2017

Prediction Intervals For Lasso and Relaxed Lasso Using D Variables

Craig J. Bartelsmeyer
Southern Illinois University Carbondale, cjbartelsmeyer@siu.edu

PREDICTION INTERVALS FOR LASSO AND RELAXED LASSO USING D VARIABLES

by

Craig Bartelsmeyer

B.S., University of Alabama, 2015

A Research Paper
Submitted in Partial Fulfillment of the Requirements for the
Master of Science

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

RESEARCH PAPER APPROVAL

PREDICTION INTERVALS FOR LASSO AND RELAXED LASSO USING D VARIABLES

by

Craig Bartelsmeyer

A Research Paper Submitted in Partial

Fulfillment of the Requirements

for the Degree of

Master of Science

in the field of Mathematics

Approved by:

David J. Olive

Bhaskar Bhattacharya

Kwangho Choiy

Graduate School
Southern Illinois University Carbondale
April 3, 2017

AN ABSTRACT OF THE RESEARCH PAPER OF

CRAIG BARTELSMEYER, for the Master of Science degree in MATHEMATICS, presented on APRIL 3, 2017, at Southern Illinois University Carbondale.

TITLE: PREDICTION INTERVALS FOR LASSO AND RELAXED LASSO USING D VARIABLES

MAJOR PROFESSOR: Dr. David J. Olive

This paper, taken from Olive (2017c), presents and examines a prediction interval for the multiple linear regression model $Y = \beta_1 x_1 + \cdots + \beta_p x_p + e$, where the lasso or relaxed lasso model is selected using d = $\min(\lceil n/J \rceil, p)$ variables for some positive integer J such as 10 or 20.

ACKNOWLEDGMENTS

I would like to take this opportunity to thank my research advisor, Dr. David Olive for overseeing my Master's project and Dr. Bhaskar Bhattacharya and Dr. Kwangho Choiy for sitting on my committee. I would also like to thank the professors of Southern Illinois University for their instruction and care over the past two years. Finally I want to say to my family (especially the TeaMaster) thank you for all of your support and encouragement. You have pushed me to succeed and I could not have done it without you!

TABLE OF CONTENTS

LIST OF TABLES

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 $\boldsymbol{x}_i^T = (x_{i,1}, ..., x_{i,p}) = (1 \;\; \boldsymbol{u}_i^T)$ 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. Ordinary least squares (OLS) regression is often used for inference if n/p is large.

It is often convenient to use the centered response $\boldsymbol{Z} = \boldsymbol{Y} - \overline{\boldsymbol{Y}}$ 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

$$W_{ij} = \frac{x_{i,j+1} - \overline{x}_{j+1}}{\tilde{\sigma}_{j+1}}$$

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

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, lasso (due to Tibshirani (1996)), and relaxed lasso (due to Meinshausen (2007)). Lasso produces an $\hat{\boldsymbol{\eta}}$ where typically some of the estimated coefficients $\hat{\eta}_i$ are equal to 0. Relaxed lasso applies OLS to the variables estimated to have nonzero coefficients as well as a constant. These three 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, $\hat{\boldsymbol{\eta}} = \mathbf{0}$ and the submodel $Y_i = \beta_1 + e_i$ has fitted values $\hat{Y}_i \equiv \overline{Y}$. Lasso has a parameter λ such that when $\lambda = 0$, the full OLS model is used. Lasso uses 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 $\lambda_1 = 0$. Then λ_M is the smallest value of λ such that $\hat{\boldsymbol{\eta}}_{\lambda_M} = \mathbf{0}$. Hence $\hat{\boldsymbol{\eta}}_{\lambda_i} \neq \mathbf{0}$ for $i < M$. See James, Witten, Hastie, and Tibshirani (2013, ch. 6), Olive (2017c), Pelawa Watagoda (2017), and Pelawa Watagoda and Olive (2017) for more details about these three methods.

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 two paragraphs assume p is fixed. Knight and Fu (2000) proved that:
i) $\hat{\boldsymbol{\eta}}$ 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).

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

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

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 paragraphs hold after model selection if $\lambda_{1,n}$ is replaced by $\hat{\lambda}_{1,n}$, o by o_P, and O by 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 $\lambda_1, \lambda_2, \cdots, \lambda_M$ where $\lambda_i = \lambda_{1,n,i}$ depends on n for $i = 1, \cdots, M$. If λ_s corresponds to the model selected, then $\hat{\lambda}_{1,n} = \lambda_s$.

Variable 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

$$\boldsymbol{x}^T\boldsymbol{\beta} = \boldsymbol{x}_S^T\boldsymbol{\beta}_S + \boldsymbol{x}_E^T\boldsymbol{\beta}_E = \boldsymbol{x}_S^T\boldsymbol{\beta}_S, \tag{1.5}$$

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 $(p - k_S) \times 1$ 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 (1.5) holds. Then

$$\boldsymbol{x}^T\boldsymbol{\beta} = \boldsymbol{x}_S^T\boldsymbol{\beta}_S = \boldsymbol{x}_S^T\boldsymbol{\beta}_S + \boldsymbol{x}_{I/S}^T\boldsymbol{\beta}_{(I/S)} + \boldsymbol{x}_O^T\boldsymbol{0} = \boldsymbol{x}_I^T\boldsymbol{\beta}_I, \tag{1.6}$$

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 $S \subseteq I$.

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 $a \times 1$ vectors. Then 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 that can be 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; this is the method that we will investigate. If p is fixed, the method will use the full OLS model once $n/J \geq p$. Hence the prediction interval (2.5) described below will be asymptotically optimal for a wide class of zero mean error distributions.

Consider predicting a future test response variable Y_f given a $p \times 1$ vector of predictors $\boldsymbol{x}_f$ and training data $(\boldsymbol{x}_1, Y_1), \cdots, (\boldsymbol{x}_n, Y_n)$. 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 $n \to \infty$.

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

$$\text{shorth(c)} = [\text{Z}_{(\text{s})}, \text{Z}_{(\text{s+c}-1)}]. \tag{1.7}$$

Let

$$k_n = \lceil n(1-\delta) \rceil, \tag{1.8}$$

then Frey (2013) showed that for large $n\delta$ and iid data the shorth(k_n) PI has maximum undercoverage $\approx 1.12\sqrt{\delta/n}$.

He used the shorth(c) estimator as the large sample $100(1-\delta)\%$ PI where

$$c = \min(n, \lceil n[1 - \delta + 1.12\sqrt{\delta/n}]\rceil). \tag{1.9}$$

A problem with the prediction intervals that cover $\approx 100(1-\delta)\%$ of the training data cases Y_i (such as the shorth(k_n) PI) 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.

