INFERENCE AFTER VARIABLE SELECTION

by

Lasanthi C. R. Pelawa Watagoda

M.S., Southern Illinois University Carbondale, 2013

A Dissertation
Submitted in Partial Fulfillment of the Requirements for the
Doctor of Philosophy Degree

Department of Mathematics
in the Graduate School
Southern Illinois University Carbondale
August 2017

DISSERTATION APPROVAL

INFERENCE AFTER VARIABLE SELECTION

By

Lasanthi C. R. Pelawa Watagoda

A Dissertation Submitted in Partial

Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

in the field of Mathematics

Approved by:

David Olive, Chair

Bhaskar Bhattacharya

Kathy Pericak-Spector

Randy Hughes

Yanyan Sheng

Graduate School
Southern Illinois University Carbondale
DATE OF DEFENCE

AN ABSTRACT OF THE DISSERTATION OF

Lasanthi C. R. Pelawa Watagoda, for the Doctor of Philosophy degree in Mathematics, presented on TBA, at Southern Illinois University Carbondale.

TITLE: Inference after variable selection

MAJOR PROFESSOR: Prof. D. J. Olive

This thesis presents inference for the multiple linear regression model $Y = \beta_1 x_1 + \cdots + \beta_p x_p + e$ after model or variable selection, including prediction intervals for a future value of the response variable Y_f, and testing hypotheses with the bootstrap. If n is the sample size, most results are for n/p large, but prediction intervals are developed that may increase in average length slowly as p increases for fixed n if the model is sparse: k predictors have nonzero coefficients β_i where n/k is large.

KEY WORDS: Bootstrap; Forward Selection; Lasso; Partial Least Squares; Prediction Interval; Principal Components Regression; Relaxed Lasso; Ridge Regression.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Suppose that the response variable Y_i and at least one predictor variable $x_{i,j}$ are quantitative with $x_{i,1} \equiv 1$. Let $\boldsymbol{x}_i^T = (x_{i,1}, ..., x_{i,p}) = (1 \quad \boldsymbol{u}_i^T)$ and $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)^T$ where β_1 corresponds to the intercept. Then the multiple linear regression (MLR) model is

$$Y_i = \beta_1 + x_{i,2}\beta_2 + \cdots + x_{i,p}\beta_p + e_i = \boldsymbol{x}_i^T\boldsymbol{\beta} + e_i \tag{1.1}$$

for $i = 1, ..., n$. This model is also called the full model. Here n is the sample size and the random variable e_i is the ith error. In matrix notation, these n equations become

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}, \tag{1.2}$$

where $\boldsymbol{Y}$ is an $n \times 1$ vector of dependent variables, $\boldsymbol{X}$ is an $n \times p$ matrix of predictors, $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown coefficients, and $\boldsymbol{e}$ is an $n \times 1$ vector of unknown errors. The ith fitted value $\hat{Y}_i = \boldsymbol{x}_i^T\hat{\boldsymbol{\beta}}$ and the ith residual $r_i = Y_i - \hat{Y}_i$ where $\hat{\boldsymbol{\beta}}$ is an estimator of $\boldsymbol{\beta}$. Ordinary least squares (OLS) is often used for inference if n/p is large.

It is often convenient to use the centered response $\boldsymbol{Z} = \boldsymbol{Y} - \overline{\boldsymbol{Y}}$ where $\overline{\boldsymbol{Y}} = \overline{Y}\boldsymbol{1}$, and the $n \times (p-1)$ matrix of standardized nontrivial predictors $\boldsymbol{W} = (W_{ij})$. For $j = 1, ..., p-1$, let W_{ij} denote the $(j+1)$th variable standardized so that $\sum_{i=1}^n W_{ij} = 0$ and $\sum_{i=1}^n W_{ij}^2 = n$. Hence

$$W_{ij} = \frac{x_{i,j+1} - \overline{x}_{j+1}}{\tilde{\sigma}_{j+1}} \quad \text{where} \quad \tilde{\sigma}^2_{\mathrm{j+1}} = \frac{1}{\mathrm{n}}\sum_{\mathrm{i=1}}^{\mathrm{n}}(\mathrm{x_{i,j+1}} - \overline{\mathrm{x}}_{\mathrm{j+1}})^2.$$

Note that the sample correlation matrix of the nontrivial predictors $\boldsymbol{u}_i$ is

$$\boldsymbol{R}_{\boldsymbol{u}} = \frac{\boldsymbol{W}^T\boldsymbol{W}}{n}.$$

Then regression through the origin is used for the model

$$\boldsymbol{Z} = \boldsymbol{W}\boldsymbol{\eta} + \boldsymbol{e} \tag{1.3}$$

where the vector of fitted values $\hat{\boldsymbol{Y}} = \overline{\boldsymbol{Y}} + \hat{\boldsymbol{Z}}$.

There are many alternative methods for estimating $\boldsymbol{\beta}$, including forward selection with OLS, principal component regression (PCR), partial least squares (PLS) due to Wold (1975), lasso due to Tibshirani (1996), relaxed lasso due to Meinshausen (2007), and ridge regression (RR): see Hoerl and Kennard (1970). These six methods produce M models and use a criterion to select the final model (e.g. C_p or 10-fold cross validation (CV)). The number of models M depends on the method. The full model is (approximately) fit with OLS. For one of the M models, some of the methods use $\hat{\boldsymbol{\eta}} = \boldsymbol{0}$ and fit the model $Y_i = \beta_1 + e_i$ with $\hat{Y} = \overline{Y}$. Lasso and ridge regression have a parameter λ. When $\lambda = 0$, the full OLS model is used. These methods also use a maximum value λ_M of λ and a grid of M λ values $0 \leq \lambda_1 < \lambda_2 < \cdots < \lambda_{M-1} < \lambda_M$ where often $\lambda_1 = 0$. For lasso, λ_M is the smallest value of λ such that $\hat{\boldsymbol{\eta}}_{\lambda_M} = \boldsymbol{0}$. Hence $\hat{\boldsymbol{\eta}}_{\lambda_i} \neq \boldsymbol{0}$ for $i < M$. For forward selection, PCR, and PLS, $M \leq p$. See James, Witten, Hastie, and Tibshirani (2013, ch. 6) for more details about these six methods.

1.1 INFERENCE FOR RIDGE REGRESSION AND LASSO

Consider choosing $\hat{\boldsymbol{\eta}}$ to minimize the criterion

$$Q(\boldsymbol{\eta}) = \frac{1}{a}(\boldsymbol{Z} - \boldsymbol{W}\boldsymbol{\eta})^T(\boldsymbol{Z} - \boldsymbol{W}\boldsymbol{\eta}) + \frac{\lambda_{1,n}}{a}\sum_{i=1}^{p-1}|\eta_i|^j \tag{1.4}$$

where $\lambda_{1,n} \geq 0$, $a > 0$, and $j > 0$ are known constants. Then $j = 2$ corresponds to ridge regression, $j = 1$ corresponds to lasso, and $a = 1, 2, n$, and $2n$ are common. The residual sum of squares $RSS(\boldsymbol{\eta}) = (\boldsymbol{Z} - \boldsymbol{W}\boldsymbol{\eta})^T(\boldsymbol{Z} - \boldsymbol{W}\boldsymbol{\eta})$, and $\lambda_{1,n} = 0$ corresponds to the OLS estimator $\hat{\boldsymbol{\eta}}_{OLS} = (\boldsymbol{W}^T\boldsymbol{W})^{-1}\boldsymbol{W}^T\boldsymbol{Z}$.

In the following three paragraphs, assume p is fixed. Knight and Fu (2000) prove i) that $\hat{\boldsymbol{\eta}}$ is a consistent estimator of $\boldsymbol{\eta}$ if $\lambda_{1,n} = o(n)$ so $\lambda_{1,n}/n \to 0$ as $n \to \infty$, ii) $\hat{\boldsymbol{\eta}}_{OLS}$ and $\hat{\boldsymbol{\eta}}$ are asymptotically equivalent if $\lambda_{1,n} \to \infty$ too slowly as $n \to \infty$, iii) $\hat{\boldsymbol{\eta}}$ is a $\sqrt{n}$ consistent estimator of $\boldsymbol{\eta}$ if $\lambda_{1,n} = O(\sqrt{n})$ (so $\lambda_{1,n}/\sqrt{n}$ is bounded), and iv) if $\lambda_{1,n}/\sqrt{n} \to \tau \geq 0$, then

$$\sqrt{n}(\hat{\boldsymbol{\eta}}_{RR} - \boldsymbol{\eta}) \xrightarrow{D} N_{p-1}(-\tau \boldsymbol{V}\boldsymbol{\eta}, \sigma^2 \boldsymbol{V})$$

where

$$\boldsymbol{R_u} = \frac{\boldsymbol{W}^T\boldsymbol{W}}{n} \xrightarrow{P} \boldsymbol{V}^{-1} \tag{1.5}$$

as $n \to \infty$. If $\tau = 0$, then OLS and ridge regression have the same limiting distribution. Note that $\boldsymbol{V}^{-1} = \boldsymbol{\rho_u}$ if the $\boldsymbol{u}_i$ are a random sample from a population with a nonsingular population correlation matrix $\boldsymbol{\rho_u}$. Under (1.5), if $\lambda_{1,n}/n \to 0$ then

$$\frac{\boldsymbol{W}^T\boldsymbol{W} + \lambda_{1,n}\boldsymbol{I}_{p-1}}{n} \xrightarrow{P} \boldsymbol{V}^{-1}, \quad \text{and} \quad \mathrm{n}(\boldsymbol{W}^{\mathrm{T}}\boldsymbol{W} + \lambda_{1,\mathrm{n}}\boldsymbol{I}_{\mathrm{p}-1})^{-1} \xrightarrow{\mathrm{P}} \boldsymbol{V}.$$

The following identity from Gunst and Mason (1980, p. 342) is useful for ridge regression inference:

$$\hat{\boldsymbol{\eta}}_{RR} = (\boldsymbol{W}^T\boldsymbol{W} + \lambda_{1,n}\boldsymbol{I}_{p-1})^{-1}\boldsymbol{W}^T\boldsymbol{Z} = (\boldsymbol{W}^T\boldsymbol{W} + \lambda_{1,n}\boldsymbol{I}_{p-1})^{-1}\boldsymbol{W}^T\boldsymbol{W}\hat{\boldsymbol{\eta}}_{OLS} = \boldsymbol{A}_n\hat{\boldsymbol{\eta}}_{OLS}$$

$$= [\boldsymbol{I}_{p-1} - \lambda_{1,n}(\boldsymbol{W}^T\boldsymbol{W} + \lambda_{1,n}\boldsymbol{I}_{p-1})^{-1}]\hat{\boldsymbol{\eta}}_{OLS} = \boldsymbol{B}_n\hat{\boldsymbol{\eta}}_{OLS}$$

since $\boldsymbol{A}_n - \boldsymbol{B}_n = \boldsymbol{0}$. If $\lambda_{1,n}/\sqrt{n} \to \tau \geq 0$, then

$$\sqrt{n}(\hat{\boldsymbol{\eta}}_{RR} - \boldsymbol{\eta}) = \sqrt{n}(\hat{\boldsymbol{\eta}}_{RR} - \hat{\boldsymbol{\eta}}_{OLS} + \hat{\boldsymbol{\eta}}_{OLS} - \boldsymbol{\eta}) =$$

$$\sqrt{n}(\hat{\boldsymbol{\eta}}_{OLS} - \boldsymbol{\eta}) - \sqrt{n}\frac{\lambda_{1,n}}{n}n(\boldsymbol{W}^T\boldsymbol{W} + \lambda_{1,n}\boldsymbol{I}_{p-1})^{-1}\hat{\boldsymbol{\eta}}_{OLS}$$

$$\xrightarrow{D} N_{p-1}(\boldsymbol{0}, \sigma^2\boldsymbol{V}) - \tau\boldsymbol{V}\boldsymbol{\eta} \sim N_{p-1}(-\tau\boldsymbol{V}\boldsymbol{\eta}, \sigma^2\boldsymbol{V}).$$

Theorem 1.1. *Let $\hat{\boldsymbol{\eta}}_L$ be the lasso estimator. Then*

$$\sqrt{n}(\hat{\boldsymbol{\eta}}_L - \boldsymbol{\eta}) \sim N_{p-1}\left(\frac{-\tau}{2}\boldsymbol{V}\boldsymbol{s}, \sigma^2\boldsymbol{V}\right).$$

where

$$\boldsymbol{R_u} = \frac{\boldsymbol{W}^T\boldsymbol{W}}{n} \xrightarrow{P} \boldsymbol{V}^{-1}$$

as $n \to \infty$. If $\tau = 0$, then OLS and lasso regression have the same limiting distribution.

Proof. The following identity from Efron and Hastie (2016, p. 308), for example, is useful for inference for the lasso estimator $\hat{\boldsymbol{\eta}}_L$:

$$\frac{-1}{n}\boldsymbol{W}^T(\boldsymbol{Z}-\boldsymbol{W}\hat{\boldsymbol{\eta}}_L)+\frac{\lambda_{1,n}}{2n}\boldsymbol{s}_n=\boldsymbol{0} \quad \text{or} \quad -\boldsymbol{W}^{\mathrm{T}}(\boldsymbol{Z}-\boldsymbol{W}\hat{\boldsymbol{\eta}}_{\mathrm{L}})+\frac{\lambda_{1,\mathrm{n}}}{2}\boldsymbol{s}_{\mathrm{n}}=\boldsymbol{0}$$

where $s_{in}\in[-1,1]$ and $s_{in}=\text{sign}(\hat{\eta}_{i,L})$ if $\hat{\eta}_{i,L}\neq 0$. Here $\text{sign}(\eta_i)=1$ if $\eta_i>1$ and $\text{sign}(\eta_i)=-1$ if $\eta_i<1$. Note that $\boldsymbol{s}_n=\boldsymbol{s}_{n,\hat{\boldsymbol{\eta}}_L}$ depends on $\hat{\boldsymbol{\eta}}_L$. Thus

$$\hat{\boldsymbol{\eta}}_L=(\boldsymbol{W}^T\boldsymbol{W})^{-1}\boldsymbol{W}^T\boldsymbol{Z}-n(\boldsymbol{W}^T\boldsymbol{W})^{-1}\ \frac{\lambda_{1,n}}{2n}\ \boldsymbol{s}_n.$$

If $\lambda_{1,n}/\sqrt{n}\to\tau\geq 0$ and $\boldsymbol{s}_n\xrightarrow{P}\boldsymbol{s}=\boldsymbol{s}_{\boldsymbol{\eta}}$, then

$$\sqrt{n}(\hat{\boldsymbol{\eta}}_L-\boldsymbol{\eta})=\sqrt{n}(\hat{\boldsymbol{\eta}}_L-\hat{\boldsymbol{\eta}}_{OLS}+\hat{\boldsymbol{\eta}}_{OLS}-\boldsymbol{\eta})=$$

$$\sqrt{n}(\hat{\boldsymbol{\eta}}_{OLS}-\boldsymbol{\eta})-\sqrt{n}\frac{\lambda_{1,n}}{2n}n(\boldsymbol{W}^T\boldsymbol{W})^{-1}\boldsymbol{s}_n\xrightarrow{D}N_{p-1}(\boldsymbol{0},\sigma^2\boldsymbol{V})-\frac{\tau}{2}\boldsymbol{V}\boldsymbol{s}\sim N_{p-1}\left(\frac{-\tau}{2}\boldsymbol{V}\boldsymbol{s},\sigma^2\boldsymbol{V}\right).$$

□

If none of the elements of $\boldsymbol{\eta}$ are zero, and if $\hat{\boldsymbol{\eta}}_L$ is a consistent estimator of $\boldsymbol{\eta}$, then $\boldsymbol{s}_n\xrightarrow{P}\boldsymbol{s}=\boldsymbol{s}_{\boldsymbol{\eta}}$. If $\lambda_{1,n}/\sqrt{n}\to 0$, then OLS and lasso are asymptotically equivalent even if $\boldsymbol{s}_n$ does not converge to a vector $\boldsymbol{s}$ as $n\to\infty$ since $\boldsymbol{s}_n$ is bounded.

The results in the above three paragraphs hold after model selection if $\lambda_{1,n}$ is replaced by $\hat{\lambda}_{1,n}$ and o and O are replaced by o_P and O_P, e.g. $\hat{\lambda}_{1,n}=o_P(\sqrt{n})$ makes lasso or ridge regression asymptotically equivalent to OLS. For model selection, the M values of λ are denoted by $\lambda_1,\lambda_2,...,\lambda_M$ where $\lambda_i=\lambda_{1,n,i}$ depends on n for $i=1,...,M$. If λ_s corresponds to the model selected, then $\hat{\lambda}_{1,n}=\lambda_s$.

CHAPTER 2

VARIABLE SELECTION

Variable selection, also called subset or model selection, is the search for a subset of predictor variables that can be deleted without important loss of information. Following Olive and Hawkins (2005), a *model for variable selection* can be described by

$$\boldsymbol{x}^T\boldsymbol{\beta} = \boldsymbol{x}_S^T\boldsymbol{\beta}_S + \boldsymbol{x}_E^T\boldsymbol{\beta}_E = \boldsymbol{x}_S^T\boldsymbol{\beta}_S \tag{2.1}$$

where $\boldsymbol{x} = (\boldsymbol{x}_S^T, \boldsymbol{x}_E^T)^T$, $\boldsymbol{x}_S$ is a $k_S \times 1$ vector and $\boldsymbol{x}_E$ is a $(p - k_S - 1) \times 1$ vector. Given that $\boldsymbol{x}_S$ is in the model, $\boldsymbol{\beta}_E = \boldsymbol{0}$ and E denotes the subset of terms that can be eliminated given that the subset S is in the model. Let $\boldsymbol{x}_I$ be the vector of k terms from a candidate subset indexed by I, and let $\boldsymbol{x}_O$ be the vector of the remaining predictors (out of the candidate submodel). Suppose that S is a subset of I and that model (2.1) holds. Then

$$\boldsymbol{x}^T\boldsymbol{\beta} = \boldsymbol{x}_S^T\boldsymbol{\beta}_S = \boldsymbol{x}_S^T\boldsymbol{\beta}_S + \boldsymbol{x}_{I/S}^T\boldsymbol{\beta}_{(I/S)} + \boldsymbol{x}_O^T\boldsymbol{0} = \boldsymbol{x}_I^T\boldsymbol{\beta}_I \tag{2.2}$$

where $\boldsymbol{x}_{I/S}$ denotes the predictors in I that are not in S. Since this is true regardless of the values of the predictors, $\boldsymbol{\beta}_O = \boldsymbol{0}$ if $S \subseteq I$.

Forward selection forms a sequence of submodels $I_1, ..., I_M$ where I_j uses j predictors including the constant. Let I_1 use $x_1^* = x_1 \equiv 1$: the model has a constant but no nontrivial predictors. To form I_2, consider all models I with two predictors including x_1^*. Compute $Q_2(I) = SSE(I) = RSS(I) = \boldsymbol{r}^T(I)\boldsymbol{r}(I) = \sum_{i=1}^n r_i^2(I) = \sum_{i=1}^n (Y_i - \hat{Y}_i(I))^2$. Let I_2 minimize $Q_2(I)$ for the $p-1$ models I that contain x_1^* and one other predictor. Denote the predictors in I_2 by x_1^*, x_2^*. In general, to form I_j consider all models I with j predictors including variables $x_1^*, ..., x_{j-1}^*$. Compute $Q_j(I) = \boldsymbol{r}^T(I)\boldsymbol{r}(I) = \sum_{i=1}^n r_i^2(I) = \sum_{i=1}^n (Y_i - \hat{Y}_i(I))^2$. Let I_j minimize $Q_j(I)$ for the $p - j + 1$ models I that contain $x_1^*, ..., x_{j-1}^*$ and one other predictor not already selected. Denote the predictors in I_j by $x_1^*, ..., x_j^*$. Continue in this manner for $j = 2, ..., M$. Often $M = \min(\lceil n/J \rceil, p)$ for some integer J such as $J = 5, 10$, or 20. Here $\lceil x \rceil$ is the smallest integer $\geq x$, e.g., $\lceil 7.7 \rceil = 8$.

2.1 CRITERIA FOR FINAL MODEL SELECTION

When there is a sequence of M submodels, the final submodel I_d needs to be selected. Let $\boldsymbol{x}_I$ and $\hat{\boldsymbol{\beta}}_I$ be $a \times 1$. Hence the candidate model contains a terms, including a constant. Suppose the e_i are independent and identically distributed (iid) with variance $V(e_i) = \sigma^2$. Then there are many criteria used to select the final submodel I_d. A simple method is to take the model that uses $d = M = \min(\lceil n/J \rceil, p)$ variables. If p is fixed, the method will use the full OLS model once $n/J \geq p$. For a given data set, p, n and $\hat{\sigma}^2$ act as constants, and a criterion below may add a constant or be divided by a constant without changing the subset I_{min} that minimizes the criterion.

2.1.1 Criterion when there is a good estimator for σ^2

Let criteria $C_S(I)$ have the form

$$C_S(I) = SSE(I) + aK_n\hat{\sigma}^2.$$

These criteria need a good estimator of σ^2. The criterion $C_p(I) = AIC_S(I)$ uses $K_n = 2$ while the $BIC_S(I)$ criterion uses $K_n = \log(n)$. Typically $\hat{\sigma}^2$ is the full OLS model

$$MSE = \sum_{i=1}^{n} \frac{r_i^2}{n-p}$$

when n/p is large. Then $\hat{\sigma}^2 = MSE$ is a $\sqrt{n}$ consistent estimator of σ^2 under mild conditions by Su and Cook (2012).

2.1.2 $AIC(I)$ and $BIC(I)$

It is hard to get a good estimator of σ^2 when n/p is not large. The following criterion are described in Burnham and Anderson (2004), but still need n/p large. AIC is due to Akaike (1973) and BIC to Schwarz (1978).

$$AIC(I) = n \log\left(\frac{SSE(I)}{n}\right) + 2a, \quad \text{and}$$

$$BIC(I) = n \log\left(\frac{SSE(I)}{n}\right) + a \log(n).$$

Let I_{min} be the submodel that minimizes the criterion. Following Seber and Lee (2003, p. 448) and Nishi (1984), the probability that model I_{min} from C_p or AIC underfits goes to zero as $n \to \infty$. If $\hat{\boldsymbol{\beta}}_I$ is $a \times 1$, form the $p \times 1$ vector $\hat{\boldsymbol{\beta}}_{I,0}$ from $\hat{\boldsymbol{\beta}}_I$ by adding 0s corresponding to the omitted variables. Since there are a finite number of regression models I that contain the true model, and each such model gives a $\sqrt{n}$ consistent estimator $\hat{\boldsymbol{\beta}}_{I,0}$ of $\boldsymbol{\beta}$, the probability that I_{min} picks one of these models goes to one as $n \to \infty$. Hence $\hat{\boldsymbol{\beta}}_{I_{min},0}$ is a $\sqrt{n}$ consistent estimator of $\boldsymbol{\beta}$ under model (2.1). Olive (2017b: ∮ 5.3.4, 2017c ∮ 3.4.1) showed that $\hat{\boldsymbol{\beta}}_{I_{min},0}$ is a consistent estimator.

2.1.3 *EBIC*

The EBIC criterion given in Luo and Chen (2012) may work when n/p is not large. Let $0 \leq \gamma \leq 1$ and $|I| = a \leq \min(n,p)$ if $\hat{\boldsymbol{\beta}}_I$ is $a \times 1$. We may use $a \leq \min(n/5, p)$. Then

$$EBIC(I) = n \log\left(\frac{SSE(I)}{n}\right) + a \log(n) + 2\gamma \log\left[\binom{p}{a}\right] = BIC(I) + 2\gamma \log\left[\binom{p}{a}\right].$$

This criterion can give good results if $p = p_n = O(n^k)$ and $\gamma > 1 - 1/(2k)$. Hence we will use $\gamma = 1$.

The above criteria can be applied to forward selection and relaxed lasso. The C_p criterion can also be applied to lasso. See, for example, Efron and Hastie (2016, pp. 221, 231).

2.2 R FUNCTIONS FOR VARIABLE SELECTION

Many methods for variable selection have been suggested. We will consider several R functions including i) forward selection with the minimum C_p criterion as computed with `regsubsets` function from the `leaps` library. The remaining methods often use 10 fold cross validation (CV) and include ii) principal components regression (PCR) with the `pcr` function from the `pls` library, iii) partial least squares (PLS) with the `plsr` function from the `pls` library, iv) ridge regression with the `cv.glmnet` function from the `glmnet` library, and v) lasso with the `cv.glmnet` function from the `glmnet` library.

2.3 OLS SUBMODEL THEOREM

Theorem 2.1.

