



2007-01-31

# Selecting the Best Linear Mixed Model Using Predictive Approaches

Jun Wang

*Brigham Young University - Provo*

Follow this and additional works at: <https://scholarsarchive.byu.edu/etd>



Part of the [Statistics and Probability Commons](#)

---

## BYU ScholarsArchive Citation

Wang, Jun, "Selecting the Best Linear Mixed Model Using Predictive Approaches" (2007). *All Theses and Dissertations*. 841.  
<https://scholarsarchive.byu.edu/etd/841>

This Selected Project is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in All Theses and Dissertations by an authorized administrator of BYU ScholarsArchive. For more information, please contact [scholarsarchive@byu.edu](mailto:scholarsarchive@byu.edu), [ellen\\_amatangelo@byu.edu](mailto:ellen_amatangelo@byu.edu).

SELECTING THE BEST LINEAR MIXED MODEL  
USING PREDICTIVE APPROACHES

By

Jun Wang

A project submitted to the faculty of

Brigham Young University

in partial fulfillment of the requirements for the degree of

Master of Science

Department of Statistics

Brigham Young University

April 2007



BRIGHAM YOUNG UNIVERSITY

GRADUATE COMMITTEE APPROVAL

of a project submitted by

Jun Wang

This project has been read by each member of the following graduate committee and by majority vote has been found to be satisfactory.

\_\_\_\_\_

Date

\_\_\_\_\_

G. Bruce Schaalje, Chair

\_\_\_\_\_

Date

\_\_\_\_\_

Gilbert W. Fellingham

\_\_\_\_\_

Date

\_\_\_\_\_

H. Dennis Tolley



BRIGHAM YOUNG UNIVERSITY

As chair of the candidate's graduate committee, I have read the project of Jun Wang in its final form and have found that (1) its format, citations, and bibliographical style are consistent and acceptable and fulfill university and department style requirements; (2) its illustrative materials including figures, tables, and charts are in place; and (3) the final manuscript is satisfactory to the graduate committee and is ready for submission to the university library.

---

Date

---

G. Bruce Schaalje  
Chair, Graduate Committee

Accepted for the Department

---

Scott D. Grimshaw  
Graduate Coordinator

Accepted for the College

---

Thomas W. Sederberg  
Associate Dean, College of Physical and  
Mathematical Sciences



## ABSTRACT

### SELECTING THE BEST LINEAR MIXED MODEL USING PREDICTIVE APPROACHES

Jun Wang

Department of Statistics

Master of Science

The linear mixed model is widely implemented in the analysis of longitudinal data. Inference techniques and information criteria are available and well-studied for goodness-of-fit within the linear mixed model setting. Predictive approaches such as  $R_{adj}^2$ , PRESS, and  $CCC_{adj}$  are available for the linear mixed model but require more research (Edward, 2005). This project used simulation to investigate the performance of  $R_{adj}^2$ , PRESS,  $CCC_{adj}$ , Pseudo F-test and information criterion for goodness-of-fit within the linear mixed model framework. Marginal and conditional approaches for these predictive statistics were studied under different variance-covariance structures.

For compound symmetry structure, the success rates for all 17 statistics (marginal and conditional PRESS, CCC, and  $R^2$ , F test, AIC and BIC) were high. The study suggested using marginal rather than conditional residuals for PRESS, CCC and  $R^2$ . It suggested using REML likelihood function which has the term  $\frac{1}{2} \log \left| \sum_{i=1}^m X_i' X_i \right|$  for AIC and BIC. For CCC,  $R^2$ , and the information criterion, there was no difference for the various parameter number adjustments.



For autoregressive order 1 plus random effect, the study suggested using conditional residuals for PRESS, marginal residuals for CCC and  $R^2$ , and using REML function with the term  $\frac{1}{2} \log \left| \sum_{i=1}^m X_i' X_i \right|$  for AIC and BIC. Also there was no difference for the different parameter number adjustments.

The F-test performed well for all covariance structures. The study also indicated that characteristics of the data, such as the covariance structure, parameter values, and sample size, can greatly impact performance of various statistics. No one criterion is consistently better than the others in terms of selection performance in the simulation study.

## Acknowledgements

I would like to thank my advisor, Dr. Schaalje for his time and effort throughout this project. I am also grateful for the members of my graduate committee, and the Department of Statistics for all the help and support. Additionally, I would like to thank my family for all the love, support and encouragement they have given me these years.



# Contents

## Chapter

1	Introduction .....	1
2	Literature Review .....	3
	2.1. Linear Mixed Models for Repeated measures .....	3
	2.2. Model selection in linear mixed models .....	4
	2.2.1 AIC and BIC .....	5
	2.2.2 Pseudo F-tests and Likelihood Ratio Tests .....	7
	2.2.3 Predictive Approaches .....	8
3	Methodology .....	12
	3.1. Data Generation .....	12
	3.2. Modeling the Data .....	14
4	Results .....	18
	4.1. Statistics .....	18
	4.2. Sample Size .....	20
	4.3. Covariance Structure .....	20
5	Conclusions .....	34
6	Bibliography.....	36
	Appendix A: Sample SAS Code .....	38

## List of Tables

3.1: Parameter Values for Covariance Structures Used In the Simulations .....	13
3.2: Values of $\beta$ Used In the Simulation .....	15
3. 3: Example of SAS PROC MIXED code.....	15
4.1: Success Rate of Choosing the Correct Model Using Different Statistics for Covariance Structure=CS (0.5) .....	21
4.2: Success Rate of Choosing the Correct Model Using Different Statistics for Covariance Structure=CS (0.9) .....	23
4.3: Success Rate of Choosing the Correct Model Using Different Statistics for Covariance Structure=ARRE (0.5) .....	25
4.4: Success Rate of Choosing the Correct Model Using Different Statistics for Covariance Structure=ARRE (0.9) .....	27
4.5: Average Success Rate of Choosing the Correct Model Using Different Statistics .....	33

## List of Figures

Figure 1: Correct model selection rates for 17 criteria for the compound symmetry structure involving within subject correlation of 0.5.....	29
Figure 2: Correct model selection rates for 17 criteria for the compound symmetry structure involving within subject correlation of 0.9.....	30
Figure 3: Correct model selection rates for 17 criteria for the autoregressive order 1 plus random effect structure with autocorrelation of 0.5.....	31
Figure 4: Correct model selection rates for 17 criteria for the autoregressive order 1 plus random effect structure with autocorrelation of 0.9.....	32



# Chapter 1

## Introduction

The linear mixed model is a powerful tool for the analysis of longitudinal data. Longitudinal data involves repeated measurements on the same individual or experimental unit. For the purposes of estimation and inference, the linear mixed model extends the ordinary linear model with independently and identically distributed Gaussian errors to a wide variety of correlated Gaussian data.

Information criteria such as AIC (Akaike, 1973) and BIC (Schwarz, 1978) are available as goodness-of-fit statistics for model selection in the linear mixed model setting (Ferron et al., 2002; Gomez et al., 2005). Pseudo F-tests for linear mixed model selection are also available using various small sample approximations. The performance of these tests for complex mixed linear models was studied by Schaalje et al. (2002).

Predictive criteria such as the  $R^2$ , PRESS (Allen, 1974), and CCC (Lin, 1989) statistics have been well established in model selection for standard linear models but have received little study for linear mixed models (Edwards et al., 2005). However, many of these statistics can be adapted to the mixed model setting (Vonesh et al., 1996). For

example,  $R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2}$  is defined in linear regression as the proportion of total

variation in  $y$  that is explained by the model.  $R^2$  can be modified in the longitudinal setting by using either marginal or conditional predicted values (Schabenberger, 2004).

CCC, an alternative measure of agreement, was modified by Vonesh et al. (1992) for use



in mixed-effects regression settings. The PRESS statistic, a cross-validated measure of goodness-of-fit, was modified for the mixed linear model setting by Schabenberger (2004) and Christensen et al. (1992).

This study uses simulations to investigate and compare the performance of AIC, BIC, pseudo  $F$ ,  $R^2$ , PRESS and CCC for model selection within the linear mixed model framework. Marginal and conditional versions for these statistics (Vonesh et al., 1996) are studied under different variance-covariance structures.

## Chapter 2

### Literature Review

#### 2.1. Linear Mixed Models for Repeated Measures

In the context of repeated measures data, let  $y_i$  be an  $(n_i \times 1)$  vector of observed data on the  $i^{th}$  subject where  $i \in \{1, \dots, m\}$ , and  $m$  is the number of independent units (subjects). The mixed model can be written as:

$$y_i = X_i \beta + Z_i u_i + \varepsilon_i \quad , \quad (1)$$

where  $X_i$  is an  $(n_i \times p)$  known constant design matrix for the  $i^{th}$  subject,  $\beta$  is a  $(p \times 1)$  vector of unknown coefficients of the fixed effects,  $Z_i$  is a  $(n_i \times q)$  known constant design matrix for the  $i^{th}$  subject with rank  $q$ ,  $u_i$  is a  $(q \times 1)$  vector of unknown individual-specific random effects, and  $\varepsilon_i$  is a  $(n_i \times 1)$  vector of random errors (Gurka, 2006). The vector of coefficients of the random effects,  $u_i$ , is assumed to follow the normal distribution with mean vector 0 and covariance matrix  $G$ .  $\varepsilon_i$  is assumed to follow a normal distribution with mean vector 0 and covariance matrix  $R_i$ . The vectors  $u_i$  and  $\varepsilon_i$  are assumed independent; consequently,  $\text{cov}(u_i, \varepsilon_i) = 0$  and  $\text{var}(y_i) = Z_i G Z_i' + R_i = V_i$ .

Let  $N = \sum_{i=1}^m n_i$ , and denote the unique parameters of  $R_i$  and  $G$  by the  $(k \times 1)$  vector  $\theta$ .

The log-likelihood function for the linear mixed model is then

$$l_{ML}(\beta, \theta) = -\frac{N}{2} \log(2\pi) - \frac{1}{2} \sum_{i=1}^m \log(V_i(\theta)) - \frac{1}{2} \sum_{i=1}^m (y_i - X_i \beta)' V_i(\theta)^{-1} (y_i - X_i \beta) \quad . \quad (2)$$

The maximum likelihood estimator (MLE) is produced by numerically maximizing  $l_{ML}(\beta, \theta)$  with respect to the unknown parameters. Since the MLE of  $\theta$  is biased (Gurka, 2006), the Restricted Maximum Likelihood estimator (REML) of  $\theta$  is typically used. Assuming that  $\text{rank}(X_1' X_2' \cdots X_m')$  is  $p$ , the REML estimator of  $\theta$  is calculated by maximizing the likelihood function of a transformation of the original  $N$  observations to a new set of  $N - p$  observations (Gurka, 2006). The transformation is constructed such that the restricted log-likelihood function can be written as

$$l_{REML}(\theta) = \frac{N-p}{2} \log(2\pi) + \frac{1}{2} \log \left| \sum_{i=1}^m X_i' X_i \right| - \frac{1}{2} \sum_{i=1}^m \log |V_i(\theta)| - \frac{1}{2} \log \left| \sum_{i=1}^m X_i' V_i(\theta)^{-1} X_i \right| - \frac{1}{2} \sum_{i=1}^m (y_i - X_i \hat{\beta})' V_i(\theta)^{-1} (y_i - X_i \hat{\beta}) \quad , \quad (3)$$

where  $\hat{\beta}$  is

$$\hat{\beta} = \left( \sum_{i=1}^m X_i' \hat{V}_i(\hat{\theta})^{-1} X_i \right)^{-1} \sum_{i=1}^m X_i' \hat{V}_i(\hat{\theta})^{-1} y_i \quad , \quad (4)$$

and  $\hat{V}_i$  is the estimated covariance matrix  $V_i$ . In practice, iterative procedures are used to maximize (2) or (3).

It is interesting to note that SAS PROC MIXED (2003) computes the REML estimates with a version of equation (3) that leaves out the term  $\frac{1}{2} \log \left| \sum_{i=1}^m X_i' X_i \right|$  (Gurka, 2006). The same estimates are obtained, but the information criteria (to be discussed in section 2.2) are different. The log-likelihood function used by SAS will be denoted  $l_{REML2}$ .

## 2.2 Model selection in linear mixed models

### 2.2.1 AIC and BIC

Information criteria such as AIC (Akaike, 1973), BIC (Schwarz, 1978), AICC (Hurvich Tsai, 1989), and CAIC (Bozdogan, 1987) are available for selecting the covariance structure ( $Z_i$ ,  $G$ , and  $R_i$ ) and the mean model ( $X_i$  and  $\beta$ ) in linear mixed models. In general, these information criteria are functions of both the maximized log-likelihood for a given model ( $l$ ) and a penalty term based on the number of parameters ( $s$ ) in the model (Gurka, 2006). The usual formulas for AIC and BIC are:

$$AIC = l - s \quad , \quad (5)$$

$$BIC = l - s(\log N^*)/2 \quad , \quad (6)$$

where  $N^*$  is a function of  $N$ ,  $p$  or  $m$  (usually  $N^* = N$  or  $m$  under ML, and  $N^* = N - p$  or  $m$  under REML), and  $s = p + k$ . When using ML estimation, generally  $s = p + k$ , the total number of parameters in the model. However, under REML estimation,  $s = k$ , the number of covariance parameters only (Gurka, 2006). A larger value of the information criterion for a given model indicates a preference for that model.

Keselman et al. (1998) compared the effectiveness of AIC and BIC for detecting various population covariance structures. In particular, the criteria were compared in unbalanced nonspherical repeated measures designs having equal/unequal group sizes, various covariance matrices, and normally/nonnormally distributed responses. Their results show that neither approach uniformly selected the correct covariance structure. Indeed, for most of the structures investigated in the study, AIC and particularly BIC more frequently picked the wrong covariance structure than the correct covariance structure. Keselman's results show that AIC chose the correct covariance structure 47%

of the time while BIC was right 35% of the time. The negative results obtained for BIC could be due to the severe penalty imposed for the number of parameters (Keselman et al., 1998).

