ABSTRACT

Name: Cathy D. Poliak                      Department: Mathematical Sciences

Title: Observed Confidence Levels for Regression Parameters

Major: Mathematical Sciences                          Degree: Doctor of Philosophy

Approved by:                                         Date:

Dissertation Director

NORTHERN ILLINOIS UNIVERSITY
ABSTRACT

Regression models are used in various subjects such as economics, chemistry and engineering. Suppose a sample is taken from a population in which regression parameters are of interest and the parameter space has been divided into a countable set of possible overlapping subsets. The problem of regions is concerned with which of the subsets contains the true parameter. To solve this problem a measure of confidence is applied to these regions. This paper develops a measure of confidence for regions of a parameter space for regression parameters. Some literature that investigates the problem of regions and the development of the confidence levels are discussed in this paper. We will also look at the asymptotic properties of these measures. Then the methods are applied to examples.
NORTHERN ILLINOIS UNIVERSITY

OBSERVED CONFIDENCE LEVELS FOR REGRESSION PARAMETERS

A DISSERTATION SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE
DOCTOR OF PHILOSOPHY

DEPARTMENT OF
MATHEMATICAL SCIENCES

BY
CATHY D. POLIAK
© 2007 Cathy D. Poliak

DEKALB, ILLINOIS
AUGUST 2007
Certification: In accordance with departmental and Graduate School policies, this dissertation is accepted in partial fulfillment of degree requirements.

[Signature]

Dissertation Director

Date: 7/27/2007
ACKNOWLEDGMENTS

I would first like to thank my dissertation advisor, Dr. Alan Polansky, whose help and encouragement allowed me to finish this task. I am also grateful to the faculty and staff in the Division of Statistics at Northern Illinois University. This experience was enhanced by their support.

I especially want to thank my family, who has encouraged me in whatever I endeavor to do. Finally, I would like to thank my husband, John, for his unfailing love and support.
DEDICATION

To my father, Rev. Daniel J. Darulla, May 20, 2004
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF TABLES</td>
<td>vii</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>viii</td>
</tr>
<tr>
<td>Chapter</td>
<td></td>
</tr>
<tr>
<td>1. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>2. REVIEW OF LITERATURE</td>
<td>3</td>
</tr>
<tr>
<td>2.1 Problem of Regions</td>
<td>3</td>
</tr>
<tr>
<td>2.2 Critical Points</td>
<td>7</td>
</tr>
<tr>
<td>2.3 Edgeworth Expansion</td>
<td>10</td>
</tr>
<tr>
<td>2.4 Observed Confidence Levels</td>
<td>11</td>
</tr>
<tr>
<td>2.5 Regression</td>
<td>18</td>
</tr>
<tr>
<td>3. OBSERVED CONFIDENCE LEVELS FOR SIMPLE LINEAR MODEL PARAMETERS</td>
<td>29</td>
</tr>
<tr>
<td>3.1 Observed Confidence Levels</td>
<td>29</td>
</tr>
<tr>
<td>3.2 Asymptotic Accuracy</td>
<td>35</td>
</tr>
<tr>
<td>3.3 Bias Corrected</td>
<td>44</td>
</tr>
<tr>
<td>3.4 Intercept Parameter</td>
<td>50</td>
</tr>
<tr>
<td>3.5 Example</td>
<td>54</td>
</tr>
<tr>
<td>4. OBSERVED CONFIDENCE LEVELS FOR MULTIPLE REGRESSION PARAMETERS</td>
<td>57</td>
</tr>
<tr>
<td>Chapter</td>
<td>Page</td>
</tr>
<tr>
<td>---------</td>
<td>------</td>
</tr>
<tr>
<td>4.1 Confidence Regions</td>
<td>58</td>
</tr>
<tr>
<td>4.2 Observed Confidence Levels</td>
<td>60</td>
</tr>
<tr>
<td>4.3 Asymptotic Analysis</td>
<td>63</td>
</tr>
<tr>
<td>4.4 Example</td>
<td>68</td>
</tr>
<tr>
<td>4.5 Model Selection</td>
<td>73</td>
</tr>
<tr>
<td>5. EMPIRICAL STUDY</td>
<td>74</td>
</tr>
<tr>
<td>6. FUTURE WORK</td>
<td>83</td>
</tr>
<tr>
<td>BIBLIOGRAPHY</td>
<td>84</td>
</tr>
<tr>
<td>APPENDIX: R CODE</td>
<td>86</td>
</tr>
</tbody>
</table>
LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1. Least squares estimates for fitting a linear regression to each tree type.</td>
<td>71</td>
</tr>
<tr>
<td>4.2. Observed confidence levels for the six different regions of the ordering of the slope parameters.</td>
<td>73</td>
</tr>
<tr>
<td>5.1. Distributions for residuals $\varepsilon_i$.</td>
<td>76</td>
</tr>
<tr>
<td>5.2. Observed confidence levels using the standard normal distribution.</td>
<td>77</td>
</tr>
<tr>
<td>5.3. Observed confidence levels using the skewed unimodal distribution.</td>
<td>77</td>
</tr>
<tr>
<td>5.4. Observed confidence levels using strongly skewed distribution.</td>
<td>78</td>
</tr>
<tr>
<td>5.5. Observed confidence levels using the kurtotic unimodal distribution.</td>
<td>78</td>
</tr>
<tr>
<td>5.6. Observed confidence levels using the outlier distribution.</td>
<td>79</td>
</tr>
</tbody>
</table>
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>The measured and estimated GI values</td>
<td>55</td>
</tr>
<tr>
<td>4.1</td>
<td>Felling time and breast height (diameter) for maple, oak, and birch</td>
<td>69</td>
</tr>
<tr>
<td>5.1</td>
<td>The distribution shapes</td>
<td>75</td>
</tr>
<tr>
<td>5.2</td>
<td>Normal distribution</td>
<td>79</td>
</tr>
<tr>
<td>5.3</td>
<td>Skewed unimodal distribution</td>
<td>80</td>
</tr>
<tr>
<td>5.4</td>
<td>Strongly skewed distribution</td>
<td>80</td>
</tr>
<tr>
<td>5.5</td>
<td>Kurtotic unimodal distribution</td>
<td>81</td>
</tr>
<tr>
<td>5.6</td>
<td>Outlier distribution</td>
<td>81</td>
</tr>
</tbody>
</table>
CHAPTER 1

INTRODUCTION

Suppose we wish to estimate a set of parameters $\theta = (\theta_1, \theta_2, \ldots, \theta_p)$ based upon $X = \{X_1, \ldots, X_n\}$, a set of independent and identically distributed $d$-dimensional random vectors that follow a distribution $F \in \mathcal{F}$, where $\mathcal{F}$ is a family of distributions. Define $T : \mathcal{F} \to \mathbb{R}^p$ such that $\theta = T(F)$ where $\Theta = \{T(F) : F \in \mathcal{F}\}$. The set $\Theta$ is called the parameter space of $\theta$. Suppose that the parameter space $\Theta_k$, $k = 1, 2, \ldots$ is a countable collection of subsets of $\Theta$ such that

$$\bigcup_{k=1}^{\infty} \Theta_k = \Theta,$$

where the sequence $\{\Theta_k\}_{k=1}^{\infty}$ may not necessarily be a partition of $\Theta$ but may overlap. The problem of interest is to determine which of the subsets the parameter $\theta$ belongs to based on the sample $X$. This problem of regions was introduced by Efron and Tibshirani (1996, 1998). It has been explored in many different applications, for example, in model selection and in determining the number of modes for an unknown density. This research will consider applying the problem of regions to regression problems.

Define $\hat{\theta}$ to be the “plug-in” estimator of $\theta$ such that $\hat{\theta} = T[\hat{F}(X)]$, where $\hat{F} : \mathbb{R}^{dn} \to \mathcal{F}$ is a function that is the estimator of $F$ based on the sample $X$. The basic approach to this problem is to conclude that the parameter $\theta$ is in the region $\Theta_k$ whenever $\hat{\theta}$ is also in that region. However, that may not always be the case.
because of the inherent variability of \( \hat{\theta} \). Another method developed by Efron and Tibshirani (1996, 1998) to determine if the parameter is in the region is to compute a measure of confidence for \( \theta \) being in the region. This is usually computed using the bootstrap. An alternative approach to this problem is observed confidence levels introduced by Polansky (2003a,b).

This research will consider regions of a parameter space for regression parameters. Given a region in a regression parameter space and regression data, we wish to determine how confident we are that the true regression parameters are within a specified region. To find these confidence levels we will use bootstrap procedures that were introduced by Efron (1981) and Polansky (2003a,b and 2005a,b).

In Chapter 2 we review the literature that has been published that is related to the topic of the observed confidence levels and estimating regression parameters based on resampling methods. First, we will review literature that describes the problem of regions. Then, we will describe the critical points using bootstrap methods and determine their accuracy. Finally, we will review literature that define observed confidence levels for scalar and vector parameters. In Chapter 3 we introduce the regression model and define the observed confidence levels for the parameters of a simple linear model. Chapter 4 explores finding observed confidence levels for regression parameters of a multiregression model. In Chapter 5 we demonstrate the accuracy of the observed confidence levels by simulations. Finally, Chapter 6 explains some future work on the observed confidence levels for the regression parameters.
CHAPTER 2

REVIEW OF LITERATURE

This chapter reviews the literature related to the problem of interest. First, this review will look at early work that was done using bootstrap procedure to produce confidence levels for parameters and the problem of regions. Second, we will look at papers and books that introduce the critical points for confidence regions. Then we will look at the work that was done with observed confidence levels. Finally, we will look at work that has been done for regression parameters.

2.1 Problem of Regions

The primary solution to the problem of regions started with the paper by Felsenstein (1985). Felsenstein’s approach was to use the bootstrap method to find the confidence levels that the parameter $\theta$ is in specified regions, based on the sample $\mathcal{X} = \{x_1, x_2, \ldots, x_n\}$. First, $\hat{\theta}$ is calculated from the sample. Then $B$ resamples are taken from $\mathcal{X}$. A resample is produced by sampling from the original sample, with replacement, $B$ times to get $\{x^*_1, x^*_2, \ldots, x^*_n\}$. The resamples may not have all of the initial sample values and some sample values may be in a resample more than once. Then the estimator is computed on each resample to yield $\{\hat{\theta}_1^*, \hat{\theta}_2^*, \ldots, \hat{\theta}_B^*\}$. Once these are calculated it is determined which of the $\hat{\theta}_k^*$ are in the region of interest. Then the confidence level is estimated to be the proportion of $\hat{\theta}_k^*$ that is in the desired region. This proportion is then considered the confidence level of the region. This concept introduced by Felsenstein (1985) was developed in order to
assign confidence levels to phylogenetic trees constructed by genetic sequence data.

However, there were some questions about this method. Some researchers, Hillis and Bull (1993) for example, claimed that Felsenstein's method was biased. The critics were saying that the probability that the bootstrap estimate is in the same region as the parameter is less than the probability that the estimate from the sample is in the same region as the parameter. Efron et al. (1996) explain this through an example. They looked at the estimation of a phylogenetic tree for the malaria parasite _Plasmodium_. The data are from 11 malaria species of the genus _Plasmodium_ and 221 polytypic sites. Thus there is an $11 \times 221$ data matrix. Using Felsenstein's method they resampled the 221 column vectors of length 11. This gives a possible $K = 4^{11} - 4$ column vectors for the sample. Let these vectors be $X_1, X_2, \ldots, X_K$. Each observed column of the data matrix, $X_k$, is an independent selection from $X_1, X_2, \ldots, X_K$ with probabilities $\pi_1, \ldots, \pi_K$, following a multinomial probability model. The data matrix $X$ then can be represented by the proportion of the $n$ columns that equal each possible $X_k$,

$$\hat{\pi}_k = \#\{\text{columns of } X \text{ equaling } X_k\}/n.$$

The possible space of $\pi = (\pi_1, \pi_2, \ldots, \pi_K)$ is divided into regions $\Theta_1, \Theta_2, \ldots$. The probability that $\hat{\pi}_k$ is in each region differs from the probability that

$$\hat{\pi}_k^* = \#\{\text{columns of } X^* \text{ equaling } X_k\}/n$$

is in each region. Thus, $\hat{\pi}_k^*$ has less probability than $\hat{\pi}$ of lying in the same region as $\pi$.

We can demonstrate the bias of Felsenstein's method by another example. Suppose $X \sim N(\theta = 0, 1)$ where $\Theta = \mathbb{R}$. Separate $\mathbb{R}$ into two regions $\Theta_1 = \mathbb{R}^-$ and
\( \Theta_2 = \mathbb{R}^+ \), then let the estimate of \( \theta \) be \( \hat{\theta} = x \) where

\[
P(\hat{\theta} \in \Theta_1) = P(\hat{\theta} \in \Theta_2) = \frac{1}{2}.
\]

However, the bootstrap estimate \( \hat{\theta}^* = X^* \) is generated conditional on the sample \( X \), that is \( X^* \sim N(\hat{\theta}, 1)|X \). Thus,

\[
P(\hat{\theta}^* \in \Theta_1) = P(\hat{\theta}^* \in \Theta_2) = \frac{1}{2}
\]

if and only if \( \hat{\theta} = 0 \), but \( P(\hat{\theta} = 0) = 0 \) since \( \hat{\theta} \sim N(0, 1) \).

Efron et al. (1996) give a simple example to explain why these confidence values are reasonable. Let \( X_1 \) and \( X_2 \) be two independent normal random variables with means \( \mu_1 \) and \( \mu_2 \) and variances equal to 1. That is \( X \sim N_2(\mu, I) \) where \( X = (X_1, X_2) \). The parameter space for \( \mu \), which is the real plane, is partitioned into several regions \( \Theta_1, \Theta_2, \ldots \). We wish to estimate the parameter \( \mu \) and determine which region \( \mu \) belongs to. Suppose that the observed value \( \hat{\mu} \) is in the region \( \Theta_1 \), then a confidence value can be found for the event that the parameter \( \mu \) lies in any region \( \Theta_i \). Felsenstein’s bootstrap confidence value is the bootstrap probability that \( \hat{\mu}^* \) lies in a region \( \Theta_i \), that is,

\[
\tilde{\alpha} = \text{Prob}_{\hat{\mu}} \{ \hat{\mu}^* \in \mathcal{R} \},
\]

where \( \hat{\mu}^* \) is \( \hat{\mu} \) computed on \( X^* \sim N_2(\hat{\mu}, I) \), conditional on \( X \). Efron et al. (1996) state that

We can use a Bayesian model to show that \( \tilde{\alpha} \) is a reasonable assessment of the probability that \( \mathcal{R}_1 \) contains \( \mu \). Suppose we believe apriori that \( \mu \) could lie anywhere in the plane with equal probability. Then having observed \( \hat{\mu} \), the aposteriori distribution of \( \mu \) given \( \hat{\mu} \) is \( N_2(\hat{\mu}, I) \), exactly the same as the bootstrap distribution of \( \hat{\mu}^* \). In other words, \( \tilde{\alpha} \) is the aposteriori probability of the event \( \mu \in \mathcal{R}_1 \), if we begin with an ‘uninformative’ prior density for \( \mu \),(p.13431)
This also applies to the multinomial model for the phylogenetic trees.

Efron et al. (1996) also compare Felsenstein’s bootstrap confidence value with a confidence level motivated by the definition of \( p \)-values in hypothesis testing. There is an approximate formula for converting Felsenstein’s confidence value \( \hat{\alpha} \) to a hypothesis-testing confidence level \( \tilde{\alpha} \). For example, consider two regions \( \Theta_1 \) and \( \Theta_2 \) where \( \hat{\mu} \in \Theta_1 \). Let \( \hat{\mu}_0 \) be the closest point in \( \Theta_2 \) to \( \hat{\mu} \), \( \Phi(z) \) be the cumulative distribution function of a standard normal random variable, \( \tilde{z} = \Phi^{-1}(\tilde{\alpha}) \), and \( \tilde{z} = \Phi^{-1}(\hat{\alpha}) \). Define

\[ z_0 = \Phi^{-1}(\text{Prob}_{\hat{\mu}_0} \{\hat{\mu}^{**} \in \Theta\}) , \]

where \( \hat{\mu}^{**} \) is \( \hat{\mu} \) computed on \( X^{**} \sim N_2(\hat{\mu}_0, I) \), conditional on \( X \). Then \( \tilde{z} \) can be approximated in terms of \( \tilde{z} \) and \( z_0 \) as

\[ \tilde{z} \approx \tilde{z} - 2z_0. \]

Efron and Tibshirani (1998) called Felsenstein’s bootstrap confidence value \( \hat{\alpha} \) a “first-order” bootstrap analysis. We calculate \( \hat{\alpha} \) as the first-order bootstrap method, then obtain \( z_0 \) from \( X^{**} \), which is called the “second-order” bootstrap. The confidence level is then calculated as

\[ \hat{\alpha} \approx \Phi\{\Phi^{-1}(\hat{\alpha}) - 2z_0\} . \]

This measure is second-order accurate to \( \hat{\alpha} \). Efron and Tibshirani (1998) uses this method to calculate the frequentest confidence levels and also the Bayesian measures of confidence. However, the frequentest method works well for two regions but gives unreasonable answers for more than two regions. The Bayesian method heavily relies on the choice of prior.
2.2 Critical Points

To obtain confidence intervals we use critical points that depend on the distribution of $X$. In the case of unknown distributions there is some discussion as to what type of critical points should be used. Hall (1988, 1992) provides the theoretical motivation for using these critical points through Edgeworth expansion theory.

Efron (1981) introduced bootstrap critical points for confidence regions. The following are five critical points that will be used in this research, percentile method ($\hat{\theta}_{\text{back}}$), the percentile-$t$ method ($\hat{\theta}_{\text{stud}}$), the hybrid method ($\hat{\theta}_{\text{hyb}}$), the bias-corrected method ($\hat{\theta}_{\text{bc}}$) and the accelerated bias-corrected method ($\hat{\theta}_{\text{abc}}$). The following description is from Hall (1988). Assume the following distribution functions are known:

\[
G(t) = P\left\{ \frac{n^{1/2}(\hat{\theta} - \theta)}{\sigma} \leq t \right\},
\]

and

\[
H(t) = P\left\{ \frac{n^{1/2}(\hat{\theta} - \theta)}{\hat{\sigma}} \leq t \right\},
\]

where the asymptotic variance is

\[
\sigma^2 = \lim_{n \to \infty} V(n^{1/2}\hat{\theta})
\]

and $\hat{\sigma}$ is the estimate of $\sigma$. Let $g_\alpha = G^{-1}(\alpha)$ and $h_\alpha = H^{-1}(\alpha)$ be the $\alpha$-level quantiles of $G$ and $H$, respectively, and assume $G$ and $H$ are continuous.

Hall (1988) defines the different critical points $\hat{\theta}(\alpha)$ with the property

\[
P\{\theta \leq \hat{\theta}(\alpha)\} \cong \alpha.
\]

If $\sigma$ is known, the ordinary critical point can be used:

\[
\hat{\theta}_{\text{ord}}(\alpha) = \hat{\theta} - n^{-1/2}\sigma g_{1-\alpha}.
\]
If \( \sigma \) is unknown, the Studentized critical point is

\[
\hat{\theta}_{\text{stud}}(\alpha) = \hat{\theta} - n^{-1/2} \hat{\sigma} h_{1-\alpha}.
\]

These points are exact because

\[
P\{\theta \leq \hat{\theta}_{\text{ord}}(\alpha)\} = P\{\theta \leq \hat{\theta}_{\text{stud}}(\alpha)\} = \alpha.
\]

If the quantiles \( g_{1-\alpha} \) and \( h_{1-\alpha} \) are mixed up, instead of using the Studentized critical point, the hybrid critical point is

\[
\hat{\theta}_{\text{hyb}}(\alpha) = \hat{\theta} - n^{-1/2} \hat{\sigma} g_{1-\alpha}.
\]

This is similar to looking up the normal tables instead of the Student’s \( t \) table. Also, if we look at the tables backward we would confuse \( h_{1-\alpha} \) with \( -g_{\alpha} \) and get the backwards critical point:

\[
\hat{\theta}_{\text{back}}(\alpha) = \hat{\theta} + n^{-1/2} \hat{\sigma} g_{\alpha}.
\]

Therefore, \( \hat{\theta}_{\text{back}} \) is the result of looking at the wrong tables backwards. There might be some error in using \( \hat{\theta}_{\text{back}}(\alpha) \) to construct confidence intervals. However, we might be able to reduce some of these errors by using another choice of \( \alpha \), such as \( \beta \in (0, 1) \), such that \( -g_{\beta} \simeq h_{1-\alpha} \). This is called a bias-corrected critical point. Using Edgeworth expansions, the theoretical bias-corrected critical point is

\[
\hat{\theta}_{\text{bc}}(\alpha) = \hat{\theta}_{\text{back}}(\beta) = \hat{\theta} + n^{-1/2} \hat{\sigma} \left[ z_{\alpha} + n^{-1/2} \left\{ 2p_1(0) - p_1(z_{\alpha}) \right\} + O(n^{-1}) \right]
\]

and the theoretical accelerated bias-corrected critical point is

\[
\hat{\theta}_{\text{abc}}(\alpha) = \hat{\theta}_{\text{back}}(\beta) = \hat{\theta} + n^{-1/2} \hat{\sigma} \left\{ z_{\alpha} + n^{-1/2} q_1(z_{\alpha}) + O(n^{-1}) \right\},
\]
where $p_1$ and $q_1$ are polynomials whose coefficients are functions of the moments of $F$. Note that $\hat{\theta}_{abc}(\alpha) = \hat{\theta}_{\text{stud}}(\alpha) + O(n^{-3/2})$. Efron (1985) describes the bias-corrected bootstrap method where the distribution function $F$ of the sample $X$ is multivariate normal.