Suppose the usual linear model $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{e}$ *, with* $E\boldsymbol{Y} = \boldsymbol{X\beta}$ *and* $E(\boldsymbol{e}) = \boldsymbol{0}$*.*

Then $Cov(\boldsymbol{Y}) = Cov(\boldsymbol{e}) = \sigma^2\boldsymbol{I}$*. If we break down* $\boldsymbol{X}$ *and* $\boldsymbol{\beta}$ *as follows;*

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_I & \boldsymbol{X}_0 \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_I \\ \\ \boldsymbol{\beta}_0 \end{bmatrix},$$

$\boldsymbol{X\beta} = \boldsymbol{X}_I\boldsymbol{\beta}_I + \boldsymbol{X}_0\boldsymbol{\beta}_0$ *where* $\boldsymbol{\beta}_I = \left[\boldsymbol{X}_I^T\boldsymbol{X}_I\right]^{-1}\boldsymbol{X}_I^T\boldsymbol{Y} = \boldsymbol{AY}$*, then*

$E(\hat{\boldsymbol{\beta}}_I) = \boldsymbol{\beta}_I + \left[\boldsymbol{X}_I^T\boldsymbol{X}_I\right]^{-1}\boldsymbol{X}_I^T\boldsymbol{X}_0\boldsymbol{\beta}_0$ *and* $Cov(\hat{\boldsymbol{\beta}}_I) = \sigma^2\left(\boldsymbol{X}_I^T\boldsymbol{X}_I\right)^{-1}$*.*

Proof. Assume this is an arbitrary submodel. When the submodel contains the set of predictors $\boldsymbol{S}$ then $\hat{\boldsymbol{\beta}}_I$ works well and estimates $\boldsymbol{\beta}_I$, but if I does not contain S then

$E(\hat{\boldsymbol{\beta}}_I) = E\left(\left[\boldsymbol{X}_I^T\boldsymbol{X}_I\right]^{-1}\boldsymbol{X}_I^T\boldsymbol{Y}\right)$

$= E\left(\boldsymbol{AY}\right) = \boldsymbol{A}E\left(\boldsymbol{Y}\right) = \boldsymbol{AX\beta}$

$= \boldsymbol{A}\left(\boldsymbol{X}_I\boldsymbol{\beta}_I + \boldsymbol{X}_0\boldsymbol{\beta}_0\right) = \left[\boldsymbol{X}_I^T\boldsymbol{X}_I\right]^{-1}\boldsymbol{X}_I^T\left(\boldsymbol{X}_I\boldsymbol{\beta}_I + \boldsymbol{X}_0\boldsymbol{\beta}_0\right) = \left[\boldsymbol{X}_I^T\boldsymbol{X}_I\right]^{-1}\boldsymbol{X}_I^T\boldsymbol{X}_I\boldsymbol{\beta}_I + \left[\boldsymbol{X}_I^T\boldsymbol{X}_I\right]^{-1}\boldsymbol{X}_I^T\boldsymbol{X}_0\boldsymbol{\beta}_0 = \boldsymbol{\beta}_I + \left[\boldsymbol{X}_I^T\boldsymbol{X}_I\right]^{-1}\boldsymbol{X}_I^T\boldsymbol{X}_0\boldsymbol{\beta}_0$.

When $\boldsymbol{\beta}_0 = 0$ then $E(\hat{\boldsymbol{\beta}}_I)$ is equal to $\hat{\boldsymbol{\beta}}_I$.

Now consider the $Cov(\hat{\boldsymbol{\beta}}_I)$:

$Cov(\hat{\boldsymbol{\beta}}_I) = Cov(\boldsymbol{AY}) = \boldsymbol{A}Cov(\boldsymbol{Y})\boldsymbol{A}^T = \boldsymbol{A}\sigma^2\boldsymbol{I}\boldsymbol{A}^T = \sigma^2\boldsymbol{AA}^T = \sigma^2\left(\left[\boldsymbol{X}_I^T\boldsymbol{X}_I\right]^{-1}\boldsymbol{X}_I^T\right)\left(\left[\boldsymbol{X}_I^T\boldsymbol{X}_I\right]^{-1}\boldsymbol{X}_I^T\right)^T = \sigma^2\left[\boldsymbol{X}_I^T\boldsymbol{X}_I\right]^{-1}\boldsymbol{X}_I^T\boldsymbol{X}_I\left(\left[\boldsymbol{X}_I^T\boldsymbol{X}_I\right]^{-1}\right)^T = \sigma^2\left(\left[\boldsymbol{X}_I^T\boldsymbol{X}_I\right]^T\right)^{-1} = \sigma^2\left(\boldsymbol{X}_I^T\boldsymbol{X}_I\right)^{-1}$.

$\square$

The above results shows why OLS does not work well if the submodel does not contains enough predictors.

CHAPTER 3

PREDICTION INTERVALS

Consider predicting a future test response variable Y_f given a $p \times 1$ vector of predictors $\boldsymbol{x}_f$ and training data $(\boldsymbol{x}_1, Y_1), ..., (\boldsymbol{x}_n, Y_n)$. A large sample $100(1-\delta)\%$ prediction interval (PI) has the form $[\hat{L}_n, \hat{U}_n]$ where $P[\hat{L}_n \leq Y_f \leq \hat{U}_n] \to 1 - \delta$ as the sample size $n \to \infty$.

3.1 SHORTH PI

The shorth(c) estimator is useful for making prediction intervals. Let $Z_{(1)}, ..., Z_{(n)}$ be the order statistics of $Z_1, ..., Z_n$. Then let the shortest closed interval containing at least c of the Z_i be

$$\text{shorth(c)} = [\text{Z}_{(\text{s})}, \text{Z}_{(\text{s+c}-1)}]. \tag{3.1}$$

Let

$$k_n = \lceil n(1-\delta) \rceil \tag{3.2}$$

where $\lceil x \rceil$ is the smallest integer $\geq x$, e.g., $\lceil 7.7 \rceil = 8$. Frey (2013) showed that for large $n\delta$ and iid data, the shorth(k_n) PI has maximum undercoverage $\approx 1.12\sqrt{\delta/n}$, and used the shorth($c$) estimator as the large sample $100(1-\delta)\%$ PI where

$$c = \min(n, \lceil n[1 - \delta + 1.12\sqrt{\delta/n}\ ]\ \rceil). \tag{3.3}$$

A problem with the prediction intervals that cover $\approx 100(1-\delta)\%$ of the training data cases Y_i (such as (3.1) using $c = k_n$ given by (3.2)), is that they have coverage lower than the nominal coverage of $1-\delta$ for moderate n. This result is not surprising since empirically statistical methods perform worse on test data. Increasing c will improve the coverage for moderate samples. Let df be the model degrees of freedom. Then empirically for many models, for $n \approx 20df$, prediction intervals such as (3.1) applied to iid data or pseudodata using $c = k_n$ tend to have undercoverage as high as 5%. The undercoverage decreases rapidly as n increases. Let $q_n = \min(1 - \delta + 0.05, 1 - \delta + p/n)$ for $\delta > 0.1$ and

$$q_n = \min(1 - \delta/2, 1 - \delta + 10\delta p/n), \quad \text{otherwise.} \tag{3.4}$$

If $1 - \delta < 0.999$ and $q_n < 1 - \delta + 0.001$, set $q_n = 1 - \delta$. Using

$$c = \lceil n q_n \rceil \tag{3.5}$$

decreased the undercoverage. For $p = 1$ and $n \geq 20$, the correction factors c/n for c given by (3.3) and (3.5) do not differ by much more than 3% for $0.01 \leq \delta \leq 0.5$.

3.2 OLIVE (2013) PI

Olive (2013) developed prediction intervals for models of the form $Y_i = m(\boldsymbol{x}_i) + e_i$, and variable selection models for (1.1) have this form, as noted by Olive (2017a). Let c be given by (3.5), and let

$$b_n = \left(1 + \frac{15}{n}\right) \sqrt{\frac{n + 2p}{n - p}}. \tag{3.6}$$

Compute the shorth(c) of the residuals $= (r_{(d)}, r_{(d+c-1)}) = (\tilde{\xi}_{\delta_1}, \tilde{\xi}_{1-\delta_2})$. Then a 100 $(1-\delta)$% large sample PI for Y_f is

$$[\hat{m}(\boldsymbol{x}_f) + b_n \tilde{\xi}_{\delta_1}, \hat{m}(\boldsymbol{x}_f) + b_n \tilde{\xi}_{1-\delta_2}], \tag{3.7}$$

3.3 TWO NEW PREDICTION INTERVALS

3.3.1 The First new prediction Interval

Results from Hastie, Tibshirani, and Wainwright (2015, pp. 20, 296, ch. 6, ch. 11) suggest that lasso can perform well for sparse models: the subset S in (2.2) contains $k_S = a_S$ predictors where $a_S/n \to 0$ as $n \to \infty$. Let d be a crude estimate of the model degrees of freedom. With the exception of ridge regression, d is the number of "variables" used by the method. Forward selection, lasso, and relaxed lasso use variables $x_1^*, ..., x_d^*$ while PCR and PLS use variables that are linear combinations of the predictors $V_j = \boldsymbol{\gamma}_j^T \boldsymbol{x}$ for $j = 1, ..., d$.

See Efron and Hastie (2016, pp. 221, 222, 231) and Tibshirani (2015) for lasso degrees of freedom.

For n/p large, Olive (2013) developed prediction intervals for models of the form $Y_i = m(\boldsymbol{x}_i) + e_i$, and variable selection models for (1.1) have this form, as noted by Olive (2017a). The first new PI, that can be useful even if n/p is not large, is defined below. The PI is similar to the Olive (2013) PI with p replaced by d, if d is not too large.

Let $q_n = \min(1 - \delta + 0.05, 1 - \delta + d/n)$ for $\delta > 0.1$ and

$$q_n = \min(1 - \delta/2, 1 - \delta + 10\delta d/n), \quad \text{otherwise.} \tag{3.8}$$

If $1 - \delta < 0.999$ and $q_n < 1 - \delta + 0.001$, set $q_n = 1 - \delta$. Let

$$c = \lceil n q_n \rceil, \tag{3.9}$$

and let

$$b_n = \left(1 + \frac{15}{n}\right) \sqrt{\frac{n + 2d}{n - d}} \tag{3.10}$$

if $d \leq 8n/9$, and

$$b_n = 5 \left(1 + \frac{15}{n}\right),$$

otherwise. Compute the shorth(c) of the residuals $= [r_{(s)}, r_{(s+c-1)}] = [\tilde{\xi}_{\delta_1}, \tilde{\xi}_{1-\delta_2}]$. Then the first new 100 $(1 - \delta)$% large sample PI for Y_f is

$$[\hat{m}(\boldsymbol{x}_f) + b_n \tilde{\xi}_{\delta_1}, \hat{m}(\boldsymbol{x}_f) + b_n \tilde{\xi}_{1-\delta_2}]. \tag{3.11}$$

3.3.2 The Second new prediction Interval (Validation PI)

The second new PI randomly divides the data into two half sets H and V where H has $n_H = \lceil n/2 \rceil$ of the cases and V has the remaining $n_V = n - n_H$ cases $i_1, ..., i_{n_V}$. The estimator $\hat{m}_H(\boldsymbol{x})$ is computed using the training data set H. Then the validation residuals $v_j = Y_{i_j} - \hat{m}_H(\boldsymbol{x}_{i_j})$ are computed for $j = 1, ..., n_V$ cases in the validation set V. Find the Frey PI $[v_{(s)}, v_{(s+c-1)}]$ of the validation residuals (replacing n in (3.3) by $n_V = n - n_H$).

Then second new $100(1-\delta)\%$ large sample PI for Y_f is

$$[\hat{m}_H(\boldsymbol{x}_f) + v_{(s)}, \hat{m}_H(\boldsymbol{x}_f) + v_{(s+c-1)}]. \tag{3.12}$$

The PIs (3.11) and (3.12) are asymptotically equivalent if p is fixed and $n \to \infty$, but $\hat{m}_H$ has about half the efficiency of $\hat{m}$. When PI (3.11) has severe undercoverage because $\hat{m}$ is a poor estimator of m, it is expected that PI (3.12) may have coverage closer to the nominal coverage. For example, if $\hat{m}$ interpolates the data and $\hat{m}_H$ interpolates the training data from H, then the validation residuals will be huge. Hence PI (3.12) will be long compared to PI (3.11).

We can also motivate PI (3.12) by modifying the justification for the Lei, G'Sell, Rinaldo, Tibshirani, and Wasserman (2016) split conformal prediction interval $[\hat{m}_H(\boldsymbol{x}_f) - a_q, \hat{m}_H(\boldsymbol{x}_f) + a_q]$ where a_q is an appropriate quantile of the absolute validation residuals. Suppose $(Y_i, \boldsymbol{x}_i)$ are iid for $i = 1, ..., n, n+1$ where $(Y_f, \boldsymbol{x}_f) = (Y_{n+1}, \boldsymbol{x}_{n+1})$. Compute $\hat{m}_H(\boldsymbol{x})$ from the cases in H. For example, get $\hat{\boldsymbol{\beta}}_H$ from the cases in H. Consider the validation residuals v_i for $i = 1, ..., n_V$ and the validation residual v_{n_V+1} for case $(Y_f, \boldsymbol{x}_f)$. Since these $n_V + 1$ cases are iid, the probability that v_t has rank j for $j = 1, ..., n_V + 1$ is $1/(n_V+1)$ for each t, i.e., the ranks follow the discrete uniform distribution. Let $t = n_V + 1$ and let $v_{(i)}$ be the ordered residuals using $i = 1, ..., n_V$. That is, get the order statistics without using the unknown validation residual v_{n_V+1}. Then $v_{(i)}$ has rank i if $v_{(i)} < v_{n_V+1}$ but rank $i+1$ if $v_{(i)} > v_{n_V+1}$. Thus

$$P(Y_f \in [\hat{m}_H(\boldsymbol{x}_f) + v_{(k)}, \hat{m}_H(\boldsymbol{x}_f) + v_{(k+b-1)}]) = P(v_{(k)} \le v_{n_V+1} \le v_{(k+b-1)}) \ge$$

$P(v_{n_V+1}$ has rank between $k+1$ and $k+b-1$ and there are no tied ranks$) \ge$ $(b-1)/(n_V+1) \approx 1-\delta$ if $b = \lceil (n_V+1)(1-\delta) \rceil + 1$ and $k+b-1 \le n_V$. This probability statement holds for a fixed k such as $k = \lceil n_V \ \delta/2 \rceil$. The statement is not true when the shorth(b) estimator is used since the shortest interval using $k = s$ can have s change with the data set. That is, s is not fixed. Hence if PI's were made from J independent data sets, the PI's with fixed k would contain Y_f about $J(1-\delta)$ times, but this value would be smaller for the shorth(b) prediction intervals where s can change with the data set.

The above argument works if the estimator $\hat{m}(\boldsymbol{x})$ is "symmetric in the data." The assumption of iid cases is stronger than that of iid errors e_i. The split conformal PI can have good coverage, but PI (3.12) does not need the error distribution to be symmetric to be asymptotically optimal.

The PIs (3.11) and (3.12) can be used with $\hat{m}(\boldsymbol{x}) = \hat{Y}_f = \boldsymbol{x}_{I_d}^T \hat{\boldsymbol{\beta}}_{I_d}$ where I_d denotes the index of predictors selected from the model or variable selection method. The PIs (3.11) and (3.12) need the shorth of the residuals to be a consistent estimator of the population shorth of the error distribution. Olive and Hawkins (2003) show that if the $\|\boldsymbol{x}_i\|$ are bounded and $\hat{\boldsymbol{\beta}}$ is a consistent estimator of $\boldsymbol{\beta}$, then $\max_{i=1,...,n} |r_i - e_i| \xrightarrow{P} 0$ and the sample quantiles of the residuals estimate the population quantiles of the error distribution. For OLS with fixed p, each submodel I produces a $\sqrt{n}$ consistent estimator provided that $S \subseteq I$.

The Cauchy Schwartz inequality says $|\boldsymbol{a}^T \boldsymbol{b}| \leq \|\boldsymbol{a}\| \ \ \|\boldsymbol{b}\|$. Suppose $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = O_P(1)$ is bounded in probability. This will occur if $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{D} N_p(\boldsymbol{0}, \boldsymbol{\Sigma})$, e.g. if $\hat{\boldsymbol{\beta}}$ is the OLS estimator. Then

$$|r_i - e_i| = |Y_i - \boldsymbol{x}_i^T \hat{\boldsymbol{\beta}} - (Y_i - \boldsymbol{x}_i^T \boldsymbol{\beta})| = |\boldsymbol{x}_i^T (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})|.$$

Hence

$$\sqrt{n} \max_{i=1,...,n} |r_i - e_i| \leq (\max_{i=1,...,n} \|\boldsymbol{x}_i\|) \ \ \|\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})\| = O_P(1)$$

since $\max \|\boldsymbol{x}_i\| = O_P(1)$ or there is extrapolation. Hence OLS residuals behave well if the zero mean error distribution of the iid e_i has a finite variance σ^2.

Note that correction factors $b_n \to 1$ are used in large sample confidence intervals and tests if the limiting distribution is N(0,1) or χ^2_p, but a t_{d_n} or pF_{p,d_n} cutoff is used: $t_{d_n,1-\delta}/z_{1-\delta} \to 1$ and $pF_{p,d_n,1-\delta}/\chi^2_{p,1-\delta} \to 1$ if $d_n \to \infty$ as $n \to 1$. Using correction factors for prediction intervals and bootstrap confidence regions improves the performance for moderate sample size n.

3.3.3 PI after model selection

The PI (3.7) was used for the variable selection estimators.

Heuristically, the other methods consider a small number of models including the full OLS model, and models that beat the full OLS model for 10 fold CV likely fit the data well when p is fixed and $n \to \infty$. Hence under regularity conditions, the PI (3.7) is likely to perform well for the other methods.

As shown in simulations, lasso and ridge regression tended to have prediction intervals that were too long when $p \geq 20$, $a = p$ or $p - 1$, and the predictor variables were correlated (with $\psi \geq 0.5$). As a possible remedy, consider 10 fold CV where the data set has been randomly divided into 10 groups of approximately equal size. For $j = 1, ..., 10$, compute the estimator when the jth group is left out and compute the PIs for the $Y_f = Y_i$ in the left out group j. After obtaining the n PIs, one for each Y_i, compute the proportion of times Y_i was in its PI and the average length of the PIs. Consider the λ_i where the proportion $\geq 1 - \delta$, and use λ_a that had the shortest average PI length as the λ for the variable selection estimator. This technique changes the CV criterion to average PI length, given that the observed coverage was at least as large as the nominal coverage.

CHAPTER 4

BOOTSTRAPPING HYPOTHESIS TESTS

We also want to use bootstrap tests. Consider testing $H_0 : \boldsymbol{\mu} = \boldsymbol{c}$ versus $H_1 : \boldsymbol{\mu} \neq \boldsymbol{c}$ where $\boldsymbol{c}$ is a known $r \times 1$ vector. Given training data $\boldsymbol{z}_1, ..., \boldsymbol{z}_n$, a large sample $100(1-\delta)\%$ confidence region for $\boldsymbol{\mu}$ is a set $\mathcal{A}_n$ such that $P(\boldsymbol{\mu} \in \mathcal{A}_n) \to 1-\delta$ as $n \to \infty$. Then reject H_0 if $\boldsymbol{c}$ is not in the confidence region $\mathcal{A}_n$. For example, let $\boldsymbol{\mu} = \boldsymbol{A\beta}$ where $\boldsymbol{\beta}$ is a $p \times 1$ vector of parameters, and $\boldsymbol{A}$ is a known full rank $r \times p$ matrix with $1 \leq r \leq p$.

To bootstrap a confidence region, Mahalanobis distances and prediction regions will be useful. Consider predicting a future test value $\boldsymbol{z}_f$, given past training data $\boldsymbol{z}_1, ..., \boldsymbol{z}_n$ where the $\boldsymbol{z}_i$ are $r \times 1$ random vectors. A *large sample* $100(1-\delta)\%$ *prediction region* is a set $\mathcal{A}_n$ such that $P(\boldsymbol{z}_f \in \mathcal{A}_n) \to 1-\delta$ as $n \to \infty$. Let the $r \times 1$ column vector T be a multivariate location estimator, and let the $r \times r$ symmetric positive definite matrix $\boldsymbol{C}$ be a dispersion estimator. Then the ith *squared sample Mahalanobis distance* is the scalar

$$D_i^2 = D_i^2(T, \boldsymbol{C}) = D^2_{\boldsymbol{z}_i}(T, \boldsymbol{C}) = (\boldsymbol{z}_i - T)^T \boldsymbol{C}^{-1} (\boldsymbol{z}_i - T) \tag{4.1}$$

for each observation $\boldsymbol{z}_i$. Notice that the Euclidean distance of $\boldsymbol{z}_i$ from the estimate of center T is $D_i(T, \boldsymbol{I}_r)$ where $\boldsymbol{I}_r$ is the $r \times r$ identity matrix. The classical Mahalanobis distance D_i uses $(T, \boldsymbol{C}) = (\overline{\boldsymbol{z}}, \boldsymbol{S})$, the sample mean and sample covariance matrix where

$$\overline{\boldsymbol{z}} = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{z}_i \quad \text{and} \quad \boldsymbol{S} = \frac{1}{\mathrm{n}-1} \sum_{\mathrm{i}=1}^{\mathrm{n}} (\boldsymbol{z}_\mathrm{i} - \overline{\boldsymbol{z}})(\boldsymbol{z}_\mathrm{i} - \overline{\boldsymbol{z}})^\mathrm{T}. \tag{4.2}$$

Let q_n and c be given by (3.4) and (3.5) with p replaced by r. Let $(T, \boldsymbol{C}) = (\overline{\boldsymbol{z}}, \boldsymbol{S})$, and let $D_{(U_n)}$ be the $100q_n$th sample quantile of the D_i. Then the Olive (2013) large sample $100(1-\delta)\%$ nonparametric prediction region for a future value $\boldsymbol{z}_f$ given iid data $\boldsymbol{z}_1, ..., \boldsymbol{z}_n$ is

$$\{\boldsymbol{z} : D^2_{\boldsymbol{z}}(\overline{\boldsymbol{z}}, \boldsymbol{S}) \leq D^2_{(U_n)}\}, \tag{4.3}$$

while the classical large sample $100(1-\delta)\%$ prediction region is

$$\{\boldsymbol{z} : D^2_{\boldsymbol{z}}(\overline{\boldsymbol{z}}, \boldsymbol{S}) \leq \chi^2_{r,1-\delta}\}. \tag{4.4}$$

The following theorem is proved in Olive (2017bde) and shows that the hyperellipsoid R_c centered at the statistic T_n is a large sample $100(1-\delta)\%$ confidence region for $\boldsymbol{\mu}$, but the hyperellipsoid centered at known $\boldsymbol{\mu}$ is a large sample $100(1-\delta)\%$ prediction region for a future value of the statistic $T_{f,n}$.

Theorem 4.1. *Let the* $100(1-\delta)$*th percentile* $D^2_{1-\delta}$ *be a continuity point of the distribution of* D^2. *Assume that* $D^2_{\boldsymbol{\mu}}(T_n, \boldsymbol{\Sigma}_T) \xrightarrow{D} D^2$, $D^2_{\boldsymbol{\mu}}(T_n, \hat{\boldsymbol{\Sigma}}_T) \xrightarrow{D} D^2$, *and* $\hat{D}^2_{1-\delta} \xrightarrow{P} D^2_{1-\delta}$ *where* $P(D^2 \leq D^2_{1-\delta}) = 1-\delta$. *i) Then* $R_c = \{\boldsymbol{w} : D^2_{\boldsymbol{w}}(T_n, \hat{\boldsymbol{\Sigma}}_T) \leq \hat{D}^2_{1-\delta}\}$ *is a large sample* $100(1-\delta)\%$ *confidence region for* $\boldsymbol{\mu}$, *and if* $\boldsymbol{\mu}$ *is known, then* $R_p = \{\boldsymbol{w} : D^2_{\boldsymbol{w}}(\boldsymbol{\mu}, \hat{\boldsymbol{\Sigma}}_T) \leq \hat{D}^2_{1-\delta}\}$ *is a large sample* $100(1-\delta)\%$ *prediction region for a future value of the statistic* $T_{f,n}$. *ii) Region* R_c *contains* $\boldsymbol{\mu}$ *iff region* R_p *contains* T_n.

Hence if there was an iid sample $T_{1,n}, ..., T_{B,n}$ of the statistic, the prediction region (4.3) for $T_{f,n}$ contains $E(T_n) = \boldsymbol{\mu}$ with asymptotic coverage $\geq 1-\delta$. Often the n is suppressed. To make the asymptotic coverage equal to $1-\delta$, use the large sample $100(1-\delta)\%$ confidence region $\{\boldsymbol{w} : D^2_{\boldsymbol{w}}(T_{1,n}, \boldsymbol{S}_T) \leq D^2_{(U_B)}\}$. The prediction region method bootstraps this procedure by using a bootstrap sample of the statistic $T^*_{1,n}, ..., T^*_{B,n}$. Let $\overline{T^*}$ and $\boldsymbol{S}^*_T$ be the sample mean and sample covariance matrix of $T^*_{1,n}, ..., T^*_{B,n}$. Centering the region at $T^*_{1,n}$ instead of $\overline{T}^*$ is not needed since the bootstrap sample is centered near T_n: the distribution of $T_n - \boldsymbol{\mu}$ is approximated by the distribution of $T^* - T_n$ or by the distribution of $T^* - \overline{T}^*$.