Example 1. (Example 5.3 from Olive(2017b)) Given below were votes for preseason 1A basketball poll from Nov. 22, 2011 WSIL News where the 778 was a typo; the actual value was 78. As shown below, finding the shorth(3) from the ordered data is simple. If the outlier is corrected, shorth(3) = [76,78]. Otherwise we have the following:

```
111   89   778   78   76


order data: 76 78 89 111 778


                  13 = 89 - 76


                      33 = 111 - 78


                          689 = 778 - 89
shorth(3) = [76, 89]
```

Olive (2007) developed prediction intervals for the full MLR model. Olive (2013) developed prediction intervals for models of the form $Y_i = m(\boldsymbol{x}_i) + e_i$. Variable selection models for (1.1) have this form, as noted by Olive(2017a). Both these PIs need n/p to be large. Let c be given by (2.4) with d replaced by p, and let

$$b_n = (1 + \frac{15}{n})\sqrt{\frac{n+2p}{n-p}} \tag{1.10}$$

Compute the shorth(c) of the residuals $= [r_{(s)}, r_{(s+c-1)}] = [\tilde{\xi}_{\delta_1}, \tilde{\xi}_{1-\delta_2}]$ where the ith residual $r_i = Y_i - \hat{Y}_i = Y_i - \hat{m}(\boldsymbol{x}_i)$.

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

$$[\hat{m}(\boldsymbol{x}_f) + b_n\tilde{\xi}_{\delta_1}, \hat{m}(\boldsymbol{x}_f) + b_n\tilde{\xi}_{1-\delta_2}]. \tag{1.11}$$

Note that correction factors $b_n \to 1$ are used in large sample confidence intervals and test 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 as follows: $\frac{t_{d_n,1-\delta}}{z_{1-\delta}} \to 1$ and $\frac{pF_{p,d_n,1-\delta}}{\chi^2_{p,1-\delta}} \to 1$ 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.

CHAPTER 2

PREDICTION INTERVALS AFTER VARIABLE SELECTION

If n/p is large, the PI (1.11) can be used for the variable selection estimators with

$$\hat{m}(\boldsymbol{x}_f) = \hat{Y}_f = \boldsymbol{x}_{f,I_d}^T \hat{\boldsymbol{\beta}}_{I_d}, \tag{2.1}$$

where I_d denotes the index of predictors selected from the variable selection method. For example, $I_d = I_{min}$ is the model that minimizes C_p for forward selection. Now we want I_d to use $d = M = \min(\lceil n/J \rceil, p)$ variables where n/p is not necessarily large.

PI (1.11) needs 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 $\max_{i=1,\cdots,n} |r_i - e_i| \xrightarrow{P} 0$ and the sample quantiles of the residuals estimate the population quantiles of the error distribution.

The Cauchy Schwartz inequality says $|\boldsymbol{a}^T \boldsymbol{b}| \leq ||\boldsymbol{a}|| \cdot ||\boldsymbol{b}||$.

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

$$|r_i - e_i| = |Y_i - \boldsymbol{x}_i^T \hat{\boldsymbol{\beta}} - (Y_i - \boldsymbol{x}_i^T \boldsymbol{\beta})| = |\boldsymbol{x}_i^T (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})| \tag{2.2}$$

and

$$\sqrt{n} \max_{i=1,\cdots,n} |r_i - e_i| \leq (\max_{i=1,\cdots,n} ||\boldsymbol{x}_i||) ||\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})|| = O_P(1) \tag{2.3}$$

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

Let d be a crude estimate of the model degrees of freedom. For forward selection with OLS, lasso, and relaxed lasso, $\hat{\boldsymbol{\beta}}_{I_d}$ is a $d \times 1$ vector. This means that $\hat{\boldsymbol{\eta}}_{I_d}$ is of dimension $(d-1) \times 1$. The Olive (2017c) and Pelawa Watagoda and Olive (2017) PI that can work if $n \gg p$ or $p > n$ is defined below. This PI is similar to the Olive (2013) PI (1.11) with p replaced by d, but some care needs to be taken so that the PI is well defined and does not have infinite length.

Let $q_n = \min(1-\delta+0.05, 1-\delta+\frac{d}{n})$ for $\delta > 0.1$ and $q_n = \min(1-\frac{\delta}{2}, 1-\delta+\frac{10\delta d}{n})$ otherwise.

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

$$c = \lceil nq_n \rceil \tag{2.4}$$

and

$$b_n = \begin{cases} (1+\frac{15}{n})\sqrt{\frac{n+2d}{n-d}}, & \text{if } d \leq \frac{8n}{9}, \\ 5(1+\frac{15}{n}), & \text{otherwise.} \end{cases}$$

Now if the shorth(c) of the residuals is computed as $[r_{(s)}, r_{(s+c-1)}] = [\tilde{\xi}_{\delta_1}, \tilde{\xi}_{1-\delta_2}]$, then a $100(1-\delta)\%$ large sample PI for Y_f is

$$[\hat{m}(\boldsymbol{x}_f) + b_n\tilde{\xi}_{\delta_1}, \hat{m}(\boldsymbol{x}_f) + b_n\tilde{\xi}_{1-\delta_2}]. \tag{2.5}$$

CHAPTER 3

EXAMPLES AND SIMULATIONS

Let $\boldsymbol{x} = (1\ \boldsymbol{u}^T)^T$ where $\boldsymbol{u}$ is the $(p-1) \times 1$ vector of nontrivial predictors. In order to simulate the data we generated $\boldsymbol{w}_i \sim N_{p-1}(\boldsymbol{0}, \boldsymbol{I})$, for $i = 1, \cdots, n$, where the $m = p-1$ elements of the vector $\boldsymbol{w}_i$ were (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 $\boldsymbol{u} = \boldsymbol{A}\boldsymbol{w}_i$ so that $Cov(\boldsymbol{u}) = \sum_{\boldsymbol{u}} = \boldsymbol{A}\boldsymbol{A}^T = (\sigma_{ij})$, where the diagonal entries $\sigma_{ii} = [1 + (m-1)\psi^2]$ and the off diagonal entries $\sigma_{ij} = [2\psi + (m-2)\psi^2]$.

This means the correlations are $cor(x_i, x_j) = \rho = \frac{(2\psi + (m-2)\psi^2)}{(1+(m-1)\psi^2)}$ for $i \neq j$, where x_i and x_j are nontrivial predictors.

If $\psi = \frac{1}{\sqrt{cp}}$, then $\rho \to \frac{1}{(c+1)}$ as $p \to \infty$ where $c > 0$. As ψ gets close to 1, the predictor vectors cluster about the line in the direction $(1, ..., 1)^T$. Then $Y_i = 1 + x_{i,2} + \cdots + x_{i,k} + e_i$ 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 different types:

i) $N(0,1)$

ii) t_3

iii) $EXP(1) - 1$

iv) $uniform(-1, 1)$

v) $0.9N(0,1) + 0.1N(0,100)$

The lengths of the asymptotically optimal 95% PIs are:

i) $3.92 = 2(1.96)$

ii) 6.365

iii) 2.996

iv) $1.90 = 2(0.95)$

v) 13.490.

