VARIABLE SELECTION WHEN CONFRONTED WITH MISSING DATA

by

Melissa L. Ziegler

B.S. Mathematics, Elizabethtown College, 2000

M.A. Statistics, University of Pittsburgh, 2002

Submitted to the Graduate

Faculty of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2006

UNIVERSITY OF PITTSBURGH

FACULTY OF ARTS AND SCIENCES

This dissertation was presented

by

Melissa L. Ziegler

It was defended on

June 2, 2006

and approved by

Satish Iyengar

Leon J. Gleser

Henry Block

Douglas E. Williamson

Dissertation Director: Satish Iyengar

VARIABLE SELECTION WHEN CONFRONTED WITH MISSING DATA

Melissa L. Ziegler, PhD

University of Pittsburgh, 2006

Variable selection is a common problem in linear regression. Stepwise methods, such as forward selection, are popular and are easily available in most statistical packages. The models selected by these methods have a number of drawbacks: they are often unstable, with changes in the set of variable selected due to small changes in the data, and they provide upwardly biased regression coefficient estimates. Recently proposed methods, such as the lasso, provide accurate predictions via a parsimonious, interpretable model.

Missing data values are also a common problem, especially in longitudinal studies. One approach to account for missing data is multiple imputation. The simulation studies were conducted comparing the lasso to standard variable selection methods under different missing data conditions, including the percentage of missing values and the missing data mechanism. Under missing at random mechanisms, missing data were created at the 25 and 50 percent levels with two types of regression parameters, one containing large effects and one containing several small, but nonzero, effects. Five correlation structures were used in generating the data: independent, autoregressive with correlation 0.25 and 0.50, and equicorrelated again with correlation 0.25 and 0.50. Three different missing data mechanisms were used to create the missing data: linear, convex and sinister. These mechanisms

Least angle regression performed well under all conditions when the true regression parameter vector contained large effects, with its dominance increasing as the correlation between the predictor variables increased. This is consistent with complete data simulations studies suggesting the lasso performed poorly in situations where the true beta vector contained small, nonzero effects. When the true beta vector contained small, nonzero effects, the performance of the variable selection methods considered was situation dependent.

Ordinary least squares had superior performance in terms confidence interval coverage under the independent correlation structure and with correlated data when the true regression parameter vector consists of small, nonzero effects. A variety of methods performed well when the regression parameter vector consisted of large effects and the predictor variables were correlated depending on the missing data situation.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1.0 OVERVIEW OF VARIABLE SELECTION METHODS

1.1 INTRODUCTION

The goal of this research is to study and understand the properties of modern variable selection methods, to assess their performance in the presence of missing data, and ultimately to apply variable selection methodology to the motivating data set to find the covariates most closely related to, and predictive of, major depressive disorder (MDD). More details about the study and the data analysis are in chapter 6.

The following chapters will: (i) summarize the development of variable selection, with special attention paid to modern methods (Chapter 1); (ii) provide a detailed analysis of the properties and implementation of the least absolute shrinkage and selection operator (lasso) (chapter 2); (iii) review missing data terminology and methods that will be applied in this research (chapter 3); (vi) detail the results of the simulation study that will examine the performance of variable selection methods when data are missing; (v) provide background on the psychobiology of depression in children and adolescents (chapter 6); and, finally (vi) highlight directions for future research (chapter 7).

1.2 VARIABLE SELECTION

One of the most common model building problems is the variable selection problem [18]. In modeling the relationship between a response variable, Y, and a set of potential predictor variables, $X_1, \cdots, X_p$, what is desired is to select a subset of the possible predictors that explains the relationship with Y, provides accurate predictions of future observations, and has a simple, scientifically plausible interpretation. Many methods have been, and continue to be, developed to address this problem. This chapter will focus on variable selection methods, paying special attention to some of the more recent advances in this area. Additionally, the concept of shrinkage of parameter estimates will be introduced to provide a basis for understanding the most recent advances in variable selection methodology. These newer methods attempt to capitalize on the variance reduction provided by shrinkage methods to improve the performance of variable selection methods.

The history of selection methods is outlined in a 2000 review paper by George [18]. The development of these methods began in the 1960's with methods designed to handle the linear regression problem which, due to its wide applicability, is still the focus of much of the new methodology. The early methods focused on reducing the rather imposing 2^p possible subsets of covariates to a manageable size using, for example, the residual sum of squares (RSS) to either identify the 'best' subset of a given size or to proceed in a stepwise manner to select covariates. Refinements of these methods add a dimensionality or complexity penalty to the RSS to penalize models with a large number of covariates. Examples of such penalties include Akaike's Information Criterion (AIC) and the Bayesian Information Criterion (BIC). Advances in computing have expanded the use of variable selection methods to models such as nonlinear and generalized linear models. An overview and more detailed description of these methods can be found in Miller's *Subset Selection in Regression* [29].

Stepwise variable selection methods are some of the most widely taught and implemented selection methods in regression. These methods are attractive because they provide an automatic solution and thus are available in virtually every general statistics software package. Three types of stepwise methods are: forward selection, backward elimination and general stepwise regression (which combines forward selection and backward elimination). Forward selection starts with the model containing no predictors and adds one predictor at each step. Within a given step, the variable selected for inclusion in the model is that which minimizes the RSS. The process stops when some prespecified stopping criterion based on the amount of reduction in RSS between steps is met. On the other hand, backward elimination starts with the model containing all predictors under consideration and at each step the predictor that minimizes the RSS upon its removal. Again, the process continues until a prespecified stopping criterion is met. Finally, general stepwise regression proceeds as in forward selection, but with an added check at each step for covariates that can be removed from the model.

Stepwise selection methods have a number of drawbacks. Miller notes, "forward selection and backward elimination can fare arbitrarily badly in finding the best fitting subsets" [29]. The estimates of the regression coefficients of the selected variables are often too large in absolute value, leading to false conclusions about the actual importance of the corresponding predictors in the model. The value of R^2 is often upwardly biased, overstating the accuracy of the overall model fit. The estimates of the regression coefficients and the set of variables selected may be highly sensitive to small changes in the data. Selection bias and overfitting resulting from the use of the same data to select the model and to estimate the regression coefficients can be difficult to control [29]. Methods proposed to address one or more of these deficiencies will be considered in the next two sections (1.3, 1.4).

The following notation will be used in the description of the variable selection methods. Consider the usual regression model $\mathbf{Y} = \beta'\mathbf{X} + \epsilon$ with $\mathbf{Y}$ the $n \times 1$ vector of responses, ϵ the $n \times 1$ vector of random errors, and $\mathbf{X}$ the $n \times p$ matrix of predictors with row vector $\mathbf{x}_i$ the values for the i^{th} subject.

1.3 SHRINKAGE METHODS

Shrinkage estimators introduce a small amount of bias into a parameter estimate in an attempt to reduce its variance so that there is an overall reduction in the mean squared error (MSE). Some of the best known shrinkage methods are the James-Stein estimator and ridge regression.

The James-Stein result [23] demonstrates that the application of shrinkage can improve estimation under squared error loss. Let $X \sim N_p(\xi, \mathrm{I}_p)$, that is, $\xi = \mathrm{E(X)}$ and $\mathrm{E(X} - \xi)'(\mathrm{X} - \xi) = \mathrm{I}_p$. The goal is to estimate ξ, say by $\hat{\xi}$, under squared error loss, $\mathrm{L}(\xi, \hat{\xi}(X)) = \| \xi - \hat{\xi} \|^2$. The usual estimator, $\xi_0(X) = X$, has expected loss $\mathrm{E\,L}(\xi, \xi_0(X)) = E \| \xi - \xi_0 \|^2 = p$. James and Stein showed that for $p \geq 3$, there exists an estimator,

$$\xi_1(X) = \left(1 - \frac{p-2}{||X||^2}\right)^+ X \tag{1.1}$$

that has smaller expected loss than $\xi_0(X)$ for all ξ, where $(\cdot)^+$ denotes the positive part [21].

1.3.0.1 Ridge Regression Ridge regression was proposed by Hoerl and Kennard [21] as a way to improve the estimation of regression parameters in the case where the predictor variables are highly correlated. The method introduces bias into the estimation process with the goal of reducing the overall mean square error. The ridge regression parameter estimates are given by

$$\hat{\beta}^{RR}(k) = (\mathbf{X}'\mathbf{X} + k\mathbf{I}_p)^{-1}\mathbf{X}'\mathbf{Y} \tag{1.2}$$

where $k \geq 0$ and $\beta = (\beta_1, \cdots, \beta_p)'$. Setting k equal to zero gives the usual ordinary least squares (OLS) estimators and for $k > 0$ some bias is introduced into the estimates [21].

Hoerl and Kennard use the ridge trace, a plot constructed by simultaneously plotting each element of $\hat{\beta}^{RR}(k)$ versus k, to estimate the optimal value of k. The value of k is selected at the initial point where the $\hat{\beta}^{RR}(k)$ estimates all appear to stabilize.

The ridge regression estimates can also be expressed as a constrained minimization

$$\hat{\beta} = \text{argmin}_\beta \sum_{i=1}^{N} (y_i - \beta'\mathbf{x}_i)^2 \text{ subject to } \sum_j \beta_j^2 \leq t. \tag{1.3}$$

where $t \geq 0$ is a tuning parameter which controls the amount of shrinkage applied to the regression parameter estimates. By rewriting the ridge regression parameter estimates as

$$\hat{\beta}^{RR}(k) = [\mathbf{I}_p + k(\mathbf{X}'\mathbf{X})^{-1}]^{-1}\hat{\beta}^{OLS} = \mathbf{Z}\hat{\beta}^{OLS}, \tag{1.4}$$

the standard errors of the parameters can be obtained, as in linear regression, as

$$\text{var}(\hat{\beta}^{RR}) = \text{var}(\mathbf{Z}\hat{\beta}^{OLS}) = \mathbf{Z}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\text{var}(\mathbf{Y})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{Z}' = \hat{\sigma}^2\mathbf{Z}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{Z}'. \tag{1.5}$$

This standard error estimate for ridge regression will be useful later to approximate the standard errors of parameter estimates in other variable selection methods.

1.3.0.2 Other Shrinkage Methods A variety of other shrinkage methods have been proposed for the regression situation. See Dempster, Schatzoff and Wermuth [15] for an extensive overview of past methods. Two of the more recent methods are highlighted in a comparison paper by Vach, Sauerbrei and Schumacher [43]. In the global shrinkage factor method, each of the regression coefficients is shrunk by a common shrinkage factor c which is estimated by cross-validation calibration. To obtain the cross-validated estimate $\hat{c}$, an OLS regression of Y on X is performed with the i^{th} observation removed resulting in $\hat{\beta}^{OLS}_{(-i)}$ for $i = 1, \cdots, n$. Using these estimated regression coefficients, predictions $\hat{Y}_{(-i)} = (\hat{\beta}^{OLS}_{(-i)})' X_i$ are computed. A simple linear regression of the orignal Y on the predictions $\hat{Y}_{(-i)}$ is then performed and the resulting regression coefficient is used as the estimate $\hat{c}$. Note that the X_i are assumed to be standardized so that $\sum_i x_{ij}/N = 0$ and $\sum_i x_{ij}^2/N = 1$ prior to any analyses. The shrunken regression coefficients are then obtained from the OLS estimators by $\hat{\beta}_j^{global} = \hat{c}\,\hat{\beta}_j^{OLS}$.

The second method extends the first by allowing parameter-wise shrinkage factors, that is, a different value of the shrinkage factor for each regression coefficient. Estimates of these parameter-wise shrinkage factors are again obtained by cross-validation calibration after standardizing the X_i. Parameter estimates can be obtained simply from the OLS estimates as $\hat{\beta}_j^{PW} = \hat{c}_j\,\hat{\beta}_j^{OLS}$ [43]. The parameterwise shrinkage method addresses one drawback of the global method, namely that it may shrink small coefficients too much. It is recommended that parameterwise shrinkage be applied subsequent to standard backward elimination due to the large number of parameters to be estimated if one starts with the full model. In this way, the pool of possible predictor variables is reduced first and then shrinkage is applied to provide some variance reduction [35]. This parameter-wise shrinkage can also be used directly as a technique for variable selection by setting coefficient estimates, $\hat{\beta}_j$ with negative shrinkage factors, that is $\hat{c}_j < 0$, to zero. [43].

Shrinkage methods provide some improvement over OLS in terms of mean square error of prediction, but generally do not reduce the number of predictors in the model. The variable selection methods discussed in section 1.2 reduce the number of predictors, but may not do so in an optimal way. In the next section, the ideas of shrinkage and variable selection are combined to develop an improved variable selection method.

1.4 COMBINING SHRINKAGE AND SELECTION

Newer methods in variable selection have attempted to combine shrinkage with variable selection in regression to address some of the drawbacks of standard variable selection methods. A number of methods which combine shrinkage and selection will be introduced here. The lasso, which will be the focus of this dissertation, will be introduced briefly here and described in more detail in Chapter 2.

1.4.1 Nonnegative Garrote

Breiman's nonnegative garrote [8] is similar in form to parameter-wise shrinkage proposed by Sauerbrei [35] in that each parameter coefficient is shrunk by some factor $\hat{c}_j$. Let $\hat{\beta}^{OLS}$ be the vector of OLS parameter estimates. Then the nonnegative garrote shrinkage factors, $\hat{c}_j$ minimize

$$\sum_k \left(y_n - \sum_k c_k \hat{\beta}_k^{OLS} x_{kn} \right)^2 \text{ subject to } \sum_{j=1}^{p} c_j \leq t \text{ and } c_j > 0, \tag{1.6}$$

where $t \geq 0$ is the shrinkage threshold [8]. Variable selection is achieved when the coefficient associated with a particular variable is shrunk to zero, removing it from the model. One potential drawback of the nonnegative garrote is that the parameter estimates depend on both the sign and magnitude of the OLS estimates, causing this method to perform poorly in situations where the OLS estimates perform poorly, for example in situations involving high correlation among predictor variables [41]. The lasso estimates are not based on the OLS estimates and in fact the lasso estimates may differ in sign from the OLS estimates.

1.4.2 Least Absolute Shrinkage and Selection Operator

The least absolute shrinkage and selection operator, or lasso, is a penalized regression method, where the L_1 norm of the regression parameters is constrained below a tuning parameter t, which controls the amount of shrinkage applied and the number of variables selected for inclusion in the model [41]. As in the nonnegative garrote, variable selection occurs when regression coefficients are shrunk to zero. The lasso parameter estimates are given by:

$$\hat{\beta} = \arg\min_{\beta} \sum_{i=1}^{N} (y_i - \beta_j' \mathbf{x}_i)^2 \text{ subject to } \sum_{j=1}^{p} |\beta_j| \leq t \quad (1.7)$$

where $t \geq 0$ is a tuning parameter. As this method will be the focus of this dissertation, further details on the lasso are reserved for Chapter 2.

1.4.3 Least Angle Regression and Related Approaches

The least angle regression algorithm (LARS) presented by Efron, Hastie, Johnstone, and Tibshirani [17] unites, under a common computational framework, three distinct, yet related, variable selection methodologies: forward stagewise linear regression, least angle regression, and the lasso. It is important to note the distinction between the least angle regression algorithm (LARS) and least angle regression as a model selection procedure. For each method, the algorithm proceeds in a stepwise manner through the pool of potential predictors, selecting a predictor at each step based on the correlation with the current residual vector. The lasso formulation of the LARS algorithm is of particular interest here because it provides an efficient algorithm for computing the lasso estimates needed for this research. Details of the modifications of the LARS algorithm needed to compute the lasso parameter estimates are in section 2.1.1.1.

1.4.4 Bridge Regression

Bridge regression [20] encompasses both ridge regression and the lasso as special cases by allowing the exponent in the constraint to vary. The bridge regression parameter estimates are given by

$$\hat{\beta} = \arg\min_{\beta} \sum_{i=1}^{N} (y_i - \beta_j' \mathbf{x}_i)^2 \text{ subject to } \sum_{j=1}^{p} |\beta_j|^{\gamma} \leq t \quad (1.8)$$

where the tuning parameter $t \geq 0$ and the exponent $\gamma \geq 0$ are estimated via generalized cross-validation. Ridge regression corresponds to $\gamma = 2$, and the lasso to $\gamma = 1$.

Fu [20] presents the results of a simulation study comparing bridge regression to OLS, the lasso, and ridge regression in the linear regression model. Each of the $m = 50$ data sets has $n = 30$ observations of $p = 10$ predictors. Each matrix has between-column pairwise correlation ρ_m drawn from a uniform distribution on the interval $(-1, 1)$. The true vector of regression coefficients, β_m, is drawn from the bridge prior,

$$\pi_{\lambda,\gamma}(\beta) = \frac{\gamma^{2^{-(1+1/\gamma)\lambda^{1/\gamma}}}}{\Gamma(1/\gamma)} \exp\left(-\frac{1}{2}\left|\frac{\beta}{\lambda^{-1/\gamma}}\right|^{\gamma}\right). \tag{1.9}$$

This prior distributions is a member of the class of elliptically contoured distributions [9]. As a special case, if we take $\gamma = 2$ in the bridge prior, the resulting distribution is normal with mean zero and variance λ^{-1}.

For fixed $\lambda = 1$, OLS, bridge regression, the lasso, and ridge regression are compared for $\gamma = 1, 1.5, 2, 3$, and 4. For $\gamma = 1$ and 1.5, bridge regression and the lasso have a similar significant reduction in both MSE and PSE over OLS, whereas for $\gamma = 2, 3$, and 4 both methods result in an increase in MSE over OLS. For all values of γ ridge regression has a moderate reduction in MSE and PSE, with similar amounts of reduction for all γ values. For $\gamma = 1$ and 1.5 bridge regression and the lasso outperform ridge regression. These results agree with those of Tibshirani in that the lasso method outperformed ridge regression in those cases ($\gamma = 1$ and 1.5) where the true beta values were either zero or relatively large in absolute value, and was outperformed by ridge regression when the true beta values were small, but nonzero ($\gamma = 2, 3$, and 4).

Because the performance of bridge regression and the lasso did not differ significantly for any value of γ and bridge regression provides a lesser degree of variable selection than the lasso for $1 < \gamma < 2$ and no variable selection for $\gamma \geq 2$, this method will not be considered further in this research.

1.4.5 Elastic Net

A recently proposed generalization of the lasso and LARS is the elastic net [48]. The elastic net provides variable selection in the $p > n$ case (where the lasso can select at most n predictors), improves performance in the case of highly correlated predictor variables (where the lasso is dominated by ridge regression) and improves selection when groups of predictors are highly correlated (where the lasso typically simply selects one representative predictor from the group).

The basic idea of the elastic net is to combine the ridge regression and lasso penalties. In the *näive elastic net*, a convex combination of L_1- and L_2- norms of the regression coefficients is constrained. The näive elastic net parameter estimates are obtained via the constrained minimization

$$\hat{\beta}_{nEN} = \arg\min_{\beta} \sum_{i=1}^{N} (y_i - \beta_j' \mathbf{x}_i)^2 \text{ subject to } (1-\alpha)\sum_{j=1}^{p} |\beta_j| + \alpha \sum_{j=1}^{p} \beta_j^2 \leq t \text{ for some } t. \quad (1.10)$$

Zou and Hastie [48] present empirical evidence via both a real data example and a simulation study indicating that the *näive elastic net* resulted in coefficient estimates that incurred 'double shrinkage' leading to an increase in the bias without a corresponding decrease in the variance. They modified their original procedure by rescaling to avoiding overshrinking while preserving the advantageous properties the elastic net. The elastic net is given by

$$\hat{\beta}_{EN} = \arg\min_{\beta} \beta' \left(\frac{X'X + \lambda_2 I}{1 + \lambda_2} \right) \beta - 2y'X\beta + \lambda_1 \sum_{j=1}^{p} |\beta_j| \quad (1.11)$$

Via their simulation study and real data example, Zou and Hastie [48] illustrate the properties of the elastic net and its performance relative to the lasso and ridge regression. The elastic net achieves better prediction error than both the lasso and ridge regression. The selection of groups of correlated predictors in the elastic net leads to the selection of larger models than the lasso. Whether the lasso or elastic net is a superior method depends on the goal of the analysis. If prediction is the goal, the lasso may be preferred because it selects only one representative predictor from highly correlated groups. However, if interpretation is the goal, the elastic net may be preferred because it will include all the predictors in a highly correlated group. Zou and Hastie propose the elastic net as a useful method in the analysis of microarray data, where the inclusion of highly correlated groups of predictors is preferred because these groups are biologically interesting.

2.0 LEAST ABSOLUTE SHRINKAGE AND SELECTION OPERATOR

2.1 LASSO BASICS

As described in section 1.4.2, the least absolute shrinkage and selection operator(lasso) constrains the L_1 norm of the regression parameters. Variable selection occurs when regression coefficients are shrunk to zero.

Consider the linear regression situation with $\mathbf{Y}$ the $n \times 1$ vector of responses y_i and $\mathbf{X}$ the $n \times xp$ matrix of predictors, with row vector $\mathbf{x}_i$ the values for the i^{th} subject. The lasso assumes that either the observations are independent or that the y_i are conditionally independent given the x_{ij}, where the x_{ij} have been standardized so that $\sum_i x_{ij}/n = 0$ and $\sum_i x_{ij}^2/n = 1$. In addition, the y_i have been centered to have sample mean 0. Under these assumptions, the lasso estimates are given by:

$$\hat{\beta} = \arg\min_{\beta} \sum_{i=1}^{n} (y_i - \beta_j' \mathbf{x}_i)^2 \text{ subject to } \sum_{j=1}^{p} |\beta_j| \leq t, \tag{2.1}$$

where $t \geq 0$ is a tuning parameter which controls the amount of shrinkage applied to the parameter estimates and, therefore, the degree of variable selection [41].

2.1.1 Computational Algorithms

In order for the lasso to be applicable in practical situations, an easily implemented, efficient computational algorithm is needed. In the paper introducing the lasso method, Tibshirani [41] presented two different algorithms. The first is based on a method proposed by Lawson and Hansen [24] used to solve linear least squares problems under a number of general linear inequality constraints. The second method reformulates the lasso problem to construct a quadratic programming problem with fewer constraints, but more variables, which can be solved by standard quadratic programming techniques. Many improvements of these algorithms have been suggested [30].

Osborne, Presnell, and Turlach [30] studied the lasso computations from the quadratic programming perspective, exploring the associated dual problem. The resulting algorithm was an improvement over those proposed by Tibshirani [41] and had the advantage of including the case where the number of predictors is larger than the number of observations. The exploration of the dual problem also provided improved estimates of the standard errors of the parameter estimates. Standard error estimation for the lasso parameter estimates will be discussed in detail in 2.4.

2.1.1.1 Least Angle Regression Algorithm The LARS algorithm with a small modification, provides efficient computation of the lasso parameter estimates. An additional constraint, $\text{sign}(\hat{\beta}_j) = \text{sign}(\hat{c}_j)$, where $\hat{c}_j = \mathbf{x}_j'(\mathbf{y} - \hat{\beta}'\mathbf{x}_j)$ i.e. the sign of any nonzero $\hat{\beta}_j$ in the model must agree with the sign of the current correlation is required [17] to obtain the lasso parameter estimates. The consequence of this restriction, in terms of computation, is that additional steps, compared with the unmodified LARS algorithm, may be required. In the regular LARS algorithm, once a covariate has entered the model, it cannot be removed, whereas with the lasso restriction in place, covariates can leave the model when the constraint above is violated. The LARS algorithm is easily implemented in the R software package version 2.2.1 in the *lars* library version 0.9-5 [32].

2.2 COMPARISON OF LASSO TO OTHER METHODS

The usefulness of the lasso method depends in large part on its performance in comparison with other variable selection methods and other types of parameter estimation. Simulation studies and real data examples have been used by several authors to illustrate the properties of the lasso method and to compare its performance with other standard methods. Vach, Sauerbrei, and Schumacher [43] compared four of the more recently developed variable selection methods: the global and parameterwise shrinkage factor methods (section 1.3), Breiman's nonnegative garrote (section 1.4.1), and Tibshirani's lasso (section 1.4.2) to backward elimination in ordinary least squares (OLS) in a simulation study. Four settings are considered in the simulation study, two involving independent covariates (A and B) and two involving pairwise correlated covariates (C and D). The correlated covariates condition is considered to examine the selection patterns for groups of correlated covariates. The true parameter values in each setting are:

$$\beta^{\mathrm{A}} = (0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2, 0.1, 0.0)'$$
$$\beta^{\mathrm{B}} = (0.8, 0.8, 0.6, 0.6, 0.4, 0.4, 0.2, 0.2, 0.0, 0.0)'$$
$$\beta^{\mathrm{C}} = (0.8, 0.8, 0.6, 0.6, 0.4, 0.4, 0.2, 0.2, 0.0, 0.0)'$$
$$\beta^{\mathrm{D}} = (0.8, 0.0, 0.6, 0.0, 0.4, 0.0, 0.2, 0.0, 0.0, 0.0)'$$

The methods are compared in terms of complexity of the selected model, distribution of the shrinkage parameters, selection bias, prediction error, and the bias and variance of the parameter estimates. Model complexity is measured by both the inclusion frequency of each variable, that is the $\Pr\{\hat{\beta}_j \neq 0\}$; and the average number of covariates selected. In settings A, B, and C, in terms of both measures, the lasso selected the largest models, followed by the nonnegative garrote and then both types of shrinkage and backward elimination. In setting D, the shrinkage methods select models that are larger than the nonnegative garrote. As in the other three settings, in setting D, the lasso selected the largest models and backward elimination selected the smallest models.

Selection bias is given by $E[|\hat{\beta}_j| - |\beta_j| \,|\, \hat{\beta}_j \neq 0]$ for $j = 1, \cdots, p$. This definition of selection bias differs from that typically used, for example Miller [29], in that the absolute values prevent under- and over-estimates from canceling out in small effects. [43] In terms of selection bias, the lasso is less biased in the case where the true parameter values are small, whereas the nonnegative garrote, parameterwise shrinkage, and backward elimination are least biased for large true parameter values. The authors conclude that overall the lasso performs well if one is aware of its propensity to underestimate large parameter values. The global shrinkage factor, while not a variable selection method, is useful in its reduction of average prediction error and the mean square error of the parameter estimates.

The average prediction error (APE) for a new observation (X^*, Y^*) is given by

APE($\hat{\beta}$) $= E[(Y^* - \hat{\beta}'\mathbf{X}^*)^2]$, which can be expressed as APE $= \sigma^2 +$ MSE where the mean square error (MSE) is MSE($\hat{\beta}$) $= E[\hat{\beta}'(\mathbf{X}^*) - \beta(\mathbf{X}^*)]^2$. Thus, comparisons based on MSE of predictions are the same as those based on APE.

Taken together, the results of this simulation study do not clearly identify one method as best in all circumstances. The relative performance of the various methods depends not only on the true parameter values, as seen earlier, but also on the goal(s) of the analysis. For example, none of the methods considered performs well if parsimony is the most important criterion: they all resulted, on average, in larger models than backward elimination. Vach, Sauerbrei, and Schumacher [43] hypothesize that in order to achieve a reasonable level of parsimony, some sacrifice in terms of the other criteria must be made.

2.3 SELECTION OF LASSO MODEL

The LARS algorithm provides a convenient, computationally efficient method for producing the full set of lasso coefficient estimates that avoids the computational burden of previously proposed lasso algorithms [41]. In fact, the entire set of lasso parameter estimates can be computed for an order of magnitude less computing time than previous methods [17]. However, because of the nature of the link between LARS, the forward stagewise method and the lasso, use of the LARS algorithm removes the automatic model selection provided by the direct use of the tuning parameter to control the amount of shrinkage and selection in the lasso.

In the LARS algorithm, a Mallows' C_p-type statistic is proposed for selecting the optimal model in LARS. An approximation of this statistic is given by

$$C_p(\hat{\beta}^{[\mathbf{k}]}) \cong (\| \mathbf{y} - \hat{\beta}^{[k]}\mathbf{X} \|^2)/(\bar{\sigma}^2) - n + 2k, \qquad (2.2)$$

where $\hat{\beta}^{[k]}$ is the vector of the k-step LARS parameter estimates and $\bar{\sigma}^2$ is the residual mean square error of regression on k variables. This C_p estimate applies only for the LARS selection method, not for the lasso or the forward stagewise method [17]. The proposed C_p-type statistic for model selection in LARS has been widely criticized, especially in the many discussions of LARS paper [17]. In particular, the first discussant, Ishwaran, shows in a simulation study that the use of C_p can lead to models that are too large. He suggested that accounting for model uncertainty through model averaging may improve the performance of the C_p statistic. Stine also criticizes the C_p statistic and proposes the S_p statistic, another penalized residual sum of squares estimate, to be used instead. This statistic is given by

$$S_p = RSS(p) + \hat{\sigma}^2 \sum_{j=1}^{p} 2j \log\left(\frac{j+4}{j+2}\right), \tag{2.3}$$

where p is the number of predictors in the current model and $\hat{\sigma}^2$ is "an honest estimate of σ^2" computed using the (conservative) estimated error variance from the model selected by the standard forward selection method. Using S_p to select the model size resulted in the selection of a model that is smaller than that selected by C_p and has smaller residual mean square error.

Leng, Li and Wahba [25] found that under the minimum prediction error criterion, LARS and the lasso are not consistent variable selection methods. A consistent variable selection method is one in which the probability of correctly identifying the set of important predictors tends to one as the sample size tends to infinity. Moreover, it is shown that the probability of selecting the correct model in LARS or the lasso is less than a constant not depending on the sample size. In simulation studies, the lasso method selected the *exact* true model with small probability between 10% and 30%. The authors are careful to point out that their criticisms are not with the LARS concept; they question only the validity of the use of prediction error as a criterion for selecting the tuning parameter. Other criteria may provide consistent variable selection.

The use of a form of the BIC for model selection with the lasso is proposed by Zou, Hastie and Tibshirani [49]. The authors present a more careful examination of the degrees of freedom for the lasso than do Efron, Johnstone, Hastie and Tibshirani [17]. They prove the following for the lasso: "Starting at step 0, let m_k be the index of the last model in the Lasso sequence containing k predictors. Then $\text{df}(\hat{\beta}^{[m_k]}) \cong k$." This implies that the degrees of freedom for the lasso estimates containing k nonzero coefficients, obtained by m_k steps, is approximately k. Note that the number of steps could be larger than the number of nonzero coefficients because predictors that have entered can exit the lasso model in later steps.

