Spring 4-5-2019

BOOTSTRAP CONFIDENCE INTERVALS FOR Beta _ i USING FORWARD SELECTION WITH C_ p

Mashael Abbas Alshammari
SIUC, mashael.alshammari@siu.edu

BOOTSTRAP CONFIDENCE INTERVALS FOR β_i USING FORWARD SELECTION WITH C_p

by

Mashael Alshammari

B.S., Northern Borders University, Saudi Arabia, 2014

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, 2019

RESEARCH PAPER APPROVAL

BOOTSTRAP CONFIDENCE INTERVALS FOR β_i USING FORWARD SELECTION WITH C_p

by

Mashael Alshammari

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

Randy Hughes

John Mcsorley

Graduate School
Southern Illinois University Carbondale
April 3, 2019

AN ABSTRACT OF THE RESEARCH PAPER OF

Mashael Alshammari, for the Master of Science degree in MATHEMATICS, presented on April 3, 2019, at Southern Illinois University Carbondale.

TITLE: BOOTSTRAP CONFIDENCE INTERVALS FOR β_i USING FORWARD SELECTION WITH C_p

MAJOR PROFESSOR: Dr. David J. Olive

This paper presents three large sample confidence intervals for β_i for the multiple linear regression model $Y = \beta_1 x_1 + \cdots + \beta_p x_p + e$, after forward selection with C_p criterion.

ACKNOWLEDGMENTS

I would like to take this opportunity to thank my research advisor, Dr. David Olive for overseeing my Master's project and Dr. Randy Hughes and Dr. John Mcsorley for sitting on my committee. Thanks also to the professors of Southern Illinois University for their instruction and care over the past two years. Last but not the least, I would like to say to my parents, brothers, and sisters 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

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 $\boldsymbol{x}_i^T = (x_{i,1}, ..., x_{i,p})$ and $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)^T$ where β_1 corresponds to the intercept. Then the multiple linear regression (MLR) model is

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

for $i = 1, ..., n$. This model is also called the full model. Here n is the sample size, and assume that the random variables e_i are independent and identically distributed (iid) with variance V(e_i) $=\sigma^2$. In matrix notation, these n equations become

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e} \qquad (1.2)$$

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.

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

$$\boldsymbol{x}^T\boldsymbol{\beta} = \boldsymbol{x}_S^T\boldsymbol{\beta}_S + \boldsymbol{x}_E^T\boldsymbol{\beta}_E = \boldsymbol{x}_S^T\boldsymbol{\beta}_S \tag{1.3}$$

where $\boldsymbol{x} = (\boldsymbol{x}_S^T, \boldsymbol{x}_E^T)^T$, $\boldsymbol{x}_S$ is an $a_S \times 1$ vector, and $\boldsymbol{x}_E$ is a $(p - a_S) \times 1$ 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 a 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 (1.3) holds. Then

$$\boldsymbol{x}^T\boldsymbol{\beta} = \boldsymbol{x}_S^T\boldsymbol{\beta}_S = \boldsymbol{x}_S^T\boldsymbol{\beta}_S + \boldsymbol{x}_{I/S}^T\boldsymbol{\beta}_{(I/S)} + \boldsymbol{x}_O^T\boldsymbol{0} = \boldsymbol{x}_I^T\boldsymbol{\beta}_I, \tag{1.4}$$

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 $I_1, ..., I_p$ 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 $j = 2, ..., M = p$ where $n \geq 10p$ and p is fixed.

When there is a sequence of M submodels, the final submodel I_d needs to be selected. Let the candidate model I contain a terms, including a constant. Let $\boldsymbol{x}_I$ and $\hat{\boldsymbol{\beta}}_I$ be $a \times 1$ vectors. Then there are many criteria used to select the final submodel I_d. 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 positive constant without changing the subset I_{min} that minimizes the criterion.

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 OLS full 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).

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 using variable selection with OLS. Following Nishii (1984), $P(S \subseteq I_{min}) \to 1$ as $n \to \infty$ if C_p or AIC is used for forward selection, backward elimination, or all subsets. 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 fewer than 2^p regression models I 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 (1.3). See Pelawa Watagoda and Olive (2019) and Olive (2017a: p. 123, 2017b: p. 176).

Chapter 2 describes bootstrap confidence intervals and regions, and chapter 3 gives a simulation for confidence intervals for β_i after variable selection.

CHAPTER 2

BOOTSTRAP CONFIDENCE REGIONS

Mixture distributions are useful for variable selection since asymptotically $\hat{\boldsymbol{\beta}}_{I_{min},0}$ is a mixture distribution of $\hat{\boldsymbol{\beta}}_{I_j,0}$ where $S \subseteq I_j$. See Equation (1.3). A random vector $\boldsymbol{u}$ has a mixture distribution if $\boldsymbol{u}$ equals a random vector $\boldsymbol{u}_j$ with probability π_j for $j = 1, ..., J$.

Definition 1. The distribution of a $g \times 1$ random vector $\boldsymbol{u}$ is a mixture distribution if the cumulative distribution function (cdf) of $\boldsymbol{u}$ is

$$F_{\boldsymbol{u}}(\boldsymbol{t}) = \sum_{j=1}^{J} \pi_j F_{\boldsymbol{u}_j}(\boldsymbol{t}) \tag{2.1}$$

where the probabilities π_j satisfy $0 \leq \pi_j \leq 1$ and $\sum_{j=1}^{J} \pi_j = 1$, $J \geq 2$, and $F_{\boldsymbol{u}_j}(\boldsymbol{t})$ is the cdf of a $g \times 1$ random vector $\boldsymbol{u}_j$. Then $\boldsymbol{u}$ has a mixture distribution of the $\boldsymbol{u}_j$ with probabilities π_j.

Theorem 1. Suppose $E(h(\boldsymbol{u}))$ and the $E(h(\boldsymbol{u}_j))$ exist. Then

$$E(h(\boldsymbol{u})) = \sum_{j=1}^{J} \pi_j E[h(\boldsymbol{u}_j)] \quad \text{and} \quad \mathrm{E}(\boldsymbol{u}) = \sum_{\mathrm{j=1}}^{\mathrm{J}} \pi_{\mathrm{j}} \mathrm{E}[\boldsymbol{u}_{\mathrm{j}}]. \tag{2.2}$$

Hence $\mathrm{Cov}(\boldsymbol{u}) = E(\boldsymbol{u}\boldsymbol{u}^T) - E(\boldsymbol{u})E(\boldsymbol{u}^T) = E(\boldsymbol{u}\boldsymbol{u}^T) - E(\boldsymbol{u})[E(\boldsymbol{u})]^T =$

$\sum_{j=1}^{J} \pi_j E[\boldsymbol{u}_j \boldsymbol{u}_j^T] - E(\boldsymbol{u})[E(\boldsymbol{u})]^T =$

$$\sum_{j=1}^{J} \pi_j \mathrm{Cov}(\boldsymbol{u}_j) + \sum_{j=1}^{J} \pi_j E(\boldsymbol{u}_j)[E(\boldsymbol{u}_j)]^T - E(\boldsymbol{u})[E(\boldsymbol{u})]^T. \tag{2.3}$$

If $E(\boldsymbol{u}_j) = \boldsymbol{\theta}$ for $j = 1, ..., J$, then $E(\boldsymbol{u}) = \boldsymbol{\theta}$ and

$$Cov(\boldsymbol{u}) = \sum_{j=1}^{J} \pi_j Cov(\boldsymbol{u}_j).$$

Definition 2. The *population mean* of a random $p \times 1$ vector $\boldsymbol{X} = (X_1, ..., X_p)^T$ is

$$E(\boldsymbol{X}) = (E(X_1), ..., E(X_p))^T$$

and the $p \times p$ *population covariance matrix*

$$Cov(\boldsymbol{X}) = E(\boldsymbol{X} - E(\boldsymbol{X}))(\boldsymbol{X} - E(\boldsymbol{X}))^T = (\sigma_{ij}).$$

That is, the ij entry of $Cov(\boldsymbol{X})$ is $Cov(X_i, X_j) = \sigma_{ij}$.

Note that $Cov(\boldsymbol{X})$ is a symmetric positive semidefinite matrix. The following results are useful. If $\boldsymbol{X}$ and $\boldsymbol{Y}$ are $p \times 1$ random vectors, $\boldsymbol{a}$ a conformable constant vector, and $\boldsymbol{A}$ and $\boldsymbol{B}$ are conformable constant matrices, then

$$E(\boldsymbol{a} + \boldsymbol{X}) = \boldsymbol{a} + E(\boldsymbol{X}) \ \text{ and } \ \mathrm{E}(\boldsymbol{X} + \boldsymbol{Y}) = \mathrm{E}(\boldsymbol{X}) + \mathrm{E}(\boldsymbol{Y}) \tag{2.4}$$

and

$$E(\boldsymbol{A}\boldsymbol{X}) = \boldsymbol{A}E(\boldsymbol{X}) \ \text{ and } \ \mathrm{E}(\boldsymbol{A}\boldsymbol{X}\boldsymbol{B}) = \boldsymbol{A}\mathrm{E}(\boldsymbol{X})\boldsymbol{B}. \tag{2.5}$$

Thus

$$Cov(\boldsymbol{a} + \boldsymbol{A}\boldsymbol{X}) = Cov(\boldsymbol{A}\boldsymbol{X}) = \boldsymbol{A}Cov(\boldsymbol{X})\boldsymbol{A}^T. \tag{2.6}$$

For the multivariate normal (MVN) distribution $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Then $E(\boldsymbol{X}) = \boldsymbol{\mu}$ and

$$Cov(\boldsymbol{X}) = \boldsymbol{\Sigma}.$$

Inference will consider bootstrap confidence intervals and bootstrap confidence regions for bootstrap hypothesis testing. Applying the shorth prediction interval and the Olive (2013) prediction region to the bootstrap sample will give the bootstrap confidence intervals and regions.

Consider predicting a future test random variable Z_f given iid training data $Z_1, ..., Z_n$. A large sample $100(1-\delta)\%$ prediction interval (PI) for Z_f has the form $[\hat{L}_n, \hat{U}_n]$ where $P(\hat{L}_n \leq Z_f \leq \hat{U}_n) \to 1-\delta$ as the sample size $n \to \infty$. The shorth(c) estimator is useful for making prediction intervals. Let $Z_{(1)}, ..., Z_{(n)}$ be the order statistics of $Z_1, ..., Z_n$. Then let the shortest closed interval containing at least c of the Z_i be

$$\text{shorth(c)} = [\text{Z}_{(\text{s})}, \text{Z}_{(\text{s+c}-1)}]. \tag{2.7}$$

Let $\lceil x \rceil$ be the smallest integer $\geq x$, e.g., $\lceil 7.7 \rceil = 8$. Let

$$k_n = \lceil n(1-\delta) \rceil. \tag{2.8}$$

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

$$c = \min(n, \lceil n[1 - \delta + 1.12\sqrt{\delta/n}\ ]\ \rceil). \tag{2.9}$$

Example 1. 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 shorth(3) from the ordered data is simple. If the outlier was corrected, shorth(3) = [76,78].

```
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]
```

We also want to use bootstrap tests. Consider testing $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ versus $H_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ where $\boldsymbol{\theta}_0$ is a known $g \times 1$ vector. Given training data $\boldsymbol{z}_1, ..., \boldsymbol{z}_n$, a large sample $100(1-\delta)$% confidence region for $\boldsymbol{\theta}$ is a set $\mathcal{A}_n$ such that $P(\boldsymbol{\theta} \in \mathcal{A}_n) \to 1 - \delta$ as $n \to \infty$. Then reject H_0 if $\boldsymbol{\theta}_0$ is not in the confidence region $\mathcal{A}_n$. For model (1.1), let $\boldsymbol{\theta} = \boldsymbol{A}\boldsymbol{\beta}$ where $\boldsymbol{A}$ is a known full rank $g \times p$ matrix with $1 \leq g \leq p$.

