Summer 2015

Bootstrapping Hypotheses Tests

Chathurangi H. Pathiravasan
Southern Illinois University Carbondale, chathurangi@siu.edu

BOOTSTRAPPING HYPOTHESES TESTS

by

Chathurangi Heshani Karunapala Pathiravasan

B.S., University of Sri Jayewardenepura, 2010

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

Department of Mathematics
in the Graduate School
Southern Illinois University Carbondale
August, 2015

RESEARCH PAPER APPROVAL

BOOTSTRAPPING HYPOTHESES TESTS

By Pathiravasan Chathurangi Heshani Karunapala

A Research Paper Submitted in Partial

Fulfillment of the Requirements

for the Degree of

Masters

in the field of Mathematics

Approved by:

Dr. David J. Olive, Chair

Dr. Bhaskar Bhattacharya

Dr. H. R. Hughes

Graduate School
Southern Illinois University Carbondale
May 05, 2015

AN ABSTRACT OF THE RESEARCH PAPER OF

PATHIRAVASAN CHATHURANGI HESHANI KARUNAPALA, for the Masters degree in Mathematics, presented on 5^{th} May 2015 , at Southern Illinois University.

TITLE: BOOTSTRAPPING HYPOTHESES TESTS

MAJOR PROFESSOR: Dr. D. J. Olive

The bootstrap is a general methodology to estimate the standard error of the test statistic. In fact, bootstrap methods can be applied to regression models and hypothesis testing. Consider testing $H_0 : \boldsymbol{A\theta} = \boldsymbol{c}$ versus $H_1 : \boldsymbol{A\theta} \neq \boldsymbol{c}$ where $\boldsymbol{A}$ is a known $r \times p$ matrix of rank r and $\boldsymbol{c}$ is a known $r \times 1$ vector. Let $\hat{\boldsymbol{\theta}}$ be a consistent estimator of $\boldsymbol{\theta}$ and make a bootstrap sample $\boldsymbol{w}_i = A\hat{\boldsymbol{\theta}}_i^* - \boldsymbol{c}$ for $i = 1, ..., B$. Make a prediction region for the $\boldsymbol{w}_i$ and determine whether $\mathbf{0}$ is in the prediction region.

The percentile method uses an interval that contains $d_B \approx k_B = \lceil B(1-\delta) \rceil$ of the $T_{i,n}^*$ from a bootstrap sample $T_{1,n}^*, ..., T_{B,n}^*$, where the statistic $T_{i,n}$ is an estimator of θ based on a sample of size n.

It will be shown that the Olive (2015b) prediction region method generalizes the percentile method for $r = 1$ to $r \geq 1$. This method can be widely applied, but should be regarded as exploratory unless theory shows that the prediction region method is a large sample test.

Moreover, this prediction region method will be compared to the Efron (2014) confidence interval for variable selection and used to bootstrap a correlation matrix. Indeed, the prediction region method can also be justified as a special case of the percentile method where the test statistic is the squared Mahalanobis distance $D_i^{2*} = (T_i^* - \overline{T^*})^T [\boldsymbol{S}_T^*]^{-1} (T_i^* - \overline{T^*}))$ where $\boldsymbol{w}_i = T_i^*$, and $\overline{T^*}$ and $\boldsymbol{S}_T^*$ are the sample mean and sample covariance matrix of $T_1^*, ..., T_B^*$.

ACKNOWLEDGMENTS

I would like to express my heartfelt gratitude to Dr. D. J. Olive for his invaluable assistance and insights leading to the writing of this paper. Most of the concepts and the examples used in this research paper have been quoted from his previous research works. Moreover, I would like to appreciate his great support and guidance through out the semester. My sincere thanks also goes to the members of my graduate committee, Dr. B. Bhattacharya and Dr. H. R. Hughes.

A special thanks also to my husband and my family who provide me great courage and support for the research work. Furthermore, I would like to thank my friends and people at the Department of Mathematics who contributed to my work.

TABLE OF CONTENTS

LIST OF TABLES

INTRODUCTION

The bootstrap was introduced in 1979 as a computer-based method for estimating the standard error of test statistic ($T = \hat{\theta}$). In statistics, bootstrapping can refer to any test or metric that relies on random sampling with replacement. Bootstrapping allows assigning measures of accuracy (defined in terms of bias, variance, confidence intervals, prediction error or some other such measure) to sample estimates. Bootstrap is more important when the theoretical distribution is unknown.

Bootstrap methods depend on the notion of a bootstrap sample. Let $\hat{F}$ be the empirical distribution, putting probability $1/n$ on each of the observed values y_i, $i = 1, 2, ..., n$. A bootstrap sample is defined to be a random sample of size n drawn from $\hat{F}$, say $Y^* = (y_1^*, y_2^*, ..., y_n^*)$. The star notation indicates that Y^* is not the actual data set Y, but rather a randomized psedo data for Y. In fact, the bootstrap data points $y_1^*, y_2^*, ..., y_n^*$ are a random sample of size n drawn with replacement from the population of a n objects $(y_1, y_2, ..., y_n)$.

Efron and Tibshirani (1993, p. 46) discussed the bootstrap algorithm which works by drawing many independent bootstrap samples, evaluating the corresponding bootstrap replications, and estimating the standard error of $\hat{\theta}$ by the empirical standard deviation of the replications. The result is called the bootstrap estimate of standard error, denoted by $\hat{se}_B$, where B is the number of bootstrap sample used. The limit of $\hat{se}_B$ as B goes to infinity is the ideal bootstrap estimate of $se_F(\hat{\theta})$,

$$\lim_{B\to\infty} \hat{se}_B = se_{\hat{F}} = se_{\hat{F}}(\hat{\theta^*})$$

The fact that $\hat{se}_B$ approaches $se_{\hat{F}}$ as B goes to infinity amounts to saying that an empirical standard deviation approaches the population standard deviation as the number of replications grows large. The "population" in this case is the population of values $\hat{\theta^*} = s(X^*)$, where $\hat{F} \to (x_1^*, x_2^*, ..., x_n^*) = X^*$.

Bootstrap methods are widely used for many areas in statistics. The application of bootstrap methods to regression models approximate the distribution of the coefficients and the distribution of the prediction errors. It may also be used for constructing hypothesis tests. In the following chapters we describe bootstrap methods that are directly designed for hypothesis testing. Indeed, the chapter 1 compares percentile method and prediction region method. Chapter 2 examines the method for multiple linear regression and method for variable selection, and chapter 3 gives an example and some simulations.

CHAPTER 1

PERCENTILE METHOD VS PREDICTION REGION METHOD

Consider testing $H_0 : \boldsymbol{A\theta} = \boldsymbol{c}$ versus $H_1 : \boldsymbol{A\theta} \neq \boldsymbol{c}$ where $\boldsymbol{A}$ is a known $r \times p$ matrix of rank r. If a confidence region can be constructed for $\boldsymbol{A\theta} - \boldsymbol{c}$, then fail to reject H_0 if $\boldsymbol{0}$ is in the confidence region, and reject H_0 if $\boldsymbol{0}$ is not in the confidence region. Given training data $\boldsymbol{w}_1, ..., \boldsymbol{w}_n$, a large sample $100(1-\delta)\%$ prediction region for a future test value $\boldsymbol{w}_f$ is a set $\mathcal{A}_n$ such that $P(\boldsymbol{w}_f \in \mathcal{A}_n) \rightarrow 1-\delta$ as $n \rightarrow \infty$, while a large sample confidence region for a parameter $\boldsymbol{\theta}$ is a set $\mathcal{A}_n$ such that $P(\boldsymbol{\theta} \in \mathcal{A}_n) \rightarrow 1-\delta$ as $n \rightarrow \infty$. The region $\mathcal{A}_n$ is typically constructed using the training data.

The percentile method, which is an interval that contains $d_B \approx k_B = \lceil B(1-\delta) \rceil$ of the $T^*_{i,n}$ from a bootstrap sample $T^*_{1,n}, ..., T^*_{B,n}$ where the statistic $T_{i,n}$ is an estimator of θ based on a sample of size n. Often the n is suppressed. Here $\lceil x \rceil$ is the smallest integer $\geq x$, e.g. $\lceil 7.8 \rceil = 8$. Let $T^*_{(1)}, T^*_{(2)}, ..., T^*_{(B)}$ be the order statistics of the bootstrap sample. Then one version of the percentile method discards the largest and smallest $\lceil B\delta/2 \rceil$ order statistics, resulting in an interval (L_B, U_B) that is a large sample $100(1-\delta)\%$ confidence interval for θ, and also a large sample $100(1-\delta)\%$ prediction interval for a future bootstrap value $T^*_{f,n}$.