4.1 PREDICTION REGION METHOD

The prediction region method is simple. Let $\hat{\boldsymbol{\mu}}$ be a consistent estimator of $\boldsymbol{\mu}$ and make a bootstrap sample $\boldsymbol{w}_i = \hat{\boldsymbol{\mu}}^*_i - \boldsymbol{c}$ for $i = 1, ..., B$. Using the nonparametric prediction region (4.3) for the $\boldsymbol{w}_i$ as a large sample $100(1-\delta)\%$ confidence region, fail to reject H_0 if $\boldsymbol{0}$ is in the confidence region (if $D_{\boldsymbol{0}} \leq D_{(U_B)}$), and reject H_0 otherwise. The method tends to work well in simulations if $\sqrt{n}(T_n - \boldsymbol{\mu}) \xrightarrow{D} \boldsymbol{z}$ where the random vector $\boldsymbol{z}$ has a nonsingular covariance matrix.

Following Bickel and Ren (2001), let the vector of parameters $\boldsymbol{\mu} = T(F)$, the statistic $T_n = T(F_n)$, and $T^* = T(F_n^*)$ where F is the cdf of iid $\boldsymbol{x}_1, ..., \boldsymbol{x}_n$, F_n is the empirical cdf, and F_n^* is the empirical cdf of $\boldsymbol{x}_1^*, ..., \boldsymbol{x}_n^*$, a sample from F_n using the nonparametric bootstrap. If $\sqrt{n}(F_n - F) \xrightarrow{D} \boldsymbol{z}_F$, a Gaussian random process, and if T is sufficiently smooth (Hadamard differentiable with a Hadamard derivative $\dot{T}(F)$), then $\sqrt{n}(T_n - \boldsymbol{\mu}) \xrightarrow{D} \boldsymbol{X}$ and $\sqrt{n}(T_i^* - \overline{T}^*) \xrightarrow{D} \boldsymbol{X}$ with $\boldsymbol{X} = \dot{T}(F)\boldsymbol{z}_F$. Olive (2017be) used these results to show that if $\boldsymbol{X} \sim N_r(\mathbf{0}, \boldsymbol{\Sigma}_T)$, then $\sqrt{n}(\overline{T}^* - T_n) \xrightarrow{D} \mathbf{0}$, $\sqrt{n}(\overline{T}^* - \boldsymbol{\mu}) \xrightarrow{D} \boldsymbol{X}$, and that the prediction region method large sample $100(1-\delta)\%$ confidence region for $\boldsymbol{\mu}$ is

$$\{\boldsymbol{w} : (\boldsymbol{w} - \overline{T}^*)^T [\boldsymbol{S}_T^*]^{-1} (\boldsymbol{w} - \overline{T}^*) \le D^2_{(U_B)}\} = \{\boldsymbol{w} : D^2_{\boldsymbol{w}}(\overline{T}^*, \boldsymbol{S}_T^*) \le D^2_{(U_B)}\} \tag{4.5}$$

where $D^2_{(U_B)}$ is computed from $D_i^2 = (T_i^* - \overline{T}^*)^T [\boldsymbol{S}_T^*]^{-1} (T_i^* - \overline{T}^*)$ for $i = 1, ..., B$. Note that the corresponding test for $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ rejects H_0 if $(\overline{T}^* - \boldsymbol{\mu}_0)^T [\boldsymbol{S}_T^*]^{-1} (\overline{T}^* - \boldsymbol{\mu}_0) > D^2_{(U_B)}$.

The prediction region method for testing $H_0 : \boldsymbol{\mu} = \boldsymbol{c}$ versus $H_1 : \boldsymbol{\mu} \neq \boldsymbol{c}$ is simple. Let $\hat{\boldsymbol{\mu}}$ be a consistent estimator of $\boldsymbol{\mu}$ and make a bootstrap sample $\boldsymbol{w}_i = \hat{\boldsymbol{\mu}}_i^* - \boldsymbol{c}$ for $i = 1, ..., B$. Make the nonparametric prediction region (4.3) for the $\boldsymbol{w}_i$ and fail to reject H_0 if $\mathbf{0}$ is in the prediction region (if $D_{\mathbf{0}} \le D_{(U_B)}$), reject H_0 otherwise.

The Bickel and Ren (2001) hypothesis testing method is equivalent to using confidence region (4.3) with $\overline{T}^*$ replaced by T_n and U_B replaced by $k_B = \lceil B(1-\delta) \rceil$. If region (4.3) or the Bickel and Ren (2001) region is a large sample $100(1-\delta)\%$ confidence region, then so is the other region if $\sqrt{n}(\overline{T}^* - T_n) \xrightarrow{D} \mathbf{0}$. Hadamard differentiability and asymptotic normality are sufficient conditions for both regions to be large sample confidence regions if $\boldsymbol{S}_T^* \xrightarrow{P} \boldsymbol{\Sigma}_T$, but Bickel and Ren (2001) showed that their method can work when Hadamard differentiability fails. For $r = 1$, the percentile method uses an interval that contains $U_B \approx k_B = \lceil B(1-\delta) \rceil$ of the $T_{i,n}^*$ from a bootstrap sample $T_{1,n}^*, ..., T_{B,n}^*$ where the statistic T_n is an estimator of μ based on a sample of size n. Note that the squared Mahalanobis distance $D_\mu^2 = (\mu - \overline{T^*})^2 / S_T^{2*} \le D^2_{(U_B)}$ is equivalent to $\mu \in [\overline{T^*} - S_T^* D_{(U_B)}, \overline{T^*} + S_T^* D_{(U_B)}]$, which is an interval centered at $\overline{T^*}$ just long enough to cover U_B of the T_i^*. Hence the prediction region method is a special case of the percentile method if $r = 1$. Efron (2014) uses a similar large sample $100(1-\delta)\%$ confidence interval assuming that T_n is asymptotically normal. The Frey (2013) shorth(c) interval applied to the $T^*_{i,n}$ is recommended since the shorth confidence interval can be much shorter than the Efron (2014) or prediction region method confidence intervals if $r = 1$. See Olive (2017e) for more information about the prediction region method.

Following Olive (2017bce), we describe the prediction region method for bootstrapping forward selection where 0s need to be added for omitted variables. Bootstrapping lasso and ridge regression is similar, but lasso already has the 0s and ridge regression usually does not produce 0 slope estimates. Suppose $n > 20p$. If $\hat{\boldsymbol{\beta}}_I$ is $(k \times 1) \times 1$, form $\hat{\boldsymbol{\beta}}_{I,0}$ from $\hat{\boldsymbol{\beta}}_I$ by adding 0s corresponding to the omitted variables. Then $\hat{\boldsymbol{\beta}}_{I,0}$ is a nonlinear estimator of $\boldsymbol{\beta}$, and the residual bootstrap method can be applied. Let $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_{I_{min},0}$ be formed from the forward selection model I_{min} that minimizes the C_p criterion. Instead of computing the least squares estimator from regressing $\boldsymbol{Y}^*_i$ on $\boldsymbol{X}$, perform variable selection on $\boldsymbol{Y}^*_i$ and $\boldsymbol{X}$, fit the model that minimizes the criterion, and add 0s corresponding to the omitted variables, resulting in estimators $\hat{\boldsymbol{\beta}}^*_1, ..., \hat{\boldsymbol{\beta}}^*_B$. Then test $\boldsymbol{\mu} = \boldsymbol{A\beta} = \boldsymbol{c}$ using the prediction region method for $r > 1$ and the shorth if $r = 1$.

4.2 RESIDUAL BOOTSTRAP

For models of form (1.1) with $n > 20p$, the residual bootstrap makes sense: the residuals from the full model are sampled with replacement resulting in a bootstrap sample $r^*_1, ..., r^*_n$ and $Y^*_i = \hat{Y}_i + r^*_i$ for $i = 1, ..., n$ are collected into a vector $\boldsymbol{Y}^*_j$ which is regressed on $\boldsymbol{X}$ to get $\hat{\boldsymbol{\beta}}^*_j$ for $j = 1, ..., n$. The nonparametric bootstrap selects n cases with replacement to form $(\boldsymbol{Y}^*_j, \boldsymbol{X}^*_j)$, and regresses $\boldsymbol{Y}^*_j$ on $\boldsymbol{X}^*_j$ to form $\hat{\boldsymbol{\beta}}^*_j$.

Note that if $S \subseteq I$, and $\boldsymbol{Y} = \boldsymbol{X}_I \boldsymbol{\beta}_I + \boldsymbol{e}_I$, then $\sqrt{n}(\hat{\boldsymbol{\beta}}_I - \boldsymbol{\beta}_I) \xrightarrow{D} N_{k+1}(\boldsymbol{0}, \sigma^2_I \boldsymbol{W}_I)$ under mild regularity conditions where $n(\boldsymbol{X}^T_I \boldsymbol{X}_I)^{-1} \to \boldsymbol{W}_I$. Hence $\sqrt{n}(\hat{\boldsymbol{\beta}}_{I,0} - \boldsymbol{\beta}_c) \xrightarrow{D} N_p(\boldsymbol{0}, \sigma^2_I \boldsymbol{W}_{I,0})$ where the $\boldsymbol{W}_{I,0}$ has a column and row of zeroes added for each variable not in I. Note that $\boldsymbol{W}_{I,0}$ is singular unless I corresponds to the full model. For example, if $p = 3$ and model I uses a constant and x_3 with

$$\boldsymbol{W}_I = \begin{bmatrix} W_{11} & W_{12} \\ W_{21} & W_{22} \end{bmatrix}, \quad \text{then} \quad \boldsymbol{W}_{\mathrm{I},0} = \begin{bmatrix} W_{11} & 0 & W_{12} \\ 0 & 0 & 0 \\ W_{21} & 0 & W_{22} \end{bmatrix}.$$

Hence it is reasonable to conjecture that $\sqrt{n}(\hat{\boldsymbol{\beta}}_{I_{min},0} - \boldsymbol{\beta}_c) \xrightarrow{D} \boldsymbol{X}$ where

$$\boldsymbol{X} = \sum_{i=1}^{K} \pi_i N_p(\boldsymbol{0}, \sigma^2_{I_i} \boldsymbol{W}_{I_i,0}),$$

$0 \leq \pi_i \leq 1$, $\sum_{i=1}^K \pi_i = 1$, and K is the number of subsets I_i that contain S. Note that the limiting distribution is not an elliptically contoured distribution (unless the full model has $\pi_k = 1$) since the probability of a 0 would be 0.

Prediction intervals and regions can have higher than the nominal coverage $1-\delta$ if the distribution is discrete or a mixture of a discrete distribution and some other distribution. In particular, coverage can be high if the $\boldsymbol{w}_i$ distribution is a mixture of a point mass at $\boldsymbol{0}$ and the method checks whether $\boldsymbol{0}$ is in the prediction region. Such a mixture often occurs for forward selection methods and lasso. The bootstrap sample for the $W_i = \hat{\boldsymbol{\beta}}^*_{ij}$ can contain many zeroes and be highly skewed if the jth predictor is weak. Then the computer program may fail because $\boldsymbol{S}_{\boldsymbol{w}}$ is singular, but if all or nearly all of the $\hat{\boldsymbol{\beta}}^*_{ij} = 0$, then there is strong evidence that the jth predictor is not needed given that the other predictors are in the variable selection method.

As an extreme simulation case, suppose $\hat{\boldsymbol{\beta}}^*_{ij} = 0$ for $i = 1, ..., B$ and for each run in the simulation. Consider testing $H_0 : \beta_j = 0$. Then regardless of the nominal coverage $1 - \delta$, the closed interval [0,0] will contain 0 for each run and the observed coverage will be $1 > 1 - \delta$. Using the open interval (0,0) would give observed coverage 0. Also intervals $[0, b]$ and $[a, 0]$ correctly suggest failing to reject $\beta_j = 0$, while intervals $(0, b)$ and $(a, 0)$ incorrectly suggest rejecting $H_0 : \beta_j = 0$. Hence closed regions and intervals make sense.

Following Seber and Lee (2003, p. 448) and Nishi (1984), the probability that model I_{min} from C_p or AIC underfits goes to zero as $n \to \infty$. Since there are a finite number of regression models I that contain the true model, and each model gives a consistent estimator $\hat{\boldsymbol{\beta}}_{I,0}$ of $\boldsymbol{\beta}$, the probability that I_{min} picks one of these models goes to one as $n \to \infty$. Hence $\hat{\boldsymbol{\beta}}_{I_{min},0}$ is a consistent estimator of $\boldsymbol{\beta}$ under model (2.2).

Note that when performing the prediction region method for lasso and ridge regression, we use the residual bootstrap where the residuals are from the full OLS model. Efron (1982, p. 36) notes that for the OLS residual bootstrap for the full OLS model, $E[\hat{\boldsymbol{\beta}}^*_j] = \hat{\boldsymbol{\beta}}_{OLS}$, and the sample covariance matrix of the $\hat{\boldsymbol{\beta}}^*_j$ is estimating the population bootstrap matrix $\frac{n-p}{n} MSE(\boldsymbol{X}^T\boldsymbol{X})^{-1}$ as $B \to \infty$. Hence the residual bootstrap standard error $SE(\hat{\beta}_i) \approx \sqrt{\frac{n-p}{n}}\, SE(\hat{\beta}_{i,OLS})$. Here the expectations are with respect to the bootstrap distribution. Camponovo (2015) suggests that the nonparametric bootstrap does not work for lasso.

CHAPTER 5

EXAMPLES AND SIMULATIONS

5.1 EXAMPLE

The Hebbler (1847) data was collected from $n = 26$ districts in Prussia in 1843. We will study the relationship between Y = the *number of women married to civilians* in the district with the predictors x_1 = constant, $x_2 = pop$ = the *population of the district in 1843*, $x_3 = mmen$ = the *number married civilian men* in the district, $x_4 = mmilmen$ = *number of married men in the military* in the district, and $x_5 = milwmn$ = the *number of women married to husbands in the military* in the district. Sometimes the person conducting the survey would not count a spouse if the spouse was not at home. Hence Y is highly correlated but not equal to x_3. Similarly, x_4 and x_5 are highly correlated but not equal. We expect that $Y = x_3 + e$ is a good model, but $n/p = 5.2$ is small.

Forward selection selected the model with the minimum C_p while the other methods used 10-fold CV. PLS and PCR used the OLS full model with PI length 2395.74, forward selection used a constant and *mmen* with PI length 2114.72, ridge regression had PI length 20336.58, lasso and relaxed lasso used a constant, *mmen*, and *pop* with lasso PI length 8482.62 and relaxed lasso PI length 2226.53. Figure 1 shows the response plots for forward selection, ridge regression, lasso, and relaxed lasso. The plots for PLS=PCR=OLS full model were similar to those of forward selection and relaxed lasso. The plots suggest that the MLR model is appropriate since the plotted points scatter about the identity line. The 90% pointwise prediction bands are also shown, and consist of two lines parallel to the identity line. These bands are very narrow in Figure 1 a) and d).

5.2 SIMULATION

For the simulations, for $u = 1, ..., n$, we generated $\boldsymbol{z}_u \sim N_{p-1}(\boldsymbol{0}, \boldsymbol{I})$ where the $m = p-1$ elements of the vector $\boldsymbol{z}_u$ are iid N(0,1). Let the $m \times m$ matrix $\boldsymbol{A} = (a_{ij})$ with $a_{ii} = 1$

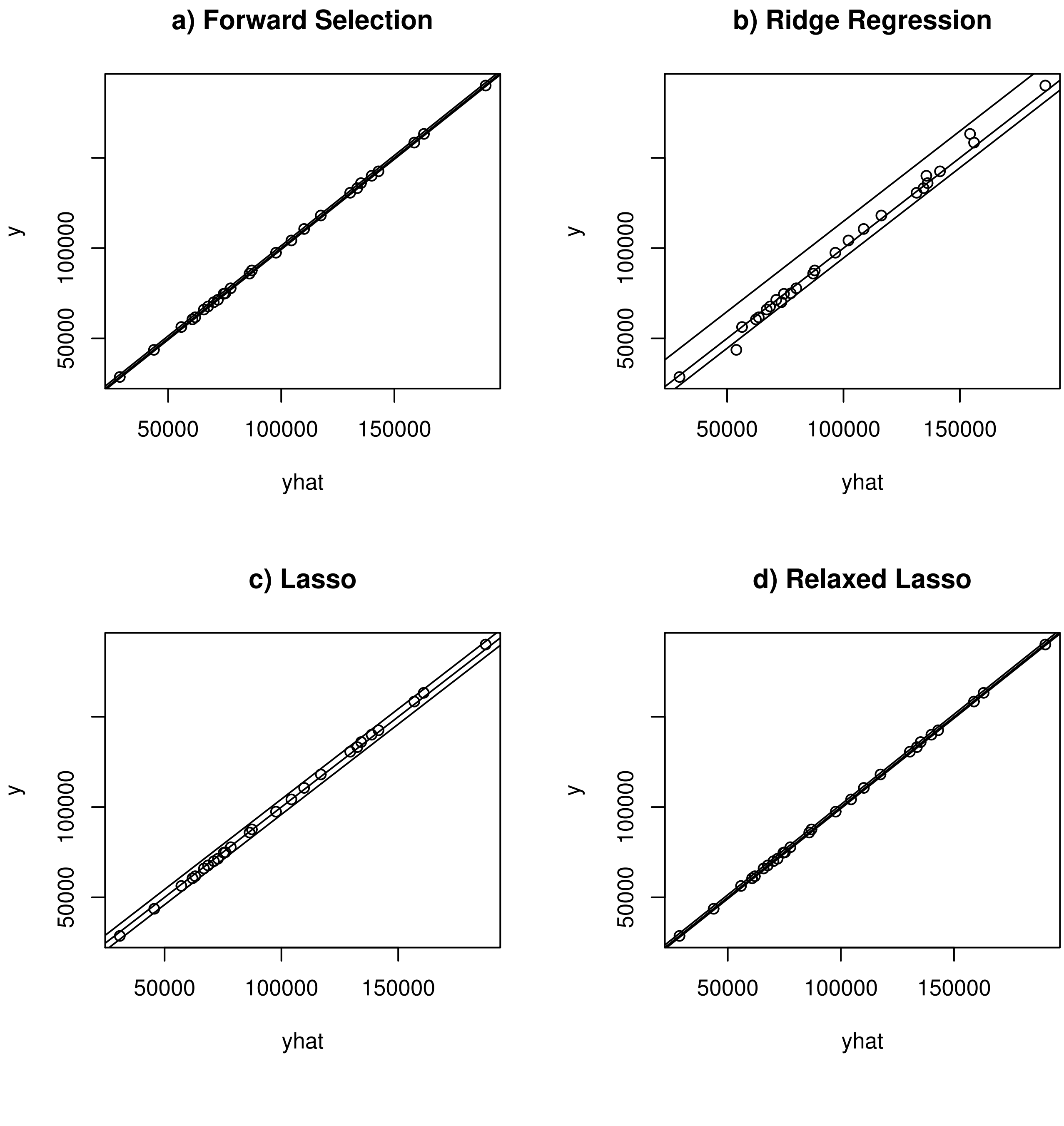

Figure 5.1. Marry Data Response Plots

and $a_{ij} = \psi$ where $0 \leq \psi < 1$ for $i \neq j$. Then the vector of predictors $\boldsymbol{w}_u = \boldsymbol{A}\boldsymbol{z}_u$ so that $Cov(\boldsymbol{w}) = \boldsymbol{\Sigma}_{\boldsymbol{w}} = \boldsymbol{A}\boldsymbol{A}^T = (\sigma_{ij})$ where the diagonal entries $\sigma_{ii} = [1 + (m-1)\psi^2]$ and the off diagonal entries $\sigma_{ij} = [2\psi + (m-2)\psi^2]$. Hence the correlations are $cor(x_i, x_j) = (2\psi + (m-2)\psi^2)/(1 + (m-1)\psi^2)$ for $i \neq j$ where x_i and x_j are nontrivial predictors. As ψ gets close to 1, the predictor vectors cluster about the line in the direction of $(1, ..., 1)^T$. The simulation used $\psi = 0, 0.5$, and 0.9. Then $Y_i = 1 + 1w_{i,1} + \cdots + 1w_{i,j} + e_i$ for $i = 1, ..., n$ with $a = k+1$ and $k = 1, p-2$, or $p-1$. Hence $\boldsymbol{\beta} = (1, .., 1, 0, ..., 0)^T$ with a ones and $p-a$ zeros. The zero mean errors e_i were iid of five types: i) N(0,1) errors, ii) t_3 errors, iii) EXP(1) - 1 errors, iv) uniform$(-1, 1)$ errors, and v) 0.9 N(0,1) + 0.1 N(0,100) errors.

The lengths of the asymptotically optimal 95% PIs are i) 3.92 = 2(1.96), ii) 6.365, iii) 2.996, iv) 1.90 = 2(0.95), and v) 13.490. The simulation used 5000 runs, so an observed coverage in (0.94, 0.96) gives no reason to doubt that the PI has the nominal coverage of 0.95. The simulations use $n = 10p, 20p$ and $100p$. Let a be the number of nonzero coefficients, including the constant, in $\boldsymbol{\beta}$. The coverage was often high for $n = 10p$ and $20p$, but close to the nominal coverage of 0.95 for $n = 100p$, where the average lengths were slightly longer than the asymptotically optimal lengths, except the lasso and ridge regression PIs were far too long when $\psi \geq 0.5$ and $a = p-1$ or $a = p$. Tables 5.1 -5.33 used 10 fold CV except forward selection used C_p with PI (3.7). Tables 5.34 -5.48 used PI (3.11) with $d = \min(\lceil n/J \rceil, p)$. The Simulation for Tables 5.16 -5.33 is described in Section 5.4.