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 $g \times 1$ 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 $g \times 1$ column vector T be a multivariate location estimator, and let the $g \times g$ symmetric positive definite matrix $\boldsymbol{C}$ be a dispersion estimator. Then the ith *squared sample Mahalanobis distance* is the scalar

$$D_i^2 = D_i^2(T, \boldsymbol{C}) = D_{\boldsymbol{z}_i}^2(T, \boldsymbol{C}) = (\boldsymbol{z}_i - T)^T \boldsymbol{C}^{-1} (\boldsymbol{z}_i - T) \tag{2.10}$$

for each observation $\boldsymbol{z}_i$. Notice that the Euclidean distance of $\boldsymbol{z}_i$ from the estimate of center T is $D_i(T, \boldsymbol{I}_g)$ where $\boldsymbol{I}_g$ is the $g \times g$ 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

$$\overline{\boldsymbol{z}} = \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{z}_i \quad \text{and} \quad \boldsymbol{S} = \frac{1}{\mathrm{n}-1} \sum_{\mathrm{i}=1}^{\mathrm{n}} (\boldsymbol{z}_{\mathrm{i}} - \overline{\boldsymbol{z}})(\boldsymbol{z}_{\mathrm{i}} - \overline{\boldsymbol{z}})^{\mathrm{T}}. \tag{2.11}$$

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

$$q_n = \min(1-\delta/2, 1-\delta+10\delta g/n), \quad \text{otherwise.} \tag{2.12}$$

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

$$c = \lceil n q_n \rceil. \tag{2.13}$$

Let $(T, \boldsymbol{C}) = (\overline{\boldsymbol{z}}, \boldsymbol{S})$, and let $D_{(U_n)}$ be the $100 q_n$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

$$\{\boldsymbol{z} : D^2_{\boldsymbol{z}}(\overline{\boldsymbol{z}}, \boldsymbol{S}) \leq D^2_{(U_n)}\}, \tag{2.14}$$

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

$$\{\boldsymbol{z} : D^2_{\boldsymbol{z}}(\overline{\boldsymbol{z}}, \boldsymbol{S}) \leq \chi^2_{g,1-\delta}\}. \tag{2.15}$$

Definition 3. Suppose that data $\boldsymbol{x}_1, ..., \boldsymbol{x}_n$ has been collected and observed. Often the data is a random sample (iid) from a distribution with cdf F. The *empirical distribution* is a discrete distribution where the $\boldsymbol{x}_i$ are the possible values, and each value is equally likely. If $\boldsymbol{w}$ is a random variable having the empirical distribution, then $p_i = P(\boldsymbol{w} = \boldsymbol{x}_i) = 1/n$ for $i = 1, ..., n$. The *cdf of the empirical distribution* is denoted by F_n.

Example 2. Let $\boldsymbol{w}$ be a random variable having the empirical distribution given by Definition 3. Show that $E(\boldsymbol{w}) = \overline{\boldsymbol{x}} \equiv \overline{\boldsymbol{x}}_n$ and $Cov(\boldsymbol{w}) = \dfrac{n-1}{n}\boldsymbol{S} \equiv \dfrac{n-1}{n}\boldsymbol{S}_n$.

Solution: Recall that for a discrete random vector, the population expected value $E(\boldsymbol{w}) = \sum \boldsymbol{x}_i p_i$ where $\boldsymbol{x}_i$ are the values that $\boldsymbol{w}$ takes with positive probability p_i. Similarly, the population covariance matrix

$$Cov(\boldsymbol{w}) = E[(\boldsymbol{w} - E(\boldsymbol{w}))(\boldsymbol{w} - E(\boldsymbol{w}))^T] = \sum (\boldsymbol{x}_i - E(\boldsymbol{w}))(\boldsymbol{x}_i - E(\boldsymbol{w}))^T p_i.$$

Hence

$$E(\boldsymbol{w}) = \sum_{i=1}^n \boldsymbol{x}_i \frac{1}{n} = \overline{\boldsymbol{x}},$$

and

$$Cov(\boldsymbol{w}) = \sum_{i=1}^{n} (\boldsymbol{x}_i - \overline{\boldsymbol{x}})(\boldsymbol{x}_i - \overline{\boldsymbol{x}})^T \frac{1}{n} = \frac{n-1}{n}\boldsymbol{S}. \ \square$$

Example 3. If $W_1, ..., W_n$ are iid from a distribution with cdf F_W, then the empirical cdf F_n corresponding to F_W is given by

$$F_n(y) = \frac{1}{n}\sum_{i=1}^{n} I(W_i \leq y)$$

where the indicator $I(W_i \leq y) = 1$ if $W_i \leq y$ and $I(W_i \leq y) = 0$ if $W_i > y$. Fix n and y. Then $nF_n(y) \sim$ binomial $(n, F_W(y))$. Thus $E[F_n(y)] = F_W(y)$ and $V[F_n(y)] = F_W(y)[1 - F_W(y)]/n$. By the central limit theorem,

$$\sqrt{n}(F_n(y) - F_W(y)) \xrightarrow{D} N(0, F_W(y)[1 - F_W(y)]).$$

Thus $F_n(y) - F_W(y) = O_P(n^{-1/2})$, and F_n is a reasonable estimator of F_W if the sample size n is large.

Suppose there is data $\boldsymbol{w}_1, ..., \boldsymbol{w}_n$ collected into an $n \times p$ matrix $\boldsymbol{W}$. Let the statistic $T_n = t(\boldsymbol{W}) = T(F_n)$ be computed from the data. Suppose the statistic estimates $\boldsymbol{\theta} = T(F)$, and let $t(\boldsymbol{W}^*) = t(F_n^*) = T_n^*$ indicate that t was computed from an iid sample from the empirical distribution F_n: a sample $\boldsymbol{w}_1^*, ..., \boldsymbol{w}_n^*$ of size n was drawn with replacement from the observed sample $\boldsymbol{w}_1, ..., \boldsymbol{w}_n$. This notation is used for von Mises differentiable statistical functions in large sample theory. See Serfling (1980, ch. 6). The *empirical* *bootstrap* or *nonparametric bootstrap* or *naive bootstrap* draws B samples of size n from the rows of $\boldsymbol{W}$, e.g. from the empirical distribution of $\boldsymbol{w}_1, ..., \boldsymbol{w}_n$. Then T^*_{jn} is computed from the jth bootstrap sample for $j = 1, ..., B$.

Example 4. Suppose the data is 1, 2, 3, 4, 5, 6, 7. Then $n = 7$ and the sample median T_n is 4. Using R, we drew $B = 2$ bootstrap samples (samples of size n drawn with replacement from the original data) and computed the sample median $T^*_{1,n} = 3$ and $T^*_{2,n} = 4$.

```
b1 <- sample(1:7,replace=T)

b1

[1] 3 2 3 2 5 2 6

median(b1)

[1] 3

b2 <- sample(1:7,replace=T)

b2

[1] 3 5 3 4 3 5 7

median(b2)

[1] 4
```

The bootstrap has been widely used to estimate the population covariance matrix

Cov(T_n) of a statistic T_n, for testing hypotheses, and for obtaining confidence regions (often confidence intervals). An iid sample $T_{1n}, ..., T_{Bn}$ of size B of the statistic would be very useful for inference, but typically we only have one sample of data and one value $T_n = T_{1n}$ of the statistic. Often $T_n = t(\boldsymbol{w}_1, ..., \boldsymbol{w}_n)$, and the bootstrap sample $T^*_{1n}, ..., T^*_{Bn}$ is formed where $T^*_{jn} = t(\boldsymbol{w}^*_{j1}, ..., \boldsymbol{w}^*_{jn})$.

The *residual bootstrap* is often useful for additive error regression models of the form $Y_i = m(\boldsymbol{x}_i) + e_i = \hat{m}(\boldsymbol{x}_i) + r_i = \hat{Y}_i + r_i$ for $i = 1, ..., n$ where the ith residual $r_i = Y_i - \hat{Y}_i$. Let $\boldsymbol{Y} = (Y_1, ..., Y_n)^T$, $\boldsymbol{r} = (r_1, ..., r_n)^T$, and let $\boldsymbol{X}$ be an $n \times p$ matrix with ith row $\boldsymbol{x}_i^T$. Then the fitted values $\hat{Y}_i = \hat{m}(\boldsymbol{x}_i)$, and the residuals are obtained by regressing $\boldsymbol{Y}$ on $\boldsymbol{X}$. Here the errors e_i are iid, and it would be useful to be able to generate B iid samples $e_{1j}, ..., e_{nj}$ from the distribution of e_i where $j = 1, ..., B$. If the $m(\boldsymbol{x}_i)$ were known, then we could form a vector $\boldsymbol{Y}_j$ where the ith element $Y_{ij} = m(\boldsymbol{x}_i) + e_{ij}$ for $i = 1, ..., n$. Then regress $\boldsymbol{Y}_j$ on $\boldsymbol{X}$. Instead, draw samples $r^*_{1j}, ..., r^*_{nj}$ with replacement from the residuals, then form a vector $\boldsymbol{Y}^*_j$ where the ith element $Y^*_{ij} = \hat{m}(\boldsymbol{x}_i) + r^*_{ij}$ for $i = 1, ..., n$. Then regress $\boldsymbol{Y}^*_j$ on $\boldsymbol{X}$.

The Olive (2017ab, 2018) prediction region method obtains a confidence region for $\boldsymbol{\theta}$ by applying the nonparametric prediction region (2.15) to the bootstrap sample $T^*_1, ..., T^*_B$. Let $\overline{T}^*$ and $\boldsymbol{S}^*_T$ be the sample mean and sample covariance matrix of the bootstrap sample. Assume $n\boldsymbol{S}^*_T \xrightarrow{P} \boldsymbol{\Sigma}_A$. See Machado and Parente (2005) for regularity conditions for this assumption.

Following Bickel and Ren (2001), let the vector of parameters $\boldsymbol{\theta} = T(F)$, 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 (has a Hadamard derivative $\dot{T}(F)$), then $\sqrt{n}(T_n - \boldsymbol{\theta}) \xrightarrow{D} \boldsymbol{u}$ and $\sqrt{n}(T_i^* - T_n) \xrightarrow{D} \boldsymbol{u}$ with $\boldsymbol{u} = \dot{T}(F)\boldsymbol{z}_F$. Olive (2017b) used these results to show that if $\boldsymbol{u} \sim N_g(\boldsymbol{0}, \boldsymbol{\Sigma}_A)$, then $\sqrt{n}(\overline{T}^* - T_n) \xrightarrow{D} \boldsymbol{0}$, $\sqrt{n}(T_i^* - \overline{T}^*) \xrightarrow{D} \boldsymbol{u}$, $\sqrt{n}(\overline{T}^* - \boldsymbol{\theta}) \xrightarrow{D} \boldsymbol{u}$, and that the prediction region method large sample $100(1-\delta)\%$ confidence region for $\boldsymbol{\theta}$ is

$$\{\boldsymbol{w} : (\boldsymbol{w} - \overline{T}^*)^T [\boldsymbol{S}^*_T]^{-1} (\boldsymbol{w} - \overline{T}^*) \leq D^2_{(U_B)}\} = \{\boldsymbol{w} : D^2_{\boldsymbol{w}}(\overline{T}^*, \boldsymbol{S}^*_T) \leq D^2_{(U_B)}\} \qquad (2.16)$$

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 $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ rejects H_0 if $(\overline{T}^* - \boldsymbol{\theta}_0)^T [\boldsymbol{S}^*_T]^{-1} (\overline{T}^* - \boldsymbol{\theta}_0) > D^2_{(U_B)}$.

The modified Bickel and Ren (2001) large sample $100(1-\delta)\%$ confidence region is

$$\{\boldsymbol{w} : (\boldsymbol{w} - T)^T [\boldsymbol{S}^*_T]^{-1} (\boldsymbol{w} - T_n) \leq D^2_{(U_B,T)}\} = \{\boldsymbol{w} : D^2_{\boldsymbol{w}}(T_n, \boldsymbol{S}^*_T) \leq D^2_{(U_B,T)}\} \qquad (2.17)$$

where $D^2_{(U_B,T)}$ is computed from $D_i^2 = (T_i^* - T_n)^T [\boldsymbol{S}^*_T]^{-1} (T_i^* - T_n)$. See Olive (2017b, p. 170).

Since (2.17) is a large sample confidence region by Bickel and Ren (2011), so is (2.16) if $\sqrt{n}(\overline{T}^* - T_n) \xrightarrow{P} \mathbf{0}$. Olive (2017b, pp. 171-172) proved (2.16) is a large sample confidence region. Pelawa Watagoda and Olive (2019) have a simpler proof.