Olive (2014, p. 283) recommends using the shorth(c) estimator for the percentile method. Let $c = k_B$, and let $W_i = T^*_{i,n}$. Let $W_{(1)}, ..., W_{(B)}$ be the order statistics of the W_i. Compute $W_{(c)} - W_{(1)}, W_{(c+1)} - W_{(2)}, ..., W_{(B)} - W_{(B-c+1)}$. Let $[W_{(s)}, W_{(s+c-1)}]$ correspond to the closed interval with the smallest distance. Then reject $H_0 : \theta = \theta_0$ if θ_0 is not in the interval. The shorth interval tends to be shorter than the interval that deletes the smallest and largest $\lceil B\delta/2 \rceil$ observations W_i when the W_i do not come from a symmetric distribution. Frey (2013) showed that for large $B\delta$ and iid data, the shorth(k_B) PI has undercoverage $\approx 1.12\sqrt{\delta/B}$, and used the shorth$(c)$ estimator as the large sample $100(1-\delta)\%$ prediction interval where $c = \lceil B[1 - \delta + 1.12\sqrt{\delta/B}\ ]\ \rceil$. Hence if $B = 1000$, there may be about 1% undercoverage using $c = k_B$.

Typically expect a large sample $100(1-\delta)\%$ prediction region for a future value of a statistic $T_{f,n}$ to have higher coverage for $\boldsymbol{\theta}$ than a large sample $100(1-\delta)\%$ confidence region for $\boldsymbol{\theta}$, although values of $T_{f,n}$ are reasonable values of $\boldsymbol{\theta}$. To see this claim, assume $\sqrt{n}(T-\boldsymbol{\theta}) \xrightarrow{D} N_p(\mathbf{0}, \boldsymbol{\Sigma})$. Then a large sample $100(1-\delta)\%$ confidence region for $\boldsymbol{\theta}$ is $\mathcal{A}_n = \{\boldsymbol{w} : (\boldsymbol{w}-T)^T \hat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{w}-T) \leq \chi^2_{p,1-\delta}\}$. If T_f is independent of T, then $\sqrt{n}(T_f - T) \xrightarrow{D} N_p(\mathbf{0}, 2\boldsymbol{\Sigma})$. Hence $P(T_f \in \mathcal{A}_n) = P[(T_f-T)^T \hat{\boldsymbol{\Sigma}}^{-1}(T_f-T) \leq \chi^2_{p,1-\delta}] \rightarrow P(X \leq \chi^2_{p,1-\delta}/2) < 1-\delta$ where $X \sim \chi^2_p$ and $P(X \leq \chi^2_{p,1-\delta}) = 1-\delta$. Hence this large sample prediction region for T_f needs a cutoff twice as large as the cutoff for the confidence region. Thus a large sample prediction region for T_f tends to be liberal (coverage is higher than the nominal coverage) as a confidence region for $\boldsymbol{\theta}$ when T_f is a consistent estimator of $\boldsymbol{\theta}$.

The percentile method is an exception since, heuristically, the bootstrap distribution tends to be centered about the statistic T rather than $\boldsymbol{\theta}$. "Bad samples" are less likely to cover θ, but across many independent samples the coverage probability tends to $1-\delta$. (Also the percentile method is a large sample $100(1-\delta)\%$ prediction region for a future value $T^*_{f,n}$ of the bootstrap statistic, not for a future value of the statistic $T_{f,n}$.)

Several additional approximations are needed for the bootstrap. Suppose the $\boldsymbol{z}_i$ are iid from a distribution with cdf F, and F_n is the empirical cdf that puts probability $1/n$ on each observed value of $\boldsymbol{z}_i$ for $i = 1, ..., n$. Let $T(F)$ denote the statistic computed from a sample of size n from F, and let $T(F_n)$ denote the statistic computed from a sample of size n from F_n. Want $T(F_n) - T(F) \xrightarrow{P} 0$ as $n \rightarrow \infty$ so that iid samples from the empirical distribution can be used in probability calculations. If $E(T_{in}) = \boldsymbol{\theta}_n \rightarrow \boldsymbol{\theta}$, need n large enough so that $\boldsymbol{\theta}_n \approx \boldsymbol{\theta}$.

Some notation is needed to give the prediction region used to bootstrap a hypothesis test. Suppose $\boldsymbol{w}_1, ..., \boldsymbol{w}_n$ are iid $p \times 1$ random vectors with mean $\boldsymbol{\mu}$ and nonsingular covariance matrix $\boldsymbol{\Sigma}_{\boldsymbol{W}}$. Let a future test observation $\boldsymbol{w}_f$ be independent of the $\boldsymbol{w}_i$ but from the

same distribution. Let $(\overline{\boldsymbol{w}}, \boldsymbol{S})$ be the sample mean and sample covariance matrix where

$$\overline{\boldsymbol{w}} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{w}_i \text{ and } \boldsymbol{S} = \boldsymbol{S}_{\boldsymbol{w}} = \frac{1}{\mathrm{n}-1}\sum_{\mathrm{i}=1}^{\mathrm{n}} (\boldsymbol{w}_{\mathrm{i}} - \overline{\boldsymbol{w}})(\boldsymbol{w}_{\mathrm{i}} - \overline{\boldsymbol{w}})^{\mathrm{T}}. \tag{1.1}$$

Then the ith *squared sample Mahalanobis distance* is the scalar

$$D^2_{\boldsymbol{w}} = D^2_{\boldsymbol{w}}(\overline{\boldsymbol{w}}, \boldsymbol{S}) = (\boldsymbol{w} - \overline{\boldsymbol{w}})^T \boldsymbol{S}^{-1} (\boldsymbol{w} - \overline{\boldsymbol{w}}). \tag{1.2}$$

Let $D_i^2 = D^2_{\boldsymbol{w}_i}$ for each observation $\boldsymbol{w}_i$. Let $D_{(c)}$ be the cth order statistic of $D_1, ..., D_n$. Following Olive (2013), a large sample $100(1-\delta)\%$ prediction region for $\boldsymbol{w}_f$ is the hyperellipsoid

$$\mathcal{A}_n = \{\boldsymbol{w} : D^2_{\boldsymbol{w}}(\overline{\boldsymbol{w}}, \boldsymbol{S}) \leq D^2_{(c)}\} = \{\boldsymbol{w} : D_{\boldsymbol{w}}(\overline{\boldsymbol{w}}, \boldsymbol{S}) \leq D_{(c)}\}. \tag{1.3}$$

If n is large, can use $c = k_n = \lceil n(1-\delta) \rceil$. If n is not large, using $c = d_n$ where d_n decreases to k_n, can improve small sample performance. Olive (2013) showed that this prediction region is a large sample $100(1-\delta)\%$ prediction region for a large class of distributions, although regions with smaller volumes may exist. Note that the result follows since if $\boldsymbol{\Sigma}_{\boldsymbol{w}}$ and $\boldsymbol{S}$ are nonsingular, then the Mahalanobis distance is a continuous function of $(\overline{\boldsymbol{w}}, \boldsymbol{S})$. Let $D = D(\boldsymbol{\mu}, \boldsymbol{\Sigma}_{\boldsymbol{w}})$. Then $D_i \xrightarrow{D} D$ and $D_i^2 \xrightarrow{D} D^2$. Hence the sample percentiles of the D_i are consistent estimators of the population percentiles of D at continuity points of cumulative distribution function (cdf) of D. The prediction region estimates the highest density region for a large class of elliptically contoured distributions. See Olive (2015a) for more on prediction regions.

Following Olive (2015b), the prediction region method makes a bootstrap sample $\boldsymbol{w}_i = A\hat{\boldsymbol{\theta}}_i^* - \boldsymbol{c}$ for $i = 1, ..., B$. Make the prediction region (1.3) for the $\boldsymbol{w}_i$ and determine whether $\mathbf{0}$ is in the prediction region. The prediction region method can also be justified as being a special case of the percentile method as follows.

Consider testing $H_0 : \boldsymbol{A\theta} = \boldsymbol{c}$ versus $H_1 : \boldsymbol{A\theta} \neq \boldsymbol{c}$, and the statistic $T_i = \boldsymbol{A}\hat{\boldsymbol{\theta}} - \boldsymbol{c}$. If $E(T_i) = \boldsymbol{\mu}$ and $\mathrm{Cov}(T_i) = \boldsymbol{\Sigma}_T$ were known, then the squared Mahalanobis distance $D_i^2(\boldsymbol{\mu}, \boldsymbol{\Sigma}_T) = (T_i - \boldsymbol{\mu})^T \boldsymbol{\Sigma}_T^{-1} (T_i - \boldsymbol{\mu})$ would be a natural statistic to use if the percentile

$D^2_{1-\delta}(\boldsymbol{\mu}, \boldsymbol{\Sigma}_T)$ was known. The prediction region method bootstraps the squared Mahalanobis distances, forming the bootstrap sample $\boldsymbol{w}_i = T^*_i = \boldsymbol{A}\hat{\boldsymbol{\theta}}^*_i - \boldsymbol{c}$ and the squared Mahalanobis distances $D^{2*}_i = D^2_i(\overline{T^*}, \boldsymbol{S}^*_T) = (T^*_i - \overline{T^*})^T[\boldsymbol{S}^*_T]^{-1}(T^*_i - \overline{T^*})$ where $\overline{T^*} = \frac{1}{B}\sum_{i=1}^{B} T^*_i$ and $\boldsymbol{S}^*_T = \frac{1}{B-1}\sum_{i=1}^{B}(T^*_i - \overline{T^*})(T^*_i - \overline{T^*})^T$ are the sample mean and sample covariance matrix of $T^*_1, ..., T^*_B$. Then the percentile method that contains the smallest $d_B \approx B(1-\delta)$ distances is used to get the closed interval $[0, D_{(d_B)}] = [0, D^*_{(d_B)}]$. If H_0 is true and $E[\hat{\boldsymbol{\theta}}] = \boldsymbol{\theta}$, then $\boldsymbol{\mu} = \boldsymbol{0}$. Let $D^2_{\boldsymbol{0}} = \overline{T^*}^T[\boldsymbol{S}^*_T]^{-1}\overline{T^*}$ and fail to reject H_0 if $D_{\boldsymbol{0}} \leq D_{(d_B)}$ and reject H_0 if $D_{\boldsymbol{0}} > D_{(d_B)}$. This percentile method is equivalent to computing the prediction region (1.3) on the $\boldsymbol{w}_i = T^*_i$ and checking whether $\boldsymbol{0}$ is in the prediction region.