Given this approximation for the degrees of freedom, selection methods based on Akaike's Information Criterion (AIC) and the Bayesian Information Criterion (BIC) are derived for the lasso. Based on the properties of AIC and BIC as described in section 1.2 and supported by the results of their simulation study, Zou, Hastie and Tibshirani [49] recommend the use of BIC in selecting the lasso model when variable selection is the primary goal. BIC was shown to select the exact correct model with higher probability than AIC which conservatively included additional covariates. In comparison with the C_p-type statistic suggested in LARS, the BIC criterion selected the same 7 covariates from among 10 predictors and a smaller 11 variable model compared with the C_p 15 variable model from among 64 predictors.

Recall that the LARS algorithm produces the complete set of lasso parameter estimates, providing the 'best' model of each size (number of nonzero coefficients) k for the computation cost of the fit of a single least squares regression model. A theorem proved by Zou, Hastie and Tibshirani shows that the optimal lasso model is among the models in the LARS algorithm output, thus we need only choose between them. Computation of the BIC based on the output of the LARS is simplified by the following result. Let $\beta^{[m_k]}$ be the vector of lasso parameter estimates at the m^{th} step in the algorithm with k nonzero coefficients at a given iteration. To find the optimal number of nonzero coefficients, we need only solve [49]

$$k_{opt} = \arg\min_{k} \frac{\| \mathbf{y} - \beta^{[m_k]\prime}\mathbf{X} \|^2}{n\sigma^2} + \frac{\log(n)}{n}k. \tag{2.4}$$

Because of the easy of implementation using the lasso estimates provided by the LARS algorithm and the evidence pointing to the BIC as the 'best' stopping criterion proposed thus far, BIC for the lasso will be used to select final models in this research.

2.4 STANDARD ERRORS FOR LASSO

The usefulness of the lasso method in practice depends in part on the accuracy of the parameter estimates. In order to perform significance testing for individual parameter estimates, estimation of the standard errors of the lasso parameter estimates will be required. A number of standard error estimates have been proposed in the literature.

Tibshirani [41] presents two standard error estimates: one based on bootstrap resampling, and a closed form expression using an approximation based on ridge regression. The standard error estimate based on bootstrap resampling. The second standard error estimate is developed by exploiting the connection between the lasso and ridge regression. This method has the undesirable property of giving the estimate zero for any regression coefficient that was shrunk to zero by the lasso. Improvements on these standard error estimates have been proposed by Osborne, Presnell and Turlach [30] (see equation 2.7).

Tibshirani's estimate is

$$\hat{\text{var}}(\hat{\beta}_{lasso})^{TIBS} = (\mathbf{X}'\mathbf{X} + \alpha\mathbf{W}^{-})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X} + \alpha\mathbf{W}^{-})^{-1}\hat{\sigma}^2 \tag{2.5}$$

where $\hat{\sigma}^2$ is an estimate of the error variance, $\mathbf{W} = diag(|\hat{\beta}_j^{lasso}|)$ and α is chosen so that $\sum_j |\hat{\beta}_j^{lasso}| = t$. The improved estimate [30] is given by

$$\hat{\text{var}}(\hat{\beta}_{lasso})^{OPT} = (\mathbf{X}'\mathbf{X} + \mathbf{V})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X} + \mathbf{V})^{-1}\hat{\sigma}^2 \tag{2.6}$$

where, again, $\hat{\sigma}^2$ is an estimate of the error variance and $\mathbf{V}$ is a slightly more complicated expression than W given by

$$\mathbf{V} = \mathbf{X}' \left(\frac{1}{\| \hat{\beta}_{lasso} \|_1 \| \mathbf{X}'\mathbf{r} \|_\infty} \mathbf{r}\mathbf{r}' \right) \mathbf{X} \tag{2.7}$$

where $\mathbf{r} = \mathbf{r}(\tilde{\beta}) = (\mathbf{Y} - \tilde{\beta}'\mathbf{X})$ is the vector of residuals corresponding to β. The standard error estimate, $\hat{\text{var}}(\hat{\beta}_{lasso})^{OPT}$, presented by Osborne, Presnell and Turlach has been shown to be superior to that of Tibshirani and will be used in this research.

3.0 MISSING DATA METHODS

Almost all longitudinal data sets have missing data values, and the motivating data for this study is no exception. The goal of this research is to assess the performance of the variable selection methods described in chapters 1 and 2 in the presence of missing data. Major texts on missing data methods include *Analysis of Incomplete Multivariate Data* by J.L. Schafer [36] and *Statistical Analysis with Missing Data* by Roderick J.A. Little and Donald B. Rubin [26].

The following notation, taken from Little and Rubin [26], will be used in the discussion of missing data. Let $Y = \{y_{ij}\}$ denote an n by k rectangular data set without missing values. Define the *missing-data indicator matrix* $M = (m_{ij})$, such that $m_{ij} = 1$ if y_{ij} is missing and $m_{ij} = 0$ if y_{ij} is observed. The matrix M then describes the pattern of missing data.

3.1 CATEGORIZATION OF MISSINGNESS

Missing data are commonly classified based on two characteristics: the pattern of missing values and the missingness mechanism. Together, these classifications can indicate which method is appropriate for the missing values in a data set. Some methods are developed to be applied only with data that follow a specific pattern. For example, many methods are useful only in the case of a monotone missing data pattern. Other methods can be used with a general pattern of missingness, but some computational savings can be achieved if the data follow a special pattern.

3.1.1 Missing Data Patterns

Three broad categories of missing data patterns: monotone missingness, file matching and general missingness, are defined by Little and Rubin [26]. Consider a set of variables $y_1, \cdots, y_k$ observed on a set of individuals. A monotone missing data pattern is one in which the variables can be ordered in such a way that when y_j is missing for a given individual the variables y_{j+1} through y_k are also missing [37]. Subject attrition in longitudinal studies is one example of a monotone pattern of missing data. It is important to note that the covariates need not be collected over time for a monotone pattern of missing to exist. The file matching pattern of missingness occurs when two variables (or two sets of variables) are never jointly observed. Arbitrary missingness describes any pattern that cannot be classified as either monotone or file matching.

3.1.2 Missingness Mechanisms

The missingness mechanism attempts to answer, from a statistical perspective, the question of why data is missing. Meng [28] describes the missingness mechanism as the process that prevents us from observing the intended data. What is of central importance is the probabilistic relationship between the value that should have been observed (the intended data) and the fact that it was not observed. This relationship is defined statistically in terms of the conditional distribution of the missing data indicator matrix given the observed data.

The three general types of missing data mechanisms defined by Little and Rubin [26] are missing completely at random (MCAR), missing at random (MAR), and not missing at random (NMAR). To characterize the distinctions between these categories, let the conditional distribution of the missing data mechanism M, given the data $Y = (Y_{obs}, Y_{mis})$ be denoted by $f(M|Y_{obs}, Y_{mis}, \phi)$, where ϕ denotes unknown parameters related to the missing data mechanism.

Missing completely at random is the case when nonresponse and the data values (both missing and observed) are unrelated; that is, nonresponse is unrelated to both the value that should have been observed and was not, and to the other values in the data set. Under the MCAR assumption, the conditional distribution of the missing data mechanism given the data Y is given by

$$f(M|Y_{obs}, Y_{mis}, \phi) = f(M|\phi) \text{ for all } Y_{obs}, Y_{mis}, \phi. \tag{3.1}$$

The MCAR assumption is often too strong to be plausible in practical situations [28], except in the case where data is missing by design [26]. An example of data missing by design is the double sampling method, often used in survey sampling, where the entire sample is asked one set of questions and only a *preselected* subsample of respondents is asked an additional set of questions.

A more plausible, but weaker assumption, is that the data is missing at random (MAR). Under the MAR mechanism, the missingness depends only on the observed components of the data, Y_{obs}, not on the missing values, Y_{mis}. That is,

$$f(M, Y_{obs}, Y_{mis}|, \phi) = f(M|Y_{obs}, \phi) \text{ for all } Y_{mis}, \phi. \tag{3.2}$$

In other words, after other variables in the analysis have been controlled for, the missingness is unrelated to Y_{mis} [1].

If, in addition to meeting the MAR assumption, the parameters governing the complete data model, θ, and those governing the missing data mechanism, ϕ, are distinct in the sense that the joint parameter space of θ and ϕ is the Cartesian product of the parameter space of θ (Ω_θ) and the parameter space of ϕ (Ω_ϕ), i.e., ($\Omega_{(\theta,\phi)} = \Omega_\theta \times \Omega_\phi$), the missing data mechanism is called *ignorable*. Ignorability does not remove the need for missing data techniques, it simply means that an explicit model of the missingness mechanism is not required. In both Allison [1], and Little and Rubin [26], the distinctness assumption is essentially ignored, and ignorability is taken as an equivalent condition to MAR. If ignorability is erroneously assumed, the resulting inference is not improper, however, a loss of efficiency is incurred.

Data that does not meet the MAR criteria is said to be not missing at random (NMAR). Under this assumption, the fact that an observation is missing is related to the value of the intended data. Specification of a model for the missingness mechanism is difficult because in most situations the observed data provide little or no information about the missing data mechanism [27].

3.2 OVERVIEW OF METHODOLOGY

3.2.1 Deletion Methods

The simplest missing data method is complete case analysis. Observations that are not complete are simply deleted from the data set. This solves the problem of how to handle those cases where data are missing, but can lead to substantial bias in any resulting inference because the cases with complete data may not be a random subsample of all cases. Equally disconcerting, large quantities of data are likely to be discarded and a loss of precision is incurred due to the reduction in sample size. A similar method, called available case analysis, attempts to reduce the amount of data deleted. In this strategy, summary statistics are computed using all the data that is available for that particular statistic. For example, to compute the correlation between U and V, all observed pairs (U, V) are used, regardless of whether other variables in the data set are observed or not. Note that in this example, available case analysis may result in a covariance matrix that is not positive definite.

3.2.2 Likelihood-Based Methods

Maximum likelihood and Bayesian inference in the incomplete data case is similar to that in the complete data case. The likelihood function is derived and the maximum likelihood parameter estimates or posterior distributions are obtained. The difference is that the missing data mechanism must be accounted for in some way in the likelihood function, depending on the type of missingness mechanism.

Recall, the data is denoted $Y = (Y_{obs}, Y_{mis})$, where Y_{obs} is the observed data and Y_{mis} denotes the missing values. The joint probability distribution of Y_{obs} and Y_{mis} is given by $f(Y_{obs}, Y_{mis}|\theta)$.

For ignorable mechanisms, the likelihood is proportional to the marginal distribution of the observed data because the missingness does not depend on the unobserved values. Then the marginal density of Y_{obs} is given by

$$f(Y_{obs}|\theta) = \int f(Y_{obs}, Y_{mis}|\theta)\mathrm{d}Y_{mis}. \tag{3.3}$$

Then, under ignorability, the likelihood of θ based on the observed data Y_{obs} is

$$L_{\text{ign}}(\theta|Y_{obs}) \propto f(Y_{obs}|\theta) \text{ for } \theta \in \Omega_\theta. \tag{3.4}$$

From a Bayesian perspective, the posterior distribution for inference on θ based on the data Y_{obs}, and assuming a prior distribution $p(\theta)$ for θ is given by $p(\theta|Y_{obs}) \propto p(\theta) \times L_{\text{ign}}(\theta|Y_{obs})$.

When ignorability does not hold, the missing data mechanism must be explicitly modeled. Let $f(M, Y|\theta, \phi)$ be the joint distribution of M, the missing data indicator matrix, and $Y = (Y_{obs}, Y_{mis})$, where $f(M, Y|\theta, \phi) = f(Y|\theta)f(M|Y, \phi)$ for $(\theta, \phi) \in \Omega_{\theta,\phi}$. Then, the marginal distribution of the observed data is given by

$$f(Y_{obs}, M|\theta, \phi) = \int f(Y_{obs}, Y_{mis}|\theta)f(M|Y_{obs}, Y_{mis}), \tag{3.5}$$

which involves the term $f(M|Y_{obs}, Y_{mis})$ not included in equation 3.3 under the ignorability assumption. Specification of this term makes ML inference under nonignorable mechanisms difficult. The likelihood function for inference on θ is given by

$$L(\theta, \phi|Y_{obs}, M) \propto f(Y_{obs}, M|\theta, \phi) \text{ for } (\theta, \phi) \in \Omega_{\theta,\phi}. \tag{3.6}$$

From a Bayesian perspective, the posterior distribution of $p(\theta, \phi|Y_{obs}, M)$ is obtained by combining the likelihood in equation 3.6 with a prior distribution $p(\theta, \phi)$,
i.e., $p(\theta, \phi|Y_{obs}, M) \propto p(\theta, \phi) \times L(\theta, \phi|Y_{obs}, M)$.

3.2.2.1 EM Algorithm The maximization of the likelihood function in missing data cases often requires special computational techniques. The expectation and maximization (EM) algorithm is a popular tool for computing ML estimates with incomplete data proposed by Dempster, Laird and Rubin in 1977 [14]. The algorithm consists of two steps, the expectation (E) step and the maximization (M) step, which are repeated iteratively until convergence. A set of starting parameter values are required and are often obtained using complete-case analysis or available case analysis. While the choice of starting values for the algorithm is often not crucial when there is a low to moderate amount of missing information, using a number of different sets of starting values can be informative, illustrating features of the complete-data likelihood and can serve as a diagnostic tool.

In the notation of Little and Rubin [26], let $l(\theta|Y_{obs}, Y_{mis}) = \ln L(\theta|Y_{obs}, Y_{mis})$ denote the complete data log-likelihood and $\theta^{(t)}$ the current estimate of θ.

The E step computes the expected complete-data log-likelihood if $\theta^{(t)}$ were the true value of θ.

$$Q(\theta|\theta^{(t)}) = \int l(\theta|Y_{obs}, Y_{mis}) f(Y_{mis}|Y_{obs}, \theta = \theta^{(t)}) dY_{mis}. \tag{3.7}$$

This step does not fill in the individual data values that are missing rather, the functions of the data (sufficient statistics) appearing in the likelihood function are estimated [26].

The M step consists simply of the standard maximum likelihood estimates based on the estimated functions of the missing data and the observed data. The next value in the sequence, $\theta^{(t)}$ is found by maximizing $Q(\theta|\theta^{(t)})$; that is, finding the value $\theta^{(t+1)}$ such that $Q(\theta^{(t+1)}|\theta^{(t)}) \geq Q(\theta|\theta^{(t)})$ for all θ. The estimated parameter values obtained in the M step are then used in a subsequent E step. The algorithm continues iteratively until the parameter estimates converge.

3.2.3 Imputation

The basic premise of imputation is to fill in the missing data with plausible values and then to proceed with the analysis as if the data were completely observed [1]. One advantage of imputation methods is that once the missing values have been filled in, existing statistical software can be used to apply any statistical model or method. Imputation is a flexible method which can be used with any type of data and for any kind of model. Methods for generating imputations will be discussed in section 3.3.

Single imputation methods construct and analyze one completed data set. For example, mean imputation replaces the unobserved values of each variable with the mean of the available cases for that variable. The major drawback of single imputation methods is that the standard analytic techniques applied to the completed data set fail to account for the fact that the imputation process involves uncertainty about the imputed values [1]. The failure to account for this uncertainty leads to the underestimation of variances and the distortion of the correlation structure of the data, biasing the correlations towards zero. For this reason, single imputation is not recommended.

The uncertainty resulting from the missing values and the imputation process can be properly accounted for by creating multiple imputed data sets. Multiple imputation repeats the single imputation process a number of times creating several filled-in data sets which are each analyzed separately. The parameter estimates obtained from each of the filled in data sets are then combined in a way that incorporates the added uncertainty due to the missing values.

3.2.3.1 Combination Rules Once parameter estimates have been obtained for each of the completed data sets, a single combined parameter estimate, along with an appropriately adjusted variance estimate, are computed. The following notation for the combination rules is taken from Little and Rubin [26]. Let θ_d and W_d be the parameter estimate and associated variance for the parameter θ calculated from completed data set d for $d = 1, \ldots, D$.
The combined estimate is

$$\bar{\theta}_D = \frac{1}{D}\sum_{d=1}^{D}\hat{\theta}_d \tag{3.8}$$

The variability associated with this estimate has two components: the average within-imputation variance,

$$\bar{W}_D = \frac{1}{D}\sum_{d=1}^{D} W_d \tag{3.9}$$

and the between-imputation variance component,

$$B_D = \frac{1}{D-1}\sum_{d=1}^{D}(\hat{\theta}_d - \bar{\theta}_D)^2 \tag{3.10}$$

The total variability associated with $\bar{\theta}_D$ is

$$T_D = \bar{W}_D + \left(\frac{D+1}{D}\right) B_D \tag{3.11}$$

where $(D+1)/D$ is an adjustment for the finite number of imputations D.

3.2.4 Nonignorable Missingness Mechanisms

The nonignorable or NMAR mechanism is the most difficult missing data problem. Unlike under the MAR assumption, data involving nonignorable missing data mechanisms require an explicit model for the missingness mechanism. Because the observed data provide little information about the nature of the missingness mechanism, subjective information about the data and its collection must be used. Sensitivity analysis is recommended in any NMAR model because results can depend greatly on the choice of model for the missingness mechanism.

The use of nonignorable models in some situations has been controversial. Schafer points out that "with the complicated patterns of missingness often encountered in multivariate datasets, it may be quite difficult to specify any realistic mechanism for the nonresponse, ignorable or otherwise." [36] In their review article, Schafer and Graham [37] claim that the complex modeling required for a nonignorable mechanism may not be worth the resulting reduction in bias. The work of Collins, Schafer and Kam supports this claim, finding that simply implementing an inclusive strategy when building an imputation model may result in an acceptable amount of bias without the hard work and potential for model misspecification [11]. An inclusive imputation model strategy involves including additional variables in the imputation model that are not of interest in the complete data model, but may provide useful information to improve the imputation process. More details on the inclusive strategy and the study by Collins, Schafer and Kam are found in section 3.3.3.

3.3 GENERATING IMPUTATIONS

The generation of imputed values is most easily motivated from a Bayesian perspective. A parametric complete data model is combined with a prior distribution to obtain the posterior predictive distribution of the missing values conditioned on the observed data. Imputed values are then generated by sampling from this distribution. If the posterior distribution is of a simple form, such as a normal distribution, sampling from it is straightforward. However, in many cases sampling from the posterior predictive distribution is a difficult task requiring the use of sophisticated techniques (see section 3.3.2). In cases where the missing data are generated by a nonignorable mechanism, a model for the nonresponse mechanism is also incorporated into the posterior predictive distribution. In practice, the prior distribution of the parameters is often assumed to be a noninformative prior or conjugate with the likelihood.

Two broad types of methods for generating imputations are to assume the posterior distribution is of a standard form from which sampling is straightforward or to use sophisticated techniques for obtaining a sample from a complex distribution. On example of each approach will be discussed in detail.

3.3.1 Assuming Normality

Multivariate normality is one of the most common assumptions for the posterior predictive distribution. It has been shown in many situations that methods based on the normality assumption perform well even in cases where the data are far from normally distributed [36]. Well-known transformation techniques can also be applied to variables that clearly violate the multivariate normality assumption to improve the performance of the imputation procedure. The imputation model is applied only to the missing values in the data set; the normality assumption has no impact on variables that have no missing data. Under both the multivariate normal and multivariate t distribution assumptions, reliable parameter estimation can be obtained using a 'surprisingly' small, between 2 and 10, number of imputed data sets in cases where the fraction of missing information is not too large [26]. Little and Rubin [26] define the fraction of missing information about θ due to the missing data as, $\hat{\gamma}_D$, the ratio of the estimated between-imputation variance and the total variance. That is,

$$\hat{\gamma}_D = \left(\frac{D+1}{D}\right)\left(\frac{B_D}{T_D}\right) \tag{3.12}$$

3.3.2 Multiple Imputation by Chained Equations

Multivariate imputation by chained equations (MICE) is essentially a Gibbs sampler modified to provide imputations from missing data values. Methods of this type are also known as variable-by-variable imputation methods or regression switching methods. Before giving details of the MICE algorithm, some background on the Gibbs sampler will be presented.

3.3.2.1 Gibbs Sampler The Gibbs sampler is an algorithm that allows for the generation of random variables from a complicated joint probability distribution (or target distribution) by generating draws from a series of full conditional distributions of this target distribution. The algorithm is useful in those situations in which it is difficult to sample from the target distribution, but draws from the conditional distributions are easily obtained. The Gibbs sampler framework converts a k-dimensional problem into k 1-dimensional problems.

The following description of the Gibbs sampler is taken from Tanner [39]. The target distribution is $p(\theta)$ where $\theta = (\theta_1, \theta_2, \cdots, \theta_d)$. Given a starting point $\theta^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \cdots, \theta_d^{(0)})$ sampling is done systematically from the conditional distributions by the following scheme.

$$\begin{array}{ll} \text{Sample } \theta_1^{(i+1)} & \text{from } p(\theta_1|\theta_2^{(i)}, \cdots, \theta_d^{(i),Y}) \\ \text{Sample } \theta_2^{(i+1)} & \text{from } p(\theta_2|\theta_1^{(i+1)}, \theta_3^{(i)}, \cdots, \theta_d^{(i),Y}) \\ \quad\vdots & \\ \text{Sample } \theta_d^{(i+1)} & \text{from } p(\theta_d|\theta_1^{(i+1)}, \cdots, \theta_{d-1}^{(i+1),Y}) \end{array}$$

The above sample scheme is known as a *systematic scan* Gibbs sampler because the algorithm proceeds from the first component of θ to the last component with each component visited once. More complicated visiting schemes may improve computational efficiency and allow for the preservation of transformations, constraints and interactions in the data set[46].

3.3.2.2 MICE The MICE algorithm and uses a Gibbs sampler to obtain random draws from the target distribution, treating the missing values as parameters. The imputations are then random draws from the joint distribution of the missing data and the observed data, $p(X_{mis}, X_{obs})$. The observed data set, X, is partitioned into $(d+1)$ parts with X_0 representing the completely observed variables and $(X_1, \cdots, X_d)$ the variables with missing values. In many situations obtaining random draws directly from this joint distribution is difficult, however, draws from the conditional distribution $p(X_i|X_j \text{ for all } j \neq i)$ are more easily obtained. In these situations, a Gibbs sampler can be constructed to generate imputations more easily.

$$\begin{aligned}
&\text{For } X_1 \text{ draw } X_1^{t+1} \text{ from } p(X_1|X_2^t, X_3^t, \cdots, X_k^t)\\
&\text{For } X_2 \text{ draw } X_2^{t+1} \text{ from } p(X_2|X_1^{t+1}, X_3^t, \cdots, X_k^t)\\
&\qquad\qquad\vdots\\
&\text{For } X_k \text{ draw } X_k^{t+1} \text{ from } p(X_k|X_1^{t+1}, X_2^{t+1}, \cdots, X_{k-1}^{t+1})
\end{aligned}$$

A set of starting values $(X_1^{(0)}, \cdots, X_d^{(0)})$ are obtained as a random draw from some known probability distribution.

The major assumption of MICE is that a multivariate distribution to which the specified set of conditional distributions converges to exists theoretically. The existence of this distribution is not guaranteed. A set of conditional distributions with no corresponding multivariate distribution are said to be incompatible conditional distributions. More precisely, two conditional distributions $f(x|y)$ and $g(y|x)$ are compatible if and only if their density ratio can be factor into the product of two functions, that is, $\frac{f(x|y)}{g(y|x)} = u(x)v(y)$, for some integrable functions u and v[45]. The use of multiple imputation with incompatible conditional distributions has been shown to be reasonably robust. Imputations generated with incompatible distributions has been shown via simulation to provide reasonable results [45], although more work in this area is needed.

The variable-by-variable approach taken by the MICE algorithm allows for the inclusion of variables of mixed type: both continuous and categorical [31]. This is an improvement over the normality assumption under which handling categorical variables is more difficult.

Assessing the convergence of the MICE algorithm is can be difficult because what is required is to assess convergence in distribution rather than assessing convergence to a particular value. Several strategies have been proposed for assessing convergence of the Gibbs sampler and the MICE algorithm. The number of iterations required for convergence is surprisingly small in comparison to most Markov chain Monte Carlo methods, in part because no burn in period is required to ensure the independence of successive draws, the imputed values are independent because for a given variable all draws are independent[46]. One method of assessing convergence is to increase the number of iterations and check for noticeable differences in results[44]. Another method for convergence compares parallel sequences, with convergence achieved when the sequences overlap and are free of trend. This method is easily implemented with the MICE algorithm because parallel draws are automatically produced.

3.3.3 Included Variables

An imputation model consists of two parts: the choice of a set of donor variables and the choice of a statistical model representing the relationship between the variable with missing data and its donors. A donor variable is one that is known to be associated with a variable with missing data (the target variable), and is either completely observed or is observed in more cases than its target [31].

The choice of donor variables to include in the imputation model is crucial to providing accurate imputed values. Collins, Schafer and Kam [11] conducted an extensive simulation study investigating the inclusion of what they term auxiliary variables in the imputation model. Auxiliary variables are those variables included in the analysis solely for the purpose of improving the missing data model. Such variables improved the performance of both maximum likelihood and multiple imputation. Because similar results were obtained by both methods, only the multiple imputation results were presented.

While auxiliary variables are not informative in terms of the hypothesis of interest, they may provide useful information regarding the missing data mechanism. A restrictive variable selection strategy incorporates few auxiliary variables, while an inclusive strategy utilizes all or almost all of the available auxiliary variables. In assessing and comparing the performance of both strategies, standardized bias, root mean square error, coverage of confidence intervals and the average length of confidence intervals were compared. Values of the standardized bias greater than 40% were considered to be a significant amount of bias and coverage levels for a 95% confidence interval were considered poor if they dropped below 90%.

Van Buuren, et al. [44] present a strategy for selecting a group of donor variables from a large data set. First, all variables of interest in the ultimate analysis should be included in the imputation model. Second, add those variables that are related to the 'cause' of missingness. Third, variables that are highly correlated with the target variables. In the final step, those variables added as donor variables that contain a high amount of missing data must be removed from the model. One advantage of the MICE approach is that because it proceeds variable by variable, different donor variables can be used to impute the value of each target variable, potentially allowing the use of donor variables that might otherwise have been excluded. In addition, the variable by variable approach allows for the easy inclusion of both continuous and categorical target variables because the type of model used for each target variable can be adjusted to match its measurement type.

In the Collins, Schafer and Kam study, the simulated data consist of 1000 samples of size 500 of 3 variables X, Y, and Z from a multivariate normal distribution. Variables X and Z are always observed and variable Y is observed or missing, with variable Z as a possible 'cause' of the missingness of Y. Missing data was created within the simulated data sets, by four different mechanisms: MCAR, MAR-linear, MAR-convex and MAR-sinister, as described below, at 25% and 50% missingness rates. Two different correlation structures were considered in the data generation: $\rho_{XY} = 0.6$, $\rho_{YZ} = 0.4$, $\rho_{XZ} = 0.24$; and $\rho_{XY} = 0.6$, $\rho_{YZ} = 0.9$, $\rho_{XZ} = 0.54$. As a result of these correlation choices, X and Z are conditionally independent given Y. Thus, when Z is not observed, the missingness in Y will appear to depend only on Y and not on X. Thus, when Z is included in the data set, the data are truly MAR because the 'cause' of missingness is included in the model. Moreover, when Z is excluded from the model, the data are truly NMAR because missingness now depends solely on the value of Y. Note that the names given to the missing data mechanisms (MCAR, MAR-linear, MAR-convex and MAR-sinister) are as the data are only truly MAR when Z is included in the imputation model.

The MCAR condition creates missingness in Y at the specified probabilities independently of X, Y, and Z. In the MAR-linear condition, the probability of missingness is taken to be linearly related to the value of Z. Specifically, the values of Z are divided into quartiles and missingness probabilities $(0.1, 0.2, 0.3, 0.4)$ are assigned to each quartile for 25% missingness and probabilities $(0.2, 0.4, 0.6, 0.8)$ are assigned for 50% missing data. In the MAR-convex condition, the values of Z are again divided into quartiles with the probabilities of missingness set to make values at the tails less observed than those at the center of the distribution of Z. The missingness probabilities assigned are $(0.4, 0.1, 0.1, 0.4)$ and $(0.8, 0.2, 0.2, 0.8)$ for 25%

missing data and 50% missing data, respectively. The MAR-sinister condition was created specifically to introduce bias into the relationship between X and Y when Z is not included in the missing data model. The probability of missingness is not a function of Z directly, but of the correlation between X and Z. The implementation of this mechanism is a bit more complicated. To start, the 500 data points are randomly divided into 10 groups of 50 points each. The sample correlation between X and Z is computed within each group. Based on the degree of correlation, the groups are assigned to either a high or low correlation group. The probabilities of missingness for the low and high correlation strata are $(0.1, 0.4)$ and $(0.2, 0.8)$, respectively.

The simulation study focused on four questions of interest, addressing the impact of the inclusion or exclusion of particular categories of variables from the analysis. The results of each question are summarized below.

> Question 1: What is the impact of the omission of auxiliary variables that are both correlated with Y and related to missingness?

When the missing data mechanism was truly MAR (Z was included in the imputation model) under all structures (linear, convex or sinister), both correlation structures and both missing data percentatges, multiple imputation performed well in estimating all parameters. When Z was excluded from the imputation model, i.e. the data were NMAR, the results were 'surprisingly robust' suggesting that the use of methods intended for situations where the missing data mechanism is ignorable (or MAR) may provide acceptable results even in cases where the missing data mechanism is truly nonignorable. Recall, the omission of Z from the missing data model makes the missing data mechanism nonignorable and biased estimates more likely.

Multiple imputation did not perform uniformly well under all types of missing data mechanisms. Under the linear mechanism, the estimation of the mean of Y was affected in all situations, with high levels of bias in parameter estimates and low coverage probabilities for confidence intervals. The estimation of standard deviations, regression parameters and correlations performed reasonably well in most cases with the exception of the case where $\rho_{YZ} = 0.9$ with 50% missingness. In both the convex and sinister mechanisms, the estimation of the mean of Y was largely unaffected by the missing data. The case of $\rho_{YZ} = 0.4$ with 25%

missingness produced good results. Other correlation and rate of missingness combinations were not as good in terms of bias and coverage.

The results for question one indicate that the structure of the missingness mechanism impacts the estimation of population quantities in different fashions. The type of MAR mechanism has an impact on the effectiveness of missing data methods and, therefore, simulation studies should examine a number of different MAR mechanisms. In most studies, the MAR linear mechanism is the only type of MAR mechanism considered. This may give a false measure of the performance of an estimate of the standard deviation, regression parameters and correlations when the MAR structure is not linear.