The remainder of this section follows Pelawa Watagoda and Olive (2019) closely. For OLS variable selection with C_p, let $\hat{\boldsymbol{\beta}}_{I_j} = (\boldsymbol{X}_{I_j}^T \boldsymbol{X}_{I_j})^{-1} \boldsymbol{X}_{I_j}^T \boldsymbol{Y} = \boldsymbol{D}_j \boldsymbol{Y}$, $T_n = \hat{\boldsymbol{\beta}}_{I_{min},0}$ and $T_{jn} = \hat{\boldsymbol{\beta}}_{I_j,0} = \boldsymbol{D}_{j,0} \boldsymbol{Y}$ where $\boldsymbol{D}_{j,0}$ adds rows of zeroes to $\boldsymbol{D}_j$ corresponding to the x_i not in I_j. Let $T_n = T_{kn} = \hat{\boldsymbol{\beta}}_{I_k,0}$ with probabilities π_{kn} where $\pi_{kn} \to \pi_k$ as $n \to \infty$. Denote the π_k with $S \subseteq I_k$ by π_j. The other $\pi_k = 0$ by Nishii (1984). Then $\sqrt{n}(\hat{\boldsymbol{\beta}}_{I_j} - \boldsymbol{\beta}_{I_j}) \xrightarrow{D} N_{a_j}(\mathbf{0}, \sigma^2 \boldsymbol{V}_j)$ and $\boldsymbol{u}_{jn} = \sqrt{n}(\hat{\boldsymbol{\beta}}_{I_j,0} - \boldsymbol{\beta}) \xrightarrow{D} \boldsymbol{u}_j \sim N_p(\mathbf{0}, \sigma^2 \boldsymbol{V}_{j,0})$ where $n(\boldsymbol{X}_{I_j}^T \boldsymbol{X}_{I_j})^{-1} \xrightarrow{P} \boldsymbol{V}_j$ and $\boldsymbol{V}_{j,0}$ adds columns and rows of zeroes corresponding to the x_i not in I_j. Hence $\boldsymbol{\Sigma}_j = \sigma^2 \boldsymbol{V}_{j,0}$ is singular unless I_j corresponds to the full model.

Then Pelawa Watagoda and Olive (2019) showed

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{I_{min},0} - \boldsymbol{\beta}) \xrightarrow{D} \boldsymbol{u} \tag{2.18}$$

where the cdf of $\boldsymbol{u}$ is $F_{\boldsymbol{u}}(\boldsymbol{z}) = \sum_j \pi_j F_{\boldsymbol{u}_j}(\boldsymbol{z})$. Thus $\boldsymbol{u}$ is a mixture distribution of the $\boldsymbol{u}_j$ with probabilities π_j, $E(\boldsymbol{u}) = \mathbf{0}$, and $\text{Cov}(\boldsymbol{u}) = \boldsymbol{\Sigma}_{\boldsymbol{u}} = \sum_j \pi_j \sigma^2 \boldsymbol{V}_{j,0}$. The values of π_j depend on the OLS variable selection method with C_p, such as backward elimination, forward selection, and all subsets. Let $\boldsymbol{A}$ be a $g \times p$ full rank matrix with $1 \leq g \leq p$.

Then

$$\sqrt{n}(\boldsymbol{A}\hat{\boldsymbol{\beta}}_{I_{min},0} - \boldsymbol{A}\boldsymbol{\beta}) \xrightarrow{D} \boldsymbol{A}\boldsymbol{u} = \boldsymbol{v} \tag{2.19}$$

where $\boldsymbol{Au}$ has a mixture distribution of the $\boldsymbol{Au}_j \sim N_g(\boldsymbol{0}, \sigma^2 \boldsymbol{A}\boldsymbol{V}_{j,0}\boldsymbol{A}^T)$ with probabilities π_j.

Two special cases are interesting. First, suppose $\pi_d = 1$ so $\boldsymbol{u} \sim \boldsymbol{u}_d \sim N_p(\boldsymbol{0}, \boldsymbol{\Sigma}_d)$. This special case occurs for C_p if $a_S = p$ so S is the full model, and for methods like BIC that choose I_S with probability going to one.

The second special case occurs if for each $\pi_j > 0$, $\boldsymbol{Au}_j \sim N_g(\boldsymbol{0}, \boldsymbol{A\Sigma}_j\boldsymbol{A}^T) = N_g(\boldsymbol{0}, \boldsymbol{A\Sigma A}^T)$. Then $\sqrt{n}(\boldsymbol{A}\hat{\boldsymbol{\beta}}_{I_{min},0} - \boldsymbol{A}\boldsymbol{\beta}) \xrightarrow{D} \boldsymbol{Au} \sim N_g(\boldsymbol{0}, \boldsymbol{A\Sigma A}^T)$. This special case occurs for $\hat{\boldsymbol{\beta}}_S$ if the nontrivial predictors are orthogonal or uncorrelated with zero mean 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 multivariate normal limiting distribution for I_{min} and for the OLS full model.

For $g = 1$, the percentile method uses an interval that contains $U_B \approx k_B = \lceil B(1-\delta) \rceil$ of the T_i^* from a bootstrap sample $T_1^*, ..., T_B^*$ where the statistic T_n is an estimator of θ based on a sample of size n. Note that the squared Mahalanobis distance $D_\theta^2 = (\theta - \overline{T^*})^2 / S_T^{2*} \leq D_{(U_B)}^2$ is equivalent to $\theta \in [\overline{T^*} - S_T^* D_{(U_B)}, \overline{T^*} + S_T^* D_{(U_B)}]$, which is an interval centered at $\overline{T^*}$ just long enough to cover U_B of the T_i^*. If D is the $100q_B$th sample quantile of $|T_i^* - \overline{T}^*|$, then the prediction region method large sample CI for

θ is

$$\overline{T}^* \pm D.$$

Similarly, the Bickel and Ren CI is an interval centered at T_n just long enough to cover $U_{B,T} \approx k_B$ of the T_i^*. Hence the prediction region method CI and Bickel and Ren CI are both special cases of the percentile method if $g = 1$. Efron (2014) used a similar large sample $100(1-\delta)\%$ confidence interval centered at $\overline{T}^*$ assuming that $\overline{T}^*$ is asymptotically normal. The Frey (2013) shorth(c) interval (2.8) (with c given by (2.10)) applied to the T_i^* gives a confidence interval that is a practical implementation of the Hall (1988) shortest bootstrap interval based on all possible bootstrap samples.

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.

Note that if $\sqrt{n}(T_n - \theta) \stackrel{D}{\to} U$ and $\sqrt{n}(T_i^* - T_n) \stackrel{D}{\to} U$ where U has a unimodal probability density function symmetric about zero, then the confidence intervals from the two confidence regions, the shorth confidence interval, and the "usual" percentile method confidence interval are asymptotically equivalent (use the central proportion of the bootstrap sample, asymptotically).

A geometric argument is useful. Assume $T_1, ..., T_B$ are iid with nonsingular covariance matrix $\boldsymbol{\Sigma}_{T_n}$. Then the large sample $100(1-\delta)\%$ prediction region $R_p = \{\boldsymbol{w} : D^2_{\boldsymbol{w}}(\overline{T}, \boldsymbol{S}_T) \leq D^2_{(U_B)}\}$ centered at $\overline{T}$ contains a future value of the statistic T_f with probability $1 - \delta_B \to 1 - \delta$ as $B \to \infty$. Hence the region $R_c = \{\boldsymbol{w} : D^2_{\boldsymbol{w}}(T_n, \boldsymbol{S}_T) \leq D^2_{(U_B)}\}$ centered at a randomly selected T_n contains $\overline{T}$ with probability $1-\delta_B$. If $\sqrt{n}(T_n - \boldsymbol{\theta}) \xrightarrow{D} \boldsymbol{u}$ with $E(\boldsymbol{u}) = \boldsymbol{0}$ and $\text{Cov}(\boldsymbol{u}) = \boldsymbol{\Sigma}_{\boldsymbol{u}}$, then for fixed B with $\boldsymbol{v}_i \sim \boldsymbol{u}$,

$$\sqrt{n}(\overline{T} - \boldsymbol{\theta}) \xrightarrow{D} \frac{1}{B}\sum_{i=1}^{B} \boldsymbol{v}_i \sim AN_g\left(\boldsymbol{0}, \frac{\boldsymbol{\Sigma}_{\boldsymbol{u}}}{B}\right).$$

Hence $(\overline{T} - \boldsymbol{\theta}) = O_P((nB)^{-1/2})$, and $\overline{T}$ gets arbitrarily close to $\boldsymbol{\theta}$ compared to T_n as $B \to \infty$. Hence R_c is a large sample $100(1-\delta)\%$ confidence region for $\boldsymbol{\theta}$ as $n, B \to \infty$. We also need $(n\boldsymbol{S}_T)^{-1}$ to be fairly well behaved (not too ill conditioned) for each $n \geq 20g$, say. This condition is weaker than $(n\boldsymbol{S}_T)^{-1} \xrightarrow{P} \boldsymbol{\Sigma}^{-1}_{\boldsymbol{u}}$.

If $\sqrt{n}(T_n - \boldsymbol{\theta})$ and $\sqrt{n}(T^*_i - T_n)$ both converge in distribution to $\boldsymbol{u} \sim N_g(\boldsymbol{0}, \boldsymbol{\Sigma}_A)$, say, then the bootstrap sample data cloud of $T^*_1, ..., T^*_B$ is like the data cloud of iid $T_1, ..., T_B$ shifted to be centered at T_n. Then region (2.17) is a confidence region by the geometric argument since $D_{(U_B,T)}$ tends to be larger than $D_{(U_B)}$, and (2.16) is a confidence region if $\sqrt{n}(\overline{T}^* - T_n) \xrightarrow{P} \boldsymbol{0}$.

Much of the bootstrap confidence region theory does not apply to the variable selection estimator $T_n = \boldsymbol{A}\hat{\boldsymbol{\beta}}_{I_{min},0}$ with $\boldsymbol{\theta} = \boldsymbol{A}\boldsymbol{\beta}$, because T_n is not smooth since T_n is equal to the estimator T_{jn} with probability π_{jn} for $j = 1, ..., J$. Here $\boldsymbol{A}$ is a known full rank $g \times p$ matrix with $1 \leq g \leq p$. We have $\sqrt{n}(T_n - \boldsymbol{\theta}) \xrightarrow{D} \boldsymbol{v}$ by (2.19) where $E(\boldsymbol{v}) = \boldsymbol{0}$, and $\boldsymbol{\Sigma_v} = \sum_j \sigma^2 \boldsymbol{A}\boldsymbol{V}_{j,0}\boldsymbol{A}^T$. Hence the geometric argument holds: applying the prediction region (2.14) to an iid sample $T_1, ..., T_B$ and then centering the region at T_n gives a large sample confidence region for $\boldsymbol{\theta}$. For variable selection, we will next show that the bootstrap sample data cloud $T_1^*, ..., T_B^*$ tends to be slightly more variable than the data cloud of iid $T_1, ..., T_B$ for large n.

Assume p is fixed, $n \geq 20p$, 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 the $r = 0$ line, respectively. If the error distribution is skewed or multimodal, then much larger sample sizes may be needed.

For the bootstrap, suppose that T_i^* is equal to T_{ij}^* with probability ρ_{jn} for $j = 1, ..., J$ where $\sum_j \rho_{jn} = 1$, and $\rho_{jn} \to \pi_j$ as $n \to \infty$. Let B_{jn} count the number of times $T_i^* = T_{ij}^*$ in the bootstrap sample. Then the bootstrap sample $T_1^*, ..., T_B^*$ can be written as

$$T_{1,1}^*, ..., T_{B_{1n},1}^*, ..., T_{1,J}^*, ..., T_{B_{Jn},J}^*$$

where the B_{jn} follow a multinomial distribution and $B_{jn}/B \xrightarrow{P} \rho_{jn}$ as $B \to \infty$.

Denote $T^*_{1j}, ..., T^*_{B_{jn},j}$ as the jth bootstrap component of the bootstrap sample with sample mean $\overline{T}^*_j$ and sample covariance matrix $\boldsymbol{S}^*_{T,j}$. Then

$$\overline{T}^* = \frac{1}{B}\sum_{i=1}^{B} T^*_i = \sum_j \frac{B_{jn}}{B}\frac{1}{B_{jn}}\sum_{i=1}^{B_{jn}} T^*_{ij} = \sum_j \hat{\rho}_{jn}\overline{T}^*_j.$$

Similarly, we can define the jth component of the iid sample $T_1, ..., T_B$ to have sample mean $\overline{T}_j$ and sample covariance matrix $\boldsymbol{S}_{T,j}$.