Note that the percentile method makes an interval that contains $d_B \approx B(1-\delta)$ of the scalar valued T^*_i. The prediction region method makes a hyperellipsoid that contains d_B of the $r \times 1$ vectors $T^*_i = \boldsymbol{w}_i$, and equivalently, makes an interval $[0, D_{(d_B)}]$ that contains d_B of the $D^*_i = D_i$.

When $r = 1$, a hyperellipsoid is an interval, so the prediction region method is a special case of the percentile method. Suppose the parameter of interest is θ, and there is a bootstrap sample $T^*_1, ..., T^*_B$. Let $\overline{T}^*$ and S^{2*}_T be the sample mean and variance of the T^*_i. Then the squared Mahalanobis distance $D^2_\theta = (\theta - \overline{T}^*)^2/S^{2*}_T \leq D^2_{(d_B)}$ is equivalent to $\theta \in [\overline{T}^* - S^*_T D_{(d_B)}, \overline{T}^* + S^*_T D_{(d_B)}]$, which is an interval centered at $\overline{T}^*$ just long enough to cover $d_B \approx B(1-\delta)$ of the T^*_i. Hence this interval is a version of the percentile method.

The point of the above discussion is that prediction region method can be thought of as a variant of the percentile method applied to vector valued statistics, and is likely widely applicable. The method should be regarded as exploratory until theory proves that the method is a large sample test, but similar remarks apply to other bootstrap methods such as the percentile method.

CHAPTER 2

BOOTSTRAP METHODS

2.1 BOOTSTRAP TEST FOR MULTIPLE LINEAR REGRESSION

Consider the multiple linear regression model $Y_i = \boldsymbol{x}_i^T\boldsymbol{\beta} + e_i$ for $i = 1, ..., n$, written in matrix form as $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}$ where $\boldsymbol{Y}$ is $n \times 1$ and $\boldsymbol{X}$ is $n \times p$. Consider testing $H_0 : \boldsymbol{A}\boldsymbol{\beta} = \boldsymbol{c}$ where $\boldsymbol{A}$ is an $r \times p$ matrix with full rank r. To perform the test, suppose a bootstrap sample $\hat{\boldsymbol{\beta}}_1^*, ..., \hat{\boldsymbol{\beta}}_B^*$ has been generated. Form the prediction region (1.3) for $\boldsymbol{w}_1 = \boldsymbol{A}\hat{\boldsymbol{\beta}}_1^* - \boldsymbol{c}, ..., \boldsymbol{w}_B = \boldsymbol{A}\hat{\boldsymbol{\beta}}_B^* - \boldsymbol{c}$. If $\boldsymbol{0}$ is in the prediction region, fail to reject H_0, otherwise reject H_0.

It is useful to compare the bootstrap tests with classical tests. Methods for bootstrapping this model are well known. The estimated covariance matrix of the (ordinary) least squares estimator is

$$Cov(\hat{\boldsymbol{\beta}}_{OLS}) = MSE(\boldsymbol{X}^T\boldsymbol{X})^{-1}.$$

The residual bootstrap computes the least squares estimator and obtains the n residuals and fitted values $r_1, ..., r_n$ and $\hat{Y}_1, ..., \hat{Y}_n$. Then a sample of size n is selected with replacement from the residuals resulting in $r_{11}^*, ..., r_{1n}^*$. Hence the empirical distribution of the residuals is used. Then a vector $\boldsymbol{Y}_1^* = (Y_{11}^*, ..., Y_{1n}^*)^T$ is formed where $Y_{1j}^* = \hat{Y}_j + r_{1j}^*$. Then $\boldsymbol{Y}_1^*$ is regressed on $\boldsymbol{X}$ resulting in the estimator $\hat{\boldsymbol{\beta}}_1^*$. This process is repeated B times resulting in the estimators $\hat{\boldsymbol{\beta}}_1^*, ..., \hat{\boldsymbol{\beta}}_B^*$. This method should have $n > 10p$ so that the residuals r_i are close to the errors e_i.

Efron (1982, p. 36) notes that for the residual bootstrap, the sample covariance matrix of the $\hat{\boldsymbol{\beta}}_i^*$ is estimating the population bootstrap matrix $\frac{n-p}{n} MSE(\boldsymbol{X}^T\boldsymbol{X})^{-1}$ as $B \to \infty$. Hence the residual bootstrap standard error $SE(\hat{\beta}_i) \approx \sqrt{\frac{n-p}{n}}\, SE(\hat{\beta}_{i,OLS})$.

If the $\boldsymbol{z}_i = (Y_i, \boldsymbol{x}_i^T)^T$ are iid observations from some population, then a sample of size n can be drawn with replacement from $\boldsymbol{z}_1, ..., \boldsymbol{z}_n$. Then the response and predictor

variables can be formed into vector $\boldsymbol{Y}_1^*$ and design matrix $\boldsymbol{X}_1^*$. Then $\boldsymbol{Y}_1^*$ is regressed on $\boldsymbol{X}_1^*$ resulting in the estimator $\hat{\boldsymbol{\beta}}_1^*$. This process is repeated B times resulting in the estimators $\hat{\boldsymbol{\beta}}_1^*, ..., \hat{\boldsymbol{\beta}}_B^*$. If the $\boldsymbol{z}_i$ are the rows of a matrix $\boldsymbol{Z}$, then this rowwise bootstrap uses the empirical distribution of the $\boldsymbol{z}_i$. This method appears to need a larger sample size n than the residual bootstrap if $n > 10p$, but may be useful if n is large but $n < 5p$.

Following Seber and Lee (2003, p. 100), the classical test statistic for testing H_0 is

$$F = \frac{(\boldsymbol{A}\hat{\boldsymbol{\beta}} - \boldsymbol{c})^T [MSE \;\; \boldsymbol{A}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{A}^T]^{-1}(\boldsymbol{A}\hat{\boldsymbol{\beta}} - \boldsymbol{c})}{r},$$

and when H_0 is true, $rF_R \xrightarrow{D} \chi_r^2$ for a large class of error distributions. The sample covariance matrix $\boldsymbol{S}_{\boldsymbol{w}}$ of the $\boldsymbol{w}_i$ is estimating $\frac{n-p}{n} MSE \;\; \boldsymbol{A}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{A}^T$, and $\overline{\boldsymbol{w}} \approx \boldsymbol{0}$ when H_0 is true. Thus under H_0, the squared distance $D_i^2 = (\boldsymbol{w}_i - \overline{\boldsymbol{w}})^T \boldsymbol{S}_{\boldsymbol{w}}^{-1} (\boldsymbol{w}_i - \overline{\boldsymbol{w}}) \approx$

$$\frac{n}{n-p}(\boldsymbol{A}\hat{\boldsymbol{\beta}}^* - \boldsymbol{c})^T [MSE \;\; \boldsymbol{A}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{A}^T]^{-1}(\boldsymbol{A}\hat{\boldsymbol{\beta}}^* - \boldsymbol{c}),$$

and expect $D_{(d_B)}^2 \approx \frac{n}{n-p}\chi_{r,1-\delta}^2$, for large n and B and small p.

2.2 BOOTSTRAPPING THE VARIABLE SELECTION ESTIMATOR

Variable selection, also called subset or model selection, is the search for a subset of predictor variables that can be deleted without important loss of information. By treating a variable selection estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$ as a shrinkage estimator, the bootstrap can be used to examine variable selection. Forward selection, backward elimination, stepwise regression, and all subsets variable selection can be used if there is a criterion that selects the submodel, such as AIC or C_p. Similar ideas can be used to bootstrap other shrinkage estimators.