> Question 2: Will including variables that are correlated with Y, but *not* related to missingness, improve the precision of estimates without negatively impacting bias or coverage?

The second question addresses the improvements in the precision of parameter estimates obtained by including covariates correlated with Y, but *not* correlated with the 'cause' of missingness Z, as auxiliary variables. The simulation study used to address this question focused on a data set with missing rate of 50% imposed in a MCAR fashion and included one of two Z variables, the first with $\rho_{YZ} = 0.4$ and the second with $\rho_{YZ} = 0.9$. Multiple imputation based inference was obtained both with and without the Z variable in the imputation model. In all cases the bias of the estimates was within the acceptable range and the coverage percentage of the confidence intervals was not adversely impacted by the inclusion of the auxiliary variable. Confidence interval coverage actually increased above the nominal level with the inclusion of Z particularly in the $\rho_{YZ} = 0.9$ case.

> Question 3: Will including variables correlated with Y, but *not* related to missingness under a nonignorable missingess mechanism reduce bias?

Because of the difficulties encountered when the missing data mechanism is nonignorable, the use of auxiliary variables correlated with Y as a way of obtaining reasonable parameter estimates without explicitly modeling the missing data mechanism is considered. The simulation study addressing this question considered cases with missing values in Y at a 50% rate created under all three MAR mechanisms. Three sets of variables were considered: X and Y only; X, Y and Z where $\rho_Z = 0.4$; and X, Y and Z where $\rho_Z = 0.9$. Estimation of all parameters with only X and Y in the model had considerable bias, often in a negative direction implying that the parameter estimates were too small. This bias was reduced with the inclusion of the auxiliary covariate, Z. This suggests that an inclusive strategy is the best course when building a missing data model either using MI or ML under a nonignorable mechanism, however, the incorporation of these extra covariates is most easily performed with MI due to the limitations of the software packages currently available.

> Question 4: Is there any disadvantage to including variables that are *uncorrelated* with Y, i.e. what is the negative impact of including extraneous variables?

The results of the third question suggest an improvement due to the inclusion of an auxiliary covariate that is correlated with Y. Question 4 addresses the potential cost of including covariates which are completely *uncorrelated* with X and Y. Comparisons were made between cases with 5, 25 and 50 'junk' variables included. While the estimation in the case of 5 extra covariates is within the acceptable range, in the cases with 25 and 50 extra variables, there is a noticeable increased in the bias of the estimates and reduction in the coverage of the confidence intervals for σ_Y^2, ρ_{XY} and β_{XY}. This is likely due in part to the increasing number of parameters with a fixed sample size of 500. The authors suggest that, as the effective sample size is increased, these biases should disappear.

4.0 SIMULATION STUDY OVERVIEW

The major goal of this research is to examine the impact of missing data on the performance of variable selection methods, in particular the lasso and stepwise regression. Ultimately, the results of the simulation study will be used to inform the application of the lasso to the psychobiological data described in Chapter 6.

4.1 SIMULATION OF DATA

The simulation study will focus on the performance of variable selection in the multiple linear regression model. The performance of ordinary least squares regression, stepwise regression, ridge regression and the lasso method will be examined, with each method applied to the complete data and data sets containing missing data. The simulated data will cover a variety of conditions that are often encountered in real data situations. The factors and factor levels considered: sample size, number of predictor variables, correlation structure and missing data characteristics, are intended to provide information about the performance of variable selection in a number of practical situations.

In each combination of factors considered, $N = 1000$ sets of simulated covariates, $X_1, \cdots, X_p$, will be generated from a multivariate normal distribution. Five different correlation structures were selected for consideration to attempt to mirror the degree of correlation often encountered in real data. The independent correlation structure agrees with the usual assumption in linear regression. The equicorrelated structures ($\rho = 0.25$, $\rho = 0.5$) are considered to mirror those cases where there is a moderate to high correlation between all covariates under consideration. The autoregressive structures ($\rho = 0.25$, $\rho = 0.5$) were selected as a convenient way to create data sets with a higher level of correlation between some covariates and little to no correlation between other covariates. The balance between sample size, n, and the number of covariates, p, is known to influence parameter estimation and variable selection in regression. The combinations we will consider in this study are: $n = 50$, $p = 5$ and $p = 10$; $n = 100$, $p = 10$ and $p = 20$; and $n = 200$ and $p = 20$.

The multiple linear regression model is of the form $\mathbf{Y} = \mathbf{X}\beta + \epsilon$ where y_i is the response for subject i and $\mathbf{x}_i$ is the vector of predictor variables for the i^{th} subject, for $i = 1, \cdots, n$. It is usually assumed $E(\epsilon) = 0$ and $E(\epsilon'\epsilon) = \sigma^2\mathbf{I}_n$. The least squares estimates of the regression parameters β are given by $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$.

The rows of the design matrix are assumed to be normal random vectors, generated from a distribution with mean zero and variance one and are then transformed to exhibit the given correlation structure. Due to the assumptions of the lasso method, the variables are centered and scaled prior to parameter estimation. The vector used in the estimation of the prediction error, X^*, is generated in the same manner, with one X^* generated in each of the 1000 simulations. The entries of the error vector are assumed to have unit common variance. Investigation of the impact of other values of this variance is left to future research.

In the generation of the response variable Y based on the simulated covariates, two beta vectors will be considered: β_1 consists of several large nonzero coefficients and coefficients that are exactly zero, while β_2 includes coefficients that are small in magnitude along with coefficients that are exactly zero. For $p = 5$, $\beta_1 = (3, 1.5, 0, 2, 0)$ and $\beta_2 = (0.85, 0.85, 0, 2, 0)$. For $p > 5$, the beta vectors are constructed by repeating the pattern for $p = 5$ to obtain a vector of the needed length, i.e. for $p = 10$, $\beta_1 = (3, 1.5, 0, 2, 0, 3, 1.5, 0, 2, 0)$.

4.1.1 Missing Data

The missing data mechanisms used to simulate missing data will be based on those used by Collins, Schafer, and Kam [11] as outlined in 3.3.3. The linear, convex and sinister mechanisms will be extended to the case of more than three covariates. These mechanisms were considered both under MAR and NMAR conditions, that is both including and excluding the cause of missingness from the imputation model. The MCAR assumption is too strong to be practical in most real data situations and therefore will not be considered [28]. Missing data will be created under each mechanism at 25 percent and 50 percent levels.

The data analysis will be completed using multiple imputation to account for the missing data. This method was selected because the variable selection methods under consideration can easily be performed on the imputed data sets and their results compared.

MI will be implemented using the *MICE* package developed for R [32] by Van Buuren and Oudshoorn which provides MI as described in 3.3.2 [46].

4.2 COMPARISON OF RESULTS

The assessment of the performance of the variable selection methods in the various missing data situations will focus on the accuracy of predictions based on future observations using the estimated model and on the accuracy with which the regression parameters are estimated.

The average mean square error of the estimation of the regression parameters will be computed as

$$\hat{\text{MSE}}_{\tilde{\beta}} = \frac{1}{B}\sum_{i=1}^{B} \parallel \tilde{\beta}_i - \beta \parallel^2, \tag{4.1}$$

where $\tilde{\beta}$ is the estimate of β obtained from the variable selection method under consideration. In addition, the actual coverage probability of a nominal 95% confidence interval for β will be computed.

Mean square error of prediction will be used to assess the accuracy of predictions based on future observations. The average MSE of prediction will be computed by

$$\text{MSE}_{pred} = \parallel \tilde{\beta}'\mathbf{x}^* - \beta'\mathbf{x}^* \parallel^2, \tag{4.2}$$

where $\mathbf{x}^*$ is a vector 'future' observations.

Assessment of the degree to which each method selects the correct model or a model containing the correct model in quantitative terms is an area for future research. The number of methods considered and the number of conditions under which these methods were compared made this assessment difficult. In selected cases, boxplots illustrating the variation of hte parameter estimates about their ture values will be presented.

5.0 RESULTS UNDER THE MISSING AT RANDOM ASSUMPTION

Recall that the simulated data consist of $N = 1000$ data sets of size n with p parameters from a multivariate normal distribution. Two types of beta vectors were considered, beta 1 consists of repetitions of $(3, 1.5, 0, 2, 0)$ and the second, β_2, consists of repetitions of $(0.85, 0.85, 0, 2, 0)$. Missing data was created under 3 different mechanisms: linear, convex and sinister. This chapter summarizes the performance of the various models considered when the data were missing at random, that is the 'cause' of missingness was included in the imputation model. Because of computational difficulties, parameter estimates for 50 percent missing data with $n = 50$, $p = 5$ and $n = 100$, $p = 20$ were unattainable.

5.1 PREDICTION ERROR

One goal of a regression model is to provide accurate predictions of future outcome based on the selected set of predictor variables. The mean square error of prediction is one measure of the accuracy of predictions. In this study, prediction mean square error was computed as the mean of the squared deviation between the true outcome y and the predicted outcome $\hat{y}$, where $\hat{y}$ was computed using the estimated regression parameters on a new data set. Tables containing the prediction error results for the $n = 50$, $p = 5$ case are presented in the text here, the entire set of prediction error tables can be found in Appendices A and B. In comparing the prediction error, five different correlation structures are considered: independent, autoregressive with $\rho = 0.25$ and $\rho = 0.50$ and equicorrelated with $\rho = 0.25$ and $\rho = 0.50$.

Table 1: MAR, Beta 1, independent, n=50, p=5, MSE of Prediction

		Linear Missing		Convex Missing		Sinister Missing	
	Complete	25%	50%	25%	50%	25%	50%
OLS	54.6914	68.2929	126.1785	68.3981	132.0979	68.6678	136.1288
Stepwise	53.0424	65.6315	122.9691	68.7254	129.7173	67.0359	133.8803
Ridge	54.8543	67.6266	121.8536	67.5476	130.0766	67.3922	133.5684
LASSO	55.1775	64.1683	119.6536	66.2485	127.2515	65.3432	127.3544

5.1.1 Beta 1 - Independent

For the complete data case, there is little difference between the methods considered. Excluding the $n = 100$, $p = 10$ case, the percent difference between the best and worst performing methods was small ranging from 0.2 to 3.9 percent. In the $n = 100$, $p = 10$ case, ridge regression performed poorly compared to the other methods, with a 10.9 percent difference between it and the best method. Excluding ridge regression, the percentage change from the best to the worst method drops to 4.9 percent.

In the incomplete data case, the lasso performed best in the $n = 50, p = 5$ situation in terms of mean square error of prediction. The lasso resulted in a decrease of between 1.8 and 4.7 percent over its closest competitor. Stepwise regression performed best in the $n = 50$, $p = 10$; $n = 100$, $p = 10$; and $n = 100$, $p = 20$ situations, with the percent decrease in MSE prediction ranging from 1.9 to 3.6 percent, 0.1 to 4.5 percent, and 0.2 to 1.4 percent, respectively. In the $n = 200$, $p = 20$ case, stepwise regression performed best in the 25 percent missing data cases, with percent decreases ranging from 2.7 to 3.8 percent across missing data mechanisms. In the 50 percent missing data cases, OLS outperformed stepwise regression by between 0.2 and 2.0 percent across the missing data mechanisms.

Table 2: MAR, Beta 1, autoregressive 0.25, n=50, p=5, MSE of Prediction

		Linear Missing		Convex Missing		Sinister Missing	
	Complete	25%	50%	25%	50%	25%	50%
OLS	55.5467	65.7170	125.7308	66.1434	123.1692	68.6117	132.3267
Stepwise	53.7703	64.9161	124.8871	63.6737	122.2519	68.0792	131.6722
Ridge	56.1997	65.8453	121.9880	65.3392	120.2885	68.3793	126.6165
LASSO	53.5634	63.4803	111.1292	62.2114	113.6424	66.4535	121.6055

5.1.2 Beta 1 - Autoregressive 0.25

For the complete data analysis as under the independent correlation structure, the methods performed similarly, with the exception of one case in which ridge regression performed poorly. In the $n = 100$, $p = 20$ case, ridge regression yields a prediction mean square error 9.8 percent higher than the best method. Excluding ridge regression, this percentage drops to 5.5 percent.

The lasso performs best in $n = 50$ $p = 5$ situation, with a small decrease in MSE prediction. For the 25% missing data cases, the lasso yields around a 2.0 percent reduction, where in the 50% missing cases, the decrease in between 4 and 9 percent. In $n = 50$, $p = 10$ case, ridge regression seemed to perform best, but with only a slightly smaller MSE prediction than the lasso (less than one percent). In the $n = 100$, $p = 10$ case, the lasso performs best followed closely by ridge regression in all but the convex and sinister 50% cases, where it is slightly outperformed by ridge regression. The percentage difference is between 1.0 and 2.5 percent. In the $n = 100$, $p = 20$ case, ridge regression performs best in the linear and convex cases, and stepwise performs best in the sinister mechanism with 25 percent missing data. In the $n = 200$, $p = 20$ case, OLS performs best in the linear mechanism but by less than one percent over its closest competitor. The difference between the best and worst case in the linear 25% and 50% are 2 percent and 4 percent, respectively. Stepwise regression performs best in the convex and sinister situations.

5.1.3 Beta 1 - Autoregressive 0.50

For the complete data case, again all methods perform similarly with the exception of OLS in the $n = 50$, $p = 10$ case, in which it has a mean square error of prediction 7.0 percent higher than the best method. Excluding OLS in the $n = 50$, $p = 10$ case, the percent difference between the best and worst method ranges from 1.9 to 3.6 percent.

The lasso performs best in all missing data cases, except the $n = 100$, $p = 20$ case, where it is outperformed in the 25 percent linear missing data case by stepwise regression (0.4 percent) and by ridge regression in the 25 convex missing data case (1.1 percent). In the complete data case, in all but the $n = 50$, $p = 10$ case, the methods perform similarly, with between 1.9 and 3.5 percent difference between the best and worst methods. In the $n = 50$, $p = 10$ case, OLS performs poorly compared to the other methods, differing from the best method by seven percent. Excluding OLS, this percentage is only 3.6 percent.

The largest percentage decrease in MSE prediction exhibited by the lasso is in the $n = 50$, $p = 5$ case. For the 25% missing data situations, the percent decrease is between 1.2 and 3.4 percent. For the 50 percent missing data cases, the percentage jumps to between 7.8 and 12 percent. For the other parameter-sample size combinations the 50 percent missing data cases yield a larger percent decrease for the lasso, but not as dramatically as in the $n = 50$, $p = 5$ case. For example, in the $n = 100$, $p = 10$ case, the percentages range from 2.6 and 3.4 percent for the 25 percent missing data cases and from 5.3 and 6.9 percent in the 50 percent missing data cases.

5.1.4 Beta 1 - Equicorrelated 0.25

In the complete data cases, stepwise regression performs uniformly best, but all methods performed similarly with percent differences between the best and worst method ranging from 2 to 3.6 percent across sample sizes and number of parameters. In the incomplete data cases, the lasso method performs best in terms of mean square error of prediction under all conditions. For the $n = 50$, $p = 5$ case, the lasso outperforms its closest competitor by a small amount (between 1.9 and 3.4 percent) in the 25 percent missing data cases and by a slightly larger amount (6.8 to 9.0 percent) in the 50 percent missing data cases. For

Table 3: MAR, Beta 1, autoregressive 0.50, n=50, p=5, MSE of Prediction

		Linear Missing		Convex Missing		Sinister Missing	
	Complete	25%	50%	25%	50%	25%	50%
OLS	54.8208	67.2951	115.0031	66.0225	112.9062	66.9540	118.0505
Stepwise	53.7677	66.2884	112.6266	65.4333	118.0544	66.9004	116.2644
Ridge	53.9582	67.1818	112.1185	66.9024	113.3605	67.3742	115.0649
LASSO	53.8426	65.5285	102.3797	63.8824	99.3837	64.7271	106.0956

$n = 50$, $p = 10$, the percent change is similar to the 50 percent missing data case for $p = 5$, ranging from 6.1 to 10.4 percent. The $n = 100$, $p = 10$ case follows a similar patter to the $n = 50$, $p = 5$ case, with slightly higher percentages, ranging from 3.1 to 6.4 percent in the 25 percent missing data situation and 8.3 to 10.5 percent in for 50 percent missing data. The $n = 200$, $p = 20$ case also follows the same general pattern and again has an increased percent decrease over the $n = 100$, $p = 10$ case, for 25 percent missing data ranging from 8.9 to 9.9 percent and for 50 percent missing data ranging from 13.5 to 14.8 percent.

Table 4: MAR, Beta 1, equicorrelated 0.25, n=50, p=5, MSE of Prediction

		Linear Missing		Convex Missing		Sinister Missing	
	Complete	25%	50%	25%	50%	25%	50%
OLS	55.1837	68.1253	124.6244	68.0459	120.3575	67.5179	126.0433
Stepwise	53.7651	67.1208	122.8441	65.8799	117.7967	66.0484	124.1935
Ridge	54.2000	67.9600	120.7277	67.9454	118.4358	67.6775	122.5943
LASSO	54.6877	64.8556	109.9059	64.6156	109.7634	63.8032	112.3873

5.1.5 Beta 1 - Equicorrelated 0.50

For the complete data the results under this correlation structure reveal larger differences in performance between the methods. The percent difference between the best and worst methods ranges from 1.4 to 8.6 percent. Excluding OLS, the worst performer, this percentage drops to between 1.2 and 5.7 percent.

The results of this correlation structure are similar to the equicorrelated $\rho = 0.25$ structure. When $n = 50$ and $p = 5$ in the 25 percent missing data case, the lasso is the best method in the linear and sinister missing data mechanisms, with a percent decrease over its closest competitor, stepwise regression of 4.3 and 3.5 percent, respsectively. In the convex missing data case, ridge regression outperforms stepwise regression by 4.1 percent. In the 50 percent missing data case the lasso is the best performing method under all missing data mechanisms, with percent decreases of 8.5, 9.4 and 9.2 percent in the linear, convex and sinister missing data mechanisms, respectively.

The lasso is again the best performer in the $n = 50$, $p = 10$ case with percentage decreases in MSE of prediction similar to the $n = 50$, $p = 5$ 50 percent missing data case, ranging from 8.9 to 9.4 percent over its closest competitor. For $n = 100$, $p = 10$, the percent decreases in MSE of prediction of the lasso over stepwise regression for the 25 percent and 50 percent missing data situations are 1.3 to 6.4 and 9.6 to 13.4, respectively. For $n = 100$, $p = 20$ with 25 percent missing data the lasso with the best performing method with the percent decrease in mean square error of prediction over the closest competitor ranging from 13.5 to 16.5 percent. For $n = 200$, $p = 20$, the percent decrease in MSE of prediction of the lasso over its closest competitor for 25 percent missing data ranges from 7.5 to 8.1 percent and for 50 percent missing data it ranges from 11.5 to 12 percent. Both the $n = 100$, $p = 20$ and $n = 200$, $p = 20$ follow a pattern similar to the $n = 50$, $p = 5$ case.

5.1.6 Beta 2 - Independent

For the complete data cases, the percent difference between the best and worst performing methods ranges from 1.7 to 5.7 percent. The highest percentage difference occurs in the $n = 100$, $p = 10$ case, where ridge regression and the lasso perform poorly compared to OLS

Table 5: MAR, Beta 1, equicorrelated 0.50, n=50, p=5, MSE of Prediction

		Linear Missing		Convex Missing		Sinister Missing	
	Complete	25%	50%	25%	50%	25%	50%
OLS	56.5982	63.0679	130.3273	66.5053	114.2230	68.1162	118.4149
Stepwise	51.7356	64.0486	124.0407	62.2755	114.7393	66.5011	117.4116
Ridge	53.4086	65.1481	117.2939	59.7011	113.0369	67.1418	114.6097
LASSO	54.8507	60.3818	107.3554	63.9047	102.4125	64.1699	104.1159

and stepwise regression. The percent difference between OLS and stepwise regression is 3.0 percent.

Under the independent correlation structure in the 25% missing data situation with $n = 50$ and $p = 5$ in the linear and sinister cases, the lasso outperforms stepwise regression only slightly by 0.7 percent and stepwise and ridge regression by 2.0 percent, respectively. In the 50 percent missing data case, the lasso method was the best performer, with percent increases in the prediction MSE, of 3.5, 3.5 and 3.2 percent over ridge regression in the linear, convex and sinister missing data mechanisms, respectively.

The 25 percent linear and sinister missing data cases with $n = 50$ and $p = 10$, ridge and stepwise regression are tied for best performer. In the 25% convex missing data situation when $n = 50$ and $p = 10$ shows stepwise regression with a 3.0 percent decrease in prediction mean square error over OLS, its closest competitor.

In the other sample size, parameter combinations, there was no clearly best method and the difference between the best method and its closest competitor is less than 2 percent, with the exception of a few select cases. For $n = 100$, $p = 10$, in the 25 percent linear and convex situations and for all mechanisms with 50 percent missing data, stepwise regression and the lasso have similar performance, while the lasso method outperforms stepwise regression by 1.2 percent in the 25 percent sinister missing data case.

In the highest sample size situation with $n = 200$ and $p = 20$, stepwise regression is the dominant method in the 25 percent linear and convex cases, while OLS and ridge regression

Table 6: MAR, Beta 2, independent, n=50, p=5, MSE of Prediction

		Linear Missing		Convex Missing		Sinister Missing	
	Complete	25%	50%	25%	50%	25%	50%
OLS	55.4664	61.9923	83.4059	61.7500	86.5321	63.2000	84.4271
Stepwise	54.2861	60.2779	83.4562	61.1181	85.1575	62.4822	83.5028
Ridge	55.8484	62.5328	80.7426	59.9136	84.8708	62.5470	80.7851
Lasso	54.1250	59.8385	77.9304	61.0997	81.9225	61.2615	78.1704

are tied as the best methods in the 50 percent linear and convex missing data cases. In the 25 percent sinister missing data case, stepwise regression and the lasso perform similarly, while in the 50 percent sinister missing data case, OLS and ridge regression are tied as the best methods.

5.1.7 Beta 2 - Autoregressive 0.25

For the complete data case, the percent difference between the best and worst performing methods ranges from 1.7 to 6.4 percent across the number of parameter, sample size combinations.

In the $n = 50$, $p = 5$ case, the lasso has the best performance for both the 25 percent and 50 percent missing data cases. In the 25 percent missing data case the lasso outperforms its closest competitor by between 0.1 and 2 percent, while in the 50 percent missing data case, the decrease in MSE of prediction ranges from 2.7 to 4.2 percent. For the $n = 50$, $p = 10$ case, the performance of ridge regression and the lasso is quite similar in the linear and convex conditions with 25 percent missing data, with ridge regression beating the lasso by 0.3 and 0.1 percent, respectively. In the 25 percent sinister missing data case, however, ridge regression outperforms the lasso by 2.7 percent.

When $n = 100$ and $p = 10$, the lasso shows a small degree of improvement in MSE of prediction over closest competitor, with percent decreases ranging from 0.3 to 1.8 percent.

Table 7: MAR, Beta 2, autoregressive 0.25, n=50, p=5

		Linear Missing		Convex Missing		Sinister Missing	
	Complete	25%	50%	25%	50%	25%	50%
OLS	53.3920	61.3784	81.0871	62.2971	84.5883	62.9392	84.6691
Stepwise	50.8729	64.9137	79.0989	60.4346	82.6154	61.9896	83.5984
Ridge	54.3701	63.1675	78.7286	60.5885	80.7941	62.7609	81.4928
Lasso	52.8584	61.3054	75.4485	59.1969	78.6286	61.2443	79.0849

In the 50 percent sinister missing data case, the lasso is outperformed slightly, 0.5 percent, by ridge regression. In the $n = 100$, $p = 20$ case with 25 percent linear missing data stepwise regression performs best, 2.9 percent decrease over ridge regression. Ridge regression enjoys a 8.6 percent decrease in MSE of prediction over the lasso in the 25 percent convex missing data case and a 0.8 percent decrease over stepwise regression in the 25 percent sinister missing data case.

In the final situation with $n = 200$, $p = 20$, with 25 percent missing data, the lasso is the best method by a small margin in the linear and convex missing data types, outperforming stepwise regression by one percent and ridge regression by 0.2 percent, respectively. In the 25 percent sinister missing data case, the lasso is the worst method, whereas OLS is the best method, outperforming stepwise regression by a tiny 0.05 percent margin. Under 50 percent missing data OLS is the best method by a small margin, outperforming its closest competitor by between 0.3 and 1.3 percent.

5.1.8 Beta 2 - Autoregressive 0.50

For the complete data case, the percent difference between the best and worst method ranges from 3 to 4.2 percent across the number of parameter, sample size combinations.

In the $n = 50$, $p = 5$ case, the lasso demonstrates the best performance of the methods considered. This difference is small in the 25% missing data cases, at less than 2 percent, and

Table 8: MAR, Beta 2, autoregressive 0.50, n=50, p=5, MSE of Prediction

		Linear Missing		Convex Missing		Sinister Missing	
	Complete	25%	50%	25%	50%	25%	50%
OLS	53.8760	61.5450	82.4335	62.1179	84.2433	62.2811	86.7815
Stepwise	51.9671	63.0262	80.7499	60.0994	83.7700	61.7423	85.5888
Ridge	54.2307	62.5154	79.6514	60.2089	81.2988	63.4752	81.8031
Lasso	53.4871	60.5544	76.6797	58.9193	78.5940	62.5772	79.3891

only slightly larger in the 50% missing and complete data cases, at about 3%. The lasso was also the best method in the $n = 50$, $p = 10$ case, with decreases over its closest competitor of 0.3%, 2.6%, 1.1% and 2.8% in the complete data, 25% linear, convex and sinister missing data cases, respectively. Recall that the 50% missing data case was not estimable.

The lasso again performed well in the $n = 100$, $p = 10$ case, with a slightly larger percent decrease. The differences ranged from 0.7% in the complete data case, to 5.8% in the 50% linear missing data case. The largest differences were seen in the 50% missing data cases. Stepwise and ridge regression performed almost identically in the $n = 100$, $p = 20$ case, the largest difference occurring in the complete data where stepwise regression was 2.6% lower in terms of mean square error of prediction. For $n = 100$, $p = 20$ in the complete data and all missing data types, stepwise and ridge regression performed almost identically. The lasso method performed poorly, stepwise and ridge regression had prediction mean square errors between 2.6 and 4.7 percent less than the lasso.

In the final parameter combination, $n = 200$, $p = 20$ there was no clear best method across all situations. In the 25% missing case, stepwise regression performed best with the lasso a close second. In the 50% missing data case, OLS performed best with ridge regression a close second. The largest difference between closest competitors was seen in the complete data case, where the prediction MSE for stepwise regresion was 2.3% less than that for OLS.

5.1.9 Beta 2 - Equicorrelated 0.25

With complete data, the percent difference between the best and worst method ranges from 2.9 to 4.5 percent across the number of parameter, sample size combinations.

Under the equal correlation structure with $\rho = 0.25$, the performance of the various methods depends on the sample size, number of parameters combination being considered. When $n = 50$ and $p = 5$ the lasso method performed best, with the greatest amount of decrease in mean square error of prediction in the 50% missing data case, where the percent decrease ranged from 2.7% to 4.2%. In the $n = 50$, $p = 10$ case, the lasso was again the best performer, with a larger amount of decrease, ranging from 7.1% to 9.4%, over its closest competitor.

In the $n = 100$, $p = 10$ case, there is a small amount of difference between the methods considered. In the linear missing data cases both 25% and 50% the lasso was 1.2% and 1.6%, respectively better. In the convex missing data and 25% sinister missing data cases, the lasso and ridge regression had similar performance with less than 0.5 percent difference between methods. There was a moderate 2.5% decrease in mean square error of prediction in the sinister 50% missing data situation, with the lasso outperforming stepwise regression.

In the $n = 100$, $p = 20$ case with 25 percent linear missing data stepwise regression has a prediction MSE 1.1 percent lower than ridge regression, while in the 25 percent convex and sinister cases, ridge regression and stepwise regression differ by 0.3 and 0.4 percent, respectively.

In the $n = 200$, $p = 20$ case, the type and percentage of missing data had a differing impact on the performance of the methods. In the complete data case, stepwise regression had a mean square error of prediction 2.3% lower than its closest competitor, OLS. The linear cases shows a 0.6 percent difference between stepwise regression and ridge regression under 25 percent linear missing data, while OLS enjoys the same percentage decrease in prediction MSE over ridge regression under 50 percent linear missing data. In the convex and sinister methods, however, there was a large percentage difference between methods. The lasso method outperformed its closest competitor by 6.2% and 11.9% in the convex cases and 6.7% and 14.4% in the sinister cases with 25 and 50 percent missing data, respectively.

Table 9: MAR, Beta 2, equicorrelated 0.25, n=50, p=5, MSE of Prediction

		Linear Missing		Convex Missing		Sinister Missing	
	Complete	25%	50%	25%	50%	25%	50%
OLS	54.2568	61.5160	81.0871	62.2971	84.5883	62.9392	84.6691
Stepwise	52.1352	63.0013	79.0989	60.4346	82.6154	61.9896	83.5984
Ridge	54.2883	62.7029	78.7286	60.5885	80.7941	62.7609	81.4928
Lasso	53.5933	60.4873	75.4485	59.1969	78.6286	61.2443	79.0849

5.1.10 Beta 2 - Equicorrelated 0.50

For the complete data case, the percent difference between the best and worst method ranges from 2.6 to 7.3 percent across the number of parameter, sample size combinations. In the $n = 50$, $p = 5$ case ridge regression is the worst performer with a MSE of prediction 7.3 percent higher than the best performer. Excluding this method, the percentage difference drops to 5.2 percent.

Under this correlation structure the $n = 50$ sample size had the greatest percent difference between methods. For $p = 5$, the complete data case had stepwise regression 3.9% smaller than the lasso, its closest competitor. In the 25% missing data percentages, there was a small percent difference in MSE prediction, from 1.1 to 1.8 percent across all missing data types. OLS performed best in the linear missing data case, while the lasso performed bed in the convex and sinister mechanisms. In the 50% missing data percentages the lasso exhibited a larger percentage difference, 4.9 percent under the linear mechanism, to 2.9% in the convex missing data type and 5.8 percent in the sinister missing data mechanism. For $p = 10$, the difference in the complete data case is only 0.6%. In the 25% missing data cases, the lasso performed 8.4%, 5.2% and 9.3% better than its closest competitor in the linear, convex and sinister missing data types, respectively.