For the residual bootstrap, we use the fitted values and residuals from the OLS full model to obtain $\boldsymbol{Y}^*$, but fit $\hat{\boldsymbol{\beta}}$ for a method such as forward selection, lasso, et cetera. Consider forward selection where each component uses a $\hat{\boldsymbol{\beta}}_{I_j}$. Let $\hat{\boldsymbol{Y}} = \hat{\boldsymbol{Y}}_{OLS} = \boldsymbol{X}\hat{\boldsymbol{\beta}}_{OLS} = \boldsymbol{H}\boldsymbol{Y}$ be the fitted values from the OLS full model where $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T$. Let $\boldsymbol{r}^W$ denote an $n \times 1$ random vector of elements selected with replacement from the OLS full model residuals. Following Freedman (1981) and Efron (1982, p. 36), $\boldsymbol{Y}^* = \boldsymbol{X}\hat{\boldsymbol{\beta}}_{OLS} + \boldsymbol{r}^W$ follows a standard linear model where the elements r^W_i of $\boldsymbol{r}^W$ are iid from the empirical distribution of the OLS full model residuals r_i. Hence

$$E(r^W_i) = \frac{1}{n}\sum_{i=1}^{n} r_i = 0, \quad V(r^W_i) = \sigma^2_n = \frac{1}{n}\sum_{i=1}^{n} r^2_i = \frac{n-p}{n}MSE,$$

$$E(\boldsymbol{r}^W) = \boldsymbol{0}, \text{ and } \operatorname{Cov}(\boldsymbol{Y}^*) = \operatorname{Cov}(\boldsymbol{r}^{\mathrm{W}}) = \sigma^2_{\mathrm{n}}\boldsymbol{I}_{\mathrm{n}}.$$

Then $\hat{\boldsymbol{\beta}}^*_{I_j} = (\boldsymbol{X}^T_{I_j}\boldsymbol{X}_{I_j})^{-1}\boldsymbol{X}^T_{I_j}\boldsymbol{Y}^* = \boldsymbol{D}_j\boldsymbol{Y}^*$ with $\operatorname{Cov}(\hat{\boldsymbol{\beta}}^*_{I_j}) = \sigma^2_n(\boldsymbol{X}^T_{I_j}\boldsymbol{X}_{I_j})^{-1}$ and $E(\hat{\boldsymbol{\beta}}^*_{I_j}) =$

$(\boldsymbol{X}_{I_j}^T \boldsymbol{X}_{I_j})^{-1} \boldsymbol{X}_{I_j}^T E(\boldsymbol{Y}^*) = (\boldsymbol{X}_{I_j}^T \boldsymbol{X}_{I_j})^{-1} \boldsymbol{X}_{I_j}^T \boldsymbol{H} \boldsymbol{Y} = \hat{\boldsymbol{\beta}}_{I_j}$ since $\boldsymbol{H}\boldsymbol{X}_{I_j} = \boldsymbol{X}_{I_j}$. The expectations are with respect to the bootstrap distribution where $\hat{\boldsymbol{Y}}$ acts as a constant.

For the above residual bootstrap with C_p, let $T_n = \boldsymbol{A}\hat{\boldsymbol{\beta}}_{I_{min},0}$ and $T_{jn} = \boldsymbol{A}\hat{\boldsymbol{\beta}}_{I_j,0} = \boldsymbol{A}\boldsymbol{D}_{j,0}\boldsymbol{Y}$ where $\boldsymbol{D}_{j,0}$ adds rows of zeroes to $\boldsymbol{D}_j$ corresponding to the x_i not in I_j. If $S \subseteq I_j$, then $\sqrt{n}(\hat{\boldsymbol{\beta}}_{I_j} - \boldsymbol{\beta}_{I_j}) \xrightarrow{D} N_{a_j}(\boldsymbol{0}, \sigma^2 \boldsymbol{V}_j)$ and $\sqrt{n}(\hat{\boldsymbol{\beta}}_{I_j,0} - \boldsymbol{\beta}) \xrightarrow{D} \boldsymbol{u}_j \sim N_p(\boldsymbol{0}, \sigma^2 \boldsymbol{V}_{j,0})$ where $\boldsymbol{V}_{j,0}$ adds columns and rows of zeroes corresponding to the x_i not in I_j. Using Theorem 1, $E(T^*) = \sum_j \rho_{jn} T_{jn} = \sum_j \rho_{jn} \boldsymbol{A}\hat{\boldsymbol{\beta}}_{I_j,0}$ and $\boldsymbol{S}_T^*$ is a consistent estimator of

$$\text{Cov}(T^*) = \sum_j \rho_{jn} \text{Cov}(T_{jn}^*) + \sum_j \rho_{jn} \boldsymbol{A}\hat{\boldsymbol{\beta}}_{I_j,0} \hat{\boldsymbol{\beta}}_{I_j,0}^T \boldsymbol{A}^T - E(T^*)[E(T^*)]^T$$

where asymptotically the sum is over $j : S \subseteq I_j$. If $\boldsymbol{\theta}_0 = \boldsymbol{0}$, then $n\boldsymbol{S}_T^* = \boldsymbol{\Sigma}_A + O_P(1)$ where

$$n\text{Cov}(T_n) \xrightarrow{P} \boldsymbol{\Sigma}_A = \sum_j \sigma^2 \pi_j \boldsymbol{A}\boldsymbol{V}_{j,0} \boldsymbol{A}^T.$$

Then $(n\boldsymbol{S}_T^*)^{-1}$ tends to be "well behaved" if $\boldsymbol{\Sigma}_A$ is nonsingular.

For the residual bootstrap with forward selection $n\text{Cov}(T_{jn})$ and $n\text{Cov}(T_{jn}^*)$ both converge in probability to $\sigma^2 \boldsymbol{A}\boldsymbol{V}_{j,0}\boldsymbol{A}^T$, and are close for $n \geq 20p$ since $\text{Cov}(T_{jn}^*) \approx (n-p)\text{Cov}(T_{jn})/n$. Hence the jth component of an iid sample $T_1, ..., T_B$ and the jth component of the bootstrap sample $T_1^*, ..., T_B^*$ have the same variability asymptotically. Since $E(T_{jn}) = \boldsymbol{\theta}$, each component of the iid sample is centered at $\boldsymbol{\theta}$. Since $E(T_{jn}^*) = T_{jn} = \boldsymbol{A}\hat{\boldsymbol{\beta}}_{I_j,0}$, the bootstrap components are centered at T_{jn}. Geometrically, separating the component clouds so that they are no longer centered at one value makes the overall data cloud larger. Thus the variability of T_n^* is larger than that of T_n for variable selection, asymptotically. Hence the prediction region applied to the bootstrap sample is slightly larger than the prediction region applied to the iid sample, asymptotically (we want $n \geq 20p$). Hence cutoff $\hat{D}^2_{1,1-\delta} = D^2_{(U_B)}$ gives coverage close to or higher than the nominal coverage for confidence regions (2.16) and (2.17), using the geometric argument. The deviation $T_i^* - T_n$ tends to be larger in magnitude than the deviations $\overline{T}^* - \boldsymbol{\theta}$, $T_n - \boldsymbol{\theta}$, and $T_i^* - \overline{T}^*$. Hence the cutoff $\hat{D}^2_{2,1-\delta} = D^2_{(U_B,T)}$ tends to be larger than $D^2_{(U_B)}$. The bootstrap sample data cloud is centered at $\overline{T}^* \approx \sum_j \rho_{jn} T_{jn}$. The T_{jn} are computed from the same data set and hence correlated. In simulations for $n \geq 20p$ and (2.16) and (2.17), the coverage tends to get close to or higher than $1-\delta$ for $B \geq \max(400, 50p)$ so that $\boldsymbol{S}_T^*$ is a good estimator of $\text{Cov}(T^*)$.

Undercoverage can occur if bootstrap sample data cloud is less variable than the iid data cloud, e.g., if $(n-p)/n$ is not close to one. Coverage can be higher than the nominal coverage for two reasons: i) the bootstrap data cloud is more variable than the iid data cloud of $T_1, ..., T_B$, and ii) zero padding.

To see the effect of zero padding, consider $H_0 : \boldsymbol{A\beta} = \boldsymbol{\beta}_O = \boldsymbol{0}$ where $\boldsymbol{\beta}_O = (\beta_{i_1}, ..., \beta_{i_g})^T$ and $O \subseteq E$ in (1.3) so that H_0 is true. Suppose a nominal 95% confidence region is used and $U_{(B)} = 0.96$. Hence the confidence region (2.16) or (2.17) covers at least 96% of the bootstrap sample. If $\hat{\boldsymbol{\beta}}^*_{O,j} = \boldsymbol{0}$ for more than 4% of the $\hat{\boldsymbol{\beta}}^*_{O,1}, ..., \hat{\boldsymbol{\beta}}^*_{O,B}$, then $\boldsymbol{0}$ is in the confidence region and the bootstrap test fails to reject H_0. If this occurs for each run in the simulation, then the observed coverage will be 100%.

If $\hat{\boldsymbol{\beta}}^*_j = 0$ for $j = 1, ..., B$, then the CI using the shorth, (2.16), or (2.17) is $[0, 0]$, and the pvalue for $H_0 : \beta_j = 0$ is one. (This result holds since $[0, 0]$ contains 100% of the $\hat{\beta}^*_j$ in the bootstrap sample.) For large sample theory tests, the pvalue estimates the population pvalue.

Note that there are several important variable selection models, including the model given by Equation (1.3). Another model is $\boldsymbol{x}^T\boldsymbol{\beta} = \boldsymbol{x}^T_{S_i}\boldsymbol{\beta}_{S_i}$ for $i = 1, ..., J$. Then there are $J \geq 2$ competing "true" nonnested submodels where $\boldsymbol{\beta}_{S_i}$ is $a_{S_i} \times 1$. For example, suppose the $J = 2$ models have predictors x_1, x_2, x_3 for S_1 and x_1, x_2, x_4 for S_2. Then x_3 and x_4 are likely to be selected and omitted often by forward selection for the B bootstrap samples. Hence omitting all predictors x_i that have a $\beta^*_{ij} = 0$ for at least one of the bootstrap samples $j = 1, ..., B$ could result in underfitting, e.g. using just x_1 and x_2 in the above $J = 2$ example. Regions (2.16) and (2.17) should still be useful.

Suppose the predictors x_i have been standardized. Then another important regression model has the β_i taper off rapidly, but no coefficients are equal to zero. For example,

$\beta_i = e^{-i}$ for $i = 1, ..., p$.

CHAPTER 3

EXAMPLE AND SIMULATIONS

Figure 1 shows 10%, 30%, 50%, 70%, 90% and 98% prediction regions for a future value of T_f for two multivariate normal distributions. The plotted points are iid $T_1, ..., T_B$ with $B = 100$.

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 of 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 and X_3 are highly correlated but not equal. Similarly, x_4 and x_5 are highly correlated but not equal. We expect that $Y = x_3 + e$ is a good model. Forward selection with C_p selected the model with a constant and *mmen*.

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 $a_{ij} = \psi$

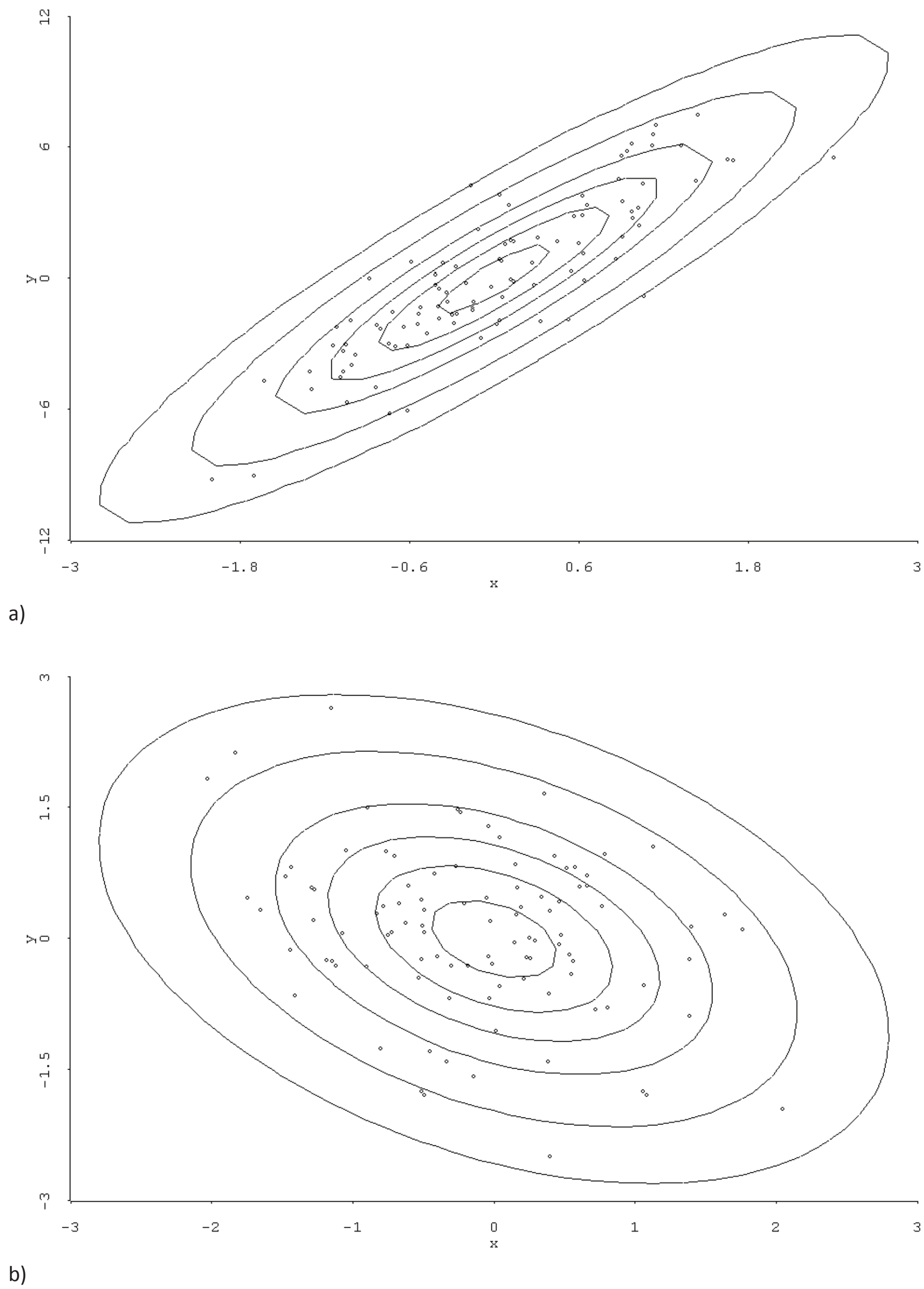

Figure 3.1. Prediction Regions

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 $(1, ..., 1)^T$. Let $Y_i = 1 + 1x_{i,2} + \cdots + 1x_{i,k+1} + e_i$ 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 from five distributions: i) N(0,1), ii) t_3, iii) EXP(1) - 1, iv) uniform$(-1, 1)$, and v) 0.9 N(0,1) + 0.1 N(0,100). Only distribution iii) is not symmetric.

