3.1.1)

where;

$f(R^2|\rho^2,n,q)$ = the distribution of the sample squared multiple correlation,

R_o^2 = the observed sample squared multiple correlation,

ρ^2 = the population multiple correlation,

ρ_u^2, ρ_l^2 = upper and lower limits for the population squared multiple correlation,

n = the number of cases, and

q = the number of predictor variables,

The concept, logic and use of the expressions in (3.1.1) can be found in Mood, Graybill and Boes (1974, p. 387-397).

For the three ad hoc procedures investigated, the R_o^2 values were obtained from pseudo complete data sets. The first ad hoc method creates a pseudo complete data set using listwise deletion. In other words, all the missing cases are dropped, thus using fewer cases for the analysis. The observed sample squared multiple correlation was computed on the complete cases of the sample. Thus the sample size was equal to the number of complete cases. The second ad hoc method, fill-in, creates a pseudo complete data set by using the mean of the complete cases for the sample to substitute for each missing data point. In effect, it was assumed that there were no missing data points. The sample size was equal to the total number of cases in the data set. The third is pairwise deletion. Pairwise solutions to missing data problems allow one to use all cases where the variable(s) under investigation are observed. For example, the mean for a variable is estimated using the cases observed on the variable. However, statistics involving pairs of measures, (e.g., a multiple correlation) are estimated using cases observed on the pair of variables. The observed sample squared multiple correlation was computed on the modal number of cases observed on the pairs of variables under investigation.

In the present dissertation, solutions for ρ_u^2, ρ_l^2 were found using a FORTRAN program. This program allows one to compute an interval estimate for ρ^2 given R_o^2, level of confidence $(1 - \alpha)$, number of predictor variables and initial 'guesses' for ρ_u^2 and ρ_l^2 (see appendix C for the actual source code).

3.2 The Gibbs Method

As previously stated, the Gibbs sampler requires iteratively sampling from the following full conditional distributions, all of which have standard forms (De Groot, 1970):

(i)$P(M|\underline{\mu},\Sigma,O)$ = the normal distribution of the missing data conditional on current parameter values for $\underline{\mu}$, Σ, and the observed data.

(ii)$P''(\Sigma|\underline{\mu},M,O)$ = the posterior distribution of Σ conditional the current values for $\underline{\mu}$, M, and the observed data.

(iii) $P''(\underline{\mu}|\Sigma,M,O)$ = the posterior distribution of $\underline{\mu}$ conditional on the current values for Σ,M, and the the observed data.

In the present dissertation, the Gibbs sampler was implemented using the BUGS computer program. This program yields interval estimates of ρ^2 in terms of central posterior intervals.

The following structure chart, Figure 1, summarizes the entire method section:

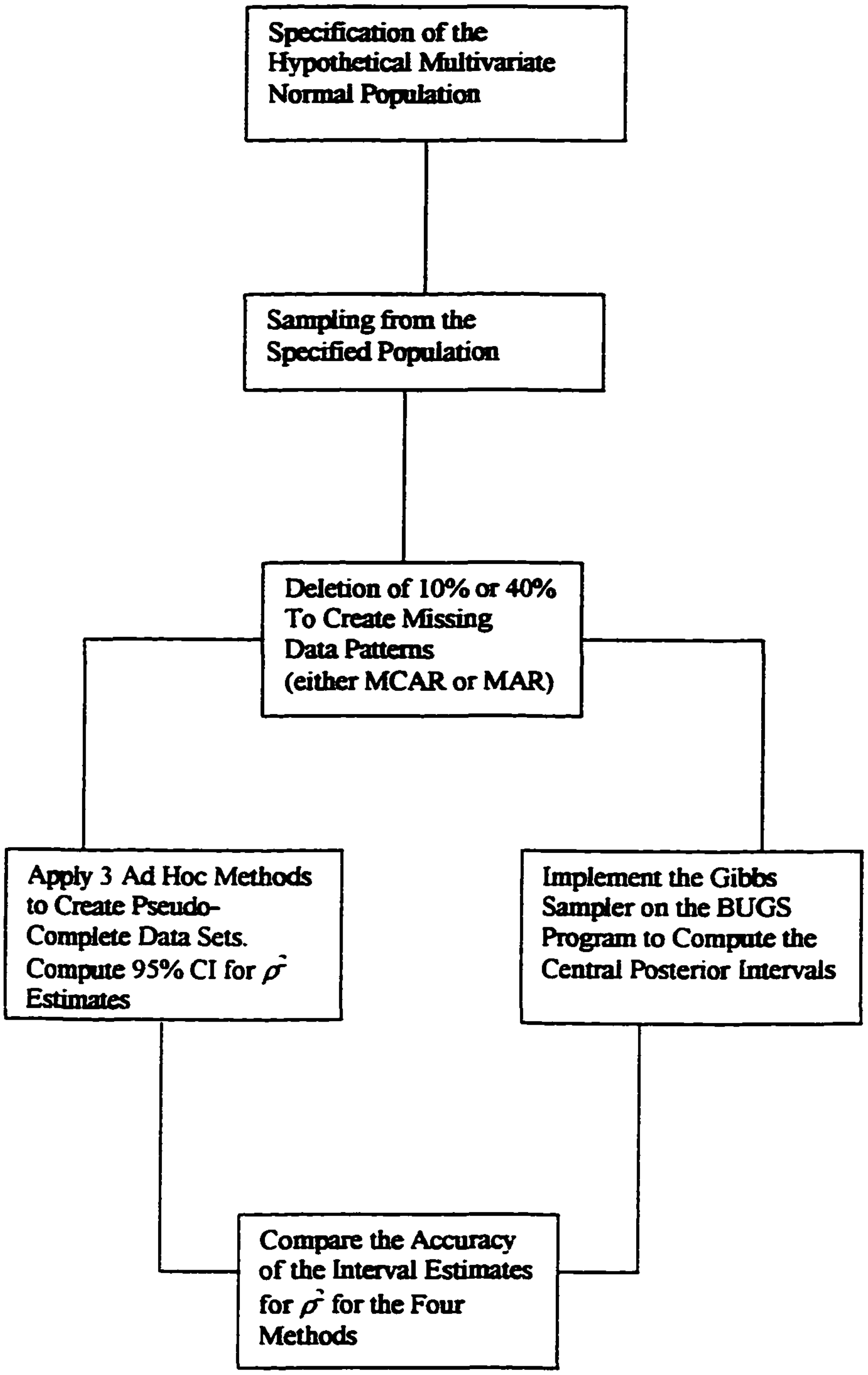


Figure 1: Structure chart summarizing the entire method section

CHAPTER IV

RESULTS

In this section, the results of the analyses of the different methods implemented in this dissertation are presented. The key-motivating factor here is to compare the results from ad hoc methods and the Bayesian method of analysis in estimating the population squared multiple correlation values. Another focus is to compare the ad hoc methods for estimating ρ^2 values. The bases of the comparisons include the effects of sample sizes, the number predictors used in the analyses, the population squared multiple correlation value and the proportion of complete data for the various samples drawn and used for our analyses. The basic results are presented in 12 tables (3A – 3C, 4A – 4C, 5A – 5C, 6A – 6C) and then summarized in additional tables (Tables 7, 8, 9, & 10).

In order to acquaint the reader with the contents of the basic tables, Table 3A is explained. Table 3A presents thirty-six results of ad hoc and Bayesian interval estimates for ρ^2 values when the proportion of complete data (P_c) is 90%, the number of predictors is two and the missing data are MCAR. The sample data sets used for estimating ρ^2 were varied as a function sample size (50,100, & 250) and population squared correlation ρ^2 value (.1, .25, .5). More specifically, within each sample size, three different values of ρ^2 were used. Thus Table3A contains nine data sets which were used to obtain interval estimates of ρ^2 using three ad hoc methods and the Bayesian method, yielding thirty-six interval estimates for ρ^2. The differences in the setup of Table 3A and the remaining tables can be found in at least one of the following: the number of predictor variables, the proportion of complete data (P_c), and missing data type.