5.2.1 Error type 1

Table 5.1. $p = 50$, *error type* $= 1$

			LASSO			RIDGE		
n	a	psi	coverage	length	penalty	coverage	length	penalty
500	1	0	0.9882	5.1527	0.0788	0.9812	4.9571	0.1557
		0.5	0.9890	5.1630	0.0260	0.9886	5.5414	0.3945
		0.9	0.9894	5.1823	0.0455	0.9898	5.2125	2.3158
	49	0	0.9780	4.9323	0.0069	0.9794	5.0308	0.1934
		0.5	0.9838	29.4716	5.4232	0.9878	29.1538	276.2661
		0.9	0.9884	51.5459	8.8077	0.9874	50.2634	492.4197
	50	0	0.9776	4.9325	0.0069	0.9784	5.0323	0.1946
		0.5	0.9840	30.0982	5.5362	0.9884	29.6323	282.0203
		0.9	0.9880	52.6117	8.9912	0.9868	51.3007	502.6784
1000	1	0	0.9796	4.7745	0.0559	0.9738	4.6880	0.1097
		0.5	0.9808	4.7813	0.0260	0.9784	5.1593	0.3954
		0.9	0.9798	4.7936	0.0456	0.9808	4.8177	2.3211
	49	0	0.9744	4.6794	0.0071	0.9770	4.7408	0.1668
		0.5	0.9848	27.1686	5.4327	0.9860	26.9734	276.7502
		0.9	0.9850	47.7251	8.8261	0.9852	46.5427	493.4512
	50	0	0.9746	4.6795	0.0071	0.9756	4.7414	0.1679
		0.5	0.9836	27.7414	5.5459	0.9862	27.4171	282.5143
		0.9	0.9850	48.7126	9.0100	0.9850	47.4989	503.7313
5000	1	0	0.9510	4.0554	0.0251	0.9480	4.0489	0.1097
		0.5	0.9514	4.0575	0.0260	0.9520	4.3926	0.3956
		0.9	0.9516	4.0617	0.0456	0.9506	4.0814	2.3223
	49	0	0.9498	4.0435	0.0072	0.9486	4.0726	0.1336
		0.5	0.9576	22.9195	5.4337	0.9582	22.8305	276.8005
		0.9	0.9588	40.4261	8.8322	0.9580	39.4116	493.7890
	50	0	0.9508	4.0437	0.0073	0.9486	4.0728	0.1339
		0.5	0.9606	23.3987	5.5469	0.9610	23.2087	282.5654
		0.9	0.9586	41.2620	9.0162	0.9584	40.2207	504.0763

Table 5.2. $p = 50$, *error type* $= 1$

			PLS			PCR		
n	a	psi	coverage	length	penalty	coverage	length	penalty
500	1	0	0.9776	4.9243	44.1168	0.9788	4.9360	0.5632
		0.5	0.9788	4.9304	45.6042	0.9814	4.9969	7.5334
		0.9	0.9804	4.9498	46.9952	0.9882	5.1708	43.8786
	49	0	0.9782	4.9237	41.0860	0.9778	4.9273	0.0012
		0.5	0.9782	4.9296	45.5902	0.9786	4.9979	7.5416
		0.9	0.9800	4.9493	46.9958	0.9880	5.1702	43.8168
	50	0	0.9778	4.9237	41.1072	0.9776	4.9277	0.0010
		0.5	0.9802	4.9471	46.9990	0.9886	5.1583	44.7416
		0.9	0.9798	4.9474	46.9992	0.9880	5.1580	44.7980
1000	1	0	0.9738	4.6734	43.9162	0.9748	4.6763	0.1322
		0.5	0.9734	4.6734	44.5844	0.9738	4.6831	1.6540
		0.9	0.9740	4.6787	46.9940	0.9792	4.7906	43.0660
	49	0	0.9736	4.6734	42.2408	0.9738	4.6734	0.0000
		0.5	0.9744	4.6739	44.5664	0.9734	4.6829	1.7140
		0.9	0.9740	4.6785	46.9924	0.9784	4.7923	43.1088
	50	0	0.9736	4.6734	42.2468	0.9738	4.6734	0.0000
		0.5	0.9732	4.6780	46.9994	0.9790	4.7790	45.0346
		0.9	0.9740	4.6780	46.9994	0.9782	4.7787	45.0822
5000	1	0	0.9498	4.0386	44.2560	0.9498	4.0387	0.0052
		0.5	0.9498	4.0386	43.7142	0.9488	4.0391	0.0892
		0.9	0.9506	4.0388	46.3834	0.9482	4.0561	26.7282
	49	0	0.9498	4.0386	43.5068	0.9498	4.0386	0.0000
		0.5	0.9498	4.0386	43.6980	0.9496	4.0389	0.0818
		0.9	0.9494	4.0387	46.4070	0.9490	4.0562	27.1914
	50	0	0.9498	4.0386	43.5122	0.9498	4.0386	0.0000
		0.5	0.9502	4.0386	47.0000	0.9522	4.0565	45.4298
		0.9	0.9500	4.0387	47.0000	0.9522	4.0563	45.4428

Table 5.3. $p = 50$, *error type* $= 1$

			Ii			I-Min		
n	a	psi	coverage	length	penalty	coverage	length	penalty
500	1	0	0.9850	5.0625	6.3984	0.9856	5.0360	8.5492
		0.5	0.9842	5.0604	6.5596	0.9842	5.0368	8.6302
		0.9	0.9848	5.0617	6.3518	0.9848	5.0378	8.4302
	49	0	0.9778	4.9277	0.0742	0.9778	4.9265	0.1498
		0.5	0.9780	4.9274	0.0752	0.9782	4.9264	0.1522
		0.9	0.9742	4.9935	-10.7068	0.9762	4.9705	-8.8512
	50	0	0.9778	4.9238	0.0000	0.9778	4.9238	0.0000
		0.5	0.9778	4.9238	0.0000	0.9778	4.9238	0.0000
		0.9	0.9756	4.9942	-11.0666	0.9762	4.9695	-9.2146
1000	1	0	0.9758	4.7353	6.3868	0.9756	4.7215	8.6396
		0.5	0.9772	4.7335	6.5390	0.9764	4.7219	8.6216
		0.9	0.9776	4.7337	6.5002	0.9766	4.7219	8.5936
	49	0	0.9740	4.6747	0.0836	0.9742	4.6743	0.1532
		0.5	0.9736	4.6746	0.0844	0.9738	4.6743	0.1580
		0.9	0.9728	4.6870	-2.9844	0.9734	4.6795	-1.8396
	50	0	0.9738	4.6734	0.0000	0.9738	4.6734	0.0000
		0.5	0.9738	4.6734	0.0000	0.9738	4.6734	0.0000
		0.9	0.9718	4.6855	-3.1798	0.9728	4.6788	-2.0432
5000	1	0	0.9502	4.0491	6.3052	0.9504	4.0471	8.5714
		0.5	0.9508	4.0490	6.4562	0.9500	4.0472	8.5294
		0.9	0.9512	4.0490	6.4586	0.9494	4.0472	8.5278
	49	0	0.9502	4.0387	0.0828	0.9500	4.0386	0.1592
		0.5	0.9504	4.0388	0.0848	0.9502	4.0387	0.1586
		0.9	0.9504	4.0388	0.0852	0.9502	4.0387	0.1584
	50	0	0.9498	4.0386	0.0000	0.9498	4.0386	0.0000
		0.5	0.9498	4.0386	0.0000	0.9498	4.0386	0.0000
		0.9	0.9498	4.0386	0.0000	0.9498	4.0386	0.0000

5.2.2 Error type 2

Table 5.4. $p = 5$, *error type* $= 2$

			LASSO			RIDGE		
n	a	psi	coverage	length	penalty	coverage	length	penalty
50	1	0	0.9772	9.9328	0.1824	0.9736	9.8355	8.3426
		0.5	0.9746	9.8959	0.0921	0.9740	9.9275	1.6133
		0.9	0.9758	9.9943	0.0294	0.9772	10.1353	0.6892
	4	0	0.9742	9.7695	0.0568	0.9744	9.7885	1.4834
		0.5	0.9730	9.7707	0.0527	0.9758	9.9376	0.5609
		0.9	0.9756	9.9452	0.0458	0.9776	10.1609	0.6115
	5	0	0.9734	9.7372	0.0219	0.9746	9.7845	1.4030
		0.5	0.9734	9.7361	0.0451	0.9758	9.9101	0.6962
		0.9	0.9760	9.9013	0.0566	0.9784	10.1766	0.8057
100	1	0	0.9714	9.1887	0.1283	0.9702	9.1209	1.2608
		0.5	0.9718	9.1752	0.0623	0.9710	9.1811	0.3677
		0.9	0.9728	9.2326	0.0198	0.9726	9.2914	0.2566
	4	0	0.9704	9.1076	0.0363	0.9704	9.1178	0.1793
		0.5	0.9700	9.1127	0.0359	0.9714	9.2040	0.4420
		0.9	0.9712	9.1821	0.0407	0.9726	9.3169	0.6064
	5	0	0.9700	9.0958	0.0110	0.9702	9.1195	0.1703
		0.5	0.9702	9.0957	0.0358	0.9710	9.1814	0.5732
		0.9	0.9714	9.1522	0.0533	0.9732	9.3268	0.8065
500	1	0	0.9564	6.6879	0.0583	0.9566	6.6803	0.1150
		0.5	0.9554	6.6851	0.0281	0.9556	6.7018	0.1482
		0.9	0.9554	6.6912	0.0136	0.9544	6.7046	0.2058
	4	0	0.9554	6.6784	0.0166	0.9560	6.6837	0.1190
		0.5	0.9566	6.6805	0.0266	0.9556	6.7236	0.3965
		0.9	0.9556	6.6825	0.0400	0.9554	6.7227	0.6081
	5	0	0.9564	6.6768	0.0074	0.9550	6.6837	0.1218
		0.5	0.9554	6.6768	0.0343	0.9538	6.6995	0.5219
		0.9	0.9544	6.6803	0.0533	0.9536	6.7347	0.8105

Table 5.5. $p = 5$, *error type* $= 2$

			PLS			PCR		
n	a	psi	coverage	length	penalty	coverage	length	penalty
50	1	0	0.9734	9.7343	1.2448	0.9748	9.9860	0.6706
		0.5	0.9726	9.7428	1.3232	0.9752	9.9234	1.0894
		0.9	0.9742	9.7454	1.7674	0.9758	9.9200	1.5528
	4	0	0.9732	9.7337	0.8578	0.9754	10.1383	0.3900
		0.5	0.9722	9.7444	1.3014	0.9746	9.9137	1.0718
		0.9	0.9738	9.7491	1.7742	0.9752	9.9247	1.5602
	5	0	0.9728	9.7335	0.8100	0.9736	10.2180	0.3730
		0.5	0.9738	9.7473	1.8096	0.9754	9.9202	1.5706
		0.9	0.9736	9.7485	1.8006	0.9750	9.9231	1.5876
100	1	0	0.9704	9.0976	0.9366	0.9704	9.2133	0.3752
		0.5	0.9696	9.0949	1.0148	0.9704	9.1838	0.8324
		0.9	0.9696	9.0997	1.7724	0.9712	9.1874	1.5566
	4	0	0.9700	9.0968	0.7708	0.9710	9.3033	0.2260
		0.5	0.9698	9.0993	0.9948	0.9698	9.1934	0.8400
		0.9	0.9696	9.0992	1.7594	0.9708	9.1909	1.5576
	5	0	0.9698	9.0966	0.7370	0.9702	9.3822	0.2244
		0.5	0.9696	9.0988	1.8188	0.9710	9.1862	1.6026
		0.9	0.9696	9.0992	1.8058	0.9712	9.1877	1.6066
500	1	0	0.9558	6.6770	0.6646	0.9560	6.7100	0.0962
		0.5	0.9558	6.6770	0.4510	0.9560	6.6979	0.2624
		0.9	0.9558	6.6770	1.6196	0.9556	6.6883	1.4342
	4	0	0.9558	6.6770	0.6704	0.9560	6.7325	0.0490
		0.5	0.9558	6.6770	0.4456	0.9556	6.6967	0.2498
		0.9	0.9556	6.6771	1.6180	0.9554	6.6891	1.4276
	5	0	0.9560	6.6770	0.6410	0.9554	6.7353	0.0454
		0.5	0.9556	6.6770	1.8196	0.9554	6.6879	1.6340
		0.9	0.9556	6.6770	1.8100	0.9552	6.6873	1.6378

Table 5.6. $p = 5$, *error type* $= 2$

			Ii			I-Min		
n	a	psi	coverage	length	penalty	coverage	length	penalty
50	1	0	0.9762	9.9737	1.2850	0.9756	9.9096	1.4938
		0.5	0.9754	9.9523	1.2292	0.9750	9.8985	1.4116
		0.9	0.9740	9.9293	1.2100	0.9742	9.8714	1.3724
	4	0	0.9732	9.8293	0.0044	0.9740	9.7976	0.1108
		0.5	0.9732	9.8391	-0.4678	0.9728	9.7962	-0.2490
		0.9	0.9750	9.9112	-1.9072	0.9746	9.8848	-1.8164
	5	0	0.9730	9.7616	-0.1250	0.9732	9.7456	-0.0772
		0.5	0.9728	9.7794	-0.8186	0.9724	9.7506	-0.6098
		0.9	0.9748	9.9013	-2.9142	0.9740	9.8778	-2.7928
100	1	0	0.9722	9.2154	1.2606	0.9716	9.1810	1.4812
		0.5	0.9708	9.2085	1.2518	0.9704	9.1769	1.4584
		0.9	0.9714	9.1908	1.1746	0.9710	9.1664	1.3160
	4	0	0.9708	9.1401	0.0742	0.9706	9.1261	0.1482
		0.5	0.9706	9.1371	-0.0608	0.9708	9.1237	0.0614
		0.9	0.9722	9.1798	-1.9200	0.9716	9.1681	-1.7926
	5	0	0.9698	9.1007	-0.0144	0.9698	9.0984	-0.0088
		0.5	0.9696	9.1017	-0.2460	0.9698	9.0989	-0.1474
		0.9	0.9710	9.1626	-2.7926	0.9712	9.1454	-2.5192
500	1	0	0.9564	6.6901	1.2574	0.9566	6.6863	1.4778
		0.5	0.9564	6.6885	1.2706	0.9560	6.6848	1.4766
		0.9	0.9560	6.6876	1.1408	0.9558	6.6849	1.2690
	4	0	0.9554	6.6815	0.0782	0.9554	6.6800	0.1592
		0.5	0.9558	6.6813	0.0780	0.9558	6.6797	0.1518
		0.9	0.9570	6.6866	-1.1268	0.9564	6.6848	-0.9246
	5	0	0.9558	6.6770	0.0000	0.9558	6.6770	0.0000
		0.5	0.9558	6.6771	-0.0010	0.9558	6.6771	-0.0004
		0.9	0.9566	6.6858	-1.7416	0.9558	6.6826	-1.4552

5.2.3 Error type 3

Table 5.7. $p = 10$, *error type* $= 3$

			LASSO			RIDGE		
n	a	psi	coverage	length	penalty	coverage	length	penalty
100	1	0	0.9826	5.0712	0.1104	0.9824	5.1036	0.1668
		0.5	0.9836	5.0281	0.0397	0.9830	5.3268	0.2090
		0.9	0.9852	5.0034	0.0205	0.9862	5.0745	0.3013
	9	0	0.9804	5.0877	0.0130	0.9814	5.1694	0.1590
		0.5	0.9814	5.1293	0.0929	0.9838	5.4815	1.4125
		0.9	0.9832	5.3474	0.1613	0.9864	5.4732	2.3956
	10	0	0.9798	5.0819	0.0081	0.9816	5.1779	0.1644
		0.5	0.9808	5.1406	0.1044	0.9842	5.1754	1.5884
		0.9	0.9826	5.4734	0.2028	0.9870	5.5471	2.7749
200	1	0	0.9762	4.2971	0.0798	0.9748	4.3708	0.1128
		0.5	0.9764	4.2725	0.0282	0.9756	4.6320	0.1899
		0.9	0.9766	4.2469	0.0197	0.9780	4.3240	0.2992
	9	0	0.9734	4.3620	0.0099	0.9718	4.4411	0.1429
		0.5	0.9732	4.4064	0.0929	0.9760	4.7785	1.4123
		0.9	0.9754	4.6542	0.1573	0.9774	4.7147	2.3893
	10	0	0.9732	4.3619	0.0074	0.9734	4.4484	0.1467
		0.5	0.9738	4.4271	0.1044	0.9760	4.4684	1.5872
		0.9	0.9738	4.7459	0.1794	0.9774	4.7585	2.6975
1000	1	0	0.9548	3.2772	0.0357	0.9562	3.3704	0.1097
		0.5	0.9550	3.2658	0.0140	0.9548	3.7015	0.1900
		0.9	0.9552	3.2490	0.0197	0.9546	3.3741	0.3001
	9	0	0.9544	3.3299	0.0073	0.9550	3.4351	0.1231
		0.5	0.9540	3.4106	0.0930	0.9540	3.8415	1.4131
		0.9	0.9542	3.7372	0.1578	0.9586	3.7640	2.3978
	10	0	0.9552	3.3331	0.0073	0.9548	3.4386	0.1247
		0.5	0.9544	3.4342	0.1045	0.9574	3.5554	1.5880
		0.9	0.9558	3.8301	0.1775	0.9590	3.8024	2.6975

Table 5.8. $p = 10$, *error type* $= 3$

			PLS			PCR		
n	a	psi	coverage	length	penalty	coverage	length	penalty
100	1	0	0.9804	5.0809	4.6928	0.9814	5.2110	0.4524
		0.5	0.9808	5.0840	5.2992	0.9824	5.2301	1.9986
		0.9	0.9816	5.0806	6.8046	0.9862	5.0805	5.7256
	9	0	0.9802	5.0794	3.7304	0.9792	5.2952	0.0992
		0.5	0.9804	5.0842	5.2704	0.9814	5.2292	1.9652
		0.9	0.9814	5.0819	6.8188	0.9846	5.0809	5.7016
	10	0	0.9804	5.0794	3.6872	0.9800	5.2914	0.0894
		0.5	0.9814	5.0804	6.8760	0.9846	5.0418	5.9098
		0.9	0.9818	5.0812	6.8840	0.9858	5.0300	5.9556
200	1	0	0.9742	4.3602	4.5072	0.9740	4.4271	0.1650
		0.5	0.9744	4.3601	4.7540	0.9730	4.4286	0.8790
		0.9	0.9738	4.3581	6.7838	0.9756	4.3359	5.5262
	9	0	0.9738	4.3600	4.0600	0.9738	4.4630	0.0374
		0.5	0.9734	4.3605	4.7258	0.9734	4.4273	0.8574
		0.9	0.9748	4.3582	6.7866	0.9766	4.3355	5.5400
	10	0	0.9738	4.3599	3.9968	0.9738	4.4561	0.0360
		0.5	0.9740	4.3584	6.8838	0.9760	4.2779	6.0038
		0.9	0.9742	4.3584	6.8954	0.9758	4.2720	6.0278
1000	1	0	0.9556	3.3283	4.6964	0.9556	3.3351	0.0088
		0.5	0.9554	3.3283	4.2268	0.9552	3.3425	0.1080
		0.9	0.9556	3.3284	6.1330	0.9546	3.3436	3.8246
	9	0	0.9554	3.3283	4.4694	0.9560	3.3380	0.0030
		0.5	0.9554	3.3283	4.1932	0.9548	3.3453	0.1192
		0.9	0.9554	3.3283	6.1082	0.9546	3.3436	3.8692
	10	0	0.9556	3.3282	4.4778	0.9556	3.3367	0.0022
		0.5	0.9556	3.3278	6.9114	0.9550	3.2683	6.0526
		0.9	0.9558	3.3279	6.9136	0.9548	3.2649	6.0808

Table 5.9. $p = 10$, *error type* $= 3$

			Ii			I-min		
n	a	psi	coverage	length	penalty	coverage	length	penalty
100	1	0	0.9854	5.0381	1.7550	0.9840	5.0599	2.2854
		0.5	0.9842	5.0401	1.8302	0.9840	5.0561	2.3010
		0.9	0.9842	5.0676	1.5234	0.9828	5.0783	1.9142
	9	0	0.9806	5.0846	0.0880	0.9808	5.0831	0.1588
		0.5	0.9808	5.0876	0.0742	0.9812	5.0842	0.1558
		0.9	0.9794	5.2621	-4.4998	0.9790	5.2020	-4.0242
	10	0	0.9804	5.0793	0.0000	0.9804	5.0793	0.0000
		0.5	0.9802	5.0834	-0.0182	0.9804	5.0809	-0.0086
		0.9	0.9792	5.2740	-5.1806	0.9792	5.2123	-4.6892
200	1	0	0.9760	4.2766	1.7322	0.9750	4.3042	2.2608
		0.5	0.9764	4.2799	1.8004	0.9752	4.3031	2.2700
		0.9	0.9758	4.3098	1.5284	0.9752	4.3243	1.9450
	9	0	0.9744	4.3547	0.0868	0.9746	4.3567	0.1616
		0.5	0.9742	4.3545	0.0882	0.9744	4.3558	0.1624
		0.9	0.9724	4.4999	-3.3820	0.9726	4.4541	-2.8298
	10	0	0.9740	4.3598	0.0000	0.9740	4.3598	0.0000
		0.5	0.9740	4.3598	0.0000	0.9740	4.3598	0.0000
		0.9	0.9728	4.5147	-3.9214	0.9730	4.4668	-3.3652
1000	1	0	0.9562	3.2716	1.7262	0.9560	3.2901	2.2778
		0.5	0.9546	3.2739	1.8216	0.9542	3.2901	2.2988
		0.9	0.9550	3.2761	1.8126	0.9540	3.2909	2.2830
	9	0	0.9546	3.3229	0.0816	0.9552	3.3246	0.1616
		0.5	0.9544	3.3229	0.0856	0.9544	3.3244	0.1576
		0.9	0.9534	3.3448	-0.3118	0.9536	3.3338	-0.0584
	10	0	0.9556	3.3283	0.0000	0.9556	3.3283	0.0000
		0.5	0.9556	3.3283	0.0000	0.9556	3.3283	0.0000
		0.9	0.9536	3.3531	-0.4700	0.9550	3.3387	-0.2510

5.2.4 Error type 4

Table 5.10. $p = 20$, *error type* $= 4$

			LASSO			RIDGE		
n	a	psi	coverage	length	penalty	coverage	length	penalty
200	1	0	0.9982	2.4205	0.0576	0.9866	2.5036	0.1094
		0.5	0.9978	2.4080	0.0183	0.9886	3.1587	0.2565
		0.9	0.9992	2.3978	0.0284	0.9978	2.5048	0.5142
	19	0	0.9856	2.4677	0.0070	0.9834	2.6787	0.1746
		0.5	0.9842	7.0022	1.1892	0.9872	7.0821	21.6156
		0.9	0.9866	12.1538	2.0630	0.9870	12.4390	45.4908
	20	0	0.9856	2.4703	0.0071	0.9834	2.6859	0.1778
		0.5	0.9836	7.4428	1.2763	0.9862	7.5363	24.9747
		0.9	0.9860	12.8003	2.1776	0.9866	13.0792	48.0181
400	1	0	0.9964	2.1784	0.0414	0.9850	2.2766	0.1097
		0.5	0.9976	2.1710	0.0171	0.9822	2.9235	0.2568
		0.9	0.9986	2.1667	0.0284	0.9922	2.2744	0.5332
	19	0	0.9850	2.2293	0.0071	0.9794	2.4025	0.1538
		0.5	0.9782	6.4046	1.1795	0.9816	6.5021	21.5611
		0.9	0.9792	11.2415	2.0654	0.9786	11.4380	45.5442
	20	0	0.9850	2.2314	0.0072	0.9810	2.4071	0.1559
		0.5	0.9776	6.7623	1.2569	0.9796	6.9301	24.9894
		0.9	0.9792	11.8362	2.1801	0.9788	12.0315	48.0743
2000	1	0	0.9686	1.9424	0.0186	0.9596	1.9998	0.1097
		0.5	0.9686	1.9432	0.0169	0.9586	2.5376	0.2571
		0.9	0.9676	1.9465	0.0285	0.9624	2.0167	0.5691
	19	0	0.9630	1.9517	0.0072	0.9584	2.0545	0.1278
		0.5	0.9574	5.4004	1.1782	0.9558	5.4893	21.5697
		0.9	0.9598	9.5338	2.0682	0.9558	9.6498	45.6055
	20	0	0.9638	1.9528	0.0073	0.9598	2.0558	0.1287
		0.5	0.9574	5.6501	1.2436	0.9578	5.8439	24.9855
		0.9	0.9580	10.0399	2.1831	0.9566	10.1477	48.1391

Table 5.11. $p = 20$, *error type* $= 4$

			PLS			PCR		
n	a	psi	coverage	length	penalty	coverage	length	penalty
200	1	0	0.9862	2.4621	13.3200	0.9860	2.4862	0.1076
		0.5	0.9874	2.4630	13.8342	0.9838	2.4944	0.8248
		0.9	0.9890	2.4612	16.8248	0.9960	2.4840	13.5116
	19	0	0.9868	2.4616	11.9970	0.9862	2.5079	0.0128
		0.5	0.9866	2.4624	13.8370	0.9856	2.4954	0.8518
		0.9	0.9888	2.4622	16.8392	0.9950	2.4849	13.5948
	20	0	0.9864	2.4616	11.9792	0.9862	2.5082	0.0138
		0.5	0.9898	2.4606	16.9512	0.9976	2.4205	15.0250
		0.9	0.9896	2.4606	16.9618	0.9986	2.4111	15.2100
400	1	0	0.9864	2.2220	13.6014	0.9864	2.2345	0.0334
		0.5	0.9868	2.2220	13.4870	0.9866	2.2368	0.2674
		0.9	0.9880	2.2208	16.6384	0.9926	2.2440	11.9354
	19	0	0.9868	2.2219	12.7944	0.9864	2.2409	0.0042
		0.5	0.9872	2.2221	13.4604	0.9846	2.2373	0.2860
		0.9	0.9872	2.2209	16.6488	0.9906	2.2440	11.9068
	20	0	0.9864	2.2219	12.7938	0.9864	2.2388	0.0028
		0.5	0.9878	2.2201	16.9692	0.9964	2.1778	15.2684
		0.9	0.9878	2.2199	16.9754	0.9976	2.1725	15.4520
2000	1	0	0.9634	1.9447	14.2694	0.9630	1.9452	0.0008
		0.5	0.9626	1.9447	13.5912	0.9626	1.9461	0.0184
		0.9	0.9634	1.9448	15.1278	0.9634	1.9495	2.3234
	19	0	0.9630	1.9447	13.6926	0.9630	1.9447	0.0000
		0.5	0.9628	1.9447	13.6106	0.9628	1.9460	0.0162
		0.9	0.9620	1.9447	15.1040	0.9626	1.9495	2.2778
	20	0	0.9626	1.9447	13.6742	0.9630	1.9447	0.0000
		0.5	0.9624	1.9447	16.9770	0.9672	1.9427	15.5024
		0.9	0.9626	1.9447	16.9836	0.9664	1.9421	15.7158