Ferron et al. (2002) also studied the performance of AIC and BIC for choosing a covariance structure. Data were generated following an AR(1) structure with different sample sizes, numbers of repeated measures, and levels of autocorrelation. AIC and BIC were computed for the true covariance structure, AR(1), and for the independent structure,  $\sigma^2 I$ . The results show that AIC was more successful than BIC for every combination of sample size, number of repeated measures, and level of autocorrelation. AIC chose the right covariance structure 79% of the time versus 66% for BIC. AIC and BIC both performed better when the sample size, the number of the repeated measures, and the level of autocorrelation were higher. The number of the repeated measures was the most influential factor and the effect of sample size was greater when the number of repeated measures was smaller.

Gomez et al. (2005) studied the performance of AIC and BIC criteria in selecting the true covariance structure. They generated data using 15 covariance structures. The results show that the success rate of AIC and BIC in choosing the correct covariance structure depends greatly on the sample size and complexity of the covariance structures. Success rates were generally low, but they were higher for larger sample sizes and simpler covariance structures. Pairing (a positive pairing refers to the case in which the largest sample size was associated with the covariance matrix containing the largest values) had only a small effect on the performance of AIC and BIC.

Gurka (2006) compared the performance of AIC and BIC for  $l_{ML}$ ,  $l_{REML}$ , and  $l_{REML2}$  (see section 2.1). He assessed linear mixed model selection under three scenarios: (1) selection of the fixed effects when the covariance structure is known, (2) selection of the random effects when the fixed effects are known, and (3) simultaneous selection of the fixed and random effects. The results show that all versions of the two criteria selected the proper random effects structure over 90% of the time, no matter the true variance or correlation. BIC performed slightly better than AIC. The results also showed that BIC performed better overall when using  $l_{REML}$ , while for AIC  $l_{REML2}$  selection was superior to  $l_{REML}$  as well as  $l_{ML}$ .

Keselman et al. (1998) investigated the performance of AIC and BIC in the context of covariance structures common in social science. Their results provide an interesting comparison to those reported by Gurka (Keselman et al. 2006). The differences between Gurka's results and Keselman's results show that the performance of AIC and BIC depends on the conditions investigated.

### 2.2.2 Pseudo F-tests and Likelihood Ratio Tests

If  $C$  is a contrast matrix of rank  $q$ , a commonly used test statistic for  $H_0: C\beta = 0$  is the Wald F statistic  $F_{q,ddf} = W / q$ , where

$$W = (C\hat{\beta})'(C(X'\hat{V}^{-1}X)^{-1}C')^{-1}(C\hat{\beta}) \quad (7)$$

and ddf is the denominator degrees of freedom. Under the null hypothesis, the Wald  $F$  approximately follows the  $F$  distribution with  $q$  and ddf degrees of freedom (Schaalje et al., 2002). There are two commonly used methods to compute ddf for the Pseudo  $F$ -test. The Fai and Cornelius (FC) method computes the ddf using spectral decomposition of the

approximate covariance matrix for the contrasts, together with repeated application of a method for single degree of freedom tests (Fai and Cornelius, 1996; Giesbrecht and Burns, 1985; Schaalje et al., 2002). Kenward and Roger (1997) suggested an improved method (KR) to calculate the Pseudo F-test. The KR method adjusts the estimated covariance matrix of the parameter estimates, computes a scale adjustment to the test statistic and computes the degrees of freedom of the Pseudo  $F$  test.

The performance of these tests for selecting the model for the means (“mean structure”) in complex mixed linear models was studied by Schaalje et al. (2002). The result shows that Type I error rates of both methods are affected by covariance structure complexity, sample size, and imbalance. The KR method performs well in situations with complicated covariance structures when sample sizes are moderate to small and the design is balanced. The KR method is preferable to the FC method, although it also inflates Type I error rates for complex covariance structures combined with small sample sizes.

Gomez et al. (2005) investigated Type I error rates after using AIC and BIC to select the covariance structures. The results show that Type I error rates of KR method hypothesis tests for best AIC and best BIC models were always higher than the target values. Type of selection criterion, number of subjects per treatment, number of repeated measures, effect tested, covariance structure, and pairing all affected the Type I error rates.

### 2.2.3 Predictive Approaches

There are two main distinctions between mixed models and linear models. First, in a mixed model, the data can be considered in a conditional and a marginal sense. The conditional distribution of  $y|u$  uses a particular realization of the random effects. The marginal distribution of  $y$  uses quantities averaged over all possible values of the random effects. Correspondingly, there are conditional and marginal predictive approaches for mixed models. Second, the estimates of the fixed effects  $\beta$  depend on the estimates of the covariance parameters (Schabenberger, 2004). Predictive approaches for model selection in mixed models reflect these distinctions.

$R^2$ , the squared correlation coefficient, is defined in the simple linear model as

$$R^2 = 1 - \frac{\sum_{i=1}^n r_i^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad (8)$$

where  $r_i = y_i - \hat{y}_i$ ,  $y_i$  is the  $i$ th observation,  $\hat{y}_i$  is the  $i$ th predicted value and  $\bar{y}$  is the grand mean. The squared correlation coefficient can also be calculated in the mixed model setting, but differs depending on whether marginal residuals  $r_m$  or conditional residuals  $r_c$  are used (Vonesh, 1996). A marginal residual,  $r_{mi} = y_i - x_i' \hat{\beta}$ , is the difference between the observed data and the estimated (marginal) mean. A conditional residual,  $r_{ci} = y_i - x_i' \hat{\beta} - z_i' \hat{u}$ , is the difference between the observed data and the predicted value of the observation. Mixed model versions of  $R^2$  can be adjusted for the number of parameters to obtain:

$$R_{adj}^2 = 1 - k^* (1 - R^2), \quad (9)$$



where  $k^* = \frac{N}{N-p}$  or  $k^* = \frac{N}{N-(p+k)}$ .

The PRESS residual (Allen 1974) is the difference between the observed value and the predicted value, where the predicted value is obtained without the observation in question. The PRESS statistic for the simple linear model can be computed without re-fitting the model by using the update formula (Schabenberger, 2004). In the longitudinal setting such update formulas are not available because the observations are correlated (Schabenberger, 2004). The PRESS statistic can also be calculated for mixed models, but differs depending on whether the marginal or conditional residual is used.

The concordance correlation coefficient (CCC) was described by Lin (1989) and modified by Vonesh (1992) for use in mixed-effects regression settings. Given bivariate measurements  $Y_1$  and  $Y_2$  with means  $\mu_1$ ,  $\mu_2$ , variances  $\sigma_1^2$ ,  $\sigma_2^2$ , and correlation  $\rho$ , Lin (1989) defined the concordance correlation between  $Y_1$  and  $Y_2$  to be

$$\rho_c = 1 - \frac{E[(Y_1 - Y_2)^2]}{\sigma_1^2 + \sigma_2^2 + (\mu_1 - \mu_2)^2} = \frac{2\rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 + (\mu_1 - \mu_2)^2}. \quad (10)$$

CCC is a measure of the degree of accuracy and precision to which pairs of observations correspond to the line of identity. The point estimate of  $\rho_c$  based on a random sample of  $n$  observations is

$$\hat{\rho}_c = 1 - \frac{\sum_{i=1}^n (Y_{i1} - Y_{i2})^2}{\sum_{i=1}^n (Y_{i1} - \bar{Y}_1)^2 + \sum_{i=1}^n (Y_{i2} - \bar{Y}_2)^2 + n(\bar{Y}_1 - \bar{Y}_2)^2}. \quad (11)$$

Vonesh (1992) suggested adapting CCC to measure model fit. He defined the “model CCC” as

$$r_c = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)'(y_i - \hat{y}_i)}{\sum_{i=1}^n (y_i - \bar{y}_j)'(y_i - \bar{y}_j) + \sum_{i=1}^n (\hat{y}_i - \hat{y}_j)'(\hat{y}_i - \hat{y}_j) + N(\bar{y} - \hat{y})^2}. \quad (12)$$

There are several advantages of using  $r_c$  as a measure of model fit. First,  $r_c$  is directly interpretable as a concordance correlation coefficient between observed and predicted values. Second, it does not require specification of a null model since the line of identity serves as a point of reference. In the mixed model setting, we can compute a marginal or conditional model CCC by using marginal or conditional estimates (Vonesh, 1996). Also, CCC can be adjusted for the number of parameters as

$$CCC_{adj} = 1 - k^*(1 - CCC), \quad (13)$$

where  $k^* = \frac{N}{N - p}$  or  $k^* = \frac{N}{N - (p + k)}$ .

## Chapter 3

### Methodology

This chapter describes the methods used for the simulation study. First, data were generated with a specified model for the means and with repeated measurements on the same individual, correlated according to some known covariance structure. Then, two mixed models (a full model and a reduced model) were fit to the data, and several predictive approaches were used to select one of the models. The proportion of runs in which the correct model was selected was recorded.

An explanation of the data generation process is outlined in the first section of this chapter. The second section discusses the use of simulations to investigate the performance of  $R_{adj}^2$ , PRESS,  $CCC_{adj}$ , information criteria, and the F test for model selection. Marginal and conditional approaches for these statistics were studied under different variance-covariance structures. Data were simulated using PROC IML of SAS and analyzed with PROC MIXED.

#### 3.1. Data Generation

The simulated data followed a repeated measures design. Each data set involved  $m=10$  or  $20$  subjects, and a constant number of observations per subject ( $n_i = n = 3$  or  $5$  for  $i = 1$  to  $m$ ). The repeated measures followed one of four covariance structures, and the fixed effects followed one of six possible models. Thus,  $2 \times 2 \times 4 \times 6 = 96$  simulation studies were carried out.

The covariance structures were the same for each subject ( $V_i = V$  for  $i = 1$  to  $m$ ). The four structures are denoted CS (0.5), CS (0.9), ARRE (0.5), and ARRE (0.9). Details on these covariance structures when the number of repeated measures was five are in Table 1. The upper left  $3 \times 3$  matrix was used when the number of repeated measures was three.

Table 3.1: Parameter Values for Covariance Structures Used In the Simulations

---

Compound symmetry (CS (0.5))	Compound symmetry (CS (0.9))
$\begin{pmatrix} 1 & & & & \\ .5 & 1 & & & \\ .5 & .5 & 1 & & \\ .5 & .5 & .5 & 1 & \\ .5 & .5 & .5 & .5 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & & & & \\ .9 & 1 & & & \\ .9 & .9 & 1 & & \\ .9 & .9 & .9 & 1 & \\ .9 & .9 & .9 & .9 & 1 \end{pmatrix}$
Autoregressive order 1 plus random effect (ARRE( $\rho=0.5$ ))	Autoregressive order 1 plus random effect (ARRE( $\rho=0.9$ ))
$\begin{pmatrix} 1 & & & & \\ .6 & 1 & & & \\ .4 & .6 & 1 & & \\ .3 & .4 & .6 & 1 & \\ .25 & .3 & .4 & .6 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & & & & \\ .92 & 1 & & & \\ .848 & .92 & 1 & & \\ .7832 & .848 & .92 & 1 & \\ .72488 & .7832 & .848 & .92 & 1 \end{pmatrix}$

---

Data for the four described covariance structures were generated following the two step method of Ripley (1987). In the first step, a random multivariate normal vector was generated with  $E(y) = 0$  and  $\text{Var}(y) = I$ , where  $I$  is the identity matrix. In the second

step, this vector was multiplied by the Cholesky decomposition of the covariance matrix corresponding to the covariance structure in question. The resulting random vector had a mean vector of zero and covariance matrix of  $V$  (Table 1).

### 3.2. Modeling the Data

The fixed effects part of the models were defined by a combination of design matrices ( $X_i = [X_0 \ X_1 \ X_2]$ ) and parameter vectors ( $\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}$ ). In the studies with

$n=5$ ,  $X_i$  was the same for all subjects, and was given by  $X_i = \begin{pmatrix} 1 & -3 & 1 \\ 1 & -2 & 2 \\ 1 & 0 & 3 \\ 1 & 2 & 1 \\ 1 & 3 & 2 \end{pmatrix}$ . In the

studies with  $n=3$ ,  $X_i = \begin{pmatrix} 1 & -3 & 0 \\ 1 & 0 & 1 \\ 1 & 3 & 0 \end{pmatrix}$ . These design matrices were combined with values

of the parameter vector given in Table 2. For  $\beta_2 = 0.6$ , some of the expected values change by as much as two standard deviations.

Two mixed models were fitted to each data set using REML as implemented in PROC MIXED. For both models, the correct covariance structure was specified using RANDOM and REPEATED statements. Also, two models for the fixed effects were specified using the MODEL statement, the full model, and the reduced model. For the full model, the MODEL statement was specified as  $y = X_1 \ X_2$ , whereas for the reduced model the MODEL statement was  $y = X_1$ .