The following conclusions can be drawn from the results in Table 3A: all but one of the intervals in this table cover the population squared multiple correlation ρ^2 values. However, a closer look at these results in Table 3A shows that in most cases (n=100,250) the fill-in method has the shortest interval. For example, (with n=100, ρ^2 = .1) listwise deletion has an interval width of 0.28, pairwise has a width of 0.28, fill-in 0.27, and Bayesian 0.28. The inflated sample size for fill-in method may explain the reason for this narrow interval.

Another clear result seen in Table 3A is that as the sample size increases, given the same conditions for Table 3A, the width of the interval estimates tends to shrink. In other words, the intervals tend to get narrower as the sample sizes increase thus increasing the accuracy of interval estimates for ρ^2 values within the same method. This observation is true across all the four methods used in the dissertation.

Rather than considering the basic results, one can consider summary tables by collapsing over the factors.

In Table 7 the proportion of times each method covered the true ρ^2 values and (average interval width) as a function of method and sample size are considered. Each proportion is based on 36 studies collapsed over missing data type, the proportion of complete data (P_c), the number of predictor variables, and the population squared multiple correlation value (ρ^2).

Table 8 shows coverage probabilities as a function of method and missing data type (MCAR or MAR). Each proportion is based on 54 studies collapsed over sample size, the proportion of complete data (P_c), the number of predictor variables, and the population squared multiple correlation value (ρ^2).

Table 9 considers for each estimation method the effect that the proportion of complete data (P_c) has on the empirical coverage probabilities. Each proportion is based on 54 studies collapsed over sample size, missing data type (MCAR or MAR), the number of predictor variables, and the population squared multiple correlation value (ρ^2).

Table 10 considers for each estimation method the effect that the value of the squared multiple correlation (ρ^2) has on the empirical coverage probabilities. Each proportion is based on 36 studies collapsed over sample size, missing data type (MCAR or MAR), the number of predictor variables, and the proportion of missing data (P_c).

The following conclusions can be drawn from tables 7, 8, 9 and 10:

1) In Table 7, listwise deletion covered the parameter value almost all the time whereas pairwise deletion and the fill-in method were less accurate in covering the parameter value for all the sample sizes considered. Listwise deletion is clearly the best Ad Hoc method and fill-in is the worst.

However, a close look at Table 7 shows that for sample sizes (n= 50, 100) the Bayesian method did worse than all ad hoc methods. For the large sample size (n=250), the Bayesian method performed better than pairwise and fill-in methods and covered the true ρ^2 value 100% of time just as listwise deletion. Further, the average interval width decreased as the sample size increased for all the methods considered.

2) In Table 8, ad hoc methods produce worse results when data are MAR than when they are MCAR. Unlike ad hoc methods, the Bayesian method on average worked better under MAR than under MCAR. This result is rather surprising and it is not the theoretical expectation.

Table 8 shows that listwise and pairwise methods include the true value of ρ^2 all the time, fill-in and the Bayesian intervals cover the parameter value 98% and 72% of the time respectively when data are MCAR. When data are MAR listwise, pairwise, fill-in, and the Bayesian intervals cover the parameter value 98%, 89%, 74%, and 89% of the time respectively.

3) Table 9 shows that for the ad hoc methods, as the proportion of missing data increases the interval estimates for ρ^2 become less accurate.

In table 9 all ad hoc methods cover the true ρ^2 value all the time and the Bayesian method covers the true ρ^2 value only 80% of the time when P_{miss} = .10. However, when P_{miss} = .40, listwise, pairwise, fill-in, and the Bayesian cover the parameter value 98%, 89%, 72%, and 81% of the time respectively.

4) Table 10 indicates that as the population squared

multiple correlation value increases the proportion of times the Bayesian method covers the parameter value increases. Unlike the Bayesian method, the proportion of times pairwise and fill-in covered the true ρ^2 value decreases as the population squared multiple correlation value increases. Listwise deletion had the best result among all the methods.

5) In general, it is surprising at first glance that the Bayesian did not dominate the ad hoc procedures across the conditions. A major reason for this result lies in the form of the prior distributions. This will be the discussed in the next section.

Table 3A

Ad Hoc and Bayesian Interval Estimates for ρ^2 (Missing Data = MCAR, Number of predictors = 2, Proportion of complete Data = .90) as a Function of Sample Size (n = 50, 100, 250) and Population Squared correlation ρ^2 value (.1, .25, .5).

	Method			
ρ^2	Listwise	Pairwise	Fill-in	Bayesian
		n = 50		
.1	.07, .51	.05, .46	.04, .43	.12, .55
.25	.16, .60	.12, .55	.12, .54	.21, .63
.5	.39, .72	.33, .72	.34, .71	.43, .78
		n = 100		
.1	.06, .34	.06, .34	.06, .33	.08, .36
.25	.19, .50	.19, .49	.19, .49	.21, .51
.5	.42, .69	.42, .68	.40, .67	.44, .69
		n = 250		
.1	.04, .20	.04, .19	.04, .18	.05, .21
.25	.17, .37	.17, .36	.16, .34	.18, .38
.5	.42, .60	.41, .59	.39, .57	.43, .61

Table 3B

Ad Hoc and Bayesian Interval Estimates for ρ^2 (Missing Data = MCAR. Number of predictors = 5. Proportion of complete Data = .90) as a Function of Sample Size (n = 50. 100. 250) and Population Squared correlation ρ^2 value (.1, .25, .5).

	Method			
ρ^2	Listwise	Pairwise	Fill-in	Bayesian
		n = 50		
.1	.03, .48	.04, .48	.04, .46	.17, .57
.25	.11, .58	.13, .57	.12, .56	.26, .65
.5	.33, .74	.35, .73	.33, .72	.46, .78
		n = 100		
.1	.03, .32	.05, .33	.04, .32	.10, .38
.25	.15, .48	.17, .48	.16, .47	.23, .53
.5	.40, .68	.40, .67	.39, .66	.46, .71
		n = 250		
.1	.05, .21	.06, .21	.05, .21	.08, .25
.25	.19, .39	.20, .39	.19, .38	.22, .41
.5	.45, .61	.46, .60	.45, .60	.47, .64

Table 3C

Ad Hoc and Bayesian Interval Estimates for ρ^2 (Missing Data = MCAR, Number of predictors = 7, Proportion of complete Data = .90) as a Function of Sample Size (n = 50, 100, 250) and Population Squared correlation ρ^2 value (.1, .25, .5).

	Method			
ρ^2	Listwise	Pairwise	Fill-in	Bayesian
		n = 50		
.1	.04, .51	.02, .47	.01, .46	.25, .64
.25	.13, .61	.13, .59	.12, .57	.35, .71
.5	.38, .77	.39, .76	.37, .75	.55, .82
		n = 100		
.1	.07, .40	.08, .39	.08, .38	.19, .48
.25	.19, .53	.19, .51	.19, .50	.31, .59
.5	.42, .71	.41, .68	.40, .68	.51, .75
		n = 250		
.1	.05, .21	.05, .21	.05, .21	.09, .26
.25	.18, .38	.19, .38	.19, .38	.22, .43
.5	.42, .61	.44, .60	.43, .59	.46, .63

Table 4A

Ad Hoc and Bayesian Interval Estimates for ρ^2 (Missing Data = MCAR, Number of predictors = 2, Proportion of complete Data = .60) as a Function of Sample Size (n = 50, 100, 250) and Population Squared correlation ρ^2 value (.1, .25, .5).

	Method			
ρ^2	Listwise	Pairwise	Fill-in	Bayesian
		n = 50		
.1	.01, .59	.01, .43	.00, .34	.08, .63
.25	.09, .68	.09, .57	.07, .48	.17, .71
.5	.31, .82	.34, .75	.27, .67	.40, .84
		n = 100		
.1	.08, .44	.09, .40	.06, .33	.11, .47
.25	.22, .59	.23, .55	.16, .46	.26, .61
.5	.45, .75	.46, .72	.34, .62	.48, .77
		n = 250		
.1	.01, .19	.03, .17	.02, .13	.03, .21
.25	.13, .38	.14, .34	.09, .27	.14, .39
.5	.40, .62	.38, .58	.26, .45	.41, .63

Table 4B

Ad Hoc and Bayesian Interval Estimates for ρ^2 (Missing Data = MCAR. Number of predictors = 5. Proportion of complete Data = .60) as a Function of Sample Size (n = 50. 100. 250) and Population Squared correlation ρ^2 value (.1. .25. .5).