In the $n = 100$ and $n = 200$ cases, there was little difference between methods. The largest differences occurred in the complete data cases, where stepwise regression was ap-

Table 10: MAR, Beta 2, equicorrelated 0.50, n=50, p=5, MSE of Prediction

		Linear Missing		Convex Missing		Sinister Missing	
	Complete	25%	50%	25%	50%	25%	50%
OLS	52.9322	60.9278	84.6885	62.3759	83.4234	62.9092	89.5624
Stepwise	50.1963	66.2506	83.6840	60.4004	82.6522	62.4638	88.3228
Ridge	54.1270	63.6931	81.9715	60.7036	80.1711	62.8453	84.6423
Lasso	52.2217	61.6363	77.9630	59.2910	77.8270	61.5606	79.7528

proximately 2.0% than ridge regression. In the missing data cases, the percent difference between closest competitors ranged from 0.2 to 1.0%. The difference between the best and worst methods in the missing data cases ranged between 0.5% and 6.9%. In the $n = 100$, 10 cases the lasso was the dominant method with a slight decrease, between 0.2 and 0.8 percent in prediction MSE over the other methods, with the exception of the 50 percent sinister missing data case, where ridge regression was 0.8 percent below the lasso in terms of MSE. In the $n = 100$, $p = 20$ case, stepwise regression and ridge regression have similar performance, differing by less and 1.0 percent in all missing data mechanisms.

Finally, for $n = 200$, $p = 20$, stepwise regression enjoys a slight edge, less than 0.7 percent, over the lasso in the linear and sinister 25 percent missing data cases, while OLS enjoys a less than 0.7 percent decrease over ridge regression in the linear and convex 50 percent missing data cases. In the 25 percent convex missing data case, the lasso, stepwise regression and ridge regression have essentially the same prediction MSE. In the sinister 50 percent missing data case, OLS has a 0.7 percent lower MSE than stepwise regression.

5.1.11 Overall Results

In the complete data, there was little variation between methods in terms of performance as measured by the mean square error of prediction. This is not true under the various missing data conditions. The performance varied across the true values of beta considered. Under beta 1, composed of larger nonzero coefficients made up of repetitions of $(3, 1.5, 0, 2, 0)$, the lasso method was the sole best performing method in the majority of cases, with a higher degree of correlation between the covariates. Under beta 2 consisting of smaller nonzero coefficients made up of repetitions of $(0.85, 0.85, 0, 2, 0)$, no one method is dominant as in the case of beta 1. For the smallest sample size $n = 50$, the performance of the methods was similar for beta 1 and beta 2, with the lasso performing best in most case for $p = 5$ and for the stronger correlation structures, autoregressive with $\rho = 0.50$ and equicorrelated with $\rho = 0.25$ and $\rho = 0.50$. Under the autoregressive with $\rho = 0.25$ correlation structure, beta 1 and beta 2 both show ridge regression and the lasso as the best methods. Under the independent correlation structure the prediction error patterns differ slightly for beta 1 and beta 2 with stepwise regression the best method under beta 1 and ridge regression and stepwise regression tie for the best method under beta 2. For the moderate sample size, $n = 100$ the methods under which beta 1 and beta 2 are more accurately estimated differs more than in the small sample size case. When $p = 10$ with $n = 100$, under beta 1 the independent correlation structure, stepwise regression performs best, while under the autoregressive structure with $\rho = 0.25$ the lasso and ridge regression perform similarly well and under the remaining structures the lasso is the best method. While for beta 2, under all correlation structures, the lasso, in some cases tied with one or more other methods, is the best performer. With $p = 20$ for $n = 100$, there is more variation in performance across missing data types in both beta 1 and beta 2. In both cases, mixtures of methods perform well, with little agreement. For beta 1, the lasso seems to be the best method under the equicorrelated structures, whereas for beta 2 ridge regression and stepwise regression are better performers. For $n = 200$, $p = 20$, there is little agreement on the best performers between beta 1 and beta 2 in the weaker correlation structures, with a variety of methods performing well. However, for the equicorrelated structures, the lasso again performs well under beta 1 for both $\rho = 0.25$ and $\rho = 0.50$, while under beta 2 the lasso performs well only in the $\rho = 0.25$ case. There is no clear best method in the $\rho = 0.50$ case, with the lasso tied or outperformed by other methods.

Based on these results, the lasso method is preferable for more highly correlated data with missing data and a true beta vector containing a few large effects. The preferred method is more situation dependent when the true beta vector consists of smaller effects. The percent of missing data did not have a large impact on the performance pattern, but the missing data mechanism did have an effect, most predominantly for beta 2.

5.2 CONFIDENCE INTERVAL COVERAGE

For each regression parameter the actual coverage of a nominal 95% confidence interval was computed. The minimum of the true coverage probabilities was selected as a summary measure to compare the results of the methods across conditions. The results for the two extremes of correlation considered, the independent and equicorrelated with $\rho = 0.5$, are compared in detail below. Tables containing the parameter estimates, standard errors and confidence interval coverage probabilities are presented in appendices C and D.

The standard errors of the regression parameter estimates for the incomplete data case are larger than those in the complete data case because of the added uncertainty resulting from the use of multiple imputation. This increase in the standard errors for the parameter estimate affects the true coverage probabilities of the associated confidence intervals because it increases the length of the confidence interval itself, making it more likely to cover the true parameter value. The increase in the standard errors due to the imputation method was consistent across methods, for a fixed sample size a number of parameters the increase in the standard errors due to imputation was approximately constant across methods. The ratio was not constant between the two beta vectors under consideration. The ratio was larger when beta 1 was the true beta vector, which is reflected in the confidence interval minimum coverage probabilities. A summary of the ratio of the complete data standard errors to the incomplete standard errors for both β_1 and β_2 under the missing at random condition is given in table 11 (see 5.2). Because of the consistency across methods, the ratio in OLS for the regression coefficient for X_1 is used as a representative of each sample size, number of parameters combination.

Table 11: MAR - Incomplete Data versus Complete Data Standard Error Ratio

MAR Beta 1										
	Independent					**Equicorrelated 0.50**				
	n50 p5	n50 p10	n100 p10	n100 p20	n200 p20	n50 p5	n50 p10	n100 p10	n100 p20	n200 p20
Linear 25%	1.5	1.7	2.4	3.0	4.3	1.4	3.4	2.0	6.2	3.1
Linear 50%	4.7	na	6.7	na	11.6	3.9	na	7.0	na	11.5
Convex 25%	1.6	4.1	2.6	3.1	4.2	1.5	3.6	2.0	6.1	3.3
Convex 50%	4.6	na	6.5	na	10.7	3.9	na	7.0	na	11.7
Sinister 25%	1.6	4.9	2.4	3.0	4.3	1.5	3.9	2.0	6.4	3.4
Sinister 50%	4.9	na	7.0	na	12.1	4.3	na	7.4	na	12.0
MAR Beta 2										
	Independent					**Equicorrelated 0.50**				
	n50 p5	n50 p10	n100 p10	n100 p20	n200 p20	n50 p5	n50 p10	n100 p10	n100 p20	n200 p20
Linear 25%	1.3	2.3	1.6	3.3	2.3	1.3	2.2	1.6	3.3	2.3
Linear 50%	2.6	na	3.8	na	5.0	2.6	na	3.8	na	5.2
Convex 25%	1.3	2.4	1.6	3.3	2.2	1.3	2.3	1.6	3.3	2.3
Convex 50%	2.6	na	3.7	na	5.5	2.6	na	3.7	na	5.5
Sinister 25%	1.3	2.4	1.6	3.4	2.3	1.4	2.3	2.3	3.4	2.3
Sinister 50%	2.7	na	4.0	na	5.4	2.8	na	5.6	na	5.4

5.2.1 Beta 1 - Independent

For the complete data, the coverage probabilities were below the nominal level, however, in all but a small number of cases the minimum coverage probabilities remained above 90 percent. The lasso method performed poorly with the complete data, it had the lowest minimum coverage probability in every case and in all but two cases was below 90 percent. Across all combinations of sample size and number of parameters with the complete data, OLS, stepwise regression and ridge regression had similar minimum coverage probabilities.

For $n = 50$, $p = 5$, with 25 percent missing data, all methods performed poorly, with all minimum coverage probabilities below 90 percent. There was little variation between the methods and no clearly dominant method. For the 50 percent missing data, the performance of all methods improves. OLS has the highest minimum coverage probability in each case with all probabilites increasing over the complete data. The $n = 50$ $p = 10$ case with 25 percent missing data is similar to the 50 percent missing data case with $n = 50$, $p = 5$. OLS is again the best performer with increased minimum coverage over the complete data. The $n = 100$, $p = 10$ case is similar to $n = 50$, $p = 5$. With 25 percent missing data, all methods perform poorly, in all but two cases dropping below 90 percent. With 50 percent missing data, the performance again improves, with OLS as the dominant method. In the $n = 100$, $p = 20$ case, the performance of all methods in all conditions is the worst of any sample size number of parameter combinations. In both missing data percentages, the coverage probabilities are all below 77 percent, with many below 65 percent with 50 percent missing data. OLS is again the dominant method, with the best performance in all but one case. In the $n = 200$, $p = 20$ case, the performance of all methods improves. With 25 percent missing data, the coverage probabilities remain below 90 percen, but are above 79 percent in all cases. OLS remains the dominant method.

5.2.2 Beta 1 - Equicorrelated 0.50

The comparative performance of the methods considered across situations in the equicorrelated $\rho = 0.5$ case differs from the independent case. The decreased performance of the lasso in the independent case does not occur here.

For the complete data, the methods perform similarly well with a few exceptions. In the complete data generated for the convex missing data, ridge regression exhibited poor performance compared with the other methods in that case. In addition, all methods performed poorly in the data generated for the linear 25 percent missing data case compared with the data generated for the convex and sinister 25 percent missing data mechanisms.

In the $n = 50$, $p = 5$ case with 25 percent missing data, there is a decrease in performance from the linear to convex to sinister missing data mechanisms. In the linear case, stepwise regression was the best performer, but all methods have minimum coverage probability above 90 percent. In the convex case, OLS and the lasso are tied for the best method with stepwise and ridge regression drop below 90 percent. In the sinister case, all methods drop below 90 percent with the lasso as the best performer. With 50 percent missing data, OLS is the best method in the linear case, while the lasso is the best performer in the convex and sinister cases. In the convex and sinister mechanisms, there was little variation between the methods. In the linear case there was more variation, with stepwise regression and the lasso lagging behind OLS and ridge regression. For $n = 50$, $p = 10$, stepwise regression lags behind the other methods under all three missing data mechanisms. OLS and the lasso are tied for the best method in the linear case, with ridge regression a close second. In the convex case, OLS is the best method with the lasso and ridge regression close behind. The lasso is the best performing method in the sinister case, with OLS and ridge regression close behind.

For the $n = 100$ $p = 10$ case, with 25 percent missing data, under all mechanisms the minimum coverage probabilities are all below 90 percent with all values at or below 80 percent in the linear case. In all three cases, the lasso is the best performing method, although there is not much variation across methods. With 50 percent missing data in the linear mechanism the lasso is the best performer, with OLS and ridge regression a close second and third. Stepwise regression lags behind the other methods. In the convex missing data mechanism, with 50 percent missing data OLS is the best method with a minimum coverage probability 1.6 percent higher than the closest competitor. Under the sinister missing data mechansim with 50 percent missing data stepwise regression is the best performer with a 2.8 percent increase over its closest competitor. With $n = 100$ and $p = 20$ there is little variation across methods in this case. In the linear case, ridge regression is the best performer by one percent over stepwise regression. In the convex case, ridge regression is the best method with a 0.2 percent increase over OLS. In the sinister case, the OLS is the best method by 0.1 percent over the lasso and ridge regression.

In the $n = 200$, $p = 20$ case, there is again little variation across methods in this case. In the 25 percent missing data case, under the linear mechanism, OLS is the best method with OLS and stepwise regression performing similarly well. Under the convex and sinister mechanisms stepwise regression is the best method with 2.3 and 1.0 percent difference between stepwise regression and the worst performing method. With 50 percent missing data, OLS is the best method in the linear and sinister cases, with a 0.4 percent advantage over ridge regression in the linear case and a 0.2 percent advantage over the lasso in the sinister case. The lasso is tie with ridge regression and 0.1 percent higher than OLS in terms of minimum coverage probability in the convex missing data case.

5.2.3 Beta 2 - Independent

As in the beta 1 independent case, the lasso method lags behind the other methods in terms of minimum coverage probability in the complete data cases with beta 2, repetitions of $(0.85, 0.85, 0, 2, 0)$. The other methods perform similarly in most cases. All methods perform poorly, with minimum coverage probabilities below 90 percent with the complete data generated for the $n = 50$, $p = 10$ and $n = 100$, $p = 20$ 25 percent convex missing data and for the $n = 100$, $p = 10$ 25 percent linear missing data case.

For the $n = 50$ $p = 5$ case, all methods perform similarly in each of the missing data mechanisms in both the 25 and 50 percent missing data percentages, no one method outperforms the other methods by a great margin. With 25 percent missing data, the minimum coverage probabilities range from 87.8 to 90.1 percent. The lasso method is the best method in both the linear and convex missing data mechanisms, while ridge regression is best in the sinister case. For the 50 percent missing data case, OLS is the best performer in each case. The minimum coverage probabilities are more variable for the 50 percent missing data, ranging from 83.3 to 91.7 percent. There is also more variability between the methods, with OLS outperforming its closest competitor by between 1 and 3 percent in the 50 percent missing data, compared with only 0.3 to 1.9 percent edge in the 25 percent missing data.

For the $n = 50$ $p = 10$ case, OLS outperforms the other methods under each missing data mechanism by a substantial margin, ranging from 5.8 to 6.0 percent higher than the closest competitor. OLS is the only method for which the minimum coverage probability remains above 90 percent, whereas the other methods have poor performance, with minimum coverage probabilities between 85 and 88.5 percent.

In the $n = 100$, $p = 10$ case with 25 percent missing data there is little variation between the methods. OLS, the best performer under each missing data mechanism, outperforms its closest competitor by between 0.2 and 0.8 percent. With 50 percent missing data, the degree of variability increases. In the linear case, stepwise regression outperforms its closest competitor by 1.2 percent, while in the convex and sinister cases OLS outperforms its closest competitor by 3.5 and 1.8 percent, respectively. For $n = 100$ $p = 20$, OLS is again the dominant method, outperforming its closest competitor by between 2.2 and 3.5 percent. The three other methods, with the exception of stepwise regression under the sinister missing data mechanism, have minimum coverage probabilities below 90 percent, ranging from 81.2 to 88.9 percent.

The performance of all methods decreases substantially in the $n = 200$, $p = 20$ case with all minimum coverage probabilities below 90 percent. With 25 percent missing data, the minimum coverage probabilities range from 78.3 percent to 87.2 percent. The lasso method is by far the worst method, with minimum coverage probability between 2.9 and 6.6 percent below its closest competition. OLS is the best method, resulting in coverage probabilities between 1.7 and 2.0 percent larger than its closest competitor. With 50 percent missing data, OLS is again the dominant method. OLS outperforms its closest competitor by 5.8, 8.1 and 4.6 percent in the linear, convex and sinister mechanisms, respectively. The lasso is again the worst method, with minimum coverage probabilities in the 60 to 70 percent range.

5.2.4 Beta 2 - Equicorrelated 0.50

With complete data, most methods perform similarly within each missing data mechanism and missing data percentage combination. In the $n = 100$, $p = 20$ case with 25 percent missing data, and in the $n = 200$, $p = 20$ case with both 25 and 50 percent missing data, the performance of all methods is decreased in the data sets generated for the convex missing data, in each case the minimum coverage probabilities drop below 90 percent, while they are above 90 percent in the corresponding linear and sinister mechanisms.

With incomplete data, for the $n = 50$, $p = 5$ case with 25 percent missing data, all methods perform poorly, with the minimum coverage probability exceeding 90 percent in only one case. The lasso is the best performing method in the linear and convex cases, outperforming its nearest competitor by 2.5 percent in both cases. Under the sinister missing data mechanism, OLS is the best method, with ridge regression a close second. With 50 percent missing data, OLS is the best performing method under each missing data mechanism, outperforming the closest method by 0.4, 3.4 and 4.5 percent in the linear, convex and sinister mechanisms, respectively over ridge regression.

For $n = 50$, $p = 10$, stepwise regression is the worst performer, with minimum coverage probabilities below 90 percent in each case. The lasso is the best performer in both the linear and convex cases, outperforming ridge regression by 1.1 and 1.8 percent respectively. Under the sinister missing data mechanism, OLS is the best performer, outperforming the lasso by 0.6 percent.

When $n = 100$ and $p = 10$ and 25 percent missing data is imposed all method have coverage probabilities below 90 percent. OLS is the best method, outperforming its closest competitor by 0.8, 1.1 and 2.5 percent in the linear, convex and sinister mechanisms, respectively. With 50 percent missing data imposed in most cases the minimum coverage probabilities exceed 90 percent, and OLS remains the best method under the linear and sinister mechanisms, outperforming stepwise regression, its closest competitor, by 4.1 percent in both cases. Under the convex mechanism, ridge regression outperforms OLS by 1.7 percent. For $n = 100$ with $p = 20$, OLS is the best performer under each missing data mechanism, exceeding stepwise regression by 2.1, 3.6 and 2.5 percent in the linear, convex and sinister mechanisms, respectively. In most cases, the minimum coverage probability is below 90 percent.

For $n = 200$, $p = 20$, the minimum coverage probabilities are below 90 percent with 25 percent or 50 percent missing data imposed. Under the 25 percent missing data condition, minimum coverage probabilities range from 76.5 to 87 percent. OLS is the best method under each missing data mechanism, outperforming stepwise regression by 3.1 and 2.5 percent, respectively, in the linear and convex mechanism, and outperforming ridge regression by 3 percent in the sinister mechanism. Under 50 percent missing data imposed, the coverage probabilities decrease from the 25 percent missing data case, ranging from 60.9 to 83.8 percent. OLS remains the dominant method, outperforming stepwise regression by 5.1, 7.6 and 7.6 percent in the linear, convex and sinister mechanisms, respectively.

5.2.5 Overall Results

Under the independent correlation structure for both beta 1 and beta 2, OLS has dominant performance in terms of the minimum coverage probability of a nominal 95 percent confidence interval. For beta 1, OLS dominates for all sample size and number of parameter combinations, whereas for beta 2 OLS dominates for $n = 100$ and $n = 20$ with $p = 20$ but OLS is tied or outperformed in a few cases with $n = 50$ and $n = 100$ with $p = 10$. Under the equal correlated structure with $\rho = 0.50$, OLS remains the dominant method for beta 2, being outperformed by the lasso in a few cases with $n = 50$ and by ridge regression in one case with $n = 100$ and $p = 10$. For beta 1, under the equicorrelated $\rho = 0.50$ structure, the dominant method is situation dependent. In many cases, OLS is tied with other methods.

Based on these results, OLS is the dominant method under the independent correlation structure regardless of the true beta vector and is the dominant method for correlated data with a true beta vector consisting of small effects. Again, the missing data percentage did not have a large impact on performance, while the type of missing data did have an impact. Missing data type was more important with correlated data under beta 1.

6.0 ANALYSIS OF MOTIVATING DATA

6.1 MOTIVATION

The data motivating this research was collected as part of a study entitled "Neurobehavioral Changes in Pediatric Affective Disorder." The main goals of this study are to understand the causes of pediatric affective disorders and their interaction with the developmental changes of childhood and adolescence; and to determine possible improvements in the treatment of such disorders. As part of these main goals, focus was placed on identifying the psychobiological, psychosocial and other correlates of MDD in children and adolescents. The data collected related to this subgoal consist of biological and EEG sleep data collected at the time of intake into the study; and symptomatology, life events and psychosocial measurements detailing the subject's relationship with his or her family and friends, collected at intake and at follow-up visits.

Previous analysis of this data set had focused on differences between subgroups of subjects on various psychobiological and sleep measurements; and on the time-to-event (MDD) outcome. In particular, survival analysis using Cox proportional hazards regression was conducted using the one-at-a-time approach, with each model containing one covariate of interest and a set of demographic variables, including age, gender, body mass index (BMI), Tanner stage of pubertal development and the socioeconomic status (SES) of the subject's family. This one-at-a-time approach was employed in an attempt to address the research questions without addressing the missing data values that exist in the data set.

The original motivation of this dissertation research was to address the missing data problem, subsequently employing variable selection methods, such as stepwise selection, to select a relevant set of predictors focusing on the time-to-MDD outcome. Over the course of time, the direction of the research changed to focus on linear regression models rather than survival analysis techniques. Additionally, the focus of new research at WPIC has shifted to anxiety disorders in children and adolescents. In particular researchers were interested in the relationships between the psychobiological predictors of depression and anxiety disorders. A data set suitable for linear regression and addressing this new focus on anxiety disorders was selected for analysis as part of this dissertation. The outcome of interest will be an index of the severity of anxiety symptoms as determined by the Screen for Child Anxiety Related Emotional Disorders (SCARED). More details about the psychobiological predictors can be found in Appendix C.

6.1.1 Background

In the overall study, psychobiological data were collected on about 200 subjects between the ages of 6 and 13 years and at Tanner stage I or II at the time of intake into the study. Tanner stages are a measure of sexual maturation with Stage I corresponding to pre-pubertal and Stage V being adult or fully mature. Body mass index (BMI), a measure of body fat based on height and weight, was computed. The Hollingshead four-factor index was used to determine the family's socioeconomic status (SES). Because the SCARED was not developed until the late 1990's, it was added more recently to the study protocol and, therefore, only a subsample of subjects were administered the SCARED assessment.

Three diagnostic groups are considered within this study: the MDD group consisting of children with a current episode of MDD, the children at high-risk for MDD and a control group at low risk for MDD. The depressed children were within episode at the time psychobiological and sleep measurements were collected; either in their first MDD episode or have one or more prior episodes. To be classified as at high-risk to develop depression, subjects were required to have never been depressed but to have at least one first-degree (parent or sibling) and one second-degree (grandparent, aunt or uncle) relative with a history of childhood-onset, recurrent, bipolar, or psychotic depression. Children classified as at low-risk of developing depression (or normal controls) are those who had not developed a psychiatric disorder at intake and had no first-degree relatives and less than 20% of their second degree relatives with a lifetime history of an affective disorder. The Schedule for Affective Disorders and Schizophrenia for School-age Children Present Episode Version (K-SADS-P) was used to assess the depressed children and the K-SADS-E epidemiological version was used to assess the high- and low-risk subjects and their families.

6.2 DATA ANALYSIS

The data set will be analyzed employing multiple imputation via chained equations MICE to account for the missing data values. Each of the multiple imputed data sets was analyzed using ordinary least squares, stepwise regression and the lasso. The imputed data sets were constructed using the MICE implementation in Stata coded by Patrick Royston [38]. Stata was selected to perform the multiple imputations because it is easily accessible to the researchers, provides an graphic user interface that allows for easy manipulation of the imputation model, and allows more easily for the inclusion of categorical and ordinal covariates in their correct form. The subsequent regression analyses and the combination of the parameter estimates was performed in R [32].

The outcome of interest in this analysis is the severity of anxiety symptoms as assessed by the Screen for Anxiety Related Emotional Disorders (SCARED). The SCARED is a self-report questionnaire consisting of separate child and parent report forms designed to screen subjects for the presence of anxiety disorders; including general anxiety disorder, separation anxiety disorder, panic disorder, social phobia and school phobia [6]. The total score of a subject on this assessment can be used as an overall measure of the severity of their anxiety disorder, with a score of 25 used on the child assessment as the threshold for anxiety disorder [7]. The measure has been repeatedly studied and has been shown to possess good psychometric properties and to exhibit sensitivity to treatment effects [10]. Both the parent and child SCARED have been shown to discriminate between subjects with anxiety and those without, and between subjects with anxiety and those with disruptive disorders. In addition, the child SCARED scores have been shown to discriminate between anxiety and depression[7]. The child SCARED scores will be used as the outcome in the analysis.

6.2.1 Missing Data

Missing values arise for two general reasons. The first is due to the typical data collection problems, such as difficulties with blood samples or assays, subject discomfort, failure to appear for a follow-up interview, etc. The second reason for missing data is that modifications have been made to the protocol over time that led to the discontinuation of some measurements. The missing data values in the data made it impossible to employ variable selection techniques to identify a relevant set of predictor variables related to MDD and directly address the questions of interest.

6.2.1.1 Amount of Missing Data The amount of missing data varies across variables, with approximately 25 percent missing values overall. Some individual measurements, including growth hormone, cortisol and prolactin measurements, have considerably higher percentages of missing data. Figure 6.2.1.1 gives the number of complete observations and missing values for each variable included in the final data analysis. Figure 6.2.1.1 presents the missing data patterns for the variables, in the order they appear in the first figure.

6.2.1.2 Data Characteristics A subset of the variables collected are used in the subsequent data analysis. Details about these variables can be found in Appendix E. The other variables were used as auxiliary variables, in the context of Collins, Schafer and Kam, to improve the performance of the multiple imputation [11]. The variables included in the imputation model as auxiliary covariates are: Tanner PH, dhea, DHEAS, androstendione, estradiol, testosterone and the mean level, peak level during sleep and levels 2 hours before and after sleep onset of cortisol (cortmsl, cortpksl, cortpre2, cortpo2), and growth hormone (ghmsleep, ghpksl, ghpre2, ghpo2) and cortisol (crfprecrt, crfpocrt, crfpecrt) and adrenocorticotrophic hormone (ACT) (crfpreact, crfpoact, crfpeact) response to corticotropin releasing factor (CRF).

There is a high degree of correlation between the variables included in the model and there is approximately 25 percent missing data. The OLS parameter estimates contain values that are small to moderate in size, suggesting that the true beta vector may be similar to

Figure 1: Motivating Data Missing Data Pattern

```
+-------------------------------------+
|               _pattern   _mv   _freq |
|-------------------------------------|
| +++++ +++++ ++++ +++     0      25 |
| ..... .++++ ++++ +++     6       9 |
| ..... ....+ ++++ +++     9       6 |
| +++++ +...+ ++++ +++     3       4 |
| +++++ ++++. .+++ +++     2       3 |
|-------------------------------------|
| +++++ +...+ ++.. +++     5       3 |
| +++++ +++++ +.++ +++     1       1 |
| +++++ ++++. ++++ +++     1       1 |
| +++++ +...+ ++++ +++     3       1 |
| +++++ +++++ +... +++     3       1 |
|-------------------------------------|
| +++++ ++++. ..++ +++     3       1 |
| +++++ +.... .+++ +++     5       1 |
| +++++ ++++. .+++ ...     5       1 |
| +++++ ++++. ..++ ...     6       1 |
| ..... .+++. ..++ +++     9       1 |
|-------------------------------------|
| ..... ....+ ++++ +++     9       1 |
| ..... ....+ ++.. +++    11       1 |
| ..... ..... ..++ +++    12       1 |
+-------------------------------------+
```

Table 12: Motivating Data Missing Data Percents

Variable	Complete Data	Missing Values
l5peprl	44	19
lrpoprl	44	19
l5preprl	44	19
l5pecrt	44	19
l5pocrt	44	19
l5precrt	44	19
clonpegh	45	18
clonpogh	45	18
clonpregh	45	18
ses	52	11
grfpregh	58	5
grfpogh	58	5
grfpegh	58	5
crfprecrt	61	2
crfpocrt	61	2
crfpecrt	61	2

the beta 2 case considered in the simulation study. The $n = 50$, $p = 10$ and $n = 100$, $p = 20$ cases match most closely with the motivating data set. Based on the simulation study results, OLS would be expected to have good performance in terms of coverage probability for confidence intervals, but may suffer from increase variability of parameter estimates, which in part accounts for this improvement. In terms of the prediction accuracy, the lasso would be expected to have good performance based on the simulation study.

The variables included for possible selection in the data analysis are: body mass index (BMI); age; gender; Tanner stage of development; socioeconomic status (SES); pre infusion, post infusion and peak after infusion levels of growth hormone released in response to growth hormone releasing hormone (grfpregh, grfpogh, grfpegh)and clonidine hydrochloride (clonpregh, clonpogh, clonpegh); prolactin (L5HTP) (l5preprl, l5poprl, l5peprl), and cortisol (l5precrt, l5pocrt, l5pecrt) response to L–5–Hydroxytryptophan and diagnostic group predicting the combined anxiety score as measure by the SCARED diagnostic tool.

6.2.2 Motivating Data Results

Using multiple imputation via chained equations implemented in the Stata software package, a set of 10 multiply imputed data sets were constructed. The combined parameter estimates using ordinary least squares, stepwise regression, ridge regression and the lasso were obtained for the set of imputed data sets and the results are given in tables 6.2.2, 6.2.2, 6.2.2 and 6.2.2.

The parameter estimates using ordinary least squares have extremely large standard errors, resulting in large confidence intervals. The mean square error computed as $\frac{1}{n}\sum_{i=1}^{n}(y)_i - \hat{y_i})^2$, for OLS is 329.8. In the ridge regression case, the parameter shrinkage imposed by the ridge constraint yields parameter estimates with much smaller standard errors. The mean square error for the ridge regression parameter estimates is 81.3 a significant reduction over ordinary least squares. For stepwise regression, the standard errors for the parameter estimates are extremely large as in ordinary least squares, however the mean square error is dramatically larger than the ordinary least squares case at 12575.83. The lasso has a similar decrease in standard error as in ridge regression. The mean square error, at 1283.2 is larger than in OLS and much larger than in ridge regression.

In both selection methods, stepwise regression and the lasso, the models selected are larger, with only 2 variables not included in the model. Because of the small degree of variable selection and the superior performance of ridge regression in terms of standard errors of parameter estimates and mean square error, ridge regression is the best method in the case.