The bootstrap estimates of the distribution functions $G$ and $H$ from Hall (1988) are

$$
\hat{G}(t) = P \left\{ \frac{n^{1/2}(\hat{\theta}^* - \hat{\theta})}{\hat{\sigma}} \leq t \ \big| \ X \right\}
$$

and

$$
\hat{H}(t) = P \left\{ \frac{n^{1/2}(\hat{\theta}^* - \hat{\theta})}{\hat{\sigma}^*} \leq t \ \big| \ X \right\}.
$$

The bootstrap estimates of $g_\alpha$ and $h_\alpha$ are $\hat{g}_\alpha = \hat{G}^{-1}(\alpha)$ and $\hat{h}_\alpha = \hat{G}^{-1}(\alpha)$. Thus the bootstrap versions of the critical points are obtained by replacing $g_\alpha$ and $h_\alpha$ with their bootstrap estimates. The bootstrap critical points are

$$
\hat{\theta}_{\text{ord}}^*(\alpha) = \hat{\theta} - n^{-1/2} \hat{\sigma}\hat{g}_{1-\alpha}, \quad \hat{\theta}_{\text{stud}}^*(\alpha) = \hat{\theta} - n^{-1/2} \hat{\sigma}\hat{h}_{1-\alpha},
$$

$$
\hat{\theta}_{\text{hyb}}^*(\alpha) = \hat{\theta} - n^{-1/2} \hat{\sigma}\hat{g}_{1-\alpha}, \quad \hat{\theta}_{\text{back}}^*(\alpha) = \hat{\theta} + n^{-1/2} \hat{\sigma}\hat{g}_{1-\alpha}.
$$

Hall (1988) uses Edgeworth expansion theory to compare the properties of these bootstrap critical values. When $\sigma$ is unknown, these critical values are $k^{th}$-order correct if the critical value is equivalent to $\hat{\theta}_{\text{stud}}(\alpha) + O(n^{-(k+1)/2})$. Using the Edgeworth expansions and Cornish-Fisher inversion, Hall (1988) found that $\hat{\theta}_{\text{stud}}$ and $\hat{\theta}_{abc}$ are second-order correct and that $\hat{\theta}_{\text{hyb}}$, $\hat{\theta}_{\text{back}}$ and $\hat{\theta}_{abc}$ are usually only first-order correct.
2.3 Edgeworth Expansion

In this section we will give the general idea of the Edgeworth expansions from Hall (1988, 1992) and Bhattacharya and Ghosh (1978). The calculations of Edgeworth expansions will assume the following given by Hall (1988, 1992). Let the sample \( \mathcal{X} = \{X_1, X_2, \ldots, X_n\} \) be independent and identically distributed \( d \)-dimensional vectors with mean \( \mu \) and put \( \tilde{X} = n^{-1} \sum_{i=1}^{n} X_i \). Define \( \theta = f(\mu) \) and \( \sigma^2 = g(\mu) \), where \( f \) and \( g \) are known real-valued smooth functions. Then estimates of \( \theta \) and \( \sigma^2 \) are \( \hat{\theta} = f(\tilde{X}) \) and \( \hat{\sigma}^2 = g(\tilde{X}) \), respectively. Let \( f_{(i_1 \ldots i_p)} = (\partial^{p}/\partial x^{(i_1)} \ldots \partial^{(i_p)}) f(x) \), \( a_{i_1 \ldots i_p} = f_{(i_1 \ldots i_p)}(\mu) \), \( \mu_{i_1 \ldots i_p} = E\{(X - \mu)^{(i_1)} \ldots (X - \mu)^{(i_p)}\} \), \( c_i = g_{(i)}(\mu) \) and \( A(x) = f(x) - f(\mu) \). Then let \( A: \mathbb{R}^d \rightarrow \mathbb{R} \) be a smooth function satisfying \( A(\mu) = 0 \). Given \( U \equiv n^{-1/2}A(\tilde{X}) \), then the cumulants of \( U \) are

\[
\begin{align*}
k_1(U) & \equiv E(U) = n^{-1/2}A_1 + O(n^{-3/2}), \\
k_2(U) & \equiv E(U^2) - [E(U)]^2 = \sigma^2 + O(n^{-1}),
\end{align*}
\]

and

\[
k_3(U) \equiv E(U^3) - 3E(U^2)E(U) + 2[E(U)]^3 = n^{-1/2}A_2 + O(n^{-3/2}),
\]

where \( a_{i_1 \ldots i_p} = A_{(i_1 \ldots i_p)}(\mu) \), \( \sigma^2 \equiv \sum \sum a_i a_j \mu_{ij} \), \( A_1 \equiv \frac{1}{2} \sum \sum a_{ij} \mu_{ij} \) and

\[
A_2 \equiv \sum \sum \sum a_i a_j a_k \mu_{ijk} + 3 \sum \sum \sum \sum a_i a_j a_k \mu_{ijk} \mu_{jl}.
\]

The Edgeworth expansions for the distributions of \( U/\sigma \) and \( U/\hat{\sigma} \) are given by

\[
P(U/\sigma \leq x) = \Phi(x) + \sum_{i=1}^{\nu} n^{-i/2} p_i(x) \phi(x) + O(n^{-(\nu+1)/2})
\]

and

\[
P(U/\hat{\sigma} \leq x) = \Phi(x) + \sum_{i=1}^{\nu} n^{-i/2} q_i(x) \phi(x) + O(n^{-(\nu+1)/2}),
\]

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
respectively, for some $\nu \leq 1$ where $\phi(x) = \Phi'(x)$, $\Phi(x)$ is the standard normal distribution function, and $p_i$ and $q_i$ are polynomials of degree $3i - 1$. The polynomials $p_i$ and $q_i$ are odd for even $i$ and even for odd $i$. Also, $-p_1(x) = \sigma^{-1}A_1 + \frac{1}{6}\sigma^{-3}A_2(x^2 - 1)$ and $p_1(x) - q_1(x) = -\frac{1}{2}\sigma^{-3}(\sum a_i c_{ij} \mu_j x^2)$.

Define $x_\alpha$, $y_\alpha$ and $z_\alpha$ by

$$P(U/\sigma \leq x_\alpha) = P(U/\hat{\sigma} \leq y_\alpha) = \Phi(z_\alpha) = \alpha.$$  

The inversion of the Edgeworth expansions are the Cornish-Fisher expansions defined as

$$x_\alpha = z_\alpha + \sum_{i=1}^{\nu} n^{-i/2}p_{ii}(z_\alpha) + O(n^{-(\nu+1)/2})$$

and

$$y_\alpha = z_\alpha + \sum_{i=1}^{\nu} n^{-i/2}q_{ii}(z_\alpha) + O(n^{-(\nu+1)/2}),$$

where the polynomials $p_{ii}$ and $q_{ii}$ are of degree $i + 1$ and are related to the polynomials $p_i$ and $q_i$, respectively. For example,

$$p_{11}(x) = -p_1(x)$$

and

$$p_{21}(x) = p_1(x)p_1'(x) - \frac{1}{2}xp_1(x)^2 - p_2(x).$$

There are similar results for $q_{11}$ and $q_{21}$.

2.4 Observed Confidence Levels

Once the critical values are known then we can determine the observed confidence levels of the regions desired. The method for calculating the observed confidence levels is based on the research by Polansky (2003 a,b). Let $C(\omega_1, \ldots, \omega_n)$ be a 100\%
confidence region for the parameter $\theta$ based on the sample $X$. The parameters $\omega_i$, $i = 1, \ldots, q$, control the confidence level, location, shape, and orientation of the confidence region. Suppose there exist $\alpha_i \in (0, 1)$ and $\omega_{i1}, \ldots, \omega_{iq} \in \Omega$ such that $C(\omega_{i1}, \ldots, \omega_{iq}) = \Theta_i$ for a specified parameter region $\Theta_i$ for $i = 1, 2, \ldots$, conditional on the sample $X$. Then $\alpha_i$ is the observed confidence level for the region $\Theta_i$.

First, we will look at the observed confidence level for a scalar parameter that was introduced by Polansky (2005a). The development of the observed confidence levels is computed using the theoretical critical points defined by Hall (1988) given in Section 2.2. Polansky (2005a) develops a two-sided confidence interval from these critical points. Let $\omega_1 \in (0, 1)$ and $\omega_2 \in (0, 1)$ such that $\omega_2 - \omega_1 \in (0, 1)$. Then a $100(\omega_2 - \omega_1)\%$ confidence interval for $\theta$ when $\sigma$ is known is $C_{\text{ord}}(\omega_1, \omega_2) = \left[\hat{\theta}_{\text{ord}}(\omega_1), \hat{\theta}_{\text{ord}}(\omega_2)\right]$. In the same way, the critical points $\hat{\theta}_{\text{stud}}, \hat{\theta}_{\text{hyb}}$ and $\hat{\theta}_{\text{back}}$ yield the confidence intervals $C_{\text{stud}}(\omega_1, \omega_2)$, $C_{\text{hyb}}(\omega_1, \omega_2)$ and $C_{\text{back}}(\omega_1, \omega_2)$, respectively.

Consider an interval region $\Theta_k = (t_L, t_U)$ where $t_L < t_U$. To obtain the observed confidence level for this region using the ordinary critical point, let

$$t_L = \hat{\theta} - n^{-1/2} \sigma g_{1-\omega_1}$$

and

$$t_U = \hat{\theta} - n^{-1/2} \sigma g_{1-\omega_2}.$$

Solving for $\omega_1$ and $\omega_2$ then setting $\alpha_{\text{ord}}(\Theta_k) = \omega_2 - \omega_1$ yields

$$\alpha_{\text{ord}}(\Theta_k) = G_n \left[\frac{n^{1/2}(\hat{\theta} - t_L)}{\sigma}\right] - G_n \left[\frac{n^{1/2}(\hat{\theta} - t_U)}{\sigma}\right],$$

where $G_n$ and $g_{\alpha}$ are as defined in Section 2.2. The observed confidence levels corresponding to the Studentized, hybrid and backwards theoretical points are derived in a similar way.
If the distribution $F$ is unknown, Polansky (2005a) shows how to use the normal approximation to find these observed critical values. This follows from the Central Limit Theorem. Assuming a smooth function $\sigma^2 > 0$ and $\frac{\partial}{\partial \mu} g(\mu) \neq 0$ we have

$$\frac{n^{1/2}(\hat{\theta} - \theta)}{\sigma} \xrightarrow{w} Z$$

and

$$\frac{n^{1/2}(\hat{\theta} - \theta)}{\hat{\sigma}} \xrightarrow{w} Z,$$

as $n \to \infty$, where $Z$ is the standard normal random variable. This implies that $G_n \approx \Phi(t)$ and $H_n \approx \Phi(t)$ for large $n$. Using the Studentized critical point, the normal approximation for the critical level to compute the observed confidence level of $\Theta_k$ yields

$$\hat{\alpha}_{\text{stud}}(\Theta_k) = \Phi \left[ \frac{n^{1/2}(\hat{\theta} - t_L)}{\hat{\sigma}} \right] - \Phi \left[ \frac{n^{1/2}(\hat{\theta} - t_U)}{\hat{\sigma}} \right].$$

If the normal distribution is not an accurate approximation to $G_n$, the bootstrap method can be used. Consider a resample $X^* = \{X^*_1, \ldots, X^*_n\}$, conditional on the original sample $X$, where the distribution of $X^*$ is $\tilde{F}$ based on the original sample $X$. Let $\hat{\theta}^*$ be the estimate of the parameter computed on $X^*$ and $\hat{\sigma}^*$ be the estimated standard error of the parameter $\hat{\theta}^*$ computed on the sample $X^*$. Thus the bootstrap estimate of $G_n$ and $H_n$ is given by

$$\hat{G}_n(t) = P \left[ \frac{n^{1/2}(\hat{\theta}^* - \hat{\theta})}{\hat{\sigma}} \leq t \left| X \right. \right]$$

and

$$\hat{H}_n(t) = P \left[ \frac{n^{1/2}(\hat{\theta}^* - \hat{\theta})}{\hat{\sigma}^*} \leq t \left| X \right. \right].$$
Then the estimated observed confidence levels based on the bootstrap ordinary theoretical point is

$$\hat{\alpha}_{ord}^* (\Theta_k) = \hat{G}_n \left[ \frac{n^{1/2}(\hat{\theta} - t_L)}{\sigma} \right] - \hat{G}_n \left[ \frac{n^{1/2}(\hat{\theta} - t_U)}{\sigma} \right].$$

Using the bootstrap Studentized, hybrid and backwards critical points would result in similar expressions.

Polansky (2005a) gives an order of accuracy for these observed confidence levels. Suppose a region $\Theta_k$ corresponds to a 100$\alpha$% confidence interval for $\theta$; the observed confidence level then should be $\alpha$. Let $C(\omega_1, \omega_2)$ be a confidence interval for $\theta$. Then a measure of the observed confidence level, $\hat{\alpha}$, is considered accurate if $\hat{\alpha} [C(\omega_1, \omega_2)] = \alpha$, whenever $\alpha = \omega_2 - \omega_1$. This measure $\hat{\alpha}$ is said to be $k^{th}$ order accurate if $\hat{\alpha} [C(\omega_1, \omega_2)] = \alpha + O(n^{-k/2})$. This order of accuracy is derived from the Edgeworth expansion method. Using the normal approximations the confidence level $\alpha_{hyb}$ yields first-order accuracy. Similarly, the confidence level $\alpha_{back}$ also yields first-order accuracy. When using the bootstrap approximation when $G_n$ and $H_n$ are unknown, the Student confidence level $\alpha_{stud}$ is second-order accurate and $\alpha_{hyb}$ and $\alpha_{back}$ are both first-order accurate.

There are two other critical points that were previously mentioned. These are the bias-corrected critical point $\theta_{bc}$ and the accelerated bias-corrected critical point $\theta_{abc}$. The observed confidence level based on the bias-corrected critical point from Polansky (2005a) is

$$\alpha_{bc} (\Theta_k) = \Phi \left\{ \Phi^{-1} \left[ G_n \left( \frac{n^{1/2}(t_U - \hat{\theta})}{\hat{\sigma}} \right) \right] - \delta_{bc} \right\} \quad \Phi \left\{ \Phi^{-1} \left[ G_n \left( \frac{n^{1/2}(t_L - \hat{\theta})}{\hat{\sigma}} \right) \right] - \delta_{bc} \right\}.$$
where $\delta_{bc} = 2\Phi^{-1} [G_n(0)]$ is the bias correction. Then the observed confidence level based on the accelerated bias-corrected critical point is

$$
\alpha_{abc} (\Theta_k) = \Phi \left\{ \frac{\Phi^{-1} \left[ G_n \left( \frac{n^{1/2} (u - \hat{\theta})}{\hat{\sigma}} \right) \right] - \frac{1}{2} \delta_{bc}}{1 + a \Phi^{-1} \left[ G_n \left( \frac{n^{1/2} (u - \hat{\theta})}{\hat{\sigma}} \right) \right] - \frac{1}{2} \delta_{bc}} - \frac{1}{2} \delta_{bc} \right\},
$$

where $a$ is the acceleration coefficient and is defined by Hall (1988) to be

$$
a = n^{-1/2} z^2_\alpha [p_1(z_\alpha) + q_1(z_\alpha) - 2p_1(0)].
$$

Polansky (2005a) shows that $\alpha_{bc}$ is first-order accurate and that $\alpha_{abc}$ is second-order accurate.

The previous methods were for scalar parameters. Polansky (2005b) derives observed confidence levels for vector parameters $\theta$ where $X$ is a set of random vectors. Let $\Sigma$ be the asymptotic covariance of $n^{1/2} \hat{\theta}$ such that

$$
\Sigma = \lim n \to \infty V(n^{1/2} \hat{\theta}).
$$

Define $G_n$ and $H_n$ as

$$
G_n(t) = P[n^{1/2} \Sigma^{-1/2} (\hat{\theta} - \theta) \leq t \bigg| X \sim F],
$$

and

$$
H_n(t) = P[n^{1/2} \Sigma^{-1/2} (\hat{\theta} - \theta) \leq t \bigg| X \sim F],
$$

and let $g_n(t)$ and $h_n(t)$ be the $p$-dimensional densities corresponding to $G_n(t)$ and $H_n(t)$, respectively. Define $G_\alpha$ to be any region in $\mathbb{R}^p$ that satisfies

$$
P[n^{1/2} \Sigma^{-1/2} (\hat{\theta} - \theta) \in G_\alpha \big| X \sim F] = \alpha.$$
and $\mathcal{H}_\alpha$ to be a region of $\mathbb{R}^p$ that satisfies

$$P[n^{1/2} \hat{\Sigma}^{-1/2} (\hat{\theta} - \theta) \in \mathcal{H}_\alpha | \mathcal{X} \sim F] = \alpha,$$

where $\hat{\Sigma}$ is the estimated covariance of $\theta$ based on the sample $\mathcal{X}$. The multivariate analog of the ordinary confidence interval is

$$C_{\text{ord}}(\alpha) = \{\hat{\theta} - n^{-1/2} \Sigma^{1/2} t : t \in G_\alpha\},$$

and when $\Sigma$ is unknown the multivariate analog of the Studentized confidence interval is given by

$$C_{\text{stud}}(\alpha) = \{\hat{\theta} - n^{-1/2} \hat{\Sigma}^{1/2} t : t \in \mathcal{H}_\alpha\}.$$

Then an observed confidence level for $\Theta_k$, given that the density $g_n(t)$ corresponding to $G_n(t)$ is known, yields

$$\alpha_{\text{ord}}(\Theta_k) = \int_{n^{1/2} \Sigma^{-1/2} (\hat{\theta} - \theta_k)} g_n(t) dt.$$

If $g_n(t)$ is unknown then we can approximate it with a $p$-variate standard normal density. Similar results are given for the other confidence levels using the Student and hybrid critical points.

If $g_n(t)$ is unknown and the normal approximation is not considered accurate enough, Polansky (2005b) proposes using the bootstrap method to compute the observed critical values. Let $\hat{\theta}^*$ be the estimate of the parameter $\theta$ based on the $b$ resamples $\mathcal{X}_i^*$, $i = 1, \ldots, b$. Then the bootstrap estimates of $G_n$ and $H_n$ are

$$\hat{G}_n(t) = P[n^{1/2} \hat{\Sigma}^{-1/2} (\hat{\theta}^* - \hat{\theta}) \leq t| \mathcal{X}^* \sim \hat{F}, \mathcal{X}]$$

and

$$\hat{H}_n(t) = P[n^{1/2} \hat{\Sigma}^{-1/2} (\hat{\theta}^* - \hat{\theta}) \leq t| \mathcal{X}^* \sim \hat{F}, \mathcal{X}],$$
where $\hat{\Sigma}^*$ is the asymptotic covariance matrix computed on the resample $X^*$. Then the bootstrap estimates of $\alpha_{\text{ord}}(\Theta_k)$ and $\alpha_{\text{stud}}(\Theta_k)$ are

$$
\hat{\alpha}_{\text{ord}}^*(\Theta_k) = \int_{R^{n_1/2}\Sigma^{-1/2}(\hat{\theta} - \Theta_k)} \hat{g}_n(t)dt,
$$

$$
\hat{\alpha}_{\text{stud}}^*(\Theta_k) = \int_{R^{n_1/2}\Sigma^{-1/2}(\hat{\theta} - \Theta_k)} \hat{h}_n(t)dt,
$$

and

$$
\hat{\alpha}_{\text{hyb}}^*(\Theta_k) = \int_{R^{n_1/2}\Sigma^{-1/2}(\hat{\theta} - \Theta_k)} \hat{g}_n(t)dt.
$$

An additional measure is based on the backwards (percentile) method introduced by Efron and Tibshirani (1998). Given $V_n(t) = P(\hat{\theta} \leq t | X \sim F)$ with a density (or mass function) $v_n(t)$, the bootstrap estimate is $\hat{V}_n(t) = P(\hat{\theta}^* \leq t | X^* \sim \hat{F}, X)$ with density $\hat{v}_n(t)$. Then a confidence region based on the bootstrap percentile method is given by any region $V_\alpha$ such that

$$
\int_{V_\alpha} \hat{v}_n(t)dt = \alpha.
$$

Setting $V_\alpha = \Theta_k$ we get the percentile observed confidence level,

$$
\hat{\alpha}_{\text{perc}}^*(\Theta_k) = \int_{\Theta_k} \hat{v}_n(t)dt.
$$

The bootstrap estimates of $g_n(t)$ and $h_n(t)$ may not exist in a closed form. To approximate these estimates, let $X_i^* \sim F$ conditional on $X$ for $i = 1, \ldots, b$ and let $\hat{\theta}_i^*$ be $\hat{\theta}$ computed on $X_i^*$. Then $g_n(t)$ and $h_n(t)$ can be approximated by

$$
\hat{g}_n(t) \approx b^{-1} \sum_{i=1}^b 1 \left[ n^{1/2} \hat{\Sigma}^* \left( \hat{\theta}_i^* - \hat{\theta} \right) = t \right]
$$

and

$$
\hat{h}_n(t) \approx b^{-1} \sum_{i=1}^b 1 \left[ n^{1/2} \hat{\Sigma}^* \left( \hat{\theta}_i^* - \hat{\theta} \right) = t \right].
$$
Let $I$ be the indicator function. The following are approximations of the observed confidence levels for $\Theta_k$:

$$
\hat{\alpha}_{\text{ord}}^* (\Theta_k) \approx b^{-1} \sum_{i=1}^{b} I \left[ \hat{\theta} - \Sigma^{1/2} \hat{\Sigma}^{-1/2} \left( \hat{\theta}_i - \hat{\theta} \right) \in \Theta_k \right],
$$

$$
\hat{\alpha}_{\text{hyb}}^* (\Theta_k) \approx b^{-1} \sum_{i=1}^{b} I \left[ 2 \hat{\theta} - \hat{\theta}_i \in \Theta_k \right],
$$

$$
\hat{\alpha}_{\text{perc}}^* (\Theta_k) \approx b^{-1} \sum_{i=1}^{b} I (\hat{\theta}_i \in \Theta_k),
$$

and

$$
\hat{\alpha}_{\text{stud}}^* (\Theta_k) \approx b^{-1} \sum_{i=1}^{b} I \left[ \hat{\theta} - \Sigma^{1/2} \hat{\Sigma}^{-1/2} \left( \hat{\theta}_i - \hat{\theta} \right) \in \Theta_k \right].
$$

Polansky (2005b) defines a measure of accuracy for these observed confidence levels which is similar to the univariate case. Using the Edgeworth expansion theory from Hall (1992), Polansky shows that $\hat{\alpha}_{\text{ord}}^*$ and $\hat{\alpha}_{\text{stud}}^*$ are second-order accurate in probability. The hybrid bootstrap expansion $\hat{\alpha}_{\text{hyb}}^*$, the percentile $\hat{\alpha}_{\text{perc}}^*$ and the normal approximation $\hat{\alpha}_{\text{stud}}^*$ are all first-order accurate in probability.