Table 5.12. $p = 20$, *error type* $= 4$

			Ii			I-Min		
n	a	psi	coverage	length	penalty	coverage	length	penalty
200	1	0	0.9956	2.4331	2.8434	0.9932	2.4429	3.8994
		0.5	0.9948	2.4346	2.9980	0.9924	2.4434	3.9148
		0.9	0.9930	2.4448	2.8534	0.9910	2.4480	3.8098
	19	0	0.9880	2.4605	0.0832	0.9876	2.4603	0.1640
		0.5	0.9864	2.4603	0.0814	0.9860	2.4602	0.1618
		0.9	0.9830	2.5714	-3.2118	0.9840	2.5302	-2.3506
	20	0	0.9864	2.4616	0.0000	0.9864	2.4616	0.0000
		0.5	0.9864	2.4616	0.0000	0.9864	2.4616	0.0000
		0.9	0.9830	2.5764	-3.5258	0.9832	2.5336	-2.6620
400	1	0	0.9930	2.1908	2.7674	0.9912	2.1998	3.8244
		0.5	0.9934	2.1923	2.9510	0.9918	2.1998	3.8662
		0.9	0.9942	2.1931	2.9484	0.9924	2.2003	3.8608
	19	0	0.9874	2.2204	0.0820	0.9876	2.2207	0.1548
		0.5	0.9866	2.2206	0.0804	0.9868	2.2209	0.1594
		0.9	0.9842	2.2457	-0.5718	0.9858	2.2313	-0.1816
	20	0	0.9864	2.2219	0.0000	0.9864	2.2219	0.0000
		0.5	0.9864	2.2219	0.0000	0.9864	2.2219	0.0000
		0.9	0.9840	2.2486	-0.7088	0.9858	2.2331	-0.3698
2000	1	0	0.9642	1.9426	2.7344	0.9672	1.9430	3.8056
		0.5	0.9660	1.9426	2.9172	0.9660	1.9430	3.8524
		0.9	0.9670	1.9426	2.9142	0.9666	1.9430	3.8568
	19	0	0.9628	1.9446	0.0844	0.9628	1.9446	0.1654
		0.5	0.9634	1.9446	0.0856	0.9632	1.9447	0.1662
		0.9	0.9634	1.9446	0.0862	0.9632	1.9447	0.1656
	20	0	0.9630	1.9447	0.0000	0.9630	1.9447	0.0000
		0.5	0.9630	1.9447	0.0000	0.9630	1.9447	0.0000
		0.9	0.9630	1.9447	0.0000	0.9630	1.9447	0.0000

5.2.5 Error type 5

Table 5.13. $p = 50$, *error type* $= 5$

			LASSO			RIDGE		
n	a	psi	coverage	length	penalty	coverage	length	penalty
500	1	0	0.9758	24.4032	0.2589	0.9730	23.6871	7.4834
		0.5	0.9762	24.3648	0.0434	0.9758	24.3581	1.2459
		0.9	0.9758	24.6002	0.0456	0.9764	24.6575	2.1344
	49	0	0.9698	22.4348	0.0076	0.9704	22.4989	0.2220
		0.5	0.9808	30.8675	4.5726	0.9820	31.4847	230.3264
		0.9	0.9886	52.3307	8.2588	0.9886	53.3699	491.2959
	50	0	0.9698	22.4328	0.0072	0.9700	22.4986	0.2217
		0.5	0.9786	31.4094	4.7065	0.9818	32.0013	239.0520
		0.9	0.9884	53.4085	8.4634	0.9888	54.3898	502.2058
1000	1	0	0.9694	23.1464	0.1855	0.9688	22.6177	2.2190
		0.5	0.9694	23.1346	0.0328	0.9688	23.0961	0.6114
		0.9	0.9700	23.2336	0.0456	0.9698	23.2634	2.3809
	49	0	0.9684	22.2102	0.0072	0.9682	22.2374	0.1738
		0.5	0.9710	28.5856	4.4955	0.9738	28.9385	228.1795
		0.9	0.9816	47.8869	8.1421	0.9810	49.3925	492.7493
	50	0	0.9686	22.2091	0.0072	0.9682	22.2369	0.1748
		0.5	0.9726	29.0773	4.6300	0.9740	29.4207	236.8647
		0.9	0.9816	48.8570	8.3412	0.9808	50.3181	503.4000
5000	1	0	0.9542	14.8694	0.0834	0.9546	14.7622	0.4135
		0.5	0.9552	14.8746	0.0261	0.9542	14.8697	0.3959
		0.9	0.9548	14.8844	0.0456	0.9542	14.8858	2.5505
	49	0	0.9544	14.7483	0.0072	0.9540	14.7515	0.1360
		0.5	0.9544	22.8692	4.4705	0.956	22.8481	227.795
		0.9	0.9506	39.9074	8.0520	0.9538	41.5863	494.0449
	50	0	0.9540	14.7478	0.0073	0.9538	14.7510	0.1363
		0.5	0.9534	23.3133	4.5913	0.9558	23.2098	234.3730
		0.9	0.9506	40.6480	8.2208	0.9538	42.3589	504.3374

Table 5.14. $p = 50$, *error type* $= 5$

			PLS			PCR		
n	a	psi	coverage	length	penalty	coverage	length	penalty
500	1	0	0.9700	22.5805	47.4532	0.9730	23.6569	28.1072
		0.5	0.9708	22.6512	46.9796	0.9758	24.3090	42.0884
		0.9	0.9708	22.6268	46.9996	0.9756	24.4456	44.8080
	49	0	0.9694	22.4335	42.5238	0.9694	22.4347	0.0542
		0.5	0.9712	22.6565	46.9800	0.9758	24.3149	42.0892
		0.9	0.9708	22.6268	46.9992	0.9764	24.4510	44.8520
	50	0	0.9694	22.4332	42.4880	0.9690	22.4335	0.0562
		0.5	0.9706	22.6253	46.9996	0.9760	24.4552	44.8904
		0.9	0.9706	22.6253	46.9994	0.9758	24.4533	44.8832
1000	1	0	0.9692	22.2150	46.8178	0.9672	22.4034	11.3746
		0.5	0.9684	22.2686	46.9288	0.9698	23.0245	39.0052
		0.9	0.9688	22.2598	46.9992	0.9700	23.1708	45.0028
	49	0	0.9682	22.2096	43.0464	0.9682	22.2097	0.0106
		0.5	0.9692	22.2727	46.9288	0.9686	23.0158	38.8416
		0.9	0.9692	22.2606	46.9994	0.9704	23.1692	45.0360
	50	0	0.9682	22.2096	43.0038	0.9682	22.2100	0.0108
		0.5	0.9692	22.2591	46.9994	0.9702	23.1707	45.1392
		0.9	0.9692	22.2591	46.9994	0.9702	23.1720	45.1586
5000	1	0	0.9546	14.7486	45.0298	0.9550	14.7492	0.3450
		0.5	0.9546	14.7485	45.3106	0.9552	14.7594	5.7424
		0.9	0.9552	14.7496	46.9996	0.9546	14.8731	44.9304
	49	0	0.9546	14.7486	43.8572	0.9546	14.7486	0.0000
		0.5	0.9546	14.7486	45.31	0.955	14.7578	5.7258
		0.9	0.9556	14.7500	46.9998	0.9548	14.8722	44.8754
	50	0	0.9546	14.7486	43.8502	0.9546	14.7486	0.0002
		0.5	0.9556	14.7496	47.0000	0.9546	14.8750	45.4536
		0.9	0.9554	14.7496	47.0000	0.9548	14.8754	45.4920

Table 5.15. $p = 50$, *error type* $= 5$

			Ii			I-Min		
n	a	psi	coverage	length	penalty	coverage	length	penalty
500	1	0	0.9738	23.6191	6.3784	0.9728	23.3927	8.5734
		0.5	0.9738	23.6044	6.5390	0.9732	23.3900	8.5750
		0.9	0.9736	23.6133	5.8324	0.9724	23.4058	7.8870
	49	0	0.9704	22.4600	0.0916	0.9700	22.4512	0.1714
		0.5	0.9708	22.2286	-2.4042	0.9706	22.3165	-1.4500
		0.9	0.9716	22.7959	-33.2134	0.9706	22.7720	-31.4164
	50	0	0.9692	22.4328	-0.0012	0.9692	22.4329	-0.0004
		0.5	0.9698	22.1960	-2.5964	0.9694	22.2966	-1.6686
		0.9	0.9716	22.7763	-33.9796	0.9710	22.7567	-32.1914
1000	1	0	0.9692	22.7667	6.3366	0.9692	22.6584	8.5568
		0.5	0.9700	22.7658	6.4900	0.9692	22.6609	8.5340
		0.9	0.9696	22.7649	5.8338	0.9692	22.6602	7.9064
	49	0	0.9686	22.2227	0.0896	0.9682	22.2191	0.1646
		0.5	0.9686	22.2136	-0.0078	0.9682	22.2161	0.1256
		0.9	0.9682	22.2265	-28.2904	0.9684	22.2407	-26.3920
	50	0	0.9682	22.2098	0.0000	0.9682	22.2098	0.0000
		0.5	0.9682	22.1999	-0.1024	0.9682	22.2065	-0.0386
		0.9	0.9690	22.2135	-28.9516	0.9682	22.2288	-27.0778
5000	1	0	0.9542	14.8215	6.2834	0.9540	14.8076	8.5384
		0.5	0.9552	14.8224	6.4320	0.9546	14.8071	8.5568
		0.9	0.9546	14.8208	6.2042	0.9546	14.8067	8.3478
	49	0	0.9546	14.7505	0.0818	0.9546	14.7500	0.1582
		0.5	0.9548	14.7506	0.0818	0.9548	14.7499	0.1610
		0.9	0.9534	14.7183	-10.7004	0.9530	14.7254	-8.7588
	50	0	0.9546	14.7486	0.0000	0.9546	14.7486	0.0000
		0.5	0.9546	14.7486	0.0000	0.9546	14.7486	0.0000
		0.9	0.9534	14.7163	-11.0292	0.9532	14.7244	-9.1016

5.3 NEW PREDICTION INTERVAL SIMULATION

Table 5.16. Etype = 1, J=10, k=1, $\psi = 0$

n	p	ψ		Lasso	RL	PLS	PCR	FS
100	20	0	Cov	0.9440	0.9468	0.9162	0.9386	0.9244
			Len	4.1888	4.2620	4.0218	5.0190	4.0739
100	40	0	Cov	0.9516	0.9424	0.8198	0.9510	0.9056
			Len	4.1679	4.1531	3.4835	5.4312	3.8646
100	100	0	Cov	0.9428	0.9168	0.2762	0.9464	0.8506
			Len	4.1594	3.9978	1.4048	5.7042	3.5653
100	200	0	Cov	0.9462	0.9148	0.0340	0.9502	0.8078
			Len	4.1878	3.9064	0.1381	5.8263	3.3569
400	20	0	Cov	0.9396	0.9420	0.9390	0.9390	0.9390
			Len	3.9155	3.9629	3.9155	3.9155	3.9155
400	40	0	Cov	0.9304	0.9438	0.9304	0.9304	0.9304
			Len	3.8128	3.9135	3.8127	3.8127	3.8127
400	100	0	Cov	0.9284	0.9264	0.8724	0.9204	0.8892
			Len	3.7120	3.7927	3.4769	4.5815	3.5604
400	200	0	Cov	0.9230	0.9068	0.6780	0.9226	0.8332
			Len	3.6776	3.6872	2.8430	4.8829	3.3347
1000	20	0	Cov	0.9484	0.9486	0.9470	0.9470	0.9470
			Len	3.9029	3.9209	3.9024	3.9024	3.9024
1000	40	0	Cov	0.9450	0.9514	0.9438	0.9438	0.9438
			Len	3.8616	3.8992	3.8615	3.8615	3.8615
1000	100	0	Cov	0.9252	0.9396	0.9248	0.9248	0.9248
			Len	3.7383	3.8389	3.7383	3.7383	3.7383
1000	200	0	Cov	0.9248	0.9314	0.8868	0.9162	0.8956
			Len	3.6537	3.7575	3.5263	4.3486	3.5590

Table 5.17. Etype = 1, J=10, k=1, $\psi = 1/\sqrt{p}$

n	p	ψ		Lasso	RL	PLS	PCR	FS
100	20	0.2236	Cov	0.9464	0.9600	0.9114	0.9440	0.9224
			Len	4.1943	4.3740	4.0244	4.7389	4.0838
100	40	0.1581	Cov	0.9502	0.9624	0.8076	0.9518	0.9016
			Len	4.1757	4.3547	3.4863	5.1220	3.8685
100	100	0.1	Cov	0.9504	0.9544	0.3054	0.9522	0.8532
			Len	4.1731	4.3061	1.4898	5.4600	3.5917
100	200	0.07	Cov	0.9444	0.9482	0.0506	0.9540	0.8180
			Len	4.1890	4.2468	0.1974	5.6412	3.3618
400	20	0.2236	Cov	0.9474	0.9504	0.9462	0.9462	0.9462
			Len	3.9164	3.9829	3.9158	3.9158	3.9158
400	40	0.1581	Cov	0.9290	0.9480	0.9272	0.9272	0.9272
			Len	3.8162	3.9670	3.8160	3.8160	3.8160
400	100	0.1	Cov	0.9300	0.9498	0.8714	0.9196	0.8862
			Len	3.7121	3.9625	3.4785	4.4436	3.5612
400	200	0.07	Cov	0.9200	0.9430	0.6846	0.9216	0.8368
			Len	3.6744	3.9545	2.8388	4.7549	3.3313
1000	20	0.2236	Cov	0.9370	0.9406	0.9358	0.9358	0.9358
			Len	3.9034	3.9297	3.9031	3.9031	3.9031
1000	40	0.1581	Cov	0.9454	0.9460	0.9460	0.9460	0.9460
			Len	3.8613	3.9215	3.8611	3.8611	3.8611
1000	100	0.1	Cov	0.9234	0.9448	0.9230	0.9230	0.9230
			Len	3.7374	3.8952	3.7372	3.7372	3.7372
1000	200	0.07	Cov	0.9188	0.9398	0.8868	0.9178	0.8866
			Len	3.6540	3.9039	3.5263	4.2717	3.5587

Table 5.18. Etype = 1, J=10, k=1, $\psi = 0.9$

n	p	ψ		Lasso	RL	PLS	PCR	FS
100	20	0.9	Cov	0.9638	0.9624	0.9102	0.9530	0.9198
			Len	4.4110	4.4052	4.0230	4.2754	4.0707
100	40	0.9	Cov	0.9632	0.9620	0.8164	0.9546	0.8972
			Len	4.4206	4.4006	3.4918	4.2898	3.8545
100	100	0.9	Cov	0.9622	0.9534	0.2912	0.9544	0.8392
			Len	4.8641	4.3694	1.4668	4.3063	3.5606
100	200	0.9	Cov	0.9710	0.9572	0.0528	0.9654	0.8084
			Len	6.4575	4.3565	0.1735	4.3384	3.3489
400	20	0.9	Cov	0.9510	0.9532	0.9396	0.9396	0.9396
			Len	4.0018	3.9980	3.9191	3.9191	3.9191
400	40	0.9	Cov	0.9506	0.9516	0.9272	0.9272	0.9272
			Len	4.0035	3.9919	3.8110	3.8110	3.8110
400	100	0.9	Cov	0.9514	0.9482	0.8652	0.9324	0.8832
			Len	4.0116	3.9785	3.4758	3.8157	3.5584
400	200	0.9	Cov	0.9532	0.9474	0.6806	0.9328	0.8314
			Len	4.0226	3.9769	2.8386	3.8246	3.3312
1000	20	0.9	Cov	0.9484	0.9486	0.9436	0.9436	0.9436
			Len	3.9367	3.9331	3.9030	3.9030	3.9030
1000	40	0.9	Cov	0.9488	0.9476	0.9376	0.9376	0.9376
			Len	3.9389	3.9298	3.8618	3.8618	3.8618
1000	100	0.9	Cov	0.9574	0.9542	0.9262	0.9262	0.9262
			Len	3.9480	3.9313	3.7410	3.7410	3.7410
1000	200	0.9	Cov	0.9512	0.9494	0.8922	0.9300	0.8940
			Len	3.9494	3.9239	3.5252	3.7451	3.5572

Table 5.19. Etype = 1, J=10, k=19, $\psi = 0$

n	p	ψ		Lasso	RL	PLS	PCR	FS
100	20	0	Cov	0.9828	0.9520	0.9666	0.9672	0.9430
			Len	19.6154	13.2498	4.9831	15.0115	13.8313
100	40	0	Cov	0.9800	0.9560	0.8970	0.9722	0.9398
			Len	19.6622	13.3625	4.2952	18.7955	13.8378
100	100	0	Cov	0.9824	0.9492	0.3466	0.9790	0.9384
			Len	19.5710	13.4861	2.1035	20.7464	13.6517
100	200	0	Cov	0.9780	0.9374	0.0408	0.9740	0.9070
			Len	19.1486	13.5431	0.3431	20.6873	13.0348
400	20	0	Cov	0.9764	0.9772	0.9772	0.9772	0.9772
			Len	4.7007	4.6990	4.6990	4.6990	4.6990
400	40	0	Cov	0.9708	0.9742	0.9700	0.9700	0.9700
			Len	4.5742	4.6335	4.5725	4.5725	4.5725
400	100	0	Cov	0.9608	0.9608	0.9222	0.9422	0.9426
			Len	4.6381	4.5222	4.1752	14.0265	4.3326
400	200	0	Cov	0.9654	0.9546	0.7672	0.9596	0.9248
			Len	4.7627	4.4426	3.4101	16.6155	4.1621
1000	20	0	Cov	0.9552	0.9554	0.9554	0.9554	0.9554
			Len	4.1784	4.1766	4.1766	4.1766	4.1766
1000	40	0	Cov	0.9562	0.9560	0.9560	0.9560	0.9560
			Len	4.1414	4.1605	4.1392	4.1392	4.1392
1000	100	0	Cov	0.9442	0.9520	0.9416	0.9416	0.9416
			Len	4.0090	4.0946	4.0067	4.0067	4.0067
1000	200	0	Cov	0.9428	0.9432	0.9044	0.9100	0.9132
			Len	3.9423	4.0049	3.7761	11.0840	3.8189

Table 5.20. Etype = 1, J=10, k=19, $\psi = 1/\sqrt{p}$

n	p	ψ		Lasso	RL	PLS	PCR	FS
100	20	0.2236	Cov	0.9900	0.9612	0.9652	0.9846	0.9378
			Len	74.0058	15.0151	4.9754	5.3109	14.6745
100	40	0.1581	Cov	0.9868	0.9574	0.9008	0.9808	0.9310
			Len	76.6257	17.6423	4.3069	12.2907	16.1301
100	100	0.1	Cov	0.9844	0.9338	0.3936	0.9808	0.9008
			Len	78.5815	20.1753	2.1790	17.4904	17.6116
100	200	0.07	Cov	0.9864	0.9262	0.0604	0.9822	0.8792
			Len	82.8101	22.4902	0.4862	19.6893	18.1515
400	20	0.2236	Cov	0.9734	0.9750	0.9750	0.9750	0.9750
			Len	4.7504	4.6969	4.6969	4.6969	4.6969
400	40	0.1581	Cov	0.9748	0.9750	0.9702	0.9702	0.9702
			Len	4.7065	4.6548	4.5788	4.5788	4.5788
400	100	0.1	Cov	0.9732	0.9682	0.9320	0.9546	0.9490
			Len	4.7559	4.5721	4.1774	11.7475	4.3342
400	200	0.07	Cov	0.9718	0.9636	0.7636	0.9542	0.9232
			Len	6.5855	4.5206	3.4146	14.9415	4.1719
1000	20	0.2236	Cov	0.9612	0.9598	0.9598	0.9598	0.9598
			Len	4.2192	4.1744	4.1744	4.1744	4.1744
1000	40	0.1581	Cov	0.9518	0.9524	0.9484	0.9484	0.9484
			Len	4.2074	4.1633	4.1367	4.1367	4.1367
1000	100	0.1	Cov	0.9554	0.9542	0.9392	0.9392	0.9392
			Len	4.1762	4.1344	4.0038	4.0038	4.0038
1000	200	0.07	Cov	0.9516	0.9474	0.9034	0.9150	0.9096
			Len	4.1528	4.1093	3.7772	10.0048	3.8203

Table 5.21. Etype = 1, J=10, k=19, $\psi = 0.9$

n	p	ψ		Lasso	RL	PLS	PCR	FS
100	20	0.9	Cov	0.9842	0.9710	0.9608	0.9780	0.9626
			Len	45.5575	5.2322	4.7840	5.0756	4.9286
100	40	0.9	Cov	0.9866	0.9712	0.8760	0.9814	0.9492
			Len	42.4335	5.5499	4.0903	5.1879	4.7386
100	100	0.9	Cov	0.9836	0.9604	0.3412	0.9796	0.9088
			Len	123.8216	5.4754	1.7120	5.2963	4.4809
100	200	0.9	Cov	0.9890	0.9564	0.0578	0.9786	0.8778
			Len	335.8509	5.1968	0.2096	5.3283	4.2461
400	20	0.9	Cov	0.9720	0.9602	0.9602	0.9602	0.9602
			Len	10.2370	4.6849	4.3789	4.3789	4.3789
400	40	0.9	Cov	0.9698	0.9616	0.9536	0.9536	0.9536
			Len	14.2814	5.0807	4.1792	4.1792	4.1792
400	100	0.9	Cov	0.9696	0.9582	0.8886	0.9468	0.9040
			Len	22.9928	5.1238	3.7880	4.2913	3.9122
400	200	0.9	Cov	0.9670	0.9564	0.7188	0.9508	0.8748
			Len	31.8672	5.0395	3.0756	4.3547	3.7014
1000	20	0.9	Cov	0.9628	0.9580	0.9590	0.9590	0.9590
			Len	9.6734	4.4579	4.1692	4.1692	4.1692
1000	40	0.9	Cov	0.9606	0.9546	0.9444	0.9444	0.9444
			Len	13.5609	4.9206	4.0799	4.0799	4.0799
1000	100	0.9	Cov	0.9580	0.9576	0.9376	0.9376	0.9376
			Len	21.9414	4.9994	3.9317	3.9317	3.9317
1000	200	0.9	Cov	0.9576	0.9546	0.8958	0.9366	0.9028
			Len	30.2701	4.9982	3.7022	4.0430	3.7452

Table 5.22. Etype = 1, J=10, k=p-1, $\psi = 0$

n	p	ψ		Lasso	RL	PLS	PCR	FS
100	40	0	Cov	0.9788	0.9320	0.8792	0.9714	0.9222
			Len	28.7166	21.4549	4.2040	25.6768	21.6615
100	100	0	Cov	0.9510	0.8676	0.2784	0.9492	0.8402
			Len	40.4687	30.9453	2.6627	39.3382	30.9364
100	200	0	Cov	0.9500	0.8434	0.0304	0.9442	0.7964
			Len	56.7788	43.8308	0.8607	56.3640	42.2023
400	40	0	Cov	0.9770	0.9786	0.9786	0.9786	0.9786
			Len	4.9127	4.9071	4.9071	4.9071	4.9071
400	100	0	Cov	0.9786	0.9398	0.9408	0.9670	0.9118
			Len	41.0935	29.3346	4.4795	32.7964	28.7049
400	200	0	Cov	0.9300	0.8498	0.6936	0.9140	0.8118
			Len	49.5575	37.7931	2.9462	44.8739	36.5594
1000	40	0	Cov	0.9760	0.9756	0.9756	0.9756	0.9756
			Len	4.4901	4.4868	4.4868	4.4868	4.4868
1000	100	0	Cov	0.9820	0.9816	0.9816	0.9816	0.9816
			Len	4.8973	4.8849	4.8849	4.8849	4.8849
1000	200	0	Cov	0.9772	0.9422	0.9562	0.9528	0.9296
			Len	56.0230	38.0741	4.6018	41.0338	37.1150

Table 5.23. Etype = 1, J=10, k=p-1, $\psi = 1/\sqrt{p}$

n	p	ψ		Lasso	RL	PLS	PCR	FS
100	40	0.1581	Cov	0.9872	0.9470	0.8962	0.9844	0.9044
			Len	160.2461	36.4689	4.3119	5.4261	33.5587
100	100	0.1	Cov	0.9868	0.9280	0.3576	0.9824	0.8650
			Len	411.7656	97.0153	1.8100	5.8605	85.7560
100	200	0.07	Cov	0.9872	0.9166	0.0642	0.9854	0.8184
			Len	871.3292	210.0977	0.2412	6.7370	166.1195
400	40	0.1581	Cov	0.9780	0.9774	0.9774	0.9774	0.9774
			Len	6.9829	4.9084	4.9084	4.9084	4.9084
400	100	0.1	Cov	0.9888	0.9556	0.9450	0.9790	0.8966
			Len	258.3051	45.6545	4.4795	4.9824	37.9356
400	200	0.07	Cov	0.9854	0.9408	0.7880	0.9800	0.8570
			Len	567.3311	104.8690	3.6487	5.1686	80.8306
1000	40	0.1581	Cov	0.9756	0.9694	0.9694	0.9694	0.9694
			Len	6.1908	4.4878	4.4878	4.4878	4.4878
1000	100	0.1	Cov	0.9830	0.9772	0.9772	0.9772	0.9772
			Len	16.3258	4.8815	4.8815	4.8815	4.8815
1000	200	0.07	Cov	0.9838	0.9606	0.9572	0.9814	0.9028
			Len	362.0462	56.1382	4.6009	4.9373	46.3527