Table 13: Ordinary Least Squares for motivating data

	Intercept	bmi	white	age	gender	tannerb	ses	grfpregh
Estimate	0	0.53	-0.72	-0.62	-3.23	1.08	-0.73	-1.02
Standard Error	1.22	5.62	1.89	4.24	2.7	6.19	4.55	12.98
LCL	-2.39	-10.48	-4.44	-8.94	-8.52	-11.06	-9.64	-26.45
UCL	2.39	11.54	2.99	7.69	2.06	13.22	8.19	24.41
	grfpogh	grfpegh	clonpregh	clonpogh	clonpegh	l5preprl	lrpoprl	l5peprl
Estimate	-2.62	3.51	-1.06	3.49	-2.84	-10.26	0.35	-17.48
Standard Error	23.97	24.85	19.2	114.25	121.82	689.25	3065.71	2846.57
LCL	-49.6	-45.19	-38.68	-220.43	-241.6	-1361.2	-6008.44	-5596.77
UCL	44.36	52.22	36.57	227.41	235.93	1340.67	6009.13	5561.8
	l5precrt	l5pocrt	l5pecrt	crfprecrt	crfpocrt	crfpecrt	dx	dx
Estimate	-8.47	2.83	-13.2	-10.89	41.28	-31.51	-0.96	2.38
Standard Error	1331.92	3003.12	13251.67	982.99	14585.4	7919.54	2.87	2.38
LCL	-2619.04	-5883.28	-25986.47	-1937.55	-28546.11	-15553.8	-6.58	-2.29
UCL	2602.1	5888.94	25960.07	1915.78	28628.67	15490.78	4.66	7.05

Table 14: Ridge regression results for motivating data

	intercept	bmi	race	age	gender	tannerb	ses	grfpregh
Estimate	0	0.18	-0.34	-0.22	-1.77	0.29	-0.27	-0.1
Standard Error	0.74	0.67	0.72	0.62	0.72	0.63	0.89	1.23
LCL	-1.45	-1.13	-1.76	-1.42	-3.17	-0.94	-2.02	-2.51
UCL	1.45	1.48	1.08	0.99	-0.37	1.51	1.48	2.31
	grfpogh	grfpegh	clonpregh	clonpogh	clonpegh	l5preprl	lrpoprl	l5peprl
Estimate	0.28	0.41	-0.25	0.12	-0.04	0.13	-0.03	-0.12
Standard Error	0.63	0.65	1.04	0.7	0.73	0.46	0.55	0.46
LCL	-0.95	-0.86	-2.3	-1.25	-1.46	-0.77	-1.11	-1.02
UCL	1.52	1.68	1.79	1.48	1.38	1.02	1.05	0.78
	l5precrt	l5pocrt	l5pecrt	crfprecrt	crfpocrt	crfpecrt	dx	dx
Estimate	-0.24	-0.03	-0.05	-0.22	0.23	0.2	-0.73	1.05
Standard Error	0.58	0.67	0.4	0.73	0.53	0.5	0.72	0.71
LCL	-1.37	-1.34	-0.82	-1.64	-0.8	-0.79	-2.14	-0.35
UCL	0.89	1.29	0.73	1.2	1.27	1.19	0.68	2.44

Table 15: Stepwise regression results for motivating data

	intercept	bmi	race	age	gender	tanner	ses	grfpregh
Estimate	0	2.24	0	-2.58	-3.31	2.79	-3.25	-2.74
Standard Error	1.1	1.38	0	1.5	1.99	1.74	2.61	1.86
LCL	-2.15	-0.47	0	-5.53	-7.2	-0.62	-8.37	-6.39
UCL	2.15	4.95	0	0.36	0.59	6.21	1.86	0.92
	grfpogh	grfpegh	clonpregh	clonpogh	clonpegh	l5preprl	lrpoprl	l5peprl
Estimate	-14.71	6.48	-0.05	14.06	-12.11	-20.1	9.01	-35.05
Standard Error	5.77	32.45	9.98	94.74	135.78	1013.6	6739.83	8551.51
LCL	-26.01	-57.13	-19.61	-171.63	-278.23	-2006.75	-13201.05	-16796
UCL	-3.41	70.09	19.51	199.74	254.01	1966.55	13219.07	16725.91
	l5precrt	l5pocrt	l5pecrt	crfprecrt	crfpocrt	crfpecrt	dx	dx
Estimate	-33.55	-6.57	-9.54	-39.44	157.41	-235.61	0	2.69
Standard Error	1717.64	1060.27	7161.08	1545.55	26831.6	180.42	0	1.44
LCL	-3400.12	-2084.69	-14045.26	-3068.72	-52432.52	-589.24	0	-0.14
UCL	3333.02	2071.55	14026.17	2989.84	52747.33	118.02	0	5.52

Table 16: Lasso complete data results

	Intercept	bmi	white	age	gender	tannerb	ses	grfpregh
Estimate	na	0.08	-3.3	-0.19	-6.82	0.51	-0.02	-0.25
Standard Error	na	0.53	7.21	1.11	4.45	2.26	0.11	2.65
LCL	na	-0.96	-17.44	-2.36	-15.55	-3.92	-0.23	-5.44
UCL	na	1.13	10.84	1.98	1.9	4.94	0.19	4.95
	grfpogh	grfpegh	clonpregh	clonpogh	clonpegh	l5preprl	lrpoprl	l5peprl
Estimate	-0.02	0.07	-1.22	0.04	0	-0.01	-0.03	0
Standard Error	0.47	0.32	9.41	0.83	0.49	0.2	0.58	0.37
LCL	-0.95	-0.57	-19.67	-1.58	-0.96	-0.4	-1.16	-0.72
UCL	0.9	0.7	17.22	1.66	0.97	0.38	1.11	0.73
	l5precrt	l5pocrt	l5pecrt	crfprecrt	crfpocrt	crfpecrt	dx	dx
Estimate	-0.02	0.01	-0.01	-0.21	0.31	-0.12	-1.72	6.1
Standard Error	0.18	0.43	0.31	0.82	1.46	1.16	6.01	6.62
LCL	-0.38	-0.84	-0.62	-1.81	-2.56	-2.4	-13.5	-6.87
UCL	0.34	0.85	0.59	1.39	3.18	2.15	10.05	19.07

6.3 CONCLUSIONS

The analysis of the motivating data set was not as fruitful as may have been expected. The degree of variable selection was not great and the parameter estimates had extremely large standard errors making interpretation difficult. In attempting to control the overall percentage of missing data in the motivating data set to be analyzed, the degree of correlation between predictor variables was quite high. A more careful selection of covariates for the initial pool on which to perform variable selection, with input from the researchers may yield more satisfactory and interpretable results.

The performance of ridge regression, resulting in significantly smaller standard errors for the parameter estimates, illustrates its importance as a tool for accounting for multicollinearity within the predictor variables. It also shows that the lasso may not provide a sufficient degree of shrinkage especially in the case of correlated predictor variables. The elastic net may provide a useful improvement over the lasso in this case.

7.0 FUTURE RESEARCH

As with all academic research, many new questions have been raised as a consequence of this research. An obvious starting point would be an analysis of the impact of different choices for the parameters considered here. Time and space constraints restricted the scope of the project. Different choices for the beta vectors, the variance of the error term, and the correlation structure exhibited by the predictor variables, among others, could impact the results. Under the MAR assumption, variations of the imputation model could impact the results. In many cases, the MAR assumption is not valid, the data under consideration are truly NMAR. The use of specialized methods for handling NMAR data could be considered in a subsequent simulation study. Perhaps more informative would be an analysis of the impact of the incorrect, but often more tractable use of MAR methods are truly NMAR data.

A careful analysis of the accuracy of the model selected by the various methods would add to our understanding of the impact of missing data on variable selection methods. The selection of the correct model, or a model containing the correct model would add to our understanding of how well each of the selection methods is able to determine which variables are important predictors of the response variable.

Subsequent to the development of the lasso, the elastic net has been proposed. Inclusion of this method in the simulation study would yield additional insight. Extension of variable selection methods to include models such as the logistic regression, survival analysis, etc. would also be an interesting question.

BIBLIOGRAPHY

[1] Allison, P.D. (2001), *Missing Data,* Thousand Oaks, CA: Sage.

[2] Birmaher, B., R. Dahl, J. Perel, et al. (1996), "Corticotropin-Releasing Hormone Challenge in Prepubertal Major Depression," *Biological Psychiatry,* 39,267-277.

[3] Birmaher, B., R. Dahl, D. Williamson, et al. (2000), "Growth Hormone Secretion on Children and Adolescents at High Risk for Major Depressive Disorder," *Archives of General Psychiatry,* 57,867-872.

[4] Birmaher, B., J. Kaufman, D. Brent, et al. (1997), "Neuroendocrine Response to 5-Hydroxy-L-Tryptophan in Prepubertal Children at High Risk of MDD," *Archives of General Psychiatry,* 54,1113-1119.

[5] Birmaher, B., N. Ryan, D. Williamson, et al. (1996), "Childhood and Adolescent Depression: A Review of the Past 10 Years. Part I," *Journal of the American Academy of Child & Adolescent Psychiatry,* 35,1427-1439.

[6] Birmaher, B., S. Khetarpal, D. Brent, et al. (1997), "The Screen for Child Anxiety Related Emotional Disorders (SCARED): Scale Construction and Psychometric Characteristics," *Journal of the American Academy of Child % Adolescent Psychiatry,* 36(4):545-553.

[7] Birmaher, B., D. Brent, L. Chiappetta, et al. (1999), "Psychometric Properties of the Screen for Child Anxiety Related Emotional Disorders (SCARED): A Replication Study," *Journal of the American Academy of Child and Adolescent Psychiatry,* 38(10):1230-1236.

[8] Breiman, L. (1995), "Better Subset Selection Using the Nonnegative Garrote," *Technometrics*, 373-384.

[9] Cambanis, S., S. Huang, and G. Simons (1981), "On the Theory of Elliptically Contoured Distributions, *Journal of Multivariate Analysis,* 11, 368-385.

[10] Clark, D., B. Birmaher, D. Axelson, et al. (2005), "Fluoxetine for the Treatment of Childhood Anxiety Disorders: Open-Label, Long-Term Extension to a Controlled Trial,"

Journal of the American Academy of Child and Adolescent Psychiatry, 44(12),1263-1270.

[11] Collins, L., J. Schafer, C. Kam (2000), "A Comparison of Inclusive and Restrictive Strategies in Modern Missing Data Procedures," *Psychological Methods,* 6(4),330-351.

[12] Dahl, R., B. Birmaher, D. Williamson, et al. (2000), "Low Growth Hormone Response to Growth Hormone Releasing Hormone in Child Depression," *Biological Psychiatry,* 48,981-988.

[13] Dahl, R., N. Ryan, B. Birmaher, et al. (1991), "Electroencephalographic Sleep Measures in Prepubertal Depression," *Psychiatry Research,* 38,201-214.

[14] Dempster, A.P., N.M. Laird, D.B. Rubin (1977. "Maximum Likelihood from Incomplete Data via the EM Algorithm." *Journal of the Royal Statistical Society, Series B,* 39(1),1-38.

[15] Dempster, A.P., M. Schatzoff, et al. (1977). "A Simulation Study of Alternatives to Ordinary Least Squares (with discussion)." *Journal of the American Statistical Association* 72(357): 77-106.

[16] Dorn, L., R. Dahl, B. Birmaher, et al. (1997), "Baseline Thyroid Hormones in Depressed and Non-depressed Pre- and Early Pubertal Boys and Girls," *Journal of Psychiatric Research,* 31,555-567.

[17] Efron, B., T. Hastie, I. Johnstone, et al. (2004) "Least Angle Regression," *Annals of Statistics* 32, 407-499.

[18] George, E. (2000), "The Variable Selection Problem," *Journal of the American Statistical Association,* 95,1304-1308.

[19] Frank, I. and J. Friedman (1993). "A Statistical View of Some Chemometrics Regression Tools." *Technometrics* 35(2): 109-135.

[20] Fu, W. (1998), "Penalized Regressions: The Bridge versus the Lasso," *Journal of Computational and Graphical Statistics,* 7(3).

[21] Hoerl, A. and R. Kennard (1970), "Ridge Regression: Biased Estimation for Nonorthogonal Problems," *Technometrics,* 12, 55-67.

[22] Kadane, J. and N. Lazar (2004) "Methods and Criteria for Model Selection." *Journal of the American Statistical Association* 99(465) 279-290.

[23] James, W., C. Stein (1961), "Estimation with Quadratic Loss," *Proceedings of the Fourth Berkeley Symposium,* 1, 361-379.

[24] Lawson, C. and R. Hansen (1974), *Solving Least Squares Problems,* New York: Chapman and Hall.

[25] Leng, C., Y. Lin, G. Wahba "A Note on the LASSO and Related Procedures in Model Selection," *University of Wisconsin-Madison Statistics Department Technical Report 1091,* April 2004.(to appear *Statistica Sinica* 2005)

[26] Little, R. and D. Rubin (2002), *Statistical Analysis with Missing Data,* 2nd edition, Hoboken, NJ: Wiley.

[27] Lu, G. and J. Copas (2004), "Missing at Random, Likelihood Ignorability and Model Completeness." *Annals of Statistics* 32(2), 754-765.

[28] Meng, X. (2000), "Missing Data: Dial M for ???," *Journal of the American Statistical Association,* 95 (452), 1325-1330.

[29] Miller, A. (2002), *Subset Selection in Regression,* London: Chapman and Hall.

[30] Osborne, M., Presnell, B., Turlach, B., (2000) "On the LASSO and its Dual," *Journal of Computational and Graphical Statistics,* 9(2):319-337.

[31] Oudshoorn, K., S. van Buuren, and J. van Rijckevorsel (1999), "Flexible Multiple Imputation by Chained Equations of the AVO-95 Survey," *Report PG/VGZ/99.045* Leiden.

[32] R Development Core Team (2005), *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL: http://www.R-project.org.

[33] Ryan, N., Birmaher, B., Perel, J., et al. (2000), "Neuroendocrine Response to L–5–Hydroxytryptophan Challenge in Prepubertal Major Depression," *Archives of General Psychiatry,* 49,843-851.

[34] Ryan, N., R. Dahl, B. Birmaher, et al. (1994), "Stimulatory Tests of Growth Hormone Secretion in Prepubertal Major Depression. *Journal of the American Academy of Child & Adolescent Psychiatry,* 33,824-833.

[35] Sauerbrei, W. (1999), "The Use of Resampling Methods to Simplify Regression Models in Medical Statistics," *Applied Statistics,* 48, 313-329.

[36] Schafer, J. (1997), *Analysis of Incomplete Multivariate Data,* New York:CRC Press.

[37] Schafer, J.L. and J.W. Graham, (2002), "Missing Data: Our View of State of the Art," *Psychological Methods,* 7(2), 147-177.

[38] Royston, P., (2004), "Multiple imputation of missing values," *Stata Journal* 4(3), 227241.

[39] Tanner, M.A. (1991), *Tools for Statistical Inference,* New York: Springer-Verlag.

[40] Tanner, M.A., W.H. Wong, (1987), "The Calculation of Posterior Distributions by Data Augmentation (with discussion)," *Journal of the American Statistical Association,* 82, 528-550.

[41] Tibshirani, R. (1996), "Regression Shrinkage and Selection via the Lasso," *Journal of the Royal Statistical Society, Series B,* 58, 267-288.

[42] Tibshirani, R. (1997), "The Lasso Method For Variable Selection in the Cox Model," *Statistics in Medicine,* 16, 385-395.

[43] Vach, K., W. Sauerbrei, and M. Schumacher (2001), "Variable Selection and Shrinkage: Comparison of Some Approaches" *Statistica Neerlandica* 55, 53-75.

[44] van Buuren, S., H.C. Boshuizen, and D.L. Knook (1999), "Mutliple Imputation of Missing Blood Pressure Covariates in Survival Analysis," *Statistics in Medicine,* 18: 681-694.

[45] van Buuren, S., J.P.L. Brand, C.G.M. Groothuis-Oudshoorn, and D.B. Rubin (2005), "Fully Conditional Specification in Multivariate Imputation," *Need reference here.*

[46] van Buuren, S., K. Oudshoorn (1999), "Flexible Multivariate Imputation by MICE," *Report PG/VGZ/99.054* Leiden.

[47] van Houwelingen, J. (2001), "Shrinkage and Penalized Likelihood as Methods to Improve Prediction Accuracy," *Statistica Neerlandica,* 55(1).

[48] Zou, H. and T. Hastie (2005). "Regularization and Variable Selection via the Elastic Net." *Journal of the Royal Statistical Society, Series B,* 67(2),301-320.

[49] Zou, H., T. Hastie, and R. Tibshirani, (2004), "On the "Degrees of Freedom" of the Lasso,"*submitted.*

APPENDIX A

MAR PREDICTION ERROR TABLES- BETA 1

Table 17: MAR, Beta 1, independent, n=50, p=5, MSE of Prediction

		Linear Missing		Convex Missing		Sinister Missing	
	Complete	25%	50%	25%	50%	25%	50%
OLS	54.6914	68.2929	126.1785	68.3981	132.0979	68.6678	136.1288
Stepwise	53.0424	65.6315	122.9691	68.7254	129.7173	67.0359	133.8803
Ridge	54.8543	67.6266	121.8536	67.5476	130.0766	67.3922	133.5684
LASSO	55.1775	64.1683	119.6536	66.2485	127.2515	65.3432	127.3544

Table 18: MAR, Beta 1, independent, n=50, p=10, MSE of Prediction

		Linear Missing	Convex Missing	Sinister Missing
	Complete	25%	25%	25%
OLS	60.5565	185.7087	183.3191	192.9225
Stepwise	59.0670	180.2669	177.2087	185.2573
Ridge	60.1811	184.4257	186.0256	192.1265
LASSO	61.0846	188.4579	190.0852	202.1298

Table 19: MAR, Beta 1, independent, n=100, p=10, MSE of Prediction

		Linear Missing		Convex Missing		Sinister Missing	
	Complete	25%	50%	25%	50%	25%	50%
OLS	107.2888	186.5416	490.7070	183.7854	501.4683	185.4017	509.5673
Stepwise	102.1218	181.1232	490.1174	179.2369	506.4319	179.0049	502.0201
Ridge	114.5674	188.0288	498.7418	180.2499	514.8287	184.9006	516.3445
LASSO	102.0426	177.9425	513.7385	171.2251	535.8941	180.8131	538.3556

Table 20: MAR, Beta 1, independent, n=100, p=20, MSE of Prediction

		Linear Missing		Convex Missing		Sinister Missing	
	Complete	25%	50%	25%	50%	25%	50%
OLS	518.2671	901.4794	2534.1985	912.6544	2624.2234	906.5727	2688.9700
Stepwise	517.9139	890.2509	2527.2605	901.9295	2598.9921	894.0207	2658.8069
Ridge	517.7248	907.5595	2621.5305	912.1938	2691.5473	911.5118	2784.5724
LASSO	518.8361	938.6454	2892.1337	940.3793	2996.5917	949.6254	3075.2901

Table 21: MAR, Beta 1, independent, n=200, p=20, MSE of Prediction

		Linear Missing		Convex Missing		Sinister Missing	
	Complete	25%	50%	25%	50%	25%	50%
OLS	221.9523	692.7812	2012.8564	667.4870	2052.7942	682.0423	2147.7561
Stepwise	218.3911	673.8847	2013.2491	649.5072	2085.8392	656.1844	2152.2926
Ridge	221.7789	703.3806	2084.2540	673.0977	2149.6785	688.3415	2230.8038
LASSO	220.8626	713.3534	2276.7400	700.3502	2309.2951	714.3915	2428.8632

Table 22: MAR, Beta 1, autoregressive 0.25, n=50, p=5, MSE of Prediction

		Linear Missing		Convex Missing		Sinister Missing	
	Complete	25%	50%	25%	50%	25%	50%
OLS	55.5467	65.7170	125.7308	66.1434	123.1692	68.6117	132.3267
Stepwise	53.7703	64.9161	124.8871	63.6737	122.2519	68.0792	131.6722
Ridge	56.1997	65.8453	121.9880	65.3392	120.2885	68.3793	126.6165
LASSO	53.5634	63.4803	111.1292	62.2114	113.6424	66.4535	121.6055

Table 23: MAR, Beta 1, autoregressive 0.25, n=50, p=10, MSE of Prediction

		Linear Missing	Convex Missing	Sinister Missing
	Complete	25%	25%	25%
OLS	61.3045	178.9880	179.9235	184.0779
Stepwise	58.8851	172.8878	176.1373	181.2408
Ridge	59.9302	170.4082	165.6729	178.6823
LASSO	60.9794	171.3020	174.6882	178.9026

Table 24: MAR, Beta 1, autoregressive 0.25, n=100, p=10, MSE of Prediction

		Linear Missing		Convex Missing		Sinister Missing	
	Complete	25%	50%	25%	50%	25%	50%
OLS	110.0577	177.0543	454.2815	179.6269	465.0719	185.3140	485.3400
Stepwise	108.6512	174.8191	455.2907	175.7548	462.5354	175.3018	486.5769
Ridge	109.8092	176.7754	452.1787	176.7148	457.7949	181.4086	476.8104
LASSO	110.2223	170.8154	443.5816	171.5941	465.4816	172.7067	486.3923

Table 25: MAR, Beta 1, autoregressive 0.25, n=100, p=20, MSE of Prediction

		Linear Missing	Convex Missing	Sinister Missing
	Complete	25%	25%	25%
OLS	120.1216	721.7036	654.2493	686.4536
Stepwise	113.4667	692.8222	651.1590	674.9235
Ridge	125.7404	681.8509	647.0110	696.2115
LASSO	123.1268	682.9706	684.0428	714.1092

Table 26: MAR, Beta 1, autoregressive 0.25, n=200, p=20, MSE of Prediction

		Linear Missing		Convex Missing		Sinister Missing	
	Complete	25%	50%	25%	50%	25%	50%
OLS	219.5702	609.8340	1975.7054	628.3953	1782.4162	636.9647	1967.4435
Stepwise	219.1817	610.9640	2052.5850	599.1572	1758.4160	621.8766	1960.7872
Ridge	227.7375	619.1731	1993.3627	615.4282	1797.2597	636.7636	1978.0901
LASSO	222.1771	610.3345	2041.6512	620.1907	1872.1956	636.6527	2072.8716

Table 27: MAR, Beta 1, autoregressive 0.50, n=50, p=5, MSE of Prediction

		Linear Missing		Convex Missing		Sinister Missing	
	Complete	25%	50%	25%	50%	25%	50%
OLS	54.8208	67.2951	115.0031	66.0225	112.9062	66.9540	118.0505
Stepwise	53.7677	66.2884	112.6266	65.4333	118.0544	66.9004	116.2644
Ridge	53.9582	67.1818	112.1185	66.9024	113.3605	67.3742	115.0649
LASSO	53.8426	65.5285	102.3797	63.8824	99.3837	64.7271	106.0956

Table 28: MAR, Beta 1, autoregressive 0.50, n=50, p=10, MSE of Prediction

		Linear Missing	Convex Missing	Sinister Missing
	Complete	25%	25%	25%
OLS	62.0978	150.0650	155.6568	159.4755
Stepwise	57.7598	151.2437	157.0554	158.4860
Ridge	59.5965	150.0448	150.1503	155.4602
LASSO	59.8987	147.8080	145.2792	147.0119

Table 29: MAR, Beta 1, autoregressive 0.50, n=100, p=10, MSE of Prediction

		Linear Missing		Convex Missing		Sinister Missing	
	Complete	25%	50%	25%	50%	25%	50%
OLS	110.4489	171.1741	407.8571	172.8261	407.4906	173.5972	428.7948
Stepwise	106.6021	165.7086	407.2357	167.8701	403.5122	169.4762	424.2551
Ridge	108.5540	166.6632	392.7161	172.7042	400.3027	171.3150	412.9140
LASSO	107.8430	161.4140	371.8090	162.2257	378.6345	164.0478	384.6215

Table 30: MAR, Beta 1, autoregressive 0.50, n=100, p=20, MSE of Prediction

		Linear Missing	Convex Missing	Sinister Missing
	Complete	25%	25%	25%
OLS	123.3967	544.7807	571.1112	578.7699
Stepwise	121.8138	515.9043	562.4940	571.9295
Ridge	122.3753	517.8149	548.9973	567.6115
LASSO	119.6102	535.5514	555.2152	558.4912

Table 31: MAR, Beta 1, autoregressive 0.50, n=200, p=20, MSE of Prediction

		Linear Missing		Convex Missing		Sinister Missing	
	Complete	25%	50%	25%	50%	25%	50%
OLS	222.0819	582.7092	1588.6875	551.6158	1596.3002	564.8337	1674.6146
Stepwise	216.4531	558.2296	1588.8780	539.6859	1602.5897	551.2713	1679.0479
Ridge	220.0712	575.1249	1552.6291	546.9128	1581.5668	559.5157	1628.4656
LASSO	222.2617	543.1309	1478.4914	518.4604	1535.1320	536.8010	1572.6396

Table 32: MAR, Beta 1, equicorrelated 0.25, n=50, p=5, MSE of Prediction

		Linear Missing		Convex Missing		Sinister Missing	
	Complete	25%	50%	25%	50%	25%	50%
OLS	55.1837	68.1253	124.6244	68.0459	120.3575	67.5179	126.0433
Stepwise	53.7651	67.1208	122.8441	65.8799	117.7967	66.0484	124.1935
Ridge	54.2000	67.9600	120.7277	67.9454	118.4358	67.6775	122.5943
LASSO	54.6877	64.8556	109.9059	64.6156	109.7634	63.8032	112.3873

Table 33: MAR, Beta 1, equicorrelated 0.25, n=50, p=10, MSE of Prediction

		Linear Missing	Convex Missing	Sinister Missing
	Complete	25%	25%	25%
OLS	61.1546	175.5798	177.4967	173.3270
Stepwise	59.0908	170.2885	170.6252	169.0561
Ridge	61.2798	167.9351	173.0097	166.6010
LASSO	59.0763	152.0506	160.1521	149.2375

Table 34: MAR, Beta 1, equicorrelated 0.25, n=100, p=10, MSE of Prediction

		Linear Missing		Convex Missing		Sinister Missing	
	Complete	25%	50%	25%	50%	25%	50%
OLS	110.2774	178.5675	450.5528	175.6482	442.0942	182.3321	468.1980
Stepwise	107.4045	174.4353	443.2301	170.4247	434.5593	176.9143	465.4317
Ridge	110.2621	177.3621	435.2475	173.5612	429.2071	179.9986	447.1488
LASSO	109.0274	166.7707	389.7753	165.1115	393.4913	165.6768	406.9605

Table 35: MAR, Beta 1, equicorrelated 0.25, n=100, p=20, MSE of Prediction

		Linear Missing	Convex Missing	Sinister Missing
	Complete	25%	25%	25%
OLS	123.3997	723.2553	676.5317	688.6427
Stepwise	119.9972	694.8354	644.0297	679.7634
Ridge	121.8646	676.3651	647.8424	675.2992
LASSO	120.2513	571.9898	563.7816	574.2475

Table 36: MAR, Beta 1, equicorrelated 0.25, n=200, p=20, MSE of Prediction

		Linear Missing		Convex Missing		Sinister Missing	
	Complete	25%	50%	25%	50%	25%	50%
OLS	220.5843	632.5637	1965.4283	632.0738	1998.8823	654.2681	2159.6610
Stepwise	216.9133	620.0131	1932.2014	624.0122	1987.2831	634.7871	2147.1978
Ridge	221.2414	632.4439	1924.1423	616.2748	1960.7702	651.5749	2110.4514
LASSO	218.1037	565.0684	1663.7628	555.2961	1671.4737	575.9436	1809.8219

Table 37: MAR, Beta 1, equicorrelated 0.50, n=50, p=5, MSE of Prediction

		Linear Missing		Convex Missing		Sinister Missing	
	Complete	25%	50%	25%	50%	25%	50%
OLS	56.5982	63.0679	130.3273	66.5053	114.2230	68.1162	118.4149
Stepwise	51.7356	64.0486	124.0407	62.2755	114.7393	66.5011	117.4116
Ridge	53.4086	65.1481	117.2939	59.7011	113.0369	67.1418	114.6097
LASSO	54.8507	60.3818	107.3554	63.9047	102.4125	64.1699	104.1159

Table 38: MAR, Beta 1, equicorrelated 0.50, n=50, p=10, MSE of Prediction

		Linear Missing	Convex Missing	Sinister Missing
	Complete	25%	25%	25%
OLS	60.0907	152.3349	146.5578	164.5213
Stepwise	60.2239	147.3591	144.1540	160.9219
Ridge	62.7391	146.5271	141.6429	159.8848
LASSO	58.3289	132.7502	129.0917	145.1945

Table 39: MAR, Beta 1, equicorrelated 0.50, n=100, p=10, MSE of Prediction

		Linear Missing		Convex Missing		Sinister Missing	
	Complete	25%	50%	25%	50%	25%	50%
OLS	109.1157	186.2294	415.4110	167.8291	404.0633	167.8750	429.3474
Stepwise	106.6308	178.8417	404.4096	166.7391	405.8736	166.4944	424.3720
Ridge	110.2445	178.5652	398.2767	166.2341	401.4705	167.1899	424.5093
LASSO	105.3876	176.1694	360.0292	155.5453	347.6478	156.8301	372.0573

Table 40: MAR, Beta 1, equicorrelated 0.50, n=100, p=20, MSE of Prediction

		Linear Missing	Convex Missing	Sinister Missing
	Complete	25%	25%	25%
OLS	123.6626	587.5774	580.1153	605.2985
Stepwise	119.4273	578.8255	566.1265	594.8556
Ridge	121.87	577.7622	566.2002	596.6994
LASSO	120.4136	482.2336	489.4524	512.4638

Table 41: MAR, Beta 1, equicorrelated 0.50, n=200, p=20, MSE of Prediction

		Linear Missing		Convex Missing		Sinister Missing	
	Complete	25%	50%	25%	50%	25%	50%
OLS	220.1670	534.0368	1861.8943	535.9510	1814.8956	551.2635	1996.2509
Stepwise	219.3279	523.0566	1840.6909	532.1103	1806.3048	545.6349	1984.5290
Ridge	219.8033	533.8635	1842.0740	539.1664	1792.0623	556.2916	1971.9494
LASSO	217.0954	483.8311	1620.0228	488.8379	1586.2408	503.8077	1740.7745

APPENDIX B

MAR PREDICTION ERROR TABLES - BETA 2

Table 42: MAR, Beta 2, independent, n=50, p=5, MSE of Prediction

		Linear Missing		Convex Missing		Sinister Missing	
	Complete	25%	50%	25%	50%	25%	50%
OLS	55.4664	61.9923	83.4059	61.7500	86.5321	63.2000	84.4271
Stepwise	54.2861	60.2779	83.4562	61.1181	85.1575	62.4822	83.5028
Ridge	55.8484	62.5328	80.7426	59.9136	84.8708	62.5470	80.7851
Lasso	54.1250	59.8385	77.9304	61.0997	81.9225	61.2615	78.1704

Table 43: MAR, Beta 2, independent, n=50, p=10, MSE of Prediction

		Linear Missing	Convex Missing	Sinister Missing
	Complete	25%	25%	25%
OLS	61.9424	108.4076	103.4098	108.6080
Stepwise	59.9775	104.7471	100.2806	105.9119
Ridge	61.3262	104.7898	107.4464	105.9647
Lasso	60.7566	109.6482	108.0634	106.8093