2.5 Regression

This section reviews some of the literature that uses the bootstrap method with regression models. Wu (1986a) demonstrates the effect of different resampling methods in the context of regression models. Liu (1988) shows a weighted variance estimate of the regression parameters. Hall (1992) develops the accuracy of the critical points for the regression parameters.

Wu (1986a) first considers the linear regression model $Y = \beta X + \varepsilon$. Define $\beta$ as a $k \times 1$ vector of parameters and $\varepsilon$ as an $n \times 1$ vector of errors with mean zero and covariance matrix $\Sigma$. The least squares estimate of $\beta$ is $\hat{\beta} = (X'X)^{-1} X'Y$, 

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
where $X' = (x_1, \ldots, x_n)$ and $Y' = (y_1, \ldots, y_n)$. One proposal is to resample the pairs $z_i = (x_i, y_i)$ for $i = 1, \ldots, n$.

Wu (1986a) defines resampling as follows. A resample of $z_1 = (x_1, y_1), \ldots, z_n = (x_n, y_n)$ is a reweighted version of $(z_i)^n$ with weight $P_i^* > 0$, where the vector $P^* = (P_1^*, \ldots, P_n^*)$ is called a resampling vector. For each $P^*$, the corresponding least squares estimate $\beta^*$ is based on the $P_i^*$ copies of the pair $(x_i, y_i)$. That is,

$$\beta^* = (X'D^*X)^{-1} X'D^*Y,$$

where $D^* = \text{diag}(P_1^*, \ldots, P_n^*)$. In the case of a bootstrap sample, $P^* = (P_1^*, \ldots, P_n^*)$ has a multinomial distribution with $P_i^* = 1/n$, where * indicates that the random mechanism in the association is computed with respect to the joint distribution of $(P_i^*)^n$ based on a resampling procedure, conditional on $X$ and $Y$. Let the expectation under this distribution be represented by $E_\ast$. Then Wu (1986a) suggests the following assumptions for these resampling procedures:

1. $E_\ast \left( \prod_{j=1}^{k} P_{i_j}^* \right) = a_k > 0$ is independent of the subset $(i_1, \ldots, i_k)$, where $k$ = number of unknown parameters in the regression model.

2. The $n$ random variables $\{P_i^*\}_{i=1}^{n}$ are exchangeable and

$$\text{Prob}_\ast(\text{support size of } P^* \geq k) > 0,$$

where the support size of $P^*$ is the total number of $i$'s with $P_i^* > 0$.

One of the resampling methods that meets these assumptions is the standard bootstrap resampling method. Wu (1986a) describes this bootstrap method as drawing an independent and identically distributed sample $(x_1^*, y_1^*), \ldots, (x_n^*, y_n^*)$ from the
original sample \((x_i, y_i), i = 1, \ldots, n\) with replacement. In this case, the resampling vector \(P^*\) has a multinomial distribution. Because Assumption 1 is satisfied for this method of resampling and Wu (1986a) shows that if \(X' D^* X\) is nonsingular then

\[
\hat{\beta} = \frac{E_* |X' D^* X| \beta^*}{E_* |X' D^* X|}
\]

and \(|X' D^* X|\) is defined to be 0 if \(X' D^* X\) is singular, where \(\hat{\beta}\) is a reweighted average for \(\beta^*\).

Wu (1986a) derives a bootstrap variance estimator for \(\beta\). The unweighted \(v^*\) and weighted bootstrap variance \(v^*\) estimators are

\[
v^* = E_* (\beta^* - \hat{\beta})(\beta^* - \hat{\beta})',
\]

\[
v_{w}^* = E_* |X' D^* X| (\beta^* - \hat{\beta})(\beta^* - \hat{\beta})'/E_* |X' D^* X|,
\]

where

\[
\beta^* = \left(\sum_{i=1}^{n} x_i x_i'\right)^{-1} \sum_{i=1}^{n} x_i y_i^*.
\]

Wu (1986a) points out that these variance estimators are generally biased. The reason for this is because the least squares estimator \(\beta^*\) is not exchangeable. However, Wu (1986a) states that \(v^*\) may be robust against error variance heteroscedasticity.

Because of the so-called bias in the bootstrap, Wu (1986a) also presents the jackknife variance estimator for \(\beta\) by retaining \(r\) samples:

\[
\tilde{v}_{J,r}(\hat{\beta}) = \sum_r w_s (\tilde{\beta}_s - \hat{\beta})(\tilde{\beta}_s - \hat{\beta}),
\]

where \(w_s \propto |X_s' X_s|\), \(\sum_r w_s = 1\) and

\[
\tilde{\beta}_s = \hat{\beta} + \left(\frac{r - k + 1}{n - r}\right)^{1/2} (\tilde{\beta}_s - \hat{\beta}).
\]
Wu (1986a) uses these variance estimators to derive the following confidence intervals for $\beta$ where $t_\alpha$ is the upper $\alpha$ percentage point of a $t$-distribution with $n - 3$ degrees of freedom:

Delete-1 jackknife: $\hat{\beta} \pm t_\alpha \sqrt{v_{t,n-1}(\hat{\beta})}$,

Retain-8 jackknife: $\hat{\beta} \pm t_\alpha \sqrt{v_{8}(\hat{\beta})}$,

Bootstrap variance: $\hat{\beta} \pm t_\alpha \sqrt{v_0}$,

and

Bootstrap percentile: $[\text{CDFB}^{-1}(\alpha), \text{CDFB}^{-1}(1 - \alpha)]$.

Liu (1988) describes a bootstrap estimate for $\text{Var}(\hat{\beta})$ for a simple linear model $Y_i = \beta x_i + \epsilon_i$, where $\hat{\beta}$ is the least squares estimate of $\beta$ and

$$\text{Var}(\hat{\beta}) = \frac{\sum_{i=1}^{n} x_i^2 \sigma_i^2}{\left(\sum_{j=1}^{n} x_j^2 \right)^2}.$$  

Let $r_i = Y_i - x_i \hat{\beta}$ be the residuals, then a bootstrap sample is $Y_i^* = x_i \hat{\beta} + r_i^*$ where $r_1^*, \ldots, r_n^*$ is a random sample from the empirical distribution function of $(r_1 - \bar{r}_n), \ldots, (r_n - \bar{r}_n)$. Then the bootstrap estimate of $\beta$ is $\hat{\beta}^*$ based on the $Y_i^*$'s. Thus the bootstrap variance of $\hat{\beta}^*$ is

$$\text{Var}(\hat{\beta}^*) = \frac{n^{-1} \sum_{i=1}^{n} (r_i - \bar{r}_n)^2}{\sum_{j=1}^{n} x_j^2},$$

which is asymptotically equal to $\frac{\sum_{i=1}^{n} x_i^2 \sigma_i^2}{\left(\sum_{j=1}^{n} x_j^2 \right)^2}$. Thus this is not a consistent estimator for the standard error of $\hat{\beta}$. Liu (1988) suggests that instead of taking a random sample from the empirical distribution function of $(r_1 - \bar{r}_n), \ldots, (r_n - \bar{r}_n)$, the empirical
distribution function of
\[ \left\{ \frac{x_i}{\bar{x}_n} \right\}^{1/2} (r_i - \bar{r}_n) \]
is used. This results in the variance of this bootstrap estimate to be
\[
\text{Var}(\hat{\beta}^*) = \frac{\sum_{i=1}^{n} x_i^2 r_i^2}{\left( \sum_{j=1}^{n} x_j^2 \right)^2} - \frac{n^{-1} \left( \sum_{i=1}^{n} x_i^2 r_i^2 \right)^2}{\left( \sum_{j=1}^{n} x_j^2 \right)^2} = \frac{\sum_{i=1}^{n} x_i^2 \epsilon_i^2}{\left( \sum_{j=1}^{n} x_j^2 \right)^2} + O_p \left( n^{-3/2} \right).
\]

Consider a linear multivariate regression model of the form
\[ Y_i = \beta_0 + x_i \beta + \epsilon_i, \]
where \( Y_i, \beta_0, \) and \( \epsilon_i \) are \( q \times 1 \) vectors; \( \beta \) is a \( p \times 1 \) vector; and \( x_i \) is a \( q \times p \) matrix.

We will first look at the confidence intervals for the slope parameter \( \beta \), then look at the intercept parameter \( \beta_0 \).

Freedman (1981) makes a distinction between regression and correlation models, for the regression model \( F \) is the distribution function of \( \epsilon_i \). We assume that \( F \) has mean 0 and finite unknown variance \( \sigma^2 \). Using the bootstrap idea we replace the unknown distribution function \( F \) by the empirical distribution function \( \hat{F} \). Then \( \hat{F} \) is the empirical distribution function for the residuals
\[ \hat{\epsilon}_i = Y_i - \hat{\beta}_0 - x_i \hat{\beta}, \]
for \( i = 1, \ldots, n \), where \( \hat{\beta}_0 \) and \( \hat{\beta} \) are the least squares estimates of \( \beta_0 \) and \( \beta \). Thus, in the case of the multivariate linear regression model, the residuals \( \epsilon_i \) are resampled as suggested by Liu (1988). Whereas in the case of the correlation model \( F \) is the distribution function of \( (X_i, Y_i), i = 1, \ldots, n \), so in this case the pairs \( (X_i, Y_i) \) are resampled.
Hall (1992) determines that confidence intervals for the intercept of the regression model and regression mean have the same coverage error of at most $O(n^{-1})$. However, in the case of slope parameters the coverage error for confidence intervals are of order of at most $O(n^{-2})$. Hall (1992) discusses in detail the bootstrap estimation of the slope parameter for the regression model. We will first look at the arguments that Hall (1992) made for the bootstrap estimate of the slope parameter for the regression model. Define

$$Y_i^* = \hat{\beta}_0 + x_i \hat{\beta} + \hat{\epsilon}^*$$

for $i = 1, \ldots, n$ as the bootstrap realization of $Y_i^*$ conditional on

$$\mathcal{X} = \{(x_1, Y_1), \ldots, (x_n, Y_n)\}.$$

First consider the case of simple linear regression. Let $p = q = 1$ so that

$$\hat{\epsilon}_i = Y_i - \bar{Y} - (x_i - \bar{x}) \hat{\beta}_1,$$

$$\bar{x} = n^{-1} \sum x_i,$$

$$\hat{\sigma}^2 = n^{-1} \sum \hat{\epsilon}_i^2,$$

and

$$\sigma_x^2 = n^{-1} \sum (x_i - \bar{x})^2.$$

Let $\theta = \beta_1$; thus the slope estimate in this case is

$$\hat{\theta} = \sigma_x^{-2} n^{-1} \sum (x_i - \bar{x}) (Y_i - \bar{Y})$$

and the bootstrap estimate of the slope parameter is

$$\hat{\theta}^* = \sigma_x^{-2} n^{-1} \sum (x_i - \bar{x}) (Y_i^* - \bar{Y}^*).$$
Notice that \( \text{Var}(\hat{\theta}) = n^{-1/2}\sigma_x^{-2}\sigma^2 \).

Hall (1992) defines \( g_\alpha \) and \( h_\alpha \) as quantiles of the distributions of
\[
\frac{n^{1/2}(\hat{\theta} - \theta)\sigma_x}{\sigma}
\]
and
\[
\frac{n^{1/2}(\hat{\theta} - \theta)\sigma_x}{\hat{\sigma}},
\]
so that
\[
P\left\{ \frac{n^{1/2}(\hat{\theta} - \theta)\sigma_x}{\sigma} \leq g_\alpha \right\} = P\left\{ \frac{n^{1/2}(\hat{\theta} - \theta)\sigma_x}{\hat{\sigma}} \leq h_\alpha \right\} = \alpha,
\]
and their bootstrap estimates \( \hat{g}_\alpha \) and \( \hat{h}_\alpha \) so that
\[
P\left\{ \frac{n^{1/2}(\hat{\theta}^* - \theta)\sigma_x}{\hat{\sigma}} \leq \hat{g}_\alpha \right\} = P\left\{ \frac{n^{1/2}(\hat{\theta}^* - \theta)\sigma_x}{\hat{\sigma}^*} \leq \hat{h}_\alpha \right\} = \alpha
\]
for \( \alpha \in (0, 1) \).

Hall (1992) defines the one-sided hybrid confidence interval denoted \( \hat{I}_1 \), the Studentized confidence intervals denoted \( \hat{J}_1 \) and the backwards confidence interval denoted \( \hat{I}_{12} \) as the bootstrap confidence intervals for the slope parameter \( \theta \). The intervals below demonstrate how the critical points for these confidence intervals are used:

\[
\hat{I}_1 = (-\infty, \hat{\theta} - n^{-1/2}\sigma_x^{-1}\hat{\sigma}\hat{v}_{1-\alpha}),
\]

\[
\hat{J}_1 = (-\infty, \hat{\theta} - n^{-1/2}\sigma_x^{-1}\hat{\sigma}\hat{v}_{1-\alpha}),
\]

and

\[
\hat{I}_{12} = (-\infty, \hat{\theta} + n^{-1/2}\sigma_x^{-1}\hat{\sigma}\hat{v}_{\alpha}).
\]

Hall (1992) describes the Studentized interval as
\[
\hat{J}_1 = (-\infty, \hat{\theta} - n^{-1/2}\sigma_x^{-1}\hat{\sigma}\hat{v}_{1-\alpha})
\]
where the bootstrap Studentized interval $\hat{J}_1$ is the bootstrap estimate of $J_1$. The coverage for the Studentized interval is exact. That is, $P(\theta \in J_1) = \alpha$. Hall (1992) shows using Edgeworth expansions that $\hat{J}_1$ is second-order correct for $J_1$ and $\hat{I}_1$ is second-order correct of $J_1$. An interval is $k^{th}$-order correct for $J_1$ if the endpoints of the intervals agree up to and including terms of order $n^{k/2}$. That is, the difference between the endpoints of the intervals are $O(n^{-(k+1)/2})$.

The backwards confidence interval is not second-order correct. However, Hall (1992) mentions that this can be overcome by using the bias-corrected confidence interval. Therefore, Hall (1992) uses the accelerated bias-corrected interval defined by

$$\hat{I}_{abc} = (-\infty, \hat{y}_{abc,\alpha}),$$

where

$$\hat{y}_{abc,\alpha} = \hat{H}^{-1} \left( \Phi \left[ \hat{m} + (\hat{m} + z_\alpha) \{1 - \hat{a} (\hat{m} + z_\alpha)\}^{-1} \right] \right),$$

$$\hat{H}(x) = P \left( \hat{\theta}^* \leq x | X \right),$$

and

$$\hat{m} = \Phi^{-1} \left\{ \hat{H}(\hat{\theta}) \right\},$$

where $\hat{a} = -n^{-1/2} \frac{1}{3} \hat{\gamma} \gamma_x$ is called the acceleration coefficient, $\hat{\gamma} = n^{-1} \sum_{i=1}^{n} (\hat{\epsilon}_i / \hat{\sigma})^3$ and $\gamma_x = n^{-1} \sum_{i=1}^{n} \{(x_i - \bar{x}) / \sigma_x \}^3$. Hall (1992) shows that $\hat{I}_{ABC}$ is third-order correct.

Hall (1992) also considers the coverage accuracy of these confidence intervals. It appears that $\hat{I}_1$ has coverage error $O(n^{-1})$. Hall (1992) states that this is no surprise because $\hat{I}_1$ is second-order correct to the Studentized interval $J_1$. Since $\hat{I}_{ABC}$ is third-
order correct to $\hat{I}_1$, that implies that $\hat{I}_{abc}$ also has coverage error $O(n^{-1})$. Even though $\hat{J}_1$ is second-order correct for $J_1$, the coverage error is actually $O(n^{-3/2})$.

Hall (1992) remarks that since $\sum_{i=1}^{n} (x_i - \bar{x}) = 0$ there is symmetry in the design points and the slope has some unusual properties. These unusual properties cause the order of accuracy to be higher than normal. The unusual properties of bootstrap confidence intervals for a slope parameter do not apply to the confidence interval for the intercept parameter or $E(Y|x)$. Let

$$y_0 = E(Y|x = x_0) = \beta_0 + x_0 \beta_1,$$

so the estimate of the mean value would be

$$\hat{y}_0 = \hat{\beta}_0 + x_0 \hat{\beta}_1.$$

If we take $x_0 = 0$ then $\hat{y}_0 = \hat{\beta}_0$. Thus estimating the intercept parameter is a special case of estimating the mean.

Define for general $y_0$,

$$y_i = \sigma_x^{-2} \left\{ (x_0 - \bar{x}) (x_i - \bar{x}) + \sigma_x^2 \right\}$$

for $i = 1, \ldots, n$ and

$$\sigma_y^2 = n^{-1} \sum_{i=1}^{n} y_i^2 = 1 + \sigma_x^{-2} (x_0 - \bar{x})^2.$$

The variance of $\hat{y}_0$ is $n^{-1} \sigma^2 \sigma_y^2$, thus a one-sided bootstrap Studentized confidence interval for $y_0$ is

$$\hat{J}_1 = (-\infty, \hat{y}_0 - n^{-1} \hat{\sigma} \sigma_y \hat{\nu}_{1-\alpha}).$$

Hall (1992) also examines the multivariate regression model defined earlier. The
least squares estimate of $\beta_0$, $\beta$, and $\Sigma$ are

$$\hat{\beta} = \Sigma_x^{-1}n^{-1}\sum_{i=1}^{n}(x_i - \bar{x})^T(Y_i - \bar{Y}),$$

$$\hat{\beta}_0 = \bar{Y} - \bar{x}\hat{\beta},$$

and

$$\hat{\Sigma} = n^{-1}\sum_{i=1}^{n}\hat{\epsilon}_i\hat{\epsilon}_i^T,$$

where $\Sigma_x = n^{-1}\sum (x_i - \bar{x})^T(x_i - \bar{x})$ and $\hat{\epsilon}_i = Y_i - \hat{\beta}_0 - x_i\hat{\beta}$, for $i = 1, \ldots, n$.

Define

$$V = n^{-1}\sum_{i=1}^{n}(x_i - \bar{x})^T\Sigma(x_i - \bar{x})$$

and

$$\hat{V} = n^{-1}\sum_{i=1}^{n}(x_i - \bar{x})^T\hat{\Sigma}(x_i - \bar{x}).$$

Thus the variance matrix of $\hat{\beta}$ is $n^{-1}\Sigma_x^{-1}V\Sigma_x^{-1}$, the standardized version of $\hat{\beta}$ is

$$n^{1/2}(\Sigma_x^{-1}V\Sigma_x^{-1})^{-1/2}(\hat{\beta} - \beta),$$

and the Studentized version is

$$n^{1/2}(\Sigma_x^{-1}\hat{V}\Sigma_x^{-1})^{-1/2}(\hat{\beta} - \beta).$$

The conclusion of accuracy is similar to the univariate case.