Suppose the simulation uses K runs with $W_i = 1$ if Y_f is in the ith PI, and $W_i = 0$ otherwise, for $i = 1, ..., K$. Then the W_i are (iid) $binomial(1, 1 - \delta_n)$ where $1 - \delta_n = \rho_n$ is the true coverage of the PI for a sample size of n.

Let $\hat{\rho}_n = \overline{W}$. Since $\sum_{i=1}^{K} W_i \sim binomial(K, \rho_n)$, the standard error $SE(\overline{W}) = \sqrt{\rho_n \frac{1-\rho_n}{K}}$. For $K = 5000$ runs and ρ_n near 0.9, we have $3SE(\overline{W}) \approx 0.01$.

Therefore, an observed coverage of $\hat{\rho}_n$ within 0.01 of the nominal coverage of $1 - \delta$ suggests that there is no reason to doubt that the nominal PI coverage is the same as the observed coverage. This means that for a large sample 95% PI we want the observed coverage to be between 0.94 and 0.96. Coverage slightly higher than the nominal coverage is better than coverage slightly lower than the nominal coverage.

Lasso uses the model that combines $d - 1$ nonzero $\hat{\eta}_i$ with the λ closest to 0 and does not necessarily fit the full OLS model if $\frac{n}{J} \geq p$. Values of J used were $5, 10, 20, 50,$ and $\lceil \frac{n}{p} \rceil$ as long as $J \leq \frac{n}{p}$ since $\frac{n}{J} \geq p$ uses the $\approx$ full model. The selected model then has d variables total with the intercept added.

Each simulation was run with 5000 iterations. Parameter values used were

$$\begin{cases} p = 20,\ 40,\ n,\ 2n \\ k = 1,\ 19,\ p - 1 \\ \psi = 0,\ \frac{1}{\sqrt{p}},\ 0.9 \end{cases}$$

An observed coverage in [0.94, 0.96] gives no reason to doubt that the PI has an actual nominal coverage of 0.95, as discussed above.

Some sample R code is shown below. For the 5000 runs of the nominal large sample 95% PI, the observed coverage for lasso regression was 0.960 and the average length was 4.174 while the observed coverage for relaxed lasso regression was 0.961 and the average length was 4.194.

```
library(glmnet)
drelpisim(n=1000,p=20,k=1,J=10,nruns=5000,psi=0,type=1)
$lassopicov
[1] 0.9596
$lassopimenlen
[1] 4.174159
$relpicov
[1] 0.961
$relpimenlen
[1] 4.194305
```

CHAPTER 4

ERROR TYPE 1 EXAMPLES

Table 4.1. PI coverage and length for error type 1 (runs = 5000)

n	p	k	J	ψ	lassoPIcov	lassoPIlen	relPIcov	relPIlen
100	20	19	50	0	0.9582	18.08926599	0.917	14.24679612
100	40	1	50	$1/\sqrt{p}$	0.9638	4.744004538	0.9586	4.363513327
100	40	1	10	0.9	0.9868	5.45556982	0.9866	5.434346541
100	40	1	10	0.9	0.9868	5.45556982	0.9866	5.434346541
100	40	1	50	0.9	0.9692	16.57428973	0.9646	4.377930386
100	40	1	5	0.9	0.9968	6.248819654	0.997	6.225604981
100	100	1	20	$1/\sqrt{p}$	0.9788	5.106903347	0.9758	4.95681708
100	200	1	10	0.9	0.9898	7.950753255	0.9828	5.352689147
100	200	19	10	$1/\sqrt{p}$	0.987	83.37070413	0.9308	22.54299607
400	20	1	5	0	0.9746	4.691673	0.9784	4.747554
400	20	1	10	$1/\sqrt{p}$	.9756	4.693017	0.9802	4.776718
400	20	19	20	$1/\sqrt{p}$	0.9762	4.745157	0.9768	4.691703
400	40	1	5	0	0.9762	4.90066	0.984	5.035935
400	40	19	5	$1/\sqrt{p}$	0.984	5.036201	0.9836	4.97781
400	40	39	10	0.9	0.988	35.33387	0.9798	8.118856
400	400	1	5	$1/\sqrt{p}$	0.4062	3.540859	1	16.34712
400	400	19	10	$1/\sqrt{p}$	1	20.61424	1	20.41068
400	800	19	10	$1/\sqrt{p}$	1	20.54067	1	19.94047
400	800	799	20	0	0.3956	27.93375	0.3612	29.0368
1000	20	1	5	0	0.9568	4.176917	0.9568	4.196162
1000	20	1	5	$1/\sqrt{p}$	0.958	4.176156849	0.961	4.204120602

Table 4.2. PI coverage and length for error type 1 (runs = 5000)

n	p	k	J	ψ	lassoPIcov	lassoPIlen	relPIcov	relPIlen
1000	20	1	5	0.9	0.9652	4.212432	0.9652	4.208432
1000	20	1	10	0	0.9628	4.177364	0.9622	4.206683
1000	20	1	10	$1/\sqrt{p}$	0.9612	4.214461	0.9618	4.210421
1000	20	1	10	0.9	0.9616	4.174927	0.9636	4.194475
1000	20	1	50	0.9	0.9626	4.211135	0.9618	4.206931
1000	20	19	5	0	0.9614	4.176652	0.962	4.174929
1000	20	19	5	$1/\sqrt{p}$	0.959	4.551397	0.9588	4.176465
1000	20	19	5	0.9	0.9636	9.683286	0.9614	4.468264
1000	20	19	10	0	0.9534	4.177427	0.9544	4.176055
1000	20	19	10	$1/\sqrt{p}$	0.9588	4.217715	0.9602	4.173525
1000	20	19	10	0.9	0.964	9.689478	0.959	4.469343
1000	20	19	50	0.9	0.96	9.6802	0.96	4.469252
1000	40	1	5	0	0.9732	4.489033	0.978	4.534692
1000	40	1	5	$1/\sqrt{p}$	0.9766	4.491209	0.9782	4.559454
1000	40	1	5	0.9	0.9774	4.578295	0.9772	4.467579
1000	40	1	10	0	0.9674	4.491446	0.9722	4.535254
1000	40	1	10	$1/\sqrt{p}$	0.9688	4.487312	0.974	4.556397
1000	40	1	10	0.9	0.9766	4.578448	0.9776	4.568738
1000	40	1	20	0	0.9706	4.486926	0.9742	4.530423
1000	40	1	20	$1/\sqrt{p}$	0.9728	4.490799	0.9764	4.561005
1000	40	1	20	0.9	0.9784	4.5779	0.9784	4.568175
1000	40	1	25	0	0.9726	4.486757	0.9748	4.532523
1000	40	1	25	$1/\sqrt{p}$	0.9746	4.489086	0.977	4.558176
1000	40	1	25	0.9	0.9774	4.57928	0.978	4.469908
1000	40	19	5	0	0.977	16.49239	0.9762	4.578055