Consider testing $H_0 : \boldsymbol{A}\boldsymbol{\beta} = \boldsymbol{c}$ where $\boldsymbol{A}$ is an $r \times p$ matrix with full rank r. Now let $\hat{\boldsymbol{\beta}}$ be a variable selection estimator of $\boldsymbol{\beta}$. To perform the test, suppose a bootstrap sample $\hat{\boldsymbol{\beta}}_1^*, ..., \hat{\boldsymbol{\beta}}_B^*$ has been generated. Form the prediction region (1.3) for $\boldsymbol{w}_1 = \boldsymbol{A}\hat{\boldsymbol{\beta}}_1^* - \boldsymbol{c}, ..., \boldsymbol{w}_B = \boldsymbol{A}\hat{\boldsymbol{\beta}}_B^* - \boldsymbol{c}$. If $\boldsymbol{0}$ is in the prediction region, fail to reject H_0, otherwise reject H_0.

A *model for variable selection* in multiple linear regression can be described by

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

where e is an error, Y is the response variable, $\boldsymbol{x} = (\boldsymbol{x}_S^T, \boldsymbol{x}_E^T)^T$ is a $p \times 1$ vector of predictors, $\boldsymbol{x}_S$ 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.

Since S is unknown, candidate subsets will be examined. Following Olive and Hawkins (2005), 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). Then

$$Y = \boldsymbol{x}_I^T \boldsymbol{\beta}_I + \boldsymbol{x}_O^T \boldsymbol{\beta}_O + e. \tag{2.2}$$

The model $Y = \boldsymbol{x}^T \boldsymbol{\beta} + e$ that uses all of the predictors is called the *full model.* A model $Y = \boldsymbol{x}_I^T \boldsymbol{\beta}_I + e$ that only uses a subset $\boldsymbol{x}_I$ of the predictors is called a *submodel.*

Suppose that S is a subset of I and that model (2.1) holds. Then

$$\boldsymbol{x}^T \boldsymbol{\beta} = \boldsymbol{x}_S^T \boldsymbol{\beta}_S = \boldsymbol{x}_S^T \boldsymbol{\beta}_S + \boldsymbol{x}_{I/S}^T \boldsymbol{\beta}_{(I/S)} + \boldsymbol{x}_O^T \boldsymbol{0} = \boldsymbol{x}_I^T \boldsymbol{\beta}_I \tag{2.3}$$

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 $S \subseteq I$.

For multiple linear regression, if the candidate model of $\boldsymbol{x}_I$ has k terms (including the constant), then the partial F statistic for testing whether the $p - k$ predictor variables in $\boldsymbol{x}_O$ can be deleted is

$$F_I = \frac{SSE(I) - SSE}{(n-k) - (n-p)} / \frac{SSE}{n-p} = \frac{n-p}{p-k} \left[\frac{SSE(I)}{SSE} - 1 \right]$$

where SSE is the error sum of squares from the full model and SSE(I) is the error sum of squares from the candidate submodel. An important criterion for variable selection is the C_p criterion

$$C_p(I) = \frac{SSE(I)}{MSE} + 2k - n = (p-k)(F_I - 1) + k$$

where MSE is the error mean square for the full model. Olive and Hawkins (2005) show that submodels with $C_p(I) \leq \min(2k, p)$ are especially interesting. The AIC is criterion similar to C_p.

Suppose the variable selection method, such as forward selection or all subsets, produces J models. Let model I_{min} be the model that minimizes the criterion, e.g. $C_p(I)$ or $AIC(I)$. Following Seber and Lee (2003, p. 448) and Nishi (1984), the probability that model I_{min} from C_p or AIC underfits goes to zero as $n \to \infty$. Since there are a finite number of regression models I that contain the true model, and each model gives a consistent estimator of $\boldsymbol{\beta}$, the probability that I_{min} picks one of these models goes to one as $n \to \infty$. Hence $\hat{\boldsymbol{\beta}}_{I_{min}}$ is a consistent estimator of $\boldsymbol{\beta}$ under model (2.1).

Other automated variable selection methods may work better than I_{min}. For the C_p criterion, find the submodel I_I with the fewest number of predictors such that $C_p(I_I) \leq C_p(I_{min}) + 1$. For AIC, Burnham and Anderson (2004) suggest that if $\Delta(I) = AIC(I) - AIC(I_{min})$, then models with $\Delta(I) \leq 2$ are good. Find the submodel I_I with the smallest number of predictors such that $\Delta(I_I) \leq 2$. It is possible that $I_I = I_{min}$ or that I_I is the full model. Do not use more predictors than model I_I to avoid overfitting.

Suppose model I is selected after variable selection. Then least squares output for the model $\boldsymbol{Y} = \boldsymbol{X}_I\boldsymbol{\beta}_I + \boldsymbol{e}$ can be obtained, but the least squares output is not correct for inference. In particular, $MSE(I)(\boldsymbol{X}_I^T\boldsymbol{X}_I)^{-1}$ is not the correct estimated covariance matrix of $\hat{\boldsymbol{\beta}}_I$. The selected model tends to fit the data too well, so $SE(\hat{\beta}_i)$ from the incorrect estimated covariance matrix is too small. Hence the confidence intervals for β_i are too short, and hypotheses tests reject $H_0 : \beta_i = 0$ too often.

Hastie, Tibshirani, and Friedman (2009, p. 57) note that variable selection is a shrinkage estimator: the coefficients are shrunk to 0 for the omitted variables. Suppose $n > 10p$. If $\hat{\boldsymbol{\beta}}_I$ is $k \times 1$, form $\hat{\boldsymbol{\beta}}$ from $\hat{\boldsymbol{\beta}}_I$ by adding 0s corresponding to the omitted variables. Then $\hat{\boldsymbol{\beta}}$ is a nonlinear estimator of $\boldsymbol{\beta}$, and the residual bootstrap method can be applied. For example, suppose $\hat{\boldsymbol{\beta}}$ is formed from model I_{min} that minimizes C_p from some variable selection method such as forward selection, backward elimination, stepwise selection, or all subsets variable selection. Instead of computing the least squares estimator from regression $\boldsymbol{Y}_i^*$ on $\boldsymbol{X}$, perform variable selection on $\boldsymbol{Y}_i^*$ and $\boldsymbol{X}$, fit the model that minimizes the criterion,

and add 0s corresponding to the omitted variables, resulting in estimators $\hat{\boldsymbol{\beta}}_1^*, ..., \hat{\boldsymbol{\beta}}_B^*$. Also see Efron (2014).

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

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

CHAPTER 3

RESULTS

3.1 EXAMPLE

Cook and Weisberg (1999, pp. 351, 433, 447) gives a data set on 82 mussels sampled off the coast of New Zealand. Let the response variable be the logarithm $\log M$ of the *muscle mass*, and the predictors are the *length* L and *height* H of the shell in mm, the logarithm $\log W$ of the *shell width* W, the logarithm $\log S$ of the *shell mass* S and a constant. The R code used to produce the Table 3.1 and Table 3.2 is shown below. The mussels data was obtained from (http://lagrange.math.siu.edu/Olive/lregdata.txt).

```
library(leaps)
y <- log(mussels[,5])
x <- mussels[,1:4]
x[,4] <- log(x[,4])
x[,2] <- log(x[,2])
out <- regboot(x,y,B=1000)
tem <- rowboot(x,y,B=1000)
outvs <- vselboot(x,y,B=1000) #get bootstrap CIs,
apply(out$betas,2,shorth2); apply(tem$betas,2,shorth2);
apply(outvs$betas,2,shorth2)
ls.print(outvs$full)
ls.print(outvs$sub) #test if beta_2 = beta_3 = beta_4 = 0
Abeta <- out$betas[,2:4]  #method with residual bootstrap
predreg(Abeta)
Abeta <- outvs$betas[,2:4] #prediction region method with Imin
predreg(Abeta)
```

Table 3.1. Large Sample Full Model Inference

model	Estimate	Std.Err	$t-value$	$Pr(>\|t\|)$	rowboot	resboot
constant	-1.2493	0.8388	-1.4894	0.1405	[-2.720,-0.015]	[-3.065,0.110]
L	-0.0006	0.0023	-0.2829	0.7780	[-0.005,0.003]	[-0.005,0.003]
log W	0.1298	0.3738	0.3471	0.7295	[-0.390,0.710]	[-0.549,0.885]
H	0.0075	0.0050	1.5044	0.1366	[-0.001,0.017]	[-0.002,0.016]
log S	0.6404	0.1686	3.7989	0.0003	[0.209,1.025]	[0.337,0.947]

Inference for the full model is shown along with the shorth(c) nominal 95% confidence intervals for β_i computed using the rowwise and residual bootstraps. As expected, the residual bootstrap intervals are close to the classical least squares confidence intervals $\approx \hat{\beta}_i \pm 2SE(\hat{\beta}_i)$.

The minimum C_p model uses a constant, H and $\log S$. The shorth(c) nominal 95% confidence intervals for β_i using the residual bootstrap are shown. Note that the intervals for H and $\log(W)$ are right skewed and contain 0 when closed intervals are used instead of open intervals.