Table 5.24. Etype = 1, J=10, k=p-1, $\psi = 0.9$

n	p	ψ		Lasso	RL	PLS	PCR	FS
100	40	0.9	Cov	0.9918	0.9686	0.8908	0.9846	0.9304
			Len	191.3079	7.7374	4.3038	5.2792	5.8268
100	100	0.9	Cov	0.9904	0.9624	0.3562	0.9870	0.8748
			Len	1635.0250	15.6466	1.7977	5.2822	10.5333
100	200	0.9	Cov	0.9902	0.9532	0.0578	0.9868	0.8394
			Len	7603.4220	28.5576	0.2127	5.2776	19.0490
400	40	0.9	Cov	0.9866	0.9760	0.9708	0.9708	0.9708
			Len	33.5314	7.6866	4.6538	4.6538	4.6538
400	100	0.9	Cov	0.9900	0.9766	0.9446	0.9820	0.9274
			Len	142.0953	16.9348	4.4710	4.9001	5.8303
400	200	0.9	Cov	0.9870	0.9754	0.7922	0.9784	0.8648
			Len	392.9536	31.1084	3.6533	4.9087	9.4875
1000	40	0.9	Cov	0.9794	0.9704	0.9680	0.9680	0.9680
			Len	31.4676	7.3765	4.4822	4.4822	4.4822
1000	100	0.9	Cov	0.9914	0.9822	0.9762	0.9762	0.9762
			Len	141.6904	17.3494	4.8835	4.8835	4.8835
1000	200	0.9	Cov	0.9898	0.9840	0.9584	0.9842	0.9222
			Len	391.2974	32.1430	4.6027	4.8814	6.4369

Table 5.25. Etype = 3, J=50, k=1, $\psi = 0$

n	p	ψ		Lasso	RL	PLS	PCR	FS
100	20	0	Cov	0.9624	0.9602	0.9386	0.9602	0.9638
			Len	3.9944	3.8048	4.1532	6.0891	3.7380
100	40	0	Cov	0.9614	0.9568	0.9120	0.9622	0.9614
			Len	4.0635	3.8167	4.0964	6.1446	3.7435
100	100	0	Cov	0.9650	0.9578	0.8504	0.9656	0.9692
			Len	4.1285	3.8177	3.7153	6.1969	3.7442
100	200	0	Cov	0.9592	0.9532	0.7758	0.9608	0.9612
			Len	4.1968	3.8145	3.1906	6.1966	3.7401
400	20	0	Cov	0.9476	0.9494	0.9410	0.9412	0.9462
			Len	3.2873	3.3461	3.4134	4.8534	3.3967
400	40	0	Cov	0.9514	0.9484	0.9314	0.9412	0.9424
			Len	3.2865	3.3750	3.4474	5.1800	3.4227
400	100	0	Cov	0.9440	0.9368	0.8710	0.9398	0.9308
			Len	3.2983	3.4080	3.3279	5.3826	3.4427
400	200	0	Cov	0.9484	0.9394	0.7072	0.9450	0.9316
			Len	3.3079	3.4203	2.8149	5.4615	3.4476
1000	20	0	Cov	0.9442	0.9474	0.9446	0.9446	0.9446
			Len	3.2496	3.2063	3.2584	3.2584	3.2584
1000	40	0	Cov	0.9466	0.9458	0.9410	0.9406	0.9424
			Len	3.2024	3.2551	3.3201	4.5350	3.3126
1000	100	0	Cov	0.9512	0.9480	0.9310	0.9448	0.9404
			Len	3.1813	3.3082	3.3811	5.0626	3.3602
1000	200	0	Cov	0.9488	0.9460	0.8894	0.9420	0.9346
			Len	3.1709	3.3313	3.3249	5.2323	3.3717

Table 5.26. Etype = 3, J=50, k=1, $\psi = 1/\sqrt{p}$

n	p	ψ		Lasso	RL	PLS	PCR	FS
100	20	0.2236	Cov	0.9674	0.9650	0.9622	0.9616	0.9682
			Len	4.3200	3.8327	5.3951	5.4754	3.7457
100	40	0.1581	Cov	0.9640	0.9582	0.9598	0.9602	0.9630
			Len	4.3323	3.8258	5.6169	5.6879	3.7268
100	100	0.1	Cov	0.9592	0.9586	0.9608	0.9622	0.9654
			Len	4.4333	3.8560	5.8365	5.8983	3.7444
100	200	0.07	Cov	0.9620	0.9564	0.9554	0.9568	0.9642
			Len	4.5215	3.8601	5.9394	6.0002	3.7434
400	20	0.2236	Cov	0.9496	0.9518	0.9430	0.9454	0.9426
			Len	3.2828	3.2824	3.4065	4.4104	3.3897
400	40	0.1581	Cov	0.9538	0.9552	0.9360	0.9504	0.9478
			Len	3.2816	3.2980	3.4459	4.8107	3.4179
400	100	0.1	Cov	0.9502	0.9514	0.8816	0.9480	0.9372
			Len	3.3054	3.3314	3.3274	5.1294	3.4463
400	200	0.07	Cov	0.9460	0.9462	0.7204	0.9446	0.9296
			Len	3.3126	3.3426	2.8288	5.2755	3.4453
1000	20	0.2236	Cov	0.9472	0.9490	0.9470	0.9470	0.9470
			Len	3.2464	3.1684	3.2589	3.2589	3.2589
1000	40	0.1581	Cov	0.9508	0.9528	0.9460	0.9450	0.9468
			Len	3.2020	3.1730	3.3185	4.2709	3.3108
1000	100	0.1	Cov	0.9464	0.9474	0.9288	0.9414	0.9392
			Len	3.1811	3.1880	3.3781	4.8283	3.3567
1000	200	0.07	Cov	0.9446	0.9482	0.8904	0.9410	0.9322
			Len	3.1721	3.1981	3.3247	5.0570	3.3711

Table 5.27. Etype = 3, J=50, k=1, $\psi = 0.9$

n	p	ψ		Lasso	RL	PLS	PCR	FS
100	20	0.9	Cov	0.9626	0.9590	0.9626	0.9626	0.9590
			Len	10.9414	3.7808	3.8323	3.8323	3.7985
100	40	0.9	Cov	0.9634	0.9626	0.9654	0.9654	0.9638
			Len	16.4666	3.7920	3.8418	3.8418	3.8058
100	100	0.9	Cov	0.9680	0.9608	0.9650	0.9650	0.9630
			Len	27.1510	3.8277	3.8333	3.8333	3.7936
100	200	0.9	Cov	0.9642	0.9548	0.9638	0.9638	0.9596
			Len	46.3657	3.8531	3.8387	3.8387	3.7939
400	20	0.9	Cov	0.9502	0.9486	0.9422	0.9492	0.9422
			Len	3.2605	3.2743	3.4101	3.3706	3.3958
400	40	0.9	Cov	0.9512	0.9510	0.9314	0.9510	0.9418
			Len	3.4655	3.2881	3.4455	3.3820	3.4289
400	100	0.9	Cov	0.9522	0.9512	0.8770	0.9530	0.9364
			Len	4.6748	3.3082	3.3300	3.3912	3.4523
400	200	0.9	Cov	0.9516	0.9450	0.7190	0.9486	0.9222
			Len	7.7790	3.3177	2.8191	3.3909	3.4567
1000	20	0.9	Cov	0.9538	0.9528	0.9510	0.9510	0.9510
			Len	3.1219	3.1452	3.2583	3.2583	3.2583
1000	40	0.9	Cov	0.9512	0.9504	0.9420	0.9480	0.9408
			Len	3.1219	3.1543	3.3182	3.2879	3.3102
1000	100	0.9	Cov	0.9494	0.9474	0.9302	0.9458	0.9370
			Len	3.1593	3.1692	3.3769	3.3095	3.3576
1000	200	0.9	Cov	0.9540	0.9508	0.8954	0.9512	0.9362
			Len	3.2081	3.1826	3.3277	3.3192	3.3751

Table 5.28. Etype = 3, J=50, k=19, $\psi = 0$

n	p	ψ		Lasso	RL	PLS	PCR	FS
100	20	0	Cov	0.9568	0.9254	0.9322	0.9612	0.9502
			Len	18.0945	14.2843	7.4280	18.9841	18.3152
100	40	0	Cov	0.9626	0.9278	0.8946	0.9672	0.9580
			Len	18.0581	14.2584	8.5690	19.3488	18.3198
100	100	0	Cov	0.9534	0.9136	0.8184	0.9584	0.9508
			Len	18.0820	14.2339	9.1398	19.6271	18.3224
100	200	0	Cov	0.9678	0.9460	0.7438	0.9684	0.9606
			Len	19.0913	16.1341	8.5696	19.6777	18.2836
400	20	0	Cov	0.9548	0.9390	0.9542	0.9422	0.9384
			Len	16.9204	11.3395	3.6648	14.2100	13.7638
400	40	0	Cov	0.9550	0.9358	0.9446	0.9510	0.9332
			Len	16.8844	11.3228	3.7000	16.3942	13.7577
400	100	0	Cov	0.9568	0.9416	0.8952	0.9558	0.9398
			Len	16.8874	11.2838	3.5654	17.6544	13.7700
400	200	0	Cov	0.9542	0.9418	0.7514	0.9552	0.9368
			Len	16.8928	11.2742	3.0564	18.1475	13.7715
1000	20	0	Cov	0.9618	0.9620	0.9620	0.9620	0.9620
			Len	3.5654	3.5587	3.5587	3.5587	3.5587
1000	40	0	Cov	0.9684	0.9638	0.9618	0.9502	0.9654
			Len	4.0219	3.5686	3.6227	12.7115	3.5575
1000	100	0	Cov	0.9590	0.9602	0.9456	0.9482	0.9624
			Len	4.2084	3.5768	3.6801	16.1112	3.5598
1000	200	0	Cov	0.9598	0.9606	0.9170	0.9562	0.9620
			Len	4.3099	3.5756	3.6105	17.2043	3.5552

Table 5.29. Etype = 3, J=50, k=19, $\psi = 1/\sqrt{p}$

n	p	ψ		Lasso	RL	PLS	PCR	FS
100	20	0.2236	Cov	0.9648	0.9216	0.9650	0.9646	0.9364
			Len	70.0260	16.0576	3.9333	3.9585	46.7489
100	40	0.1581	Cov	0.9604	0.9084	0.9662	0.9670	0.9316
			Len	69.0525	17.3747	12.2804	12.5084	48.8759
100	100	0.1	Cov	0.9596	0.8822	0.9634	0.9648	0.9320
			Len	67.5682	18.1072	16.0544	16.2718	50.1554
100	200	0.07	Cov	0.9626	0.8858	0.9656	0.9662	0.9288
			Len	78.0310	25.2198	17.4941	17.7004	50.2342
400	20	0.2236	Cov	0.9588	0.9524	0.9616	0.9628	0.9310
			Len	65.8737	10.5683	3.6687	3.6172	16.5396
400	40	0.1581	Cov	0.9600	0.9476	0.9422	0.9548	0.9262
			Len	64.8400	11.3780	3.7012	10.6168	17.8030
400	100	0.1	Cov	0.9580	0.9474	0.8962	0.9548	0.9276
			Len	63.4644	12.3146	3.5616	14.6268	18.9488
400	200	0.07	Cov	0.9568	0.9372	0.7402	0.9592	0.9268
			Len	61.9664	12.8776	3.0998	16.2209	19.6442
1000	20	0.2236	Cov	0.9660	0.9630	0.9630	0.9630	0.9630
			Len	3.7035	3.5576	3.5576	3.5576	3.5576
1000	40	0.1581	Cov	0.9626	0.9566	0.9546	0.9436	0.9568
			Len	18.3475	3.5911	3.6191	8.4917	3.5561
1000	100	0.1	Cov	0.9634	0.9560	0.9372	0.9500	0.9562
			Len	31.4030	3.7841	3.6748	13.3632	3.5626
1000	200	0.07	Cov	0.9666	0.9596	0.9176	0.9538	0.9630
			Len	40.4800	4.2393	3.6119	15.3912	3.5693

Table 5.30. Etype = 3, J=50, k=19, $\psi = 0.9$

n	p	ψ		Lasso	RL	PLS	PCR	FS
100	20	0.9	Cov	0.9706	0.9388	0.9650	0.9650	0.9366
			Len	230.6223	5.0067	3.7450	3.7450	8.1529
100	40	0.9	Cov	0.9642	0.9400	0.9662	0.9662	0.9270
			Len	329.9553	4.9791	4.2715	4.2715	8.0814
100	100	0.9	Cov	0.9694	0.9292	0.9618	0.9618	0.9308
			Len	523.5096	4.7851	4.4544	4.4544	7.9819
100	200	0.9	Cov	0.9680	0.9298	0.9610	0.9610	0.9178
			Len	893.6305	4.7265	4.5118	4.5118	7.8609
400	20	0.9	Cov	0.9636	0.9572	0.9576	0.9646	0.9496
			Len	96.2901	4.3825	3.6590	3.5715	4.3506
400	40	0.9	Cov	0.9594	0.9518	0.9424	0.9584	0.9414
			Len	69.2012	4.7417	3.6920	3.9697	4.3524
400	100	0.9	Cov	0.9588	0.9498	0.8956	0.9576	0.9414
			Len	124.7301	4.8335	3.5631	4.1959	4.3771
400	200	0.9	Cov	0.9614	0.9508	0.7554	0.9598	0.9378
			Len	215.4489	4.7771	3.0108	4.2744	4.3809
1000	20	0.9	Cov	0.9664	0.9598	0.9602	0.9602	0.9602
			Len	9.6969	4.1974	3.5466	3.5466	3.5466
1000	40	0.9	Cov	0.9582	0.9586	0.9592	0.9614	0.9592
			Len	13.5855	4.8134	3.5632	3.7576	3.5731
1000	100	0.9	Cov	0.9624	0.9560	0.9394	0.9592	0.9514
			Len	21.9236	4.9075	3.5995	4.0062	3.6186
1000	200	0.9	Cov	0.9606	0.9528	0.9108	0.9584	0.9504
			Len	30.2919	4.9066	3.5292	4.1001	3.6562

Table 5.31. Etype = 3, J=50, k=p-1, $\psi = 0$

n	p	ψ		Lasso	RL	PLS	PCR	FS
100	40	0	Cov	0.9584	0.9106	0.8906	0.9646	0.9508
			Len	25.9217	21.2087	11.6953	27.3291	26.2453
400	40	0	Cov	0.9554	0.9288	0.9424	0.9502	0.9292
			Len	24.7895	19.5493	3.7066	23.0648	21.8866
1000	40	0	Cov	0.9580	0.9430	0.9576	0.9438	0.9420
			Len	23.7430	15.2729	3.6203	17.6395	17.2337
100	100	0	Cov	0.9500	0.8800	0.8100	0.9582	0.9376
			Len	41.2669	34.0048	19.7201	43.8646	41.9770
400	100	0	Cov	0.9542	0.9170	0.8954	0.9554	0.9360
			Len	40.0461	34.1398	3.5688	39.3103	36.8825
1000	100	0	Cov	0.9576	0.9332	0.9496	0.9526	0.9314
			Len	39.4560	32.0500	3.6792	35.7837	33.7183
400	200	0	Cov	0.9462	0.9108	0.6976	0.9472	0.9226
			Len	55.1043	47.9555	3.3893	55.1655	51.5128
1000	200	0	Cov	0.9622	0.9306	0.9176	0.9552	0.9280
			Len	56.7019	48.7998	3.6097	54.1929	50.5728

Table 5.32. Etype = 3, J=50, k=p-1, $\psi = 1/\sqrt{p}$

n	p	ψ		Lasso	RL	PLS	PCR	FS
100	40	0.1581	Cov	0.9642	0.9006	0.9658	0.9662	0.9400
			Len	140.3552	34.2855	4.1832	4.2165	99.0408
100	100	0.1	Cov	0.9584	0.8626	0.9618	0.9632	0.9218
			Len	345.5432	85.3149	4.8078	4.8526	254.1073
100	200	0.07	Cov	0.9640	0.8790	0.9580	0.9588	0.9194
			Len	799.7382	236.6724	5.6739	5.7304	506.2869
400	40	0.1581	Cov	0.9588	0.9376	0.9476	0.9606	0.9162
			Len	131.8074	24.5767	3.6941	3.6915	38.8234
400	100	0.1	Cov	0.9554	0.9284	0.8930	0.9616	0.9094
			Len	324.7592	65.1255	3.5620	3.9346	105.7167
400	200	0.07	Cov	0.9584	0.9152	0.7336	0.9606	0.8996
			Len	636.9782	130.6055	3.0075	4.2697	215.5844
1000	40	0.1581	Cov	0.9640	0.9520	0.9582	0.9596	0.9346
			Len	128.9778	18.0767	3.6194	3.5833	20.0816
1000	100	0.1	Cov	0.9550	0.9414	0.9488	0.9618	0.9164
			Len	320.4386	52.9085	3.6781	3.7016	63.0990
1000	200	0.07	Cov	0.9582	0.9364	0.9136	0.9594	0.9086
			Len	625.8618	107.7568	3.6102	3.8816	133.2308

Table 5.33. Etype = 3, J=50, k=p-1, $\psi = 0.9$

n	p	ψ		Lasso	RL	PLS	PCR	FS
100	40	0.9	Cov	0.9660	0.9296	0.9608	0.9608	0.9328
			Len	675.2869	7.2529	3.7490	3.7490	15.0066
400	40	0.9	Cov	0.9676	0.9480	0.9472	0.9624	0.9248
			Len	295.8083	6.6366	3.6964	3.5787	6.0805
1000	40	0.9	Cov	0.9594	0.9532	0.9566	0.9580	0.9478
			Len	28.9976	6.8519	3.6209	3.5550	4.4977
100	100	0.9	Cov	0.9684	0.9194	0.9658	0.9658	0.9214
			Len	2723.3590	14.6029	3.7405	3.7405	36.3488
400	100	0.9	Cov	0.9642	0.9522	0.8908	0.9616	0.9048
			Len	1849.4080	13.7998	3.5630	3.5818	12.8216
1000	100	0.9	Cov	0.9648	0.9504	0.9400	0.9596	0.9210
			Len	122.0355	14.1967	3.6803	3.5590	8.0451
100	200	0.9	Cov	0.9672	0.9192	0.9640	0.9640	0.9152
			Len	9360.4090	28.6047	3.7406	3.7406	71.3896
400	200	0.9	Cov	0.9626	0.9390	0.7484	0.9638	0.8986
			Len	6157.3260	25.8033	3.0048	3.5753	24.7109
1000	200	0.9	Cov	0.9618	0.9456	0.9212	0.9656	0.9040
			Len	425.1361	26.2287	3.6143	3.5604	15.0387

5.4 EBIC WITH FORWARD SELECTION SIMULATION

Table 5.34. Etype = 1, J=5, k=1

n	p		$\psi = 0$	$\psi = \frac{1}{\sqrt{p}}$	$\psi = 0.9$
100	20	Cov	0.9634	0.9650	0.9620
		Len	4.4384	4.4430	4.3939
100	40	Cov	0.9600	0.9618	0.9600
		Len	4.4467	4.4393	4.3791
100	100	Cov	0.9658	0.9630	0.9580
		Len	4.4396	4.4280	4.3682
100	200	Cov	0.9606	0.9652	0.9598
		Len	4.4219	4.4256	4.3427
400	20	Cov	0.9538	0.9484	0.9548
		Len	4.0071	4.0058	3.9979
400	40	Cov	0.9446	0.9538	0.9524
		Len	4.0055	4.0064	3.9944
400	100	Cov	0.9496	0.9490	0.9484
		Len	4.0080	4.0094	3.9929
400	200	Cov	0.9522	0.9534	0.9518
		Len	4.0033	4.0061	3.9870
1000	20	Cov	0.9470	0.9414	0.9496
		Len	3.9379	3.9382	3.9394
1000	40	Cov	0.9446	0.9488	0.9494
		Len	3.9367	3.9368	3.9372
1000	100	Cov	0.9464	0.9512	0.9494
		Len	3.9349	3.9352	3.9336
1000	200	Cov	0.9506	0.9528	0.9530
		Len	3.9344	3.9393	3.9352

Table 5.35. Etype = 1, J=5, k=19

n	p		$\psi = 0$	$\psi = \frac{1}{\sqrt{p}}$	$\psi = 0.9$
100	20	Cov	0.9832	0.9862	0.9616
		Len	5.7094	5.7056	5.1729
100	40	Cov	0.9796	0.9706	0.9516
		Len	5.9440	6.7200	5.1444
100	100	Cov	0.9680	0.9306	0.9334
		Len	6.6805	8.7804	5.0738
100	200	Cov	0.9392	0.8760	0.9110
		Len	8.3810	10.4374	4.9965
400	20	Cov	0.9758	0.9792	0.9634
		Len	4.6929	4.6887	4.4468
400	40	Cov	0.9756	0.9702	0.9502
		Len	4.6857	4.6855	4.4108
400	100	Cov	0.9746	0.9736	0.9388
		Len	4.6774	4.6791	4.3971
400	200	Cov	0.9756	0.9708	0.9414
		Len	4.6730	4.6753	4.3848
1000	20	Cov	0.9604	0.9662	0.9610
		Len	4.1742	4.1791	4.1663
1000	40	Cov	0.9602	0.9578	0.9554
		Len	4.1781	4.1782	4.1481
1000	100	Cov	0.9592	0.9580	0.9562
		Len	4.1775	4.1778	4.1539
1000	200	Cov	0.9598	0.9576	0.9504
		Len	4.1739	4.1783	4.1495

Table 5.36. Etype = 1, J=5, k=p-1

n	p		$\psi = 0$	$\psi = \frac{1}{\sqrt{p}}$	$\psi = 0.9$
100	40	Cov	0.9040	0.8858	0.9278
		Len	17.4837	19.4947	5.4377
100	100	Cov	0.9394	0.8000	0.8230
		Len	40.6694	50.2127	6.8075
100	200	Cov	0.9452	0.7186	0.7248
		Len	58.9667	92.8953	10.7159
400	40	Cov	0.9792	0.9824	0.9614
		Len	4.8971	4.9038	4.8126
400	100	Cov	0.9252	0.9212	0.9366
		Len	17.7642	17.8330	5.2853
400	200	Cov	0.9384	0.8320	0.8642
		Len	53.8183	48.1354	6.4799
1000	40	Cov	0.9696	0.9710	0.9702
		Len	4.4919	4.4908	4.4896
1000	100	Cov	0.9764	0.9748	0.9790
		Len	4.8752	4.8825	4.8765
1000	200	Cov	0.9826	0.9800	0.9826
		Len	5.2696	5.2731	5.2713

Table 5.37. Etype = 2, J=5, k=1

n	p		$\psi = 0$	$\psi = \frac{1}{\sqrt{p}}$	$\psi = 0.9$
100	20	Cov	0.9564	0.9538	0.9496
		Len	7.2198	7.2275	7.1927
100	40	Cov	0.9510	0.9578	0.9504
		Len	7.2597	7.2496	7.1912
100	100	Cov	0.9492	0.9534	0.9462
		Len	7.2354	7.2028	7.1998
100	200	Cov	0.9490	0.9496	0.9500
		Len	7.2521	7.2308	7.1480
400	20	Cov	0.9470	0.9446	0.9454
		Len	6.4632	6.4483	6.4468
400	40	Cov	0.9516	0.9520	0.9458
		Len	6.4629	6.4492	6.4404
400	100	Cov	0.9528	0.9460	0.9496
		Len	6.4616	6.4504	6.4514
400	200	Cov	0.9494	0.9454	0.9440
		Len	6.4509	6.4768	6.4113
1000	20	Cov	0.9482	0.9500	0.9466
		Len	6.3494	6.3564	6.3449
1000	40	Cov	0.9454	0.9438	0.9430
		Len	6.3465	6.3578	6.3420
1000	100	Cov	0.9480	0.9498	0.9480
		Len	6.3603	6.3612	6.3478
1000	200	Cov	0.9550	0.9482	0.9438
		Len	6.3571	6.3606	6.3464

Table 5.38. Etype = 2, J=5, k=19

n	p		$\psi = 0$	$\psi = \frac{1}{\sqrt{p}}$	$\psi = 0.9$
100	20	Cov	0.9760	0.9780	0.9620
		Len	10.0696	10.0334	8.2869
100	40	Cov	0.9726	0.9676	0.9520
		Len	10.0310	10.0748	8.0663
100	100	Cov	0.9700	0.9712	0.9420
		Len	10.3542	10.0511	7.8225
100	200	Cov	0.9612	0.9692	0.9316
		Len	11.2014	10.2027	7.6814
400	20	Cov	0.9712	0.9754	0.9532
		Len	8.4250	8.4276	6.9967
400	40	Cov	0.9688	0.9714	0.9488
		Len	8.3983	8.3948	6.9465
400	100	Cov	0.9698	0.9738	0.9398
		Len	8.3662	8.3874	6.9018
400	200	Cov	0.9738	0.9710	0.9414
		Len	8.3995	8.4170	6.8531
1000	20	Cov	0.9580	0.9606	0.9560
		Len	6.9778	6.9918	6.7105
1000	40	Cov	0.9598	0.9572	0.9518
		Len	7.0023	7.0053	6.6845
1000	100	Cov	0.9606	0.9530	0.9506
		Len	6.9960	6.9858	6.6595
1000	200	Cov	0.9590	0.9622	0.9470
		Len	6.9924	6.9981	6.6346