Table 3.2: Values of  $\beta$  Used In the Simulation

$$\beta = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}$$

$$\beta = \begin{pmatrix} 1 \\ 1 \\ .05 \end{pmatrix}$$

$$\beta = \begin{pmatrix} 1 \\ 1 \\ .1 \end{pmatrix}$$

$$\beta = \begin{pmatrix} 1 \\ 1 \\ .2 \end{pmatrix}$$

$$\beta = \begin{pmatrix} 1 \\ 1 \\ .4 \end{pmatrix}$$

$$\beta = \begin{pmatrix} 1 \\ 1 \\ .6 \end{pmatrix}$$

Table 3.3: Example of SAS PROC MIXED code

---

```

*SAS code for compound symmetry structure;
proc mixed data=datanew;
by iter;
class subj;
model y=x1 x2/ S influence(iter=5 est) outpred=pred_full(keep=y iter x1
x2 subj Pred Resid)
outpredm=predm_full(keep=y iter x1 x2 subj Pred Resid)
DDFM=kenwardroger;
random subj/s;
ods output fitstatistics=cs;
ods output tests3=pval_cs;
ods output SolutionF=coef_fixed (keep=estimate iter) ;
ods output SolutionR=coef_random;
ods output influence=press (keep=iter pressres); run;

*SAS code for autoregressive order 1 plus random effect structure;
proc mixed data=datanew;
by iter;
class subj;
model y=x1 x2/ S influence(iter=5 est) outpred=pred_full(keep=y iter x1
x2 subj Pred Resid)
outpredm=predm_full(keep=y iter x1 x2 subj Pred Resid)
DDFM=kenwardroger;
random subj/s;
repeated /type=ar(1) subject=subj;
ods output fitstatistics=arre;
ods output tests3=pval_arre;
ods output SolutionF=coef_fixed (keep=estimate iter) ;
ods output SolutionR=coef_random;
ods output influence=press (keep=iter pressres); run;

```

---

Several versions of PRESS,  $CCC_{adj}$ ,  $R_{adj}^2$ , the pseudo F-test and information criteria (AIC and BIC) were calculated and recorded for each model. Some of the information criteria and the F-test were given by default by PROC MIXED. The Kenward-Roger method for the F-test was specified by using the `ddfm=kenwardroger` option in the MODEL statement. For PRESS,  $CCC_{adj}$  and  $R_{adj}^2$ , both marginal and conditional residuals were considered. The marginal residual  $r_m = y_i - x_i' \hat{\beta}$  is the difference between the observed data and the estimated (marginal) mean. The conditional residual  $r_c = y_i - x_i' \hat{\beta} - z_i' \hat{u}$  is the difference between the observed data and the predicted value of the observation. Marginal and conditional residuals were given by using `outpredm` and `outpred` options in the model statement of PROC MIXED. The ODS (output delivery system) was used to capture the fitted statistics and the residuals. PRESS,  $CCC_{adj}$  and  $R_{adj}^2$  were then calculated using PROC IML.

The PRESS statistic is the sum of the squared differences between the observed values and the predicted values, where each predicted value is obtained without the observations in question. Two versions of PRESS were calculated, the marginal and conditional PRESS. A PROC MIXED option was used to compute the marginal PRESS residuals. A SAS MACRO (Appendix A) was used to calculate the conditional PRESS statistics.

$CCC_{adj}$  was calculated as

$$CCC = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)' (y_i - \hat{y}_i)}{\sum_{i=1}^n (y_i - \bar{y}_j)' (y_i - \bar{y}_j) + \sum_{i=1}^n (\hat{y}_i - \hat{y}_j)' (\hat{y}_i - \hat{y}_j) + N(\bar{y} - \hat{y})^2} \quad \text{and}$$

$CCC_{adj} = 1 - k^*(1 - CCC)$  where  $k^* = \frac{N}{N - s}$  and  $s$  is the number of parameters in the model. Four versions of  $CCC_{adj}$  were calculated: marginal  $CCC_{adj}$  with  $s = p$ , conditional  $CCC_{adj}$  with  $s = p$ , marginal  $CCC_{adj}$  with  $s = p + k$ , and conditional  $CCC_{adj}$  with  $s = p + k$ .

The equations for adjusted  $R^2$  are:  $R^2 = 1 - \frac{\sum_{i=1}^n r_i^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2}$  and  $R_{adj}^2 = 1 - k^*(1 - R^2)$

where  $k^* = \frac{N}{N - s}$ . Four versions of  $R_{adj}^2$  were calculated: marginal  $R_{adj}^2$  with  $s = p$ , conditional  $R_{adj}^2$  with  $s = p$ , marginal  $R_{adj}^2$  with  $s = p + k$ , and conditional  $R_{adj}^2$  with  $s = p + k$ .

For AIC and BIC, the smaller-is-better formulas were used:  $AIC = -2l + 2s$  and  $BIC = -2l + s(\log N^*)$ , where  $l$  is the restricted log-likelihood function with or without the term  $\frac{1}{2} \log |\sum_{i=1}^m X_i' X_i|$ ,  $s = p + k$  and  $N^* = m$  or  $N^* = N - p$ . Therefore, two versions of AIC and four versions of BIC were considered.

Seventeen goodness-of-fit statistics were calculated for each iteration and used to select either the full or reduced model. The number of times these criteria selected the true model (full model if  $\beta_2 > 0$ , reduced model if  $\beta_2 = 0$ ) was recorded in each situation. Rates of success of these criteria were then compared.



## Chapter 4

### Results

Correct model selection rates were tabulated for the seventeen criteria (Tables 3-6) and displayed graphically (Figures 1-4). For all statistics, numbers of subjects, numbers of measurements, and covariance structures, the success rate was high for  $\beta_2 = 0$ . Overall, the statistics performed better when the number of subjects and/or the number of measurements increased. The conditional statistics performed better than the marginal statistics overall (Table 7), but not for  $\beta_2 = 0$ . On average, the first type of CCC and  $R^2$  (calculated using  $s = p$ ) performed slightly better than the second type (calculated using  $s = p + k$ ) (Table 7). The criterion of better performance was higher success rate for  $\beta_2 = 0$  and faster increase rate (steeper slope) as  $\beta_2$  increased from 0.05 to 0.6.

#### 4.1. Statistics

PRESS statistics performed better for ARRE than for CS. When the covariance structure was CS, the performance of marginal and conditional PRESS was similar, but when the covariance structure was ARRE, conditional PRESS performed much better than marginal PRESS.

For CCC, adjustment 1 (calculated using  $s = p$ ) was the same as adjustment 2 (calculated using  $s = p + k$ ). Marginal CCC was better than conditional CCC for ARRE (1). For CS, marginal CCC was better in that conditional too often chose the more complex model.

$R^2$  performance was very similar to CCC: adjustment 1 and adjustment 2 were the same and marginal  $R^2$  was better than conditional  $R^2$ . Conditional PRESS was better than any CCC or  $R^2$ .  $R^2$  and CCC (especially marginal) had a “bounce” at  $\beta_2 = 0.1$  for subject=10 and observation=3 (small sample size).  $R^2$  and CCC conditional performed very poorly for observation=3 for ARRE(1).

For AIC, AIC1 (with the term  $\frac{1}{2} \log | \sum_{i=1}^m X_i' X_i |$ ) performed better for  $\beta_2 = 0$ , especially for the lower correlation. AIC2 (without the term  $\frac{1}{2} \log | \sum_{i=1}^m X_i' X_i |$ ) was better for  $\beta_2 > 0$ . AIC2 was similar to PRESS for the higher correlation, and worse than PRESS for the lower correlation.

F-test with KR achieved 95% for CS when  $\beta_2 = 0$ , and slightly more than 95% for ARRE(1) for  $\beta_2 = 0$ . The F-test was similar to PRESS for CS, but not for ARRE(1). The F-test was similar to AIC1 but lower than AIC2 for  $\beta_2 > 0$ .

All four kinds of BIC were different. BIC1 (with the term  $\frac{1}{2} \log | \sum_{i=1}^m X_i' X_i |$  and  $N^* = N - p$ ) and BIC2 (with the term  $\frac{1}{2} \log | \sum_{i=1}^m X_i' X_i |$  and  $N^* = m$ ) were best for  $\beta_2 = 0$ , and BIC3 (without the term  $\frac{1}{2} \log | \sum_{i=1}^m X_i' X_i |$  and  $N^* = N - p$ ) and BIC4 (without the term  $\frac{1}{2} \log | \sum_{i=1}^m X_i' X_i |$  and  $N^* = m$ ) were best for  $\beta_2 > 0$ . BIC1 and BIC2 were quite similar to each other. BIC4 was better than BIC3 for  $\beta_2 > 0$ . For  $\beta_2 > 0$ ,

BIC4 was better than BIC3, BIC3 was better than BIC2, and BIC2 was better than BIC1. The opposite was true for  $\beta_2 = 0$ .

#### 4.2. Sample Size

All 17 statistics performed much better for subjects=20 and observations=5 per subject than subjects=10 and observations=3 per subject. The performance for subjects=20 and observations=5 was similar for subjects=20 and observations=3, and better than that for subjects=10 and observations=3. All statistics performed better for larger numbers of subjects and performed much better for larger numbers of observations.

#### 4.3. Covariance Structure

$R^2$ , CCC, and PRESS performed better for Compound Symmetry than for Autoregressive order 1 plus random effect. Other statistics (AIC, BIC, F-test) were not greatly affected. All statistics performed better for higher correlations.

Covariance Structure	Parameter	# of subject	# of obs	beta	Marginal Press	Conditional Press	Conditional CCC	Marginal CCC	Conditional R <sup>2</sup>	Marginal R <sup>2</sup>	Conditional CCC_2	Marginal CCC_2	Conditional R <sup>2</sup> _2
CS	0.5	10	3	0	0.870800	0.8321	0.6165	0.8112	0.6213	0.8246	0.6298	0.8269	0.6348
CS	0.5	10	3	0.05	0.136400	0.1737	0.3879	0.1969	0.3842	0.1853	0.3763	0.1833	0.3723
CS	0.5	10	3	0.1	0.146100	0.1893	0.4114	0.2113	0.4080	0.1968	0.3999	0.1942	0.3958
CS	0.5	10	3	0.2	0.214500	0.2617	0.4942	0.2893	0.4900	0.2731	0.4804	0.2713	0.4765
CS	0.5	10	3	0.4	0.452900	0.5107	0.7230	0.5521	0.7201	0.5308	0.7141	0.5288	0.7113
CS	0.5	10	3	0.6	0.720000	0.7684	0.8979	0.7928	0.8964	0.7783	0.8931	0.7783	0.8920
CS	0.5	10	5	0	0.864800	0.8345	0.6315	0.8141	0.6378	0.8293	0.6402	0.8227	0.6491
CS	0.5	10	5	0.05	0.161500	0.1886	0.3944	0.2176	0.3881	0.2009	0.3854	0.2091	0.3791
CS	0.5	10	5	0.1	0.234100	0.2679	0.4787	0.2972	0.4719	0.2790	0.4708	0.2861	0.4644
CS	0.5	10	5	0.2	0.482700	0.5241	0.7184	0.5612	0.7139	0.5400	0.7130	0.5508	0.7074
CS	0.5	10	5	0.4	0.919400	0.9347	0.9792	0.9480	0.9784	0.9406	0.9778	0.9442	0.9771
CS	0.5	10	5	0.6	0.998500	0.9987	0.9999	0.9992	0.9999	0.9991	0.9999	0.9992	0.9999
CS	0.5	20	3	0	0.883400	0.8344	0.6203	0.8152	0.6246	0.8307	0.6263	0.8231	0.6307
CS	0.5	20	3	0.05	0.128835	0.1807	0.4042	0.1993	0.3992	0.1841	0.3973	0.1927	0.3925
CS	0.5	20	3	0.1	0.161910	0.2216	0.4500	0.2413	0.4468	0.2237	0.4448	0.2320	0.4403
CS	0.5	20	3	0.2	0.296760	0.3659	0.5959	0.3882	0.5930	0.3713	0.5913	0.3801	0.5878
CS	0.5	20	3	0.4	0.674975	0.7357	0.8800	0.7595	0.8786	0.7435	0.8776	0.7518	0.8760
CS	0.5	20	3	0.6	0.933150	0.9526	0.9855	0.9588	0.9850	0.9540	0.9851	0.9562	0.9845
CS	0.5	20	5	0	0.870900	0.8354	0.6322	0.8176	0.6404	0.8348	0.6374	0.8221	0.6447
CS	0.5	20	5	0.05	0.187300	0.2261	0.4367	0.2506	0.4310	0.2328	0.4327	0.2461	0.4265
CS	0.5	20	5	0.1	0.323300	0.3707	0.5869	0.4020	0.5808	0.3822	0.5830	0.3968	0.5764
CS	0.5	20	5	0.2	0.704700	0.7488	0.8828	0.7704	0.8797	0.7530	0.8808	0.7657	0.8784
CS	0.5	20	5	0.4	0.996500	0.9978	0.9995	0.9985	0.9995	0.9983	0.9995	0.9984	0.9995
CS	0.5	20	5	0.6	1	1	1	1	1	1	1	1	1

Table 4.1: Success Rate of Choosing the Correct Model Using Different Statistics for Covariance Structure=Compound Symmetry (0.5)

Covariance Structure	Parameter	# of subject	# of obs	beta	Marginal R <sup>2</sup> _2	fctest	aic1	aic2	bic1	bic2	bic3	bic4
CS	0.5	10	3	0	0.8401	0.9495	0.894	0.626	0.9478	0.9123	0.839	0.7047
CS	0.5	10	3	0.05	0.1683	0.0542	0.1089	0.3736	0.0561	0.0908	0.1742	0.3018
CS	0.5	10	3	0.1	0.1803	0.0665	0.1227	0.3985	0.0647	0.104	0.1954	0.3193
CS	0.5	10	3	0.2	0.2560	0.108	0.189	0.4831	0.1127	0.1637	0.2708	0.4083
CS	0.5	10	3	0.4	0.5093	0.2842	0.4217	0.7292	0.2961	0.3883	0.5175	0.6684
CS	0.5	10	3	0.6	0.7617	0.5434	0.7003	0.8983	0.5716	0.6659	0.7826	0.87
CS	0.5	10	5	0	0.8379	0.9484	0.9567	0.6336	0.9839	0.9636	0.8906	0.7117
CS	0.5	10	5	0.05	0.1907	0.0662	0.0576	0.3906	0.0233	0.0489	0.1327	0.317
CS	0.5	10	5	0.1	0.2697	0.1126	0.0985	0.478	0.0461	0.0879	0.2043	0.4057
CS	0.5	10	5	0.2	0.5256	0.3022	0.284	0.7218	0.1625	0.2598	0.4496	0.6644
CS	0.5	10	5	0.4	0.9383	0.8275	0.8183	0.9797	0.6785	0.7989	0.912	0.9728
CS	0.5	10	5	0.6	0.9990	0.9932	0.992	1	0.9692	0.9901	0.9973	0.9998
CS	0.5	20	3	0	0.8397	0.9511	0.9346	0.64	0.9802	0.9646	0.909	0.8259
CS	0.5	20	3	0.05	0.1756	0.0583	0.0741	0.3743	0.026	0.0427	0.1047	0.1895
CS	0.5	20	3	0.1	0.2167	0.0817	0.105	0.4193	0.0366	0.0612	0.1387	0.2343
CS	0.5	20	3	0.2	0.3621	0.1734	0.2096	0.5805	0.1013	0.1438	0.2727	0.3899
CS	0.5	20	3	0.4	0.7353	0.518	0.5734	0.8772	0.3837	0.4749	0.6456	0.7616
CS	0.5	20	3	0.6	0.9516	0.8499	0.8837	0.9865	0.7634	0.8249	0.922	0.9583
CS	0.5	20	5	0	0.8397	0.9517	0.9733	0.6367	0.9944	0.9841	0.9344	0.8251
CS	0.5	20	5	0.05	0.2276	0.0828	0.0506	0.4324	0.0143	0.0314	0.1078	0.2408
CS	0.5	20	5	0.1	0.3758	0.1768	0.1233	0.58	0.0466	0.084	0.2211	0.3959
CS	0.5	20	5	0.2	0.7475	0.5394	0.4529	0.8799	0.2577	0.3627	0.603	0.7691
CS	0.5	20	5	0.4	0.9980	0.9863	0.9731	0.9996	0.9211	0.9578	0.9908	0.9984
CS	0.5	20	5	0.6	1	1	1	1	0.9999	1	1	1