It was expected that $\log(S)$ may be the only predictor needed, along with a constant, since $\log(S)$ and $\log(M)$ are both log(mass) measurements and likely highly correlated. Hence want to test $H_0 : \beta_2 = \beta_3 = \beta_4 = 0$ with the I_{min} model selected by all subsets variable selection. Of course this test would be easy to do with the full model using least squares theory. Then $H_0 : \boldsymbol{A\beta} = (\beta_2, \beta_3, \beta_4)^T = \mathbf{0}$. Using the prediction region method with least squares gave an interval [0,2.937] with $D_{\mathbf{0}} = 1.594$. Note that $\sqrt{\chi^2_{3,0.95}} = 2.795$. So fail to reject H_0. The prediction region method using I_{min} had $[0, D_{(d_B)}] = [0, 3.282]$ while $D_{\mathbf{0}} = 1.137$. So fail to reject H_0.

Table 3.2. Incorrect min C_p submodel Inference

model	Estimate	Std.Err	$t-value$	$Pr(>\|t\|)$	resboot
constant	-0.9573	0.1519	-6.3018	0.0000	[-3.214,-0.593]
L	0				[-0.005, 0.003]
log W	0				[0.000, 0.977]
H	0.0072	0.0047	1.5490	0.1254	[0.000, 0.015]
log S	0.6530	0.1160	5.6297	0.0000	[0.358, 0.933]

3.2 SIMULATIONS

3.2.1 Bootstrapping Regression and Variable Selection

A small simulation study was done in R using $B = \max(1000, n)$ and 5000 runs. The regression model used $\boldsymbol{\beta} = (1, 1, 0, 0)^T$ with $n = 100$, $p = 4$ and various zero mean iid error distributions. The design matrix $\boldsymbol{X}$ consisted of iid N(0,1) random variables. Hence the full model least squares confidence intervals for β_i should have length near $2t_{99,0.975}\sigma/\sqrt{n} \approx 2(1.96)\sigma/10 = 0.392\sigma$ when the iid zero mean errors have variance σ^2. The simulation computed the shorth(k_B) interval for each β_i and used the prediction region method to test $H_0 : \beta_3 = \beta_4 = 0$. The nominal coverage was 0.95 with $\delta = 0.05$. Observed coverage between 0.94 and 0.96 would suggest coverage is close to the nominal value. Observed coverage near 0.94 would not be surprising since with $B = 1000$, expect about 1% undercoverage.

The function `regbootsim` is used to simulates residual bootstrap for multiple linear regression. The function `vsbootsim` is used to simulates bootstrap for all subsets variable selection. So need p small. For both these cases, five iid error distributions were used.

$Type = 1$ for $N(0,1)$ errors.

$Type = 2$ for t_3 errors.

$Type = 3$ for $exp(1) - 1$ errors.

$Type = 4$ for $uniform(-1,1)$ errors.

$Type = 5$ for [9.9 N(0,1) + 0.1 N(0,100)] errors.

The regression models used the residual bootstrap on the full model least squares estimator and on the all subsets variable selection estimator for the model I_{min}. The residuals were from least squares applied to the full model in both cases. Results are shown for when the iid errors $e_i \sim N(0,1)$. Table 3.3 shows two rows for each model giving the observed confidence interval coverages and average lengths of the confidence intervals for $Type = 1$. The term "reg" is for the full model regression, and the term "vs" is for the all subsets variable selection. The column for the "test" gives the length and coverage = P(fail to reject H_0) for the interval $[0, D_{(d_n)}]$ where $D_{(d_n)}$ is the cutoff for the prediction region. The volume of the prediction region will decrease to 0 as $n \to \infty$. The cutoff will often be near $\sqrt{\chi^2_{r,0.95}}$ if the statistic T is asymptotically normal.

Note that $\sqrt{\chi^2_{2,0.95}} = 2.448$ is close to 2.4503 for the full model regression bootstrap test. The coverages were near 0.94 for the regression bootstrap on the full model. For I_{min} the coverages were near 0.94 for β_1 and β_2, but higher for the other 3 tests since zeroes often occurred for $\hat{\beta}^*_j$ for $j = 3, 4$. The average lengths and coverages were similar for the full model and all subsets variable selection I_{min} for β_1 and β_2, but the lengths are shorter for I_{min} for β_3 and β_4. Volumes of the hyperellipsoids were not computed, but the average cutoff of 2.6859 for the variable selection test suggests that the test statistic was not multivariate normal, which is not surprising since many zeroes were produced for $\hat{\beta}^*_j$ for $j = 3, 4$.

Table 3.3. Bootstrapping Regression and Variable Selection for Type 1

model	cov/len	β_1	β_2	β_3	β_4	test
reg	cov	0.9322	0.9342	0.9354	0.9374	0.9386
	len	0.3823	0.3852	0.3862	0.3852	2.4503
vs	cov	0.9332	0.9358	0.9982	0.9956	0.9936
	len	0.3823	0.3847	0.3033	0.3035	2.6859

Furthermore, Table 3.4, Table 3.5 , Table 3.6 and Table 3.7 used Type 2, Type 3, Type 4 and Type 5 errors, respectively. According to the Table 3.4 average cutoff of 2.4914 for the full model regression bootstrap test while 2.7154 for the variable selection test. It emphasis that variable selection test statistic was not normal. Same result holds for Type 3 and Type 4 which is not surprising since many zeros were produced for $\hat{\beta}_j^*$ for $j = 3, 4$.. Furthermore, cutoff for Type 5 full model regression bootstrap test as depicted in Table 3.7 is also 2.55, suggest that the statistic T is not asymptotically normal.

Table 3.4. Bootstrapping Regression and Variable Selection for Type 2

model	cov/len	β_1	β_2	β_3	β_4	test
reg	cov	0.9334	0.9412	0.9362	0.9388	0.9422
	len	0.6387	0.6515	0.6511	0.6506	2.4914
vs	cov	0.9332	0.9344	0.9970	0.9960	0.9928
	len	0.6355	0.6508	0.5102	0.5130	2.7154

As far as coverages of Type 2 are concerned, it is closed to the nominal value for the regression bootstrap on the full model because the value is near 0.95. The average length and coverages were similar for the full model and all subsets variable selection I_{min} for β_1 and β_2, but the lengths are shorter for I_{min} for β_3 and β_4. In fact, the afore mentioned result was true for Type 3, Type 4 and Type 5 as depicted in Table 3.5 , Table 3.6 and Table 3.7 respectively. Nevertheless, lengths for regression and variable selection of Type 2 is considerably higher than all other five types while lowest lengths are for Type 5 in both cases.

Table 3.5. Bootstrapping Regression and Variable Selection for Type 3

model	cov/len	β_1	β_2	β_3	β_4	test
reg	cov	0.9344	0.9342	0.9364	0.9414	0.9388
	len	0.3791	0.3850	0.3844	0.3847	2.4742
vs	cov	0.9256	0.9394	0.9962	0.9976	0.9922
	len	0.3789	0.3845	0.3038	0.3026	2.7119

Table 3.6. Bootstrapping Regression and Variable Selection for Type 4

model	cov/len	β_1	β_2	β_3	β_4	test
reg	cov	0.9348	0.9426	0.9308	0.9398	0.9360
	len	0.2207	0.2224	0.2226	0.2224	2.4437
vs	cov	0.9390	0.9426	0.9966	0.9966	0.9930
	len	0.2211	0.2225	0.1756	0.1755	2.6963

Table 3.7. Bootstrapping Regression and Variable Selection for Type 5

model	cov/len	β_1	β_2	β_3	β_4	test
reg	cov	0.9300	0.9456	0.9438	0.9344	0.9500
	len	1.2201	1.2577	1.2532	1.2548	2.5500
vs	cov	0.9336	0.9308	0.9974	0.9972	0.9948
	len	1.2264	1.2594	1.0137	1.0153	2.7444

3.2.2 Bootstrapping the Correlation Matrix

One problem with the method is larger sample sizes n are needed as r increases. Olive (2013) suggested that for iid data $\boldsymbol{x}_i$ where $\boldsymbol{x}_i$ is $p \times 1$, the coverage started to get close to the nominal when $n > 20p$, but volume ratios needed $n > 50p$.

Consider testing whether correlations in a correlation matrix are 0. Let $\boldsymbol{\theta} = (\rho_{12}, ..., \rho_{1p}, \rho_{23}, ..., \rho_{2p}, ..., \rho_{p-1,p})^T$. There are $r = p(p-1)/2$ correlations $\rho_{i,j} = cor(X_i, X_j)$ where $i < j$. The simulation simulated iid data $\boldsymbol{w}$ with $\boldsymbol{x} = \boldsymbol{A}\boldsymbol{w}$ and $\boldsymbol{A}_{ij} = \psi$ for $i \neq j$ and $\boldsymbol{A}_{ii} = 1$. Hence

$$cor(X_i, X_j) = [2\psi + (p-2)\psi^2]/[1 + (p-1)\psi^2].$$

The function `corbootsim` is used to simulates bootstrap for correlation matrix. It stacks entries above the diagonal into a vector β. Make X for 10 different types of distribution.

$Type = 1$; for MVN $N_q(0, I)$.