	Method			
ρ^2	Listwise	Pairwise	Fill-in	Bayesian
		n = 50		
.1	.00, .45	.03, .46	.00, .41	.16, .60
.25	.02, .54	.11, .56	.07, .51	.20, .64
.5	.19, .69	.34, .73	.26, .67	.38, .76
		n = 100		
.1	.00, .32	.07, .38	.07, .35	.10, .45
.25	.11, .51	.20, .52	.19, .50	.22, .58
.5	.35, .71	.44, .71	.42, .68	.45, .75
		n = 250		
.1	.05, .26	.06, .22	.05, .20	.09, .30
.25	.19, .43	.20, .40	.18, .37	.23, .46
.5	.44, .65	.46, .61	.41, .58	.47, .67

Table 4C

Ad Hoc and Bayesian Interval Estimates for ρ^2 (Missing Data = MCAR, Number of predictors = 7, Proportion of complete Data = .60) as a Function of Sample Size (n = 50, 100, 250) and Population Squared correlation ρ^2 value (.1, .25, .5).

	Method			
ρ^2	Listwise	Pairwise	Fill-in	Bayesian
		n = 50		
.1	.00, .44	.00, .52	.02, .46	.22, .64
.25	.05, .59	.13, .59	.13, .58	.31, .71
.5	.31, .76	.39, .76	.38, .75	.53, .83
		n = 100		
.1	.00, .37	.12, .45	.08, .39	.17, .55
.25	.06, .51	.23, .54	.18, .49	.25, .62
.5	.26, .68	.43, .70	.37, .65	.44, .75
		n = 250		
.1	.03, .23	.08, .25	.06, .22	.09, .29
.25	.17, .41	.20, .40	.19, .38	.23, .45
.5	.43, .64	.45, .61	.42, .59	.47, .67

Table 5A

Ad Hoc and Bayesian Interval Estimates for ρ^2 (Missing Data = MAR, Number of predictors = 2, Proportion of complete Data = .90) *as a Function of Sample Size* (n = 50, 100, 250) *and Population Squared correlation* ρ^2 *value* (.1, .25, .5).

	Method			
ρ^2	Listwise	Pairwise	Fill-in	Bayesian
		n = 50		
.1	.00, .23	.00, .22	.00, .21	.01, .32
.25	.04, .47	.05, .47	.04, .43	.08, .50
.5	.28, .69	.29, .69	.23, .64	.33, .72
		n = 100		
.1	.00, .12	.00, .12	.00, .11	.00, .17
.25	.05, .33	.06, .33	.04, .30	.07, .36
.5	.31, .61	.31, .61	.26, .55	.33, .62
		n = 250		
.1	.05, .20	.05, .20	.04, .18	.05, .21
.25	.17, .37	.17, .37	.15, .33	.18, .38
.5	.43, .60	.43, .60	.36, .54	.44, .62

Table 5B

Ad Hoc and Bayesian Interval Estimates for ρ^2 (Missing Data = MAR, Number of predictors = 5, Proportion of complete Data = .90) as a Function of Sample Size (n = 50, 100, 250) and Population Squared correlation ρ^2 value (.1, .25, .5).

	Method			
ρ^2	Listwise	Pairwise	Fill-in	Bayesian
		n = 50		
.1	.00, .32	.00, .32	.00, .31	.10, .48
.25	.07, .53	.08, .53	.07, .51	.21, .61
.5	.33, .73	.35, .74	.32, .71	.46, .78
		n = 100		
.1	.01, .26	.01, .26	.01, .26	.07, .33
.25	.12, .44	.13, .43	.12, .42	.20, .49
.5	.38, .66	.39, .67	.37, .65	.44, .69
		n = 250		
.1	.03, .14	.01, .14	.01, .13	.04, .18
.25	.13, .32	.13, .32	.04, .27	.15, .35
.5	.40, .58	.38, .57	.31, .50	.41, .60

Table 5C

Ad Hoc and Bayesian Interval Estimates for ρ^2 (Missing Data = MAR. Number of predictors = 7. Proportion of complete Data = .90) as a Function of Sample Size (n = 50. 100. 250) and Population Squared correlation ρ^2 value (.1, .25, .5).

	Method			
ρ^2	Listwise	Pairwise	Fill-in	Bayesian
		n = 50		
.1	.00, .35	.00, .37	.00, .29	.17, .55
.25	.05, .53	.06, .54	.01, .44	.27, .65
.5	.29, .72	.30, .72	.16, .61	.48, .79
		n = 100		
.1	.00, .28	.01, .27	.00, .25	.11, .39
.25	.10, .43	.12, .43	.10, .40	.21, .51
.5	.35, .64	.36, .64	.31, .60	.43, .69
		n = 250		
.1	.05, .21	.05, .21	.04, .19	.09, .26
.25	.20, .40	.20, .40	.14, .33	.25, .44
.5	.48, .62	.48, .61	.44, .59	.50, .66

Table 6A

Ad Hoc and Bayesian Interval Estimates for ρ^2 (Missing Data = MAR, Number of predictors = 2, Proportion of complete Data = .60) as a Function of Sample Size (n = 50, 100, 250) and Population Squared correlation ρ^2 value (.1, .25, .5).

	Method			
ρ^2	Listwise	Pairwise	Fill-in	Bayesian
		n = 50		
.1	.00, .11	.00, .10	.00, .10	.01, .25
.25	.00, .22	.00, .20	.00, .15	.04, .32
.5	.02, .51	.03, .46	.00, .27	.08, .55
		n = 100		
.1	.00, .10	.00, .09	.00, .09	.02, .18
.25	.00, .31	.02, .29	.01, .22	.03, .33
.5	.24, .61	.29, .61	.14, .43	.27, .63
		n = 250		
.1	.01, .17	.02, .15	.01, .12	.02, .18
.25	.08, .31	.09, .28	.05, .19	.09, .31
.5	.28, .53	.28, .49	.13, .29	.29, .54

Table 6B

Ad Hoc and Bayesian Interval Estimates for ρ^2 (Missing Data = MAR, Number of predictors = 5, Proportion of complete Data = .60) as a Function of Sample Size (n = 50, 100, 250) and Population Squared correlation ρ^2 value (.1, .25, .5).

	Method			
ρ^2	Listwise	Pairwise	Fill-in	Bayesian
		n = 50		
.1	.00, .34	.00, .36	.00, .27	.10, .57
.25	.00, .58	.05, .54	.00, .40	.23, .68
.5	.31, .79	.40, .79	.16, .60	.50, .84
		n = 100		
.1	.01, .36	.04, .35	.01, .25	.10, .44
.25	.06, .44	.10, .43	.05, .33	.16, .51
.5	.27, .64	.32, .63	.15, .46	.37, .69
		n = 250		
.1	.00, .12	.00, .13	.00, .08	.04, .31
.25	.06, .28	.07, .26	.03, .17	.10, .32
.5	.27, .51	.27, .48	.11, .28	.31, .54

Table 6C

Ad Hoc and Bayesian Interval Estimates for ρ^2 (Missing Data = MAR. Number of predictors = 7. Proportion of complete Data = .60) as a Function of Sample Size (n = 50. 100, 250) and Population Squared correlation ρ^2 value (.1, .25, .5).

	Method			
ρ^2	Listwise	Pairwise	Fill-in	Bayesian
		n = 50		
.1	.00, .40	.00, .43	.00, .21	.20, .63
.25	.00, .53	.05, .57	.00, .28	.31, .71
.5	.14, .69	.17, .67	.00, .33	.43, .80
		n = 100		
.1	.00, .30	.03, .35	.00, .17	.13, .47
.25	.04, .44	.10, .45	.02, .28	.20, .55
.5	.26, .64	.31, .63	.11, .42	.41, .70
		n = 250		
.1	.00, .11	.00, .11	.00, .08	.04, .21
.25	.05, .25	.06, .24	.02, .15	.10, .31
.5	.27, .51	.25, .46	.10, .27	.33, .55

Table 7

Empirical Coverage Probabilities and (Average Interval Width) as a Function of Method and Total Sample Size.