Table 5.39. Etype = 2, J=5, k=p-1

n	p		$\psi = 0$	$\psi = \frac{1}{\sqrt{p}}$	$\psi = 0.9$
100	40	Cov	0.9098	0.9060	0.9490
		Len	18.8294	20.3581	8.7305
400	40	Cov	0.9720	0.9720	0.9536
		Len	8.7203	8.7201	7.6968
1000	40	Cov	0.9732	0.9710	0.9538
		Len	7.8911	7.9020	7.1201
100	100	Cov	0.9300	0.8096	0.8702
		Len	41.0859	50.5781	9.1799
400	100	Cov	0.9288	0.9202	0.9394
		Len	18.8219	18.8528	8.5608
1000	100	Cov	0.9730	0.9722	0.9668
		Len	8.7675	8.7732	8.3899
100	200	Cov	0.9490	0.7192	0.7474
		Len	59.2288	93.0165	11.7748
400	200	Cov	0.9190	0.9062	0.9354
		Len	19.2643	21.6358	8.6117
1000	200	Cov	0.9748	0.9752	0.9502
		Len	9.3328	9.3402	8.8941

Table 5.40. Etype = 3, J=5, k=1

n	p		$\psi = 0$	$\psi = \frac{1}{\sqrt{p}}$	$\psi = 0.9$
100	20	Cov	0.9612	0.9646	0.9578
		Len	3.7945	3.7679	3.8340
100	40	Cov	0.9624	0.9634	0.9648
		Len	3.7741	3.7681	3.8388
100	100	Cov	0.9618	0.9662	0.9608
		Len	3.7837	3.7770	3.8520
100	200	Cov	0.9646	0.9610	0.9566
		Len	3.7816	3.7656	3.8651
400	20	Cov	0.9562	0.9556	0.9510
		Len	3.2180	3.2105	3.2729
400	40	Cov	0.9578	0.9576	0.9496
		Len	3.2167	3.2142	3.2913
400	100	Cov	0.9564	0.9596	0.9468
		Len	3.2142	3.2166	3.3162
400	200	Cov	0.9498	0.9522	0.9494
		Len	3.2130	3.2136	3.3201
1000	20	Cov	0.9528	0.9488	0.9524
		Len	3.0971	3.0988	3.1190
1000	40	Cov	0.9506	0.9534	0.9522
		Len	3.0988	3.0992	3.1343
1000	100	Cov	0.9518	0.9486	0.9502
		Len	3.0978	3.0953	3.1499
1000	200	Cov	0.9530	0.9498	0.9484
		Len	3.0976	3.1002	3.1649

Table 5.41. Etype = 3, J=5, k=19

n	p		$\psi = 0$	$\psi = \frac{1}{\sqrt{p}}$	$\psi = 0.9$
100	20	Cov	0.9778	0.9834	0.9576
		Len	5.6294	5.6451	5.2388
100	40	Cov	0.9782	0.9674	0.9498
		Len	5.9242	6.6955	5.2174
100	100	Cov	0.9690	0.9378	0.9392
		Len	6.7217	8.8035	5.1545
100	200	Cov	0.9412	0.8788	0.9190
		Len	8.4551	10.4310	5.0499
400	20	Cov	0.9756	0.9756	0.9570
		Len	4.3314	4.3278	4.1679
400	40	Cov	0.9728	0.9744	0.9542
		Len	4.3384	4.3435	4.2074
400	100	Cov	0.9708	0.9752	0.9482
		Len	4.3399	4.3523	4.2441
400	200	Cov	0.9752	0.9744	0.9452
		Len	4.3357	4.3480	4.2610
1000	20	Cov	0.9598	0.9596	0.9582
		Len	3.5601	3.5606	3.5657
1000	40	Cov	0.9578	0.9604	0.9526
		Len	3.5707	3.5722	3.6925
1000	100	Cov	0.9592	0.9574	0.9554
		Len	3.5653	3.5739	3.7609
1000	200	Cov	0.9600	0.9608	0.9460
		Len	3.5675	3.5706	3.7959

Table 5.42. Etype = 3, J=5, k=p-1

n	p		$\psi = 0$	$\psi = \frac{1}{\sqrt{p}}$	$\psi = 0.9$
100	40	Cov	0.9068	0.8876	0.9374
		Len	17.4717	19.4589	5.5039
400	40	Cov	0.9760	0.9754	0.9660
		Len	4.6697	4.6777	4.7592
1000	40	Cov	0.9706	0.9684	0.9674
		Len	4.0441	4.0353	4.0387
100	100	Cov	0.9366	0.7986	0.8118
		Len	40.6862	50.2978	6.8339
400	100	Cov	0.9116	0.9272	0.9408
		Len	17.7940	17.8558	5.3550
1000	100	Cov	0.9800	0.9774	0.9776
		Len	4.6590	4.6526	4.6559
100	200	Cov	0.9422	0.7270	0.7302
		Len	58.9611	92.9103	10.7338
400	200	Cov	0.9062	0.8964	0.9176
		Len	18.1816	20.8019	5.5971
1000	200	Cov	0.9778	0.9766	0.9774
		Len	5.2097	5.2074	5.2125

Table 5.43. Etype = 4, J=5, k=1

n	p		$\psi = 0$	$\psi = \frac{1}{\sqrt{p}}$	$\psi = 0.9$
100	20	Cov	0.9896	0.9910	0.9792
		Len	2.2083	2.2065	2.2245
100	40	Cov	0.9914	0.9922	0.9752
		Len	2.2081	2.2071	2.2274
100	100	Cov	0.9900	0.9914	0.9794
		Len	2.2070	2.2052	2.2238
100	200	Cov	0.9910	0.9924	0.9706
		Len	2.2049	2.2071	2.2210
400	20	Cov	0.9702	0.9702	0.9652
		Len	1.9623	1.9624	1.9673
400	40	Cov	0.9690	0.9674	0.9680
		Len	1.9621	1.9626	1.9699
400	100	Cov	0.9714	0.9704	0.9694
		Len	1.9627	1.9624	1.9736
400	200	Cov	0.9682	0.9678	0.9648
		Len	1.9626	1.9619	1.9770
1000	20	Cov	0.9574	0.9562	0.9570
		Len	1.9188	1.9189	1.9189
1000	40	Cov	0.9564	0.9548	0.9576
		Len	1.9187	1.9188	1.9190
1000	100	Cov	0.9530	0.9584	0.9550
		Len	1.9192	1.9191	1.9191
1000	200	Cov	0.9532	0.9562	0.9506
		Len	1.9192	1.9187	1.9190

Table 5.44. Etype = 4, J=5, k=19

n	p		$\psi = 0$	$\psi = \frac{1}{\sqrt{p}}$	$\psi = 0.9$
100	20	Cov	0.9912	0.9906	0.9714
		Len	2.9663	2.9670	2.9300
100	40	Cov	0.9888	0.9770	0.9428
		Len	3.3131	4.7542	3.0338
100	100	Cov	0.9772	0.9398	0.9108
		Len	4.3395	7.5802	3.0910
100	200	Cov	0.9550	0.8766	0.8896
		Len	5.9387	9.5670	3.1067
400	20	Cov	0.9834	0.9848	0.9828
		Len	2.2223	2.2216	2.2229
400	40	Cov	0.9870	0.9796	0.9738
		Len	2.2303	2.2313	2.3126
400	100	Cov	0.9810	0.9810	0.9652
		Len	2.2299	2.2308	2.3874
400	200	Cov	0.9816	0.9804	0.9558
		Len	2.2304	2.2335	2.4401
1000	20	Cov	0.9734	0.9666	0.9678
		Len	2.0060	2.0062	2.0056
1000	40	Cov	0.9682	0.9620	0.9644
		Len	2.0081	2.0083	2.0089
1000	100	Cov	0.9662	0.9660	0.9694
		Len	2.0083	2.0086	2.0132
1000	200	Cov	0.9656	0.9680	0.9670
		Len	2.0088	2.0083	2.0181

Table 5.45. Etype = 4, J=5, k=p-1

n	p		$\psi = 0$	$\psi = \frac{1}{\sqrt{p}}$	$\psi = 0.9$
100	40	Cov	0.9060	0.8960	0.9210
		Len	17.0568	19.1970	3.4131
400	40	Cov	0.9854	0.9848	0.9846
		Len	2.4102	2.4096	2.4109
1000	40	Cov	0.9772	0.9776	0.9794
		Len	2.1266	2.1262	2.1260
100	100	Cov	0.9376	0.8008	0.8120
		Len	40.5312	50.1514	6.0249
400	100	Cov	0.9132	0.9136	0.9422
		Len	17.4085	17.4846	3.1723
1000	100	Cov	0.9836	0.9854	0.9826
		Len	2.3854	2.3852	2.3844
100	200	Cov	0.9448	0.7238	0.7258
		Len	58.8565	93.0229	10.3869
400	200	Cov	0.9334	0.8280	0.8462
		Len	53.6547	48.0680	5.6165
1000	200	Cov	0.9836	0.9846	0.9838
		Len	2.7135	2.7143	2.7150

Table 5.46. Etype = 5, J=5, k=1

n	p		$\psi = 0$	$\psi = \frac{1}{\sqrt{p}}$	$\psi = 0.9$
100	20	Cov	0.9416	0.9418	0.9430
		Len	13.7744	13.7533	14.0450
100	40	Cov	0.9412	0.9424	0.9498
		Len	13.7591	13.8346	14.2835
100	100	Cov	0.9440	0.9450	0.9466
		Len	13.6856	13.5354	14.1634
100	200	Cov	0.9368	0.9426	0.9456
		Len	13.4667	13.5818	14.4944
400	20	Cov	0.9444	0.9462	0.9366
		Len	12.5336	12.5414	12.5615
400	40	Cov	0.9398	0.9400	0.9420
		Len	12.5302	12.5153	12.5147
400	100	Cov	0.9446	0.9424	0.9470
		Len	12.4439	12.4857	12.5657
400	200	Cov	0.9396	0.9400	0.9478
		Len	12.4129	12.4970	12.5170
1000	20	Cov	0.9422	0.9480	0.9426
		Len	12.7085	12.6851	12.6970
1000	40	Cov	0.9464	0.9460	0.9444
		Len	12.6467	12.6349	12.6916
1000	100	Cov	0.9422	0.9432	0.9450
		Len	12.6234	12.6757	12.6424
1000	200	Cov	0.9504	0.9424	0.9474
		Len	12.7262	12.6543	12.5888

Table 5.47. Etype = 5, J=5, k=19

n	p		$\psi = 0$	$\psi = \frac{1}{\sqrt{p}}$	$\psi = 0.9$
100	20	Cov	0.9568	0.9706	0.9534
		Len	21.8404	22.0011	15.9242
100	40	Cov	0.9476	0.9560	0.9510
		Len	20.2484	19.9555	15.3632
100	100	Cov	0.9440	0.9222	0.9414
		Len	21.1782	19.8807	15.0603
100	200	Cov	0.9420	0.8844	0.9438
		Len	21.8344	20.3214	14.6771
400	20	Cov	0.9718	0.9702	0.9456
		Len	21.2147	21.2535	13.4103
400	40	Cov	0.9724	0.9638	0.9394
		Len	21.0893	21.1343	13.3680
400	100	Cov	0.9700	0.9690	0.9450
		Len	20.9233	20.8773	13.0638
400	200	Cov	0.9630	0.9644	0.9422
		Len	20.8480	20.9003	13.0574
1000	20	Cov	0.9566	0.9582	0.9458
		Len	15.7497	15.7355	13.2877
1000	40	Cov	0.9518	0.9538	0.9436
		Len	15.8588	15.8285	13.1573
1000	100	Cov	0.9534	0.9524	0.9442
		Len	15.7830	15.8592	13.0396
1000	200	Cov	0.9574	0.9548	0.9454
		Len	15.7961	15.7520	12.9725

Table 5.48. Etype = 5, J=5, k=p-1

n	p		$\psi = 0$	$\psi = \frac{1}{\sqrt{p}}$	$\psi = 0.9$
100	40	Cov	0.9318	0.9224	0.9504
		Len	25.2335	24.6307	17.4242
400	40	Cov	0.9696	0.9676	0.9474
		Len	21.7556	21.6811	14.5232
1000	40	Cov	0.9676	0.9630	0.9484
		Len	19.8813	19.8024	13.9840
100	100	Cov	0.9438	0.8054	0.9228
		Len	43.2638	51.3765	17.6832
400	100	Cov	0.9502	0.9470	0.9546
		Len	23.7750	23.7304	17.5792
1000	100	Cov	0.9742	0.9716	0.9584
		Len	22.7117	22.6312	16.0372
100	200	Cov	0.9486	0.7130	0.8302
		Len	60.5080	93.6231	17.2067
400	200	Cov	0.9366	0.8364	0.9404
		Len	55.8214	49.6055	17.9916
1000	200	Cov	0.9710	0.9738	0.9656
		Len	23.2567	23.2984	19.1344

5.5 SIMULATIONS FOR THE NEW PI

For the model $\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{e}$, methods such as forward selection, PCR, PLS, ridge regression, relaxed lasso, and lasso each generate M fitted models $I_1, ..., I_M$, where M depends on the method, and we considered several methods for selecting the final submodel I_d. Only method i) needs n/p large.

i) Let $I_d = I_{min}$ be the model that minimizes C_p for forward selection, relaxed lasso, or lasso.

ii) Let I_d use $d = \min(\lceil n/J \rceil, p)$ variables where J is a positive integer. We used J = 5, 10, 20, and 50. Forward selection used $M = d$. (For ridge regression, we used the model I_c with the "degrees of freedom" closest to d.) For PCR and PLS, the "variables" were the $v_j = \gamma_j^T \boldsymbol{x}$. This method uses the full OLS model if $n/p \geq J$ for forward selection, PCR, PLS, ridge regression, relaxed lasso, and lasso. Hence large sample inference is simple for these six model selection estimators if p is fixed. For lasso, several values of λ may have the same degrees of freedom: we chose the model with the smallest λ value. In the simulation for lasso with $d = p$, we used the lasso model with λ_0 instead of OLS. See Section 5.2.

iii) Let $I_d = I_{min}$ be the model that minimizes $EBIC$ for forward selection or relaxed lasso. For forward selection, we used $M = \min(\lceil n/5 \rceil, p)$. See Section 5.3.

iv) Choose I_d using k–fold cross validation (CV). We used 10-fold CV.

The following method is currently slow to simulate, but is a useful diagnostic. When the model underfits, PI (3.7) tends to have coverage near or greater than the nominal 0.95 coverage, but the PI length is long. When the model severely overfits, the PI length is short, but the coverage is less than 0.95. See Section 5.1.

v) Modify k–fold cross validation to compute the PI coverage and average PI length on all M models. Then n PIs are made for Y_i using $\boldsymbol{x}_f = \boldsymbol{x}_i$ for $i = 1, ..., n$. The coverage is the proportion of times the n PIs contained Y_i. For example, choose the model I_d with the shortest average PI length given that the nominal large sample $100(1 - \delta)\%$ PI had coverage

$$\geq c_n = \max(1 - \delta - \frac{1}{\sqrt{n}}, 1 - \delta - 0.01).$$

If no model I_i had coverage $\geq c_n$, pick the model with the largest coverage.

The simulation for the new PIs (3.11) and (3.12) was similar to that in section 5.1. Let $\boldsymbol{x} = (1\ \boldsymbol{u}^T)^T$ where $\boldsymbol{u}$ is the $p-1 \times 1$ vector of nontrivial predictors. In the simulations, for $i = 1, ..., n$, we generated $\boldsymbol{w}_i \sim N_{p-1}(\boldsymbol{0}, \boldsymbol{I})$ where the $m = p - 1$ elements of the vector $\boldsymbol{w}_i$ are iid N(0,1). Let the $m \times m$ matrix $\boldsymbol{A} = (a_{ij})$ with $a_{ii} = 1$ and $a_{ij} = \psi$ where $0 \leq \psi < 1$ for $i \neq j$. Then the vector $\boldsymbol{u}_i = \boldsymbol{A}\boldsymbol{w}_i$ so that $Cov(\boldsymbol{u}_i) = \boldsymbol{\Sigma}_{\boldsymbol{u}} = \boldsymbol{A}\boldsymbol{A}^T = (\sigma_{ij})$ where the diagonal entries $\sigma_{ii} = [1 + (m-1)\psi^2]$ and the off diagonal entries $\sigma_{ij} = [2\psi + (m-2)\psi^2]$. Hence the correlations are $cor(x_i, x_j) = \rho = (2\psi + (m-2)\psi^2)/(1 + (m-1)\psi^2)$ for $i \neq j$ where x_i and x_j are nontrivial predictors. If $\psi = 1/\sqrt{cp}$, then $\rho \to 1/(c+1)$ as $p \to \infty$ where $c > 0$. As ψ gets close to 1, the predictor vectors cluster about the line in the direction of $(1, ..., 1)^T$. Then $Y_i = 1 + 1x_{i,2} + \cdots + 1x_{i,k} + e_i$ for $i = 1, ..., n$. Hence $\boldsymbol{\beta} = (1, .., 1, 0, ..., 0)^T$ with $k+1$ ones and $p-k-1$ zeros. The zero mean errors e_i were iid of five types: i) N(0,1) errors, ii) t_3 errors, iii) EXP(1) - 1 errors, iv) uniform$(-1, 1)$ errors, and v) 0.9 N(0,1) + 0.1 N(0,100) errors.

The lengths of the asymptotically optimal 95% PIs are i) 3.92 = 2(1.96), ii) 6.365, iii) 2.996, iv) 1.90 = 2(0.95), and v) 13.490. The simulation used 5000 runs, so an observed coverage in [0.94, 0.96] gives no reason to doubt that the PI has the nominal coverage of 0.95. The simulation used $p = 20, 40, n$, and $2n$. The simulation used $\psi = 0, 1/\sqrt{p}$, and 0.9, and $k = 1, 19$, and $p - 1$.

Tables 5.49 - 5.50 shows some simulation results. For lasso, often more than one λ value had $d - 1$ active predictors, and we used the value of λ closest to 0. If $d = p$, lasso and relaxed lasso used the selected value of λ rather than the OLS full model. For $N(0, 1)$ errors, $\psi = 0$, and $d < k$, the asymptotically optimal PI length is $3.92\sqrt{k - d + 1}$.

Table 5.49. Simulated Large Sample 95% PI Coverages and Lengths, $e_i \sim N(0,1)$

n	p	ψ	k		FS	lasso	RL	RR	PLS	PCR
100	20	0	1	Cov	0.9644	0.9570	0.9534	0.9354	0.9438	0.9772
				Len	4.4490	4.3849	4.3648	4.1441	4.4149	5.5647
100	40	0	1	Cov	0.9654	0.9522	0.9482	0.8932	0.8810	0.9882
				Len	4.4294	4.3113	4.2734	3.8982	4.0202	7.3393
100	100	0	1	Cov	0.9686	0.9494	0.9414	0.9554	0.8000	0.9932
				Len	4.4274	4.2427	4.1600	5.4422	3.5035	9.5767
100	200	0	1	Cov	0.9648	0.9332	0.9222	0.9254	0.6616	0.9922
				Len	4.4268	4.1546	4.0340	4.9843	2.7695	12.4116
200	20	0	19	Cov	0.9788	0.9766	0.9788	0.9792	0.9550	0.9786
				Len	4.9613	4.9636	4.9613	5.0458	4.3211	4.9610
200	40	0	19	Cov	0.9742	0.9650	0.9732	0.9606	0.9324	0.9792
				Len	4.9285	4.8146	4.8567	4.8044	4.2152	5.3616
200	100	0	19	Cov	0.9746	0.9456	0.9472	0.8416	0.7834	1.0000
				Len	4.9057	4.5640	4.5551	3.9090	3.4810	23.3839
200	200	0	19	Cov	0.9728	0.9124	0.9136	0.9696	0.3500	1.0000
				Len	4.8835	4.3197	4.2244	16.5887	2.1451	51.8962
400	20	0	19	Cov	0.9756	0.9756	0.9756	0.9760	0.9516	0.9756
				Len	4.6934	4.6959	4.6934	4.7504	4.0704	4.6934
400	40	0	19	Cov	0.9738	0.9748	0.9760	0.9714	0.9412	0.9790
				Len	4.6733	4.6638	4.6813	4.6776	4.0165	4.9001
400	100	0	19	Cov	0.9686	0.9554	0.9588	0.9250	0.8928	1.0000
				Len	4.6777	4.5262	4.4992	4.2544	3.7749	9.6077
400	200	0	19	Cov	0.9718	0.9528	0.9430	0.7956	0.7306	1.0000
				Len	4.6784	4.4430	4.3454	3.5541	3.1304	22.9925

Table 5.50. Simulated Large Sample 95% PI Coverages and Lengths, $e_i \sim N(0,1)$

n	p	ψ	k		FS	lasso	RL	RR	PLS	PCR
400	20	0.9	19	Cov	0.9664	0.9806	0.9686	0.981	0.9554	0.9536
				Len	4.5121	11.1707	4.8065	11.2288	4.0017	3.9771
400	40	0.9	19	Cov	0.9674	0.9792	0.9710	0.9756	0.9482	0.9646
				Len	4.5682	16.4331	5.4591	16.2248	4.0070	4.3797
400	400	0.9	19	Cov	0.9348	0.9672	0.9564	0.9666	0.9462	0.9478
				Len	4.3687	48.1703	4.9438	48.8372	4.29138	4.4764
400	800	0.9	19	Cov	0.9268	0.9660	0.9554	0.9692	0.9438	0.9554
				Len	4.3427	67.4054	4.7992	66.6885	4.2965	4.6533
100	50	0	49	Cov	0.8996	0.8416	0.8440	0.8734	0.8448	1.0000
				Len	22.0672	4.2152	4.1995	4.7628	4.2141	38.9044
100	100	0	99	Cov	0.9330	0.2838	0.1952	0.9302	0.2164	1.0000
				Len	40.4282	5.2674	3.4328	33.0614	2.0001	153.4029
100	200	0	1	Cov	0.9648	0.9332	0.9222	0.9254	0.6616	0.9922
				Len	4.4268	4.1546	4.0340	4.9843	2.7695	12.4116
200	200	0	199	Cov	0.9432	0.2370	0.1390	0.8910	0.1434	1.0000
				Len	56.2052	5.6741	3.4821	38.6765	1.4941	214.7205
400	400	0	1	Cov	0.9490	0.9442	0.9346	0.9256	0.7926	0.9914
				Len	4.0122	3.9495	3.8827	4.8334	3.3135	8.3890
400	400	0	19	Cov	0.9774	0.9384	0.9258	0.9504	0.3282	1.0000
				Len	4.6768	4.3460	4.1384	14.7190	1.9502	51.1337
400	400	0	399	Cov	0.9486	0.2330	0.1324	0.8702	0.0948	1.0000
				Len	78.4105	7.3847	4.3729	47.5899	1.1749	305.9331

5.6 SIMULATIONS FOR BOOTSTRAPPING

Assume $n \geq 20p$ and that the error distribution is unimodal and not highly skewed. The response plot and residual plot are plots with $\hat{Y} = \boldsymbol{x}^T \hat{\boldsymbol{\beta}}$ on the horizontal axis and Y or r on the vertical axis, respectively. Then the plotted points in these plots should scatter in roughly even bands about the identity line with unit slope and zero intercept and $r = 0$ lines, respectively. See Figure 5.1. If the plots for the OLS full model suggest that the error distribution is skewed or multimodal, then much larger sample sizes may be needed.

If the error distribution is unknown, then large sample theory tests are straightforward if the estimator is asymptotically equivalent to the OLS full model, e.g. $\hat{\lambda}_{1,n} = o_P(\sqrt{n})$, or choose the OLS full model if $n \geq 50p$. The latter technique may be reasonable if the large sample theory of the method is not better than that of the OLS full model (lasso and ridge regression), or if it is not known how to do inference unless the model is asymptotically equivalent to the OLS full model (PCR, PLS, forward selection).

The residual bootstrap with the residuals from the OLS full model can provide a lot of information. Olive (2017b: p. 128, 2017a) showed that the prediction region method can simulate well for the $p \times 1$ vector $\hat{\boldsymbol{\beta}}_{I_{min},0}$ obtained by adding zeroes to $\hat{\boldsymbol{\beta}}_{I_{min}}$ where I_{min} is the model that minimizes C_p for forward selection. Asymptotically, $\hat{\boldsymbol{\beta}}_{I_{min},0}$ is a mixture $\sum_j \pi_j \hat{\boldsymbol{\beta}}_{I_j,0}$ where $0 \leq \pi_j \leq 1$ and $\sum_j \pi_j = 1$ where the sum is over all 2^{p-a_S} submodels I_j that contain S. Results from Knight and Fu (2000) show that this residual bootstrap works for each component $\hat{\boldsymbol{\beta}}_{I_j,0}$, but we may need at least $50p$ bootstrap samples per component with nonnegligible π_j. The number of nonnegligible π_j can be small if $p - a_S$ is small or if a criterion that picks S with high probability, such as BIC, is used. Here $Y = \boldsymbol{x}^T \boldsymbol{\beta} + e = \boldsymbol{x}_S^T \boldsymbol{\beta}_S + e$ where $\boldsymbol{\beta}_S$ is $a_s \times 1$.