A small simulation was done using $B = \max(1000, n, 20p)$ and 5000 runs. So an observed coverage in [0.94, 0.96] gives no reason to doubt that the CI has the nominal coverage of 0.95. The simulation used $p = 7$; $n = 10p, 25p$, $n = Jp$; $\psi = 0, 1/\sqrt{p}$, and 0.9; and $k = 1$ and 2. We tried to choose J so that the shorth CIs gave coverages ≥ 0.93. Simulations in Imhoff (2018) suggested that the shorth CI may need larger sample size n than the (2.16) and (2.17) CIs to have coverage ≥ 0.93. We expect the (2.16) CI average length to be less than that of the (2.17) CI, especially when the predictors are highly correlated.

When $\psi = 0$, the full model least squares confidence intervals for β_i should have length near $2t_{n-p,0.975}\sigma/\sqrt{n} \approx 2(1.96)\sigma/\sqrt{n}$ when the iid zero mean errors have variance σ^2. The simulation computed the Frey shorth(c) CI, prediction region method CI, and Bickel and Ren CI for each β_i 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.

The regression models used the residual bootstrap on the forward selection estimator $\hat{\boldsymbol{\beta}}_{I_{min},0}$. Table 1 gives results for when the iid errors $e_i \sim N(0,1)$. Two rows for each CI giving the observed confidence interval coverages and average lengths of the confidence intervals.

```
install.packages("leaps") #one time per computer

source("http://lagrange.math.siu.edu/Olive/slpack.txt")

library(leaps);Y <- marry[,3]; X <- marry[,-3]

temp<-regsubsets(X,Y,method="forward"); out<-summary(temp)

out$cp  [1] -0.8268967  1.0151462  3.0029429  5.0000000

Selection Algorithm: forward

         pop mmen mmilmen milwmn

1  ( 1 ) " " "*"  " "     " "
```

```
2  ( 1 ) " " "*"  "*"     " "

3  ( 1 ) "*" "*"  "*"     " "

4  ( 1 ) "*" "*"  "*"     "*"

record coverages and average lengths for b1, b2, ... bp-1, bp for

shorth CIs, prediction region method CIs and Bickel and Ren CIs

vscisim(n=70,p=7,k=1,psi=0.0,type=1,nruns=5000) #2 hours

$scicov

[1] 0.9364 0.9370 0.9954 0.9956 0.9974 0.9952 0.9978

$savelen

[1] 0.4709592 0.4760565 0.3901531 0.3904700 0.3873577 0.3888251 0.3872584

$prcicov

[1] 0.9302 0.9346 0.9934 0.9932 0.9954 0.9938 0.9966

$pravelen

[1] 0.4610427 0.4660094 0.4760685 0.4760695 0.4767312 0.4760690 0.4751954

$brcicov

[1] 0.9344 0.9358 0.9930 0.9930 0.9952 0.9928 0.9954

$bravelen

[1] 0.4656170 0.4705133 0.5418484 0.5421790 0.5445051 0.5458890 0.5450134
```

```
$beta    [1] 1 1 0 0 0 0 0

$k       [1] 1
```

Table 3.1. Bootstrap CIs with C_p, $p = 7$, N(0,1) errors

n, k, ψ	β_1	β_2	β_3	β_4	β_5	β_6	β_7
70,1,0	0.9364	0.9370	0.9954	0.9956	0.9974	0.9952	0.9978
shlen	0.4710	0.4761	0.3902	0.3905	0.3874	0.3888	0.3873
70,1,0	0.9302	0.9346	0.9934	0.9932	0.9954	0.9938	0.9966
prlen	0.4610	0.4660	0.4761	0.4761	0.4767	0.4761	0.4752
70,1,0	0.9344	0.9358	0.9930	0.9930	0.9952	0.9928	0.9954
brlen	0.4656	0.4705	0.5418	0.5421	0.5445	0.5459	0.5450

Suppose $\psi = 0$. Then from chapter 2, $\hat{\boldsymbol{\beta}}_S$ has the same limiting distribution for I_{min} and the full model. Note that the average lengths and coverages for forward selection I_{min} CIs for β_1 and β_2 were close to the expected full model lengths $3.92/\sqrt{n} = 0.469$. There was slight undercoverage since $\psi = 0$ and $(n-p)/n = 0.9$ for $n = 10p$. For $k = 1$, the lengths were shorter for $\beta_3, ..., \beta_7$ and the coverages were higher than 0.95 for the inactive predictors since zeros often occurred for inactive $\hat{\beta}^*_j$.

CHAPTER 4

ERROR TYPE 1 EXAMPLES

Table 4.1. Bootstrap CIs with C_p, $p = 7$, N(0,1) errors

n, k, ψ	β_1	β_2	β_3	β_4	β_5	β_6	β_7
70,2,0	0.9376	0.9362	0.9352	0.9976	0.9966	0.9964	0.9956
shlen	0.4715	0.4769	0.4773	0.3891	0.3893	0.3902	0.3883
70,2,0	0.9304	0.9310	0.9320	0.9962	0.9954	0.9952	0.9942
prlen	0.4620	0.4668	0.4671	0.4768	0.4782	0.4775	0.4771
70,2,0	0.9330	0.9332	0.9334	0.9962	0.9954	0.9956	0.9942
brlen	0.4652	0.4703	0.4710	0.5440	0.5453	0.5431	0.5452
175,1,0	0.9448	0.9458	0.9970	0.9972	0.9988	0.9976	0.9980
shlen	0.3003	0.3020	0.2454	0.2450	0.2440	0.24411	0.2440
175,1,0	0.9408	0.9420	0.9960	0.9964	0.9974	0.9966	0.9968
prlen	0.2940	0.2951	0.3013	0.3015	0.3010	0.3020	0.3011
175,1,0	0.9428	0.9436	0.9960	0.9964	0.9974	0.9966	0.9968
brlen	0.2950	0.2963	0.3393	0.3420	0.3420	0.3414	0.3410
175,2,0	0.9458	0.9462	0.9446	0.9980	0.9974	0.9972	0.9986
shlen	0.3010	0.3020	0.3020	0.2444	0.2442	0.2450	0.2450
175,2,0	0.9450	0.9442	0.9400	0.9972	0.9958	0.9966	0.9970
prlen	0.2943	0.2953	0.2954	0.3013	0.3020	0.3020	0.3020
175,2,0	0.9452	0.9442	0.9424	0.9972	0.9956	0.9966	0.9970
brlen	0.2951	0.2963	0.2963	0.3410	0.3422	0.3420	0.3413

Table 4.2. Bootstrap CIs with C_p, $p = 7$, N(0,1) errors

n, k, ψ	β_1	β_2	β_3	β_4	β_5	β_6	β_7
70,1,0.378	0.9340	0.9342	0.9970	0.9962	0.9954	0.9964	0.9988
shlen	0.4713	0.4775	0.3891	0.3910	0.3910	0.3891	0.3895
70,1,0.378	0.9314	0.9314	0.9958	0.9956	0.9950	0.9954	0.9978
prlen	0.4612	0.4673	0.4761	0.4764	0.4761	0.4766	0.4757
70,1,0.378	0.9346	0.9344	0.9960	0.9958	0.9950	0.9956	0.9978
brlen	0.4656	0.4720	0.5420	0.5430	0.5420	0.5462	0.5435
70,2,0.378	0.9370	0.9428	0.9450	0.9964	0.9954	0.9962	0.9972
shlen	0.4713	0.7044	0.7050	0.5751	0.5759	0.5780	0.5757
70,2,0.378	0.9324	0.9424	0.9492	0.9954	0.9946	0.9956	0.9958
prlen	0.4614	0.6920	0.6920	0.7011	0.6985	0.6997	0.7020
70,2,0.378	0.9348	0.9524	0.9562	0.9956	0.9950	0.9958	0.9964
brlen	0.4650	0.7194	0.7194	0.8051	0.7942	0.7991	0.8032
175,1,0.378	0.9436	0.9608	0.9982	0.9986	0.9976	0.9986	0.9978
shlen	0.3010	0.4472	0.3641	0.3631	0.3630	0.3634	0.3650
175,1,0.378	0.9408	0.9572	0.9976	0.9984	0.9974	0.9982	0.9976
prlen	0.2942	0.4377	0.4424	0.4423	0.4420	0.4422	0.4430
175,1,0.378	0.9426	0.9670	0.9980	0.9984	0.9976	0.9984	0.9976
brlen	0.2953	0.4634	0.5003	0.5020	0.5010	0.4995	0.5010
175,2,0,378	0.9430	0.9570	0.9540	0.9958	0.9976	0.9974	0.9968
shlen	0.3004	0.4459	0.4451	0.3612	0.3620	0.3620	0.3620
175,2,0.378	0.9408	0.9530	0.9504	0.9956	0.9974	0.9968	0.9962
prlen	0.2941	0.4365	0.4356	0.4424	0.4430	0.4423	0.4420
175,2,0.378	0.9414	0.9628	0.9548	0.9958	0.9974	0.9970	0.9962
brlen	0.2951	0.4510	0.4502	0.5013	0.5014	0.5004	0.4990

Table 4.3. Bootstrap CIs with C_p, $p = 7$, N(0,1) errors

n, k, ψ	β_1	β_2	β_3	β_4	β_5	β_6	β_7
70,1,0.9	0.9368	0.9058	0.9980	0.9968	0.9974	0.9964	0.9960
shlen	0.4710	3.7221	3.6702	3.6656	3.6681	3.6768	3.6675
70,1,0.9	0.9330	0.9730	0.9968	0.9962	0.9966	0.9948	0.9956
prlen	0.4610	4.3752	4.3310	4.3321	4.3304	4.3297	4.3255
70,1,0.9	0.9342	0.9762	0.9972	0.9968	0.9972	0.9954	0.9958
brlen	0.4652	4.9420	4.9084	4.9120	4.8950	4.8901	4.9091
70,2,0.9	0.9384	0.8836	0.8848	0.9972	0.9962	0.9970	0.9972
shlen	0.4723	3.6969	3.6983	3.6158	3.6154	3.6033	3.6001
70,2,0.9	0.9354	0.9708	0.9718	0.9962	0.9954	0.9966	0.9958
prlen	0.4622	4.3903	4.3820	4.3451	4.3392	4.3289	4.3322
70,2,0.9	0.9386	0.9756	0.9752	0.9966	0.9958	0.9970	0.9966
brlen	0.4674	4.8973	4.8810	4.9087	4.9302	4.8977	4.8959
175,1,0.9	0.9464	0.9624	0.9980	0.9970	0.9972	0.9982	0.9986
shlen	0.3005	2.4174	2.2810	2.2771	2.2945	2.2814	2.2882
175,1,0.9	0.9434	0.9746	0.9968	0.9962	0.9964	0.9972	0.9976
prlen	0.2940	2.8401	2.7430	2.7420	2.7413	2.7468	2.7453
175,1,0.9	0.9434	0.9856	0.9970	0.9958	0.9966	0.9978	0.9978
brlen	0.2952	3.0882	3.0753	3.0759	3.0575	3.0710	3.0720
175,2,0.9	0.9480	0.9476	0.9480	0.9976	0.9986	0.9982	0.9980
shlen	0.3010	2.4131	2.4330	2.3050	2.3110	2.2997	2.2983
175,2,0.9	0.9414	0.9662	0.9680	0.9970	0.9978	0.9976	0.9976
prlen	0.2944	2.8410	2.8445	2.7901	2.7954	2.7894	2.7910
175,2,0.9	0.9432	0.9814	0.9842	0.9974	0.9978	0.9978	0.9976
brlen	0.2956	3.2140	3.2184	3.1469	3.1495	3.1460	3.1463