	Method			
Sample size	Listwise	Pairwise	Fill-in	Bayesian
50	.97 (.43)	.94 (.41)	.94 (.36)	.61 (.39)
100	1.00 (.32)	.97 (.29)	.89 (.27)	.81 (.30)
250	1.00 (.19)	.92 (.18)	.75 (.16)	1.00 (.20)

Table 8

The empirical coverage Probabilities as a Function of Method and Missing data type.

	Method			
Missing data type	Listwise	Pairwise	Fill-in	Bayesian
MCAR	1.00	1.00	.98	.72
MAR	.98	.89	.74	.89

Table 9

The empirical coverage Probabilities as a Function of Method and Proportion of Missing data (P_{miss})

P_{miss}	Method			
	Listwise	Pairwise	Fill-in	Bayesian
.10	1.00	1.00	1.00	.80
.40	.98	.89	.72	.81

Table 10

Empirical Coverage Probabilities as a Function of Method and Population Squared Multiple Correlation Value.

	Method			
ρ^2	Listwise	Pairwise	Fill-in	Bayesian
.1	1.00	.97	.91	.67
.25	.97	.94	.86	.81
.50	1.00	.92	.81	.94

CHAPTER V

Discussion and Conclusion

Listwise deletion is clearly the best Ad Hoc method and fill-in is the worst across all conditions considered in the dissertation. This result can be clearly seen in the coverage probabilities given in Table 7. Another major finding was the poor performance of the Bayesian method in estimating ρ^2 values for sample sizes less than 250. In these cases, even the poorest ad hoc method (fill-in) showed higher coverage probabilities (see Table 7). The bases for this surprising result are considered next.

As indicated in the literature survey, the Bayesian method was expected to be more accurate than the ad hoc methods for small sample sizes and MAR cases. The results in Table 7 show that the Bayesian method covered the ρ^2 values fewer times than pairwise and listwise methods for sample sizes less than 250. Further, listwise deletion did better than the Bayesian method when the sample data are MAR (see table 8). The argument below will show that the reason for Bayesian poor performance in estimating the true population ρ^2 values lies in the form of the prior distribution for ρ^2.

It is well known that Bayesian estimates can be sensitive to the form of the prior distributions when the sample size is small. It is of interest to consider how the non-informative prior assigned to Σ (the variance-covariance matrix), determines the implied prior for ρ^2. The priors used can be stated as follows:

$$P'(\underline{\mu},\Sigma) \propto P'(\underline{\mu})P'(\Sigma)$$

$$\propto \frac{1}{|\Sigma|^{(nvar+1)/2}} \qquad (5.1)$$

$|\Sigma|$ = determinant of the variance-covariance matrix,
nvar = the number of predictor variables plus the dependent variable.

Using the above priors, it can be shown that the implied prior for ρ^2 is

$$P'(\rho^2) \propto \frac{(\rho^2)^{nxvar/2-1}}{(1-\rho^2)^{nxvar/2+1}} \qquad (5.2)$$

nxvar = the number of predictor variables used in the analysis.

The problem here is that even though the prior for variance-covariance matrix may be non-informative, the implied prior for ρ^2 given in equation (5.2) is increasingly skewed to the left as the number of predictors increases. Figures 2 and 3 support the argument as are shown below. These figures were obtained by using the BUGS computer program to draw n = 5000 Σ matrices given that $P'(\Sigma)$ is approximately of the non-informative form,

$$P'(\Sigma) = |\Sigma|^{-(nvar + 1)/2}.$$

Given each generated Σ matrix, ρ^2 was computed. Figures 2 and 3 provide the empirical histogram of the 5000 ρ^2 values. Figure 2 has the form $1/ (1 - \rho^2)^2$ and figure 3 has the form $(\rho)^{5/2} / (1 - \rho^2)^{9/2}$. Both distributions are skewed to the left and place high probability on large ρ^2 values.

Thus, the following hypotheses can be stated:

- The worst Bayesian results should be observed when the sample sizes are small and the true population squared multiple correlation values are small. In these cases,the skewed prior for ρ^2 is inappropriate and has a dominant effect on the posterior due to the small sample sizes.

- The best Bayesian Results should be observed for cases with large sample sizes, or small sample sizes but large population ρ^2 values. With large n, the prior has little effect. For cases with small sample sizes, if ρ^2 is large the prior is appropriate.

The results in the basic tables 3A-6C support the arguments when specific cases are considered. For the 12 cases where $n = 50$ and $\rho^2 = .1$, the Bayesian did not cover the true ρ^2 value in seven out of twelve cases. However, for the 12 cases where $n = 50$ and $\rho^2 = .5$, the Bayesian method covered the true value of ρ^2 for all but one case. When all cases for which $n = 250$ are considered, the Bayesian method covered the parameter values all the time.

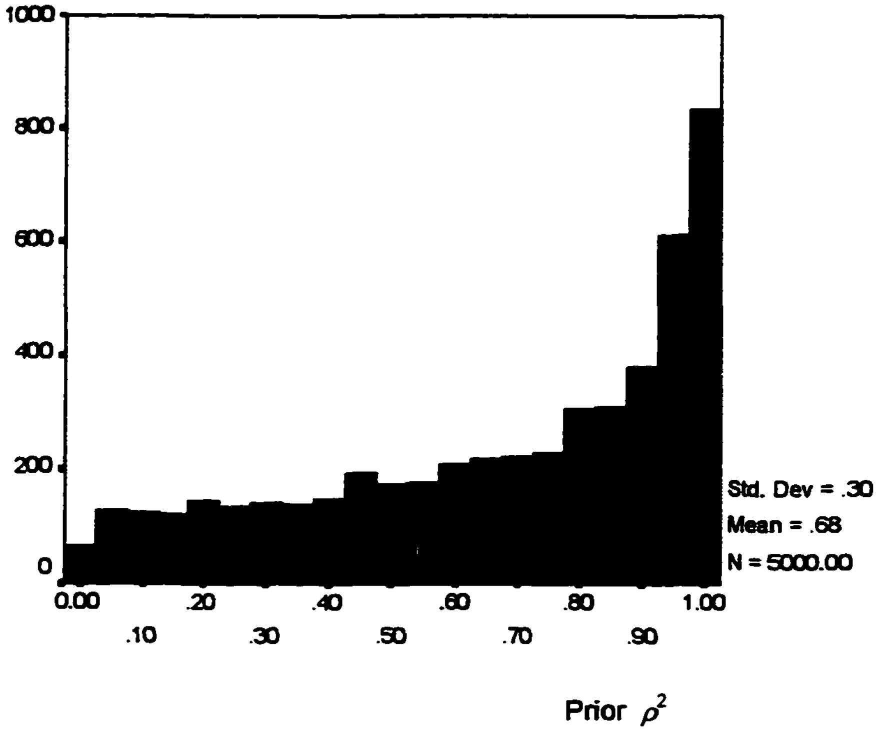


Fig. 2

The prior distribution for ρ^2 when there are 2 predictors using a non-informative prior for Σ.

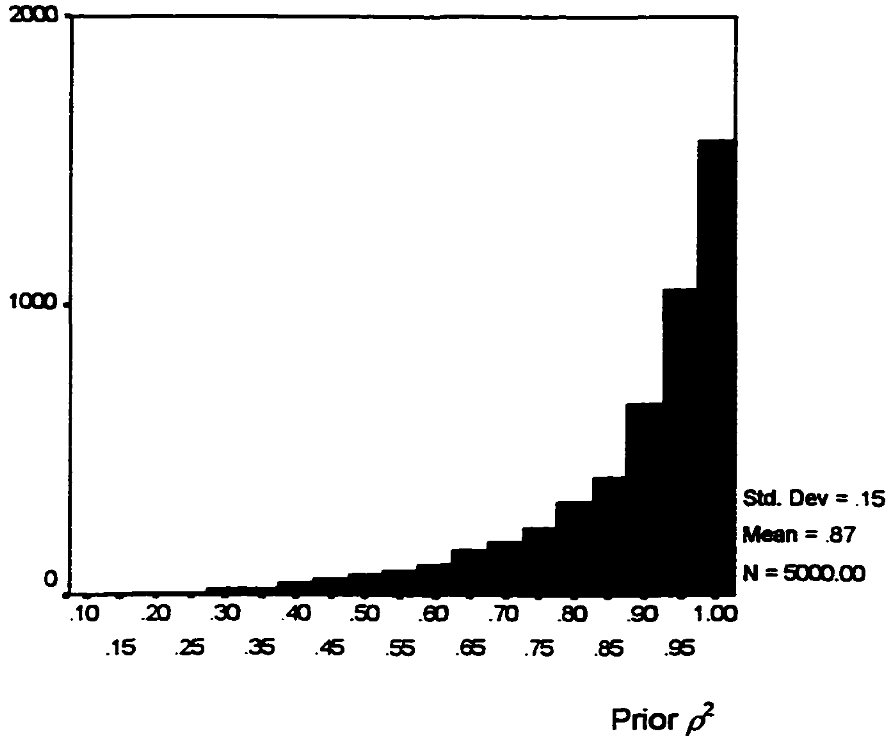


Fig. 3

The prior distribution for ρ^2 when there are 7 predictor using a non-informative prior for Σ.