Table 44: MAR, Beta 2, independent, n=100, p=10, MSE of Prediction

		Linear Missing		Convex Missing		Sinister Missing	
	Complete Data	25%	50%	25%	50%	25%	50%
OLS	105.5557	140.3998	234.7738	138.9323	236.9196	142.6805	243.6577
Stepwise	108.8456	137.4937	238.0556	139.3674	236.0774	141.4535	243.2629
Ridge	111.9339	143.7699	232.9270	138.5979	232.4102	142.4350	239.3344
Lasso	111.2963	138.5174	230.6621	137.9250	230.4343	139.7001	240.1304

Table 45: MAR, Beta 2, independent, n=100, p=20, MSE of Prediction

		Linear Missing	Convex Missing	Sinister Missing
	Complete	25%	25%	25%
OLS	123.8002	317.7941	314.0482	325.3913
Stepwise	119.1469	313.8023	306.4068	322.2278
Ridge	122.1211	316.6489	306.9949	319.0307
Lasso	123.0028	331.0238	314.9211	333.4112

Table 46: MAR, Beta 2, independent, n=200, p=20, MSE of Prediction

		Linear Missing		Convex Missing		Sinister Missing	
	Complete	25%	50%	25%	50%	25%	50%
OLS	220.9925	366.5676	746.9667	362.3835	797.0031	370.5020	831.2598
Stepwise	218.1931	357.2839	766.6740	351.6718	807.4723	361.4974	833.3316
Ridge	222.0203	364.0751	747.2426	361.7068	800.7469	366.1707	835.8743
LASSO	221.5270	362.3833	799.7593	357.2051	850.1229	363.9596	863.0756

Table 47: MAR, Beta 2, autoregressive 0.25, n=50, p=5

		Linear Missing		Convex Missing		Sinister Missing	
	Complete	25%	50%	25%	50%	25%	50%
OLS	53.3920	61.3784	81.0871	62.2971	84.5883	62.9392	84.6691
Stepwise	50.8729	64.9137	79.0989	60.4346	82.6154	61.9896	83.5984
Ridge	54.3701	63.1675	78.7286	60.5885	80.7941	62.7609	81.4928
Lasso	52.8584	61.3054	75.4485	59.1969	78.6286	61.2443	79.0849

Table 48: MAR, Beta 2, autoregressive 0.25, n=50, p=10

		Linear Missing	Convex Missing	Sinister Missing
	Complete	25%	25%	25%
OLS	61.0721	106.4161	105.4736	106.7821
Stepwise	58.5595	103.8674	101.6583	104.8263
Ridge	60.6036	101.8088	100.8084	100.6141
Lasso	61.3736	102.1518	100.8155	103.4377

Table 49: MAR, Beta 2, autoregressive 0.25, n=100, p=10

		Linear Missing		Convex Missing		Sinister Missing	
	Complete	25%	50%	25%	50%	25%	50%
OLS	109.6989	140.4010	231.4666	139.4390	239.1563	142.4680	246.5022
Stepwise	108.0884	138.1040	234.2962	138.5506	232.7717	138.8389	247.1862
Ridge	109.6723	138.8020	230.1526	139.2449	233.3611	138.5603	243.0439
Lasso	109.9075	135.6867	227.1010	137.4459	232.1775	138.1130	244.3787

Table 50: MAR, Beta 2, autoregressive 0.25, n=100, p=20

		Linear Missing	Convex Missing	Sinister Missing
	Complete	25%	25%	25%
OLS	124.0102	320.0781	331.6553	326.4945
Stepwise	118.6777	314.9973	340.0865	324.3875
Ridge	122.1751	317.0071	300.4084	321.7268
Lasso	122.4813	332.4515	328.5702	333.4550

Table 51: MAR, Beta 2, autoregressive 0.25, n=200, p=20, MSE of Prediction

		Linear Missing		Convex Missing		Sinister Missing	
	Complete	25%	50%	25%	50%	25%	50%
OLS	220.3495	369.7895	750.5181	358.0273	794.7197	368.0512	852.8311
Stepwise	215.8794	363.7206	766.8885	347.5991	810.2299	360.2178	853.2176
Ridge	219.0234	370.6777	752.4451	345.1900	804.9105	363.0765	861.5634
Lasso	223.4010	360.2185	795.6305	344.5703	853.7385	362.0528	883.5625

Table 52: MAR, Beta 2, autoregressive 0.50, n=50, p=5, MSE of Prediction

		Linear Missing		Convex Missing		Sinister Missing	
	Complete	25%	50%	25%	50%	25%	50%
OLS	53.8760	61.5450	82.4335	62.1179	84.2433	62.2811	86.7815
Stepwise	51.9671	63.0262	80.7499	60.0994	83.7700	61.7423	85.5888
Ridge	54.2307	62.5154	79.6514	60.2089	81.2988	63.4752	81.8031
Lasso	53.4871	60.5544	76.6797	58.9193	78.5940	62.5772	79.3891

Table 53: MAR, Beta 2, autoregressive 0.50, n=50, p=10, MSE of Prediction

		Linear Missing	Convex Missing	Sinister Missing
	Complete	25%	25%	25%
OLS	61.7514	98.7063	96.0502	98.0908
Stepwise	60.1054	98.2804	94.7296	97.6379
Ridge	60.9980	95.4709	93.5015	94.8410
Lasso	59.8983	93.0246	92.4652	92.1966

Table 54: MAR, Beta 2, autoregressive 0.50, n=100, p=10, MSE of Prediction

		Linear Missing		Convex Missing		Sinister Missing	
	Complete	25%	50%	25%	50%	25%	50%
OLS	61.6130	100.4288	221.0276	96.2799	218.0811	98.8063	229.4185
Stepwise	59.7946	100.0436	219.2331	95.9292	217.6744	99.1662	228.6306
Ridge	61.3432	95.8751	221.1010	93.4457	213.1906	95.7127	222.1788
Lasso	59.3541	93.6130	206.6071	92.4387	204.0954	91.6757	211.9897

Table 55: MAR, Beta 2, autoregressive 0.50, n=100, p=20, MSE of Prediction

		Linear Missing	Convex Missing	Sinister Missing
	Complete	25%	25%	25%
OLS	123.8586	317.8149	316.3727	325.7987
Stepwise	119.2574	313.4903	309.0534	318.2755
Ridge	122.4451	316.3364	308.9948	318.2255
Lasso	123.1043	330.3037	317.3816	333.7529

Table 56: MAR, Beta 2, autoregressive 0.50, n=200, p=20, MSE of Prediction

		Linear Missing		Convex Missing		Sinister Missing	
	Complete	25%	50%	25%	50%	25%	50%
OLS	220.7011	362.6640	746.3367	362.8365	792.6193	367.4292	829.8984
Stepwise	215.6149	356.4940	766.0813	352.3461	808.5918	361.5859	823.6309
Ridge	221.4842	360.8605	747.9055	361.6337	801.3906	364.6881	831.2443
Lasso	223.3223	357.1737	799.4608	357.8421	853.3585	362.8708	856.9081

Table 57: MAR, Beta 2, equicorrelated 0.25, n=50, p=5, MSE of Prediction

		Linear Missing		Convex Missing		Sinister Missing	
	Complete	25%	50%	25%	50%	25%	50%
OLS	54.2568	61.5160	81.0871	62.2971	84.5883	62.9392	84.6691
Stepwise	52.1352	63.0013	79.0989	60.4346	82.6154	61.9896	83.5984
Ridge	54.2883	62.7029	78.7286	60.5885	80.7941	62.7609	81.4928
Lasso	53.5933	60.4873	75.4485	59.1969	78.6286	61.2443	79.0849

Table 58: MAR, Beta 2, equicorrelated 0.25, n=50, p=10, MSE of Prediction

		Linear Missing	Convex Missing	Sinister Missing
	Complete	25%	25%	25%
OLS	62.3342	100.2092	106.0568	104.5109
Stepwise	59.9578	99.9685	102.9970	103.8035
Ridge	61.8396	97.7046	101.3932	100.1808
Lasso	59.8726	88.5078	93.3420	93.0878

Table 59: MAR, Beta 2,equicorrelated 0.25, n=100, p=10, MSE of Prediction

		Linear Missing		Convex Missing		Sinister Missing	
	Complete	25%	50%	25%	50%	25%	50%
OLS	110.5835	139.9455	232.5636	137.6197	240.7820	142.2173	252.9278
Stepwise	107.3460	138.4203	234.9955	138.2101	237.6844	138.5520	242.0260
Ridge	109.3694	138.9341	230.6610	138.0531	231.1490	138.5844	256.2997
Lasso	110.1011	136.7551	227.0475	137.3988	232.2463	137.7901	236.0590

Table 60: MAR, Beta 2, equicorrelated 0.25, n=100, p=20, MSE of Prediction

		Linear Missing	Convex Missing	Sinister Missing
	Complete	25%	25%	25%
OLS	124.2014	319.0681	314.8202	324.7582
Stepwise	119.0144	314.4334	307.8894	320.6041
Ridge	122.4730	317.7844	306.5390	319.5896
Lasso	123.6116	331.8201	315.1352	331.3346

Table 61: MAR, Beta 2,equicorrelated 0.25, n=200, p=20, MSE of Prediction

		Linear Missing		Convex Missing		Sinister Missing	
	Complete	25%	50%	25%	50%	25%	50%
OLS	220.2049	362.8798	792.9154	365.6803	839.5905	371.0991	895.8982
Stepwise	215.0534	356.4157	808.2478	355.5532	810.0444	365.4716	884.0049
Ridge	221.0177	362.2311	797.8845	363.8140	824.8760	370.9729	882.0749
Lasso	223.3716	358.6786	807.4707	333.6596	713.3234	340.8430	755.2949

Table 62: MAR, Beta 2, equicorrelated 0.50, n=50, p=5, MSE of Prediction

		Linear Missing		Convex Missing		Sinister Missing	
	Complete	25%	50%	25%	50%	25%	50%
OLS	52.9322	60.9278	84.6885	62.3759	83.4234	62.9092	89.5624
Stepwise	50.1963	66.2506	83.6840	60.4004	82.6522	62.4638	88.3228
Ridge	54.1270	63.6931	81.9715	60.7036	80.1711	62.8453	84.6423
Lasso	52.2217	61.6363	77.9630	59.2910	77.8270	61.5606	79.7528

Table 63: MAR, Beta 2, equicorrelated 0.50, n=50, p=10, MSE of Prediction

		Linear Missing	Convex Missing	Sinister Missing
	Complete	25%	25%	25%
OLS	61.9440	99.5823	98.7587	97.4286
Stepwise	60.4055	97.4589	99.5372	94.6496
Ridge	61.3228	94.1730	99.8796	93.4017
Lasso	60.0211	86.2777	93.6448	84.6894

Table 64: MAR, Beta 2, equicorrelated 0.50, n=100, p=10, MSE of Prediction

		Linear Missing		Convex Missing		Sinister Missing	
	Complete	25%	50%	25%	50%	25%	50%
OLS	110.2927	140.6018	234.4312	138.3587	238.4604	145.8848	246.0462
Stepwise	107.4731	138.1872	238.5790	138.1345	236.6126	142.0871	246.9823
Ridge	109.4332	138.7602	233.0569	138.2877	231.2719	143.4015	242.7919
LASSO	109.9843	137.0641	231.3474	137.6017	230.7224	141.3007	244.6732

Table 65: MAR, Beta 2, equicorrelated 0.50, n=100, p=20, MSE of Prediction

		Linear Missing	Convex Missing	Sinister Missing
	Complete	25%	25%	25%
OLS	123.4252	318.0649	311.4063	327.1608
Stepwise	118.9435	314.6298	304.9000	325.7907
Ridge	121.7989	316.0135	303.7960	322.4671
LASSO	122.7538	330.6742	313.0575	338.1091

Table 66: MAR, Beta 2, equicorrelated 0.50, n=200, p=20, MSE of Prediction

		Linear Missing		Convex Missing		Sinister Missing	
	Complete	25%	50%	25%	50%	25%	50%
OLS	220.4095	362.7299	772.2296	359.9536	798.6486	367.4292	856.9889
Stepwise	215.9701	356.2062	786.0386	353.0787	815.2694	361.5859	862.9232
Ridge	220.8759	362.5637	777.6661	353.9950	805.1890	364.6881	870.6374
LASSO	223.5058	358.7953	811.9098	353.0124	858.2234	362.8708	904.2275

APPENDIX C

MAR PARAMETER ESTIMATES TABLES - N=50, P=5, BETA 1

Table 67: OLS - MAR, Beta 1, indep, n=50, p=5, Linear Missing at 25 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	2.9903	1.4904	-0.0048	1.9923	-0.0043
Standard Error	0.1481	0.1482	0.1478	0.148	0.1482
True Beta	3	1.5	0	2	0
MSE	0.0219	0.0226	0.0233	0.0209	0.0235
Coverage of 95 CI	95.1	94.6	93.7	94.9	94.3
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	3.006	1.4557	-0.0149	1.9782	-0.0182
Standard Error	0.2154	0.229	0.2337	0.2203	0.2383
True Beta	3	1.5	0	2	0
MSE	0.0581	0.0779	0.0728	0.0579	0.0741
Coverage of 95 CI	91.4	86.8	88.7	91.6	89.8

Table 68: Stepwise - MAR, Beta 1, indep, n=50, p=5, Linear Missing at 25 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	2.9903	1.4902	-0.0011	1.9913	-6e-04
Standard Error	0.1443	0.1443	0.1406	0.1442	0.1431
True Beta	3	1.5	0	2	0
MSE	0.0218	0.0225	0.0147	0.0207	0.0152
Coverage of 95 CI	94.2	94.5	na	95.2	na
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	3.0035	1.4531	-0.0117	1.9763	-0.0159
Standard Error	0.2109	0.2222	0.2133	0.2156	0.2237
True Beta	3	1.5	0	2	0
MSE	0.0575	0.0781	0.0608	0.0569	0.0618
Coverage of 95 CI	91	86.4	87.5	91.3	87.4

Table 69: Ridge - MAR, Beta 1, indep, n=50, p=5, Linear Missing at 25 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	2.9765	1.4833	-0.0047	1.9832	-0.0041
Standard Error	0.1473	0.1475	0.1471	0.1473	0.1475
True Beta	3	1.5	0	2	0
MSE	0.0224	0.0227	0.0231	0.021	0.0233
Coverage of 95 CI	94.1	94.3	93.7	94.6	94.2
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	2.9815	1.4432	-0.0143	1.9618	-0.0177
Standard Error	0.2124	0.2255	0.2297	0.2174	0.2342
True Beta	3	1.5	0	2	0
MSE	0.0572	0.078	0.0708	0.0574	0.072
Coverage of 95 CI	90.9	86.8	88.7	90.9	89.4

Table 70: LASSO - MAR, Beta 1, indep, n=50, p=5, Linear Missing at 25 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	2.92	1.4183	8e-04	1.9217	-0.0039
Standard Error	0.1465	0.1465	0.1465	0.1464	0.147
True Beta	3	1.5	0	2	0
MSE	0.0301	0.0325	0.0143	0.0292	0.0146
Coverage of 95 CI	89.3	89.1	95.2	91.5	95.9
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	2.9328	1.3816	-0.0102	1.9052	-0.0131
Standard Error	0.2171	0.2308	0.2237	0.2226	0.2282
True Beta	3	1.5	0	2	0
MSE	0.0622	0.0918	0.0562	0.0657	0.0565
Coverage of 95 CI	90.2	84.7	92	88.9	92.4

Table 71: OLS - MAR, Beta 1, indep, n=50, p=5, Linear Missing at 50 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	2.9998	1.4861	0.003	1.9975	-0.0044
Standard Error	0.1488	0.1492	0.1487	0.149	0.1492
True Beta	3	1.5	0	2	0
MSE	0.022	0.0209	0.0237	0.0223	0.0184
Coverage of 95 CI	94.6	95.4	93	93.4	96.4
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	2.862	1.2466	-0.0051	1.7744	0.0194
Standard Error	0.6369	0.6812	0.731	0.7023	0.7161
True Beta	3	1.5	0	2	0
MSE	0.2288	0.3077	0.2374	0.3486	0.2194
Coverage of 95 CI	96.5	96	97.7	96.2	96.2

Table 72: Stepwise - MAR, Beta 1, indep, n=50, p=5, Linear Missing at 50 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	3.0002	1.4867	0.0032	1.9971	-0.0067
Standard Error	0.145	0.145	0.1421	0.1449	0.1418
True Beta	3	1.5	0	2	0
MSE	0.0215	0.0203	0.0147	0.0217	0.0099
Coverage of 95 CI	95.3	94.3	na	93.6	na
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	2.8571	1.2229	-0.003	1.7628	0.0161
Standard Error	0.5926	0.5959	0.7688	0.6179	0.7727
True Beta	3	1.5	0	2	0
MSE	0.2286	0.3282	0.2139	0.3606	0.1994
Coverage of 95 CI	95.7	91.9	93.6	94.3	93.1

Table 73: Ridge - MAR, Beta 1, indep, n=50, p=5, Linear Missing at 50 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	2.9861	1.4792	0.0031	1.9883	-0.0043
Standard Error	0.1481	0.1485	0.148	0.1483	0.1484
True Beta	3	1.5	0	2	0
MSE	0.0223	0.021	0.0234	0.0224	0.0182
Coverage of 95 CI	95.3	95.5	93	93.8	96.4
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	2.7659	1.2059	0	1.7158	0.0197
Standard Error	0.593	0.6118	0.6435	0.6314	0.6356
True Beta	3	1.5	0	2	0
MSE	0.2533	0.31	0.2095	0.3504	0.1962
Coverage of 95 CI	94.4	94.5	97.2	94.8	96.1

Table 74: LASSO - MAR, Beta 1, indep, n=50, p=5, Linear Missing at 50 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	2.932	1.4172	0.0047	1.9293	-0.0029
Standard Error	0.1473	0.1477	0.1474	0.1475	0.1479
True Beta	3	1.5	0	2	0
MSE	0.0289	0.0293	0.0137	0.0301	0.009
Coverage of 95 CI	89.9	91.4	96	88.4	98.7
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	2.6926	1.1097	0.0049	1.6173	0.0142
Standard Error	0.614	0.6176	0.591	0.6427	0.5822
True Beta	3	1.5	0	2	0
MSE	0.2891	0.3801	0.1621	0.4193	0.1558
Coverage of 95 CI	92.2	89.2	96.3	89.8	96.2

Table 75: OLS - MAR, Beta 1, indep, n=50, p=5, Convex Missing at 25 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	3.004	1.499	-0.0064	2.0167	0.0149
Standard Error	0.1495	0.1496	0.1498	0.1499	0.1497
True Beta	3	1.5	0	2	0
MSE	0.0228	0.0203	0.0238	0.0246	0.024
Coverage of 95 CI	94.1	94.1	93.1	92.2	93.2
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	3.0183	1.4503	-0.0081	1.9977	0.0015
Standard Error	0.2209	0.2295	0.2438	0.2291	0.2419
True Beta	3	1.5	0	2	0
MSE	0.0541	0.0704	0.0783	0.0685	0.0909
Coverage of 95 CI	93.6	90.9	91.4	92.1	88.2

Table 76: Stepwise - MAR, Beta 1, indep, n=50, p=5, Convex Missing at 25 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	3.0044	1.4974	-0.0045	2.0165	0.0126
Standard Error	0.1455	0.1455	0.1438	0.1456	0.1416
True Beta	3	1.5	0	2	0
MSE	0.0224	0.0202	0.0159	0.024	0.0159
Coverage of 95 CI	93	94.3	na	91.6	na
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	3.0163	1.4478	-0.0116	1.9943	0.003
Standard Error	0.2163	0.224	0.2266	0.2222	0.2135
True Beta	3	1.5	0	2	0
MSE	0.0535	0.071	0.067	0.0677	0.079
Coverage of 95 CI	93.6	89.9	90.1	91.1	85.5

Table 77: Ridge - MAR, Beta 1, indep, n=50, p=5, Convex Missing at 25 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	2.99	1.4917	-0.006	2.0072	0.0149
Standard Error	0.1488	0.1488	0.1491	0.1492	0.1489
True Beta	3	1.5	0	2	0
MSE	0.0229	0.0202	0.0234	0.0242	0.0238
Coverage of 95 CI	94.1	94.1	93.4	92.4	93.3
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	2.993	1.4377	-0.0073	1.9806	0.0018
Standard Error	0.2181	0.2262	0.2396	0.226	0.2378
True Beta	3	1.5	0	2	0
MSE	0.0527	0.0701	0.0756	0.0673	0.0887
Coverage of 95 CI	93	90.8	90.8	91.4	87.9

Table 78: LASSO - MAR, Beta 1, indep, n=50, p=5, Convex Missing at 25 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	2.9371	1.4293	-0.0032	1.9467	0.013
Standard Error	0.148	0.1479	0.1485	0.1483	0.1485
True Beta	3	1.5	0	2	0
MSE	0.029	0.0268	0.0144	0.031	0.0148
Coverage of 95 CI	89.9	91.4	95	88.5	95.2
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	2.9472	1.3783	-0.0084	1.9234	0.005
Standard Error	0.2223	0.2311	0.234	0.2314	0.2308
True Beta	3	1.5	0	2	0
MSE	0.0568	0.0829	0.0604	0.0735	0.0732
Coverage of 95 CI	92.2	88.7	94.4	91.1	89.8

Table 79: OLS - MAR, Beta 1, indep, n=50, p=5, Convex Missing at 50 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	2.9909	1.5063	-0.0012	1.9986	-9e-04
Standard Error	0.1483	0.1481	0.148	0.1484	0.1485
True Beta	3	1.5	0	2	0
MSE	0.0241	0.0215	0.0214	0.0232	0.0217
Coverage of 95 CI	92	94.1	95.4	93.3	94.6
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	2.8231	1.2052	0.0217	1.6931	0
Standard Error	0.6384	0.7118	0.7053	0.6954	0.6936
True Beta	3	1.5	0	2	0
MSE	0.2848	0.3354	0.2506	0.3702	0.2454
Coverage of 95 CI	95.3	94.7	97.1	94	98.1

Table 80: Stepwise - MAR, Beta 1, indep, n=50, p=5, Convex Missing at 50 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	2.9915	1.5064	-0.0074	1.998	0.0067
Standard Error	0.1442	0.1442	0.1431	0.1443	0.1387
True Beta	3	1.5	0	2	0
MSE	0.023	0.021	0.0124	0.0225	0.0131
Coverage of 95 CI	92	94	na	93.1	na
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	2.8207	1.1857	0.0261	1.6817	0.0011
Standard Error	0.583	0.635	0.7557	0.6224	0.7636
True Beta	3	1.5	0	2	0
MSE	0.2862	0.3514	0.2286	0.3841	0.2208
Coverage of 95 CI	94.6	91.9	95.6	92.2	95.8

Table 81: Ridge - MAR, Beta 1, indep, n=50, p=5, Convex Missing at 50 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	2.9769	1.4992	-0.0013	1.989	-6e-04
Standard Error	0.1476	0.1474	0.1473	0.1477	0.1477
True Beta	3	1.5	0	2	0
MSE	0.0245	0.0214	0.0213	0.0233	0.0214
Coverage of 95 CI	91.7	94.7	95.4	93.3	94.9
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	2.7255	1.1658	0.021	1.6379	0.0021
Standard Error	0.5908	0.6305	0.6181	0.6215	0.6164
True Beta	3	1.5	0	2	0
MSE	0.3241	0.3372	0.2212	0.3853	0.2225
Coverage of 95 CI	93.6	93.8	97	92.2	97.6

Table 82: LASSO - MAR, Beta 1, indep, n=50, p=5, Convex Missing at 50 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	2.9196	1.4348	-0.0031	1.9273	0.0027
Standard Error	0.1467	0.1465	0.1468	0.1468	0.1471
True Beta	3	1.5	0	2	0
MSE	0.0331	0.0284	0.0127	0.0306	0.0125
Coverage of 95 CI	88.6	90.5	97	89.2	95.8
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	2.6496	1.0716	0.0211	1.547	0.0034
Standard Error	0.5866	0.622	0.5617	0.6196	0.5562
True Beta	3	1.5	0	2	0
MSE	0.3874	0.399	0.1735	0.4606	0.1741
Coverage of 95 CI	89.6	87.9	94.4	87.1	96.8

Table 83: OLS - MAR, Beta 1, indep, n=50, p=5, Sinister Missing at 25 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	3.0032	1.5028	0.0058	2.0002	-0.0036
Standard Error	0.1489	0.1485	0.1486	0.1488	0.1485
True Beta	3	1.5	0	2	0
MSE	0.0228	0.021	0.0219	0.0233	0.022
Coverage of 95 CI	94.4	94.5	94.4	94	94.4
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	3.0141	1.4815	0.0048	1.9912	-5e-04
Standard Error	0.2258	0.2355	0.2439	0.2303	0.2458
True Beta	3	1.5	0	2	0
MSE	0.0651	0.0695	0.0842	0.0662	0.0741
Coverage of 95 CI	91	90.2	89.2	90.7	91.3

Table 84: Stepwise - MAR, Beta 1, indep, n=50, p=5, Sinister Missing at 25 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	3.0032	1.5033	3e-04	2.001	-0.0016
Standard Error	0.1448	0.1446	0.143	0.1447	0.1426
True Beta	3	1.5	0	2	0
MSE	0.0224	0.0207	0.0134	0.0229	0.0136
Coverage of 95 CI	93.6	93.8	na	93.4	na
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	3.0119	1.4798	0.002	1.9892	-6e-04
Standard Error	0.221	0.229	0.2232	0.2252	0.228
True Beta	3	1.5	0	2	0
MSE	0.0643	0.0691	0.0715	0.0653	0.0625
Coverage of 95 CI	90.4	89.8	87.8	90.6	88.5

Table 85: Ridge - MAR, Beta 1, indep, n=50, p=5, Sinister Missing at 25 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	2.989	1.4956	0.0058	1.9906	-0.0036
Standard Error	0.1482	0.1478	0.1479	0.1481	0.1478
True Beta	3	1.5	0	2	0
MSE	0.0229	0.0209	0.0217	0.0232	0.0218
Coverage of 95 CI	93.5	94.5	94.3	93.7	94.4
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	2.9875	1.4682	0.0043	1.9732	-7e-04
Standard Error	0.2226	0.2319	0.2396	0.2272	0.2414
True Beta	3	1.5	0	2	0
MSE	0.0632	0.0686	0.0816	0.0651	0.0719
Coverage of 95 CI	90.9	90	89.1	90.5	91.1

Table 86: LASSO - MAR, Beta 1, indep, n=50, p=5, Sinister Missing at 25 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	2.9275	1.4279	0.0032	1.9249	-0.0026
Standard Error	0.1471	0.1468	0.1473	0.1471	0.1472
True Beta	3	1.5	0	2	0
MSE	0.0306	0.028	0.0125	0.0312	0.0122
Coverage of 95 CI	90.2	91	96.2	88.8	96.5
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	2.9409	1.4099	0.0023	1.9187	8e-04
Standard Error	0.2284	0.2373	0.2347	0.2336	0.2349
True Beta	3	1.5	0	2	0
MSE	0.0665	0.0779	0.0658	0.0722	0.0578
Coverage of 95 CI	89.2	89.9	92.3	90.2	93.9

Table 87: OLS - MAR, Beta 1, indep, n=50, p=5, Sinister Missing at 50 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	3.0022	1.4973	-9e-04	2.0009	0.0024
Standard Error	0.1483	0.1483	0.1485	0.1484	0.1483
True Beta	3	1.5	0	2	0
MSE	0.0211	0.0207	0.0215	0.0204	0.0212
Coverage of 95 CI	95.4	94.9	94.1	95	95.2
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	2.8556	1.1887	0.0041	1.7269	0.0131
Standard Error	0.6648	0.7411	0.7452	0.7385	0.74
True Beta	3	1.5	0	2	0
MSE	0.2763	0.3855	0.2928	0.3576	0.2514
Coverage of 95 CI	96.5	95.3	97.5	95.1	96.3

Table 88: Stepwise - MAR, Beta 1, indep, n=50, p=5, Sinister Missing at 50 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	3.0031	1.4975	0.0032	2.0006	-0.0049
Standard Error	0.1442	0.1442	0.1419	0.1443	0.1418
True Beta	3	1.5	0	2	0
MSE	0.0207	0.0204	0.0135	0.0204	0.0124
Coverage of 95 CI	94.8	94	na	94.3	na
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	2.8501	1.1686	0.0053	1.7124	0.0159
Standard Error	0.6228	0.6647	0.7943	0.658	0.8208
True Beta	3	1.5	0	2	0
MSE	0.2786	0.405	0.2675	0.3734	0.226
Coverage of 95 CI	95.3	90	93.7	90.8	92.9

Table 89: Ridge - MAR, Beta 1, indep, n=50, p=5, Sinister Missing at 50 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	2.9882	1.4902	-8e-04	1.9914	0.0025
Standard Error	0.1476	0.1476	0.1477	0.1477	0.1476
True Beta	3	1.5	0	2	0
MSE	0.0213	0.0206	0.0214	0.0204	0.021
Coverage of 95 CI	95.3	94.5	94.1	94.7	95.2
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	2.7552	1.1509	0.0055	1.6689	0.0115
Standard Error	0.6106	0.66	0.6602	0.6622	0.656
True Beta	3	1.5	0	2	0
MSE	0.3028	0.3832	0.2565	0.3699	0.2258
Coverage of 95 CI	93.9	94	97.1	94	95.8

Table 90: LASSO - MAR, Beta 1, indep, n=50, p=5, Sinister Missing at 50 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	2.9299	1.4252	0.0016	1.9276	-0.0037
Standard Error	0.1467	0.1467	0.1472	0.1467	0.147
True Beta	3	1.5	0	2	0
MSE	0.0288	0.0282	0.0125	0.0277	0.012
Coverage of 95 CI	91.3	90.5	96.2	91.5	96.5
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	2.6827	1.0646	0.0079	1.5771	0.0076
Standard Error	0.6396	0.6693	0.6208	0.6893	0.6125
True Beta	3	1.5	0	2	0
MSE	0.3564	0.4465	0.2055	0.4452	0.1824
Coverage of 95 CI	92.5	89.3	97.6	91.4	96.4

Table 91: OLS - MAR, Beta 1, equi 0.50, n=50, p=5, Linear Missing at 25 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	2.9613	1.4995	0.0141	1.9967	-0.0016
Standard Error	0.1908	0.1952	0.1956	0.1959	0.1917
True Beta	3	1.5	0	2	0
MSE	0.0288	0.052	0.044	0.0395	0.0434
Coverage of 95 CI	97.5	92.1	97.6	98.1	92
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	2.968	1.5207	0.0581	2.0135	0.0831
Standard Error	0.2683	0.2962	0.2991	0.2841	0.2821
True Beta	3	1.5	0	2	0
MSE	0.0684	0.0872	0.0794	0.0681	0.107
Coverage of 95 CI	96	95.7	95.4	97	90.5