$Type = 2, 3, 4$ and 5 ; for $(1-\delta)\ N_q(0, I) + \delta\ N_q(0, 25I)$ with $\delta = 0.4, 0.6, 0.1, 0.25$.

$Type = 6, 7, 8, 9$; for multivariate t_d with d=3, 5, 9, or 1.

$Type = 10$; for lognormal.

Small values of n which give the test coverage near 0.93 was found by trail and error method and then used $\psi = 0.03$ and 0.1 to evaluate the power of the test. Table 3.8 shows the results for multivariate normal data with $p = 4$ so $r = 6$ for testing $H_0 : \boldsymbol{\theta} = \mathbf{0}$. The nominal coverage was 0.95. For $n = 100$ and $\psi = 0$, the test failed to reject H_0 85.54% of the time, but 92.54% of the time for $n = 400$. Note that $\sqrt{\chi^2_{6,0.95}} = 3.548$. With $n = 400$ and $\psi > 0$, for the test the coverage = 1 - power. For $\psi = 0.3$ the simulated power was 0.558, but 1.0 for $\psi = 0.1$.

Table 3.8. Bootstrapping the Correlation Matrix for Type 1, $B = 1000$

n	ψ	cov/len	ρ_{12}	ρ_{13}	ρ_{14}	ρ_{23}	ρ_{24}	ρ_{34}	test
100	0	cov	0.9264	0.9370	0.9274	0.9347	0.9308	0.9322	0.8554
		len	0.3778	0.3787	0.3773	0.3780	0.3781	0.3775	3.5508
400	0	cov	0.9410	0.9470	0.9382	0.9404	0.9360	0.9436	0.9254
		len	0.1916	0.1915	0.1917	0.1916	0.1916	0.1916	3.5581
400	0.03	cov	0.9406	0.9378	0.9418	0.9426	0.9396	0.9414	0.4420
		len	0.1908	0.1909	0.1908	0.1908	0.1909	0.1908	3.5582
400	0.1	cov	0.9440	0.9400	0.9474	0.9402	0.9422	0.9478	0.0000
		len	0.1827	0.1827	0.1826	0.1835	0.1827	0.1827	3.5620

Table 3.9 reveals about the bootstrapping correlation matrix for Type 2 data. It can be seen that when n = 100 , the coverage is 0.7152 which is not around nominal coverage. The lowest n which occur close to 0.93 is $n = 900$. Note that $\sqrt{\chi^2_{6,0.95}} = 3.548$ is close to 3.5581 when $n = 900$. Furthermore, power is increasing as ψ increases and exactly equal to 1 when $\psi = 0.1$.

Table 3.9. Bootstrapping the Correlation Matrix for Type 2, $B = 1000$

n	ψ	cov/len	ρ_{12}	ρ_{13}	ρ_{14}	ρ_{23}	ρ_{24}	ρ_{34}	test
100	0	cov	0.9190	0.9204	0.9232	0.9164	0.9184	0.9086	0.7152
		len	0.5449	0.5456	0.5449	0.5460	0.5466	0.5456	3.5727
900	0	cov	0.9364	0.9386	0.9426	0.9400	0.9402	0.9430	0.9256
		len	0.1905	0.1905	0.1906	0.1906	0.1906	0.1904	3.5604
900	0.03	cov	0.9400	0.9422	0.9380	0.9458	0.9418	0.9376	0.4510
		len	0.1898	0.1897	0.1895	0.1899	0.1894	0.1896	3.5596
900	0.1	cov	0.9436	0.9396	0.9390	0.9432	0.9356	0.9368	0.0000
		len	0.1816	0.1816	0.1819	0.1816	0.1818	0.1815	3.5635

Moreover, the bootstrapping correlation matrix for Type 3, 4, 5, 6, 7, 8 and 10 were generated and shows in Table 3.10, Table 3.11, Table 3.12, Table 3.13, Table 3.14, Table 3.15 and Table 3.16 respectively. The small n value which gives nominal coverage 0.93 could not be found for Type 9 perhaps the correlation matrix does not exist for Type 9. All other types for $n = 100$ and $\psi = 0$, the test fail to reject $H_0 : \boldsymbol{\theta} = \mathbf{0}$ and lowest n which gives the nominal coverages are: n= 650, 3000, 1500, 25000, 1800, 500 and 12000 for Type 3, 4, 5, 6, 7, 8 and 10 respectively. Indeed, all these types cutoff lengths near $\sqrt{\chi^2_{6,0.95}} = 3.548$ for aforementioned values of n. Furthermore, it is not surprising that power increased as ψ increased and becomes 1 when $\psi = 0.1$ for each type.

Table 3.10. Bootstrapping the Correlation Matrix for Type 3, $B = 1000$

n	ψ	cov/len	ρ_{12}	ρ_{13}	ρ_{14}	ρ_{23}	ρ_{24}	ρ_{34}	test
100	0	cov	0.9212	0.9230	0.9300	0.9256	0.9186	0.9268	0.7922
		len	0.4674	0.4663	0.4669	0.4683	0.4660	0.4682	3.5581
650	0	cov	0.9366	0.9360	0.9364	0.9412	0.9414	0.9398	0.9294
		len	0.1888	0.1889	0.1890	0.1891	0.1891	0.1891	3.5609
650	0.03	cov	0.9488	0.9408	0.9452	0.9458	0.9418	0.9430	0.4278
		len	0.1882	0.1887	0.1880	0.1883	0.1884	0.1882	3.5605
650	0.1	cov	0.9380	0.9396	0.9382	0.9388	0.9334	0.9386	0.0000
		len	0.1804	0.1802	0.1805	0.1804	0.1801	0.1803	3.5634

Table 3.11. Bootstrapping the Correlation Matrix for Type 4, $B = 1000$

n	ψ	cov/len	ρ_{12}	ρ_{13}	ρ_{14}	ρ_{23}	ρ_{24}	ρ_{34}	test
100	0	cov	0.8802	0.8864	0.8782	0.8810	0.8820	0.8788	0.4214
		len	0.7162	0.7141	0.7119	0.7128	0.7168	0.7123	3.6728
3000	0	cov	0.9382	0.9452	0.9422	0.9408	0.9414	0.9386	0.9302
		len	0.1629	0.1635	0.1633	0.1634	0.1631	0.1633	3.5584
3000	0.03	cov	0.9418	0.9416	0.9396	0.9374	0.9404	0.9452	0.2904
		len	0.1629	0.1625	0.1626	0.1627	0.1626	0.1625	3.5588
3000	0.1	cov	0.9428	0.9400	0.9380	0.9466	0.9342	0.9370	0.0000
		len	0.1557	0.1561	0.1557	0.1559	0.1558	0.1559	3.5609

Table 3.12. Bootstrapping the Correlation Matrix for Type 5, $B = 1000$

n	ψ	cov/len	ρ_{12}	ρ_{13}	ρ_{14}	ρ_{23}	ρ_{24}	ρ_{34}	test
100	0	cov	0.9084	0.9064	0.9018	0.9060	0.9082	0.9018	0.6130
		len	0.6325	0.6308	0.6311	0.6332	0.6330	0.6307	3.5997
1500	0	cov	0.9404	0.9412	0.9426	0.9418	0.9452	0.9408	0.9274
		len	0.1766	0.1769	0.1767	0.1767	0.1768	0.1767	3.5601
1500	0.03	cov	0.9404	0.9394	0.9390	0.9382	0.9364	0.9362	0.3782
		len	0.1762	0.1762	0.1763	0.1762	0.1762	0.1761	3.5608
1500	0.1	cov	0.9462	0.9416	0.9424	0.9426	0.9374	0.9416	0.0000
		len	0.1689	0.1689	0.1690	0.1690	0.1689	0.1687	3.5637

Table 3.13. Bootstrapping the Correlation Matrix for Type 6, $B = 1000$

n	ψ	cov/len	ρ_{12}	ρ_{13}	ρ_{14}	ρ_{23}	ρ_{24}	ρ_{34}	test
100	0	cov	0.8980	0.8990	0.8948	0.9050	0.8972	0.8994	0.6412
		len	0.5912	0.5962	0.5919	0.5957	0.5929	0.5943	3.6574
25000	0	cov	0.9412	0.9334	0.9426	0.9442	0.9416	0.9358	0.9282
		len	0.1200	0.1198	0.1204	0.1205	0.1211	0.1191	3.5714
25000	0.03	cov	0.9364	0.9400	0.9386	0.9364	0.9356	0.9354	0.0084
		len	0.1186	0.1204	0.1188	0.1200	0.1191	0.1195	3.5727
25000	0.1	cov	0.9378	0.9380	0.9408	0.9352	0.9390	0.9378	0.0000
		len	0.1148	0.1158	0.1140	0.1164	0.1154	0.1152	3.5748