This analysis of the implied prior for ρ^2 does not mean that the Bayesian method may not be useful. However, to achieve accurate results the prior for ρ^2 must be appropriate. For example, Table 10 shows that larger values of ρ^2 can yield better results given our prior distribution.

The preceding arguments show that the Bayesian method must be used with care. Standard approach to assigning non-informative priors to $\underline{\mu}$ and Σ may produce inappropriate priors for the population ρ^2, and thus poor interval estimates for the true population ρ^2 values. Clearly the issue of construction of priors directly for the population ρ^2 needs to be further studied.

Another question for further examination in this dissertation is how much missing data from a data set can become a problem. Table 9 shows that the proportion of missing data has a big effect on the number of times the estimates for each method covered the true population ρ^2 values. It is of interest to know what proportion of missing data will compel an investigator to use non-standard methods of analyses for incomplete data sets.

APPENDIX

A : Description of the full conditional posterior distributions

(I) $P(M|\underline{\mu},\Sigma ,O)$

The data set can be divided into groups, each with its own unique missing data pattern. This allows one to specify separate conditional distributions for each distinct missing data parttern. Thus each individual case i (i = 1,..., n) has a case-specific conditional distribution that is multivariate normal and determined by

$$P(\underline{M}_i \mid \underline{O}_i,\underline{\mu},\Sigma) = N[E(\underline{M}_i \mid \underline{O}_i,\underline{\mu},\Sigma),\ \Sigma_{M|O}] \tag{A.1}$$

where;

$\underline{M}_i$, $\underline{O}_i$ = represent the vectors for missing and observed variables respectively for subject i,

$E(\underline{M}_i \mid \underline{O}_i, \mu, \Sigma)$ = the expected value for each distinct missing data pattern ($\underline{M}_i$) conditional on the current values of the parameters (μ, Σ) and the observed variables(O_i), and

$\Sigma_{M|O}$ = the variance-covariance matrix for each unique missing data group conditional on specific observed variables.

Further the notation $P(\mathbf{M}|\mu,\Sigma,\mathbf{O})$ should be understood to represent a collection of conditional distributions $P(\underline{M}_i|\underline{O}_i,\mu,\Sigma)$ in this dissertation.

For a simple example, suppose we consider an incomplete data set from a trivariate normal distribution in which there are missing values for the dependent variable but the predictor variables are always observed. Let "1" denote an observed value and "0" a missing value. Thus for the y x_1 x_2 data, the data patterns are 1 1 1 and 0 1 1. The normal distribution of the missing y values conditional on the current values of the predictor variables can be represented as:

$$
\begin{aligned}
P(M|\underline{\mu},\Sigma,O) &= P(y|\underline{\mu},\Sigma,x_1,x_2) \\
&= N(\beta_0 + \beta_{1x1} + \beta_{2x2},\ \sigma_{yy\,|\,x1x2})
\end{aligned}
\tag{A .2}
$$

where,

$$\underline{\mu} = \begin{bmatrix} \mu_y \\ \underline{\mu}_x \end{bmatrix},$$

$$\Sigma = \begin{bmatrix} \sigma_{x1x1} & \sigma_{x1x2} & \sigma_{x1y} \\ & \sigma_{x2x2} & \sigma_{x2y} \\ & & \sigma_{yy} \end{bmatrix} = \begin{bmatrix} \Sigma_{xx} & \underline{\sigma}_{xy} \\ & \sigma_{yy} \end{bmatrix},$$

$$\beta_0 = \mu_y - \underline{\beta}'\,\underline{\mu}_y,$$

$$\underline{\beta} = \Sigma_{xx}^{-1}\ \underline{\sigma}_{xy},\ \text{and}$$

$$\sigma_{yy|x1x2} = \sigma_{yy} - \underline{\beta}'\ \Sigma_{xx}^{-1}\ \underline{\beta}$$

(II) $P''(\Sigma|\underline{\mu},M,O)$

The conditional posterior distribution for Σ is given as:

$$\begin{aligned} P''(\Sigma|\underline{\mu},M,O) &\propto P'(\Sigma|\underline{\mu})\ P(M,O|\underline{\mu},\Sigma) \\ &\propto P'(\Sigma|\underline{\mu})\ L(\Sigma|\ M,\underline{\mu},O) \end{aligned} \tag{A.3}$$

where;

$L(\Sigma|\ M,\underline{\mu},O)$ = the multivariate normal likelihood for Σ and $\underline{\mu}$ where $\underline{\mu}$ is fixed.

Using Jeffrey's rule for the selection of a non-informative prior distribution (Box & Tiao, 1973,P. 42, 54), the following prior can be stated :

$$\begin{aligned} P'(\underline{\mu},\ \Sigma) &\propto P'(\underline{\mu})\ P'(\Sigma) \\ &\propto k\ P'(\Sigma) \\ &\propto P'(\Sigma) \\ &\propto |\Sigma|^{-\frac{(q+1)}{2}} \end{aligned} \tag{A.4}$$

$|\Sigma|$ = determinant of the variance-covariance matrix Σ

Thus

$$P'(\Sigma|\underline{\mu}) = P'(\Sigma) \propto |\Sigma|^{-\frac{(q+1)}{2}} \qquad \text{(A.5)}$$

Suppose $\underline{z}_i = [\underline{y}_i, \underline{x}_i]$ is the pseudo-complete data set which includes current estimates for all missing values. The likelihood for Σ, fixing $\underline{\mu}$, is given by:

$$L(\Sigma \mid \underline{\mu}, \underline{z}) \propto |\Sigma|^{-\frac{n}{2}} \exp\left\{-\frac{1}{2} \sum_{i=1}^{n} (z_j - \underline{\mu})' \, \Sigma^{-1} (z_j - \underline{\mu})\right\}$$

$$\propto |\Sigma|^{-\frac{n}{2}} \exp\left\{-\frac{1}{2} \operatorname{tr} \Sigma^{-1} \sum_{i=1}^{n} (z_j - \underline{\mu}) (z_j - \underline{\mu})'\right\}$$

$$\propto |\Sigma|^{-\frac{n}{2}} \exp\left\{-\frac{1}{2} \operatorname{tr} \Sigma^{-1} A\right\}$$

(A.6)

where $A = \sum_{i=1}^{n} (z_j - \underline{\mu}) (z_i - \underline{\mu})'$.

Thus A represents the sum of the sum of squares and cross-product terms of the completed data based on deviations from the population mean $\underline{\mu}$. Combining the results in (A.5) and (A.6) and treating $\underline{\mu}$ as fixed, the posterior distribution for Σ given $\underline{\mu}$, O, M can be expressed

$$\begin{aligned} P''(\Sigma|\underline{\mu},M,O) &\propto |\Sigma|^{-\frac{(q+1)}{2}} \; |\Sigma|^{-\frac{n}{2}} \exp\{-\frac{1}{2} \operatorname{tr} \Sigma^{-1} A\} \\ &\propto |\Sigma|^{-\frac{(n+q+1)}{2}} \exp\{-\frac{1}{2} \operatorname{tr} \Sigma^{-1} A\} \end{aligned} \tag{A.7}$$

The last expression in (A.7) has the form of inverted Wishart distribution with n degrees of freedom and q x q precision matrix A (Anderson,1971, p. 268, Box & Tiao, 1973, p. 427). Thus the posterior distribution of Σ can be written

$$P''(\Sigma|\underline{\mu},M,O) = W_q^{-1}(A,n) \tag{A.8}$$

Suppose now $\psi = \Sigma^{-1}$denotes the precision matrix, the posterior distribution of ψ has a Wishart distribution with n degrees of freedom and q x q variance-covariance matrix A^{-1} denoted by

$$P''(\psi|\underline{\mu},M,O) \;=\; W_q\,(A^{-1},n) \tag{A.9}$$

Thus, if one randomly generates a ψ matrix from $W_q\,(A^{-1},n)$ the inverse of the generated value yields a value from the posterior distribution of Σ .