Table 4.3. PI coverage and length for error type 1 (runs = 5000)

n	p	k	J	ψ	lassoPIcov	lassoPIlen	relPIcov	relPIlen
1000	40	19	5	$1/\sqrt{p}$	0.9694	4.563947	0.9706	4.517777
1000	40	19	5	0.9	0.9732	14.92449	0.9778	5.403383
1000	40	19	10	0	0.9706	4.491501	0.9692	4.512656
1000	40	19	10	$1/\sqrt{p}$	0.9636	4.564619	0.974	4.519277
1000	40	19	10	0.9	0.9796	14.92232	0.9694	5.403476
1000	40	39	5	0	0.9726	4.491956	0.9718	4.488374
1000	40	39	5	$1/\sqrt{p}$	0.974	6.195057	0.9688	4.485727
1000	40	39	5	0.9	0.9782	31.49523	0.9744	7.400381
1000	40	39	10	0	0.971	4.490206	0.9722	4.48607
1000	40	39	10	$1/\sqrt{p}$	0.972	6.196794	0.9698	4.491277
1000	40	39	10	0.9	0.9762	31.49008	0.9686	7.398837
1000	1000	1	5	0	0.9756	4.851105	0.9674	4.958444
1000	1000	1	5	$1/\sqrt{p}$	0.9762	4.854484	0.9946	5.798404
1000	1000	1	10	0	0.9768	4.663095	0.9602	4.574495
1000	1000	1	10	$1/\sqrt{p}$	0.9734	4.662686	0.9874	5.09361
1000	1000	19	5	0	0.982	4.963937	0.974	4.967163
1000	1000	19	10	0	0.984	4.83071	0.98	4.605999
1000	1000	19	10	0.9	0.992	81.57082	0.984	6.097495
1000	1000	19	20	$1/\sqrt{p}$	0.99	6.188108	0.99	4.569792
1000	1000	19	20	0.9	0.99	76.12693	0.99	5.704682
1000	1000	999	5	$1/\sqrt{p}$	1	1470.192	0.91	249.2783
1000	1000	999	10	$1/\sqrt{p}$	0.97	2404.338	0.93	369.515
1000	1000	999	20	$1/\sqrt{p}$	1	3275.38	0.93	519.667
1000	2000	1999	20	0.9	0.9816	11810.65	0.9708	232.7601
1000	2000	19	1	0.9	1	522.0454	1	26.00606

CHAPTER 5

ERROR TYPE 2 EXAMPLES

Table 5.1. PI coverage and length for error type 2 (runs = 5000)

n	p	k	J	ψ	lassoPIcov	lassoPIlen	relPIcov	relPIlen
100	20	19	50	0	0.9782	9.986011	0.9786	9.982043
100	40	1	50	$1/\sqrt{p}$	0.9756	11.16431	0.9898	13.77597
100	40	1	10	0.9	0.9904	14.87895	0.9906	14.83489
100	40	1	50	0.9	0.9942	14.80818	0.9938	14.78451
100	40	1	5	0.9	0.9922	14.8138	0.9918	14.784
100	100	1	20	$1/\sqrt{p}$	0.2556	4.469624	0.9998	29.72624
100	100	19	5	0	0.298	5.059401	0.9982	29.49392
100	100	19	10	0	0.3028	5.003609	0.9978	29.47892
100	100	19	10	0.9	1	123.8077	0.9996	40.36388
100	100	19	20	$1/\sqrt{p}$	0.9774	18.10836	0.9992	30.30216
100	100	19	20	0.9	1	124.164	0.9996	40.03507
100	100	99	5	$1/\sqrt{p}$	0.9994	78.594	0.8896	28.47352
100	100	99	10	$1/\sqrt{p}$	0.999	78.53506	0.8912	28.45027
100	100	99	20	$1/\sqrt{p}$	0.9986	78.69514	0.892	28.36202
100	200	1	10	0.9	0.9996	42.88319	0.9996	41.16636
100	200	19	10	$1/\sqrt{p}$	0.999	28.28844	0.9988	27.85977
100	200	19	1	0.9	1	178.5396	0.9994	38.96447
100	200	199	20	0.9	1	1895.293	1	124.7427
400	20	1	5	0	0.9684	8.408982	0.9706	8.540036
400	20	1	10	$1/\sqrt{p}$	0.9736	8.391781	0.9762	8.582045
400	20	19	20	$1/\sqrt{p}$	0.9708	8.477878	0.9694	8.454392

Table 5.2. PI coverage and length for error type 2 (runs = 5000)

n	p	k Jj	ψ	lassoPIcov	lassoPIlen	relPIcov	relPIlen	
400	40	1	5	0	0.9734	8.735258	0.9778	9.023681
400	40	19	5	$1/\sqrt{p}$	0.9736	8.955075	0.9724	8.628296
400	40	39	10	0.9	0.9876	34.65113	0.9816	10.72025
400	400	1	5	$1/\sqrt{p}$	0.3268	5.311224	0.9996	27.90181
400	400	19	10	$1/\sqrt{p}$	0.9842	20.23788	0.9994	32.4073
400	800	19	10	$1/\sqrt{p}$	0.9998	35.3136	1	34.97075
400	800	799	20	0	0.9738	127.8741	0.9318	107.7702
1000	20	1	5	0	0.9574	6.991185	0.9582	7.028166
1000	20	1	5	$1/\sqrt{p}$	0.9556	6.978823	0.9571	7.032174
1000	20	1	5	0.9	0.9552	7.054695	0.9566	7.051979
1000	20	1	10	0	0.958	6.985568	0.9586	7.02243
1000	20	1	10	$1/\sqrt{p}$	0.9622	7.00362	0.9644	7.061007
1000	20	1	10	0.9	0.9576	7.054468	0.9582	7.051245
1000	20	1	50	0.9	0.9638	7.051513	0.9642	7.048241
1000	20	19	5	0	0.957	6.990788	0.9568	6.989791
1000	20	19	5	$1/\sqrt{p}$	0.9604	7.005239	0.96	6.983021
1000	20	19	5	0.9	0.96	9.627046	0.9576	7.121022
1000	20	19	10	0	0.9574	6.980051	0.957	6.979446
1000	20	19	10	$1/\sqrt{p}$	0.9532	7.019064	0.953	6.997742
1000	20	19	10	0.9	0.9622	9.631616	0.9556	7.129047
1000	20	19	50	0.9	0.9622	9.621963	0.9594	7.121329
1000	40	1	5	0	0.9656	7.905414	0.9674	8.0034
1000	40	1	5	$1/\sqrt{p}$	0.9698	7.905535	0.9742	8.057311
1000	40	1	5	0.9	0.972	8.086521	0.9718	8.072032
1000	40	1	10	0	0.9626	7.904883	0.9642	8.001917