Table 3.14. Bootstrapping the Correlation Matrix for Type 7, $B = 1000$

n	ψ	cov/len	ρ_{12}	ρ_{13}	ρ_{14}	ρ_{23}	ρ_{24}	ρ_{34}	test
100	0	cov	0.9188	0.9176	0.9158	0.9130	0.9176	0.9140	0.7434
		len	0.4793	0.4781	0.4773	0.4760	0.4789	0.4775	3.5646
1800	0	cov	0.9410	0.9404	0.9372	0.9420	0.9374	0.9376	0.9256
		len	0.1385	0.1385	0.1384	0.1385	0.1385	0.1382	3.5588
1800	0.03	cov	0.9394	0.9410	0.9338	0.9402	0.9368	0.9374	0.0888
		len	0.1382	0.1382	0.1380	0.1387	0.1388	0.1383	3.5590
1800	0.1	cov	0.9422	0.9418	0.9390	0.9382	0.9342	0.9400	0.0000
		len	0.1324	0.1327	0.1324	0.1318	0.1320	0.1317	3.5599

Table 3.15. Bootstrapping the Correlation Matrix for Type 8, $B = 1000$

n	ψ	cov/len	ρ_{12}	ρ_{13}	ρ_{14}	ρ_{23}	ρ_{24}	ρ_{34}	test
100	0	cov	0.9280	0.9288	0.9296	0.9232	0.9264	0.9340	0.8416
		len	0.3949	0.3966	0.3963	0.3957	0.3959	0.3962	3.5495
500	0	cov	0.9394	0.9410	0.9358	0.9376	0.9376	0.9452	0.9266
		len	0.1819	0.1818	0.1819	0.1816	0.1816	0.1818	3.5583
500	0.03	cov	0.9352	0.9446	0.9474	0.9418	0.9364	0.9416	0.4090
		len	0.1810	0.1812	0.1808	0.1812	0.1809	0.1810	3.5572
500	0.1	cov	0.9382	0.9388	0.9420	0.9446	0.9440	0.9394	0.0000
		len	0.1733	0.1734	0.1738	0.1736	0.1740	0.1736	3.5608

Table 3.16. Bootstrapping the Correlation Matrix for Type 10, $B = 1000$

n	ψ	cov/len	ρ_{12}	ρ_{13}	ρ_{14}	ρ_{23}	ρ_{24}	ρ_{34}	test
100	0	cov	0.8928	0.8932	0.8916	0.8886	0.8910	0.8930	0.7478
		len	0.3347	0.3320	0.3367	0.3302	0.3363	0.3328	3.7545
12000	0	cov	0.9382	0.9346	0.9368	0.9366	0.9384	0.9392	0.9314
		len	0.0341	0.0343	0.0343	0.0342	0.0340	0.0344	3.5740
12000	0.03	cov	0.9350	0.9408	0.9364	0.9408	0.9370	0.9394	0.0000
		len	0.0341	0.0343	0.0341	0.0340	0.0340	0.0340	3.5742
12000	0.1	cov	0.9402	0.9400	0.9438	0.9422	0.9376	0.9386	0.0000
		len	0.0330	0.0332	0.0332	0.0333	0.0332	0.0334	3.5726

On the other hand, the simulation was extended by changing the bootstrap sample size $B = 1000$ to $B = 4000$. For 5000 runs, $B = 4000$ and $\psi = 0$, the smallest n which occur nominal coverages for all types were obtained. Table 3.17 to Table 3.26 reveals about bootstrapping correlation Matrix when $B = 4000$ for Type 1 to Type 10 excluding Type 9. In this case, all the coverages and lengths are almost similar as $B = 1000$ case. In fact, the smallest n values for nominal coverages are $n = 400$, 900, 650, 3000, 1500, 25000, 1800, 500 and 12000 for Type 1, Type 2, Type 3, Type 4, Type 5, Type 6, Type 7, Type 8, Type 10 respectively. Therefore, it can be concluded that increasing the number of bootstrap samples did not much affect testing the correlation matrix.

Table 3.17. Bootstrapping the Correlation Matrix for Type 1, $B = 4000$

n	B	cov/len	ρ_{12}	ρ_{13}	ρ_{14}	ρ_{23}	ρ_{24}	ρ_{34}	test
100	1000	cov	0.9264	0.9370	0.9274	0.9347	0.9308	0.9322	0.8554
		len	0.3778	0.3787	0.3773	0.3780	0.3781	0.3775	3.5508
400	1000	cov	0.9410	0.9470	0.9382	0.9404	0.9360	0.9436	0.9254
		len	0.1916	0.1915	0.1917	0.1916	0.1916	0.1916	3.5581
100	4000	cov	0.9332	0.9328	0.9340	0.9318	0.9384	0.9348	0.8462
		len	0.3810	0.3818	0.3821	0.3818	0.3824	0.3824	3.5334
400	4000	cov	0.9418	0.9404	0.9486	0.9420	0.9494	0.9454	0.9248
		len	0.1938	0.1936	0.1934	0.1935	0.1937	0.1937	3.5420

Table 3.18. Bootstrapping the Correlation Matrix for Type 2, $B = 4000$

n	B	cov/len	ρ_{12}	ρ_{13}	ρ_{14}	ρ_{23}	ρ_{24}	ρ_{34}	test
100	1000	cov	0.9190	0.9204	0.9232	0.9164	0.9184	0.9086	0.7152
		len	0.5449	0.5456	0.5449	0.5460	0.5466	0.5456	3.5727
900	1000	cov	0.9364	0.9386	0.9426	0.9400	0.9402	0.9430	0.9256
		len	0.1905	0.1905	0.1906	0.1906	0.1906	0.1904	3.5604
100	4000	cov	0.9252	0.9138	0.9252	0.9220	0.9144	0.9194	0.7164
		len	0.5519	0.5483	0.5499	0.5491	0.5515	0.5515	3.5543
900	4000	cov	0.9430	0.9434	0.9394	0.9476	0.9400	0.9430	0.9208
		len	0.1928	0.1925	0.1920	0.1923	0.1923	0.1923	3.5438

Table 3.19. Bootstrapping the Correlation Matrix for Type 3, $B = 4000$

n	B	cov/len	ρ_{12}	ρ_{13}	ρ_{14}	ρ_{23}	ρ_{24}	ρ_{34}	test
100	1000	cov	0.9212	0.9230	0.9300	0.9256	0.9186	0.9268	0.7922
		len	0.4674	0.4663	0.4669	0.4683	0.4660	0.4682	3.5581
650	1000	cov	0.9366	0.9360	0.9364	0.9412	0.9414	0.9398	0.9294
		len	0.1888	0.1889	0.1890	0.1891	0.1891	0.1891	3.5609
100	4000	cov	0.9210	0.9380	0.9372	0.9270	0.9340	0.9334	0.8080
		len	0.4709	0.4735	0.4733	0.4722	0.4734	0.4733	3.5416
650	4000	cov	0.9454	0.9462	0.9440	0.9420	0.9408	0.9446	0.9252
		len	0.1908	0.1909	0.1908	0.1911	0.1910	0.1911	3.5436

Table 3.20. Bootstrapping the Correlation Matrix for Type 4, $B = 4000$

n	B	cov/len	ρ_{12}	ρ_{13}	ρ_{14}	ρ_{23}	ρ_{24}	ρ_{34}	test
100	1000	cov	0.8802	0.8864	0.8782	0.8810	0.8820	0.8788	0.4214
		len	0.7162	0.7141	0.7119	0.7128	0.7168	0.7123	3.6728
3000	1000	cov	0.9382	0.9452	0.9422	0.9408	0.9414	0.9386	0.9302
		len	0.1629	0.1635	0.1633	0.1634	0.1631	0.1633	3.5584
100	4000	cov	0.8842	0.8870	0.8816	0.8844	0.8806	0.8872	0.4140
		len	0.7197	0.7205	0.7181	0.7193	0.7186	0.7205	3.6492
3000	4000	cov	0.9474	0.9432	0.9498	0.9400	0.9434	0.9384	0.9142
		len	0.1650	0.1649	0.1649	0.1650	0.1650	0.1648	3.5425

Table 3.21. Bootstrapping the Correlation Matrix for Type 5, $B = 4000$

n	B	cov/len	ρ_{12}	ρ_{13}	ρ_{14}	ρ_{23}	ρ_{24}	ρ_{34}	test
100	1000	cov	0.9084	0.9064	0.9018	0.9060	0.9082	0.9018	0.6130
		len	0.6325	0.6308	0.6311	0.6332	0.6330	0.6307	3.5997
1500	1000	cov	0.9404	0.9412	0.9426	0.9418	0.9452	0.9408	0.9274
		len	0.1766	0.1769	0.1767	0.1767	0.1768	0.1767	3.5601
100	4000	cov	0.9034	0.9172	0.9038	0.9108	0.9144	0.9044	0.5998
		len	0.6404	0.6415	0.6371	0.6389	0.6418	0.6391	3.5791
1500	4000	cov	0.9466	0.9402	0.9434	0.9436	0.9468	0.9442	0.9250
		len	0.1785	0.1787	0.1790	0.1788	0.1783	0.1787	3.5441