(III) $P''(\underline{\mu}|\Sigma,M,O)$

Treating Σ as fixed, the posterior distribution of $\underline{\mu}$ is derived as:

$$P''(\underline{\mu}|\Sigma,M,O) \;\propto\; P'(\underline{\mu}|\Sigma)\; L(\underline{\mu}|\Sigma,M,O) \tag{A.10}$$

Following (2.4.1.2.) $P'(\underline{\mu}|\Sigma)$ is given by

$$P'(\underline{\mu}|\Sigma) = P'(\underline{\mu}) \;\propto\; k \tag{A.11}$$

The likelihood, fixing Σ, can be written

$$L(\underline{\mu}|\Sigma, \underline{y}) \propto \Sigma^{-\frac{n}{2}} \exp\{-\frac{1}{2}\sum_{i=1}^{n}(\underline{y}_i - \underline{\mu})'\Sigma^{-1}(\underline{y}_i - \underline{\mu})\}$$

$$\propto \Sigma^{-\frac{n}{2}} \exp\{-\frac{1}{2} \operatorname{tr} \Sigma^{-1} \sum_{i=1}^{n} (\underline{y}_i - \underline{\mu})(\underline{y}_i - \underline{\mu})'\}$$

(A.12)

The expression under the summation sign in (A.12) can be rewritten as

$$\sum_{i=1}^{n} (\underline{y}_i - \underline{\bar{y}} - \underline{\mu} + \underline{\bar{y}})(\underline{y}_i - \underline{\bar{y}} - \underline{\mu} + \underline{\bar{y}})'$$

$$= \sum_{i=1}^{n}[(\underline{y}_i - \underline{\bar{y}}) - (\underline{\mu} + \underline{\bar{y}})][(\underline{y}_i - \underline{\bar{y}}) - (\underline{\mu} + \underline{\bar{y}})']$$

$$= \sum_{i=1}^{n} (\underline{y}_i - \underline{\bar{y}})(\underline{y}_i - \underline{\bar{y}})' + \sum_{i=1}^{n} (\underline{\mu} - \underline{\bar{y}})(\underline{\mu} - \underline{\bar{y}})$$

$$- \sum_{i=1}^{n} (\underline{y}_i - \underline{\bar{y}})(\underline{\mu} - \underline{\bar{y}})' + \sum_{i=1}^{n} (\underline{\mu} - \underline{\bar{y}})(\underline{y}_i - \underline{\bar{y}})$$

(A.13)

The last two terms in (A.13) sum up to zero , thus,

$$\Sigma(\underline{y}_i-\underline{\mu})(\underline{y}_i-\underline{\mu})' \quad = \quad \Sigma(\underline{y}_i-\bar{\underline{y}})(\underline{y}_i-\bar{\underline{y}})'+n(\underline{\mu}-\bar{\underline{y}})(\underline{\mu}-\bar{\underline{y}})$$

(A.14)

Furthermore, the first term in the right hand side in (A.14) is a constant with respect to $\underline{\mu}$. It is not difficult to see that by dropping the constant term from (A.14) and substituting from (A.14) into (A.12) the expression for the likelihood reduces to

$$L(\underline{\mu}|\Sigma,\underline{y}) \quad \propto \quad |\Sigma|^{-\frac{n}{2}} \exp\{-\frac{n}{2} \operatorname{tr} \Sigma^{-1} n(\underline{\mu}-\bar{\underline{y}})(\underline{\mu}-\bar{\underline{y}})'\}$$

(A.15)

Finally, combining expressions (A.10), (A.11), and (A.15) the posterior distribution of $\underline{\mu}$ can written as

$$P''(\underline{\mu}|\Sigma,\underline{y}) \quad \propto \quad |\Sigma|^{-\frac{n}{2}} \exp\{-\frac{n}{2} \operatorname{tr} [\ \Sigma^{-1} n(\underline{\mu}-\underline{y}_i)(\underline{\mu}-\underline{y}_i)'\ ]\}$$

(A.16).

The expression in (A.16) has the form of a multivariate normal distribution with mean $\underline{y}$ and variance matrix Σ/n.

Appendix B

B: Data deletion Processes for MCAR and MAR.

For each missing data process, 10% and 40% of the sample data were deleted.

(I) MCAR

Let P_{miss} be the probability of data missing where $P_{miss} = .10, .40$. The probability of complete data across nvar variables can be expressed in the following form:

$$(1 - pmiss)^{nvar} = .90$$
$$(1 - pmiss)^{nvar} = .60 \tag{B.1}$$

Where

nvar = the number of y and x variables in the analysis.

The equations in (B.1) assume that the probability that any one variable is missing is independent. P_{miss} is computed by solving the equations in (B1). P_{miss} is thus used in the FORTRAN program to delete the desired quantity of data. The program deletes any data point for which a uniform random generated number (u) is less than p_{miss}, $(0<u>1)$.

(II) MAR

Three missing data patterns are identified:

1) All variables observed (npat = 1)
2) y is missing and all x's are observed (npat =2)
3) y is observed, all x's are missing(npat = 3)

The FORTRAN program generates random numbers (u) that determine the missing data patterns in the following way: (note that the mean of all variables is assumed to be zero)

a) $u < P_1$, npat = 1

b) $p_1 < u < p_1 + p_2$ and $x_1 < 0$, npat = 2; if $x_1 \geq 0$, npat = 1

c) $u > p_1 + p_2$ and $y < 0$, npat = 3; if $y \geq 0$, npat = 1

p_1 and p_2 are defined such that p(complete) is .90 and .60, where

$$\begin{aligned} P(\text{npat} = 1) &= P(\text{complete}) \\ &= p_1 + p_2(.50) + (1-p_1-p_2)(.50) \\ &= (p_1 + 1)(.5) \end{aligned}$$

Further P(npat =2) = $(p_2)(.5)$ and

P(npat = 3) = $(1-p_2-p_3)(.5)$

Thus the following P_1, P_2 values (.80,.10), (.20, .40) were used.