Table 5.3. PI coverage and length for error type 2 (runs = 5000)

n	p	k	J	ψ	lassoPIcov	lassoPIlen	relPIcov	relPIlen
1000	40	1	10	$1/\sqrt{p}$	0.9696	7.896571	0.9738	8.053807
1000	40	1	10	0.9	0.9714	8.084686	0.9714	8.071757
1000	40	1	20	0	0.9672	7.891193	0.9688	7.99056
1000	40	1	20	$1/\sqrt{p}$	0.9668	7.901988	0.968	8.054795
1000	40	1	20	0.9	0.9686	8.095528	0.9682	8.082035
1000	40	1	25	0	0.9674	7.891272	0.97	7.98888
1000	40	1	25	$1/\sqrt{p}$	0.971	7.901481	0.972	8.053153
1000	40	1	25	0.9	0.9736	8.094453	0.9736	8.082428
1000	40	19	5	0	0.97	7.903042	0.9708	7.950502
1000	40	19	5	$1/\sqrt{p}$	0.9724	7.987886	0.9706	7.964978
1000	40	19	5	0.9	0.974	15.10558	0.967	8.39797
1000	40	19	10	0	0.9698	7.88747	0.9714	7.935441
1000	40	19	10	$1/\sqrt{p}$	0.973	7.982947	0.9726	7.960083
1000	40	19	10	0.9	0.974	15.10891	0.9728	8.403158
1000	40	39	5	0	0.9728	7.904407	0.9722	7.901676
1000	40	39	5	$1/\sqrt{p}$	0.9662	8.001848	0.9646	7.913011
1000	40	39	5	0.9	0.9756	30.37885	0.9666	9.677735
1000	40	39	10	0	0.9652	7.893228	0.9658	7.891447
1000	40	39	10	$1/\sqrt{p}$	0.972	7.976662	0.9726	7.890466
1000	40	39	10	0.9	0.9746	30.375	0.9708	9.673675
1000	1000	1	5	0	0.2738	4.78834	0.9986	27.07884
1000	1000	1	5	$1/\sqrt{p}$	0.4202	6.086135	0.999	28.12553
1000	1000	1	10	0	0.2768	4.757204	0.9986	27.04844
1000	1000	1	10	$1/\sqrt{p}$	0.4226	6.081164	1	28.13339

CHAPTER 6

ERROR TYPE 3 EXAMPLES

Table 6.1. PI coverage and length for error type 3 (runs = 5000)

n	p	k	J	ψ	lassoPIcov	lassoPIlen	relPIcov	relPIlen
100	20	19	50	0	0.9824	5.665902	0.9828	5.661451
100	40	1	50	$1/\sqrt{p}$	0.9774	6.53667	0.9934	7.571672
100	40	1	10	0.9	0.9968	7.451577	0.9968	7.500225
100	40	1	50	0.9	0.996	7.495749	0.9958	7.549283
100	40	1	5	0.9	0.9976	7.494217	0.9976	7.54809
100	100	1	20	$1/\sqrt{p}$	0.2974	3.075235	0.9998	17.69779
100	100	19	5	0	0.4696	3.888439	0.9996	16.14255
100	100	19	10	0	0.4776	3.890055	0.9998	16.09453
100	100	19	10	0.9	1	125.3866	0.9998	25.66427
100	100	19	20	$1/\sqrt{p}$	1	18.13938	0.9996	18.60395
100	100	19	20	0.9	1	125.6306	1	25.67769
100	100	99	5	$1/\sqrt{p}$	0.9994	78.57459	0.8868	26.70603
100	100	99	10	$1/\sqrt{p}$	0.9984	78.53441	0.8838	26.73516
100	100	99	20	$1/\sqrt{p}$	0.9994	78.53441	0.8838	26.73516
100	200	1	10	0.9	1	21.98743	1	21.90229
100	200	19	10	$1/\sqrt{p}$	0.9998	18.19313	0.9998	16.5866
100	200	19	1	0.9	1	179.0192	1	24.62364
100	200	199	20	0.9	1	1891.511	1	121.1727
400	20	1	5	0	0.9796	4.321021	0.9824	4.275604
400	20	1	10	$1/\sqrt{p}$	0.976	4.320263	0.978	4.227706
400	20	19	20	$1/\sqrt{p}$	0.9714	4.44452	0.972	4.323275

Table 6.2. PI coverage and length for error type 3 (runs = 5000)

n	p	k	J	ψ	lassoPIcov	lassoPIlen	relPIcov	relPIlen
400	40	1	5	0	0.9764	4.659659	0.9814	4.642859
400	40	19	5	$1/\sqrt{p}$	0.9802	4.766706	0.9812	4.663903
400	40	39	10	0.9	0.9882	35.3124	0.9792	8.236873
400	400	1	5	$1/\sqrt{p}$	0.4152	3.53642	1	16.61972
400	400	19	10	$1/\sqrt{p}$	1	20.1722	1	19.97523
400	800	19	10	$1/\sqrt{p}$	1	20.33964	1	19.67565
400	800	799	20	0	0.9766	127.7357	0.927	107.5949
1000	20	1	5	0	0.9594	3.49285	0.9596	3.50406
1000	20	1	5	$1/\sqrt{p}$	0.9582	3.546851	0.9616	3.46505
1000	20	1	5	0.9	0.9582	3.414522	0.9582	3.440118
1000	20	1	10	0	0.9572	3.552344	0.9566	3.508031
1000	20	1	10	$1/\sqrt{p}$	0.962	3.542788	0.964	3.460628
1000	20	1	10	0.9	0.967	3.412949	0.9666	3.437629
1000	20	1	50	0.9	0.963	3.417989	0.962	3.442714
1000	20	19	5	0	0.9606	3.460349	0.9598	3.553305
1000	20	19	5	$1/\sqrt{p}$	0.9624	3.704012	0.9624	3.558968
1000	20	19	5	0.9	0.962	9.71626	0.9558	4.219468
1000	20	19	10	0	0.961	3.46658	0.9612	3.559566
1000	20	19	10	$1/\sqrt{p}$	0.959	3.702529	0.9584	3.557582
1000	20	19	10	0.9	0.9648	9.709805	0.9624	4.214765
1000	20	19	50	0.9	0.9618	9.718041	0.9568	4.212111
1000	40	1	5	0	0.9686	4.030873	0.969	3.976922
1000	40	1	5	$1/\sqrt{p}$	0.9688	4.030965	0.9714	3.915248
1000	40	1	5	0.9	0.9746	3.823763	0.974	3.856993
1000	40	1	10	0	0.969	4.032271	0.9704	3.979124