Booth and Hall (1993) develop confidence intervals for regression models that have errors in the covariates. So instead of the linear equation $Y = X\beta + \varepsilon$, the true linear relationship is given by $V = a + bU$, where $X = U + \xi$ and $Y = V + \eta$ and $U, \xi$, and $\eta$ are independent random variables with means $\mu$, 0 and 0 and variances $\sigma_U^2, \sigma_\xi^2$ and $\sigma_\eta^2$, respectively. Let $\lambda = \sigma_\eta^2/\sigma_\xi^2$; then a range of functional values can be
estimated given by \((\hat{a}_\lambda + \hat{b}_\lambda u, \lambda \in \Lambda)\). This assumes that the true ratio variance is \(\lambda\). Let \(\Lambda = [\lambda_1, \lambda_2]\) and \(\mathcal{U} = [u_1, u_2]\), \(\Lambda \subset [0, \infty)\) and \(\mathcal{U} \subset (-\infty, \infty)\). The confidence band for \(v = a + bu\) overvalues \(u\) in the interval \(\mathcal{U}\), under the hypothesis that \(\lambda \in \Lambda\) is

\[
\mathcal{B} = \{(u, v) : g_1(u) \leq v \leq g_2(u), u \in \mathcal{U}\},
\]

where

\[
g_1(u) = \inf_{\lambda \in \Lambda} \{\hat{a}_\lambda + \hat{b}_\lambda u - n^{-1/2}t\hat{\omega}_\lambda(u)\},
\]

\[
g_2(u) = \sup_{\lambda \in \Lambda} \{\hat{a}_\lambda + \hat{b}_\lambda u - n^{-1/2}t\hat{\omega}_\lambda(u)\};
\]

\(\hat{\omega}_\lambda(u)\) is the asymptotic variance of

\[
n^{1/2}[\hat{a}_\lambda + \hat{b}_\lambda u - (a_\lambda + b_\lambda u)];
\]

and \(t\) is the solution of the equation

\[
P \left\{ \sup_{u \in \mathcal{U}, \lambda \in \Lambda} \left| \frac{n^{1/2}[\hat{a}_\lambda + \hat{b}_\lambda u - (a_\lambda + b_\lambda u)]}{\hat{\omega}_\lambda(u)} \right| \leq t \right\} = \alpha.
\]

In the case where \(\lambda\) is not given, this confidence band is estimated. Thus Booth and Hall (1993) used the percentile-\(t\) bootstrap confidence band to find a bootstrap estimate for \(\mathcal{B}\). However, the percentile-\(t\) only produces first-order corrected regions.

There is additional research that addresses the issue of confidence regions for the estimation of regression parameters. Adkins and Hill (1990) demonstrate how to use the bootstrap estimate to determine a confidence ellipsoid for regression parameters. Freedman and Peters (1984) look at the standard error of the regression parameter and compares conventional asymptotic estimates of standard error to the bootstrap method.
CHAPTER 3

OBSERVED CONFIDENCE LEVELS FOR SIMPLE LINEAR MODEL PARAMETERS

In this chapter we discuss different methods for obtaining observed confidence levels of regression parameters for a simple linear model. First, we will develop the observed confidence levels based on confidence regions of regression model parameters. Second, we develop the asymptotic accuracy for these confidence levels. Last, some examples will be shown.

3.1 Observed Confidence Levels

We will look at the observed confidence levels for regression parameters of a linear model defined by

\[ Y = X\beta + \varepsilon, \]  

(3.1.1)

where \( Y \) is a \( n \times 1 \) random vector, \( \beta \) is a \( k \times 1 \) vector of unknown parameters, \( X \) is a \( n \times (k + 1) \) matrix, and \( \varepsilon \) is a \( n \times 1 \) random vector where \( E(\varepsilon) = 0 \). To find the observed confidence levels of the regression parameters we will use a method developed by Polansky (2003b). Let \( C(\omega_1, \omega_2) \) be a 100\( \alpha \)% confidence region for \( \beta \) based on the sample \( \mathcal{X} = \{(X_i, Y_i), i = 1, \ldots, n\} \). Then let \( \hat{F} \) be the empirical distribution of the residual vector

\[ \hat{\varepsilon} = Y - X\hat{\beta} \]
defined by

\[ \hat{F}(t) = n^{-1} \sum_{i=1}^{n} I(\varepsilon_i \leq t), \quad i \in \{1, \ldots, n\}, \]

where \( I \) is the indicator function and \( \hat{\beta} \) is the least squares estimate of \( \beta \).

As a special case, consider a simple linear regression model where \( k = 1 \). Let

\[ y_i = \beta_0 + x_i \beta_1 + \varepsilon_i, \quad i \in \{1, \ldots, n\}. \tag{3.1.2} \]

As previously discussed, we will develop observed confidence levels for the slope and for the intercept parameters separately. First, we will consider the observed confidence levels for the slope parameter \( \beta_1 \). Let \( \theta = \beta_1 \); from Hall (1992), \( \hat{\sigma}^2 \) and \( \sigma_x^2 \) are defined in Section 2.5. Define

\[ \hat{\theta} = \sigma_x^{-2} n^{-1} \sum_{i=1}^{n} (x_i - \bar{x}) (y_i - \bar{y}), \tag{3.1.3} \]

where the variance of \( \hat{\theta} \) is \( n^{-1} \sigma_x^{-2} \sigma^2 \). Define the distributions,

\[ G_n = \left[ \frac{n^{1/2} \sigma_x (\hat{\theta} - \theta)}{\sigma} \leq t \right], \tag{3.1.4} \]

and

\[ H_n = \left[ \frac{n^{1/2} \sigma_x (\hat{\theta} - \theta)}{\hat{\sigma}} \leq t \right], \tag{3.1.5} \]

where \( g_\alpha = G_n^{-1}(\alpha) \) and \( h_\alpha = H_n^{-1}(\alpha) \) for \( \alpha \in (0, 1) \).

To obtain the observed confidence levels for \( \theta \), consider the theoretical critical points defined by Hall (1988). First is the exact critical point for the case where \( \sigma \) is known, called the ordinary critical point, and defined by

\[ \hat{\theta}_{\text{ord}}(\alpha) = \hat{\theta} - n^{-1/2} \sigma_x^{-1} \sigma g_{1-\alpha}. \tag{3.1.6} \]
Next is the exact critical point for the case where \( \sigma \) is unknown, called the Studentized critical point, and defined by

\[
\hat{\theta}_{\text{stud}}(\alpha) = \hat{\theta} - n^{-1/2} \sigma_x^{-1} \hat{\sigma} h_{1-\alpha}. \tag{3.1.7}
\]

Other critical points also available for the case when \( \sigma \) is unknown. These are the hybrid critical point,

\[
\hat{\theta}_{\text{hyb}}(\alpha) = \hat{\theta} - n^{-1/2} \sigma_x^{-1} \hat{\sigma} g_{1-\alpha}, \tag{3.1.8}
\]

and the backwards critical point,

\[
\hat{\theta}_{\text{back}}(\alpha) = \hat{\theta} + n^{-1/2} \sigma_x^{-1} \hat{\sigma} g_{\alpha}. \tag{3.1.9}
\]

Two-sided confidence intervals can then be constructed from these critical points as demonstrated previously. Let \( \omega_1 \in (0, 1) \) and \( \omega_2 \in (0, 1) \) such that \( \omega_2 - \omega_1 \in (0, 1) \), then \( C_{\text{stud}}(\omega_1, \omega_2) = [\hat{\theta}_{\text{stud}}(\omega_1), \hat{\theta}_{\text{stud}}(\omega_2)] \) is a 100(\( \omega_2 - \omega_1 \))\% confidence interval for \( \theta \) when \( \sigma \) is unknown. Also, it follows that \( C_{\text{hyb}}(\omega_1, \omega_2) \) and \( C_{\text{back}}(\omega_1, \omega_2) \) are approximate confidence intervals for \( \theta \) based on the hybrid and backwards theoretical critical points respectively.

For a region \( \Theta_k = (t_L, t_U) \) we desire to know the observed confidence levels corresponding to the theoretical critical points that were introduced previously. Setting \( \Theta_k = C_{\text{ord}}(\omega_1, \omega_2) \) yields

\[
t_L = \hat{\theta} - n^{-1/2} \sigma_x^{-1} \sigma g_{1-\omega_1 \rightarrow \omega_1} \tag{3.1.10}
\]

and

\[
t_U = \hat{\theta} - n^{-1/2} \sigma_x^{-1} \sigma g_{1-\omega_2 \rightarrow \omega_2}. \tag{3.1.11}
\]

Solving for \( \omega_1 \) and \( \omega_2 \) in Equations (3.1.10) and (3.1.11) yields

\[
\omega_1 = 1 - G_n \left[ \frac{n^{1/2} \sigma_x (\hat{\theta} - t_L)}{\sigma} \right]
\]

\[
\omega_2 = 1 - G_n \left[ \frac{n^{1/2} \sigma_x (\hat{\theta} - t_U)}{\sigma} \right]
\]
and
\[
\omega_2 = 1 - G_n \left[ \frac{n^{1/2} \sigma_x (\hat{\theta} - t_U)}{\sigma} \right].
\]
The observed confidence level that corresponds to the ordinary critical point is
\[
\alpha_{\text{ord}}(\Theta_k) = \omega_2 - \omega_1 = G_n \left[ \frac{n^{1/2} \sigma_x (\hat{\theta} - t_L)}{\sigma} \right] - G_n \left[ \frac{n^{1/2} \sigma_x (\hat{\theta} - t_U)}{\sigma} \right].
\] (3.1.12)

A similar algorithm is used to develop the observed confidence levels based on the other critical points. Let \( t_L = \hat{\theta} - n^{-1/2} \sigma_x^{-1} \hat{\sigma} h_{1-\omega_1} \) and \( t_U = \hat{\theta} - n^{-1/2} \sigma_x^{-1} \hat{\sigma} h_{1-\omega_2} \), then the observed confidence level that corresponds to the Studentized critical point is
\[
\alpha_{\text{stud}}(\Theta_k) = \omega_2 - \omega_1 = H_n \left[ \frac{n^{1/2} \sigma_x (\hat{\theta} - t_L)}{\hat{\sigma}} \right] - H_n \left[ \frac{n^{1/2} \sigma_x (\hat{\theta} - t_U)}{\hat{\sigma}} \right].
\] (3.1.13)

Also, using the hybrid critical points, let \( t_L = \hat{\theta} - n^{-1/2} \sigma_x^{-1} \hat{\sigma} g_{1-\omega_1} \) and \( t_U = \hat{\theta} - n^{-1/2} \sigma_x^{-1} \hat{\sigma} g_{1-\omega_2} \). The observed confidence level that corresponds to the hybrid theoretical critical point is
\[
\alpha_{\text{hyb}}(\Theta_k) = \omega_2 - \omega_1 = G_n \left[ \frac{n^{1/2} \sigma_x (\hat{\theta} - t_L)}{\hat{\sigma}} \right] - G_n \left[ \frac{n^{1/2} \sigma_x (\hat{\theta} - t_U)}{\hat{\sigma}} \right].
\] (3.1.14)

Last, for the backwards critical points, let \( t_L = \hat{\theta} + n^{-1/2} \sigma_x^{-1} \hat{\sigma} g_{\omega_1} \) and \( t_U = \hat{\theta} + n^{-1/2} \sigma_x^{-1} \hat{\sigma} g_{\omega_2} \). The observed confidence level that corresponds to the backwards theoretical critical point is
\[
\alpha_{\text{back}}(\Theta_k) = \omega_2 - \omega_1 = G_n \left[ \frac{n^{1/2} \sigma_x (t_U - \hat{\theta})}{\hat{\sigma}} \right] - G_n \left[ \frac{n^{1/2} \sigma_x (t_L - \hat{\theta})}{\hat{\sigma}} \right].
\] (3.1.15)

When the distribution \( F \) is unknown, the distributions of \( G_n \) and \( H_n \) will usually also be unknown. Therefore, where \( F \) is unknown, we need to estimate or approximate the observed confidence levels. One way to approximate the observed
confidence levels is based on the Central Limit Theorem. Assuming the smooth function model described in Section 2.4, $\sigma^2 > 0$ and $\frac{\partial}{\partial \mu} g(\mu) \neq 0$, then

$$\frac{n^{1/2}(\hat{\theta} - \theta)}{\sigma} \xrightarrow{w} Z,$$

and

$$\frac{n^{1/2}(\hat{\theta} - \theta)}{\hat{\sigma}} \xrightarrow{w} Z,$$

as $n \to \infty$, where $Z$ is the standard normal random variable. This implies that $G_n(t) \simeq \Phi(t)$ and $H_n(t) \simeq \Phi(t)$, where $\Phi$ is the distribution function of the standard normal distribution. Therefore, using the normal approximations yields the following estimated observed confidence levels:

$$\hat{\alpha}_{ord}(\Theta_k) = \Phi \left[ \frac{n^{1/2} \sigma_x (\hat{\theta} - t_L)}{\sigma} \right] - \Phi \left[ \frac{n^{1/2} \sigma_x (\hat{\theta} - t_U)}{\sigma} \right], \quad (3.1.16)$$

and

$$\hat{\alpha}_{stud}(\Theta_k) = \hat{\alpha}_{hyb}(\Theta_k) = \hat{\alpha}_{back}(\Theta_k)$$

$$= \Phi \left[ \frac{n^{1/2} \sigma_x (\hat{\theta} - t_L)}{\hat{\sigma}} \right] - \Phi \left[ \frac{n^{1/2} \sigma_x (\hat{\theta} - t_U)}{\hat{\sigma}} \right]. \quad (3.1.17)$$

If the normal approximations are not accurate enough, another way to estimate the observed confidence levels is by using the bootstrap method. To use the bootstrap method for a regression model, let $\{\varepsilon_1^*, \ldots, \varepsilon_n^*\}$ be a independent and identically distributed sample from $\hat{F}_n$, the empirical distribution function of $\{\hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_n\}$. The bootstrap regression model is

$$y_i^* = \hat{\beta}_0 + x_i \hat{\theta} + \varepsilon_i^*, \quad i \in \{1, \ldots, n\},$$

and the bootstrap estimate of the slope parameter is

$$\hat{\theta}^* = \sigma_x^2 n^{-1} \sum_{i=1}^{n} (x_i - \bar{x})(y_i^* - \bar{y}^*). \quad (3.1.18)$$
Thus the bootstrap estimate of $G_n$ is given by

$$
\hat{G}_n(t) = P \left[ \frac{n^{1/2} \sigma_x (\hat{\theta}^* - \hat{\theta})}{\hat{\sigma}} \leq t \right] \{ \varepsilon_i^*, 1 \leq i \leq n \} \sim \hat{F}, X, Y.
$$

(3.1.19)

Similarly, the bootstrap estimate of $H_n$ is given by

$$
\hat{H}_n(t) = P \left[ \frac{n^{1/2} \sigma_x (\hat{\theta}^* - \hat{\theta})}{\hat{\sigma}^*} \leq t \right] \{ \varepsilon_i^*, 1 \leq i \leq n \} \sim \hat{F}, X, Y,
$$

(3.1.20)

where $\hat{\sigma}^* = n^{-1} \sum_{i=1}^{n} \{ y_i - \bar{y}^* - (x_i - \bar{x})\hat{\theta}^* \}^2$. Define $\hat{\sigma}_a = G_n^{-1}(\alpha)$ and $\hat{\sigma}_a = H_n^{-1}(\alpha)$ for $\alpha \in (0,1)$, conditional on $X$ and $Y$. The bootstrap estimates of the theoretical critical points are

$$
\hat{\theta}_{\text{ord}}^*(\alpha) = \hat{\theta} - n^{-1/2} \sigma_x^{-1} \sigma \hat{\sigma}_1 - \alpha,
$$

(3.1.21)

$$
\hat{\theta}_{\text{stud}}^*(\alpha) = \hat{\theta} - n^{-1/2} \sigma_x^{-1} \sigma \hat{\sigma}_1 - \alpha,
$$

(3.1.22)

$$
\hat{\theta}_{\text{hyb}}^*(\alpha) = \hat{\theta} - n^{-1/2} \sigma_x^{-1} \sigma \hat{\sigma}_1 - \alpha.
$$

(3.1.23)

and

$$
\hat{\theta}_{\text{back}}^*(\alpha) = \hat{\theta} + n^{-1/2} \sigma_x^{-1} \sigma \hat{\sigma}_\alpha.
$$

(3.1.24)

The estimated observed confidence levels based on the above bootstrap critical points are

$$
\hat{\alpha}_{\text{ord}}^*(\Theta_k) = \hat{G}_n \left[ \frac{n^{1/2} \sigma_x (\hat{\theta} - \theta)}{\sigma} \right] - \hat{G}_n \left[ \frac{n^{1/2} \sigma_x (\hat{\theta} - \theta)}{\sigma} \right],
$$

(3.1.25)

$$
\hat{\alpha}_{\text{stud}}^*(\Theta_k) = \hat{H}_n \left[ \frac{n^{1/2} \sigma_x (\hat{\theta} - \theta)}{\sigma} \right] - \hat{H}_n \left[ \frac{n^{1/2} \sigma_x (\hat{\theta} - \theta)}{\sigma} \right],
$$

(3.1.26)

$$
\hat{\alpha}_{\text{hyb}}^*(\Theta_k) = \hat{G}_n \left[ \frac{n^{1/2} \sigma_x (\hat{\theta} - \theta)}{\sigma} \right] - \hat{G}_n \left[ \frac{n^{1/2} \sigma_x (\hat{\theta} - \theta)}{\sigma} \right],
$$

(3.1.27)

and

$$
\hat{\alpha}_{\text{back}}^*(\Theta_k) = \hat{G}_n \left[ \frac{n^{1/2} \sigma_x (\theta - \hat{\theta})}{\sigma} \right] - \hat{G}_n \left[ \frac{n^{1/2} \sigma_x (\theta - \hat{\theta})}{\sigma} \right].
$$

(3.1.28)
3.2 Asymptotic Accuracy

The observed confidence levels defined in Section 3.1 are based on different theoretical critical points and methods of approximation. Therefore, we want to compare the different methods for computing observed confidence levels by a measure of accuracy. Given a region \( \Theta_k \) that corresponds to a 100\( \alpha \)% confidence interval for \( \theta \), the observed confidence level of \( \Theta_k \) should be \( \alpha \). That is, an observed confidence level is accurate if it is equal to \( \alpha \) when applied to a 100\( \alpha \)% confidence region. When \( \sigma \) is known, \( C_{\text{ord}}(\omega_1, \omega_2) \) will be used as the standard confidence interval for \( \theta \). Thus when \( \omega_2 - \omega_1 = \alpha \), an observed confidence level \( \tilde{\alpha} \) is accurate if \( \tilde{\alpha}[C_{\text{ord}}(\omega_1, \omega_2)] = \alpha \). Similarly, if \( \sigma \) is unknown, \( C_{\text{stud}}(\omega_1, \omega_2) \) will be used as the standard confidence interval for \( \theta \). Thus, when \( \alpha = \omega_2 - \omega_1 \), the measure \( \tilde{\alpha} \) is accurate if \( \tilde{\alpha}[C_{\text{stud}}(\omega_1, \omega_2)] = \alpha \). By this definition, \( \alpha_{\text{ord}} \) and \( \alpha_{\text{stud}} \) are accurate for the cases when \( \sigma \) is known or unknown, respectively. If \( \tilde{\alpha} \) is not accurate we can measure the order of accuracy based on asymptotic expansions. In the case where \( \sigma \) is known, \( \tilde{\alpha} \) is said to be \( k^{th} \)-order accurate if \( \tilde{\alpha}[C_{\text{ord}}(\omega_1, \omega_2)] = \alpha + O(n^{-k/2}) \) as \( n \to \infty \). In the case where \( \sigma \) is unknown, \( \tilde{\alpha} \) is said to be \( k^{th} \)-order accurate if \( \tilde{\alpha}[C_{\text{stud}}(\omega_1, \omega_2)] = \alpha + O(n^{-k/2}) \) as \( n \to \infty \).