CHAPTER 5

ERROR TYPE 2 EXAMPLE

Table 5.1. Bootstrap CIs with C_p, $p = 7$, t_3 errors

n, k, ψ	β_1	β_2	β_3	β_4	β_5	β_6	β_7
70,1,0	0.9320	0.9390	0.9970	0.9978	0.9964	0.9978	0.9958
shlen	0.7761	0.8013	0.6476	0.6460	0.6465	0.6431	0.6473
70,2,0	0.9322	0.9430	0.9966	0.9966	0.9960	0.9966	0.9950
prlen	0.7602	0.8030	0.7881	0.7892	0.7865	0.7868	0.7889
70,1,0	0.9370	0.9490	0.9966	0.9966	0.9958	0.9966	0.9950
brlen	0.7679	0.8150	0.8998	0.8996	0.8972	0.8969	0.8984
70,2,0	0.9384	0.9240	0.9288	0.9958	0.9978	0.9970	0.9980
shlen	0.7784	0.8021	0.8020	0.6489	0.6496	0.6499	0.6510
70,2,0	0.9390	0.9332	0.9350	0.9948	0.9974	0.9956	0.9972
prlen	0.7630	0.8066	0.8051	0.7924	0.7910	0.7910	0.7911
70,2,0	0.9394	0.9414	0.9426	0.9948	0.9976	0.9956	0.9970
brlen	0.7693	0.8198	0.8179	0.9025	0.9041	0.9034	0.9031
175,1,0	0.9434	0.9468	0.9982	0.9986	0.9990	0.9976	0.9966
shlen	0.5025	0.5102	0.4101	0.4095	0.4130	0.4097	0.4114
175,1,0	0.9446	0.9452	0.9974	0.9980	0.9984	0.9974	0.9954
prlen	0.4921	0.5004	0.5055	0.5050	0.5050	0.5053	0.5044
175,1,0	0.9440	0.9478	0.9972	0.9980	0.9984	0.9972	0.9954
brlen	0.4940	0.5030	0.5741	0.5730	0.5720	0.5731	0.5711

Table 5.2. Bootstrap CIs with C_p, $p = 7$, t_3 errors

n, k, ψ	β_1	β_2	β_3	β_4	β_5	β_6	β_7
175,2,0	0.9522	0.9460	0.9470	0.9974	0.9970	0.9970	0.9980
shlen	0.5030	0.5111	0.5114	0.4124	0.4125	0.4110	0.4110
175,2,0	0.9506	0.9442	0.9448	0.9966	0.9966	0.9964	0.9970
prlen	0.4924	0.5010	0.5015	0.5054	0.5052	0.5054	0.5061
175,2,0	0.9530	0.9456	0.9446	0.9966	0.9964	0.9964	0.9968
brlen	0.4940	0.5030	0.5034	0.5720	0.5710	0.5745	0.5723
70,1,0.378	0.9380	0.9458	0.9974	0.9970	0.9976	0.9976	0.9974
shlen	0.7798	1.1912	0.9715	0.9677	0.9772	0.9681	0.9689
70,1,0.378	0.9364	0.9446	0.9968	0.9966	0.9968	0.9968	0.9962
prlen	0.7640	1.2474	1.1656	1.1640	1.1678	1.1630	1.1640
70,1,0.378	0.9388	0.9508	0.9972	0.9966	0.9970	0.9974	0.9962
brlen	0.7721	1.3392	1.3340	1.3257	1.3340	1.3287	1.3278
70,2,0.378	0.9386	0.9286	0.9350	0.9978	0.9964	0.9968	0.9964
shlen	0.7778	1.1856	1.1859	0.9665	0.9710	0.9620	0.9677
70,2,0.378	0.9382	0.9310	0.9382	0.9966	0.9954	0.9956	0.9952
prlen	0.7620	1.2450	1.2450	1.1656	1.1686	1.1655	1.1661
70,2,0.378	0.9402	0.9414	0.9520	0.9964	0.9960	0.9964	0.9954
brlen	0.7686	1.3171	1.3157	1.3285	1.3356	1.3330	1.3244
175,1,0.378	0.9446	0.9546	0.9974	0.9994	0.9976	0.9988	0.9988
shlen	0.5013	0.7565	0.6104	0.6081	0.6073	0.6110	0.6086
175,1,0.378	0.9418	0.9562	0.9968	0.9986	0.9972	0.9984	0.9982
prlen	0.4910	0.7498	0.7396	0.7382	0.7375	0.7410	0.7391
175,1,0.378	0.9448	0.9652	0.9972	0.9986	0.9972	0.9984	0.9980
brlen	0.4930	0.7910	0.8398	0.8391	0.8386	0.8397	0.8402

Table 5.3. Bootstrap CIs with C_p, $p = 7$, t_3 errors

n, k, ψ	β_1	β_2	β_3	β_4	β_5	β_6	β_7
175,2,0.378	0.9466	0.9570	0.9622	0.9978	0.9982	0.9986	0.9986
shlen	0.5035	0.7585	0.7586	0.6094	0.6098	0.6110	0.6071
175,2,0.378	0.9442	0.9596	0.9616	0.9972	0.9978	0.9978	0.9984
prlen	0.4931	0.7520	0.7513	0.7450	0.7440	0.7450	0.7440
175,2,0.378	0.9438	0.9650	0.9708	0.9974	0.9978	0.9978	0.9986
brlen	0.4950	0.7774	0.7776	0.8430	0.8423	0.8450	0.8451
70,1,0.9	0.9386	0.8480	0.9972	0.9962	0.9978	0.9968	0.9958
shlen	0.7742	6.1574	6.1586	6.1540	6.1684	6.1564	6.1477
70,1,0.9	0.9378	0.9712	0.9956	0.9948	0.9968	0.9946	0.9942
prlen	0.7583	7.1210	7.1187	7.1194	7.1273	7.1230	7.1186
70,1,0.9	0.9386	0.9642	0.9960	0.9966	0.9974	0.9954	0.9948
brlen	0.7653	8.0063	7.9792	7.9768	8.0030	7.9698	7.9912
70,2,0.9	0.9330	0.8404	0.8516	0.9978	0.9958	0.9976	0.9964
shlen	0.7820	6.1610	6.1565	6.0897	6.1005	6.1166	6.1275
70,2,0.9	0.9344	0.9734	0.9770	0.9966	0.9944	0.9962	0.9958
prlen	0.7658	7.2430	7.2697	7.1910	7.2182	7.2230	7.2121
70,2,0.9	0.9370	0.9716	0.9740	0.9968	0.9948	0.9966	0.9962
brlen	0.7732	8.1010	8.1550	8.1289	8.1602	8.1750	8.1210
175,1,0.9	0.9430	0.9022	0.9984	0.9984	0.9980	0.9990	0.9994
shlen	0.5040	3.9591	3.9201	3.9230	3.9321	3.9111	3.9220
175,1,0.9	0.9396	0.9800	0.9982	0.9984	0.9976	0.9984	0.9992
prlen	0.4933	4.6521	4.6120	4.6202	4.6212	4.6073	4.6210
175,1,0.9	0.9408	0.9804	0.9982	0.9988	0.9978	0.9988	0.9988
brlen	0.4951	5.2411	5.1892	5.2004	5.1885	5.1823	5.1981

Table 5.4. Bootstrap CIs with C_p, $p = 7$, t_3 errors

n, k, ψ	β_1	β_2	β_3	β_4	β_5	β_6	β_7
175,2,0.9	0.9472	0.8950	0.8754	0.9984	0.9986	0.9986	0.9978
shlen	0.5020	3.8630	3.8533	3.7896	3.8078	3.7962	3.7964
175,2,0.9	0.9458	0.9808	0.9734	0.9976	0.9984	0.9978	0.9970
prlen	0.4915	4.6166	4.6261	4.5750	4.5801	4.5720	4.5814
175,2,0.9	0.9472	0.9800	0.9772	0.9984	0.9984	0.9980	0.9972
brlen	0.4940	5.1242	5.1322	5.1550	5.1495	5.1386	5.1534
350,1,0.9	0.9518	0.9306	0.9358	0.9992	0.9980	0.9984	0.9976
shlen	0.3604	2.8322	2.8358	2.7142	2.7304	2.7150	2.7240
350,1,0.9	0.9490	0.9736	0.9800	0.9988	0.9974	0.9982	0.9964
prlen	0.3530	3.3550	3.3541	3.3285	3.3250	3.3255	3.3277
350,1,0.9	0.9490	0.9802	0.9828	0.9990	0.9980	0.9982	0.9970
brlen	0.3540	3.8374	3.8530	3.7940	3.7943	3.7687	3.7756
350,2,0.9	0.9530	0.9430	0.9984	0.9984	0.9984	0.9988	0.9970
shlen	0.3613	2.8474	2.7710	2.7721	2.7650	2.7675	2.7640
350,2,0.9	0.9518	0.9742	0.9974	0.9976	0.9980	0.9984	0.9964
prlen	0.3540	3.3795	3.3083	3.3040	3.2991	3.3077	3.3054
350,2,0.9	0.9518	0.9850	0.9978	0.9978	0.9980	0.9988	0.9968
brlen	0.3545	3.7410	3.7052	3.6975	3.7010	3.7021	3.7166

CHAPTER 6

ERROR TYPE 3 EXAMPLE

Table 6.1. Bootstrap CIs with C_p, $p = 7$, EXP(1)-1 errors

n, k, ψ	β_1	β_2	β_3	β_4	β_5	β_6	β_7
70,1,0	0.9266	0.9354	0.9958	0.9960	0.9960	0.9970	0.9970
shlen	0.4666	0.4778	0.3868	0.3880	0.38811	0.3895	0.3888
70,1,0	0.9218	0.9302	0.9940	0.9956	0.9956	0.9958	0.9968
prlen	0.4571	0.4673	0.47345	0.4741	0.4740	0.4730	0.4730
70,1,0	0.9254	0.9328	0.9944	0.9956	0.9956	0.9956	0.9966
brlen	0.4620	0.4720	0.5420	0.5412	0.5397	0.5398	0.5382
70,2,0	0.9290	0.9454	0.9368	0.9966	0.9964	0.9972	0.9954
shlen	0.4641	0.4740	0.4750	0.3851	0.3850	0.3851	0.3862
70,2,0	0.9230	0.9416	0.9304	0.9952	0.9958	0.9964	0.9942
prlen	0.4544	0.4632	0.4641	0.4712	0.4720	0.4710	0.4711
70,2,0	0.9268	0.9422	0.9338	0.9952	0.9958	0.9964	0.9942
brlen	0.4582	0.4667	0.4676	0.5420	0.5389	0.5359	0.5378
175,1,0	0.9396	0.9426	0.9980	0.9966	0.9984	0.9974	0.9980
shlen	0.2982	0.3013	0.2434	0.2440	0.2440	0.2434	0.2440
175,1,0	0.9330	0.9386	0.9976	0.9960	0.9968	0.9970	0.9976
prlen	0.2930	0.2950	0.2995	0.2993	0.3002	0.2999	0.3001
175,1,0	0.9350	0.9398	0.9976	0.9960	0.9970	0.9970	0.9976
brlen	0.2930	0.2959	0.3391	0.3375	0.3403	0.3391	0.3410