Table 92: Stepwise - MAR, Beta 1, equi 0.50, n=50, p=5, Linear Missing at 25 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	2.96	1.501	-0.0015	1.9949	0.0139
Standard Error	0.1812	0.1839	0.1995	0.1851	0.1849
True Beta	3	1.5	0	2	0
MSE	0.0268	0.0527	0.0309	0.0371	0.0316
Coverage of 95 CI	97.2	91.7	na	97.4	na
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	2.969	1.5176	0.0597	2.0119	0.0857
Standard Error	0.2582	0.2818	0.2607	0.2687	0.234
True Beta	3	1.5	0	2	0
MSE	0.0672	0.0854	0.0663	0.068	0.0952
Coverage of 95 CI	95.1	94.8	94.2	96.7	89.7

Table 93: Ridge - MAR, Beta 1, equi 0.50, n=50, p=5, Linear Missing at 25 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	2.9522	1.4973	0.0195	1.9923	0.0034
Standard Error	0.1899	0.1941	0.1945	0.1949	0.1907
True Beta	3	1.5	0	2	0
MSE	0.0293	0.0512	0.0438	0.0389	0.0428
Coverage of 95 CI	97.7	92.1	97.4	98	91.7
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	2.9552	1.5179	0.0654	2.0068	0.0895
Standard Error	0.2655	0.2925	0.2954	0.2807	0.2786
True Beta	3	1.5	0	2	0
MSE	0.0682	0.0849	0.0795	0.0667	0.1061
Coverage of 95 CI	95.9	95.6	95.2	96.8	90.3

Table 94: LASSO - MAR, Beta 1, equi 0.50, n=50, p=5, Linear Missing at 25 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	2.8928	1.4456	0.0613	1.9367	0.0509
Standard Error	0.1894	0.1942	0.1917	0.1949	0.1875
True Beta	3	1.5	0	2	0
MSE	0.035	0.0486	0.0187	0.0434	0.0186
Coverage of 95 CI	97	92.4	98.4	97.8	92.7
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	2.9202	1.4657	0.1082	1.9599	0.1199
Standard Error	0.2634	0.2914	0.2761	0.2782	0.2672
True Beta	3	1.5	0	2	0
MSE	0.0708	0.0828	0.0647	0.0667	0.0884
Coverage of 95 CI	96.2	95.7	96.4	96.3	91

Table 95: OLS - MAR, Beta 1, equi 0.50, n=50, p=5, Linear Missing at 50 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	3.0142	1.4946	-0.0138	1.9701	0.023
Standard Error	0.1963	0.1955	0.1948	0.1976	0.1974
True Beta	3	1.5	0	2	0
MSE	0.0301	0.0268	0.0363	0.0294	0.0375
Coverage of 95 CI	98.4	99.4	94.8	95.7	95
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	3.011	1.3643	0.3664	1.9377	0.3828
Standard Error	0.7724	0.8914	0.8325	0.7798	0.8087
True Beta	3	1.5	0	2	0
MSE	0.2597	0.4258	0.6009	0.3048	0.4388
Coverage of 95 CI	98.5	95.4	96.4	98.5	96.4

Table 96: Stepwise - MAR, Beta 1, equi 0.50, n=50, p=5, Linear Missing at 50 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	3.0139	1.4941	-0.013	1.9754	0.015
Standard Error	0.183	0.1833	0.1809	0.1849	0.1942
True Beta	3	1.5	0	2	0
MSE	0.0281	0.0247	0.0199	0.0269	0.0248
Coverage of 95 CI	98.2	99.3	na	95.6	na
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	3.0214	1.3466	0.3539	1.9456	0.3747
Standard Error	0.6995	0.8076	0.8503	0.6802	0.8603
True Beta	3	1.5	0	2	0
MSE	0.2671	0.4388	0.5821	0.2979	0.4246
Coverage of 95 CI	96.9	93.9	92.1	95.7	89.4

Table 97: Ridge - MAR, Beta 1, equi 0.50, n=50, p=5, Linear Missing at 50 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	3.0044	1.4927	-0.0083	1.9655	0.0287
Standard Error	0.1953	0.1945	0.1938	0.1965	0.1963
True Beta	3	1.5	0	2	0
MSE	0.0295	0.0266	0.0357	0.0295	0.0374
Coverage of 95 CI	98.4	99.4	93.5	95.7	95
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	2.9785	1.3596	0.3778	1.9239	0.3906
Standard Error	0.7167	0.827	0.7812	0.7273	0.7545
True Beta	3	1.5	0	2	0
MSE	0.2435	0.4004	0.5756	0.2927	0.4286
Coverage of 95 CI	98.6	94.1	95	98.5	95.1

Table 98: LASSO - MAR, Beta 1, equi 0.50, n=50, p=5, Linear Missing at 50 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	2.9594	1.4502	0.0423	1.9175	0.0702
Standard Error	0.1951	0.1946	0.1914	0.1965	0.1948
True Beta	3	1.5	0	2	0
MSE	0.0272	0.0262	0.0179	0.0367	0.0214
Coverage of 95 CI	98.4	99.5	97.9	92.9	97.9
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	2.9346	1.3205	0.3932	1.862	0.4013
Standard Error	0.6906	0.7656	0.7058	0.6991	0.6669
True Beta	3	1.5	0	2	0
MSE	0.237	0.3684	0.5123	0.2769	0.3734
Coverage of 95 CI	95.8	92.6	93.9	94.5	97.7

Table 99: OLS - MAR, Beta 1, equi 0.50, n=50, p=5, Convex Missing at 25 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	3.009	1.5064	0.0358	1.9933	0.0161
Standard Error	0.1899	0.1916	0.1937	0.1908	0.1933
True Beta	3	1.5	0	2	0
MSE	0.0361	0.0347	0.0261	0.0346	0.0247
Coverage of 95 CI	97.4	97.2	97.1	97.1	96.9
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	3.0907	1.4348	0.1229	1.9911	0.0413
Standard Error	0.2754	0.2869	0.2888	0.2876	0.2931
True Beta	3	1.5	0	2	0
MSE	0.0834	0.0827	0.1041	0.0868	0.1018
Coverage of 95 CI	96.4	95	93.4	90.4	87.7

Table 100: Stepwise - MAR, Beta 1, equi 0.50, n=50, p=5, Convex Missing at 25 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	3.0161	1.5134	0.0224	2.0035	0.0021
Standard Error	0.1765	0.1772	0.1829	0.1765	0.1798
True Beta	3	1.5	0	2	0
MSE	0.0323	0.0319	0.0158	0.0359	0.011
Coverage of 95 CI	97	97.1	na	90.9	na
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	3.0971	1.4366	0.112	1.9981	0.0289
Standard Error	0.2611	0.2761	0.2834	0.2717	0.2541
True Beta	3	1.5	0	2	0
MSE	0.0837	0.0799	0.0903	0.0859	0.0817
Coverage of 95 CI	89.9	94.7	92.4	90.3	86.5

Table 101: Ridge - MAR, Beta 1, equi 0.50, n=50, p=5, Convex Missing at 25 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	3.0003	1.5045	0.0408	1.9892	0.021
Standard Error	0.189	0.1907	0.1927	0.1899	0.1923
True Beta	3	1.5	0	2	0
MSE	0.0357	0.0344	0.0263	0.0343	0.0247
Coverage of 95 CI	97.3	97.1	97.1	97	96.8
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	3.0767	1.4332	0.1298	1.9849	0.0477
Standard Error	0.2721	0.2833	0.2852	0.2841	0.2894
True Beta	3	1.5	0	2	0
MSE	0.0799	0.0813	0.1041	0.0853	0.1021
Coverage of 95 CI	96.3	95	93.2	90.4	87.5

Table 102: LASSO - MAR, Beta 1, equi 0.50, n=50, p=5, Convex Missing at 25 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	2.9639	1.4596	0.0569	1.9504	0.0434
Standard Error	0.1892	0.1909	0.1919	0.1902	0.1913
True Beta	3	1.5	0	2	0
MSE	0.0342	0.0358	0.016	0.0356	0.0136
Coverage of 95 CI	97	97.2	97.8	96.7	98.1
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	3.0448	1.3851	0.1597	1.9486	0.0752
Standard Error	0.2717	0.282	0.2685	0.2823	0.2823
True Beta	3	1.5	0	2	0
MSE	0.0712	0.0872	0.088	0.0842	0.078
Coverage of 95 CI	96.6	95.6	95	90.4	94.9

Table 103: OLS - MAR, Beta 1, equi 0.50, n=50, p=5, Convex Missing at 50 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	2.9923	1.5115	-0.0079	2.0035	0.0022
Standard Error	0.1936	0.1942	0.194	0.1925	0.1934
True Beta	3	1.5	0	2	0
MSE	0.0394	0.0368	0.0387	0.0367	0.0412
Coverage of 95 CI	93.7	93.9	94.2	94.8	92.4
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	3.0015	1.4247	0.3133	1.8479	0.3143
Standard Error	0.7529	0.8202	0.7694	0.7641	0.7687
True Beta	3	1.5	0	2	0
MSE	0.3295	0.3181	0.3833	0.3444	0.4721
Coverage of 95 CI	96.9	97.3	92.2	94.6	93.1

Table 104: Stepwise - MAR, Beta 1, equi 0.50, n=50, p=5, Convex Missing at 50 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	2.9919	1.5107	7e-04	2.0022	-0.002
Standard Error	0.1805	0.1806	0.1819	0.1792	0.1841
True Beta	3	1.5	0	2	0
MSE	0.0374	0.0353	0.0225	0.0364	0.0258
Coverage of 95 CI	92.6	93.4	na	93.5	na
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	3.0099	1.4172	0.3047	1.8462	0.3053
Standard Error	0.6842	0.7009	0.8417	0.6758	0.8223
True Beta	3	1.5	0	2	0
MSE	0.3303	0.3303	0.3649	0.3499	0.447
Coverage of 95 CI	95	94.3	85.8	92.3	86.5

Table 105: Ridge - MAR, Beta 1, equi 0.50, n=50, p=5, Convex Missing at 50 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	2.983	1.5096	-0.0026	1.9989	0.0075
Standard Error	0.1926	0.1931	0.1929	0.1915	0.1923
True Beta	3	1.5	0	2	0
MSE	0.0394	0.0364	0.0382	0.0363	0.0409
Coverage of 95 CI	93.8	93.9	94.2	94.9	92.2
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	2.9665	1.4201	0.3251	1.8332	0.3257
Standard Error	0.7077	0.769	0.7258	0.7204	0.7248
True Beta	3	1.5	0	2	0
MSE	0.3159	0.3026	0.3781	0.3321	0.4587
Coverage of 95 CI	96.7	97.3	92	94.3	92.1

Table 106: LASSO - MAR, Beta 1, equi 0.50, n=50, p=5, Convex Missing at 50 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	2.9352	1.4562	0.0408	1.9467	0.0404
Standard Error	0.1925	0.1931	0.1909	0.1914	0.1906
True Beta	3	1.5	0	2	0
MSE	0.0419	0.0356	0.0201	0.0383	0.023
Coverage of 95 CI	93	95.5	95.5	93.6	95
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	2.9234	1.3704	0.3443	1.7732	0.3415
Standard Error	0.6928	0.737	0.655	0.6961	0.6591
True Beta	3	1.5	0	2	0
MSE	0.3108	0.294	0.3435	0.3394	0.4042
Coverage of 95 CI	96.1	96.3	93	94.3	93.4

Table 107: OLS - MAR, Beta 1, equi 0.50, n=50, p=5, Sinister Missing at 25 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	3.0082	1.4926	-0.0031	2.0073	-3e-04
Standard Error	0.1937	0.1927	0.1933	0.1938	0.1931
True Beta	3	1.5	0	2	0
MSE	0.0384	0.0349	0.0342	0.0414	0.0357
Coverage of 95 CI	93.6	95.8	95.3	93.1	95.8
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	3.0375	1.4956	0.0619	1.9934	0.0674
Standard Error	0.2823	0.2929	0.3001	0.288	0.3013
True Beta	3	1.5	0	2	0
MSE	0.0981	0.1061	0.1189	0.1097	0.1201
Coverage of 95 CI	90.8	92.2	89.2	88	90.2

Table 108: Stepwise - MAR, Beta 1, equi 0.50, n=50, p=5, Sinister Missing at 25 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	3.0079	1.4905	-5e-04	2.0061	0.0022
Standard Error	0.1798	0.1795	0.1801	0.1801	0.1812
True Beta	3	1.5	0	2	0
MSE	0.0363	0.0341	0.0193	0.0385	0.0206
Coverage of 95 CI	92.1	94.4	na	91.5	na
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	3.0395	1.495	0.0584	1.9954	0.0626
Standard Error	0.2708	0.2782	0.2813	0.2748	0.274
True Beta	3	1.5	0	2	0
MSE	0.0956	0.1064	0.1009	0.1073	0.1035
Coverage of 95 CI	90.1	90.5	87.1	86.9	86.2

Table 109: Ridge - MAR, Beta 1, equi 0.50, n=50, p=5, Sinister Missing at 25 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	2.9989	1.4907	0.0023	2.0029	0.0051
Standard Error	0.1927	0.1918	0.1923	0.1928	0.1921
True Beta	3	1.5	0	2	0
MSE	0.038	0.0345	0.0338	0.041	0.0354
Coverage of 95 CI	93.3	95.8	95.2	92.8	95.9
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	3.0232	1.4929	0.0694	1.9872	0.0748
Standard Error	0.2788	0.289	0.2963	0.2842	0.2973
True Beta	3	1.5	0	2	0
MSE	0.0953	0.104	0.1178	0.1077	0.1192
Coverage of 95 CI	90.8	92.2	89	87.8	89.6

Table 110: LASSO - MAR, Beta 1, equi 0.50, n=50, p=5, Sinister Missing at 25 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	2.952	1.4344	0.0405	1.9512	0.0422
Standard Error	0.1927	0.1917	0.1904	0.1928	0.1902
True Beta	3	1.5	0	2	0
MSE	0.0379	0.0381	0.0168	0.0411	0.0179
Coverage of 95 CI	93.7	94.8	96.8	92.4	97.4
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	2.9851	1.4428	0.1053	1.9418	0.108
Standard Error	0.2782	0.2881	0.284	0.2833	0.2832
True Beta	3	1.5	0	2	0
MSE	0.0912	0.1032	0.0927	0.1064	0.0965
Coverage of 95 CI	91.5	92	91.8	88.4	92.3

Table 111: OLS - MAR, Beta 1, equi 0.50, n=50, p=5, Sinister Missing at 50 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	3.012	1.5032	-0.0032	1.9951	-0.001
Standard Error	0.1928	0.1933	0.1928	0.1933	0.1922
True Beta	3	1.5	0	2	0
MSE	0.0387	0.0339	0.0381	0.04	0.039
Coverage of 95 CI	94.1	95.4	93.8	94	94
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	3.0247	1.3983	0.3242	1.8664	0.3185
Standard Error	0.8254	0.8939	0.854	0.8662	0.8537
True Beta	3	1.5	0	2	0
MSE	0.3547	0.3401	0.442	0.3706	0.4465
Coverage of 95 CI	96.4	95.9	93.4	96.7	94.9

Table 112: Stepwise - MAR, Beta 1, equi 0.50, n=50, p=5, Sinister Missing at 50 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	3.0114	1.5038	-0.0054	1.9971	-0.0013
Standard Error	0.18	0.18	0.1799	0.1802	0.1804
True Beta	3	1.5	0	2	0
MSE	0.0367	0.0327	0.0232	0.0378	0.0236
Coverage of 95 CI	92.4	95	na	92.9	na
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	3.0335	1.388	0.3171	1.8641	0.3137
Standard Error	0.7589	0.7994	0.9094	0.7562	0.9127
True Beta	3	1.5	0	2	0
MSE	0.3531	0.3503	0.4195	0.3752	0.4236
Coverage of 95 CI	96	93.8	88.9	92.7	90.7

Table 113: Ridge - MAR, Beta 1, equi 0.50, n=50, p=5, Sinister Missing at 50 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	3.0028	1.5012	0.0022	1.9908	0.0043
Standard Error	0.1918	0.1923	0.1918	0.1923	0.1912
True Beta	3	1.5	0	2	0
MSE	0.0384	0.0336	0.0377	0.0397	0.0386
Coverage of 95 CI	93.8	95.5	93.7	94.2	94.1
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	2.9862	1.3926	0.3377	1.8532	0.3314
Standard Error	0.7666	0.8308	0.7986	0.8084	0.7995
True Beta	3	1.5	0	2	0
MSE	0.3351	0.3214	0.4322	0.3542	0.4344
Coverage of 95 CI	96.4	95.6	93.1	96	94.4

Table 114: LASSO - MAR, Beta 1, equi 0.50, n=50, p=5, Sinister Missing at 50 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	2.9557	1.4475	0.0417	1.9402	0.0428
Standard Error	0.1918	0.1923	0.1898	0.1923	0.1894
True Beta	3	1.5	0	2	0
MSE	0.0392	0.0353	0.0196	0.04	0.0213
Coverage of 95 CI	94.1	95.6	95.8	94	96.2
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	2.9386	1.3416	0.3552	1.7942	0.3517
Standard Error	0.7522	0.7817	0.7279	0.7771	0.7268
True Beta	3	1.5	0	2	0
MSE	0.3269	0.3123	0.3827	0.3579	0.386
Coverage of 95 CI	96.4	95.3	93.8	94.7	95.5

APPENDIX D

MAR PARAMETER ESTIMATES TABLES - N=50, P=5, BETA 2

Table 115: OLS - MAR, Beta 2, indep, n=50, p=5, Linear Missing at 25 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	0.8526	0.8411	0.0038	2.0048	-0.0145
Standard Error	0.1532	0.1577	0.1577	0.1579	0.1535
True Beta	0.85	0.85	0	2	0
MSE	0.0251	0.0248	0.0272	0.0265	0.0242
Coverage of 95 CI	93.2	96	93.4	93.6	93.6
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	0.8435	0.8275	0.0314	2.0024	0.0045
Standard Error	0.1993	0.2065	0.2128	0.1989	0.2048
True Beta	0.85	0.85	0	2	0
MSE	0.0537	0.0505	0.0598	0.0456	0.0598
Inclusion Frequency	95.4	95.5	9.8	100	12.2

Table 116: Stepwise - MAR, Beta 2, indep, n=50, p=5, Linear Missing at 25 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	0.8512	0.8435	0.0015	2.0042	-0.0135
Standard Error	0.1489	0.1501	0.1521	0.1463	0.1468
True Beta	0.85	0.85	0	2	0
MSE	0.0244	0.0245	0.0166	0.0253	0.015
Coverage of 95 CI	93	94.2	na	92	na
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	0.8406	0.8249	0.028	2.0015	0.0033
Standard Error	0.1906	0.1953	0.1832	0.1888	0.1762
True Beta	0.85	0.85	0	2	0
MSE	0.0555	0.0524	0.0488	0.0441	0.0482
Coverage of 95 CI	88	89.8	88.6	91	87.8

Table 117: Ridge - MAR, Beta 2, indep, n=50, p=5, Linear Missing at 25 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	0.8442	0.8329	0.0121	1.979	-0.0081
Standard Error	0.1511	0.1553	0.1553	0.1555	0.1513
True Beta	0.85	0.85	0	2	0
MSE	0.0245	0.0243	0.0266	0.0268	0.0234
Coverage of 95 CI	93.7	95.9	93.5	93.6	93.7
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	0.8314	0.8157	0.0421	1.9656	0.0117
Standard Error	0.1937	0.2003	0.206	0.1937	0.1988
True Beta	0.85	0.85	0	2	0
MSE	0.0513	0.0485	0.0571	0.046	0.0567
Coverage of 95 CI	89.5	91.6	90.2	91.1	87.9

Table 118: LASSO - MAR, Beta 2, indep, n=50, p=5, Linear Missing at 25 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	0.7923	0.7833	0.0277	1.9268	0.0016
Standard Error	0.1512	0.1559	0.155	0.1546	0.151
True Beta	0.85	0.85	0	2	0
MSE	0.0297	0.0302	0.0163	0.0326	0.0142
Coverage of 95 CI	90.7	90.1	95.4	91.2	94.7
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	0.785	0.7695	0.0506	1.9256	0.0162
Standard Error	0.1986	0.2034	0.2026	0.1989	0.1969
True Beta	0.85	0.85	0	2	0
MSE	0.0554	0.0537	0.0441	0.0492	0.0431
Coverage of 95 CI	89.7	90.6	93.4	91.6	92.8

Table 119: OLS - MAR, Beta 2, indep, n=50, p=5, Linear Missing at 50 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	0.86	0.8464	0.0079	1.9958	-0.0089
Standard Error	0.1567	0.1608	0.1611	0.1611	0.1557
True Beta	0.85	0.85	0	2	0
MSE	0.0219	0.0259	0.0249	0.024	0.0271
Coverage of 95 CI	94.3	93.7	93.8	93.8	92.7
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	0.7584	0.7959	0.1522	1.8989	0.0766
Standard Error	0.4098	0.4436	0.4256	0.3901	0.4068
True Beta	0.85	0.85	0	2	0
MSE	0.1392	0.1522	0.1486	0.1485	0.1419
Coverage of 95 CI	94.3	93.9	93.9	90.7	94.5

Table 120: Stepwise - MAR, Beta 2, indep, n=50, p=5, Linear Missing at 50 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	0.8603	0.8434	0.0106	1.997	-0.0102
Standard Error	0.1522	0.1533	0.1556	0.1487	0.15
True Beta	0.85	0.85	0	2	0
MSE	0.0215	0.0246	0.0161	0.0217	0.0183
Coverage of 95 CI	93.2	93.8	na	94.9	na
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	0.7411	0.7842	0.1437	1.9002	0.0749
Standard Error	0.3611	0.3927	0.4287	0.3538	0.4218
True Beta	0.85	0.85	0	2	0
MSE	0.1479	0.1546	0.1354	0.1458	0.1277
Coverage of 95 CI	90.5	92.5	88.5	88	88.5

Table 121: Ridge - MAR, Beta 2, indep, n=50, p=5, Linear Missing at 50 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	0.8517	0.8381	0.017	1.9694	-0.0026
Standard Error	0.1544	0.1582	0.1585	0.1585	0.1535
True Beta	0.85	0.85	0	2	0
MSE	0.0211	0.0255	0.0245	0.025	0.0262
Coverage of 95 CI	94.3	94.2	94.2	92.9	93.2
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	0.7356	0.7714	0.1606	1.8254	0.082
Standard Error	0.3617	0.3866	0.3723	0.3454	0.3616
True Beta	0.85	0.85	0	2	0
MSE	0.1262	0.1371	0.1351	0.1634	0.1258
Coverage of 95 CI	93.5	92.6	92.4	86.5	93.8

Table 122: LASSO - MAR, Beta 2, indep, n=50, p=5, Linear Missing at 50 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	0.7978	0.7831	0.0334	1.9153	0.0029
Standard Error	0.1547	0.1588	0.1583	0.1579	0.1531
True Beta	0.85	0.85	0	2	0
MSE	0.0253	0.0301	0.0137	0.0306	0.0157
Coverage of 95 CI	93.6	91.3	95.8	90.7	94.7
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	0.6702	0.7134	0.1527	1.7694	0.0738
Standard Error	0.359	0.3851	0.3526	0.3609	0.3473
True Beta	0.85	0.85	0	2	0
MSE	0.1415	0.1412	0.1043	0.1886	0.0982
Coverage of 95 CI	90.4	90.3	95.3	85.5	95.5

Table 123: OLS - MAR, Beta 2, indep, n=50, p=5, Convex Missing at 25 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	0.8434	0.8612	0.0165	1.9957	-0.0056
Standard Error	0.1525	0.1575	0.1575	0.1592	0.1551
True Beta	0.85	0.85	0	2	0
MSE	0.0229	0.027	0.0267	0.0203	0.0225
Coverage of 95 CI	96.9	94.2	95.4	96	97.6
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	0.8661	0.825	0.0395	1.9916	0.0251
Standard Error	0.2048	0.2166	0.2132	0.2034	0.2111
True Beta	0.85	0.85	0	2	0
MSE	0.0508	0.054	0.0619	0.045	0.0505
Coverage of 95 CI	91.3	93.4	87.9	93.6	91.6

Table 124: Stepwise - MAR, Beta 2, indep, n=50, p=5, Convex Missing at 25 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	0.843	0.8645	0.0141	1.9929	0.0066
Standard Error	0.1479	0.1498	0.1484	0.1464	0.1509
True Beta	0.85	0.85	0	2	0
MSE	0.0231	0.0257	0.0153	0.0188	0.0121
Coverage of 95 CI	96.8	91.9	na	96.8	na
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	0.8625	0.822	0.0331	1.9934	0.0245
Standard Error	0.1986	0.2063	0.1869	0.1928	0.18
True Beta	0.85	0.85	0	2	0
MSE	0.0535	0.0555	0.048	0.0429	0.0385
Coverage of 95 CI	88.4	93	87.9	93.6	90.3

Table 125: Ridge - MAR, Beta 2, indep, n=50, p=5, Convex Missing at 25 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	0.8358	0.8524	0.0246	1.9699	0.0015
Standard Error	0.1504	0.1551	0.1551	0.1568	0.1528
True Beta	0.85	0.85	0	2	0
MSE	0.0224	0.026	0.026	0.0208	0.0218
Coverage of 95 CI	96.8	94.2	96.5	95.8	97.8
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	0.8525	0.8111	0.0503	1.9538	0.0326
Standard Error	0.1987	0.2095	0.2064	0.1978	0.2046
True Beta	0.85	0.85	0	2	0
MSE	0.0476	0.0521	0.06	0.0452	0.0486
Coverage of 95 CI	90.2	93.6	88.1	92.6	91.6

Table 126: LASSO - MAR, Beta 2, indep, n=50, p=5, Convex Missing at 25 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	0.7814	0.8003	0.0306	1.9156	0.0183
Standard Error	0.1506	0.1555	0.1548	0.1559	0.1524
True Beta	0.85	0.85	0	2	0
MSE	0.0314	0.0302	0.0158	0.0277	0.0118
Coverage of 95 CI	91.5	92.9	95.9	92.1	98
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	0.8012	0.7612	0.0586	1.9141	0.0354
Standard Error	0.2028	0.2134	0.2043	0.2013	0.2023
True Beta	0.85	0.85	0	2	0
MSE	0.0525	0.0578	0.0442	0.0496	0.0346
Coverage of 95 CI	90.1	90.6	92.2	93.1	98

Table 127: OLS - MAR, Beta 2, indep, n=50, p=5, Convex Missing at 50 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	0.8521	0.8479	0.0057	2.0055	7e-04
Standard Error	0.1486	0.1485	0.1486	0.1487	0.1485
True Beta	0.85	0.85	0	2	0
MSE	0.0225	0.0227	0.0233	0.0232	0.0209
Coverage of 95 CI	94.6	93.6	94.9	93.8	94.4
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	0.6832	0.6881	0.0158	1.9207	-0.006
Standard Error	0.3909	0.3856	0.381	0.3474	0.3826
True Beta	0.85	0.85	0	2	0
MSE	0.1667	0.1569	0.1265	0.1136	0.1299
Coverage of 95 CI	91.9	91.5	94.2	93.4	92.1

Table 128: Stepwise - MAR, Beta 2, indep, n=50, p=5, Convex Missing at 50 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	0.8515	0.8484	0.0013	2.0055	0.0017
Standard Error	0.1444	0.1445	0.1432	0.1444	0.1419
True Beta	0.85	0.85	0	2	0
MSE	0.0222	0.0224	0.0144	0.0228	0.0124
Coverage of 95 CI	94	93.6	na	93.5	na
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	0.668	0.6716	0.0149	1.9167	-0.0076
Standard Error	0.3542	0.3477	0.3875	0.3211	0.3886
True Beta	0.85	0.85	0	2	0
MSE	0.1725	0.1611	0.112	0.1136	0.1179
Coverage of 95 CI	88.4	86.9	89.7	92.1	89.3

Table 129: Ridge - MAR, Beta 2, indep, n=50, p=5, Convex Missing at 50 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	0.841	0.8371	0.0056	1.9801	9e-04
Standard Error	0.1465	0.1465	0.1466	0.1467	0.1465
True Beta	0.85	0.85	0	2	0
MSE	0.0222	0.0226	0.0228	0.0237	0.0204
Coverage of 95 CI	94.3	93	95	93.1	94.6
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	0.6568	0.6612	0.0153	1.8413	-0.0051
Standard Error	0.3473	0.3438	0.3397	0.3207	0.3373
True Beta	0.85	0.85	0	2	0
MSE	0.1595	0.1506	0.111	0.1298	0.1144
Coverage of 95 CI	90.3	89.2	93.6	89.7	92

Table 130: LASSO - MAR, Beta 2, indep, n=50, p=5, Convex Missing at 50 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	0.7783	0.7757	0.004	1.9321	0.0018
Standard Error	0.1457	0.1457	0.1463	0.1458	0.1463
True Beta	0.85	0.85	0	2	0
MSE	0.0301	0.0303	0.0135	0.0301	0.0114
Coverage of 95 CI	89.7	90.5	96.2	89.6	96.5
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	0.5917	0.5938	0.012	1.7856	-0.0043
Standard Error	0.3447	0.3429	0.3207	0.3301	0.3185
True Beta	0.85	0.85	0	2	0
MSE	0.1812	0.1719	0.0843	0.1556	0.089
Coverage of 95 CI	83.4	83.3	94	86.4	93.4

Table 131: OLS - MAR, Beta 2, indep, n=50, p=5, Sinister Missing at 25 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	0.8511	0.8544	-0.0024	2.0075	-0.0095
Standard Error	0.1543	0.1587	0.1588	0.1594	0.1541
True Beta	0.85	0.85	0	2	0
MSE	0.0225	0.023	0.0264	0.0273	0.0214
Coverage of 95 CI	96.2	95.4	92.7	92.9	95.3
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	0.8422	0.82	0.0556	2.0014	-0.001
Standard Error	0.2097	0.2139	0.218	0.2037	0.2111
True Beta	0.85	0.85	0	2	0
MSE	0.0558	0.0679	0.062	0.0515	0.0558
Coverage of 95 CI	92.8	89.2	90.3	92.2	90.5