We will use the Edgeworth expansion theory to calculate the order of accuracy. This theory was discussed by Bhattacharya and Ghosh (1978) and Hall (1988, 1992). Following the procedure outlined in Section 2.3, let \( U = n^{-1/2}(\tilde{\theta} - \theta) \) and \( \text{Var}(\tilde{\theta}) = \)
$n^{-1} \sigma_x^{-2} \sigma^{-2}$. Then the Edgeworth expansion for $G_n(t)$ for $\nu \geq 1$ is

$$G_n(t) = P \left[ \frac{n^{1/2} \sigma_x (\hat{\theta} - \theta)}{\sigma} \leq t \right]$$

$$= \Phi(t) + \sum_{i=1}^{\nu} n^{-i/2} p_i(t) \phi(t) + O(n^{-(\nu+1)/2}), \quad (3.2.1)$$

where $\phi(t)$ is the density of the standard normal distribution. Similarly, the Edgeworth expansion for $H_n(t)$ for $\nu \geq 1$ is

$$H_n(t) = P \left[ \frac{n^{1/2} \sigma_x (\hat{\theta} - \theta)}{\tilde{\sigma}} \leq t \right]$$

$$= \Phi(t) + \sum_{i=1}^{\nu} n^{-i/2} q_i(t) \phi(t) + O(n^{-(\nu+1)/2}). \quad (3.2.2)$$

Hall (1992) derives the polynomials $p_1, q_1, p_2$ and $q_2$ for the slope parameter as

$$p_1(t) = q_1(t) = -\frac{1}{6} \gamma \gamma_x (t^2 - 1), \quad (3.2.3)$$

$$p_2(t) = -t \left[ \frac{1}{24} \kappa (\kappa_x + 3)(t^2 - 3) + \frac{1}{72} \gamma^2 \gamma_x^2 (t^4 - 10t^2 + 15) \right], \quad (3.2.4)$$

and

$$q_2(t) = -t \left[ 2 + \frac{1}{24} (\kappa \kappa_x + 6)(t^2 - 3) + \frac{1}{72} \gamma^2 \gamma_x^2 (t^4 - 10t^2 + 15) \right]. \quad (3.2.5)$$

Here $\gamma, \gamma_x, \kappa$ and $\kappa_x$ are

$$\gamma = E(\varepsilon / \sigma)^3,$$

$$\gamma_x = n^{-1} \sum_{i=1}^{n} [(x_i - \bar{x}) / \sigma_x]^3,$$

$$\kappa = E(\varepsilon / \sigma)^4 - 3,$$

and

$$\kappa_x = n^{-1} \sum_{i=1}^{n} \{x_i - \bar{x} / \sigma_x\}^4 - 3.$$
The bootstrap versions of these Edgeworth expansions are

\[
\hat{G}_n(t) = P \left[ \frac{n^{1/2} \sigma_x (\hat{\theta}^* - \hat{\theta})}{\hat{\sigma}} \leq t | \{ \varepsilon_i^*, 1 \leq i \leq n \} \sim \hat{F}, X, Y \right] \\
= \Phi(t) + \sum_{i=1}^{\nu} n^{-i/2} \hat{p}_i(t) \phi(t) + O(n^{-(\nu+1)/2}) 
\tag{3.2.6}
\]

and

\[
\hat{H}_n(t) = P \left[ \frac{n^{1/2} \sigma_x (\hat{\theta}^* - \hat{\theta})}{\hat{\sigma}^*} \leq t | \{ \varepsilon_i^*, 1 \leq i \leq n \} \sim \hat{F}, X, Y \right] \\
= \Phi(t) + \sum_{i=1}^{\nu} n^{-i/2} \hat{q}_i(t) \phi(t) + O(n^{-(\nu+1)/2}), 
\tag{3.2.7}
\]

where the polynomials \( \hat{p}_1, \hat{q}_1, \hat{p}_2 \) and \( \hat{q}_2 \) are estimated by using the sample residuals as

\[
\hat{p}_1(t) = \hat{q}_1(t) = -\frac{1}{6} \hat{\gamma} \gamma_x (t^2 - 1), \\
\hat{p}_2(t) = -t \left[ \frac{1}{24} \hat{\kappa} (\kappa_x + 3)(t^2 - 3) + \frac{1}{72} \hat{\gamma}^2 \gamma_x^2 (t^4 - 10t^2 + 15) \right]
\]

and

\[
\hat{q}_2(t) = -t \left[ 2 + \frac{1}{24} (\hat{\kappa} \kappa_x + 6)(t^2 - 3) + \frac{1}{72} \hat{\gamma}^2 \gamma_x^2 (t^4 - 10t^2 + 15) \right],
\]

where

\[
\hat{\gamma} = n^{-1} \sum_{i=1}^{n} (\hat{\varepsilon}_i / \hat{\sigma})^3
\]

and

\[
\hat{\kappa} = n^{-1} \sum_{i=1}^{n} (\hat{\varepsilon}_i / \hat{\sigma})^4 - 3.
\]

The polynomials \( \hat{p}_i \) and \( \hat{q}_i \) are similar to \( p_i \) and \( q_i \) except that we are using the moments of the sample residuals from the distribution of \( \hat{F} \) to determine these
polynomials. Therefore, \( \hat{p}_i = p_i + O_p(n^{-1/2}) \) and \( \hat{q}_i = q_i + O_p(n^{-1/2}) \). Then the bootstrap versions of \( G_n \) and \( H_n \) are second-order correct in probability since

\[
\hat{G}_n(t) = \Phi(t) + n^{-1/2}[p_1(t) + O_p(n^{-1/2})] \phi(t) + O(n^{-1}) = \Phi(t) + n^{-1/2}p_1(t) \phi(t) + O_p(n^{-1}) \quad (3.2.8)
\]

and

\[
\hat{H}_n(t) = \Phi(t) + n^{-1/2}[q_1(t) + O_p(n^{-1/2})] \phi(t) + O(n^{-1}) = \Phi(t) + n^{-1/2}q_1(t) \phi(t) + O_p(n^{-1}) \quad (3.2.9)
\]

The Cornish-Fisher inversions of \( G_n \) and \( H_n \) are

\[
g_\alpha = z_\alpha + \sum_{i=1}^{\nu} n^{-i/2} p_{1i}(z_\alpha) + O(n^{-(\nu+1)/2})
\]

\[
= z_\alpha - n^{-1/2} p_1(z_\alpha) + O(n^{-1}) \quad (3.2.10)
\]

and

\[
h_\alpha = z_\alpha + \sum_{i=1}^{\nu} n^{-i/2} q_{1i}(z_\alpha) + O(n^{-(\nu+1)/2})
\]

\[
= z_\alpha - n^{-1/2} q_1(z_\alpha) + O(n^{-1}), \quad (3.2.11)
\]

where \( \Phi(z_\alpha) = \alpha, \ p_{11}(t) = -p_1(t) \) and \( q_{11}(t) = -q_1(t) \). The bootstrap versions are

\[
\hat{g}_\alpha = z_\alpha + \sum_{i=1}^{\nu} n^{-i/2} \hat{p}_{1i}(z_\alpha) + O(n^{-(\nu+1)/2})
\]

\[
= z_\alpha - n^{-1/2} \hat{p}_1(z_\alpha) + O(n^{-1}) \quad (3.2.12)
\]
and

\[
\hat{h}_\alpha = z_\alpha + \sum_{i=1}^{\nu} n^{-i/2} \hat{q}_1(z_\alpha) + O(n^{-(\nu+1)/2})
\]
\[
= z_\alpha - n^{-1/2} \hat{q}_1(z_\alpha) + O(n^{-1}). \tag{3.2.13}
\]

Notice that \( \hat{p}_1 = p_1 + O_p(n^{-1/2}) \) and \( \hat{q}_1 = q_1 + O_p(n^{-1/2}) \), so this results in

\[
\hat{g}_\alpha = z_\alpha - n^{-1/2} [p_1(z_\alpha) + O_p(n^{-1/2})] + O(n^{-1})
\]
\[
= z_\alpha - n^{-1/2} p_1(z_\alpha) + O_p(n^{-1}) \tag{3.2.14}
\]

and

\[
\hat{h}_\alpha = z_\alpha - n^{-1/2} q_1(z_\alpha) + O_p(n^{-1/2}) + O(n^{-1})
\]
\[
= z_\alpha - n^{-1/2} q_1(z_\alpha) + O_p(n^{-1}). \tag{3.2.15}
\]

We will use these expansions to obtain the order of accuracy for the observed confidence levels based on the critical points \( \theta_{\text{ord}}, \theta_{\text{stud}}, \theta_{\text{hyb}} \) and \( \theta_{\text{back}} \).

First, we look at the observed confidence levels based on the normal approximations. Suppose \( \sigma \) is known and \( \alpha = \omega_2 - \omega_1 \), then we use the interval \( C_{\text{ord}}(\omega_1, \omega_2) = [\hat{\theta}_{\text{ord}}(\omega_1), \hat{\theta}_{\text{ord}}(\omega_2)] \) using Equation (3.1.6). Thus, the accuracy of the approximate observed confidence level is

\[
\check{\alpha}_{\text{ord}}[C_{\text{ord}}(\omega_1, \omega_2)] = \Phi \left\{ n^{-1/2} \sigma \frac{\hat{\theta} - \hat{\theta}_{\text{ord}}(\omega_1)}{\sigma} \right\} - \Phi \left\{ n^{-1/2} \sigma \frac{\hat{\theta} - \hat{\theta}_{\text{ord}}(\omega_2)}{\sigma} \right\}
\]
\[
= \Phi(g_{1-\omega_1}) - \Phi(g_{1-\omega_2}). \tag{3.2.16}
\]

Then using the Taylor expansion for the standard normal distribution \( \Phi(t) \),

\[
\Phi(g_{1-\omega}) = \Phi[z_{1-\omega} - n^{-1/2} p_1(z_{1-\omega}) + O(n^{-1})]
\]
\[
= \Phi(z_{1-\omega}) - n^{-1/2} \phi(z_{1-\omega}) p_1(z_{1-\omega}) + O(n^{-1}) \tag{3.2.17}
\]
\[
= 1 - \omega - n^{-1/2} \phi(z_{1-\omega}) p_1(z_{1-\omega}) + O(n^{-1}).
\]
Therefore,

\[
\hat{\alpha}_{\text{ord}}[C_{\text{ord}}(\omega_1, \omega_2)] = \alpha + n^{-1/2}[\phi(z_{1-\omega_2})p_1(z_{1-\omega_2}) - \phi(z_{1-\omega_1})p_1(z_{1-\omega_1})] + O(n^{-1}). \tag{3.2.18}
\]

This shows that \( \hat{\alpha}_{\text{ord}} \) is first-order accurate. When \( \sigma \) is unknown and \( \alpha = \omega_2 - \omega_1 \), we have

\[
\hat{\alpha}_{\text{stud}}[C_{\text{stud}}(\omega_1, \omega_2)] = \Phi\left\{ \frac{n^{-1/2}\sigma_x[\hat{\theta} - \hat{\theta}_{\text{stud}}(\omega_1)]}{\sigma} \right\} - \Phi\left\{ \frac{n^{-1/2}\sigma_x[\hat{\theta} - \hat{\theta}_{\text{stud}}(\omega_2)]}{\sigma} \right\} = \Phi(h_{1-\omega_1}) - \Phi(h_{1-\omega_2}). \tag{3.2.19}
\]

Then using the Taylor expansion for \( \phi \) and similar arguments to those above,

\[
\hat{\alpha}_{\text{stud}}[C_{\text{stud}}(\omega_1, \omega_2)] = \alpha + n^{-1/2}[\phi(z_{1-\omega_2})q_1(z_{1-\omega_2}) - \phi(z_{1-\omega_1})q_1(z_{1-\omega_1})] + O(n^{-1}). \tag{3.2.20}
\]

Thus \( \hat{\alpha}_{\text{stud}} \) is also first-order accurate. Therefore the methods based on the normal approximation are first-order accurate.

If the distributions \( G_n \) and \( H_n \) are known, then the normal approximations are not needed and we can use Equations (3.1.12)-(3.1.15) to compute the observed confidence levels using these distributions. If \( \sigma \) is known, we have

\[
\hat{\alpha}_{\text{ord}}[C_{\text{ord}}(\omega_1, \omega_2)] = G_n(g_{1-\omega_1}) - G_n(g_{1-\omega_2}) = 1 - \omega_1 - 1 + \omega_2 = \omega_2 - \omega_1 = \alpha. \tag{3.2.21}
\]

Similarly, if \( \sigma \) is unknown, we have

\[
\hat{\alpha}_{\text{stud}}[C_{\text{stud}}(\omega_1, \omega_2)] = H_n(h_{1-\omega_1}) - H_n(h_{1-\omega_2}) = 1 - \omega_1 - 1 + \omega_2 = \alpha. \tag{3.2.22}
\]
Thus if \( G_n \) and \( H_n \) are known, then \( \alpha_{\text{ord}} \) and \( \alpha_{\text{stud}} \) are both accurate.

To measure the order of accuracy for \( \alpha_{\text{hyb}} \) and \( \alpha_{\text{back}} \), we use the Edgeworth and Taylor expansions and Cornish-Fisher inversions. In both cases \( \sigma \) is unknown. Thus we use \( C_{\text{stud}}(\omega_1, \omega_2) \) as the standard confidence interval for \( \theta \). For \( \alpha_{\text{hyb}} \) we have

\[
\alpha_{\text{hyb}}[C_{\text{stud}}(\omega_1, \omega_2)] = G_n \left\{ \frac{n^{1/2} \sigma_x [\hat{\theta} - \hat{\theta}_{\text{stud}}(\omega_1)]}{\hat{\sigma}} \right\} - G_n \left\{ \frac{n^{1/2} \sigma_x [\hat{\theta} - \hat{\theta}_{\text{stud}}(\omega_2)]}{\hat{\sigma}} \right\}
= G_n(h_{1-\omega}) - G_n(h_{1-\omega_2}). \tag{3.2.23}
\]

Using the Edgeworth expansion yields

\[
G_n(h_{1-\omega}) = \Phi(h_{1-\omega}) + n^{-1/2} p_1(h_{1-\omega}) \phi(h_{1-\omega}) + O(n^{-1}). \tag{3.2.24}
\]

Using the Taylor expansion and Cornish-Fisher inversion yields

\[
\Phi(h_{1-\omega}) = \Phi[z_{1-\omega} - n^{-1/2} q_1(z_{1-\omega}) + O(n^{-1})]
= \Phi(z_{1-\omega}) - n^{-1/2} q_1(z_{1-\omega}) \phi(z_{1-\omega}) + O(n^{-1}) \tag{3.2.25}
= 1 - \omega - n^{-1/2} q_1(z_{\omega}) \phi(z_{\omega}) + O(n^{-1}),
\]

\[
p_1(h_{1-\omega}) = p_1[z_{1-\omega} - n^{-1/2} q_1(z_{1-\omega}) + O(n^{-1})]
= p_1(z_{1-\omega}) - n^{-1/2} q_1(z_{1-\omega}) p'_1(z_{1-\omega}) + O(n^{-1}) \tag{3.2.26}
= p_1(z_{\omega}) - n^{-1/2} q_1(z_{\omega}) p'_1(z_{1-\omega}) + O(n^{-1}),
\]

and

\[
\phi(h_{1-\omega}) = \phi[z_{1-\omega} - n^{-1/2} q_1(z_{1-\omega}) + O(n^{-1})]
= \phi(z_{\omega}) - n^{-1/2} z_{\omega} q_1(z_{\omega}) \phi(z_{\omega}) + O(n^{-1}). \tag{3.2.27}
\]

Notice that since the polynomials \( p_1(t) \) and \( q_1(t) \) are even and \( z_{1-\omega} = -z_{\omega} \),

\[
p_1(z_{1-\omega}) = p_1(-z_{\omega}) = p_1(z_{\omega})
\]
and
\[ q_1(z_{1-\omega}) = q_1(-z_{\omega}) = q_1(z - \omega). \]

Substituting Equations (3.2.25), (3.2.26), and (3.2.27) into Equation (3.2.24) yields
\[ G_n(h_{1-\omega}) = 1 - \omega - n^{-1/2}q_1(z_{\omega})\phi(z_{\omega}) + n^{-1/2}p_1(z_{\omega})\phi(z_{\omega}) + O(n^{-1}) \]
\[ = 1 - \omega + O(n^{-1}). \]

(3.2.28)

Notice that \( p_1(z_{\omega}) = q_1(z_{\omega}) \). This occurs because \( \sum (x_i - \bar{x}) = 0 \) (see Hall, 1992).

Then substituting Equation (3.2.28) into Equation (3.2.23) yields
\[ \alpha_{\text{hyb}}[C_{\text{stud}}(\omega, \omega_2)] = 1 - \omega_1 - 1 + \omega_2 + O(n^{-1}) \]
\[ = \alpha + O(n^{-1}). \]

(3.2.29)

Thus \( \alpha_{\text{hyb}} \) is second-order accurate. For \( \alpha_{\text{back}} \) we have
\[ \alpha_{\text{back}}[C_{\text{stud}}(\omega_1, \omega_2)] = G_n \left\{ \frac{n^{1/2}\sigma_x[\hat{\theta}_{\text{stud}}(\omega_2) - \hat{\theta}]}{\hat{\sigma}} \right\} - G_n \left\{ \frac{n^{1/2}\sigma_x[\hat{\theta}_{\text{stud}}(\omega_1) - \hat{\theta}]}{\hat{\sigma}} \right\} \]
\[ = G_n(-h_{1-\omega_2}) - G_n(-h_{1-\omega_1}). \]

(3.2.30)

Using the Taylor expansion and the Cornish-Fisher inversions yields
\[ \Phi(-h_{1-\omega}) = \Phi[-z_{1-\omega} + n^{-1/2}q_1(z_{1-\omega}) + O(n^{-1})] \]
\[ = \omega + n^{-1/2}q_1(z_{\omega})\phi(z_{\omega}) + O(n^{-1}), \]
\[ \Phi(-h_{1-\omega}) = n^{-1/2}p_1(z_{\omega})\phi(z_{\omega}) + O(n^{-1}). \]

(3.2.31)

where \( p_1(-h_{1-\omega}) = p_1(h_{1-\omega}) \) since \( p_1 \) is an even polynomial. Thus we can use Equation (3.2.26). Also, \( \phi(-h_{1-\omega}) = \phi(h_{1-\omega}) \), which is equal to Equation (3.2.27).

Combining Equations (3.2.26), (3.2.27) and (3.2.31) together yields
\[ G_n(-h_{1-\omega}) = \Phi(-h_{1-\omega}) + n^{-1/2}p_1(-h_{1-\omega})\phi(-h_{1-\omega}) + O(n^{-1}) \]
\[ = \omega + n^{-1/2}q_1(z_{\omega})\phi(z_{\omega}) \]
\[ + n^{-1/2}p_1(z_{\omega})\phi(z_{\omega}) + O(n^{-1}). \]

(3.2.32)
Thus

\[ \alpha_{\text{back}}[\mathcal{C}_{\text{stud}}(\omega_1, \omega_2)] = G_n(-h_{1-\omega_2}) - G_n(-h_{1-\omega_1}) \]

\[ = \omega_2 + n^{-1/2}[p_1(z_{\omega_2}) + q_1(z_{\omega_2})]\phi(z_{\omega_2}) \]

\[ - \omega_1 - n^{-1/2}[p_1(z_{\omega_1}) + q_1(z_{\omega_1})]\phi(z_{\omega_1}) + O(n^{-1}) \]

\[ = \alpha + O(n^{-1/2}). \quad (3.2.33) \]

This shows that \( \alpha_{\text{back}} \) is first-order accurate.

For the bootstrap estimates, notice that from Equation (3.2.9) \( \hat{G}_n(t) = G_n(t) + O_p(n^{-1/2}) \). Thus for the bootstrap estimate of \( \hat{\alpha}_{\text{ord}} \)

\[ \hat{\alpha}_{\text{ord}}[\mathcal{C}_{\text{ord}}(\omega_1, \omega_2)] = \hat{G}_n \left\{ \frac{n^{1/2}[\hat{\theta} - \hat{\theta}_{\text{ord}}(\omega_1)]}{\sigma} \right\} - \hat{G}_n \left\{ \frac{n^{1/2}[\hat{\theta} - \hat{\theta}_{\text{ord}}(\omega_2)]}{\sigma} \right\} \]

\[ = G_n \left\{ \frac{n^{1/2}[\hat{\theta} - \hat{\theta}_{\text{ord}}(\omega_1)]}{\sigma} \right\} - G_n \left\{ \frac{n^{1/2}[\hat{\theta} - \hat{\theta}_{\text{ord}}(\omega_2)]}{\sigma} \right\} \]

\[ + O_p(n^{-1}) \]

\[ = \alpha + O_p(n^{-1}). \quad (3.2.34) \]

Thus \( \hat{\alpha}_{\text{ord}} \) is second-order accurate in probability. Similar arguments establish the accuracy for the bootstrap estimates of \( \hat{\alpha}_{\text{stud}}, \hat{\alpha}_{\text{hyb}} \) and \( \alpha_{\text{back}} \):

\[ \hat{\alpha}_{\text{stud}}[\mathcal{C}_{\text{stud}}(\omega_1, \omega_1)] = \hat{H}_n \left\{ \frac{n^{1/2}[\hat{\theta} - \hat{\theta}_{\text{stud}}(\omega_1)]}{\sigma} \right\} - \hat{H}_n \left\{ \frac{n^{1/2}[\hat{\theta} - \hat{\theta}_{\text{stud}}(\omega_2)]}{\sigma} \right\} \]

\[ = H_n \left\{ \frac{n^{1/2}[\hat{\theta} - \hat{\theta}_{\text{stud}}(\omega_1)]}{\sigma} \right\} - H_n \left\{ \frac{n^{1/2}[\hat{\theta} - \hat{\theta}_{\text{stud}}(\omega_2)]}{\sigma} \right\} \]

\[ + O_p(n^{-1}) \]

\[ = \alpha + O_p(n^{-1}), \quad (3.2.35) \]
\[ \hat{\alpha}_{\text{hyb}}[c_{\text{stud}}(\omega_1, \omega_2)] = \hat{G}_n(h_1 - \omega_1) - \hat{G}_n(h_1 - \omega_2) \]
\[ = G_n(h_1 - \omega_1) - G_n(h_1 - \omega_2) + O_p(n^{-1}) \]  \hspace{1cm} (3.2.36)
\[ = \alpha + O_p(n^{-1}), \]

and

\[ \hat{\alpha}_{\text{back}}[c_{\text{stud}}(\omega_1, \omega_2)] = \hat{G}_n(-h_1 - \omega_2) - \hat{G}_n(-h_1 - \omega_1) \]
\[ = G_n(-h_1 - \omega_2) - G_n(-h_1 - \omega_1) + O_p(n^{-1}) \]  \hspace{1cm} (3.2.37)
\[ = \alpha + O_p(n^{-1/2}). \]

Thus, the bootstrap estimates of \( \alpha_{\text{ord}} \), \( \alpha_{\text{stud}} \) and \( \alpha_{\text{hyb}} \) are second-order accurate in probability, whereas the bootstrap estimate of \( \alpha_{\text{back}} \) is first-order accurate in probability.