Table 6.2. Bootstrap CIs with C_p, $p = 7$, EXP(1)-1 errors

n, k, ψ	β_1	β_2	β_3	β_4	β_5	β_6	β_7
175,2,0	0.9430	0.9506	0.9456	0.9972	0.9978	0.9984	0.9980
shlen	0.2991	0.3020	0.3020	0.2452	0.2440	0.2440	0.2440
175,2,0	0.9374	0.9470	0.9418	0.9972	0.9970	0.9976	0.9964
prlen	0.2930	0.2954	0.2951	0.3010	0.3010	0.3012	0.3012
175,2,0	0.9384	0.94747	0.9432	0.9972	0.9970	0.9976	0.9962
brlen	0.2940	0.2964	0.2959	0.3395	0.3421	0.3410	0.3413
70,1,0.378	0.9290	0.9532	0.9964	0.9968	0.9980	0.9970	0.9960
shlen	0.4655	0.7089	0.5758	0.5727	0.5750	0.5768	0.5757
70,1,0.378	0.9246	0.9520	0.9954	0.9962	0.9970	0.9954	0.9948
prlen	0.4558	0.6979	0.6925	0.6930	0.6940	0.6933	0.6943
70,1,0.378	0.9276	0.9640	0.9956	0.9964	0.9972	0.9956	0.9958
brlen	0.4604	0.7420	0.7895	0.7950	0.7920	0.7914	0.7933
70,2,0.378	0.9362	0.9514	0.9452	0.9966	0.9970	0.9962	0.9968
shlen	0.4659	0.7057	0.7060	0.5699	0.5731	0.5731	0.5741
70,2,0.378	0.9304	0.9528	0.9484	0.9958	0.9954	0.9954	0.9964
prlen	0.4562	0.6950	0.6951	0.6930	0.6931	0.6930	0.6930
70,2,0.378	0.9328	0.9606	0.9602	0.9962	0.9954	0.9954	0.9964
brlen	0.4601	0.7214	0.7220	0.7924	0.7910	0.7868	0.7879
175,1,0.378	0.9396	0.9604	0.9984	0.9982	0.9994	0.9986	0.9996
shlen	0.2976	0.4444	0.3603	0.3620	0.3602	0.3599	0.3610
175,1,0.378	0.9356	0.9550	0.9972	0.9972	0.9986	0.9982	0.9992
prlen	0.2914	0.4350	0.4394	0.4386	0.4384	0.4392	0.4402
175,1,0.378	0.9370	0.9644	0.9976	0.9972	0.9988	0.9984	0.9992
brlen	0.2930	0.4602	0.5001	0.4973	0.4979	0.4973	0.4993

Table 6.3. Bootstrap CIs with C_p, $p = 7$, EXP(1)-1 errors

n, k, ψ	β_1	β_2	β_3	β_4	β_5	β_6	β_7
175,2,0.378	0.9404	0.9554	0.9580	0.9980	0.9990	0.9970	0.9976
shlen	0.2985	0.4450	0.4440	0.3603	0.3614	0.3612	0.3602
175,2,0.378	0.9356	0.9524	0.9520	0.9978	0.9990	0.9956	0.9966
prlen	0.2923	0.4351	0.4343	0.4410	0.4411	0.4410	0.4410
175,2,0.378	0.9366	0.9608	0.9610	0.9978	0.9988	0.9956	0.9970
brlen	0.2932	0.4496	0.4487	0.5010	0.4996	0.5004	0.5010
70,1,0.9	0.9242	0.9056	0.9958	0.9974	0.9970	0.9968	0.9968
shlen	0.4640	3.6920	3.6457	3.6547	3.6468	3.6557	3.6460
70,1,0.9	0.9206	0.9726	0.9944	0.9962	0.9958	0.9952	0.9962
prlen	0.4544	4.3297	4.2974	4.2976	4.2830	4.2830	4.2858
70,1,0.9	0.9236	0.9766	0.9948	0.9966	0.9966	0.9956	0.9964
brlen	0.4585	4.9020	4.8685	4.8586	4.8367	4.8450	4.8631
70,2,0.9	0.9290	0.8866	0.8836	0.9974	0.9984	0.9978	0.9958
shlen	0.4677	3.6743	3.6673	3.5940	3.5892	3.6178	3.5840
70,2,0.9	0.9240	0.9694	0.9718	0.9966	0.9978	0.9968	0.9952
prlen	0.4581	4.3610	4.3625	4.3068	4.3140	4.3210	4.2984
70,2,0.9	0.9304	0.9792	0.9788	0.9970	0.9980	0.9976	0.9960
brlen	0.4634	4.8888	4.9030	4.9160	4.8786	4.8923	4.9030
175,1,0.9	0.9464	0.9608	0.9972	0.9976	0.9982	0.9982	0.9976
shlen	0.2981	2.4084	2.2731	2.2770	2.2859	2.2801	2.2830
175,1,0.9	0.9426	0.9728	0.9968	0.9970	0.9964	0.9976	0.9970
prlen	0.2920	2.8224	2.7275	2.7277	2.7250	2.7242	2.7275
175,1,0.9	0.9446	0.9850	0.9970	0.9970	0.9970	0.9976	0.9972
brlen	0.2931	3.0845	3.0594	3.0478	3.0357	3.0420	3.0571

Table 6.4. Bootstrap CIs with C_p, $p = 7$, EXP(1)-1 errors

n, k, ψ	β_1	β_2	β_3	β_4	β_5	β_6	β_7
175,2,0.9	0.9396	0.9472	0.9482	0.9982	0.9988	0.9972	0.9982
shlen	0.2996	2.4191	2.4388	2.2910	2.2910	2.2922	2.2940
175,2,0.9	0.9370	0.9722	0.9666	0.9972	0.9976	0.9960	0.9974
prlen	0.2934	2.8510	2.8551	2.7903	2.7810	2.7797	2.7820
175,2,0.9	0.9364	0.9836	0.9810	0.9972	0.9982	0.9960	0.9972
brlen	0.2950	3.2230	3.2166	3.1634	3.1485	3.1480	3.1495

CHAPTER 7

ERROR TYPE 4 EXAMPLE

Table 7.1. Bootstrap CIs with C_p, $p = 7$, uniform$(-1, 1)$errors

n, k, ψ	β_1	β_2	β_3	β_4	β_5	β_6	β_7
70,1,0	0.9392	0.9384	0.9974	0.9974	0.9966	0.9980	0.9964
shlen	0.2725	0.2751	0.2241	0.2250	0.2241	0.2250	0.2252
70,1,0	0.9316	0.9340	0.9960	0.9964	0.9960	0.9974	0.9960
prlen	0.2667	0.2694	0.2750	0.2756	0.2750	0.2753	0.2757
70,1,0	0.9368	0.9362	0.9962	0.9966	0.9960	0.9970	0.9960
brlen	0.2694	0.2720	0.3150	0.3155	0.3130	0.3140	0.3150
70,2,0	0.9336	0.9378	0.9312	0.9968	0.9950	0.9944	0.9966
shlen	0.2722	0.2753	0.2750	0.2254	0.2244	0.2251	0.2255
70,2,0	0.9308	0.9338	0.9302	0.9958	0.9932	0.9944	0.9954
prlen	0.2665	0.2695	0.2689	0.2750	0.2745	0.2753	0.2755
70,2,0	0.9336	0.9372	0.9326	0.9958	0.9934	0.9944	0.9954
brlen	0.2685	0.2720	0.2711	0.3130	0.3130	0.3140	0.3150
175,1,0	0.9466	0.9528	0.9980	0.9982	0.9992	0.9974	0.9970
shlen	0.1734	0.1742	0.1414	0.1420	0.1412	0.1410	0.1410
175,1,0	0.9398	0.9502	0.9980	0.9974	0.9986	0.9970	0.9966
prlen	0.1697	0.1710	0.1740	0.1742	0.1740	0.1740	0.1741
175,1,0	0.9432	0.9506	0.9980	0.9974	0.9986	0.9970	0.9966
brlen	0.1704	0.1711	0.1964	0.1971	0.1971	0.1961	0.1968

Table 7.2. Bootstrap CIs with C_p, $p = 7$, uniform$(-1, 1)$errors

n, k, ψ	β_1	β_2	β_3	β_4	β_5	β_6	β_7
175,2,0	0.9430	0.9440	0.9494	0.9972	0.9974	0.9974	0.9976
shlen	0.1735	0.1742	0.1741	0.1410	0.1412	0.1411	0.1410
175,2,0	0.9390	0.9394	0.9458	0.9960	0.9962	0.9960	0.9972
prlen	0.1697	0.1710	0.1705	0.1740	0.1740	0.1741	0.1740
175,2,0	0.9388	0.9388	0.9470	0.9960	0.9962	0.9960	0.9972
brlen	0.1703	0.1711	0.1710	0.1964	0.1968	0.1965	0.1971
70,1,0.378	0.9362	0.9514	0.9962	0.9962	0.9956	0.9976	0.9964
shlen	0.2730	0.4080	0.3340	0.3340	0.3340	0.3341	0.3333
70,1,0.378	0.9334	0.9464	0.9958	0.9952	0.9952	0.9970	0.9952
prlen	0.2668	0.3996	0.4041	0.4050	0.4040	0.4040	0.4031
70,1,0.378	0.9356	0.9534	0.9960	0.9960	0.9956	0.9974	0.9958
brlen	0.2694	0.4250	0.4611	0.4620	0.4620	0.4630	0.4594
70,2,0.378	0.9392	0.9396	0.9498	0.9946	0.9968	0.9968	0.9952
shlen	0.2730	0.4074	0.4075	0.3333	0.3322	0.3336	0.3330
70,2,0.378	0.9306	0.9344	0.9454	0.9940	0.9958	0.9962	0.9946
prlen	0.2672	0.3988	0.3989	0.4053	0.4054	0.4052	0.4054
70,2,0.378	0.9342	0.9448	0.9530	0.9940	0.9958	0.9960	0.9950
brlen	0.2693	0.4154	0.4150	0.4631	0.4631	0.4621	0.4641
175,1,0.378	0.9458	0.9586	0.9984	0.9978	0.9978	0.9978	0.9972
shlen	0.1735	0.2585	0.2105	0.2097	0.2098	0.2102	0.2101
175,1,0.378	0.9402	0.9540	0.9980	0.9970	0.9974	0.9976	0.9964
prlen	0.1698	0.2530	0.2557	0.2557	0.2558	0.2561	0.2556
175,1,0.378	0.9428	0.9624	0.9978	0.9972	0.9974	0.9976	0.9968
brlen	0.1705	0.2675	0.2890	0.2897	0.2910	0.2900	0.2891

Table 7.3. Bootstrap CIs with C_p, $p = 7$, uniform$(-1, 1)$errors

n, k, ψ	β_1	β_2	β_3	β_4	β_5	β_6	β_7
175,2,0.378	0.9464	0.9506	0.9498	0.9980	0.9972	0.9980	0.9976
shlen	0.1740	0.2575	0.2570	0.2083	0.2089	0.2088	0.2096
175,2,0.378	0.9420	0.9472	0.9492	0.9974	0.9968	0.9972	0.9966
prlen	0.1699	0.2521	0.2520	0.2555	0.2558	0.2559	0.2556
175,2,0.378	0.9440	0.9546	0.9560	0.9974	0.9968	0.9974	0.9966
brlen	0.1710	0.2605	0.2599	0.2894	0.2903	0.2898	0.2885
70,1,0.9	0.9434	0.9548	0.9972	0.9966	0.9958	0.9966	0.9970
shlen	0.2727	2.2300	2.1010	2.1095	2.0961	2.0963	2.0950
70,1,0.9	0.9374	0.9592	0.9966	0.9958	0.9950	0.9958	0.9964
prlen	0.2670	2.5964	2.5105	2.5154	2.5140	2.5089	2.5030
70,1,0.9	0.9392	0.9804	0.9968	0.9962	0.9954	0.9964	0.9964
brlen	0.2695	2.8222	2.8266	2.8359	2.8272	2.8172	2.8123
70,2,0.9	0.9404	0.9462	0.9406	0.9960	0.9956	0.9968	0.9946
shlen	0.2740	2.2630	2.2750	2.1187	2.1221	2.1274	2.1297
70,2,0.9	0.9358	0.9600	0.9512	0.9944	0.9944	0.9960	0.9946
prlen	0.2678	2.6323	2.6368	2.5386	2.5510	2.5459	2.5476
70,2,0.9	0.9374	0.9792	0.9792	0.9950	0.9948	0.9966	0.9946
brlen	0.2710	2.9530	2.9585	2.8801	2.8920	2.8730	2.8772
175,1,0.9	0.9502	0.9764	0.9976	0.9984	0.9978	0.9966	0.9980
shlen	0.1740	1.5605	1.3130	1.3164	1.3140	1.3150	1.3140
175,1,0.9	0.9486	0.9516	0.9970	0.9976	0.9974	0.9958	0.9978
prlen	0.1698	1.7131	1.6052	1.6013	1.6010	1.6013	1.6023
175,1,0.9	0.9488	0.9502	0.9972	0.9982	0.9978	0.9964	0.9978
brlen	0.1710	1.8850	1.8366	1.8312	1.8240	1.8310	1.8250

Table 7.4. Bootstrap CIs with C_p, $p = 7$, uniform$(-1, 1)$errors

n, k, ψ	β_1	β_2	β_3	β_4	β_5	β_6	β_7
175,2,0.9	0.9462	0.9558	0.9624	0.9978	0.9974	0.9972	0.9990
shlen	0.1740	1.5794	1.5735	1.3313	1.3297	1.3198	1.3250
175,2,0.9	0.9434	0.9392	0.9438	0.9974	0.9974	0.9964	0.9980
prlen	0.1702	1.7291	1.7233	1.6089	1.6075	1.6110	1.6065
175,2,0.9	0.9440	0.9276	0.9346	0.9976	0.9974	0.9964	0.9980
brlen	0.1710	1.8530	1.8430	1.8204	1.8150	1.8267	1.8163