Table 6.3. PI coverage and length for error type 3 (runs = 5000)

n	p	k	J	ψ	lassoPIcov	lassoPIlen	relPIcov	relPIlen
1000	40	1	10	$1/\sqrt{p}$	0.971	4.030595	0.976	3.913966
1000	40	1	10	0.9	0.9786	3.824773	0.9784	3.86008
1000	40	1	20	0	0.9694	4.031046	0.9708	3.977078
1000	40	1	20	$1/\sqrt{p}$	0.9716	4.03867	0.9734	3.91797
1000	40	1	20	0.9	0.976	3.825773	0.9754	3.858879
1000	40	1	25	0	0.9718	4.029896	0.9736	3.977106
1000	40	1	25	$1/\sqrt{p}$	0.97	4.028143	0.9708	3.911009
1000	40	1	25	0.9	0.9736	3.828638	0.9742	3.862054
1000	40	19	5	0	0.9718	4.025381	0.9722	4.014004
1000	40	19	5	$1/\sqrt{p}$	0.971	4.131319	0.9716	4.004463
1000	40	19	5	0.9	0.973	14.95493	0.9708	5.396035
1000	40	19	10	0	0.9674	4.021581	0.9684	4.010711
1000	40	19	10	$1/\sqrt{p}$	0.975	4.130375	0.9758	4.005739
1000	40	19	10	0.9	0.9754	14.95976	0.9724	5.377959
1000	40	39	5	0	0.9766	4.054429	0.9762	4.04348
1000	40	39	5	$1/\sqrt{p}$	0.9716	6.272166	0.9736	4.043151
1000	40	39	5	0.9	0.9764	31.52734	0.9708	7.48032
1000	40	39	10	0	0.9732	4.04486	0.9726	4.033237
1000	40	39	10	$1/\sqrt{p}$	0.9672	6.268707	0.9674	4.038025
1000	40	39	10	0.9	0.975	31.4999	0.9634	7.469625
1000	1000	1	5	0	0.9786	4.605857	0.9702	4.989658
1000	1000	1	5	$1/\sqrt{p}$	0.9776	4.611052	0.9888	5.26497
1000	1000	1	10	0	0.9772	4.261538	0.968	4.531791
1000	1000	1	10	$1/\sqrt{p}$	0.9764	4.265001	0.9844	4.532029

CHAPTER 7

ERROR TYPE 4 EXAMPLES

Table 7.1. PI coverage and length for error type 4 (runs = 5000)

n	p	k	J	ψ	lassoPIcov	lassoPIlen	relPIcov	relPIlen
100	20	19	50	0	0.9816	5.66	0.9812	5.65629
100	40	1	50	$1/\sqrt{p}$	0.9812	6.517709	0.9964	7.533506
100	40	1	10	0.9	0.9956	7.485936	0.9954	7.539107
100	40	1	50	0.9	0.9966	7.462941	0.9968	7.52339
100	40	1	5	0.9	0.9968	7.473067	0.9966	7.523167
100	100	1	20	$1/\sqrt{p}$	0.3284	3.066278	1	17.65001
100	100	19	5	0	0.4634	3.90054	0.9992	16.02386
100	100	19	10	0	0.4642	3.901203	0.9988	15.97603
100	100	19	10	0.9	1	125.4736	1	25.68199
100	100	19	20	$1/\sqrt{p}$	1	18.15398	1	18.63717
100	100	19	20	0.9	1	125.1457	1	25.65453
100	100	99	5	$1/\sqrt{p}$	0.9994	78.57968	0.8812	26.84516
100	100	99	10	$1/\sqrt{p}$	0.999	78.37046	0.8892	26.74216
100	100	99	20	$1/\sqrt{p}$	0.9998	78.48123	0.8776	26.74871
100	200	1	10	0.9	1	22.0277	1	21.91787
100	200	19	10	$1/\sqrt{p}$	0.9998	18.20713	0.9998	16.5942
100	200	19	1	0.9	1	179.5306	1	24.52998
100	200	199	20	0.9	1	1895.625	1	120.9471
400	20	1	5	0	0.9718	4.323666	0.9738	4.27713
400	20	1	10	$1/\sqrt{p}$	0.9736	4.311256	0.9758	4.218829
400	20	19	20	$1/\sqrt{p}$	0.9778	4.446429	0.976	4.32748

Table 7.2. PI coverage and length for error type 4 (runs = 5000)

n	p	k	J	ψ	lassoPIcov	lassoPIlen	relPIcov	relPIlen
400	40	1	5	0	0.9756	4.654701	0.9808	4.641708
400	40	19	5	$1/\sqrt{p}$	0.978	4.76252	0.9776	4.662213
400	40	39	10	0.9	0.9866	35.34412	0.982	8.196287
400	400	1	5	$1/\sqrt{p}$	0.4206	3.555375	1	16.62904
400	400	19	10	$1/\sqrt{p}$	1	20.1319	1	19.95076
400	800	19	10	$1/\sqrt{p}$	1	20.31236	1	19.66508
400	800	799	20	0	0.3932	27.9613	0.3706	29.01742
1000	20	1	5	0	0.9566	3.548062	0.959	3.503186
1000	20	1	5	$1/\sqrt{p}$	0.959	3.548597	0.9608	3.465762
1000	20	1	5	0.9	0.9616	3.410038	0.961	3.434517
1000	20	1	10	0	0.9568	3.548401	0.959	3.504275
1000	20	1	10	$1/\sqrt{p}$	0.959	3.546294	0.9624	3.463378
1000	20	1	10	0.9	0.9638	3.427024	0.9646	3.442302
1000	20	1	50	0.9	0.9652	3.419045	0.965	3.444963
1000	20	19	5	0	0.96	3.55797	0.9596	3.551597
1000	20	19	5	$1/\sqrt{p}$	0.9598	3.707907	0.961	3.560731
1000	20	19	5	0.9	0.9604	9.710046	0.9568	4.211821
1000	20	19	10	0	0.9602	3.566101	0.96	3.559749
1000	20	19	10	$1/\sqrt{p}$	0.9638	3.702234	0.9636	3.557424
1000	20	19	10	0.9	0.9608	9.707596	0.9578	4.215385
1000	20	19	50	0.9	0.9596	9.722391	0.9572	4.214272
1000	40	1	5	0	0.9726	4.028099	0.9762	3.975737
1000	40	1	5	$1/\sqrt{p}$	0.9762	4.023835	0.9772	3.910505
1000	40	1	5	0.9	0.9762	3.822411	0.9758	3.85708
1000	40	1	10	0	0.9722	4.027183	0.9742	3.973056