APPENDIX C

The FORTRAN program below was used to compute the solutions for ρ_u^2, ρ_l^2. This program allows one to compute an interval estimate for ρ^2 given R_o^2, level of confidence (1 - α), number of predictor variables and initial 'guesses' for ρ_u^2 and ρ_l^2.

```
IMPLICIT REAL*8(A-H,O-Z)
     EXTERNAL DNEQNF, FU ,FL
     DIMENSION RHO2IU(1),RHO2U(1)
     DIMENSION RHO2IL(1),RHO2L(1)
     COMMON NSUBJ, NXVAR,
R2OBS,ALPHAU,ALPHAL

     OPEN (1,FILE='CI.INP')
     OPEN (2,FILE='CI.OUT')

     READ(1,100) NSUBJ, NXVAR,
R2OBS,ALPHAU,ALPHAL
100  FORMAT(I5,I5,3F5.0)
5
C RHO2U,RHO2L CONTAINS THE UPPER AND
LOWER BOUNDS FOR THE POP. SQ. MULT
C ALPHAU AND ALPAHL ARE THE LEFT AND RIGHT
HAND PROBABILITIES
```

```
      WRITE(2,101) NSUBJ, NXVAR,
R2OBS,ALPHAU,ALPHAL
101   FORMAT(1X,'NSUBJ, NXVAR,
R20BS',I6,I6,F10.4/
     11X,'ALPHAU,ALPHAL',2F10.4//)
      NVALUE=1
      ITMAX=200
      ERRER=.000001

C OBTAIN UPPER LIMIT OF CONFIDENCE INTERVAL
C SET START VALUE FOR UPPER LIMIT
      READ(1,104) RHO2IU
104   FORMAT(F5.0)
      CALL
DNEQNF(FU,ERRER,NVALUE,ITMAX,RHO2IU,RHO2U
,FNORM)
      WRITE(2,102) RHO2U,FNORM
102   FORMAT(1X, 'UPPER
BOUND',F10.4,1X,'FNORM=',F8.4)

      READ(1,104) RHO2IL
      CALL
DNEQNF(FL,ERRER,NVALUE,ITMAX,RHO2IL,RHO2L,
FNORM)
      WRITE(2,103) RHO2L,FNORM
```

```
103   FORMAT(1X, 'LOWER
BOUND',F10.4,1X,'FNORM=',F8.4)

      STOP
      END
      SUBROUTINE FU(RHO2U,FXX,NVALUE)
      IMPLICIT REAL*8(A-H,O-Z)
      EXTERNAL DBETDF, DGAMMA, DLNGAM

      DIMENSION RHO2U(NVALUE),FXX(NVALUE)
      COMMON NSUBJ, NXVAR, R2OBS
,ALPHAU,ALPHAL

      FXX(1)=-ALPHAU

      IF(RHO2U(1) .GE. 1.0) GO TO 1000
      IF(RHO2U(1) .LT. 0.0) GO TO 1000

C     WRITE(2,300) RHO2U(1)
300   FORMAT(1X,'RHO2U VALUE',F10.4)

C COMPUTE F(RHO2U)
      XN=NSUBJ-1

      F=(XN/2)*DLOG(1-RHO2U(1))
```

```
C     write(2,340) f
340   FORMAT(1X, 'F VALUE',F10.4)

C COMPUTE DLOG OF GAMMA(XN/2)
      GAM1=DLNGAM(XN/2)

C APPROXIMATE INFINITE SERIES WITH FIRST 100 TERMS
      SUM=0.0


      nterms=150
      IF(RHO2U(1) .EQ. 0.0) NTERMS=1
      DO 1 II=1,NTERMS
C COMPUTE LOG OF I FACTORIAL
      I=II-1
      IF(I .EQ. 0) FSUM=0.0
      IF( I .EQ. 1) FSUM=0.0
      IF(I .LE. 1) GO TO 3
      FSUM=0.0
      DO 4 J=1,I
      XJ=J
      FSUM=FSUM+DLOG(XJ)
4     CONTINUE
3     CONTINUE
C      WRITE(2,350) I,FSUM
350   FORMAT(1X,'I,FSUM',I5,F10.4)
C COMPUTE I*DLOG(RHO2U)
```

```
C     COMPUTE G
      G=0.0
      IF(RHO2U(1) .EQ. 0.0) GO TO 500
      G=I*DLOG(RHO2U(1))
500   CONTINUE

C CALCULATE LOG OF GAMMA(A+B)
      A=(NXVAR)/2.0 + I
      B=(XN-NXVAR )/2.0
C     WRITE(2,379) A,B
379   FORMAT(1X,'A,B',2F10.4)
C     GAM2=DGAMMA(A+B)
      GAM2=DLNGAM(A+B)

C NOW INTEGRATE THE ITH TERM

      XINT=DBETDF(R2OBS,A,B)
      TERM=0
      IF (XINT .EQ. 0) GO TO 8888

C     WRITE(2,360) XINT,R2OBS
360   FORMAT(1X,'XINT,R2OBS',2F10.4)
      XINT=DLOG(XINT)

C     WRITE(2,390) G,FSUM,XINT,GAM2
390   FORMAT(1X,'PARTS OF TERM',4F10.4)
      TERM=G-FSUM +XINT + GAM2
```

```
      TERM=DEXP(TERM)

C     WRITE(2,380) TERM
380   FORMAT(1X,'TERM',F10.8)

8888  SUM=SUM+TERM

1     CONTINUE
      SUM=SUM*DEXP(F-GAM1)
C      WRITE(2,395) SUM
395   FORMAT(1X,'SUM',F12.4)

      FXX(1)=SUM-ALPHAU

      WRITE(2,400) RHO2U,FXX
400   FORMAT(1X,'RHO2U AND FXX VALUE',2F15.4)
1000  RETURN
      END


      SUBROUTINE FL(RHO2L,FXX,NVALUE)
      IMPLICIT REAL*8(A-H,O-Z)
      EXTERNAL DBETDF, DGAMMA, DLNGAM

      DIMENSION RHO2L(NVALUE),FXX(NVALUE)
      COMMON NSUBJ, NXVAR, R2OBS
,ALPHAU,ALPHAL
```

```
      FXX(1)=-ALPHAL

      IF(RHO2L(1) .GE. 1.0) GO TO 1000
      IF(RHO2L(1) .LE. 0.0) GO TO 1000

c     WRITE(2,300) RHO2L(1)
300   FORMAT(1X,'RHO2L VALUE',F10.4)

C COMPUTE F(RHO2L)
      XN=NSUBJ-1
      F=(XN/2)*DLOG(1-RHO2L(1))
C       write(2,340) f
340   FORMAT(1X, 'F VALUE',F10.4)

C COMPUTE DLOG OF GAMMA(XN/2)
      GAM1=DLNGAM(XN/2)

C APPROXIMATE INFINITE SERIES WITH FIRST 100
TERMS
      SUM=0.0
      nterms=150
      IF(RHO2L(1) .EQ. 0.0) NTERMS=1
      DO 1 II=1,NTERMS
C COMPUTE LOG OF I FACTORIAL
      I=II-1
      IF(I .EQ. 0) FSUM=0.0
```

```
      IF( I .EQ. 1) FSUM=0.0
      IF(I .LE. 1) GO TO 3
      FSUM=0.0
      DO 4 J=1,I
      XJ=J
      FSUM=FSUM+DLOG(XJ)
4     CONTINUE
3     CONTINUE
C       WRITE(2,350) I,FSUM
350     FORMAT(1X,'I,FSUM',I5,F10.4)
C COMPUTE I*DLOG(RHO2L)
C       COMPUTE G
      G=0.0
      IF(RHO2L(1) .EQ. 0.0) GO TO 500
      G=I*DLOG(RHO2L(1))
500   CONTINUE

C CALCULATE GAMMA(A+B)
      A=(NXVAR)/2.0 + I
      B=(XN-NXVAR )/2.0
C       WRITE(2,379) A,B
379     FORMAT(1X,'A,B',2F10.4)
C       GAM2=DGAMMA(A+B)
      GAM2=DLNGAM(A+B)

C NOW INTEGRATE THE ITH TERM
```

```
      XINT=DBETDF(R2OBS,A,B)
      XINT=1.0-XINT
      TERM=0
      IF (XINT .EQ. 0) GO TO 8888

C      WRITE(2,360) XINT,R2OBS
360   FORMAT(1X,'XINT,R2OBS',2F10.4)
      XINT=DLOG(XINT)

C      WRITE(2,390) G,FSUM,XINT,GAM2
390   FORMAT(1X,'PARTS OF TERM',4F10.4)
      TERM=G-FSUM +XINT + GAM2
      TERM=DEXP(TERM)

C     WRITE(2,380) TERM
380   FORMAT(1X,'TERM',F10.8)

8888  SUM=SUM+TERM

1     CONTINUE
      SUM=SUM*DEXP(F-GAM1)
C      WRITE(2,395) SUM
395   FORMAT(1X,'SUM',F12.4)

      FXX(1)=SUM-ALPHAL

      WRITE(2,400) RHO2L,FXX
```

```
400   FORMAT(1X,'RHO2L AND FXX VALUE',2F15.4)
1000  RETURN
      END
```

References

Anderson, W. T. (1984). An introduction to multivariate statistical analysis. New York: John Wiley & sons Inc.

Box, E. P. & Tiao, G. C. (1973). Bayesian inference in statistical analysis. New York: Wiley.

Buckle D. J. (1993). Bayesian inference for stable distribution. Journal of the American Statistical Association, 90, 605-613.

Cassella, G. & George, E. I. (1992). Explaining the Gibbs sampler. The American statistician, 46, 146 - 174.

De Groot, M. H. (1970). Optimal statistical decisions. New York: McGraw Hill.