Table 4.1 (continue): Success Rate of Choosing the Correct Model Using Different Statistics for Covariance Structure=CS (0.5)

\* R<sup>2</sup>\_2 and CCC\_2 are calculated using s=p + k

\* aic2, bic3 and bic4 are calculated without the term  $\frac{1}{2} \log |\sum_{i=1}^m X_i X_i|$

\* bic2 and bic4 used  $N^* = m$

Covariance Structure	Parameter	# of subject	# of obs	beta	Marginal Press	Conditional Press	Conditional CCC	Marginal CCC	Conditional R <sup>2</sup>	Marginal R <sup>2</sup>	Conditional CCC_2	Marginal CCC_2	Conditional R <sup>2</sup> _2
CS	0.9	10	3	0	0.8601	0.8303	0.5837	0.9891	0.5839	0.9903	0.5969	0.9912	0.5978
CS	0.9	10	3	0.05	0.1712	0.2031	0.4537	0.0178	0.4529	0.0159	0.4377	0.0135	0.4374
CS	0.9	10	3	0.1	0.2465	0.2963	0.5465	0.0288	0.5463	0.0255	0.5321	0.0242	0.5316
CS	0.9	10	3	0.2	0.5134	0.5903	0.8024	0.1333	0.8022	0.1189	0.7924	0.1143	0.7921
CS	0.9	10	3	0.4	0.9339	0.9602	0.9902	0.6003	0.9902	0.5621	0.9896	0.5594	0.9896
CS	0.9	10	3	0.6	0.9986	0.9991	0.9999	0.9513	0.9999	0.9338	0.9999	0.9384	0.9999
CS	0.9	10	5	0	0.837	0.8264	0.623	0.9875	0.6247	0.9892	0.633	0.9891	0.6345
CS	0.9	10	5	0.05	0.2804	0.3034	0.5215	0.0353	0.5197	0.0313	0.5111	0.0322	0.5095
CS	0.9	10	5	0.1	0.5439	0.5881	0.7834	0.1412	0.7827	0.1263	0.7773	0.1306	0.7764
CS	0.9	10	5	0.2	0.9526	0.9697	0.9918	0.6288	0.9918	0.5902	0.9914	0.6083	0.9914
CS	0.9	10	5	0.4	1	1	1	0.9996	1	0.999	1	0.9996	1
CS	0.9	10	5	0.6	1	1	1	1	1	1	1	1	1
CS	0.9	20	3	0	0.8838	0.8355	0.5871	0.9953	0.588	0.9964	0.596	0.9963	0.5961
CS	0.9	20	3	0.05	0.164642	0.2297	0.482	0.016	0.4815	0.0141	0.4755	0.0146	0.4746
CS	0.9	20	3	0.1	0.328086	0.4084	0.6551	0.0424	0.6547	0.036	0.6489	0.0388	0.6483
CS	0.9	20	3	0.2	0.74096	0.8125	0.9334	0.2555	0.9334	0.2272	0.9316	0.2402	0.9314
CS	0.9	20	3	0.4	0.999688	1	1	1	1	0.9999	1	0.9999	1
CS	0.9	20	3	0.6	0.999688	1	1	1	1	0.9999	1	0.9999	1
CS	0.9	20	5	0	0.8586	0.8366	0.6252	0.9942	0.6266	0.9954	0.6297	0.9949	0.6322
CS	0.9	20	5	0.05	0.3665	0.4181	0.6352	0.0511	0.6339	0.042	0.6311	0.049	0.6299
CS	0.9	20	5	0.1	0.7808	0.8215	0.9286	0.2829	0.928	0.2491	0.9269	0.2732	0.9264
CS	0.9	20	5	0.2	0.9981	0.9993	1	0.9437	1	0.9253	1	0.9403	1
CS	0.9	20	5	0.4	1	1	1	1	1	1	1	1	1
CS	0.9	20	5	0.6	1	1	1	1	1	1	1	1	1

Table 4.2: Success Rate of Choosing the Correct Model Using Different Statistics for Covariance Structure=Compound Symmetry (0.9)

Covariance Structure	Parameter	# of subject	# of obs	beta	Marginal R <sup>2</sup> _2	fctest	aic1	aic2	bic1	bic2	bic3	bic4
CS	0.9	10	3	0	0.9921	0.9511	0.9608	0.8817	0.9782	0.9658	0.9403	0.8998
CS	0.9	10	3	0.05	0.0122	0.0717	0.0582	0.1551	0.0306	0.049	0.0821	0.1317
CS	0.9	10	3	0.1	0.022	0.1221	0.1038	0.2386	0.0592	0.0913	0.1352	0.2077
CS	0.9	10	3	0.2	0.1014	0.3494	0.3249	0.5406	0.2311	0.2984	0.3972	0.4983
CS	0.9	10	3	0.4	0.5185	0.8655	0.8604	0.9521	0.7809	0.8411	0.9063	0.941
CS	0.9	10	3	0.6	0.9156	0.9953	0.9952	0.9994	0.9915	0.9939	0.9981	0.999
CS	0.9	10	5	0	0.9906	0.9526	0.984	0.8769	0.9944	0.9857	0.9596	0.8993
CS	0.9	10	5	0.05	0.0287	0.129	0.0566	0.2442	0.0303	0.0504	0.1134	0.2108
CS	0.9	10	5	0.1	0.1186	0.3654	0.2211	0.5301	0.1285	0.2022	0.3385	0.4909
CS	0.9	10	5	0.2	0.5682	0.8985	0.7997	0.9615	0.683	0.7782	0.8912	0.9515
CS	0.9	10	5	0.4	0.9987	1	1	1	0.9999	1	1	1
CS	0.9	10	5	0.6	1	1	1	1	1	1	1	1
CS	0.9	20	3	0	0.9968	0.9495	0.9747	0.8807	0.9919	0.9858	0.9606	0.9353
CS	0.9	20	3	0.05	0.0132	0.0912	0.0506	0.1872	0.02	0.0326	0.069	0.1114
CS	0.9	20	3	0.1	0.0322	0.2015	0.1392	0.3455	0.062	0.0934	0.1673	0.2408
CS	0.9	20	3	0.2	0.2139	0.6127	0.5044	0.7664	0.3436	0.413	0.5756	0.6673
CS	0.9	20	3	0.4	0.9998	1	1	1	0.966	1	0.9916	1
CS	0.9	20	3	0.6	0.9998	1	1	1	1	1	1	1
CS	0.9	20	5	0	0.9958	0.9511	0.9899	0.8808	0.9984	0.994	0.9776	0.9365
CS	0.9	20	5	0.05	0.0393	0.2162	0.0883	0.3569	0.0327	0.0614	0.1479	0.2546
CS	0.9	20	5	0.1	0.24	0.6453	0.4102	0.782	0.242	0.3383	0.5394	0.6883
CS	0.9	20	5	0.2	0.9212	0.9949	0.9793	0.9989	0.9446	0.9677	0.9915	0.9966
CS	0.9	20	5	0.4	1	1	1	1	1	1	1	1
CS	0.9	20	5	0.6	1	1	1	1	1	1	1	1

Table 4.2 (continue): Success Rate of Choosing the Correct Model Using Different Statistics for Covariance Structure=CS (0.9)

\* R<sup>2</sup>\_2 and CCC\_2 are calculated using s=p+k

\* aic2, bic3 and bic4 are calculated without the term  $\frac{1}{2} \log | \sum_{i=1}^m X_i' X_i |$

\* bic2 and bic4 used  $N^* = m$

Covariance Structure	Parameter	# of subject	# of obs	beta	Marginal Press	Conditional Press	Conditional CCC	Marginal CCC	Conditional R <sup>2</sup>	Marginal R <sup>2</sup>	Conditional CCC_2	Marginal CCC_2	Conditional R <sup>2</sup> _2
ARRE(1)	0.5	10	3	0	0.6663	0.8233	0.6436	0.8842	0.6553	0.895	0.6817	0.904	0.6922
ARRE(1)	0.5	10	3	0.05	0.3403	0.1937	0.3585	0.1302	0.3459	0.1198	0.3157	0.1092	0.3026
ARRE(1)	0.5	10	3	0.1	0.3506	0.2209	0.3565	0.148	0.3447	0.1347	0.3147	0.1249	0.3041
ARRE(1)	0.5	10	3	0.2	0.4056	0.3231	0.4044	0.2527	0.3912	0.2373	0.3633	0.2225	0.3494
ARRE(1)	0.5	10	3	0.4	0.5088	0.6192	0.4955	0.5713	0.4847	0.5482	0.4638	0.5312	0.4521
ARRE(1)	0.5	10	3	0.6	0.609	0.8662	0.5679	0.8505	0.5608	0.8369	0.55	0.8262	0.5434
ARRE(1)	0.5	10	5	0	0.879	0.8239	0.6929	0.8453	0.7032	0.8605	0.7079	0.857	0.7175
ARRE(1)	0.5	10	5	0.05	0.1369	0.2101	0.3267	0.1705	0.3245	0.1681	0.3144	0.1578	0.3116
ARRE(1)	0.5	10	5	0.1	0.1523	0.2987	0.3687	0.25	0.3748	0.2599	0.356	0.2344	0.3615
ARRE(1)	0.5	10	5	0.2	0.2329	0.5959	0.5154	0.5195	0.5287	0.5393	0.5031	0.5012	0.517
ARRE(1)	0.5	10	5	0.4	0.3851	0.9617	0.7503	0.9504	0.7565	0.9549	0.7466	0.9452	0.7525
ARRE(1)	0.5	10	5	0.6	0.5004	0.9988	0.8232	0.9996	0.8228	0.9996	0.8224	0.9993	0.8218
ARRE(1)	0.5	20	3	0	0.6122	0.8314	0.6028	0.8889	0.6156	0.8989	0.6189	0.8964	0.6336
ARRE(1)	0.5	20	3	0.05	0.3983	0.1845	0.4037	0.1314	0.3908	0.1198	0.3863	0.1212	0.3721
ARRE(1)	0.5	20	3	0.1	0.4148	0.2378	0.4069	0.1798	0.3942	0.1643	0.3883	0.1659	0.3747
ARRE(1)	0.5	20	3	0.2	0.4773	0.4359	0.4172	0.3625	0.4055	0.3438	0.3982	0.3462	0.384
ARRE(1)	0.5	20	3	0.4	0.581	0.8526	0.4739	0.8227	0.4695	0.8072	0.466	0.8108	0.4592
ARRE(1)	0.5	20	3	0.6	0.6434	0.9902	0.4632	0.9837	0.4654	0.981	0.461	0.9819	0.4632
ARRE(1)	0.5	20	5	0	0.897	0.8283	0.6614	0.8533	0.6715	0.8641	0.6683	0.8580	0.6787
ARRE(1)	0.5	20	5	0.05	0.1144	0.2411	0.3431	0.1757	0.3499	0.194	0.3369	0.1698	0.3425
ARRE(1)	0.5	20	5	0.1	0.1491	0.4209	0.396	0.3167	0.4161	0.3543	0.3895	0.3085	0.4101
ARRE(1)	0.5	20	5	0.2	0.2445	0.8256	0.5562	0.7329	0.579	0.7683	0.5527	0.7264	0.5745
ARRE(1)	0.5	20	5	0.4	0.3796	0.9992	0.692	0.9976	0.6989	0.9983	0.6912	0.9974	0.6985
ARRE(1)	0.5	20	5	0.6	0.4996	1	0.7935	1	0.7929	1	0.7935	1	0.7926

Table 4.3: Success Rate of Choosing the Correct Model Using Different Statistics for Covariance Structure= Autoregressive order 1 plus random effect (0.5)