Examining $\hat{\boldsymbol{\beta}}_S$ and $\hat{\boldsymbol{\beta}}_E$ is informative for I_{min}. First assume that the nontrivial predictors are uncorrelated or orthogonal so $\boldsymbol{X}^T \boldsymbol{X}/n \to diag(d_1, ..., d_p)$ as $n \to \infty$ where each $d_i > 0$. Then $\hat{\boldsymbol{\beta}}_S$ has the same limiting distribution for I_{min} and for the OLS full model. The bootstrap distribution for $\hat{\boldsymbol{\beta}}_E$ is a mixture of zeroes and a distribution that would produce a confidence region for $\boldsymbol{A\beta}_E = \boldsymbol{0}$ that has asymptotic coverage of $\boldsymbol{0}$ equal to $100(1-\delta)\%$. Hence the asymptotic coverage is greater than the nominal coverage provided that $\boldsymbol{S}_T$ in nonsingular with probability going to one (e.g., $p - a_S$ is small), where $T = \boldsymbol{A}\hat{\boldsymbol{\beta}}_{E,I_{min}}$. With uncorrelated predictors, the number of bootstrap samples $B \geq 50p$ may work well for the shorth confidence intervals and for testing $\boldsymbol{A\beta}_S = \boldsymbol{0}$.

We do not yet have a proof that the prediction region method works when the estimator is not asymptotically multivariate normal, but in the simulations for forward selection, coverages were similar regardless of the correlation of the predictors. Let $\boldsymbol{\beta}_O$ be a vector component of $\boldsymbol{\beta}_E$, and consider testing $H_0 : \boldsymbol{A\beta}_O = \boldsymbol{0}$. If $\boldsymbol{A}\hat{\boldsymbol{\beta}}^*_{O,i} = \boldsymbol{0}$ for greater than $B\delta$ of the bootstrap samples $i = 1, ..., B$, then the $100(1-\delta)\%$ prediction region method confidence region will contain $\boldsymbol{0}$, and the test will fail to reject H_0.

Suppose we want to bootstrap $T = \hat{\boldsymbol{\beta}}_O$, where $\boldsymbol{\beta} = (\boldsymbol{\beta}_I^T, \boldsymbol{\beta}_O^T)^T$, and all $\hat{\boldsymbol{\beta}}^*_{O,i} = \boldsymbol{0}$ for $i = 1, ..., B$. Then $\boldsymbol{S}_T$ is singular, but the singleton set $\{\boldsymbol{0}\}$ is the large sample prediction region method $100(1-\delta)\%$ confidence region for $\boldsymbol{\beta}_O$ and $\delta \in (0,1)$, and the pvalue for $H_0 : \boldsymbol{\beta}_O = \boldsymbol{0}$ is one. For large sample theory tests, the pvalue estimates the population pvalue. For the I_{min} model from forward selection, there may be strong evidence that $\boldsymbol{x}_O$ is not needed in the model given $\boldsymbol{x}_I$ is in the model if the confidence region is $\{\boldsymbol{0}\}$, $n \geq 20p$, $B \geq 50p$, and the error distribution is unimodal and not highly skewed. (Since the pvalue is one, this technique may be useful for data snooping: applying OLS theory to submodel I may have negligible selection bias.)

A small simulation was done, using the same type of data as for the prediction interval simulation, using $B = \max(1000, n, 20p)$ and 5000 runs. The regression model used $\boldsymbol{\beta} = (1,1,0,0)^T$ with $n = 100$ and $p = 4$. When $\psi = 0$, the design matrix $\boldsymbol{X}$ consisted of iid N(0,1) random variables, and the full model least squares confidence intervals for β_i should have length near $2t_{96,0.975}\sigma/\sqrt{n} \approx 2(1.96)\sigma/10 = 0.392\sigma$ when the iid zero mean errors have variance σ^2. The simulation computed the Frey shorth(c) interval for each β_i and used the prediction region method to test $H_0 : \beta_3 = \beta_4 = 0$. The nominal coverage was 0.95 with

$\delta = 0.05$. Observed coverage between 0.94 and 0.96 would suggest coverage is close to the nominal value. Models with the first $k+1$ $\beta_i = 1$ and the last $p-k-1$ $\beta_i = 0$ were also considered.

The regression models used the residual bootstrap on the full model least squares estimator and on the forward selection estimator $\hat{\boldsymbol{\beta}}_{I_{min},0}$. Results are shown for when the iid errors $e_i \sim N(0,1)$. Table 5.51 shows two rows for each model giving the observed confidence interval coverages and average lengths of the confidence intervals. The term "reg" is for the full model regression, and the term "vs" is for forward selection. The column for the "test" gives the length and coverage = P(fail to reject H_0) for the interval $[0, D_{(U_B)}]$ where $D_{(U_B)}$ is the cutoff for the confidence region. These lengths do not give information about the volume of the confidence region, which will decrease to 0 as $n \to \infty$. The cutoff will often be near $\sqrt{\chi^2_{r,0.95}}$ if the statistic T is asymptotically normal. Note that $\sqrt{\chi^2_{2,0.95}} = 2.448$ is close to 2.45 for the full model regression bootstrap test. The coverages were near 0.95 for the regression bootstrap on the full model.

Suppose $\psi = 0$. Then $\hat{\boldsymbol{\beta}}_S$ has the same limiting distribution for I_{min} and the full model. Note that the average lengths and coverages were similar for the full model and forward selection I_{min} for β_1 and β_2 and $\boldsymbol{\beta}_S = (\beta_1, \beta_2)^T$. Table 5.52 shows similar results for lasso when $\psi = 0$.

Table 5.51. Bootstrapping OLS Regression and Forward Selection

model	ψ	cov/len	β_1	β_2	β_3	β_4	test
reg	0	cov	0.9456	0.9474	0.9496	0.9474	0.9442
		len	0.3961	0.3997	0.3988	0.3992	2.4503
vs	0	cov	0.9472	0.9470	0.9980	0.9980	0.9936
		len	0.3964	0.3991	0.3246	0.3233	2.6936
reg	0.5	cov	0.9432	0.9452	0.9498	0.9506	0.9436
		len	0.3976	0.6642	0.6645	0.6637	2.4507
vs	0.5	cov	0.9458	0.9728	0.9976	0.9974	0.9926
		len	0.3966	0.6598	0.5383	0.5383	2.7055
reg	0.9	cov	0.9432	0.9512	0.9500	0.9498	0.9442
		len	0.3963	3.2621	3.2613	3.2611	2.4505
vs	0.9	cov	0.9422	0.9678	0.9944	0.9970	0.9914
		len	0.3957	2.7640	2.7356	2.7430	2.7121

Table 5.52. Bootstrap LASSO, $\psi = 0$

n	eps	type		β_1	β_2	β_3	β_4	test
100		1	cicov	0.9440	0.9376	0.9910	0.9946	0.9790
			avelen	0.4143	0.4470	0.3759	0.3763	2.6444
		2	cicov	0.9468	0.9428	0.9946	0.9944	0.9816
			avelen	0.6870	0.7565	0.6238	0.6226	2.6832
		3	cicov	0.9418	0.9408	0.9930	0.9948	0.9840
			avelen	0.4110	0.4506	0.3743	0.3746	2.6684
		4	cicov	0.9468	0.9370	0.9938	0.9948	0.9838
			avelen	0.2392	0.2578	0.2151	0.2153	2.6454
	0.5	5	cicov	0.9438	0.9344	0.9988	0.9970	0.9924
			avelen	2.9380	2.5042	2.4912	2.4715	2.8536
	0.9	5	cicov	0.9506	0.9290	0.9974	0.9976	0.9956
			avelen	3.9180	3.2760	3.2739	3.2702	2.8836
200		1	cicov	0.9494	0.9390	0.9942	0.9924	0.9802
			avelen	0.4132	0.4460	0.3754	0.3760	2.6455
		2	cicov	0.9474	0.9502	0.9966	0.9948	0.9860
			avelen	0.4902	0.5365	0.4445	0.4448	2.6726
		3	cicov	0.9432	0.9440	0.9958	0.9966	0.9852
			avelen	0.2924	0.3167	0.2641	0.2647	2.6617
		4	cicov	0.9504	0.9354	0.9952	0.9948	0.9858
			avelen	0.1699	0.1810	0.1506	0.1510	2.6429
	0.9	5	cicov	0.9486	0.9352	0.9978	0.9972	0.9936
			avelen	2.6679	2.2152	2.1943	2.2014	2.7952

CHAPTER 6

CONCLUSIONS

Let p be fixed and $n \to \infty$. For forward selection, PCR, PLS, ridge regression, relaxed lasso, and lasso, if $P(d \to p) \to 1$ as $n \to \infty$ then the six methods are asymptotically equivalent to the OLS full model, and the PIs (3.11) and (3.12) are asymptotically optimal on a large class of iid unimodal error distributions. For PCR and some constants θ_i, $\sum_{i=1}^{j} \theta_i \gamma_i^T \boldsymbol{x}_i = \sum_{i=1}^{p} \beta_i x_i$ if $j = p$, but not if $j < p$ in general. Hence PCR tends to give inconsistent estimators unless P($j = p$) = P(PCR uses the full OLS model) goes to one. Forward selection with C_p produces a $\sqrt{n}$ consistent estimator $\hat{\boldsymbol{\beta}}_{I_{min},0}$ of $\boldsymbol{\beta}$. Using $d = \min(\lceil n/J \rceil, p)$ with forward selection, PCR, PLS, ridge regression, lasso, and relaxed lasso makes large sample inference easy since the selected model is the full OLS model if $n/p \geq J$.

There is massive literature on variable selection and a fairly large literature for inference after variable selection. See, for example, Bertsimas, King, and Mazumder (2016), Fan and Lv (2010), Ferrari and Yang (2015), Fithian, Sun, and Taylor (2014), Hjort and Claeskins (2003), Knight and Fu (2000), Lee, Sun, Sun, and Taylor (2016), Leeb and Pötscher (2006), Lockhart, Taylor, Tibshirani, and Tibshirani (2014), Qi, Luo, Carroll, and Zhao (2015), and Taylor, Lockhart, Tibshirani, and Tibshirani (2014).

If n/p is large, the residual bootstrap with OLS residuals should work for lasso, relaxed lasso, and ridge regression if $\hat{\lambda}_{1,n} = o_P(\sqrt{n})$. Also see Knight and Fu (2000). Camponovo (2015) suggests that the nonparametric bootstrap does not work for lasso. Chatterjee and Lahiri (2011) state that the residual bootstrap with lasso does not work. Hall, Lee, and Park (2009) show that the residual bootstrap with full model OLS residuals does not work, but the m out of n residual bootstrap with OLS full model residuals does work. Rejchel (2016) is a good review of lasso theory. Fan and Lv (2010) review large sample theory for some alternative methods. See Lockhart, Taylor, Tibshirani, and Tibshirani (2014) for a partial remedy for hypothesis testing with lasso. Xu, Caramanis, and Mannor (2011) suggest that sparse algorithms are not stable.

Lei, G'Sell, Rinaldo, Tibshirani, and Wasserman (2016) and Wasserman (2014) suggest prediction intervals for estimators such as lasso. Also see Butler and Rothman (1980). Steinberger and Leeb (2016) use leave-one-out residuals to make a PI. Chao, Ning, and Liu (2014), assume that the e_i are iid $N(0, \sigma^2)$. Denham (1997) gave a PI for PLS when the number of PLS components V_j is selected in advance. Zhang (1992) has some k–fold CV theory.

With n/p large, C_p produced good PIs for forward selection and 10-fold CV produced good PIs for PCR and PLS. For lasso and ridge regression, 10-fold CV produced good PIs if $\psi = 0$ or if k was small. If k was larger than about 18 and the predictors were highly correlated, 10-fold CV tended to underfit and the PI lengths were too long.

When n/p is not large, inference is currently much more difficult. Zheng and Loh (1995) show that BIC_S can work if $p = o(\log(n))$ and there is a consistent estimator of σ^2. Chun and Keleş (2010) show that PLS does not give a consistent estimator of $\boldsymbol{\beta}$ unless $p/n \to 0$. Also see Cook, Helland, and Su (2013). Fan and Lv (2010) give large sample theory for some methods if $p = o(n^{1/5})$. Leeb, Pötscher, and Ewald (2015) suggests that the method Berk et al. (2013) method does not really work. Also see Dezeure et al. (2015), Javanmard and Montanari (2014), Taylor et al. (2014), and van de Geer et al. (2014).

Response plots of the fitted values $\hat{Y}$ versus the response Y are useful for checking linearity of the MLR model and for detecting outliers. Residual plots should also be made. When n is large, the points within the pointwise PI bands can be omitted, eliminating a black band about the identity line.

The simulations were done in R. See R Core Team (2016). A much larger simulation study is in Pelawa Watagoda (2017). We used several R functions including forward selection as computed with `regsubsets` function from the `leaps` library, principal components regression with the `pcr` function and partial least squares with the `plsr` function from the `pls` library, and ridge regression and lasso with the `glmnet` and `cv.glmnet` functions from the `glmnet` library.

The collection of Olive (2017d) *R* functions *slpack*, available from (http://lagrange.math.siu.edu/Olive/slpack.txt), has some useful functions for the inference. Tables 5.49 and 5.50 were made with `mspisim` while Table 5.51 was made with `regbootsim` for the OLS full model and `vsbootsim3` for forward selction. The function `lassobotsim3` uses the prediction region method for lasso for Table 5.52. For PI (3.12), the function `valvspisim` is for forward selection using the minimum C_p model, and the function `valrelpisim` simulates the relaxed lasso model corresponding to the lasso model chosen with 10-fold CV.

REFERENCES

[1] Akaike, H. (1973), "Information Theory as an Extension of the Maximum Likelihood Principle," in *Proceedings, 2nd International Symposium on Information Theory*, eds. Petrov, B.N., and Csakim F., Akademiai Kiado, Budapest, 267-281.

[2] Berk, R., Brown, L., Buja, A., Zhang, K., and Zhao, L. (2013), "Valid Post-Selection Inference," *The Annals of Statistics,* 41, 802-837.

[3] Bertsimas, D., King, A., and Mazumder, R. (2016), "Best Subset Selection Via a Modern Optimizations Lens," *Annals of Statistics*, to appear.

[4] Bickel, P.J., and Ren, J.–J. (2001), "The Bootstrap in Hypothesis Testing," in *State of the Art in Probability and Statistics: Festschrift for William R. van Zwet*, eds. de Gusnt, M., Klaassen, C., and van der Vaart, A., The Institute of Mathematical Statistics, Hayward, CA, 91-112.

[5] Burnham, K.P., and Anderson, D.R. (2004), "Multimodel Inference Understanding AIC and BIC in Model Selection," *Sociological Methods* & *Research*, 33, 261-304.

[6] Butler, R., and Rothman, E. (1980), "Predictive Intervals Based on Reuse of the Sample," *Journal of the American Statistical Association,* 75, 881-889.

[7] Camponovo, L. (2015), "On the Validity of the Pairs Bootstrap for Lasso Estimators," *Biometrika,* 102, 981-987.

[8] Chao, S.-K., Ning, Y., and Liu, H. (2014), "On High Dimensional Post-Regularization Prediction Intervals," unpublished manuscript at (http://www.stat.purdue.edu/∼skchao74/HD_PCI.pdf).

[9] Chatterjee, A., and Lahiri, S.N. (2011), "Bootstrapping Lasso Estimators," *Journal of the American Statistical Association*, 106, 608-625.

[10] Chun, H. and S. Keleş (2010), "Sparse Partial Least Squares Regression for Simultaneous Dimension Reduction and Predictor Selection," *Journal of the Royal Statistical Society, B,* 72, 3-25.

[11] Cook, R.D., Helland, I.S., and Su, Z. (2013), "Envelopes and Partial Least Squares Regression," *Journal of the Royal Statistical Society, B*, 75, 851-877.

[12] Denham, M.C., (1997), "Prediction Intervals in Partial Least Squares," *Journal of Chemometrics,* 11, 39-52.

[13] Dezeure, R., Bühlmann, P., Meier, L., and Meinshausen, N. (2015), High-Dimensional Inference: Confidence Intervals, p-Values and R-Software hdi," *Statistical Science*, 30, 533-558.

[14] Efron, B. (1982), *The Jackknife, the Bootstrap and Other Resampling Plans,* SIAM, Philadelphia, PA.

[15] Efron, B. (2014), "Estimation and Accuracy After Model Selection," (with discussion), *Journal of the American Statistical Association,* 109, 991-1007.

[16] Efron, B., and Hastie, T. (2016), *Computer Age Statistical Inference*, Cambridge University Press, New York, NY.

[17] Fan, J. and Lv, J. (2010), "A Selective Overview of Variable Selection in High Dimensional Feature Space," *Statistica Sinica*, 20, 101-148.

[18] Ferrari, D., and Yang, Y. (2015), "Confidence Sets for Model Selection by F–Testing," *Statistica Sinica*, 25, 1637-1658.

[19] Fithian, W., Sun, D., and Taylor, J. (2014), "Optimal Inference After Model Selection," ArXiv e-prints .

[20] Frey, J. (2013), "Data-Driven Nonparametric Prediction Intervals," *Journal of Statistical Planning and Inference,* 143, 1039-1048.

[21] Gunst, R.F., and Mason, R.L. (1980), *Regression Analysis and Its Application,* Marcel Dekker, New York, NY.

[22] Hall, P., Lee, E.R., and Park, B.U. (2009), "Bootstrap-Based Penalty Choice for the Lasso Achieving Oracle Performance," *Statistica Sinica*, 19, 449-471.

[23] Hastie, T., Tibshirani, R., and Wainwright, M. (2015), *Statistical Learning with Sparsity: the Lasso and Generalizations*, CRC Press Taylor & Francis, Boca Raton, FL.

[24] Hebbler, B. (1847), "Statistics of Prussia," *Journal of the Royal Statistical Society, A*, 10, 154-186.

[25] Hjort, N.L., and Claeskins, G. (2003), "Frequentist Model Average Estimators," *Journal of the American Statistical Association,* 98, 879-899.

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

[27] James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013), *An Introduction to Statistical Learning With Applications in R*, Springer, New York, NY.

[28] Javanmard, A., and Montanari, A. (2014), "Confidence Intervals and Hypothesis Testing for High-Dimensional Regression," *Journal of Machine Learning Research*, 15, 2869-2909.

[29] Knight, K., and Fu, W.J. (2000), "Asymptotics for Lasso-Type Estimators," *Annals of Statistics*, 28, 1356–1378.

[30] Lee, J., Sun, D., Sun, Y., and Taylor, J. (2016), "Exact Post–Selection Inference, with Application to the Lasso," *Annals of Statistics,* to appear.

[31] Leeb, H., and Pötscher, B.M. (2006), "Can One Estimate the Conditional Distribution of Post–Model-Selection Estimators?" *The Annals of Statistics*, 34, 2554-2591.

[32] Leeb, H., Pötscher, B.M., and Ewald, K. (2015), "On Various Confidence Intervals Post-Model-Selection," *Statistical Science*, 30, 216-227.

[33] Lei, J., G'Sell, M., Rinaldo, A., Tibshirani, R.J., and Wasserman, L. (2016), "Distribution-Free Predictive Inference for Regression," unpublished document at (http://www.stat.cmu.edu/∼ryantibs/papers/conformal.pdf).

[34] Lockhart, R., Taylor, J., Tibshirani, R. and Tibshirani, R. (2014), "A Significance Test for the Lasso," with discussion, *Annals of Statistics,* 42, 413–468.

[35] Luo, S., and Chen, Z. (2013), "Extended BIC for Linear Regression Models With Diverging Number of Relevant Features and High or Ultra-High Feature Spaces," *Journal of Statistical Planning and Inference,* 143, 494-504.

[36] Meinshausen, N. (2007), "Relaxed Lasso," *Computational Statistics & Data Analysis*, 52, 374-393.

[37] Nishi, R. (1984), "Asymptotic Properties of Criteria for Selection of Variables in Multiple Regression," *Annals of Statistics,* 12, 758-765.

[38] Olive, D.J. (2013), "Asymptotically Optimal Regression Prediction Intervals and Prediction Regions for Multivariate Data," *International Journal of Statistics and Probability*, 2, 90-100.

[39] Olive, D.J. (2017a), "Applications of Hyperellipsoidal Prediction Regions," *Statistical Learning*, to appear, see (http://lagrange.math.siu.edu/Olive/pphpr.pdf).

[40] Olive, D.J. (2017b), *Robust Multivariate Analysis*, Springer, New York, NY, to appear.

[41] Olive, D.J. (2017c), *Linear Regression*, Springer, New York, NY, to appear.

[42] Olive, D.J. (2017d), *Prediction and Statistical Learning*, online course notes, see (http://lagrange.math.siu.edu/Olive/slearnbk.htm).

[43] Olive, D.J. (2017e), "Bootstrapping Hypothesis Tests," unpublished manuscript, (http://lagrange.math.siu.edu/Olive/ppvselboot.pdf).

[44] Olive, D.J., and Hawkins, D.M. (2003), "Robust Regression with High Coverage," *Statistics & Probability Letters,* 63, 259-266.

[45] Olive, D.J., and Hawkins, D.M. (2005), "Variable Selection for 1D Regression Models," *Technometrics*, 47, 43-50.

[46] Pelawa Watagoda, L.C.R. (2017), *Simulation for Inference After Variable Selection*, unpublished manuscript online at (http://lagrange.math.siu.edu/Olive/slasanthisim.pdf).

[47] Qi, X., Luo, R., Carroll, R.J., and Zhao, H. (2015), "Sparse Regression by Projection and Sparse Discriminant Analysis," *Journal of Computational and Graphical Statistics,* 24, 416-438.

[48] R Core Team (2016), "R: a Language and Environment for Statistical Computing," R Foundation for Statistical Computing, Vienna, Austria, (www.R-project.org).

[49] Rejchel, W. (2016), "Lasso With Convex Loss: Model Selection Consistency and Estimation," *Communications in Statistics: Theory and Methods,* 45, 1989-2004.

[50] Schwarz, G. (1978), "Estimating the Dimension of a Model," *The Annals of Statistics,* 6, 461-464.

[51] Seber, G.A.F., and Lee, A.J. (2003), *Linear Regression Analysis,* 2nd ed., Wiley, New York, NY.

[52] Steinberger, L., and Leeb, H. (2016), "Leave-One-Out Prediction Intervals in Linear Regression Models With Many Variables," unpublished manuscript at (https://arxiv.org/pdf/1602.05801.pdf).

[53] Su, Z., and Cook, R.D. (2012), "Inner Envelopes: Efficient Estimation in Multivariate Linear Regression," *Biometrika*, 99, 687-702.

[54] Taylor, J., Lockhart, R., Tibshirani, R., and Tibshirani, R. (2014), "Post-Selection Adaptive Inference for Least Angle Regression and the Lasso," arXiv: 1401.3889; submitted.

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

[56] Tibshirani, R.J. (2015), "Degrees of Freedom and Model Search," *Statistica Sinica*, 25, 1265-1296.

[57] van de Geer, S., Bülhmann, P. Ritov, Y., and Dezeure, R. (2014), "On Asymptotically Optimal Confidence Regions and Tests for High-Dimensional Models," *The Annals of Statistics*, 42, 1166-1202.

[58] Wasserman, L. (2014), "Discussion: A Significance Test for the Lasso," *The Annals of Statistics,* 42, 501-508.

[59] Wold, H. (1975), "Soft Modelling by Latent Variables: the Nonlinear Partial Least Squares (NIPALS) Approach," in *Perspectives in Probability and Statistics, Papers in Honor of M.S. Bartlett,* ed. Gani, J., Academic Press, San Diego, CA, 117-144.

[60] Xu, H., Caramanis, C., and Mannor, S. (2011), "Sparse Algorithms are Not Stable:

a No-Free-Lunch Theorem," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* PP(99), 1-9.

[61] Zhang, P. (1992), "On the Distributional Properties of Model Selection Criterion," *Journal of the American Statistical Association*, 87, 732-737.

[62] Zheng, Z., and Loh, W.-Y. (1995), "Consistent Variable Selection in Linear Models," *Journal of the American Statistical Association*, 90, 151-156.

VITA

Graduate School
Southern Illinois University

Lasanthi C. R. Pelawa Watagoda

lasanthi@siu.edu (lasanthi@appstate.edu)

Southern Illinois University Carbondale
Master of Science, Mathematics, August 2013

University of Sri Jayewardenepura Sri Lanka
Bachelor of Science(Special), Mathematics, July 2010

Special Honors and Awards: Dissertation Research Assistantship Award 2017.
John M. H. Olmsted Award- Ph.D. Teaching Assistant Award for outstanding teaching performance, Department of Mathematics at SIU - 2015.

Dissertation Title:

Inference After Variable Selection.

Major Professor: Dr. D. J. Olive

Publications:

1. "Bootstrapping analogs of the Hotelling's T^2 test", Communications in Statistics: Theory and Methods, to appear, with Hasthika Rupasinghe.
2. "Visualizing and Testing the Multivariate Linear Regression Model", International Journal of Statistics and Probability January 22, 2015, with David J Olive and Hasthika Rupasinghe.
3. "Inference For Multiple Linear Regression After Model or Variable Selection", work in progress, with David J. Olive.