Dempster, A. P., Laird, N. M., & Rubin, D. B.(1977). Maximum likelihood estimation from incomplete data via the EM algorithm (with discussion). Journal of the Royal Statistical Society, Ser. B, 39, 1-38.

Dillon, W., R. & Goldstein, M. (1984). Multivariate analysis: Methods and applications. New York: John Wiley & sons Inc.

Gelfand, A. E., Hills, S. E., Raccine-Poon, A., & Smith, A. F. M. (1990a). Illustration of Bayesian inference in normal data models using Gibbs sampling. Journal of the American Statistical Association, 85, 972-985.

Gelfand, A. E. & Smith, A. F. M. (1990b). Sampling-based approahes to calculating marginal densities. Journal of the American Statistical Association, 85, 398-409.

Geman, S. & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesisan restoration of images. IEEE transactions on pattern analysis and machine intelligence, 6(6), 721-741.

George, E. I. & McCulloch, R. E. (1993). Variable selection via Gibbs sampling. Journal of the American Statistical Association, 88, 881-889.

Gross, A. L. & Torres-Quevedo, R. (1995). Estimating correlations with missing data. A Bayesian approach. Psychometrika, 60(3), 341-354.

Hobert, J. P. & Casella, G. (1996). The effect of improper priors on Gibbs sampling in hierarchical linear mixed models. Journal of the american statistical association, 91, 1461-1473.

Kolassa, J. E. & Tanner, M. A. (1994). Approximate Conditional inference in exponential families via the Gibbs sampler. Journal of the american statistical association, 89, 697-702.

Kromrey, J. D. & Hines, C. V. (1994). Nonrandomly missing data in multiple regression: An empirical comparison of common missing-data treatments. Educational and psychological measurement, 54(3), 573-593.

Lee, P. M. (1989). Bayesian statistics: An introduction. New York: Halstead Press.

Little, R. J. A. & Rubin, D. B. (1987). Statistical analysis with missing data. New York: John Wiley & sons Inc.

Liu, C. (1996). Bayesian robust multivariate linear regression with incomplete data. Journal of the American Statistical Association, 91, 1219-1227.

Liu, J. S., Wong W. H. & Kong A. (1994). Covariance structure of the Gibbs sampler with applications to the comparisons of estimators and augmentation schemes. Biometrika, 81(1), 27-40.

Lord, F. M. & Novick, M. R. (1968). Statistical theories of mental test scores. Reading, Mass.: Addison-Wesley.

Mood, A. M., Graybill, F. A., & Boes, D. C. (1974). Introduction to the theory of statistics. McGraw-Hill, Inc.

Perlis, T. (1996). A Gibbs Sampler approach to estimation of a multiple correlation coefficient in the presence of missing data. Unpublished doctoral dissertation, City University of New York, New York.

Peruggia, M. & Santner, T. (1996). Bayesian analysis of time evolution of earthquakes. Journal of the american atatistical association, 91 1209-1218.

Ritter, C. & Tanner, M. A. (1992). Faciliting the Gibbs sampler: The Gibbs stopper and the Griddy-Gibbs sample. Journal of the american statistical association, 87, 861-868.

Rubin, D. B. (1976). Inference and missing data. Biometrika, 63(3), 581-592.

Rubin, D. B. (1996). Multiple imputation after 18 + years. Journal of the american statistical association, 91, 473-489.

Samaniego, F. J. & Neath, A. A. (1996). How to be better Bayesian. Journal of the american statistical association, 91, 733-742.

Sinha, D. (1993). Semiparametric Bayesian analysis of multiple event time data. Journal of the american statistical association, 88, 779-783.

Sobel, M. J. (1993). Bayes and empirical procedures for comparing parameters. Journal of the american statistical association,88, 687-693.

Sun, D. & Ye, K. (1995). Reference prior Bayesian analysis for normal products. Journal of the american statistical association, 90, 589-597.

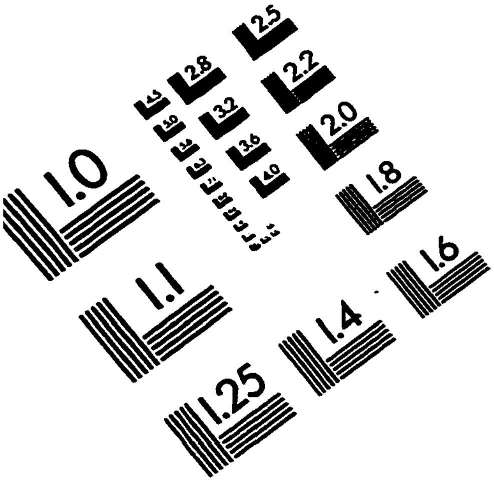
Tanner, M. A. & Wong, W. (1987). The calculation of the posterior distributions by data augmentation. Journal of the american statistical association, 82, 528-550.

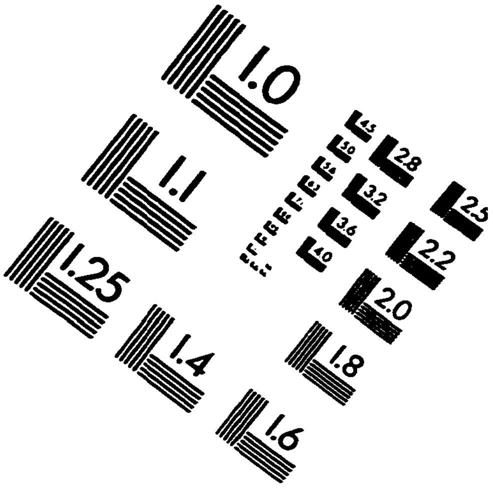
Winer, B. J. (1971). Statistical principles in experimental design. New York: McGraw - Hill.

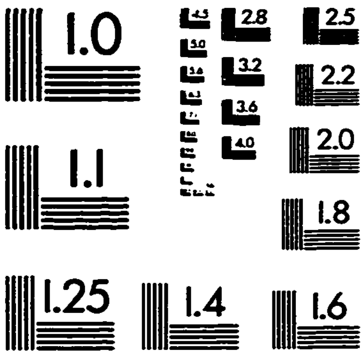
Zeger, S. L. & Karim, M. R. (1991). Generalized linear models with random effects; A Gibbs sampling approach. Journal of the american statistical association, 86, 79-86.

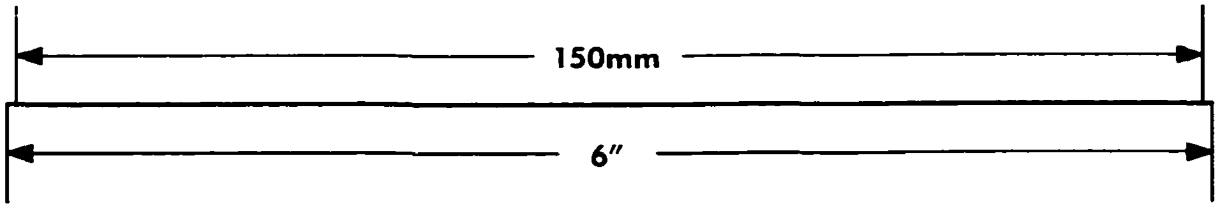
Zellner, A. & Min, C. (1995). Gibbs sampler convergence criteria. Journal of the american statistical association, 90, 921-927.

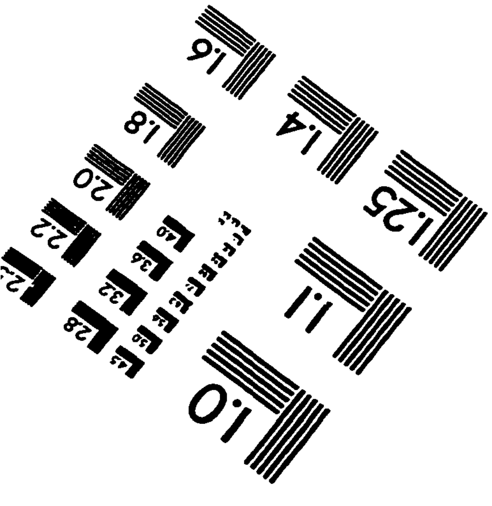
IMAGE EVALUATION
TEST TARGET (QA-3)











APPLIED IMAGE . Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989