Covariance Structure	Parameter	# of subject	# of obs	beta	Marginal R <sup>2</sup> _2	f <sub>test</sub>	aic1	aic2	bic1	bic2	bic3	bic4
ARRE(1)	0.5	10	3	0	0.9107	0.9722	0.9259	0.7267	0.9644	0.9384	0.8738	0.7835
ARRE(1)	0.5	10	3	0.05	0.0992	0.0309	0.0804	0.2831	0.0436	0.0671	0.1391	0.2241
ARRE(1)	0.5	10	3	0.1	0.1153	0.0352	0.0956	0.3143	0.0544	0.08	0.1607	0.2557
ARRE(1)	0.5	10	3	0.2	0.2075	0.0818	0.1802	0.4529	0.1067	0.1575	0.2604	0.3893
ARRE(1)	0.5	10	3	0.4	0.5104	0.2684	0.4799	0.7579	0.3411	0.441	0.5877	0.7108
ARRE(1)	0.5	10	3	0.6	0.8113	0.5575	0.7917	0.9454	0.6696	0.7637	0.86	0.9264
ARRE(1)	0.5	10	5	0	0.872	0.9637	0.9644	0.7006	0.9851	0.971	0.9065	0.7579
ARRE(1)	0.5	10	5	0.05	0.1556	0.053	0.0521	0.3358	0.0209	0.0449	0.127	0.2793
ARRE(1)	0.5	10	5	0.1	0.2428	0.1061	0.1072	0.4508	0.0498	0.0965	0.2024	0.3931
ARRE(1)	0.5	10	5	0.2	0.5188	0.3167	0.3263	0.7413	0.2047	0.298	0.4992	0.6946
ARRE(1)	0.5	10	5	0.4	0.9499	0.8775	0.8899	0.9913	0.7775	0.8716	0.9487	0.9877
ARRE(1)	0.5	10	5	0.6	0.9996	0.9967	0.9977	0.9999	0.9899	0.9973	0.9989	0.9999
ARRE(1)	0.5	20	3	0	0.9085	0.9672	0.952	0.7405	0.9858	0.9728	0.9248	0.8684
ARRE(1)	0.5	20	3	0.05	0.1099	0.0385	0.0573	0.2963	0.0181	0.0331	0.0829	0.1547
ARRE(1)	0.5	20	3	0.1	0.1534	0.0622	0.0894	0.3588	0.0356	0.0554	0.1256	0.2067
ARRE(1)	0.5	20	3	0.2	0.3232	0.1648	0.2231	0.567	0.1096	0.1534	0.2826	0.4008
ARRE(1)	0.5	20	3	0.4	0.7928	0.5962	0.6967	0.9196	0.4892	0.5932	0.7478	0.8468
ARRE(1)	0.5	20	3	0.6	0.9793	0.9226	0.961	0.9965	0.893	0.9289	0.9762	0.9868
ARRE(1)	0.5	20	5	0	0.8697	0.9565	0.9752	0.7058	0.9947	0.9870	0.9407	0.8476
ARRE(1)	0.5	20	5	0.05	0.1871	0.082	0.0517	0.3723	0.0137	0.0318	0.1017	0.2174
ARRE(1)	0.5	20	5	0.1	0.3461	0.1914	0.1366	0.5614	0.0552	0.0957	0.2381	0.3968
ARRE(1)	0.5	20	5	0.2	0.7617	0.6184	0.5379	0.9092	0.3314	0.4489	0.6813	0.8268
ARRE(1)	0.5	20	5	0.4	0.9981	0.9966	0.9932	0.9999	0.9719	0.9861	0.998	0.9997
ARRE(1)	0.5	20	5	0.6	1	1	1	1	1	1	1	1

Table 4.3 (continue): Success Rate of Choosing the Correct Model Using Different Statistics for Covariance Structure=ARRE(0.5)

\* R<sup>2</sup>\_2 and CCC\_2 are calculated using s=p + k

\* aic2, bic3 and bic4 are calculated without the term  $\frac{1}{2} \log | \sum_{i=1}^m X_i' X_i |$

\* bic2 and bic4 used  $N^* = m$

Covariance Structure	Parameter	# of subject	# of obs	beta	Marginal Press	Conditional Press	Conditional CCC	Marginal CCC	Conditional R <sup>2</sup>	Marginal R <sup>2</sup>	Conditional CCC_2	Marginal CCC_2	Conditional R <sup>2</sup> _2
ARRE(1)	0.9	10	3	0	0.6111	0.8206	0.7092	0.9985	0.712	0.9987	0.7347	0.9991	0.7385
ARRE(1)	0.9	10	3	0.05	0.402	0.2257	0.4145	0.3217	0.4134	0.3203	0.4057	0.3187	0.4048
ARRE(1)	0.9	10	3	0.1	0.4163	0.3867	0.2894	0.0156	0.288	0.0141	0.2643	0.0109	0.2632
ARRE(1)	0.9	10	3	0.2	0.4978	0.758	0.2992	0.0936	0.2998	0.0849	0.2835	0.0691	0.2839
ARRE(1)	0.9	10	3	0.4	0.6493	0.995	0.3729	0.6411	0.3744	0.5971	0.3689	0.5637	0.37
ARRE(1)	0.9	10	3	0.6	0.7448	0.9998	0.4155	0.9767	0.4168	0.9617	0.4148	0.9591	0.4157
ARRE(1)	0.9	10	5	0	0.6847	0.8297	0.7168	0.9933	0.7215	0.9947	0.7266	0.995	0.7309
ARRE(1)	0.9	10	5	0.05	0.3452	0.3646	0.2583	0.0228	0.2585	0.0229	0.2504	0.0179	0.2494
ARRE(1)	0.9	10	5	0.1	0.3833	0.7203	0.2516	0.1157	0.2449	0.1132	0.2409	0.1006	0.2378
ARRE(1)	0.9	10	5	0.2	0.4509	0.9939	0.3722	0.6093	0.3615	0.6128	0.3636	0.5731	0.3533
ARRE(1)	0.9	10	5	0.4	0.5842	1	0.3918	0.9978	0.3965	0.9983	0.3916	0.9975	0.3961
ARRE(1)	0.9	10	5	0.6	0.6894	1	0.4748	1	0.4769	1	0.4746	1	0.4766
ARRE(1)	0.9	20	3	0	0.5657	0.8386	0.6735	0.9989	0.6761	0.999	0.6859	0.9991	0.6885
ARRE(1)	0.9	20	3	0.05	0.4491	0.2742	0.2948	0.0037	0.2927	0.0035	0.2855	0.0033	0.2827
ARRE(1)	0.9	20	3	0.1	0.4736	0.5427	0.2367	0.0245	0.2372	0.0216	0.2289	0.0205	0.228
ARRE(1)	0.9	20	3	0.2	0.5353	0.9446	0.2204	0.2487	0.2215	0.2209	0.2148	0.2214	0.2162
ARRE(1)	0.9	20	3	0.4	0.6884	1	0.2972	0.9728	0.3013	0.9591	0.2966	0.9656	0.3006
ARRE(1)	0.9	20	3	0.6	0.7492	1	0.369	1	0.3722	1	0.3689	1	0.372
ARRE(1)	0.9	20	5	0	0.5752	0.8228	0.6770	0.9961	0.6803	0.9973	0.6802	0.9966	0.6832
ARRE(1)	0.9	20	5	0.05	0.4535	0.5035	0.2273	0.0414	0.2214	0.0353	0.2228	0.0375	0.2186
ARRE(1)	0.9	20	5	0.1	0.4867	0.9102	0.2151	0.2681	0.1916	0.256	0.2095	0.2526	0.187
ARRE(1)	0.9	20	5	0.2	0.5394	1	0.2544	0.9243	0.2544	0.9386	0.2537	0.917	0.2535
ARRE(1)	0.9	20	5	0.4	0.6396	1	0.3022	1	0.308	1	0.3022	1	0.3078
ARRE(1)	0.9	20	5	0.6	0.7258	1	0.4117	1	0.4163	1	0.4117	1	0.4163

Table 4.4: Success Rate of Choosing the Correct Model Using Different Statistics for Covariance Structure=Autoregressive order 1 plus random effect (0.9)

Covariance Structure	Parameter	# of subject	# of obs	beta	Marginal R <sup>2</sup> _2	f <sub>test</sub>	aic1	aic2	bic1	bic2	bic3	bic4
ARRE(1)	0.9	10	3	0	0.9991	0.9818	0.9794	0.9263	0.9880	0.9832	0.9638	0.9395
ARRE(1)	0.9	10	3	0.05	0.3175	0.0334	0.0405	0.1172	0.0232	0.0339	0.0667	0.0989
ARRE(1)	0.9	10	3	0.1	0.01	0.0842	0.1054	0.2379	0.0635	0.0928	0.1505	0.2098
ARRE(1)	0.9	10	3	0.2	0.0617	0.3002	0.4035	0.6307	0.2932	0.3701	0.4994	0.5898
ARRE(1)	0.9	10	3	0.4	0.5221	0.7789	0.9652	0.9934	0.9284	0.9565	0.982	0.9914
ARRE(1)	0.9	10	3	0.6	0.9433	0.9131	0.9998	0.9999	0.9997	0.9997	1	0.9999
ARRE(1)	0.9	10	5	0	0.9954	0.9773	0.9884	0.9073	0.9949	0.9902	0.9707	0.9237
ARRE(1)	0.9	10	5	0.05	0.0185	0.0893	0.0629	0.25	0.0322	0.0554	0.1244	0.2214
ARRE(1)	0.9	10	5	0.1	0.0979	0.3229	0.2948	0.6287	0.1841	0.2713	0.4359	0.5895
ARRE(1)	0.9	10	5	0.2	0.5765	0.8577	0.9236	0.9902	0.8478	0.9128	0.9657	0.9883
ARRE(1)	0.9	10	5	0.4	0.9981	0.9923	1	1	1	1	1	1
ARRE(1)	0.9	10	5	0.6	1	0.999	1	1	1	1	1	1
ARRE(1)	0.9	20	3	0	0.9993	0.9767	0.9862	0.9262	0.9952	0.9921	0.9790	0.9596
ARRE(1)	0.9	20	3	0.05	0.0025	0.0585	0.0459	0.1537	0.0173	0.0302	0.06	0.0986
ARRE(1)	0.9	20	3	0.1	0.0173	0.1826	0.172	0.4027	0.0832	0.1206	0.2133	0.2956
ARRE(1)	0.9	20	3	0.2	0.194	0.5965	0.7124	0.9017	0.5386	0.6174	0.7662	0.8359
ARRE(1)	0.9	20	3	0.4	0.9498	0.93	0.9997	1	0.999	0.999	0.9998	0.9999
ARRE(1)	0.9	20	3	0.6	1	0.9822	1	1	1	1	1	1
ARRE(1)	0.9	20	5	0	0.9974	0.9743	0.9917	0.9146	0.9979	0.9959	0.9791	0.9538
ARRE(1)	0.9	20	5	0.05	0.0328	0.1847	0.108	0.3976	0.0414	0.0749	0.1759	0.2891
ARRE(1)	0.9	20	5	0.1	0.2421	0.6453	0.5751	0.8819	0.3778	0.5008	0.6897	0.8125
ARRE(1)	0.9	20	5	0.2	0.9336	0.9878	0.9983	1	0.994	0.9976	0.9996	0.9998
ARRE(1)	0.9	20	5	0.4	1	1	1	1	1	1	1	1
ARRE(1)	0.9	20	5	0.6	1	1	1	1	1	1	1	1

Table 4.4 (continue): Success Rate of Choosing the Correct Model Using Different Statistics for Covariance Structure=ARRE(0.9)

\* R<sup>2</sup>\_2 and CCC\_2 are calculated using s=p + k

\* aic2, bic3 and bic4 are calculated without the term  $\frac{1}{2} \log \left| \sum_{i=1}^m X_i' X_i \right|$

\* bic2 and bic4 used  $N^* = m$

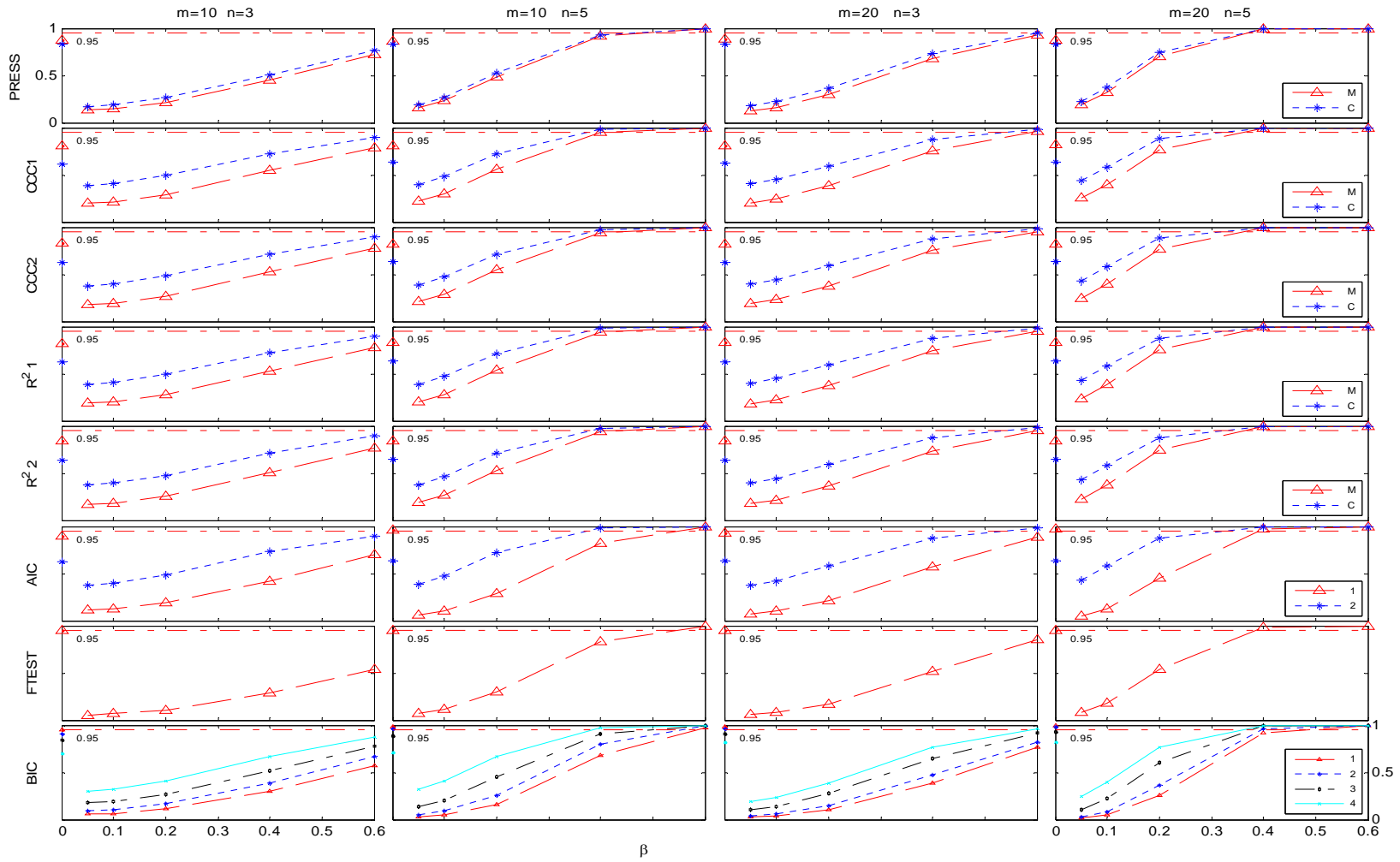


Figure 1: Correct model selection rates for 17 criteria for the compound symmetry structure involving within subject correlation of 0.5. M= calculation with marginal residuals, C= calculation with conditional residuals, AIC1= with determinat term, AIC2= without determinant term, BIC 1,2= with determinat term, BIC3,4= without determinant term, BIC1,3= $(N^* = N - p)$ , BIC2,4=  $(N^* = m)$ .