CHAPTER 8

ERROR TYPE 5 EXAMPLE

Table 8.1. Bootstrap CIs with C_p, $p = 7$, 0.9 N(0,1) + 0.1 N(0,100) errors

n, k, ψ	β_1	β_2	β_3	β_4	β_5	β_6	β_7
70,1,0	0.9324	0.9320	0.9986	0.9986	0.9978	0.9982	0.9972
shlen	1.4891	1.4923	1.2534	1.2534	1.2578	1.2510	1.2559
70,1,0	0.9364	0.9204	0.9978	0.9980	0.9970	0.9974	0.9970
prlen	1.4585	1.6445	1.5140	1.5098	1.5173	1.5120	1.5153
70,1,0	0.9432	0.9226	0.9978	0.9982	0.9972	0.9974	0.9968
brlen	1.4740	1.7262	1.7203	1.7210	1.7240	1.7225	1.7250
70,2,0	0.9342	0.9352	0.9334	0.9972	0.9972	0.9982	0.9980
shlen	1.4950	1.4962	1.4956	1.2594	1.2530	1.2570	1.2555
70,2,0	0.9408	0.9248	0.9246	0.9964	0.9964	0.9964	0.9970
prlen	1.4640	1.6510	1.6530	1.5240	1.5184	1.5220	1.5173
70,2,0	0.9432	0.9282	0.9264	0.9966	0.9966	0.9964	0.9972
brlen	1.4784	1.7520	1.7577	1.7453	1.7359	1.7358	1.7268
175,1,0	0.9460	0.9352	0.9988	0.9984	0.9980	0.9978	0.9984
shlen	0.9820	1.0212	0.8044	0.8033	0.8054	0.8074	0.8010
175,1,0	0.9486	0.9384	0.9978	0.9982	0.9970	0.9974	0.9976
prlen	0.9603	1.0476	0.9822	0.9830	0.9864	0.9856	0.9830
175,1,0	0.9522	0.9482	0.9978	0.9980	0.9970	0.9974	0.9976
brlen	0.9641	1.0621	1.1112	1.1115	1.1184	1.1157	1.1158

Table 8.2. Bootstrap CIs with C_p, $p = 7$, 0.9 N(0,1) + 0.1 N(0,100) errors

n, k, ψ	β_1	β_2	β_3	β_4	β_5	β_6	β_7
175,2,0	0.9440	0.9420	0.9352	0.9972	0.9988	0.9974	0.9984
shlen	0.9750	1.0120	1.0130	0.8033	0.7956	0.7957	0.7977
175,2,0	0.9474	0.9474	0.9394	0.9962	0.9980	0.9968	0.9980
prlen	0.9540	1.0369	1.0395	0.9820	0.9774	0.9764	0.9785
175,2,0	0.9486	0.9548	0.9454	0.9962	0.9980	0.9968	0.9980
brlen	0.9572	1.0514	1.0542	1.1114	1.1073	1.1088	1.1112
70,1,0.378	0.9344	0.9450	0.9976	0.9972	0.9986	0.9968	0.9986
shlen	1.4891	2.0492	1.8787	1.8704	1.8732	1.8844	1.8797
70,1,0.378	0.9418	0.9538	0.9964	0.9960	0.9976	0.9952	0.9968
prlen	1.4581	2.3291	2.2322	2.2268	2.2241	2.2240	2.2250
70,1,.378	0.9416	0.9638	0.9970	0.9964	0.9976	0.9956	0.9976
brlen	1.4731	2.5595	2.5420	2.5283	2.5330	2.5250	2.5342 —
70,2,0.378	0.9258	0.9408	0.9410	0.9980	0.9982	0.9976	0.9970
shlen	1.4930	2.0699	2.0643	1.8861	1.8776	1.8814	1.8810
70,2,0.378	0.9342	0.9502	0.9516	0.9974	0.9970	0.9970	0.9958
prlen	1.4620	2.3662	2.3651	2.2463	2.2441	2.2434	2.2464
70,2,0.378	0.9334	0.9586	0.9650	0.9980	0.9972	0.9970	0.9960
brlen	1.4779	2.6231	2.6261	2.5520	2.5574	2.5496	2.5581
175,1,0.378	0.9418	0.9558	0.9986	0.9978	0.9988	0.9982	0.9986
shlen	0.9820	1.4530	1.2094	1.1991	1.2021	1.2020	1.1986
175,1,0.378	0.9420	0.9428	0.9976	0.9972	0.9984	0.9978	0.9980
prlen	0.9610	1.5804	1.4572	1.4522	1.4552	1.4550	1.4512
175,1,0.378	0.9394	0.9430	0.9974	0.9976	0.9984	0.9978	0.9984
brlen	0.9650	1.6978	1.6482	1.6530	1.6521	1.6540	1.6501

Table 8.3. Bootstrap CIs with C_p, $p = 7$, 0.9 N(0,1) + 0.1 N(0,100) errors

n, k, ψ	β_1	β_2	β_3	β_4	β_5	β_6	β_7
175,2,0.378	0.9526	0.9478	0.9534	0.9978	0.9976	0.9980	0.9986
shlen	0.9777	1.4622	1.4597	1.2062	1.1999	1.2015	1.2120
175,2,0.378	0.9546	0.9360	0.9400	0.9972	0.9972	0.9970	0.9972
prlen	0.9568	1.5940	1.5895	1.4563	1.4559	1.4568	1.4588
175,2,0.378	0.9530	0.9356	0.9384	0.9972	0.9972	0.9970	0.9972
brlen	0.9602	1.6840	1.6843	1.6483	1.6493	1.6533	1.6486
70,1,0.9	0.9408	0.8078	0.9968	0.9966	0.9970	0.9988	0.9972
shlen	1.4820	11.8250	11.8430	11.8389	11.8250	11.7840	11.8710
70,1,0.9	0.9458	0.9704	0.9948	0.9952	0.9958	0.9970	0.9954
prlen	1.4514	13.5793	13.5682	13.5721	13.5865	13.5287	13.6082
70,1,0.9	0.9470	0.9560	0.9960	0.9962	0.9964	0.9974	0.9956
brlen	1.4650	15.1172	15.0769	15.0679	15.1296	15.0888	15.1130
70,2,0.9	0.9266	0.7992	0.7936	0.9986	0.9978	0.9974	0.9974
shlen	1.4985	11.9465	11.8720	11.9282	11.9010	11.9520	11.9540
70,2,0.9	0.9350	0.9712	0.9714	0.9978	0.9964	0.9962	0.9954
prlen	1.4677	13.8140	13.7714	13.8102	13.7810	13.7920	13.8230
70,2,0.9	0.9408	0.9616	0.9640	0.9980	0.9972	0.9968	0.9958
brlen	1.4820	15.5060	15.5143	15.5434	15.4675	15.5084	15.5221
175,1,0.9	0.9456	0.8334	0.9990	0.9984	0.9984	0.9982	0.9974
shlen	0.9755	7.6320	7.6420	7.6596	7.6157	7.6387	7.6385
175,1,0.9	0.9430	0.9804	0.9986	0.9974	0.9976	0.9980	0.9966
prlen	0.9544	8.8677	8.8675	8.8640	8.8410	8.8410	8.8562
175,1,0.9	0.9468	0.9668	0.9990	0.9976	0.9978	0.9980	0.9972
brlen	0.9579	9.8520	9.8640	9.8456	9.8581	9.8242	9.8240

Table 8.4. Bootstrap CIs with C_p, $p = 7$, 0.9 N(0,1) + 0.1 N(0,100) errors

n, k, ψ	β_1	β_2	β_3	β_4	β_5	β_6	β_7
175,2,0.9	0.9428	0.8340	0.8202	0.9984	0.9990	0.9980	0.9978
shlen	0.9775	7.5730	7.5750	7.5810	7.5605	7.5761	7.5650
175,2,0.9	0.9430	0.9798	0.9780	0.9972	0.9984	0.9976	0.9974
prlen	0.9564	8.9441	8.9520	8.9141	8.9073	8.9110	8.9112
175,2,0.9	0.9432	0.9704	0.9722	0.9978	0.9986	0.9978	0.9974
brlen	0.9597	10.0620	10.0730	10.0263	10.0167	10.0550	10.0459
1400,1,0.9	0.9550	0.9482	0.9988	0.9986	0.9988	0.9986	0.9978
shlen	0.3520	2.7584	2.6832	2.6803	2.6731	2.6776	2.6788
1400,1,0.9	0.9522	0.9790	0.9980	0.9980	0.9980	0.9982	0.9978
prlen	0.3442	3.2715	3.2113	3.2063	3.2076	3.2050	3.2010
1400,1,0.9	0.9532	0.9854	0.9984	0.9982	0.9986	0.9982	0.9978
brlen	0.3450	3.6059	3.5761	3.5776	3.5859	3.5783	3.5734
1400,2,0.9	0.9536	0.9366	0.9354	0.9986	0.9984	0.9982	0.9990
shlen	0.3520	2.7530	2.7443	2.6350	2.6284	2.6270	2.6313
1400,2,0.9	0.9504	0.9746	0.9778	0.9982	0.9980	0.9974	0.9988
prlen	0.3444	3.2713	3.2733	3.2524	3.2563	3.2578	3.2473
1400,2,0.9	0.9518	0.9810	0.9816	0.9980	0.9982	0.9978	0.9990
brlen	0.3450	3.7587	3.7655	3.7198	3.7190	3.7134	3.7014

CHAPTER 9

CONCLUSIONS

There is massive literature on variable selection and a fairly large literature for inference after variable selection. See references in Pelawa Watagoda and Olive (2019).

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.

For my simulations, the zero mean errors e_i were from five distributions as stated before. We chose to run the same schedule of parameters for all five error types. The simulation used $p = 7$; $n = 10p, 25p$, $n = Jp$; $\psi = 0, 1/\sqrt{p}$, and 0.9; and $k = 1$ and 2. We tried to choose J so that the shorth CIs gave coverages ≥ 0.93.

As we have seen, for the most part, we did not need J since the shorth CIs gave coverages ≥ 0.93. The only case that we needed J was when $\psi = 0.9$ in types 2 and 5. J was = 50 that means n=350 for type 2, and it was =200 which implied that n=1400 for type 5.

The 3 CIs used different correction factors. Hence, the shorth CI was not always the shortest. The shorth CIs for slopes tended to be shortest when $\boldsymbol{\beta}_i$= 0. The other 2 CIs were often longest when $\boldsymbol{\beta}_i$= 0, and the increase was larger for the Bickel and Ren

CIs.

The simulations were done in *R*. See R Core Team (2016). We used several *R* functions including forward selection as computed with the `regsubsets` function from the `leaps` library. The collection of Olive (2019) *R* functions *slpack*, available from (http://lagrange.math.siu.edu/Olive/slpack.txt), has some useful functions for the inference. Tables were made with `vscisim`.

[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] 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 Gunst, M., Klaassen, C., and van der Vaart, A., The Institute of Mathematical Statistics, Hayward, CA, 91-112.

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

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

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

[6] Freedman, D.A. (1981), "Bootstrapping Regression Models," *The Annals of Statistics*, 9, 1218-1228.

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

[8] Hall, P. (1988), "Theoretical Comparisons of Bootstrap Confidence Intervals," (with discussion), *The Annals of Statistics*, 16, 927-985.

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

[10] Imhoff, D.C. (2018), "Bootstrapping Forward Selection with C_P," Master's Research Paper, Southern Illinois University.

[11] Machado, J.A.F., and Parente, P. (2005), "Bootstrap Estimation of Covariance Matrices Via the Percentile Method," *Econometrics Journal,* 8, 70–78.

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

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

[14] Olive, D.J. (2017a), *Linear Regression*, Springer, New York, NY.

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

[16] Olive, D.J. (2018), "Applications of Hyperellipsoidal Prediction Regions," *Statistical Papers*, 59, 913-931.

[17] Olive, D.J. (2019), *Prediction and Statistical Learning*, online course notes, see

(http://lagrange.math.siu.edu/Olive/slearnbk.htm).

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

[19] Pelawa Watagoda, L.C.R., and Olive, D.J. (2019), "Bootstrapping Multiple Linear Regression After Variable Selection," preprint at (http://lagrange.math.siu.edu/Olive/ppbootCR.pdf).

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

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

[22] Serfling, R.J. (1980), *Approximation Theorems of Mathematical Statistics,* Wiley, New York, NY.

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

VITA

Graduate School

Southern Illinois University

Mashael Alshammari

Almashail00@hotmail.com

Northern Borders University, Saudi Arabia

Bachelor of Science, Mathematics, 2014

Research Paper Title:

Bootstrap Confidence Intervals For β_i Using Forward Selection With C_p

Major Professor: Dr. David J. Olive