Table 132: Stepwise - MAR, Beta 2, indep, n=50, p=5, Sinister Missing at 25 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	0.8526	0.8532	-0.0022	2.0057	-0.0046
Standard Error	0.1501	0.1512	0.1531	0.1473	0.1456
True Beta	0.85	0.85	0	2	0
MSE	0.0225	0.0229	0.0171	0.0255	0.0121
Coverage of 95 CI	95.6	93.8	na	92.6	na
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	0.8383	0.8165	0.0488	1.9991	5e-04
Standard Error	0.1995	0.2011	0.1964	0.1926	0.1846
True Beta	0.85	0.85	0	2	0
MSE	0.0576	0.0686	0.0494	0.0494	0.0431
Coverage of 95 CI	90.7	87	89.4	90.6	90.5

Table 133: Ridge - MAR, Beta 2, indep, n=50, p=5, Sinister Missing at 25 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	0.8432	0.8459	0.0063	1.9816	-0.0033
Standard Error	0.1522	0.1563	0.1563	0.1569	0.1519
True Beta	0.85	0.85	0	2	0
MSE	0.022	0.0224	0.0257	0.0277	0.0207
Coverage of 95 CI	96.1	95.5	93	92.9	95.3
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	0.8294	0.8089	0.0655	1.9631	0.0073
Standard Error	0.2031	0.2068	0.2109	0.1979	0.2044
True Beta	0.85	0.85	0	2	0
MSE	0.0534	0.0647	0.0598	0.0514	0.0525
Coverage of 95 CI	92.4	89.3	89.5	90.6	90.2

Table 134: LASSO - MAR, Beta 2, indep, n=50, p=5, Sinister Missing at 25 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	0.7909	0.7906	0.0251	1.9255	0.0052
Standard Error	0.1525	0.1567	0.1559	0.156	0.1515
True Beta	0.85	0.85	0	2	0
MSE	0.0276	0.0274	0.0144	0.033	0.0115
Coverage of 95 CI	91.4	92.4	95.8	89.7	96.7
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	0.778	0.7597	0.072	1.9178	0.0124
Standard Error	0.2076	0.2104	0.2068	0.2023	0.2018
True Beta	0.85	0.85	0	2	0
MSE	0.0594	0.0695	0.0446	0.0554	0.0383
Coverage of 95 CI	91.6	88.4	93.1	90	93.3

Table 135: OLS - MAR, Beta 2, indep, n=50, p=5, Sinister Missing at 50 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	0.8476	0.8455	0.0014	1.9973	-0.0033
Standard Error	0.1539	0.1594	0.159	0.1582	0.1535
True Beta	0.85	0.85	0	2	0
MSE	0.0263	0.0263	0.0238	0.0241	0.0241
Coverage of 95 CI	93.3	94.8	95.3	95.7	95.5
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	0.7764	0.7793	0.1382	1.8755	0.0863
Standard Error	0.4169	0.4318	0.4387	0.3963	0.42
True Beta	0.85	0.85	0	2	0
MSE	0.1397	0.1508	0.1558	0.1314	0.1341
Coverage of 95 CI	92.7	94.1	91.7	92.7	94.6

Table 136: Stepwise - MAR, Beta 2, indep, n=50, p=5, Sinister Missing at 50 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	0.8487	0.8439	0.004	1.9958	0.0013
Standard Error	0.1496	0.1506	0.15	0.1462	0.149
True Beta	0.85	0.85	0	2	0
MSE	0.0267	0.0252	0.0132	0.0221	0.0154
Coverage of 95 CI	91.8	93.9	na	94.9	na
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	0.7616	0.7659	0.1353	1.8727	0.0809
Standard Error	0.3723	0.3848	0.453	0.3623	0.4391
True Beta	0.85	0.85	0	2	0
MSE	0.1442	0.155	0.144	0.1311	0.1216
Coverage of 95 CI	91	92.8	87.8	90.9	91.4

Table 137: Ridge - MAR, Beta 2, indep, n=50, p=5, Sinister Missing at 50 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	0.8399	0.8366	0.0102	1.9714	0.0028
Standard Error	0.1517	0.1568	0.1565	0.1557	0.1514
True Beta	0.85	0.85	0	2	0
MSE	0.0257	0.0258	0.0233	0.0247	0.0233
Coverage of 95 CI	93.4	94.8	94.9	95.1	95.3
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	0.7495	0.7545	0.1487	1.7961	0.0927
Standard Error	0.3706	0.3792	0.3849	0.36	0.3713
True Beta	0.85	0.85	0	2	0
MSE	0.13	0.1342	0.1415	0.1488	0.1198
Coverage of 95 CI	92.3	93.2	90.3	89.2	93.7

Table 138: LASSO - MAR, Beta 2, indep, n=50, p=5, Sinister Missing at 50 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	0.7905	0.7839	0.0292	1.9165	0.014
Standard Error	0.1522	0.1573	0.1562	0.1548	0.1509
True Beta	0.85	0.85	0	2	0
MSE	0.0311	0.0308	0.0128	0.0305	0.0129
Coverage of 95 CI	90.1	93.3	96.6	91.8	96.2
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	0.6881	0.696	0.1459	1.7452	0.0857
Standard Error	0.3719	0.3731	0.3643	0.3712	0.3478
True Beta	0.85	0.85	0	2	0
MSE	0.1425	0.1391	0.1165	0.1714	0.096
Coverage of 95 CI	89.3	91.3	92.9	87.8	94.4

Table 139: OLS - MAR, Beta 2, equi 0.50, n=50, p=5, Linear Missing at 25 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	0.8445	0.8423	0.0101	1.9941	0.0074
Standard Error	0.1546	0.1577	0.1581	0.159	0.1538
True Beta	0.85	0.85	0	2	0
MSE	0.0158	0.023	0.0196	0.016	0.0182
Coverage of 95 CI	96.1	94.2	99.8	100	94.1
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	0.8326	0.8654	0.0233	1.9872	0.0359
Standard Error	0.2038	0.2095	0.2142	0.1934	0.2055
True Beta	0.85	0.85	0	2	0
MSE	0.0467	0.0603	0.0694	0.0442	0.0571
Coverage of 95 CI	92.2	94.1	84.5	90.4	91.6

Table 140: Stepwise - MAR, Beta 2, equi 0.50, n=50, p=5, Linear Missing at 25 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	0.8423	0.8473	0.0031	1.9891	0.0163
Standard Error	0.1495	0.1499	0.1566	0.146	0.1389
True Beta	0.85	0.85	0	2	0
MSE	0.0145	0.0238	0.0091	0.0164	0.0105
Coverage of 95 CI	96.1	94.2	na	97.9	na
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	0.8288	0.8599	0.0201	1.9867	0.0395
Standard Error	0.1961	0.1984	0.1798	0.1836	0.1659
True Beta	0.85	0.85	0	2	0
MSE	0.049	0.0634	0.0576	0.0433	0.0445
Coverage of 95 CI	86.6	90.4	86.4	90.3	91.7

Table 141: Ridge - MAR, Beta 2, equi 0.50, n=50, p=5, Linear Missing at 25 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	0.836	0.8344	0.0186	1.968	0.014
Standard Error	0.1524	0.1553	0.1557	0.1565	0.1516
True Beta	0.85	0.85	0	2	0
MSE	0.0157	0.0227	0.0191	0.0172	0.0178
Coverage of 95 CI	96.1	96.1	99.8	97.8	94.1
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	0.8198	0.8536	0.0359	1.9528	0.0431
Standard Error	0.1978	0.2028	0.2069	0.1886	0.1992
True Beta	0.85	0.85	0	2	0
MSE	0.0449	0.0568	0.0655	0.0456	0.0548
Coverage of 95 CI	92.2	92.3	84.5	90.4	89.8

Table 142: LASSO - MAR, Beta 2, equi 0.50, n=50, p=5, Linear Missing at 25 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	0.7841	0.7917	0.0222	1.9163	0.0189
Standard Error	0.1526	0.1563	0.1554	0.1557	0.1514
True Beta	0.85	0.85	0	2	0
MSE	0.0197	0.0276	0.0085	0.0261	0.0101
Coverage of 95 CI	92.1	94.2	99.8	99.6	99.7
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	0.7683	0.8012	0.0495	1.9103	0.0473
Standard Error	0.2037	0.2077	0.2041	0.1938	0.1975
True Beta	0.85	0.85	0	2	0
MSE	0.0539	0.0592	0.0502	0.0533	0.0396
Coverage of 95 CI	90.4	88.6	94.1	88.6	93.7

Table 143: OLS - MAR, Beta 2, equi 0.50, n=50, p=5, Linear Missing at 50 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	0.845	0.8639	-0.0154	2.0178	0.0022
Standard Error	0.1531	0.1582	0.1584	0.1583	0.1539
True Beta	0.85	0.85	0	2	0
MSE	0.0242	0.0265	0.0227	0.0255	0.0245
Coverage of 95 CI	94.8	95.2	96.4	95	94.6
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	0.7582	0.7769	0.1329	1.906	0.0877
Standard Error	0.4017	0.4253	0.4196	0.3624	0.4005
True Beta	0.85	0.85	0	2	0
MSE	0.1388	0.1525	0.1474	0.1241	0.1347
Coverage of 95 CI	93.3	93	91.7	93.9	92.8

Table 144: Stepwise - MAR, Beta 2, equi 0.50, n=50, p=5, Linear Missing at 50 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	0.846	0.8605	-0.0065	2.0143	0.0023
Standard Error	0.1488	0.1498	0.1513	0.1456	0.1477
True Beta	0.85	0.85	0	2	0
MSE	0.0237	0.0249	0.0124	0.0236	0.014
Coverage of 95 CI	94.1	94.5	na	93.5	na
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	0.7449	0.7672	0.1287	1.905	0.0851
Standard Error	0.3582	0.3793	0.4274	0.3333	0.4124
True Beta	0.85	0.85	0	2	0
MSE	0.144	0.1571	0.1355	0.1216	0.1228
Coverage of 95 CI	90.1	89.4	87.3	91.3	89

Table 145: Ridge - MAR, Beta 2, equi 0.50, n=50, p=5, Linear Missing at 50 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	0.8378	0.8555	-0.0064	1.9927	0.0083
Standard Error	0.151	0.1558	0.156	0.1559	0.1517
True Beta	0.85	0.85	0	2	0
MSE	0.0237	0.0257	0.0217	0.0254	0.0238
Coverage of 95 CI	94.9	94.7	96.3	95	94.3
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	0.7376	0.7556	0.1426	1.8381	0.0954
Standard Error	0.3605	0.3763	0.3723	0.3334	0.3576
True Beta	0.85	0.85	0	2	0
MSE	0.131	0.1405	0.136	0.1332	0.1226
Coverage of 95 CI	92.3	92.2	90.8	91.4	92.2

Table 146: LASSO - MAR, Beta 2, equi 0.50, n=50, p=5, Linear Missing at 50 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	0.7846	0.7982	0.0173	1.9337	0.0155
Standard Error	0.1513	0.1561	0.1555	0.1548	0.1511
True Beta	0.85	0.85	0	2	0
MSE	0.0295	0.0281	0.0115	0.0301	0.0125
Coverage of 95 CI	91.3	93.4	97.1	93.7	96.4
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	0.6762	0.6959	0.1403	1.7817	0.0906
Standard Error	0.3579	0.3726	0.3491	0.3423	0.3363
True Beta	0.85	0.85	0	2	0
MSE	0.1394	0.1476	0.1077	0.1527	0.0958
Coverage of 95 CI	89.4	89.7	93.4	88.7	94.5

Table 147: OLS - MAR, Beta 2, equi 0.50, n=50, p=5, Convex Missing at 25 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	0.8454	0.8561	-0.0019	1.9999	0.0059
Standard Error	0.1523	0.1573	0.1571	0.1572	0.152
True Beta	0.85	0.85	0	2	0
MSE	0.0229	0.0246	0.0268	0.0274	0.0224
Coverage of 95 CI	94.7	93.6	93.6	93.1	94.8
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	0.8268	0.8385	0.0412	1.9891	0.023
Standard Error	0.2018	0.2082	0.2107	0.1989	0.2034
True Beta	0.85	0.85	0	2	0
MSE	0.0508	0.0519	0.0586	0.0515	0.0541
Coverage of 95 CI	92.6	90.9	87.9	90.8	90.7

Table 148: Stepwise - MAR, Beta 2, equi 0.50, n=50, p=5, Convex Missing at 25 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	0.8461	0.8557	0.0016	1.9992	0.0051
Standard Error	0.1482	0.1497	0.1506	0.1455	0.1443
True Beta	0.85	0.85	0	2	0
MSE	0.0226	0.024	0.0173	0.0258	0.0134
Coverage of 95 CI	94.2	92.8	na	92	na
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	0.8234	0.8358	0.0348	1.989	0.0209
Standard Error	0.1924	0.1957	0.1826	0.1884	0.1759
True Beta	0.85	0.85	0	2	0
MSE	0.0533	0.0537	0.047	0.0492	0.0425
Coverage of 95 CI	89.9	88	87.7	90.1	90

Table 149: Ridge - MAR, Beta 2, equi 0.50, n=50, p=5, Convex Missing at 25 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	0.8375	0.8473	0.0068	1.9743	0.0118
Standard Error	0.1502	0.1549	0.1547	0.1548	0.1499
True Beta	0.85	0.85	0	2	0
MSE	0.0226	0.024	0.0259	0.0282	0.0219
Coverage of 95 CI	94.2	93.8	93.7	91.3	94.5
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	0.815	0.8263	0.051	1.9529	0.0304
Standard Error	0.1959	0.2017	0.204	0.1937	0.1974
True Beta	0.85	0.85	0	2	0
MSE	0.0492	0.0501	0.0568	0.0532	0.0519
Coverage of 95 CI	92.1	90.4	87.8	89.3	90.5

Table 150: LASSO - MAR, Beta 2, equi 0.50, n=50, p=5, Convex Missing at 25 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	0.7862	0.7932	0.0269	1.9197	0.0171
Standard Error	0.1505	0.1552	0.1542	0.1539	0.1495
True Beta	0.85	0.85	0	2	0
MSE	0.0281	0.0287	0.0146	0.0343	0.0123
Coverage of 95 CI	92.2	92.9	95.6	89.3	95.9
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	0.7684	0.7791	0.0591	1.9122	0.0335
Standard Error	0.2002	0.2056	0.2004	0.1985	0.195
True Beta	0.85	0.85	0	2	0
MSE	0.0558	0.0543	0.0438	0.057	0.0387
Coverage of 95 CI	90.3	90.3	91.6	90.3	93.5

Table 151: OLS - MAR, Beta 2, equi 0.50, n=50, p=5, Convex Missing at 50 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	0.8484	0.8472	7e-04	1.9998	2e-04
Standard Error	0.1541	0.1589	0.1591	0.1589	0.1541
True Beta	0.85	0.85	0	2	0
MSE	0.0233	0.0256	0.0255	0.0263	0.0232
Coverage of 95 CI	94.7	94.2	94.9	93.7	94.4
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	0.7629	0.7833	0.1235	1.8906	0.0746
Standard Error	0.4005	0.4117	0.4112	0.373	0.3991
True Beta	0.85	0.85	0	2	0
MSE	0.1453	0.1304	0.1498	0.1406	0.1292
Coverage of 95 CI	92.7	94.3	93.7	91.9	93.9

Table 152: Stepwise - MAR, Beta 2, equi 0.50, n=50, p=5, Convex Missing at 50 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	0.8476	0.8476	-0.0024	2.0002	7e-04
Standard Error	0.1498	0.1508	0.1526	0.1471	0.1479
True Beta	0.85	0.85	0	2	0
MSE	0.0226	0.025	0.0156	0.0245	0.0144
Coverage of 95 CI	94.3	93.7	na	92.9	na
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	0.7496	0.7708	0.1181	1.8895	0.0715
Standard Error	0.362	0.3622	0.4185	0.3407	0.41
True Beta	0.85	0.85	0	2	0
MSE	0.15	0.1354	0.1362	0.1399	0.1169
Coverage of 95 CI	90.2	91.5	87.5	90.3	89.1

Table 153: Ridge - MAR, Beta 2, equi 0.50, n=50, p=5, Convex Missing at 50 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	0.8408	0.8384	0.0094	1.974	0.0061
Standard Error	0.1519	0.1565	0.1566	0.1565	0.1519
True Beta	0.85	0.85	0	2	0
MSE	0.0228	0.025	0.0248	0.0269	0.0226
Coverage of 95 CI	94.5	93.9	94.9	93.3	94.4
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	0.7404	0.7578	0.1339	1.8186	0.0827
Standard Error	0.358	0.3671	0.3659	0.3426	0.3566
True Beta	0.85	0.85	0	2	0
MSE	0.134	0.1201	0.1359	0.1554	0.1154
Coverage of 95 CI	91.7	93.9	92.8	88.8	92.9

Table 154: LASSO - MAR, Beta 2, equi 0.50, n=50, p=5, Convex Missing at 50 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	0.7884	0.7851	0.0226	1.9189	0.0136
Standard Error	0.1522	0.1569	0.1563	0.1556	0.1514
True Beta	0.85	0.85	0	2	0
MSE	0.0278	0.03	0.0146	0.0323	0.0121
Coverage of 95 CI	91.5	91.5	95.9	91.2	95.9
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	0.6795	0.6985	0.1311	1.7629	0.0828
Standard Error	0.3517	0.3597	0.3398	0.3474	0.3323
True Beta	0.85	0.85	0	2	0
MSE	0.1436	0.1288	0.1092	0.178	0.0908
Coverage of 95 CI	88	89.6	93.7	87.5	94.2

Table 155: OLS - MAR, Beta 2, equi 0.50, n=50, p=5, Sinister Missing at 25 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	0.8539	0.8535	-0.0037	2.0061	-0.0137
Standard Error	0.1542	0.1586	0.1585	0.1591	0.1538
True Beta	0.85	0.85	0	2	0
MSE	0.0223	0.0248	0.0253	0.0266	0.0225
Coverage of 95 CI	95.5	94.4	93.3	92.9	95.2
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	0.8477	0.8223	0.0485	1.9948	-5e-04
Standard Error	0.2101	0.214	0.2168	0.2034	0.2114
True Beta	0.85	0.85	0	2	0
MSE	0.0559	0.069	0.0609	0.052	0.0574
Coverage of 95 CI	93	88.8	90.6	91.6	90.1

Table 156: Stepwise - MAR, Beta 2, equi 0.50, n=50, p=5, Sinister Missing at 25 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	0.8551	0.8521	-0.0018	2.0031	-0.0052
Standard Error	0.1501	0.151	0.152	0.1469	0.1451
True Beta	0.85	0.85	0	2	0
MSE	0.0223	0.0247	0.0164	0.0249	0.0127
Coverage of 95 CI	95.2	93	na	92.5	na
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	0.844	0.8188	0.0415	1.9927	5e-04
Standard Error	0.2001	0.2011	0.1946	0.1924	0.1845
True Beta	0.85	0.85	0	2	0
MSE	0.0571	0.0695	0.0485	0.0499	0.0442
Coverage of 95 CI	91.1	85.9	89.9	90.2	90

Table 157: Ridge - MAR, Beta 2, equi 0.50, n=50, p=5, Sinister Missing at 25 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	0.8459	0.8449	0.0051	1.9801	-0.0074
Standard Error	0.152	0.1561	0.156	0.1566	0.1516
True Beta	0.85	0.85	0	2	0
MSE	0.0218	0.0243	0.0246	0.027	0.0217
Coverage of 95 CI	95.6	94.6	93.8	93	95.2
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	0.8348	0.8112	0.0585	1.9566	0.008
Standard Error	0.2036	0.2069	0.2097	0.1976	0.2046
True Beta	0.85	0.85	0	2	0
MSE	0.0532	0.0659	0.0587	0.0522	0.054
Coverage of 95 CI	92.8	88.7	89.8	89.8	89.8

Table 158: LASSO - MAR, Beta 2, equi 0.50, n=50, p=5, Sinister Missing at 25 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	0.7938	0.7903	0.0227	1.9238	0.0049
Standard Error	0.1523	0.1565	0.1556	0.1556	0.1511
True Beta	0.85	0.85	0	2	0
MSE	0.0267	0.0295	0.0138	0.0325	0.0119
Coverage of 95 CI	91.1	91.9	95.7	90.2	96.6
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	0.7837	0.7621	0.0659	1.9114	0.0137
Standard Error	0.208	0.2106	0.2058	0.2019	0.2021
True Beta	0.85	0.85	0	2	0
MSE	0.0576	0.0708	0.044	0.0565	0.0392
Coverage of 95 CI	92.6	87.9	93.1	89.6	92.9

Table 159: OLS - MAR, Beta 2, equi 0.50, n=50, p=5, Sinister Missing at 50 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	0.8403	0.8643	0.0143	2.012	0.0054
Standard Error	0.1508	0.1548	0.1557	0.1567	0.1513
True Beta	0.85	0.85	0	2	0
MSE	0.0234	0.0254	0.0226	0.0216	0.0232
Coverage of 95 CI	96.5	93.1	95.2	95.9	96.6
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	0.7283	0.7503	0.1307	1.9512	0.0517
Standard Error	0.4218	0.4339	0.4342	0.4023	0.442
True Beta	0.85	0.85	0	2	0
MSE	0.1543	0.2052	0.1683	0.1494	0.1168
Coverage of 95 CI	93.1	92.5	92.5	93.7	94.5

Table 160: Stepwise - MAR, Beta 2, equi 0.50, n=50, p=5, Sinister Missing at 50 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	0.8409	0.8643	0.0016	2.0173	-0.0077
Standard Error	0.147	0.1476	0.15	0.1442	0.1418
True Beta	0.85	0.85	0	2	0
MSE	0.0224	0.0235	0.0119	0.0194	0.0109
Coverage of 95 CI	95.9	93.8	na	95.7	na
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	0.7167	0.7427	0.1209	1.9475	0.0494
Standard Error	0.376	0.3958	0.4361	0.373	0.4709
True Beta	0.85	0.85	0	2	0
MSE	0.1591	0.2053	0.1558	0.1538	0.1064
Coverage of 95 CI	87.6	85.5	87.7	92.3	91

Table 161: Ridge - MAR, Beta 2, equi 0.50, n=50, p=5, Sinister Missing at 50 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	0.8329	0.8561	0.0225	1.987	0.0112
Standard Error	0.1488	0.1526	0.1534	0.1544	0.1492
True Beta	0.85	0.85	0	2	0
MSE	0.023	0.0247	0.0221	0.0217	0.0227
Coverage of 95 CI	95.8	93.1	95.2	94.5	96.6
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	0.7082	0.7278	0.1443	1.8696	0.0608
Standard Error	0.3775	0.3851	0.3827	0.3676	0.3886
True Beta	0.85	0.85	0	2	0
MSE	0.1451	0.186	0.1548	0.1469	0.1015
Coverage of 95 CI	92.4	90.5	92.5	89.6	93.1

Table 162: LASSO - MAR, Beta 2, equi 0.50, n=50, p=5, Sinister Missing at 50 percent

Complete Data					
	X1	X2	X3	X4	X5
Estimate	0.7799	0.8011	0.033	1.9326	0.0151
Standard Error	0.1491	0.1528	0.1531	0.1536	0.1487
True Beta	0.85	0.85	0	2	0
MSE	0.0281	0.0286	0.0125	0.0264	0.0121
Coverage of 95 CI	91.6	91.7	97.4	91.6	97.3
Incomplete Data					
	X1	X2	X3	X4	X5
Estimate	0.6431	0.6784	0.1347	1.814	0.0532
Standard Error	0.3798	0.3813	0.369	0.3947	0.3674
True Beta	0.85	0.85	0	2	0
MSE	0.1625	0.1875	0.1228	0.1699	0.0805
Coverage of 95 CI	91	87.8	95.2	87.5	93.8

APPENDIX E

PSYCHOBIOLOGICAL MEASURES

The measurements collected as part of this study were selected based on the results of studies of depression in adults and a priori hypotheses of the investigators based on their knowledge of pediatric affective disorders and maturational changes during adolescence. The following sections briefly explain the action of, and results about, a subset of the measurements that were collected as part of the study, a subset of which are used in the application of variable selection methods.

One method of collection involved stimulatory tests meant to measure the body's response to stimulation by some pharmacological agent. For the purposes of this study, measurements are taken at 15 minute intervals for 30 minutes before and two to two and one half hours after infusion of the challenge agent. The mean pre-infusion and the mean and peak post-infusion levels of the hormonal response are used in the subsequent data analysis [34].

E.1 GROWTH HORMONE RESPONSE TO STIMULATORY TESTS

Growth hormone (GH), as its name implies, is involved in the growth and regeneration of body tissues. It is released by the anterior pituitary gland and acts within the body to promote protein synthesis and the breakdown of fatty cells to provide energy. Growth hormone measurements were collected by a number of different methods, including baseline hormone measurements and through stimulatory tests.

Stimulatory tests were used to measure the amount of growth hormone released in response to growth hormone releasing hormone (GHRH), and clonidine hydrochloride (CLON). GHRH is released by the hypothalamus causing the body to produce growth hormone. Clonidine acts in the brain to reduce the response of the sympathetic nervous system.

Preliminary hypotheses about GH response were that children with MDD would show less GH secretion in response to CLON and hypoglycemia than the low risk subjects and similar GH response to GHRH. Published results indicate that the MDD children had a lower response than the normal controls in all three tests with the differences in response to GHRH and hypoglycemia reaching statistical significance [34]. A second set of results with a larger sample again indicated this blunted GH response to GHRH in MDD children compared with normal controls. Additionally, the "GH response to GHRH remained low in subjects studied during clinical remission from depression [12]." When comparing normal controls and children at high-risk for depression, the blunted GH response to GHRH persisted [3].

E.2 CORTISOL AND PROLACTIN RESPONSE TO L5HTP

Cortisol is secreted by the adrenal glands in response to physical and psychological stress. Its purpose is to prepare the body to deal with stressors and to insure that the brain receives adequate energy in times of stress. Prolactin is a hormone closely related to GH. Stimulatory tests measuring cortisol response to L–5–Hydroxytryptophan (L5HTP) and corticotropin releasing hormone (CRH) were performed.

Corticotropin releasing hormone is released by the hypothalamus and stimulates the release of adrenocorticotrophic hormone (ACTH), which is released by the anterior pituitary gland and controls the secretion of cortisol. L-5-Hydroxytryptophan is an amino acid that stimulates the serotonergic system and causes the release of prolactin and cortisol. Prior to the L5HTP challenge test subjects are given oral carbidopa at intervals over the evening prior and the morning of the test. The purpose of this drug is to block the metabolism of L5HTP outside of the central nervous system allowing a lower dosage of L5HTP to be used in the challenge test itself.

After infusion with L5HTP, children with MDD when compared to normal controls had a significantly smaller cortisol response and a significantly larger prolactin response. Because of a significant gender by diagnosis interaction, analysis was performed separately for males and females and revealed that depressed females released significantly more prolactin than their control group counterparts whereas no difference was seen in the males [33]. A subsequent paper compared cortisol and prolactin response to L5HTP in children with MDD, children at high-risk for MDD and normal controls. The cortisol response was similar in the MDD and high-risk children with both groups secreting significantly less cortisol than the normal controls. The gender by diagnosis interaction was again seen in the prolactin response; MDD and high-risk girls secreted more prolactin than normal control girls with no difference seen in boys [4]. The comparisons for response to L5HTP were in terms of area-under-the-curve (AUC) and peak post-infusion measures.

E.3 CORTISOL AND ADRENOCORTICOTROPHIC HORMONE RESPONSE TO CRH

The hypothalamic-pituitary-adrenal (HPA) axis is known to be associated with adult MDD. Briefly, the HPA axis refers to a number of hormones released by the hypothalamus, the pituitary gland and the adrenal glands that work together to regulate the overall level of certain hormones in the body. Specific results in the literature suggest that dysregulations arising from the hypothalamus may be particularly important. The corticotropin-releasing hormone (CRH) stimulatory test is used to test this hypothesis. The focus of this study is on the influences of development on the HPA axis dysregulation associated with depression.

Consistent with previous studies, the results showed no significant differences in either cortisol or ACTH response to CRH in any measure considered including baseline, mean post-infusion, peak post-infusion, time to peak level and time to return to baseline level. This may indicate that the HPA axis is influenced by maturational changes that result in the dysregulation found in adults [2].

E.4 NIGHTTIME CORTISOL AND GROWTH HORMONE MEASURES

As part of their stay in the sleep laboratory, detailed in section E.5, plasma levels of cortisol and growth hormone were determined around sleep onset. The measurements were collected on the subjects' second, or baseline, night in the lab. Blood samples were collected every 20 minutes following lights out time. For the current research, summary measurements were computed including mean secretion during awake time, mean secretion during sleep, peak secretion during sleep and secretion levels in the 1 and 2 hour period before and after sleep onset.

Results published on this data used the following summary measures: area under the curve (AUC) in the 4 hours after sleep onset, AUC over the total sleep period and the peak hormonal concentrations during sleep. For the cortisol measurements, it was shown that the depressed sample had lower cortisol than normal controls in the 4 hours post sleep onset, while no difference was seen in the other measures. No significant group differences were seen in the growth hormone measurements, although within the depressed, girls secreted less growth hormone than boys.

E.5 SLEEP MEASURES

The motivation to collect electroencephalographic (EEG) sleep measures was the apparent contradictions between the results seen in adults and those seen in children and adolescents. The results of the adult studies include "decreased delta sleep, reduced rapid eye movement (REM) latency, increased sleep continuity disturbances, and accelerated accumulation of REM sleep across the night [13]." The data were collected over the two nights the subject spent in the sleep laboratory. Subjects kept a sleep diary during the week preceding their time in the sleep lab that was used to collect subjective sleep data and to determine the bedtime and wake-up times typical for the subject that were then replicated in the lab.

"Major dependent variables were defined as follows: Sleep latency was the time from lights out to sleep onset. Sleep onset was the first 10 minute stage 2 (or deeper) sleep with less than 1 minute of intervening awake time. REM period latency was the interval from sleep onset to the first REM period lasting 3 minutes. ... REM activity was an integrated estimate of eye movement frequency during each minute of REM, score on a 0 to 8 scale. Sleep maintenance was the percentage of time spent asleep from sleep onset to wake-up time [13]."

The initial analysis of the sleep measures in 1991 were concordant with the existing literature in that they resulted in no significant differences between children with MDD and normal control children. Based on these results, it is hypothesized that EEG sleep measures in children and adolescents are not affected by depression, but that sleep disturbances may increase with age thereby affecting adults more significantly than children [13].