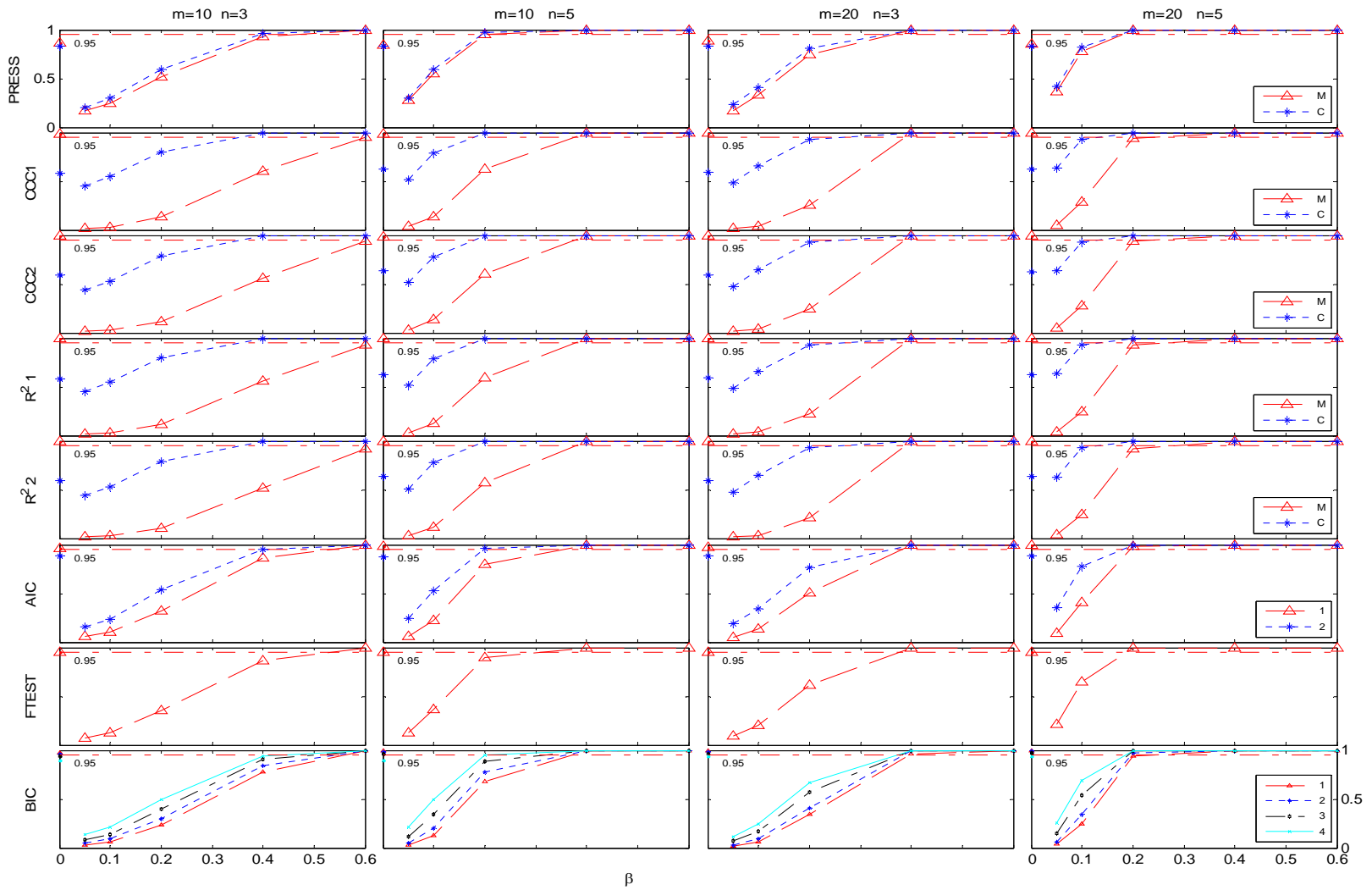


Figure 2: Correct model selection rates for 17 criteria for the compound symmetry structure involving within subject correlation of 0.9. M= calculation with marginal residuals, C= calculation with conditional residuals, AIC1= with determinat term, AIC2= without determinat term, BIC 1,2= with determinat term, BIC3,4= without determinant term, BIC1,3= $(N^* = N - p)$ , BIC2,4=  $(N^* = m)$ .

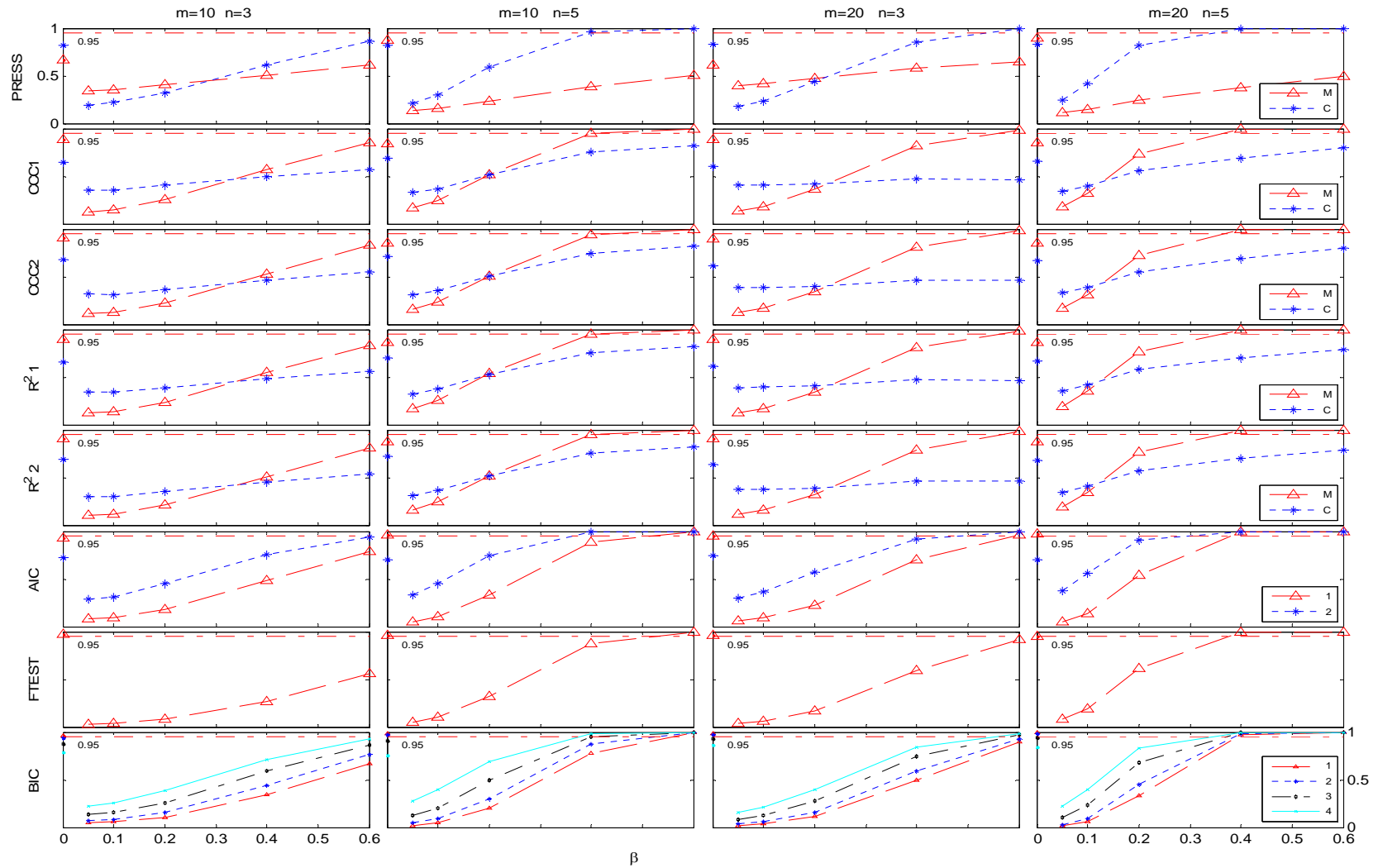


Figure 3: Correct model selection rates for 17 criteria for the autoregressive order 1 plus random effect structure with autocorrelation of 0.5. M= calculation with marginal residuals, C= calculation with conditional residuals, AIC1= with determinat term, AIC2= without determinat term, BIC 1,2= with determinat term, BIC3,4= without determinat term, BIC1,3= $(N^* = N - p)$ , BIC2,4=  $(N^* = m)$ .

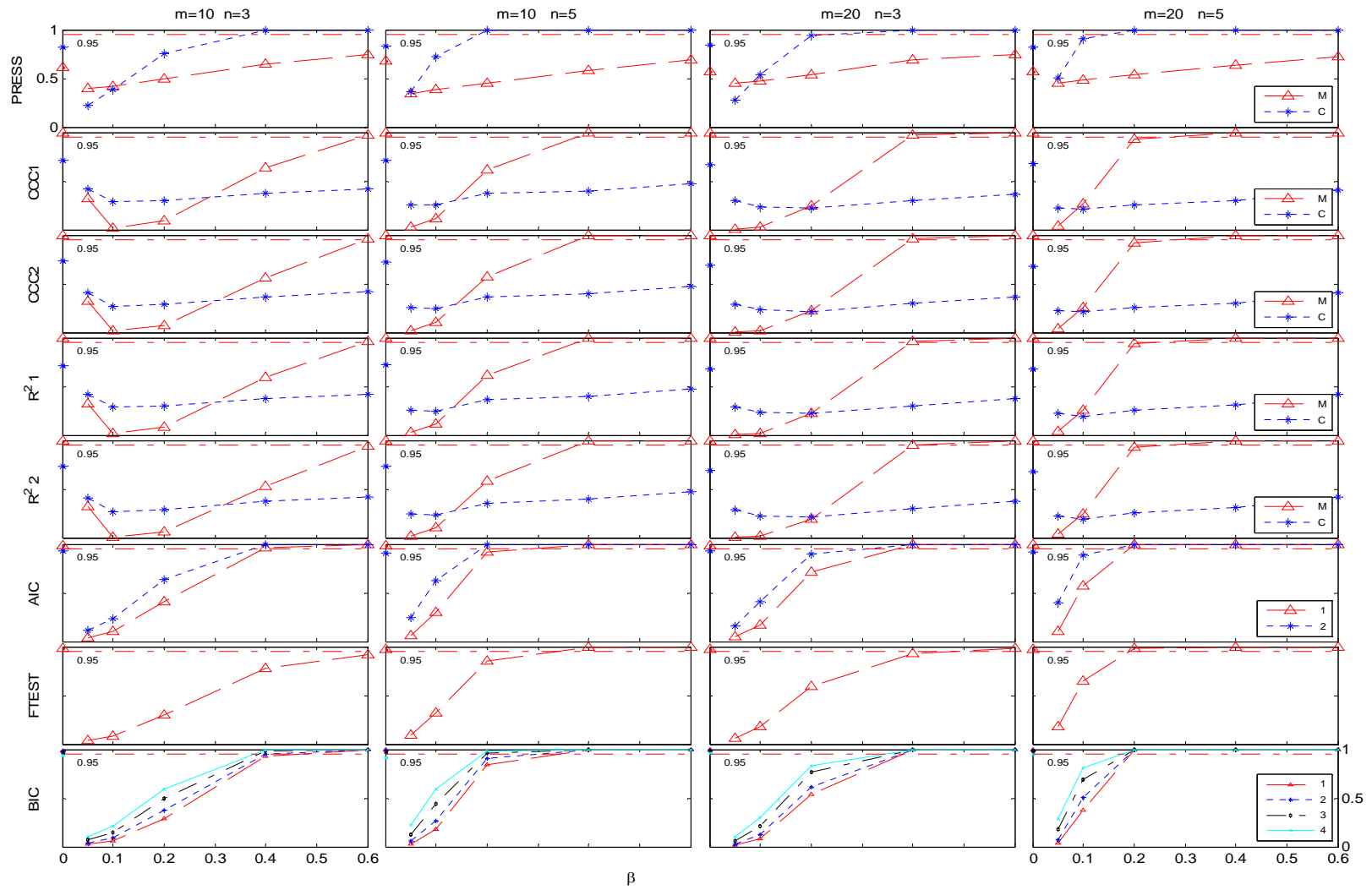


Figure 4: Correct model selection rates for 17 criteria for the autoregressive order 1 plus random effect structure with autocorrelation of 0.9. M= calculation with marginal residuals, C= calculation with conditional residuals, AIC1= with determinat term, AIC2= without determinat term, BIC 1,2= with determinat term, BIC3,4= without determinat term, BIC1,3= $(N^* = N - p)$ , BIC2,4=  $(N^* = m)$ .