### 3.3 Bias Corrected

Notice that the order of accuracy for the observed confidence levels based on the backwards critical point was only first-order, while the other observed confidence levels are second-order. Hall (1992) shows that this lower accuracy can be overcome by adjusting the critical point of the backwards method. This correction is done in two ways. First, by adjusting for the bias in the position of the critical point. This is called the bias-corrected method. The critical point for the slope parameter based on the bias-corrected method is

\[ \hat{\theta}_{bc}(\alpha) = \hat{\theta}_{\text{back}}(\psi) = \hat{\theta} + n^{-1/2} \hat{\sigma}_x^{-1} \hat{\sigma}_\psi, \] \hspace{1cm} (3.3.1)

where \( \psi = \Phi(z_\alpha + m) \) and \( m = \Phi^{-1}[G_n(0)] \). Setting

\[ \Theta_k = C(\omega_1, \omega_2) = [\hat{\theta}_{bc}(\omega_1), \hat{\theta}_{bc}(\omega_2)] = [t_L, t_U] \]
yields
\[ g_\psi = \frac{n^{1/2} \sigma_x(t_L - \hat{\theta})}{\hat{\sigma}}, \]
which then yields
\[ \Phi(z_{\omega_1} + m) = G_n \left[ \frac{n^{1/2} \sigma_x(t_L - \hat{\theta})}{\hat{\sigma}} \right], \]
so that
\[ \omega_1 = \Phi \left[ \Phi^{-1} \left\{ G_n \left[ \frac{n^{1/2} \sigma_x(t_L - \hat{\theta})}{\hat{\sigma}} \right] \right\} - m \right] = \Phi \left[ \Phi^{-1} \left\{ G_n \left[ \frac{n^{1/2} \sigma_x(t_L - \hat{\theta})}{\hat{\sigma}} \right] \right\} - m \right]. \tag{3.3.2} \]
Thus the observed confidence level based on the bias-corrected critical point is
\[ \alpha_{bc}(\Theta_k) = \omega_2 - \omega_1 = \Phi \left[ \Phi^{-1} \left\{ G_n \left[ \frac{n^{1/2} \sigma_x(t_L - \hat{\theta})}{\hat{\sigma}} \right] \right\} - m \right] \tag{3.3.3} \]
- \[ \Phi \left[ \Phi^{-1} \left\{ G_n \left[ \frac{n^{1/2} \sigma_x(t_L - \hat{\theta})}{\hat{\sigma}} \right] \right\} - m \right]. \tag{3.3.4} \]

To calculate the order of accuracy for the observed confidence level based on the bias-corrected critical points, set \( \alpha = \omega_2 - \omega_1 \) and let \( \Theta_k = C_{stud}(\omega_1, \omega_2) \). Thus
\[ \alpha_{bc}(C_{stud}(\omega_1, \omega_2)) = \Phi \left[ \Phi^{-1} \left\{ G_n \left[ \frac{n^{1/2} \sigma_x(\hat{\theta}_{stud}(\omega_2) - \hat{\theta})}{\hat{\sigma}} \right] \right\} - m \right] \]
- \[ \Phi \left[ \Phi^{-1} \left\{ G_n \left[ \frac{n^{1/2} \sigma_x(\hat{\theta}_{stud}(\omega_1) - \hat{\theta})}{\hat{\sigma}} \right] \right\} - m \right] \tag{3.3.5} \]
- \[ \Phi \{ \Phi^{-1}[G_n(-h_{1-\omega_2})] - 2\Phi^{-1}[G_n(0)] \} \]
- \[ \Phi \{ \Phi^{-1}[G_n(-h_{1-\omega_1})] - 2\Phi^{-1}[G_n(0)] \}. \]

For the slope parameter, recall that \( p_1(t) = q_1(t) \). This yields
\[ G(-h_{1-\omega}) = \omega + 2n^{-1/2}p_1(z_{\omega})\phi(z_{\omega}) + O(n^{-1}). \tag{3.3.5} \]

Then
\[ \Phi^{-1}[G_n(-h_{1-\omega})] = z_{\omega} + 2n^{-1/2}p_1(z_{\omega}) + O(n^{-1}), \tag{3.3.6} \]
since
\[
\Phi^{-1}(\omega + \lambda) = z_\omega + \frac{\lambda}{\phi(z_\omega)} + O(\lambda^2);
\]
as \(\lambda \to 0\). In the same way
\[
G_n(0) = \Phi(0) + n^{-1/2}p_1(0)\phi(0) + O(n^{-1}); \tag{3.3.7}
\]
thus
\[
\Phi^{-1}[G_n(0)] = n^{-1/2}p_1(0) + O(n^{-1}). \tag{3.3.8}
\]
Thus combining equations (3.3.6) and (3.3.8) together yields
\[
\Phi\{\Phi^{-1}[G_n(-h_{1-\omega})] - 2\Phi^{-1}[G_n(0)]\} = \Phi\{z_\omega + 2n^{-1/2}[p_1(z_\omega) - p_1(0)] + O(n^{-1})\}
\]
\[
= \omega + 2n^{-1/2}[p_1(z_\omega) - p_1(0)]\phi(z_\omega) + O(n^{-1})
\]
\[
= \omega + O(n^{-1/2}). \tag{3.3.9}
\]
This yields
\[
\alpha_{bc}[C_{\text{stud}}(\omega_1, \omega_2)] = \omega_2 - \omega_1 + O(n^{-1/2}). \tag{3.3.10}
\]
Thus the bias-corrected method is first-order accurate.

The bootstrap version of the bias-corrected method is similar to the bootstrap of the backwards method, where
\[
\alpha_{bc}^*[C_{\text{stud}}(\omega_1, \omega_2)] = \Phi\{\Phi^{-1}[\hat{G}_n(-h_{1-\omega_2})] - 2\Phi^{-1}[\hat{G}_n(0)]\}
\]
\[
- \Phi\{\Phi^{-1}[\hat{G}_n(-h_{1-\omega_1})] - 2\Phi^{-1}[\hat{G}_n(0)]\}
\]
\[
= \Phi\{\Phi^{-1}[G_n(-h_{1-\omega_2})] - 2\Phi^{-1}[G_n(0)]\}
\]
\[
- \Phi\{\Phi^{-1}[G_n(-h_{1-\omega_1})] - 2\Phi^{-1}[G_n(0)]\} + O_p(n^{-1})
\]
\[
= \alpha + O_p(n^{1/2}). \tag{3.3.11}
\]
Hence, the bootstrap estimate of $\alpha_{bc}$ is first-order accurate in probability.

The accuracy of the bias-corrected method remains the same as the backwards method. This is because this method corrects for the constant terms in $p_1 + q_1$. However, since $p_1 = q_1$, this produces a critical point that is not fully corrected, so an accelerated bias corrected version is also considered. The accelerated bias correction introduces a skewness adjustment. The critical point for the slope parameter using the accelerated bias-corrected method is

$$\hat{\theta}_{abc}(\alpha) = \hat{\theta}_{back}[\psi(\alpha)] = \hat{\theta} + n^{-1/2} \hat{\sigma} g_{\psi(\alpha)}.$$ \hfill (3.3.12)

where

$$\psi(\alpha) = \Phi \left[ m + \frac{m + z_\alpha}{1 - a(m + z_\alpha)} \right].$$ \hfill (3.3.13)

The coefficient $a$ is the acceleration constant and from Hall (1992) given by

$$a = -n^{-1/2} \frac{1}{3} \gamma_x = 2n^{-1/2} z_\alpha [p_1(z_\alpha) - p_1(0)].$$ \hfill (3.3.14)

Setting $\Theta_k = C_{abc}(\omega_1, \omega_2) = \{t_L, t_U\}$ yields

$$t_L = \hat{\theta} + n^{-1/2} \sigma_x^{-1} g_{\psi(\omega_1)}.$$

Therefore

$$\Phi \left[ m + \frac{m + z_{\omega_1}}{1 - a(m + z_{\omega_1})} \right] = G_n \left[ \frac{n^{1/2} \sigma_x (t_L - \hat{\theta})}{\hat{\sigma}} \right],$$

and then

$$z_{\omega_1} = \frac{\Phi^{-1} \left[ G_n \left( \frac{n^{1/2} \sigma_x (t_L - \hat{\theta})}{\hat{\sigma}} \right) \right] - m}{1 + a \left\{ \Phi^{-1} \left[ G_n \left( \frac{n^{1/2} \sigma_x (t_L - \hat{\theta})}{\hat{\sigma}} \right) \right] - m \} - m},$$ \hfill (3.3.15)

such that

$$\omega_1 = \Phi \left\{ \frac{\Phi^{-1} \left[ G_n \left( \frac{n^{1/2} \sigma_x (t_L - \hat{\theta})}{\hat{\sigma}} \right) \right] - m}{1 + a \left\{ \Phi^{-1} \left[ G_n \left( \frac{n^{1/2} \sigma_x (t_L - \hat{\theta})}{\hat{\sigma}} \right) \right] - m \} - m \right\}. \hfill (3.3.16)
Then the observed confidence level of the slope parameter based on the accelerated bias-corrected critical point is

\[
\alpha_{abc}(\Theta_k) = \omega_2 - \omega_1
\]

\[
= \Phi \left\{ \frac{\Phi^{-1} \left[ G_n \left( \frac{n^{1/2} \sigma_\delta (t_U - \tilde{\theta})}{\hat{m}} \right) - m \right]}{1 + a \left\{ \Phi^{-1} \left[ G_n \left( \frac{n^{1/2} \sigma_\delta (t_U - \tilde{\theta})}{\hat{m}} \right) - m \right] \right\} - m} \right\}
\]

(3.3.17)

To obtain the order of accuracy we will calculate \( \alpha[C_{\text{stud}}(\omega_1, \omega_2)] \), which is

\[
\alpha_{abc}[C_{\text{stud}}(\omega_1, \omega_2)] =
\]

\[
\Phi \left\{ \frac{\Phi^{-1} \left[ G_n \left( -h_{1-\omega_2} \right) - m \right]}{1 + a \left\{ \Phi^{-1} \left[ G_n \left( -h_{1-\omega_2} \right) - m \right] \right\} - m} \right\}
\]

(3.3.18)

where

\[
a\{\Phi^{-1}[G_n(-h_{1-\omega_2}) - m]\} = 2n^{-1/2}z_{\omega_2}^{-1}[p_1(z_{\omega_2}) - p_1(0)] + O(n^{-1})
\]

(3.3.19)

and by a Taylor expansion

\[
[1 + a\{\Phi^{-1}[G_n(-h_{1-\omega_2}) - m]\}]^{-1} = 1 - 2n^{-1/2}z_{\omega_2}^{-1}[p_1(z_{\omega_2}) - p_1(0)] + O(n^{-1}).
\]

(3.3.20)

Thus

\[
\{\Phi^{-1}[G_n(-h_{1-\omega_2})] - m\}[1 + a\{\Phi^{-1}[G_n(-h_{1-\omega_2})] - m\}]^{-1}
\]

\[
= z_{\omega_2} - 2n^{-1/2}[p_1(z_{\omega_2}) - p_1(0)] + n^{-1/2}[2p_1(z_{\omega_2}) - p_1(0)] + O(n^{-1})
\]

\[
= z_{\omega_2} + n^{-1/2}p_1(0) + O(n^{-1}),
\]

(3.3.21)
so that
\[
\frac{\Phi^{-1}[G_n(-h_{1-\omega_2})] - m}{1 + a\{\Phi^{-1}[G_n(-h_{1-\omega_2})] - m\}} - m = z_{\omega_2} + n^{-1/2}p_1(0) - n^{-1/2}p_1(0) + O(n^{-1}) \\
= z_{\omega_2} + O(n^{-1}).
\] (3.3.22)

Therefore
\[
\alpha_{abc}[C_{\text{stud}}(\omega_1, \omega_2)] = \omega_2 - \omega_1 + O(n^{-1}) \\
= \alpha + O(n^{-1}).
\] (3.3.23)

Thus \(\alpha_{abc}\) is second-order accurate.

The bootstrap version of \(\alpha_{abc}\) is
\[
\alpha_{abc}^*(\Theta_k) = \omega_2 - \omega_1 \\
= \Phi \left\{ \frac{\Phi^{-1}\left[\hat{G}_n\left(\frac{n^{1/2}\sigma(x(t_{1/2} - \hat{\theta}))}{\hat{\delta}}\right) - \hat{m}\right]}{1 + \hat{a}\left\{\Phi^{-1}\left[\hat{G}_n\left(\frac{n^{1/2}\sigma(x(t_{1/2} - \hat{\theta}))}{\hat{\delta}}\right) - \hat{m}\right]\right\}} - \hat{m} \right\} - \hat{m} \\
- \Phi \left\{ \frac{\Phi^{-1}\left[\hat{G}_n\left(\frac{n^{1/2}\sigma(x(t_{1/2} - \hat{\theta}))}{\hat{\delta}}\right) - \hat{m}\right]}{1 + \hat{a}\left\{\Phi^{-1}\left[\hat{G}_n\left(\frac{n^{1/2}\sigma(x(t_{1/2} - \hat{\theta}))}{\hat{\delta}}\right) - \hat{m}\right]\right\}} - \hat{m} \right\} - \hat{m}.
\] (3.3.24)

Here \(\hat{a} = 2n^{-1/2}z_0[\hat{p}_1(z_0) - \hat{p}_1(0)]\) and \(\hat{m} = \phi(t)G_n(0)\). Recall from Equation (3.2.9) that \(\hat{p}_1(t) = p_1(t) + O_p(n^{-1/2})\) and \(\hat{G}_n(t) = G_n(t) + O_p(n^{-1/2})\); this yields
\[
\hat{a} = a + O_p(n^{-1})
\]
and
\[
\hat{m} = m + O_p(n^{-1}).
\]

Thus
\[
\alpha_{abc}^*[C_{\text{stud}}(\omega_1, \omega_2)] = \alpha + O_p(n^{-1}).
\] (3.3.25)

Thus, the bootstrap version of \(\alpha_{abc}\) is second-order accurate in probability.
3.4 Intercept Parameter

The asymptotic properties of the confidence intervals of the slope parameter do not apply to the intercept parameter. Thus, the observed confidence levels for the intercept parameter are calculated differently. Consider again a simple linear model where \( k = 1 \). Let

\[ y_i = \beta_0 + x_i \beta_1 + \epsilon_i \]  

for \( i = 1, \ldots, n \). We desire to find the observed confidence levels for the intercept parameter \( \beta_0 \). The mean of \( Y \) given an observed value of \( x_0 \) is

\[ E(Y|x = x_0) = y_0 = \beta_0 + x_0 \beta_1. \]  

Then the estimate of \( y_0 \) is

\[ \hat{y}_0 = \hat{\beta}_0 + x_0 \hat{\beta}_1. \]  

Letting \( x_0 = 0 \), we get \( \hat{y}_0 = \hat{\beta}_0 \). Thus, the intercept parameter is the same as the expected value of \( y_0 \) given that \( x_0 = 0 \). Therefore, this discussion will look at the estimate of the mean of \( Y \). From Hall (1992), define for an estimation of a general \( y_0 \),

\[ y_i = \sigma_x^{-2} \{(x_0 - \bar{x})(x_i - \bar{x}) + \sigma_x^2\}, \]  

for \( i \in \{1, \ldots, n\} \), then

\[ \sigma_y^2 = n^{-1} \sum_{i=1}^{n} y_i^2 \]

\[ = n^{-1} \sum_{i=1}^{n} \left[ \frac{(x_0 - \bar{x})(x_i - \bar{x})}{\sigma_x^4} + 2(x_0 - \bar{x})(x_i - \bar{x}) + 1 \right] \]  

\[ = 1 + \sigma_x^{-2} (x_0 - \bar{x})^2. \]
Let
\[\gamma_y = n^{-1} \sum_{i=1}^{n} y_i^2,\]
\[\kappa_y = n^{-1} \sum_{i=1}^{n} y_i^4 - 3,\]
and
\[V(y_0) = n^{-1} \sigma^2_2 \sigma_y^2.\]

Thus define the distributions \(G_n\) and \(H_n\) as
\[G_n(t) = P\left[ \frac{n^{1/2}(\hat{y}_0 - y_0)}{\sigma \sigma_y} \leq t \right]\]
and
\[H_n(t) = P\left[ \frac{n^{1/2}(\hat{y}_0 - y_0)}{\sigma \sigma_y} \leq t \right].\]

Also, define \(g_\alpha = G_n^{-1}(\alpha)\) and \(h_\alpha = H_n^{-1}(\alpha)\) for \(\alpha \in (0,1)\). The Edgeworth expansions for \(G_n\) and \(H_n\) are
\[G_n(t) = \Phi(t) + \sum_{i=1}^{\nu} n^{-1/2} p_i(t) \phi(t) + O(n^{-(\nu+1)/2})\]
and
\[H_n(t) = \Phi(t) + \sum_{i=1}^{\nu} n^{-1/2} q_i(t) \phi(t) + O(n^{-(\nu+1)/2}).\]

where
\[p_i(t) = -\frac{1}{6} \gamma_y (t^2 - 1)\]
and
\[q_i(t) = -\frac{1}{6} \gamma (\gamma_y - 3\sigma_y^{-1}) t^2 - \frac{1}{6} \gamma \gamma_y.\]

From these functions we can get the Cornish-Fisher expansions of \(g_\alpha\) and \(h_\alpha\), given by
\[g_\alpha = z_\alpha - n^{-1/2} p_1(z_\alpha) \phi(z_\alpha) + O(n^{-1})\]
(3.4.10)
and

\[ h_\alpha = z_\alpha - n^{-1/2} q_1(z_\alpha) \phi(z_\alpha) + O(n^{-1}) \]  \hspace{1cm} (3.4.11)

Define the theoretical critical points for \( \hat{y}_0 \) as

\[ \hat{\theta}_{\text{ord}}(\alpha) = \hat{y}_0 - n^{-1/2} \sigma_y \sigma g_{1-\alpha} \] \hspace{1cm} (3.4.12)

when \( \sigma \) is known. When \( \sigma \) is unknown, define

\[ \hat{\theta}_{\text{stud}}(\alpha) = \hat{y}_0 - n^{-1/2} \sigma_y \hat{\sigma} h_{1-\alpha} \] \hspace{1cm} (3.4.13)

\[ \hat{\theta}_{\text{hyb}}(\alpha) = \hat{y}_0 - n^{-1/2} \sigma_y \hat{\sigma} g_{1-\alpha} \] \hspace{1cm} (3.4.14)

and

\[ \hat{\theta}_{\text{back}}(\alpha) = \hat{y}_0 + n^{-1/2} \sigma_y \hat{\sigma} g_\alpha \] \hspace{1cm} (3.4.15)

Given \( \omega_2 - \omega_1 = \alpha \), set \( \Theta_k = C_{\text{ord}}(\omega_1, \omega_2) \) then let

\[ t_L = \hat{y}_0 - n^{-1/2} \sigma_y \hat{\sigma} g_{1-\omega_1} \]

and

\[ t_U = \hat{y}_0 - n^{-1/2} \sigma_y \hat{\sigma} g_{1-\omega_2} \]

Thus the observed confidence levels that correspond to the theoretical observed confidence limits of \( \hat{y}_0 \) are

\[ \alpha_{\text{ord}}(\Theta_k) = G_n \left[ \frac{n^{1/2}(\hat{y}_0 - t_L)}{\sigma \sigma_y} \right] - G_n \left[ \frac{n^{1/2}(\hat{y}_0 - t_U)}{\sigma \sigma_y} \right], \]  \hspace{1cm} (3.4.16)

\[ \alpha_{\text{stud}}(\Theta_k) = H_n \left[ \frac{n^{1/2}(\hat{y}_0 - t_L)}{\hat{\sigma} \sigma_y} \right] - H_n \left[ \frac{n^{1/2}(\hat{y}_0 - t_U)}{\hat{\sigma} \sigma_y} \right], \]  \hspace{1cm} (3.4.17)

\[ \alpha_{\text{hyb}}(\Theta_k) = G_n \left[ \frac{n^{1/2}(\hat{y}_0 - t_L)}{\hat{\sigma} \sigma_y} \right] - G_n \left[ \frac{n^{1/2}(\hat{y}_0 - t_U)}{\hat{\sigma} \sigma_y} \right], \]  \hspace{1cm} (3.4.18)

\[ \alpha_{\text{back}}(\Theta_k) = G_n \left[ \frac{n^{1/2}(t_U - \hat{y}_0)}{\hat{\sigma} \sigma_y} \right] - G_n \left[ \frac{n^{1/2}(t_L - \hat{y}_0)}{\hat{\sigma} \sigma_y} \right], \]  \hspace{1cm} (3.4.19)
\[ \alpha_{abc}(\Theta_k) = \Phi \left\{ \frac{\Phi^{-1} \left[ G_n \left( \frac{n^{1/2}(y_0 - \hat{y})}{\hat{\sigma}_y} \right) \right] - m}{1 + a \left\{ \Phi^{-1} \left[ G_n \left( \frac{n^{1/2}(y_0 - \hat{y})}{\hat{\sigma}_y} \right) \right] - m \right\}} - m \right\} \]  

(3.4.20)

and

\[ \alpha_{bc}(\Theta_k) = \Phi \left\{ \frac{\Phi^{-1} \left[ G_n \left( \frac{n^{1/2}(y_0 - \hat{y})}{\hat{\sigma}_y} \right) \right] - m}{1 + \Phi^{-1} \left[ G_n \left( \frac{n^{1/2}(y_0 - \hat{y})}{\hat{\sigma}_y} \right) \right] - m} \right\} \]  

(3.4.21)

The normal and bootstrap approximations are similar to that of the slope parameter.