Table 3.22. Bootstrapping the Correlation Matrix for Type 6, $B = 4000$

n	B	cov/len	ρ_{12}	ρ_{13}	ρ_{14}	ρ_{23}	ρ_{24}	ρ_{34}	test
100	1000	cov	0.8980	0.8990	0.8948	0.9050	0.8972	0.8994	0.6412
		len	0.5912	0.5962	0.5919	0.5957	0.5929	0.5943	3.6574
25000	1000	cov	0.9412	0.9334	0.9426	0.9442	0.9416	0.9358	0.9282
		len	0.1200	0.1198	0.1204	0.1205	0.1211	0.1191	3.5714
100	4000	cov	0.9022	0.9024	0.9084	0.9024	0.9020	0.8986	0.6434
		len	0.5967	0.5978	0.6020	0.5998	0.6001	0.6036	3.6358
25000	4000	cov	0.9494	0.9424	0.9434	0.9392	0.9426	0.9466	0.9246
		len	0.1210	0.1229	0.1216	0.1225	0.1210	0.1224	3.5547

Table 3.23. Bootstrapping the Correlation Matrix for Type 7, $B = 4000$

n	B	cov/len	ρ_{12}	ρ_{13}	ρ_{14}	ρ_{23}	ρ_{24}	ρ_{34}	test
100	1000	cov	0.9188	0.9176	0.9158	0.9130	0.9176	0.9140	0.7434
		len	0.4793	0.4781	0.4773	0.4760	0.4789	0.4775	3.5646
1800	1000	cov	0.9410	0.9404	0.9372	0.9420	0.9374	0.9376	0.9256
		len	0.1385	0.1385	0.1384	0.1385	0.1385	0.1382	3.5588
100	4000	cov	0.9238	0.9212	0.9216	0.9158	0.9244	0.9174	0.7454
		len	0.4827	0.4796	0.4843	0.4798	0.4824	0.4805	3.5466
1800	4000	cov	0.9440	0.9436	0.9490	0.9480	0.9488	0.9462	0.9360
		len	0.1402	0.1402	0.1404	0.1402	0.1404	0.1402	3.54200

Table 3.24. Bootstrapping the Correlation Matrix for Type 8, $B = 4000$

n	B	cov/len	ρ_{12}	ρ_{13}	ρ_{14}	ρ_{23}	ρ_{24}	ρ_{34}	test
100	1000	cov	0.9280	0.9288	0.9296	0.9232	0.9264	0.9340	0.8416
		len	0.3949	0.3966	0.3963	0.3957	0.3959	0.3962	3.5495
500	1000	cov	0.9394	0.9410	0.9358	0.9376	0.9376	0.9452	0.9266
		len	0.1819	0.1818	0.1819	0.1816	0.1816	0.1818	3.5583
100	4000	cov	0.9304	0.9372	0.9338	0.9284	0.9322	0.9296	0.8368
		len	0.3993	0.3996	0.3996	0.4004	0.3993	0.4000	3.5318
500	4000	cov	0.9464	0.9446	0.9514	0.9488	0.9436	0.9416	0.9314
		len	0.1834	0.1836	0.1832	0.1837	0.1837	0.1837	3.5415

Table 3.25. Bootstrapping the Correlation Matrix for Type 10, $B = 4000$

n	B	cov/len	ρ_{12}	ρ_{13}	ρ_{14}	ρ_{23}	ρ_{24}	ρ_{34}	test
100	1000	cov	0.8928	0.8932	0.8916	0.8886	0.8910	0.8930	0.7478
		len	0.3347	0.3320	0.3367	0.3302	0.3363	0.3328	3.7545
12000	1000	cov	0.9382	0.9346	0.9368	0.9366	0.9384	0.9392	0.9314
		len	0.0341	0.0343	0.0343	0.0342	0.0340	0.0344	3.5740
100	4000	cov	0.8986	0.9002	0.8958	0.8954	0.8924	0.8944	0.7424
		len	0.3338	0.3351	0.3392	0.3347	0.3370	0.3387	3.7336
12000	4000	cov	0.9428	0.9440	0.9426	0.9460	0.9454	0.9406	0.9288
		len	0.0346	0.03455	0.0343	0.0346	0.0344	0.0346	3.5574

3.3 CONCLUSIONS

Applying the large sample $100(1-\delta)\%$ prediction region to the bootstrap sample $T^*_{(1)}, T^*_{(2)}, ..., T^*_{(B)}$ gives a large sample $100(1-\delta)\%$ confidence region for an $r \times 1$ parameter vector for θ, generalizing the percentile method for $r = 1$ to $r \geq 1$. Moreover, the prediction region method can be regarded as special case of the percentile method where the test statistic is the squared Mahalanobis distance $D_i^{2*} = (T_i^* - \overline{T^*})^T [\boldsymbol{S}_T^*]^{-1} (T_i^* - \overline{T^*}))$ where $\boldsymbol{w}_i = T_i^*$, and $\overline{T^*}$ and $\boldsymbol{S}_T^*$ are the sample mean and sample covariance martix of $T_1^*, ..., T_B^*$.

Applications of the prediction region method are numerous, but may need $n \geq 50r$ and $B \geq \max(1000, n)$ if the test statistic has an approximate multivariate normal distribution. Sample sizes may need to be much larger for other limiting distribution.

Example of bootstrapping hypothesis test was discussed based on Cook and Weisberg (1999) mussels data set. Simulations were done in *R*. See R Development Core Team (2011). The collection of *R* functions *lregpack*, available at

(http://lagrange.math.siu.edu/Olive/lregpack.txt), has some useful functions for the prediction region method. The function `vselboot` bootstraps the minimum C_p model from all subsets variable selection. The function `shorth2` can be used to find the shorth intervals for $\hat{\theta}_i$. The function `predreg` computes the prediction region and the Mahalanobis distance of the zero vector corresponding to $\boldsymbol{A\theta} - \boldsymbol{c} = \boldsymbol{0}$. The functions `rowboot` and `regboot` do the rowwise and residual bootstrap for the full model. The functions `regbootsim` and `vsbootsim` can be used to simulate the bootstrap tests for multiple linear regression and for the all subsets variable selection model that minimizes C_p. The functions `corboot` and `corbootsim` can be used to bootstrap the correlation matrix.

Bootstrapping regression for the full model and bootstrapping variable selection for the model I_{min} has been done for five different types of error distributions. Coverages of β_1, β_2, β_3 and β_4 are close to nominal value for the regression bootstrap on the full model. The average length and coverages were similar the full model and all subsets variable selection I_{min} for β_1 and β_2, but the lengths are smaller for I_{min} for β_3 and β_4. Also the cutoffs for the full model regression near $\sqrt{\chi^2_{2,0.95}}$ suggest that the test statistic is asymptotically normal while the average cutoff for the variable selection suggests that the test statistic was not normal, which is not surprising since many zeroes were produced for $\hat{\beta}^*_j$ for $j = 3, 4$.

Small values of n which give coverage close to the nominal coverage for 10 different types of distributions were found by generating the bootstrapping the correlation matrix. Indeed, n = 400, 900, 650, 3000, 1500, 25000, 1800, 500 and 12000 for Type 1, Type 2, Type 3, Type 4, Type 5, Type 6, Type 7, Type 8, Type 10 respectively. All these different types have cutoff near $\sqrt{\chi^2_{6,0.95}}$ and power increased as ψ increased. Furthermore, it can be observed that increasing the number of bootstrap samples B did not much affect testing the correlation matrix.

REFERENCES

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

[2] Cook, R.D., and Weisberg, S., *Applied Regression Including Computing and Graphics*, Wiley, New York, NY, 1999.

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

[4] Efron, B., and Tibshirani, R.J., *An Introduction to the Bootstrap*, Chapman and Hall, New York, NY, 1993.

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

[6] Frey, J., "Data-Driven Nonparametric Prediction Intervals," *Sociological Methods & Research*, **143** (2013), 1039–1048.

[7] Hastie, T., Tibshirani, R., and Friedman, J., *The Elements of Statistical Learning: Data Mining, Inference and Prediction*, Springer, New York, NY, 2009.

[8] Nishi, R., "Asymptotic Properties of Criteria for Selection of Variables in Multiple Regression," *Annals of Statistics*, **12** (2013), 758–765.

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

[10] Olive, D.J., *Statistical Theory and Inference*, Springer, New York, NY, 2014.

[11] Olive, D.J., *Highest Density Region Prediction*, unpublished manuscript at (http://lagrange.math.siu.edu/Olive/pphdrpred.pdf), 2015a.

[12] Olive, D.J., *Bootstrapping Hypotheses Tests*, unpublished manuscript at (http://lagrange.math.siu.edu/Olive/ppvselboot.pdf), 2015b.

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

[14] R Development Core Team., *R: a Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, (www.R-project.org), 2011.

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

VITA

Graduate School
Southern Illinois University

Chathurangi H. Karunapala Pathiravasan

email address (chathurangi_heshani@yahoo.com)

University of Sri Jayewardenepura, Sri Lanka
Bachelor of Science, Mathematics, July 2010

Research Paper Title:
BOOTSTAPPING HYPOTHESES TESTS

Major Professor: Dr. David J. Olive