Statistics	Average Success Rate for $\beta_2 = 0$	Average Success Rate for $\beta_2 > 0$
Conditional PRESS	83.14813%	65.38863%
Marginal PRESS	77.59688%	52.86574%
Conditional CCC	64.35438%	58.39%
Marginal CCC	91.76688%	52.48213%
Conditional CCC-2	65.58438%	57.6435%
Marginal CCC-2	92.32188%	51.4375%
Conditional $R^2$	64.8925%	58.25725%
Marginal $R^2$	92.49313%	51.69163%
Conditional $R^2$ -2	66.14375%	57.49638%
Marginal $R^2$ -2	93.03%	50.64013%
F-test	96.09188%	50.398%
AIC1	96.445%	51.232%
AIC2	78.77750%	69.35038%
BIC1	98.74188%	44.09075%
BIC2	97.41563%	48.71213%
BIC3	94.08313%	55.39738%
BIC4	86.07688%	63.96625%

Table 4.5: Average Success Rate of Choosing the Correct Model Using Different Statistics



## Chapter 5

### Conclusions

This study investigated the performance of PRESS, CCC,  $R^2$ , F-test, and two information criteria in selecting the better of two models for the mean in linear mixed models. Also, the study investigated the impact of marginal and conditional approaches for calculating the residual, the impact of using different REML likelihood functions for information criteria, and the impact of different adjustments for CCC,  $R^2$ , and information criteria.

The question addressed was appropriateness of the 17 statistics in model selection for the fixed effects portion of the model. Although this simulation study was based on a very simple scenario and a basic level of model selection, it provides some numerical evidence of the usefulness of the 17 statistics in selecting the best linear mixed model.

For the compound symmetry structure, the success rates for all 17 statistics were high. The study suggested using marginal rather than conditional residuals for PRESS, CCC and  $R^2$ . It suggested using REML likelihood function, which has the term  $\frac{1}{2} \log \left| \sum_{i=1}^m X_i' X_i \right|$  for AIC and BIC. For CCC,  $R^2$ , and the information criterion, there was no difference for the various parameter number adjustments.

For Autoregressive order 1 plus random effect, the study suggested using conditional residuals for PRESS, marginal residuals for CCC and  $R^2$ , and using the REML function with the term  $\frac{1}{2} \log \left| \sum_{i=1}^m X_i' X_i \right|$  for AIC and BIC. There was also no difference for the various parameter number adjustments.

The F-test performed well for all covariance structures. The study also indicated that characteristics of the data, such as the covariance structure, parameter values, and sample size, can greatly impact performance of various statistics. No one criterion is consistently better than the others in terms of selection performance in the simulation study. Further research is needed on model selection from more than 2 models (Gurka, 2006), model selection for other covariance structures, selection for both mean model and covariance structure, model selection using maximum likelihood, and the likelihood ratio test, as well as why there was a “bounce” of PRESS, CCC, and  $R^2$  for subjects=10 and observations=3.

## Bibliography

- Akaike, H. (1973), "Information theory and an extension of the maximum likelihood principle," *In Second International Symposium on Information Theory*, B. N. Petrov and F. Csaki, eds. Akademiai Kiado, Budapest, 267-281.
- Allen, D.M. (1974), "The Relationship between Variable Selection and Data Augmentation and a Method of Prediction," *Technometrics*, 16, 125-127.
- Boozdogan, H. (1987), "Model Selection and Akaike's Information Criterion (AIC): The General Theory and its Analytical extensions," *Psychometrika*, 52, 345-370.
- Christensen, R., Pearson, L. M., and Johnson, W. "Case-Deletion Diagnostics for Mixed Models," *Technometrics*, Vol. 34, No.1 (February 1992), 38-45.
- Edwards, L. J., Muller, K. E., Gurka, M. J. and Stewart, P. W. "Selecting the Best Linear Mixed Model: The State of the Art for Longitudinal Data Analysis," *2005 Joint Statistical Meetings, Minneapolis, Minnesota*.
- Fai, A.H.T. and Cornelius, P.L. (1996), "Approximate F-tests of Multiple Degree of Freedom Hypotheses in Generalized Least Squares Analyses of Unbalanced Split-plot Experiments," *Journal of Statistical Computing and Simulation* 54: 363-378.
- Ferron, J., Dailey, R., and Yi, Q. (2002), "Effects of misspecifying the first-level error structure in two-level models of change," *Multivariate Behavioral Research*, 37 (3) 379-403.
- Giesbrecht, F.G. and Burns, J.C. (1985), "Two-Stage Analysis Based on a Mixed Model: Large-sample Asymptotic Theory and Small-Sample Simulation Results," *Biometrics* 41: 853-862.
- Gomez, E. V., Schaalje, G. B. and Fellingham, G. W. "Performance of the Kenward-Roger Method when the Covariance Structure is Selected Using AIC and BIC," *Communications in Statistics-Simulation and Computation*, 34: 377-392, 2005.
- Gurka, M. J. (2006), "Selecting the Best Linear Mixed Model under REML," *The American Statistician*, 60(1): 19-26.
- Hurvich, C. M., and Tsai, C. L. (1989), "Regression and Time Series Model Selection in Small Samples," *Biometrika*, 76, 297-307.
- Keselman, H. J., Algina, J., Kowalchuk, R. K., and Wolfinger, R. D. (1998), "A comparison of two approaches for selecting covariance structures in the analysis of repeated measurements," *Communications in Statistics: Simulation and Computation*, 27 (3) 591-604.

Kenward, M. G. and Roger, J. H. (1997), "Small Sample Inference for Fixed Effects from Restricted Maximum Likelihood," *Biometrics* 53: 983-997.

Lin, L. I.-K. (1989), "A Concordance Correlation Coefficient to Evaluate Reproducibility," *Biometrics* 45, 255-268.

Schabenberger, O. "Mixed Model Influence Diagnostics," *SUGI 29, Paper 189-29*

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

Schaalje, G. B., McBride, J. B., and Fellingham, G. W. (2001), "Approximations to Distributions of Test Statistics in Complex Mixed Linear Models Using SAS Proc MIXED," Paper 262-26 in Proceedings of the SAS Users Group International 26<sup>th</sup> Annual Conference.

Schaalje, G. B., McBride, J. B., and Fellingham, G. F. (2002), "Adequacy of Approximations to Distributions of Test Statistics in Complex Mixed Linear Models," *Journal of Agricultural, Biological, and Environmental Statistics*, 7 512-524.

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

Vonesh, E. F., Chinchilli, V. M. and Pu, K. "Goodness-of-Fit in Generalized Nonlinear Mixed-Effects Models," *Biometrics*, Vol. 52, No. 2 (Jun., 1996), 575-587.

## APPENDIX A

### Sample SAS Code

```
/******  
for m=20 (subject), n=5 (observation per subject), covariance  
structure=cs(0.9),beta2=0.4  
*****/  
proc iml;  
  *set parameters;  
  seed=1.62104E9;  
  %let k=%eval(2);  
  %let nit=%eval(6000);  
  %let nobobs=%eval(5);  
  %let nsub=%eval(20);  
  %let n1=%eval(&nobobs*&nsub);  
  %let n=%eval(&nobobs*&nsub*&nit);  
  beta2=0.4;  
  covmat_cs={1 0.9 0.9 0.9 0.9,  
             0.9 1 0.9 0.9 0.9,  
             0.9 0.9 1 0.9 0.9,  
             0.9 0.9 0.9 1 0.9,  
             0.9 0.9 0.9 0.9 1};  
  ch=t(root(covmat_cs));  
  x1=j(&nobobs,1,1)||{-3,-2,0,2,3};  
  x2={1,2,3,1,2};  
  beta0={1, 1};  
  *generating data;  
  do iter=1 to &nit;  
    do subj=1 to &nsub;  
      y1=normal(repeat(seed, &nobobs,1));  
      call rannor(seed,y1);  
      ylp=ch*y1;  
      y2=x1*beta0+x2*beta2+ylp;  
      if (iter=1) & (subj=1) then y=y2;  
      else y=y//y2;  
      if (iter=1) & (subj=1) then x=x1[,2];  
      else x=x//x1[,2];  
      if (iter=1) & (subj=1) then x22=x2;  
      else x22=x22//x2;  
      if (iter=1) & (subj=1) then subj2=j(&nobobs,1,subj);  
      else subj2=subj2//j(&nobobs,1,subj);  
      if (iter=1) & (subj=1) then iter2=j(&nobobs,1,iter);  
      else iter2=iter2//j(&nobobs,1,iter);  
    end;  
  end;  
  print seed;  
  create datanew var{y iter2 x x22 subj2};  
  append;  
  quit;  
  data datanew (rename=(y=y iter2=iter x=x1 x22=x2 subj2=subj));  
  set datanew ;  
  run;  
  * fit the full model;  
  ods listing close;  
  proc mixed data=datanew;
```

```

by iter;
class subj;
model y=x1 x2/ S influence(iter=5 est) outpred=pred_full(keep=y iter x1
x2 subj Pred Resid)
outpredm=predm_full(keep=y iter x1 x2 subj Pred Resid)
DDFM=kenwardroger;
random subj/s;
ods output fitstatistics=cs;
ods output tests3=pval_cs;
ods output SolutionF=coef_fixed (keep=estimate iter) ;
ods output SolutionR=coef_random;
ods output influence=press (keep=iter pressres);
run;
ods listing;
*fit the reduced model;
ods listing close;
proc mixed data=datanew;
by iter;
class subj;
model y=x1 / S influence(iter=5 est) outpred=pred_reduced(keep=y iter
x1 subj Pred Resid)
outpredm=predm_reduced(keep=y iter x1 subj Pred Resid)
ddfm=kenwardroger;
random subj/s;
ods output fitstatistics=cs_reduced;
ods output tests3=pval_cs_reduced;
ods output SolutionF=coef_fixed_reduced (keep=estimate iter) ;
ods output SolutionR=coef_random_reduced;
ods output influence=press_redu(keep=iter pressres) ;
run;
ods listing;
/*****
marginal PRESS statistics
*****/
data press_marginal_full;
set press;
pressmf=pressres**2;
run;
proc sort data=press_marginal_full;
by iter;
run;
data press_marginal_redu;
set press_redu;
pressmr=pressres**2;
run;
proc sort data=press_marginal_redu;
by iter;
run;
data press_marginal;
merge press_marginal_full press_marginal_redu;
by iter;
keep iter pressmf pressmr;
run;
*sum the PRESS statistics over iteration;
proc means noprint sum data=press_marginal;
by iter;
var pressmf pressmr;

```

```

output out=press_margin sum=;
run;
data pressm;
set press_margin;
press_diff=pressmr-pressmf;
if press_diff>0 then presm_rate=1;
if press_diff<0 then presm_rate=0;
run;
proc means data=pressm;
var presm_rate;
run;
/*****
conditional PRESS statistics
*****/
*full model;
data pred3;
%macro mixed2;
  %do i=1 %to &n1;
    data analysis;
    set datanew;
    obsnum=mod(_N_,&n1);
    if obsnum=0 then obsnum=&n1;
    if obsnum = &i then do;
      yactual=y;
      y=.;
    end;
proc mixed data=analysis;
by iter;
class subj;
model y=x1 x2/ S outpred=pred;
random subj/s;
run;
data pred2;
set pred;
if y=.;
press=(pred-yactual)**2;
keep iter press;
run;
data pred3;
set pred3 pred2;
%end;
%mend ;
ods listing close;
%mixed2;
ods listing;
proc sort data=pred3;
by iter;
run;
*reduced model;
data pred4;
%macro mixed3;
  %do i=1 %to &n1;
    data analysis;
    set datanew;
    obsnum=mod(_N_,&n1);
    if obsnum=0 then obsnum=&n1;
    IF obsnum = &i THEN do;

```

```

        yactual=y;
        y=.;
        end;
        run;
proc mixed data=analysis;
by iter;
class subj;
model y=x1/ S outpred=predr;
random subj/s;
run;
data pred5;
set predr;
if y=.;
press2=(pred-yactual)**2;
keep iter press2;
data pred4;
set pred4 pred5;
%end;
%mend ;
ods listing close;
%mixed3;
ods listing;
proc sort data=pred4;
by iter;
run;
data press_condi;
merge pred3 pred4;
by iter;
run;
proc means sum noprint data=press_condi;
where iter~=.;
by iter;
var press press2;
output out=presc sum=;
run;
data press_c;
set presc;
press_d=press2-press;
if press_d>0 then presc_rate=1;
if press_d<0 then presc_rate=0;
run;
proc means data=press_c;
var presc_rate;
run;
/*****
CCC and adjusted R-squared
*****/
*marginlal and conditional prediction and residue for the full model ;
data new;
merge pred_full (rename=( Pred=pred_conditional
Resid=resid_conditional))
predm_full (rename=( Pred=pred_marginal Resid=resid_marginal));
by iter;
run;
proc iml;
use new;
read all var{y} into y;