The asymptotic accuracy of observed confidence limits for the theoretical confidence intervals of \( y_0 \) are somewhat similar to the accuracy for the slope parameter. For the normal approximations, \( \alpha_{ord} \) and \( \alpha_{stud} \) are first-order accurate as it was with the slope parameter. Also, \( \alpha_{abc} \) is second-order accurate and \( \alpha_{back} \) and \( \alpha_{bc} \) are both first-order accurate. These accuracies are the same as for the slope parameter. However, notice that \( \alpha_{hyb} \) is first-order accurate instead of second-order accurate as in the case of the slope parameter. Since \( p_1 \neq q_1 \) for \( y_0 \), thus

\[ \alpha_{hyb}[C_{stud}(\omega_1, \omega_2)] = G_n \left\{ \frac{n^{1/2}[y_0 - \hat{\theta}_{stud}(\omega_1)]}{\hat{\sigma}_y} \right\} - G_n \left\{ \frac{n^{1/2}[y_0 - \hat{\theta}_{stud}(\omega_2)]}{\hat{\sigma}_y} \right\} = G_n(h_{1-\omega_1}) - G_n(h_{1-\omega_2}). \]  

(3.4.22)

Notice that

\[ G_n(h_{1-\omega}) = 1 - \omega + n^{-1/2}[p_1(z_\omega) - q_1(z_\omega)]\phi(z_\omega) + O(n^{-1}). \]  

(3.4.23)

Since \( p_1 \neq q_1 \), the \( n^{-1/2} \) term does not become zero as in the case of the slope
parameter; thus,

\[
\alpha_{\text{hyb}}[C_{\text{stud}}(\omega_1, \omega_2)] = 1 - \omega_1 + n^{-1/2}[p_1(z_{\omega_1}) - q_1(z_{\omega_1})]\phi(z_{\omega_1}) \\
- \quad 1 + \omega_2 - n^{-1/2}[p_1(z_{\omega_2}) - q_1(z_{\omega_2})]\phi(z_{\omega_2}) + O(n^{-1}) \\
= \alpha + O(n^{-1/2}). \quad \text{(3.4.24)}
\]

The bootstrap estimates \(\hat{\alpha}_{\text{ord}}^*, \hat{\alpha}_{\text{stud}}^*,\) and \(\hat{\alpha}_{\text{abc}}^*\) are second-order accurate in probability, and \(\hat{\alpha}_{\text{hyb}}^*\) and \(\hat{\alpha}_{\text{back}}^*\) are first-order accurate in probability.

### 3.5 Example

Chikae et al. (2007) described an alternative technique to obtain measurements for compost maturity. A good measurement of compost maturity is called the germination index (GI). In order to calculate the GI, a water extract is taken from compost samples and put into a plastic petri dish with a filter paper. Then seeds of a certain plant are distributed on the filter paper and incubated for 48 hours. Other seeds of the same type are germinated in distilled water as a control. Then the GI calculation is

\[
GI(\%) = \frac{\text{Seed germination} \times \text{root length of treatment} \times 100}{\text{Seed germination} \times \text{root length of control}}. \quad \text{(3.5.1)}
\]

The waiting period and the complicated measurements makes it difficult to measure the germination index. Thus Chikae et al. (2007) developed a maturity sensor system based on the combination of three electrically measured parameters, pH, \(\text{NH}_4^+\) concentration, and phosphatase activity, in the water extracts of compost samples to estimate the germination index. Twenty-four compost samples were examined. Water was extracted from the samples and the variables pH, \(\text{NH}_4^+\),
phosphate activity and GI were measured to get the estimated germination index (EGI). To investigate the reliability of their system, Chikae et al. (2007) compared the germination index to the estimated germination index, found by evaluating the electrical parameters using a regression model. The germination index and estimated index for the 24 samples are plotted in Figure 3.1.

![Figure 3.1: The measured and estimated GI values.](image)

Notice that from Figure 3.1 the estimation is not a perfect predictor of the germination index. The least squares linear equation tends to overestimate for smaller values and underestimate for larger values of GI. One might ask, given this behavior, how much confidence is there that the slope of the regression line is less than 1? We can determine this confidence through the observed confidence levels.
as previously discussed. The least squares regression line plotted in Figure 3.1 is

\[ EGI = 10.912 + 0.804GI \]  \hspace{1cm} (3.5.2)

with \( \sigma_x = 22.3894 \) and \( \hat{\sigma} = 9.9531 \). Let the desired region be \( \Theta = (-\infty, 1) \), then the observed confidence level for the region \( \Theta \) using the normal approximation is

\[
\alpha(\Theta) = 1 - \Phi \left[ \frac{\sqrt{24(22.3894)(0.804 - 1)}}{9.9531} \right] \\
= 1 - \Phi(-2.160) \\
= 1 - 0.0154 = 0.9846. \hspace{1cm} (3.5.3)
\]

The bootstrap estimates of the observed confidence levels were calculated using R and the boot library. The R code is in the Appendix. The estimates of the observed confidence levels are \( \alpha_{\text{stud}} = 0.957 \), \( \alpha_{\text{hyb}} = 0.98 \) and \( \alpha_{\text{perc}} = 0.982 \). These show a large amount of confidence that the slope parameter is less than 1.
CHAPTER 4

OBSERVED CONFIDENCE LEVELS FOR MULTIPLE REGRESSION PARAMETERS

In this chapter we discuss different methods for obtaining observed confidence levels of regression parameters in a multiple regression model. From Hall (1992), define

\[ Y_i = \beta_0 + x_i \beta + \varepsilon_i, \]

where \( Y_i \) is a random variable, \( \beta_0 \) is the intercept parameter, \( \beta \) is a \( p \times 1 \) vector of unknown parameters, and \( x_i \) is a \( 1 \times p \) matrix. Also, \( \varepsilon_i \) are independent and identically distributed with \( E(\varepsilon_i) = 0 \) and variance \( E(\varepsilon_i^2) = \sigma^2 \). Then the least squares estimates of \( \beta_0, \beta \) and \( \sigma \) are

\[ \hat{\beta} = \Sigma^{-1} n^{-1} \sum_{i=1}^{n} (x_i - \bar{x})'(Y_i - \bar{Y}), \]
\[ \hat{\beta}_0 = \bar{Y} - \bar{x} \hat{\beta}, \]

and

\[ \hat{\sigma}^2 = n^{-1} \sum_{i=1}^{n} \varepsilon_i^2. \]

Define \( \bar{x} \) as \( 1 \times p \) matrix where

\[ \bar{x}_j = n^{-1} \sum_{i=1}^{n} x_{ij}. \]
Also, define $\Sigma_x$ as a $p \times p$ covariance matrix of $x_i$ where
\[
\Sigma_x = n^{-1} \sum_{i=1}^{n} (x_i - \bar{x})'(x_i - \bar{x})
\]
and the residual $\varepsilon_i$ is
\[
\varepsilon_i = Y_i - \hat{\beta}_0 - x_i\hat{\beta}.
\]

### 4.1 Confidence Regions

The intercept parameter $\beta_0$ is univariate as in the case in Section 3.4. Thus the observed confidence levels that were discussed in Section 3.4 are the same for $\beta_0$ in this model. In this chapter we will develop the observed confidence levels for the vector $\beta$, which is a generalization for the slope parameter. Let $\mathcal{C}(\alpha) \subset \Theta$ be a $100\alpha\%$ confidence region for $\beta$ based on the sample $\mathcal{X}$. As in Chapter 3, let $\hat{P}$ be the empirical distribution of the residuals $\varepsilon_i$ for $i = 1, \ldots, n$.

Hall (1992) defines the following:

\[
V = n^{-1} \sum_{i=1}^{n} (x_i - \bar{x})' \Sigma (x_i - \bar{x}) = \sigma^2 \Sigma_x
\]
and
\[
\hat{V} = n^{-1} \sum_{i=1}^{n} (x_i - \bar{x})' \hat{\Sigma} (x_i - \bar{x}) = \hat{\sigma}^2 \Sigma_x.
\]
Then $V(\hat{\beta}) = n^{-1}\sigma^2 \Sigma_x^{-1}$, thus the following standardized distributions are

\[
G_n(t) = P[n^{1/2}\sigma^{-1} \Sigma_x^{1/2}(\hat{\beta} - \beta) \leq t]
\]

and

\[
H_n(t) = P[n^{1/2}\hat{\sigma}^{-1} \Sigma_x^{1/2}(\hat{\beta} - \beta) \leq t],
\]

(4.1.1) (4.1.2)
where $t \in \mathbb{R}^p$. Let the $p$-dimensional densities of $G_n(t)$ and $H_n(t)$ be $g_n(t)$ and $h_n(t)$, respectively. Define $\mathcal{G}_\alpha$ to be any region of $\mathbb{R}^p$ that satisfies

$$P[n^{1/2}\sigma^{-1}\Sigma_x^{1/2}(\hat{\beta} - \beta) \in \mathcal{G}_\alpha] = \alpha$$

(4.1.3)

and define $\mathcal{H}_\alpha$ to be a region of $\mathbb{R}^p$ such that

$$P[n^{1/2}\hat{\sigma}^{-1}\Sigma_x^{1/2}(\hat{\beta} - \beta) \in \mathcal{H}_\alpha] = \alpha.$$  

(4.1.4)

In order to get the observed confidence levels, we first need to identify the confidence regions for $\beta$. These confidence regions are derived from the confidence regions developed by Polansky (2005b). If $\sigma$ is known, then a $100\alpha\%$ confidence region for $\beta$ that is a multivariate analog to the ordinary confidence interval is given by

$$C_{ord}(\alpha) = \{\hat{\beta} - n^{-1/2}\sigma \Sigma_x^{-1/2}t | t \in \mathcal{G}_\alpha\}.$$  

(4.1.5)

If $\sigma$ is unknown, then a $100\alpha\%$ confidence region for $\beta$ that is the multivariate analog of the Studentized confidence interval is given by

$$C_{stud}(\alpha) = \{\hat{\beta} - n^{-1/2}\hat{\sigma} \Sigma_x^{-1/2}t | t \in \mathcal{H}_\alpha\}.$$  

(4.1.6)

The region $\mathcal{H}_\alpha$ may not be known, but it might be common for the region $\mathcal{G}_\alpha$ to be known. If this is the case and $\sigma$ is unknown, then the multivariate analog for the hybrid confidence interval is the confidence region given by

$$C_{hyb}(\alpha) = \{\hat{\beta} - n^{-1/2}\hat{\sigma} \Sigma_x^{-1/2}t | t \in \mathcal{G}_\alpha\}. $$  

(4.1.7)

If both regions $\mathcal{G}_\alpha$ and $\mathcal{H}_\alpha$ are unknown, the normal approximation can be used. To use the normal approximation we assume the smooth function model and that

$$\frac{\partial g_n(t)}{\partial t_i} \bigg|_{t=\mu} \neq 0$$
where \( g(t) = [g_1(t), \ldots, g_p(t)]' \) and \( t = [t_1, \ldots, t_d]' \in \mathbb{R}^d \). From Serfling (1980) under these assumptions, it follows that

\[
n^{1/2} \sigma \Sigma_x^{-1/2}(\hat{\beta} - \beta) \xrightarrow{w} N_p(0, I) \quad (4.1.8)
\]

and

\[
n^{1/2} \hat{\sigma} \Sigma_x^{-1/2}(\hat{\beta} - \beta) \xrightarrow{w} N_p(0, I), \quad (4.1.9)
\]

as \( n \to \infty \), where \( N_p(0, I) \) is a \( p \)-variate normal distribution with mean vector at the origin and identity covariance matrix.

Let \( N_\alpha \) be any region in \( \mathbb{R}^p \) that satisfies the condition of \( P(Z \in N_\alpha) = \alpha \), where \( Z \sim N_p(0, I) \). Using the weak convergence from Equations (4.1.8) and (4.1.9), the confidence regions \( C_{\text{ord}} \) and \( C_{\text{stud}} \) can be approximated by

\[
\hat{C}_{\text{ord}}(\alpha) = \{\hat{\beta} - n^{-1/2} \sigma \Sigma_x^{-1/2} t : t \in N_\alpha\}
\]

and

\[
\hat{C}_{\text{stud}}(\alpha) = \{\hat{\beta} - n^{-1/2} \hat{\sigma} \Sigma_x^{-1/2} t : t \in N_\alpha\},
\]

where \( \sigma \) is known and unknown, respectively.

### 4.2 Observed Confidence Levels

To develop observed confidence levels of the multivariate parameters we will use the theory from Polansky (2005b). For the case when \( \sigma \) is known and based on the region \( C_{\text{ord}} \), the observed confidence level for \( \Theta_k \) is computed by solving

\[
\Theta_k = \{\hat{\beta} - n^{-1/2} \sigma \Sigma_x^{-1/2} t : t \in G_\alpha\} \quad \text{for } \alpha.
\]

Let \( n^{1/2} \sigma^{-1} \Sigma_x^{1/2}(\hat{\beta} - \Theta_k) \) be a linear transformation of \( \Theta_k \) given by \( \{n^{1/2} \sigma^{-1} \Sigma_x^{1/2}(\hat{\beta} - t) : t \in \Theta_k\} \). Then we want to find a value of \( \alpha \) such that

\[
n^{1/2} \sigma^{-1} \Sigma_x^{1/2}(\beta - \Theta_k) = G_\alpha.
\]
Thus if \( g_n(t) \) is known, then the observed confidence level for \( \Theta_k \) based on the region \( C_{\text{ord}} \) is given by

\[
\alpha_{\text{ord}}(\Theta_k) = \int_{\beta - \Theta_k \in \mathcal{C}_k} g_n(t)dt. \tag{4.2.1}
\]

Similarly, if \( \sigma \) is unknown, we want to find a value of \( \alpha \) such that

\[
n^{1/2} \sigma^{-1} \Sigma^{1/2}_k (\beta - \Theta_k) = \mathcal{H}_\alpha.
\]

Thus an observed confidence level for \( \Theta_k \) based on the region \( C_{\text{stud}} \), where \( h_n(t) \) is known, is given by

\[
\alpha_{\text{stud}}(\Theta_k) = \int_{\beta - \Theta_k \in \mathcal{C}_k} h_n(t)dt. \tag{4.2.2}
\]

However, if \( \sigma \) is unknown but \( g_n(t) \) is known, the observed confidence level for \( \Theta_k \) can be computed using the region \( C_{\text{hyb}} \), which is given by

\[
\alpha_{\text{hyb}}(\Theta_k) = \int_{\beta - \Theta_k \in \mathcal{C}_k} g_n(t)dt. \tag{4.2.3}
\]

If both \( g_n(t) \) and \( h_n(t) \) are unknown, we can use the normal approximation to compute the observed confidence levels. For \( \sigma \) known, the approximate observed confidence level can be computed as

\[
\hat{\alpha}_{\text{ord}}(\Theta_k) = \int_{\beta - \Theta_k \in \mathcal{C}_k} \phi_p(t)dt,
\]

where \( \phi_p(t) \) is the \( p \)-variate standard normal density. Also, if \( \sigma \) is unknown, an approximate observed confidence level is given by

\[
\hat{\alpha}_{\text{stud}}(\Theta_k) = \int_{\beta - \Theta_k \in \mathcal{C}_k} \phi_p(t)dt.
\]

If the normal approximations are not accurate enough, the bootstrap method can be used to estimate the observed confidence intervals. Let \( \{\varepsilon_{1k}, \ldots, \varepsilon_{nk}\} \) be an
independent and identically distributed sample from \( \hat{F}_n \), the empirical distribution function of \( \hat{\varepsilon}_1, \ldots, \hat{\varepsilon}_n \). This sample produces the estimated residuals

\[
\hat{\varepsilon}_i = Y_i - \hat{\beta}_0 - x_i \hat{\beta},
\]

where \( \beta_0 \) is the intercept parameter and \( \beta \) is the generalized form of slope. The bootstrap multivariate regression model is

\[
Y_i^* = \hat{\beta}_0 + x_i \hat{\beta} + \varepsilon_i^*,
\]

where \( i = 1, \ldots, n \), and the bootstrap estimate of \( \beta \) is

\[
\hat{\beta}^* = \Sigma^{-1} \sum_{i=1}^n (x_i - \bar{x})(Y_i^* - \bar{Y}^*).
\]

The bootstrap estimate of \( G_n \) is given by

\[
\hat{G}_n(t) = P[n^{1/2} \hat{\sigma}^{-1} \Sigma_{x}^{1/2}(\hat{\beta}^* - \bar{\beta}) \leq t|\varepsilon_1^*, \ldots, \varepsilon_n^* \sim \hat{F}, x_i, Y_i].
\]

(4.2.4)

Similarly, the bootstrap estimate of \( H_n \) is given by

\[
\hat{H}_n(t) = P[n^{1/2} \hat{\sigma}^{-1} \Sigma_{x}^{1/2}(\hat{\beta}^* - \bar{\beta}) \leq t|\varepsilon_1^*, \ldots, \varepsilon_n^* \sim \hat{F}, x_i, Y_i],
\]

(4.2.5)

where \( \hat{\sigma}^{-1} = n^{-1} \sum_{i=1}^n \varepsilon_i^2 \). Let \( \hat{g}_n(t) \) and \( \hat{h}_n(t) \) be the corresponding mass functions of \( \hat{G}_n(t) \) and \( \hat{H}_n(t) \), respectively. The bootstrap estimates of the observed confidence levels for \( \beta \) are

\[
\hat{\alpha}_{\text{ord}}^*(\Theta_k) = \int_{n^{1/2} \hat{\sigma}^{-1} \Sigma_{x}^{1/2}(\hat{\beta} - \Theta_k)}^{\hat{G}_n(t)} \hat{g}_n(t)dt, \quad (4.2.6)
\]

\[
\hat{\alpha}_{\text{stud}}^*(\Theta_k) = \int_{n^{1/2} \hat{\sigma}^{-1} \Sigma_{x}^{1/2}(\hat{\beta} - \Theta_k)}^{\hat{H}_n(t)} \hat{h}_n(t)dt, \quad (4.2.7)
\]

and

\[
\hat{\alpha}_{\text{byb}}^*(\Theta_k) = \int_{n^{1/2} \hat{\sigma}^{-1} \Sigma_{x}^{1/2}(\hat{\beta} - \Theta_k)}^{\hat{g}_n(t)} \hat{g}_n(t)dt. \quad (4.2.8)
\]
Another measure of confidence introduced by Efron and Tibshirani (1998) is based on the percentile bootstrap confidence interval, also called the backward confidence interval. Let \( V_n(t) = P(\hat{\beta} \leq t | \epsilon \sim F) \) with density \( v_n(t) \). Then the bootstrap estimate of \( V_n(t) \) is \( \hat{V}_n(t) = P(\hat{\beta}^* \leq t | \hat{\epsilon}^* \sim \hat{F}, X, Y) \) with density \( \hat{v}_n(t) \). Then any region \( V_\alpha \) is a bootstrap percentile confidence region for \( \beta \) as long as

\[
\int_{V_\alpha} \hat{v}_n(t) dt = \alpha.
\]

Setting \( V_\alpha \) to the desired region \( \Theta_k \) yields the observed confidence level based on the percentile bootstrap method:

\[
\hat{\alpha}_\text{perc}(\Theta_k) = \int_{\Theta_k} \hat{v}_n(t) dt = P(\hat{\beta}^* \in \Theta_k | \hat{\epsilon} \sim \hat{F}, X, Y) = P[n^{1/2} \hat{\sigma}^{-1} \Sigma_z^{1/2} (\hat{\beta}^* - \hat{\beta}) \in n^{1/2} \hat{\sigma}^{-1} \Sigma_z^{1/2} (\Theta_k - \hat{\beta}) | \hat{\epsilon}^*, X, Y] = \int n^{1/2} \hat{\sigma}^{-1} \Sigma_z^{1/2} (\Theta_k - \hat{\beta}) \hat{g}_n(t) dt. \tag{4.2.9}
\]

### 4.3 Asymptotic Analysis

These observed confidence levels are based on different theoretical confidence regions. Thus, we want to compare the methods for computing observed confidence levels by a measure of accuracy. Given a region \( \Theta_k \) that corresponds to a 100\( \alpha \)% confidence region for \( \beta \), the observed confidence level should be \( \alpha \). When \( \sigma \) is known, the confidence region \( C_{\text{ord}} \) will be used as the standard confidence region for \( \beta \). Then an observed confidence level \( \hat{\alpha} \) is accurate if \( \hat{\alpha}[C_{\text{ord}}(\alpha)] = \alpha \). Similarly, if \( \sigma \) is unknown, the confidence region \( C_{\text{stud}} \) will be used as the standard confidence region for \( \beta \). Thus the measure \( \hat{\alpha} \) is accurate if \( \hat{\alpha}[C_{\text{stud}}(\alpha)] = \alpha \). If \( \hat{\alpha} \) is not accurate
but \(\hat{\alpha}[C_{\text{ord}}(\alpha)] = \alpha + O(n^{-k/2})\) when \(\sigma\) is known or \(\hat{\alpha}[C_{\text{stud}}(\alpha)] = \alpha + O(n^{-k/2})\) when
\(\sigma\) is unknown as \(n \to \infty\), this measure is called \(k\)-th order accurate.