Table 7.3. PI coverage and length for error type 4 (runs = 5000)

n	p	k	J	ψ	lassoPIcov	lassoPIlen	relPIcov	relPIlen
1000	40	1	10	$1/\sqrt{p}$	0.9684	4.024621	0.9704	3.909709
1000	40	1	10	0.9	0.973	3.8287	0.9734	3.862166
1000	40	1	20	0	0.97	4.036794	0.9732	3.984156
1000	40	1	20	$1/\sqrt{p}$	0.9746	4.031043	0.9776	3.917079
1000	40	1	20	0.9	0.9756	3.828959	0.9746	3.861855
1000	40	1	25	0	0.9764	4.026741	0.978	3.973227
1000	40	1	25	$1/\sqrt{p}$	0.973	4.024812	0.9758	3.910337
1000	40	1	25	0.9	0.9758	3.824224	0.9754	3.857322
1000	40	19	5	0	0.9718	4.016771	0.9722	4.006705
1000	40	19	5	$1/\sqrt{p}$	0.9706	4.125796	0.9716	3.999824
1000	40	19	5	0.9	0.9764	14.93977	0.9724	5.378307
1000	40	19	10	0	0.9716	4.01791	0.9722	4.005528
1000	40	19	10	$1/\sqrt{p}$	0.9738	4.131782	0.9732	4.006684
1000	40	19	10	0.9	0.9738	14.94511	0.97	5.37966
1000	40	39	5	0	0.9702	4.04914	0.9708	4.038427
1000	40	39	5	$1/\sqrt{p}$	0.971	6.273491	0.971	4.039874
1000	40	39	5	0.9	0.9728	31.49195	0.9714	7.490543
1000	40	39	10	0	0.97	4.053798	0.9702	4.042019
1000	40	39	10	$1/\sqrt{p}$	0.9728	6.266155	0.97	4.039419
1000	40	39	10	0.9	0.9794	31.48421	0.9676	7.47915
1000	1000	1	5	0	0.9778	4.607652	0.9698	4.990698
1000	1000	1	5	$1/\sqrt{p}$	0.978	4.606934	0.988	5.255228
1000	1000	1	10	0	0.9742	4.262206	0.9664	4.531228
1000	1000	1	10	$1/\sqrt{p}$	0.9754	4.269198	0.9812	4.535169

CHAPTER 8

ERROR TYPE 5 EXAMPLES

Table 8.1. PI coverage and length for error type 5 (runs = 5000)

n	p	k	J	ψ	lassoPIcov	lassoPIlen	relPIcov	relPIlen
100	20	19	50	0	0.9672	22.83954	0.967	22.86759
100	40	1	50	$1/\sqrt{p}$	0.9718	23.86942	0.9814	31.50268
100	40	1	10	0.9	0.9802	34.47685	0.9802	34.46282
100	40	1	50	0.9	0.9814	34.86414	0.9816	34.84943
100	40	1	5	0.9	0.979	34.59846	0.9792	34.58271
100	100	1	20	$1/\sqrt{p}$	0.2204	7.788296	0.9972	61.71316
100	100	19	5	0	0.218	7.800493	0.997	60.18609
100	100	19	10	0	0.224	7.676551	0.9958	60.12018
100	100	19	10	0.9	1	137.3027	0.9986	91.15347
100	100	19	20	$1/\sqrt{p}$	0.7578	18.22891	0.9982	57.87845
100	100	19	20	0.9	1	137.6473	0.9976	91.11067
100	100	99	5	$1/\sqrt{p}$	0.9994	79.20089	0.927	35.3101
100	100	99	10	$1/\sqrt{p}$	0.9988	79.42998	0.9292	35.5124
100	100	99	20	$1/\sqrt{p}$	0.9988	79.50519	0.9218	35.41802
100	200	1	10	0.9	0.9978	100.0861	0.9982	98.9453
100	200	19	10	$1/\sqrt{p}$	0.9948	49.8871	0.9958	60.30151
100	200	19	1	0.9	1	185.2953	0.9984	89.57376
100	200	199	20	0.9	1	1893.2	1	139.6943
400	20	1	5	0	0.9684	21.20359	0.969	21.65993
400	20	1	10	$1/\sqrt{p}$	0.9662	21.27113	0.9686	21.94057
400	20	19	20	$1/\sqrt{p}$	0.9706	21.17575	0.9704	21.1763

Table 8.2. PI coverage and length for error type 5 (runs = 5000)

n	p	k	J	ψ	lassoPIcov	lassoPIlen	relPIcov	relPIlen
400	40	1	5	0	0.9684	21.72023	0.9704	22.75018
400	40	19	5	$1/\sqrt{p}$	0.971	22.0886	0.9708	22.31344
400	40	39	10	0.9	0.9826	36.21113	0.9692	22.90886
400	400	1	5	$1/\sqrt{p}$	0.2716	9.144157	0.9988	59.66783
400	400	19	10	$1/\sqrt{p}$	0.7908	18.15376	0.999	61.8995
400	800	19	10	$1/\sqrt{p}$	0.9994	77.35313	0.9998	84.05446
400	800	799	20	0	0.9726	128.867	0.9246	108.4209
1000	20	1	5	0	0.9534	15.7748	0.954	15.90134
1000	20	1	5	$1/\sqrt{p}$	0.9484	15.80838	0.952	16.00015
1000	20	1	5	0.9	0.952	15.98506	0.952	15.97982
1000	20	1	10	0	0.95	15.75796	0.9508	15.87743
1000	20	1	10	$1/\sqrt{p}$	0.9548	1576826	0.9556	15.95964
1000	20	1	10	0.9	0.958	16.05504	0.9584	16.05049
1000	20	1	50	0.9	0.954	16.02897	0.9542	16.02334
1000	20	19	5	0	0.956	15.81485	0.9562	15.8148
1000	20	19	5	$1/\sqrt{p}$	0.9564	15.7631	0.956	15.76005
1000	20	19	5	0.9	0.9548	16.07489	0.9526	15.7776
1000	20	19	10	0	0.9528	15.78372	0.953	15.78393
1000	20	19	10	$1/\sqrt{p}$	0.956	15.76038	0.9568	15.75749
1000	20	19	10	0.9	0.958	16.09224	0.9566	15.7966
1000	20	19	50	0.9	0.959	16.14037	0.9596	15.83275
1000	40	1	5	0	0.965	19.8792	0.9656	20.2371
1000	40	1	5	$1/\sqrt{p}$	0.9674	19.86695	0.9686	20.43901
1000	40	1	5	0.9	0.968	20.51275	0.968	20.49809
1000	40	1	10	0	0.965	19.80434	0.9678	20.16451