```



```

read all var{iter} into iter;
read all var{x1} into x1;
read all var{x2} into x2;
read all var{subj} into subj;
read all var{pred_conditional} into pred_condi;
read all var{resid_conditional} into resid_condi;
read all var{pred_marginal} into pred_margin;
read all var{resid_marginal} into resid_margin;
k_full=&n1/(&n1-3);
k_full2=&n1/(&n1-5);
*compute ccc and R-squared for each iteration (full model);
ccc_condi=j(&nit,1,0);
ccc_margin=j(&nit,1,0);
r2_condi=j(&nit,1,0);
r2_margin=j(&nit,1,0);
* conditional;
do iter2=1 to &nit;
ybar=sum(y[(iter2-1)*&n1+1:iter2*&n1,])/&n1;
yhatbar_condi=sum(pred_condi[(iter2-1)*&n1+1:iter2*&n1,])/&n1;
ccc_condi[iter2]=1-(resid_condi[(iter2-
1)*&n1+1:iter2*&n1,]^*resid_condi[(iter2-1)*&n1+1:iter2*&n1,])/
((y[(iter2-1)*&n1+1:iter2*&n1,]-ybar)^*(y[(iter2-
1)*&n1+1:iter2*&n1,]-ybar)+
(pred_condi[(iter2-1)*&n1+1:iter2*&n1,]-
yhatbar_condi)^*(pred_condi[(iter2-1)*&n1+1:iter2*&n1,]-yhatbar_condi)+
&n1*(ybar-yhatbar_condi)**2);
r2_condi[iter2]=1-((resid_condi[(iter2-
1)*&n1+1:iter2*&n1,]^*resid_condi[(iter2-1)*&n1+1:iter2*&n1,])/
((y[(iter2-1)*&n1+1:iter2*&n1,]-ybar)^*(y[(iter2-
1)*&n1+1:iter2*&n1,]-ybar)));
* marginal;
ybar=sum(y[(iter2-1)*&n1+1:iter2*&n1,])/&n1;
yhatbar_margin=sum(pred_margin[(iter2-1)*&n1+1:iter2*&n1,])/&n1;
ccc_margin[iter2]=1-(resid_margin[(iter2-
1)*&n1+1:iter2*&n1,]^*resid_margin[(iter2-1)*&n1+1:iter2*&n1,])/
((y[(iter2-1)*&n1+1:iter2*&n1,]-ybar)^*(y[(iter2-
1)*&n1+1:iter2*&n1,]-ybar)+
(pred_margin[(iter2-1)*&n1+1:iter2*&n1,]-
yhatbar_margin)^*(pred_margin[(iter2-1)*&n1+1:iter2*&n1,]-
yhatbar_margin)+
&n1*(ybar-yhatbar_margin)**2);
r2_margin[iter2]=1-((resid_margin[(iter2-
1)*&n1+1:iter2*&n1,]^*resid_margin[(iter2-1)*&n1+1:iter2*&n1,])/
((y[(iter2-1)*&n1+1:iter2*&n1,]-ybar)^*(y[(iter2-
1)*&n1+1:iter2*&n1,]-ybar)));
end;
cccadj_condi=1-k_full*(1-ccc_condi);
r2adj_condi=1-k_full*(1-r2_condi);
cccadj_margin=1-k_full*(1-ccc_margin);
r2adj_margin=1-k_full*(1-r2_margin);
ccc_r2_full=cccadj_condi||r2adj_condi||cccadj_margin||r2adj_margin;
create ccc_r2_full from ccc_r2_full;
append from ccc_r2_full;

cccadj_condi2=1-k_full2*(1-ccc_condi);
r2adj_condi2=1-k_full2*(1-r2_condi);

```

```

cccadj_margin2=1-k_full2*(1-ccc_margin);
r2adj_margin2=1-k_full2*(1-r2_margin);
ccc_r2_full2=cccadj_condi2|r2adj_condi2|cccadj_margin2|r2adj_margin2
;
create ccc_r2_full2 from ccc_r2_full2;
append from ccc_r2_full2;
quit;
data ccc_r2_full1 (rename=(col1=cccadj_condi col2=r2adj_condi
col3=cccadj_margin col4=r2adj_margin ));
set ccc_r2_full1;
run;
data ccc_r2_full2 (rename=(col1=cccadj_condi2 col2=r2adj_condi2
col3=cccadj_margin2 col4=r2adj_margin2 ));
set ccc_r2_full2;
run;
*reduced model;
data new_reduced;
merge pred_reduced (rename=( Pred=pred_conditional_reduced
Resid=resid_conditional_reduced))
predm_reduced (rename=( Pred=pred_marginal_reduced
Resid=resid_marginal_reduced));
by iter;
run;
proc iml;
use new_reduced;
read all var{y} into y;
read all var{iter} into iter;
read all var{x1} into x1;
read all var{subj} into subj;
read all var{pred_conditional_reduced} into pred_condi_reduced;
read all var{resid_conditional_reduced} into resid_condi_reduced;
read all var{pred_marginal_reduced} into pred_margin_reduced;
read all var{resid_marginal_reduced} into resid_margin_reduced;
*compute ccc and R-squared for each iteration (reduced model);
k_reduced=&n1/(&n1-2);
k_reduced2=&n1/(&n1-4);
ccc_condi_reduced=j(&nit,1,0);
ccc_margin_reduced=j(&nit,1,0);
r2_condi_reduced=j(&nit,1,0);
r2_margin_reduced=j(&nit,1,0);
do iter2=1 to &nit;
* conditional;
ybar=sum(y[(iter2-1)*&n1+1:iter2*&n1,])/&n1;
yhatbar_condi_reduced=sum(pred_condi_reduced[(iter2-
1)*&n1+1:iter2*&n1,])/&n1;
ccc_condi_reduced[iter2]=1-(resid_condi_reduced[(iter2-
1)*&n1+1:iter2*&n1,]^*resid_condi_reduced[(iter2-1)*&n1+1:iter2*&n1,])/
((y[(iter2-1)*&n1+1:iter2*&n1,]-ybar)^*(y[(iter2-
1)*&n1+1:iter2*&n1,]-ybar)+
(pred_condi_reduced[(iter2-1)*&n1+1:iter2*&n1,]-
yhatbar_condi_reduced)^*(pred_condi_reduced[(iter2-
1)*&n1+1:iter2*&n1,]-yhatbar_condi_reduced)+
&n1*(ybar-yhatbar_condi_reduced)**2);
r2_condi_reduced[iter2]=1-((resid_condi_reduced[(iter2-
1)*&n1+1:iter2*&n1,]^*resid_condi_reduced[(iter2-1)*&n1+1:iter2*&n1,])/
((y[(iter2-1)*&n1+1:iter2*&n1,]-ybar)^*(y[(iter2-
1)*&n1+1:iter2*&n1,]-ybar)));

```

```

* marginal;
ybar=sum(y[(iter2-1)*&n1+1:iter2*&n1,])/&n1;
yhatbar_margin_reduced=sum(pred_margin_reduced[(iter2-
1)*&n1+1:iter2*&n1,])/&n1;
ccc_margin_reduced[iter2]=1-(resid_margin_reduced[(iter2-
1)*&n1+1:iter2*&n1,]^*resid_margin_reduced[(iter2-
1)*&n1+1:iter2*&n1,])/
      ((y[(iter2-1)*&n1+1:iter2*&n1,]-ybar)^*(y[(iter2-
1)*&n1+1:iter2*&n1,]-ybar)+
      (pred_margin_reduced[(iter2-1)*&n1+1:iter2*&n1,]-
yhatbar_margin_reduced)^*(pred_margin_reduced[(iter2-
1)*&n1+1:iter2*&n1,]-yhatbar_margin_reduced)+
      &n1*(ybar-yhatbar_margin_reduced)**2);
r2_margin_reduced[iter2]=1-((resid_margin_reduced[(iter2-
1)*&n1+1:iter2*&n1,]^*resid_margin_reduced[(iter2-
1)*&n1+1:iter2*&n1,])/
      ((y[(iter2-1)*&n1+1:iter2*&n1,]-ybar)^*(y[(iter2-
1)*&n1+1:iter2*&n1,]-ybar)));
end;
* adjusted ccc and r-squared;
cccadj_condi_reduced=1-k_reduced*(1-ccc_condi_reduced);
r2adj_condi_reduced=1-k_reduced*(1-r2_condi_reduced);
cccadj_margin_reduced=1-k_reduced*(1-ccc_margin_reduced);
r2adj_margin_reduced=1-k_reduced*(1-r2_margin_reduced);
use ccc_r2_full;
read all var{cccadj_condi} into cccadj_condi_full;
read all var{r2adj_condi} into r2adj_condi_full;
read all var{cccadj_margin} into cccadj_margin_full;
read all var{r2adj_margin} into r2adj_margin_full;
cccadj_condi=cccadj_condi_full-cccadj_condi_reduced;
cccadj_margin=cccadj_margin_full-cccadj_margin_reduced;
r2adj_condi=r2adj_condi_full-r2adj_condi_reduced;
r2adj_margin=r2adj_margin_full-r2adj_margin_reduced;
*success rates in choosing the correct model;
ccc_con=0;
ccc_mar=0;
r2_con=0;
r2_mar=0;
do i=1 to &nit;
if (cccadj_condi[i]>0) then ccc_con=ccc_con+1;
if (cccadj_margin[i]>0) then ccc_mar=ccc_mar+1;
if (r2adj_condi[i]>0) then r2_con=r2_con+1;
if (r2adj_margin[i]>0) then r2_mar=r2_mar+1;
end;
ccc_con=ccc_con/&nit;
ccc_mar=ccc_mar/&nit;
r2_con=r2_con/&nit;
r2_mar=r2_mar/&nit;
print ccc_con ccc_mar r2_con r2_mar;
* if s=# para in beta + covariance structure;
cccadj_condi_reduced2=1-k_reduced2*(1-ccc_condi_reduced);
r2adj_condi_reduced2=1-k_reduced2*(1-r2_condi_reduced);
cccadj_margin_reduced2=1-k_reduced2*(1-ccc_margin_reduced);
r2adj_margin_reduced2=1-k_reduced2*(1-r2_margin_reduced);
use ccc_r2_full2;
read all var{cccadj_condi2} into cccadj_condi_full2;
read all var{r2adj_condi2} into r2adj_condi_full2;

```

```

read all var{cccadj_margin2} into cccadj_margin_full2;
read all var{r2adj_margin2} into r2adj_margin_full2;
cccadj_c2=cccadj_condi_full2-cccadj_condi_reduced2;
cccadj_m2=cccadj_margin_full2-cccadj_margin_reduced2;
r2adj_c2=r2adj_condi_full2-r2adj_condi_reduced2;
r2adj_m2=r2adj_margin_full2-r2adj_margin_reduced2;
*success rates in choosing the correct model;
ccc_con2=0;
ccc_mar2=0;
r2_con2=0;
r2_mar2=0;
do i=1 to &nit;
if (cccadj_c2[i]>0) then ccc_con2=ccc_con2+1;
if (cccadj_m2[i]>0) then ccc_mar2=ccc_mar2+1;
if (r2adj_c2[i]>0) then r2_con2=r2_con2+1;
if (r2adj_m2[i]>0) then r2_mar2=r2_mar2+1;
end;
ccc_con2=ccc_con2/&nit;
ccc_mar2=ccc_mar2/&nit;
r2_con2=r2_con2/&nit;
r2_mar2=r2_mar2/&nit;
print ccc_con2 ccc_mar2 r2_con2 r2_mar2;
/*****
AIC, BIC, F test
*****/
data fttest;set pval_cs; keep fvalue probf iter effect;
if effect='x1' then delete;run;
data fttest;set fttest;
if probf<=0.05 then f_cs=1;
if probf>0.05 then f_cs=0;run;
proc means data=fttest;var f_cs;run;
* AIC AND BIC;
data cs2;set cs; retain aic bic loglikelihood; keep aic bic
loglikelihood iter;
if descr='-2 Res Log Likelihood' then do;
aic=.;bic=.;loglikelihood=value;end;
if descr='AIC (smaller is better)' then do; aic=value;end;
if descr='BIC (smaller is better)' then do; bic=value;
output; end; run;
proc iml;
use cs2;
read all var{aic} into aic;
read all var{bic} into bic;
read all var{loglikelihood} into loglike;
read all var{iter} into iter;
x=j(&nobs,1,1)||{-3,-2,0,2,3}||{1,2,3,1,2};
term=log(det(&nsb*(x`*x)));
p=trace(x*sweep(x`*x)*x`);
s=p+&k;
aic1=loglike+2*s;
aic2=loglike-term+2*s;
bic1=loglike+s*log(&n1-p);
bic2=loglike+s*log(&nsb);
bic3=loglike-term+s*log(&n1-p);
bic4=loglike-term+s*log(&nsb);
info=iter||aic1||aic2||bic1||bic2||bic3||bic4;
create info_f from info;

```

```

append from info;
quit;
data info_f (rename=(coll=iter col2=aic1_f col3=aic2_f col4=bic1_f
col5=bic2_f col6=bic3_f col7=bic4_f));
set info_f;
run;
data cs_reduced1;set cs_reduced; retain aic_r bic_r loglikelihood_r;
keep aic_r bic_r loglikelihood_r iter;
if descr='-2 Res Log Likelihood' then do;
aic_r=.;bic_r=.;loglikelihood_r=value;end;
if descr='AIC (smaller is better)' then do; aic_r=value;end;
if descr='BIC (smaller is better)' then do; bic_r=value;output; end;
run;
proc iml;
use cs_reduced1;
read all var{aic_r} into aic_r;
read all var{bic_r} into bic_r;
read all var{loglikelihood_r} into loglike_r;
read all var{iter} into iter;
x=j(&nobs,1,1)||{-3,-2,0,2,3};
term=log(det(&sub*(x`*x)));
p=trace(x*sweep(x`*x)*x`);
s=p+&k;
aic1_r=loglike_r+2*s;
aic2_r=loglike_r-term+2*s;
bic1_r=loglike_r+s*log(&n1-p);
bic2_r=loglike_r+s*log(&sub);
bic3_r=loglike_r-term+s*log(&n1-p);
bic4_r=loglike_r-term+s*log(&sub);
info2=iter||aic1_r||aic2_r||bic1_r||bic2_r||bic3_r||bic4_r;
create info_r from info2;
append from info2;
quit;
data info_r (rename=(coll=iter col2=aic1_r col3=aic2_r col4=bic1_r
col5=bic2_r col6=bic3_r col7=bic4_r));
set info_r;
run;
data cs_info;
merge info_f info_r;
by iter;
run;
data test;
set cs_info;
aic1_dif=aic1_r-aic1_f;
aic2_dif=aic2_r-aic2_f;
bic1_dif=bic1_r-bic1_f;
bic2_dif=bic2_r-bic2_f;
bic3_dif=bic3_r-bic3_f;
bic4_dif=bic4_r-bic4_f;
if aic1_dif>0 then aic1_cs=1;else aic1_cs=0;
if aic2_dif>0 then aic2_cs=1;else aic2_cs=0;
if bic1_dif>0 then bic1_cs=1;else bic1_cs=0;
if bic2_dif>0 then bic2_cs=1;else bic2_cs=0;
if bic3_dif>0 then bic3_cs=1;else bic3_cs=0;
if bic4_dif>0 then bic4_cs=1;else bic4_cs=0;run;
proc means data=test;
var aic1_cs aic2_cs bic1_cs bic2_cs bic3_cs bic4_cs;run;

```