Table 8.3. PI coverage and length for error type 5 (runs = 5000)

n	p	k	J	ψ	lassoPIcov	lassoPIlen	relPIcov	relPIlen
1000	40	1	10	$1/\sqrt{p}$	0.9692	19.87526	0.9696	20.44563
1000	40	1	10	0.9	0.967	20.55532	0.9666	20.53642
1000	40	1	20	0	0.9676	19.86589	0.9686	20.23198
1000	40	1	20	$1/\sqrt{p}$	0.9664	19.85278	0.9662	20.41846
1000	40	1	20	0.9	0.9668	20.50069	0.9664	20.48688
1000	40	1	25	0	0.9618	19.86085	0.9634	20.21436
1000	40	1	25	$1/\sqrt{p}$	0.967	19.83468	0.9698	20.407
1000	40	1	25	0.9	0.9656	20.53683	0.9652	20.52101
1000	40	19	5	0	0.9688	19.87051	0.9702	20.05446
1000	40	19	5	$1/\sqrt{p}$	0.9714	20.0827	0.972	20.0752
1000	40	19	5	0.9	0.9742	20.96881	0.9734	20.25004
1000	40	19	10	0	0.966	19.80117	0.965	19.98561
1000	40	19	10	$1/\sqrt{p}$	0.9668	20.1203	0.9674	20.11521
1000	40	19	10	0.9	0.9678	21.02034	0.9668	20.30503
1000	40	39	5	0	0.9644	19.86745	0.9644	19.867
1000	40	39	5	$1/\sqrt{p}$	0.965	19.88486	0.9658	19.8689
1000	40	39	5	0.9	0.9774	31.85886	0.9654	20.24388
1000	40	39	10	0	0.9664	19.80572	0.966	19.80738
1000	40	39	10	$1/\sqrt{p}$	0.9678	19.84115	0.967	19.83012
1000	40	39	10	0.9	0.9732	31.83711	0.9668	20.28476
1000	1000	1	5	0	0.9704	22.67935	0.965	20.83877
1000	1000	1	5	$1/\sqrt{p}$	0.9744	22.68762	0.9836	27.8255
1000	1000	1	10	0	0.9738	22.2961	0.9674	20.11294
1000	1000	1	10	$1/\sqrt{p}$	0.9732	22.31307	0.978	24.61319

CHAPTER 9

CONCLUSIONS

For my simulations, I chose to run mostly the same schedule of parameters for all five error types (error type 1 has a couple different test cases than the latter four) hoping to compare the resulting data values across the spectrum of error types. At the outset of this research project I anticipated, through my readings and prior knowledge of the subject matter, that relaxed lasso would give shorter prediction intervals with better coverage than regular lasso.

For the most part, this hypothesis held and was borne out in the data gathered; relaxed lasso was often better (shorter PI) than lasso and the only times there was significant difference between the two was when relaxed lasso $\ll$ lasso. In all of the simulations I ran I came across a very few test cases where lasso performed much much better than relaxed lasso.

Some observations of the simulations:

1. When relaxed lasso $\ll$ lasso it was most often with a large parameter value of $p = n$, or $2n$. This makes intuitive sense because when n/p is small ($p \geq n$) we have a lot of potential predictors compared to the number of data points and therefore a lot of noise. The large number of potential predictors makes it difficult for the process to choose the correct variables to include in the final model. This results in large prediction intervals. This inflated PI length allows relaxed lasso to perform much better in these situations, reducing the length of the PI drastically.

Examples of this reduction include:

Table 9.1. Relaxed lasso $\ll$ lasso, small n/p

n	p	k	J	ψ	error type	lcov	llen	rlcov	rllen
100	200	19	10	$1/\sqrt{p}$	1	.987	83.37070413	.9308	22.54299607
1000	1000	19	10	.9	1	.992	81.57082	.984	6.097495
100	100	99	5	$1/\sqrt{p}$	3	.9994	78.57459	.8868	26.70603
100	100	99	10	$1/\sqrt{p}$	5	.9988	79.42998	.9292	35.5124

Of course, some of the difference in lengths can be explained by the reduced coverage observed when moving from the lasso to relaxed lasso PI's.

2. Other times when relaxed lasso $\ll$ lasso included when $k \geq 19$ (i.e. $k = 19$ or $p - 1$):

Table 9.2. Relaxed lasso $\ll$ lasso, $k \geq 19$

n	p	k	J	ψ	error type	lcov	llen	rlcov	rllen
1000	40	39	5	.9	1	.9782	31.49523	.9744	7.400381
400	40	39	10	.9	2	.9876	34.65113	.9816	10.72025
1000	20	19	10	.9	4	.9608	9.707596	.9578	4.215385

This trend, as shown, is most significant when $k = p - 1$.

3. For the majority of the simulations run lasso and relaxed lasso PI coverages and lengths were very similar. This was especially true when n/p was large, for reasons discussed in the introduction.

4. If lasso $\ll$ relaxed lasso the test cases seemed to take on a specific form:

Table 9.3. Lasso $\ll$ Relaxed lasso, lasso coverage is low

n	p	k	J	ψ	error type	lcov	llen	rlcov	rllen
1000	1000	1	5	0	1	.2738	4.78834	.9986	27.07884
1000	1000	1	5	$1/\sqrt{p}$	1	.4202	6.086135	.999	28.12553
100	100	1	20	$1/\sqrt{p}$	5	.2204	7.788296	.9972	61.71316
100	100	19	10	0	5	.224	7.676551	.9958	60.12018

specifically, if the length of the relaxed lasso PI was to be significantly larger than the length of the lasso PI it was a product of low coverage of the lasso PI and high coverage on the part of the relaxed lasso PI. Understanding this trend, we can see that the test cases that produce this data (lasso $\ll$ relaxed lasso) are rather useless, because of their poor coverages, in understanding the roles of lasso vs. relaxed lasso.

5. Overall, I found that relaxed lasso performed better in producing smaller prediction intervals except in specific cases where lasso shrunk the used variables too much and produced an under covered PI. Relaxed lasso is a very good alternative to the lasso method when trying to understand prediction intervals.

The simulations were done in *R*. See R Core Team (2016). The collection of *R* functions *slpack*, available from (http://lagrange.math.siu.edu/Olive/slpack.txt), has some useful functions for the inference. The function `drelpisim` was used to do the simulation.

REFERENCES

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

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

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

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

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

[6] Olive, D.J. (2007), "Prediction Intervals for Regression," *Computational Statistics & Data Analysis*, 51, 3115-3122.

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

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

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

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

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

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

[13] Pelawa Watagoda, L.C.R. (2017), *Inference After Variable Selection*, Ph.D. Thesis, Southern Illinois University, to appear. Preprint available (lagrange.math.siu.edu/Olive/slasanthiphd.pdf).

[14] Pelawa Watagoda, L.C.R., and Olive, D.J. (2017), "Inference for Multiple Linear Regression After Model or Variable Selection," to appear. Preprint available(lagrange.math.siu.edu/Olive/ppvsinf.pdf)

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

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

VITA

Graduate School
Southern Illinois University

Craig Bartelsmeyer

Cjbartelsmeyer@siu.edu

University of Alabama
Bachelor of Science, Mathematics, May 2015

Research Paper Title:
Prediction Intervals for Lasso and Relaxed Lasso Using D Variables

Major Professor: Dr. David J. Olive