To calculate the order of accuracy we will use the Edgeworth expansion theory based on Bhattacharya and Ghosh (1978) and Hall (1992). Recall that in Chapter 3 for the univariate case we used

\[
G_n(t) = P \left[ \frac{n^{1/2} \sigma_x (\hat{\theta} - \theta)}{\sigma} \leq t \right] = \Phi(t) + \sum_{i=1}^{\nu} n^{-i/2} p_i(t) \phi(t) + O(n^{-(\nu+1)/2}).
\]

For the multivariate case, if a set \(S \subset \mathbb{R}^p\) is a finite union of convex sets, then

\[
P[n^{1/2} \sigma^{-1} \Sigma_x^{1/2} (\hat{\beta} - \beta) \in S] = \int_S \{1 + \sum_{i=1}^{\nu} n^{-i/2} r_i(t)\} \phi(t) dt + o(n^{-\nu/2}), \quad (4.3.1)
\]

where

\[
r_i(t) = \frac{d}{dt} [p_i(t) \phi(t)] \phi(t)^{-1}.
\]

Similarly, for \(\sigma\) unknown,

\[
P[n^{1/2} \hat{\sigma}^{-1} \Sigma_x^{1/2} (\hat{\beta} - \beta) \in S] = \int_S \{1 + \sum_{i=1}^{\nu} n^{-i/2} s_i(t)\} \phi(t) dt + o(n^{-\nu/2}), \quad (4.3.2)
\]

where

\[
s_i(t) = \frac{d}{dt} [q_i(t) \phi(t)] \phi(t)^{-1}.
\]

The polynomials \(r_i\) and \(s_i\) have degree of \(3i\) with coefficients that depend on the moments of \(F\). These functions are even when \(i\) is even and odd when \(i\) is odd.

First, we will look at the normal approximations of the observed confidence level, \(\hat{\alpha}_{\text{ord}}\) and \(\hat{\alpha}_{\text{stud}}\). For \(\sigma\) known, if \(n^{1/2} \sigma^{-1} \Sigma_x^{1/2} [\hat{\beta} - C_{\text{ord}}(\alpha)]\) is a finite union of convex sets, then by Equation (4.3.1),

\[
\int_{n^{1/2} \sigma^{-1} \Sigma_x^{1/2} [\hat{\beta} - C_{\text{ord}}(\alpha)]} \phi_p(t) dt + n^{-1/2} \int_{n^{1/2} \sigma^{-1} \Sigma_x^{1/2} [\hat{\beta} - C_{\text{ord}}(\alpha)]} r_1(t) \phi_p(t) dt = \alpha + o(n^{-1/2}),
\]
which implies

$$\hat{\alpha}_{\text{ord}}[C(\alpha)] = \int_{n^{1/2} \sigma^{-1} \Sigma_x^{1/2} [\hat{\vartheta} - C_{\text{ord}}(\alpha)]} \phi_p(t) dt = \alpha + O(n^{-1/2}). \quad (4.3.3)$$

Similarly when \( \sigma \) is unknown and \( n^{1/2} \sigma^{-1} \Sigma_x^{1/2} [\hat{\vartheta} - C_{\text{stud}}(\alpha)] \) is a finite union of convex sets, then equation (4.3.2) implies

$$\hat{\alpha}_{\text{stud}}[C_{\text{stud}}(\alpha)] = \int_{n^{1/2} \sigma^{-1} \Sigma_x^{1/2} [\hat{\vartheta} - C_{\text{stud}}(\alpha)]} \phi_p(t) dt = \alpha + O(n^{-1/2}). \quad (4.3.4)$$

Hence \( \hat{\alpha}_{\text{ord}} \) and \( \hat{\alpha}_{\text{stud}} \) are both first-order accurate.

Second, to find the order of accuracy for \( \alpha_{\text{hyb}} \), notice that \( \sigma \) is unknown, so we will use \( C_{\text{stud}}(\alpha) \) in place of \( \Theta_k \) for the desired region. Then from Equation (4.1.6),

$$n^{1/2} \sigma^{-1} \Sigma_x^{1/2} [\hat{\vartheta} - C_{\text{stud}}(\alpha)] = \mathcal{H}_\alpha.$$  

Thus

$$\alpha_{\text{hyb}}[C_{\text{stud}}(\alpha)] = \int_{\mathcal{H}_\alpha} g_n(t) dt$$

$$= \int_{\mathcal{H}_\alpha} \phi_p(t) dt + n^{-1/2} \int_{\mathcal{H}_\alpha} r_1(t) \phi_p(t) dt + O(n^{-1}). \quad (4.3.5)$$

Equation (4.3.2) implies

$$\int_{\mathcal{H}_\alpha} \phi_p(t) dt = \alpha - n^{-1/2} \int_{\mathcal{H}_\alpha} s_1(t) \phi_p(t) dt + O(n^{-1}). \quad (4.3.6)$$

Hall (1992) shows that \( s_1 = r_1 \), thus combining Equations (4.3.5) and (4.3.6) yields

$$\alpha_{\text{hyb}}[C_{\text{stud}}(\alpha)] = \alpha + O(n^{-1}). \quad (4.3.7)$$

Hence \( \alpha_{\text{hyb}} \) is second-order accurate, which matches the accuracy of the univariate case.
If \( G_n(t) \) and \( H_n(t) \) are unknown and the normal approximations are not accurate enough, the bootstrap estimates of the observed confidence levels can be used. To find the accuracy of the observed confidence levels based on the bootstrap estimates we will use Edgeworth expansions. From Hall (1992), under the smoothness and conditions on the moments of \( F \), if \( S \) is a finite union of convex sets, then

\[
P[n^{1/2} \sigma^{-1} \sum_x^{1/2} (\hat{\beta}^* - \hat{\beta}) \in S | \varepsilon \sim \hat{F}, X, Y] = \int_S \hat{g}_n(t) dt
\]

and

\[
P[n^{1/2} \sigma^{-1} \sum_x^{1/2} (\hat{\beta}^* - \hat{\beta}) \in S | \varepsilon \sim \hat{F}, X, Y] = \int_S \hat{h}_n(t) dt
\]

The functions \( \hat{r}_i \) and \( \hat{s}_i \) are bootstrap versions of \( r_i \) and \( s_i \) for \( i \geq 1 \). Thus \( \hat{r}_i(t) = r_i(t) + O_p(n^{-1/2}) \) and \( \hat{s}_i(t) = s_i(t) + O_p(n^{-1/2}) \) for all \( t \in \mathbb{R}^p \). From Equations (4.3.8) and (4.3.9) it follows that

\[
P[n^{1/2} \sigma^{-1} \sum_x^{1/2} (\hat{\beta}^* - \hat{\beta}) \in S | \varepsilon \sim \hat{F}, X, Y] = \int_S \phi_p(t) dt + n^{-1/2} \int_S r_1(t) \phi_p(t) dt + O_p(n^{-1}) \quad (4.3.10)
\]

and

\[
P[n^{1/2} \sigma^{-1} \sum_x^{1/2} (\hat{\beta}^* - \hat{\beta}) \in S | \varepsilon \sim \hat{F}, X, Y] = \int_S \phi_p(t) dt + n^{-1/2} \int_S s_1(t) \phi_p(t) dt + O_p(n^{-1}). \quad (4.3.11)
\]

If \( \sigma \) is known, from Equation (4.3.10) it follows that

\[
\hat{\sigma}^{*\text{ord}}[C_{\text{ord}}](\alpha) = \int_{G_0} \hat{g}_n(t) dt = \int_{G_0} \phi_p(t) dt + n^{-1/2} \int_{G_0} r_1(t) \phi_p(t) dt + O_p(n^{-1/2}).
\]
Equation (4.3.1) implies that
\[
\int_{\mathcal{G}_\alpha} \phi_p(t)dt = \alpha - n^{-1/2} \int_{\mathcal{G}_\alpha} r_i(t)\phi_p(t)dt + O(n^{-1}).
\]
Thus,
\[
\hat{\alpha}_{\text{ord}}^*[\mathcal{C}_{\text{ord}}(\alpha)] = \alpha - n^{-1/2} \int_{\mathcal{G}_\alpha} r_i(t)\phi_p(t)dt + n^{-1/2} \int_{\mathcal{G}_\alpha} r_i(t)\phi_p(t)dt + O_p(n^{-1})
\]
\[
= \alpha + O_p(n^{-1}).
\]

Similarly, when \(\sigma\) is unknown, it follows from Equation (4.3.11) that \(\hat{\alpha}_{\text{stud}}[\mathcal{C}_{\text{stud}}(\alpha)] = \alpha + O_p(n^{-1}).\) Thus both \(\hat{\alpha}_{\text{ord}}\) and \(\hat{\alpha}_{\text{stud}}\) are second-order accurate in probability.

To find the accuracy of the bootstrap estimate of the hybrid observed confidence level, we use Equation (4.3.10) to obtain
\[
\hat{\alpha}_{\text{hyb}}^*[\mathcal{C}_{\text{stud}}(\alpha)] = \int_{\mathcal{H}_\alpha} \hat{g}_n(t)dt
\]
\[
= \int_{\mathcal{H}_\alpha} \phi_p(t)dt + n^{-1/2} \int_{\mathcal{H}_\alpha} r_i(t)\phi(t)dt + O_p(n^{-1})
\]
\[
= \alpha + O_p(n^{-1}). \tag{4.3.12}
\]

Thus \(\hat{\alpha}_{\text{hyb}}\) is also second-order accurate.

The last bootstrap estimate of the observed confidence level is the percentile bootstrap method. Similar to the hybrid observed confidence interval we will use \(\mathcal{C}_{\text{stud}}(\alpha)\) as the standard region. Then, from Equation (4.2.9),
\[
n^{1/2} \hat{\sigma}^{-1} \Sigma_x^{1/2}[\mathcal{C}_{\text{stud}}(\alpha) - \hat{\beta}] = -\mathcal{H}_\alpha.
\]

Therefore Equation (4.3.10) implies
\[
\hat{\alpha}_{\text{perc}}^*[\mathcal{C}_{\text{stud}}(\alpha)] = \int_{-\mathcal{H}_\alpha} \hat{g}_n(t)dt
\]
\[
= \int_{-\mathcal{H}_\alpha} \phi_p(t)dt + n^{-1/2} \int_{-\mathcal{H}_\alpha} r_i(t)\phi(t)dt + O_p(n^{-1}). \tag{4.3.13}
\]
Because, $\phi_p(t)$ is an even function and $r_1(t)$ is an odd function, it follows that

$$\hat{\alpha}_{perc}[C_{stud}(\alpha)] = \int_{-\tau_0}^{\tau_0} \hat{g}_n(t)dt$$

$$= \int_{\tau_0}^{\tau_0} \phi_p(t)dt - n^{-1/2} \int_{\tau_0}^{\tau_0} r_1(t)\phi_p(t)dt + O_p(n^{-1}).$$

Then Equation (4.3.10) yields

$$\hat{\alpha}_{perc}[C_{stud}(\alpha)] = \alpha - n^{-1/2} \left[ \int_{\tau_0}^{\tau_0} r_1(t)\phi_p(t)dt + \int_{\tau_0}^{\tau_0} s_1(t)\phi_p(t)dt \right] + O_p(n^{-1})$$

$$= \alpha + O_p(n^{-1/2}).$$

Thus $\hat{\alpha}_{perc}$ is first-order accurate in probability.

### 4.4 Example

In a study that compared axe heads for felling trees, Mathieu and Meyer (1997) looked at data that used flint axes to fell three different types of trees. They compared sugar maple, oak and birch trees. The data for the oak and birch trees are from Jørgensen (1985). This study compared the type of tree and the diameter to the amount of time it took to fell the tree. The results that are similar to what was reported are presented in Figure 4.1. The data were modeled by regression lines for each tree type. The slope of each of these lines indicates how difficult it is to fell each of these tree types using a flint axe. Observed confidence levels can be constructed to measure the amount of confidence there is based on the order of most difficult type of tree to fell. That is, for example, maple is the most difficult tree to fell, followed by oak then birch.

Let $Y_{ij}$ be the observed time that it takes to fell a tree with diameter $x_{ij}$ for type $i$ tree, where $j = 1, \ldots, n_i$ and $i = 1, 2, 3$. The types of tree are indexed as maple.
(1), oak (2) and birch (3). In this study let \( n_i \) represent the number of trees of each type, where \( n_1 = n_2 = n_3 = 20 \). The linear regression model to fit a separate line for each type is given by \( Y = W_0 + X_0 + \epsilon \), where

\[
Y' = (Y_{11}, \ldots, Y_{1n_1}, Y_{21}, \ldots, Y_{2n_2}, Y_{31}, \ldots, Y_{3n_3}),
\]

\[
\beta'_0 = (\beta_{01}, \beta_{02}, \beta_{03}),
\]

\[
\beta' = (\beta_1, \beta_2, \beta_3),
\]

\[
W = \begin{bmatrix}
1_{n_1} & 0_{n_1} & 0_{n_1} \\
0_{n_2} & 1_{n_2} & 0_{n_2} \\
0_{n_3} & 0_{n_3} & 1_{n_3}
\end{bmatrix},
\]

and

\[
X = \begin{bmatrix}
x_1 & 0_{n_1} & 0_{n_1} \\
0_{n_2} & x_2 & 0_{n_2} \\
0_{n_3} & 0_{n_3} & x_3
\end{bmatrix},
\]

Figure 4.1: Felling time and breast height (diameter) for maple (o), oak (△), and birch (+). The lines are the least squares regression lines for each tree type: maple (solid line), oak (dashed line) and birch (dotted line).
Here, $\beta_0_i$ is the intercept parameter for each of the regression lines, $\beta_i$ is the slope parameter for each of the regression lines, $1_{n_i}$ and $0_{n_i}$ are $n_i \times 1$ vectors and $x_i' = (x_{i1}, \ldots, x_{im_i})$. Also, $\varepsilon$ is the error term where $E(\varepsilon) = 0$ and

$$\text{Var}(\varepsilon) = \Sigma = \begin{bmatrix} \sigma_1^2 I_{n_1} & 0_{n_1} & 0_{n_1} \\ 0_{n_2} & \sigma_2^2 I_{n_2} & 0_{n_2} \\ 0_{n_3} & 0_{n_3} & \sigma_3^2 I_{n_3} \end{bmatrix}$$

is the covariance matrix. Notice from Figure 4.1 that it strongly indicates that there is a different error variance for each type of tree. Thus we will use this covariance matrix. Therefore, we can use this model as an extension of the model we introduced in this chapter.

We desire to know which tree type is most difficult to fell, then the second most difficult, then the least. Thus the parameter of interest is the slope parameter for each of these regression lines. That is, if maple is least difficult, oak is the second difficult and birch is the most difficult to fell, then we have a region $\beta_1 < \beta_2 < \beta_3$. Thus there are six different regions that are possible, $\Theta_1 : \beta_1 < \beta_2 < \beta_3$, $\Theta_2 : \beta_1 < \beta_3 < \beta_2$, $\Theta_3 : \beta_2 < \beta_1 < \beta_3$, $\Theta_4 : \beta_2 < \beta_3 < \beta_1$, $\Theta_5 : \beta_3 < \beta_1 < \beta_2$ and $\Theta_6 : \beta_3 < \beta_2 < \beta_1$.

The estimates of $\beta_1$, $\beta_2$, and $\beta_3$ are obtained by fitting a linear regression line for each tree type. The observed estimates are given in Table 4.1. Under the assumption that $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$ are independent of each other, the asymptotic variance of $n^{1/2} \hat{\beta}$ is given by

$$\Sigma_T = \begin{bmatrix} \sigma_1^2 / \sigma_{x,1}^2 & 0 & 0 \\ 0 & \sigma_2^2 / \sigma_{x,2}^2 & 0 \\ 0 & 0 & \sigma_3^2 / \sigma_{x,3}^2 \end{bmatrix},$$

where

$$\sigma_{x,i}^2 = n_i^{-1} \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2$$
for $i = 1, 2, 3$. Then the covariance estimate is

$$\hat{\Sigma}_T = \begin{bmatrix} \hat{\sigma}_{11}^2 & 0 & 0 \\ 0 & \hat{\sigma}_{22}^2 & 0 \\ 0 & 0 & \hat{\sigma}_{33}^2 \end{bmatrix},$$

where

$$\hat{\sigma}_{ii}^2 = n_i^{-1} \sum_{j=1}^{n_i} \xi_{ij}^2$$

for each $\xi_i$ corresponding to the residuals for the separate fit of each tree type.

Table 4.1: Least squares estimates for fitting a linear regression to each tree type.

<table>
<thead>
<tr>
<th>$i$</th>
<th>Tree Type</th>
<th>$\hat{\beta}_0$</th>
<th>$\hat{\beta}_1$</th>
<th>$\hat{\beta}_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Maple</td>
<td>-36.080</td>
<td>4.646</td>
<td>4.939</td>
</tr>
<tr>
<td>2</td>
<td>Oak</td>
<td>-2.893</td>
<td>1.573</td>
<td>218.077</td>
</tr>
<tr>
<td>3</td>
<td>Birch</td>
<td>1.335</td>
<td>0.889</td>
<td>61.503</td>
</tr>
</tbody>
</table>

To obtain the observed confidence levels, define

$$H_n(t) = \Pr\left[n^{1/2} \hat{\Sigma}_T^{-1/2} (\hat{\beta} - \beta) \leq t| \varepsilon_{i1}, \ldots, \varepsilon_{in_i} \sim F_i, i = 1, 2, 3\right].$$

Then for any region $\Theta_k \subset \mathbb{R}^3$, for $k = 1, 2, 3, 4, 5, 6$, the observed confidence level based on a Studentized confidence region for $\beta$ is

$$\alpha_{\text{stud}}(\Theta_k) = \int_{n^{1/2} \hat{\Sigma}_T^{-1/2} (\beta - \hat{\Theta}_k)} h_n(t) dt,$$

where $h_n(t)$ is a three-dimensional density corresponding to $H_n(t)$. In this study $F_1, F_2,$ and $F_3$ are the empirical distribution functions of $\varepsilon_{i1}, \ldots, \varepsilon_{in_i}, i = 1, 2, 3$. These distribution functions are unknown, thus $h_n(t)$ is also unknown. The normal
The bootstrap estimate of $H_n(t)$ is given by

$$H_n(t) = P\left[n^{1/2} \hat{\Sigma}_T^{-1/2} (\hat{\beta} - \hat{\beta}) \leq t | \varepsilon_{11}^*, \ldots, \varepsilon_{m_t}^* \sim \hat{F}_i, i = 1, 2, 3\right],$$

where

$$\hat{\Sigma}_T^* = \begin{bmatrix} \hat{\sigma}_1^2 / \sigma_{x,1}^2 & 0 & 0 \\ 0 & \hat{\sigma}_2^2 / \sigma_{x,2}^2 & 0 \\ 0 & 0 & \hat{\sigma}_3^2 / \sigma_{x,3}^2 \end{bmatrix}.$$

The bootstrap estimate of the observed confidence level based on the Studentized region is

$$\hat{\alpha}_{\text{stud}}^*(\Theta_k) = \int_{n^{1/2} \hat{\Sigma}_T^{-1/2} (\beta - \Theta_k)} \phi(t) dt.$$
Table 4.2: Observed confidence levels for the six different regions of the ordering of the slope parameters.

<table>
<thead>
<tr>
<th>Region</th>
<th>Ordering</th>
<th>$\hat{\alpha}_{\text{stud}}$</th>
<th>$\hat{\alpha}^*_{\text{stud}}$</th>
<th>$\hat{\alpha}^*_{\text{hyb}}$</th>
<th>$\hat{\alpha}^*_{\text{perc}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Theta_1$</td>
<td>$\beta_1 &lt; \beta_2 &lt; \beta_3$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$\Theta_2$</td>
<td>$\beta_1 &lt; \beta_3 &lt; \beta_2$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$\Theta_3$</td>
<td>$\beta_2 &lt; \beta_1 &lt; \beta_3$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$\Theta_4$</td>
<td>$\beta_2 &lt; \beta_3 &lt; \beta_1$</td>
<td>0.079</td>
<td>0.138</td>
<td>0.037</td>
<td>0.044</td>
</tr>
<tr>
<td>$\Theta_5$</td>
<td>$\beta_3 &lt; \beta_1 &lt; \beta_2$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$\Theta_6$</td>
<td>$\beta_3 &lt; \beta_2 &lt; \beta_1$</td>
<td>0.921</td>
<td>0.862</td>
<td>0.963</td>
<td>0.956</td>
</tr>
</tbody>
</table>

4.5 Model Selection

In this section we will determine if the observed confidence levels can be used for model selection. Suppose we have a linear model

$$Y_i = \beta_0 + x_i \beta + \varepsilon_i,$$

where $\beta' = (\beta_1, \beta_2)$ and $\varepsilon_1, \ldots, \varepsilon_n \sim N(0, \sigma^2)$. Consider a region where $\beta_2 = 0$, that is, the region $\Theta = \{\beta : \beta_2 = 0\}$. Thus the region is a line along one of the axes in $\mathbb{R}^2$. It follows that

$$\hat{\alpha}_{\text{stud}}(\Theta) = \int_{n^{1/2} \hat{\sigma}^{-1} \Sigma_{x}^{1/2} (\hat{\beta} - \Theta)} \phi(t) dt = 0,$$

since $\phi(t)$ is a continuous density in $\mathbb{R}^2$. This will be true for any continuous density and for any model selection problem where we desire at least one of the regression parameters to be equal to zero. Thus the observed confidence levels cannot be used for model selection.
CHAPTER 5
EMPIRICAL STUDY

This chapter compares the asymptotic behavior of the observed confidence levels for the slope parameter of a simple linear regression model as discussed in Chapter 3. This comparison was done using computer-based simulations in R. The code to these simulations are in the Appendix. The simple linear regression model is

\[ Y_i = \beta_0 + x_i \beta_1 + \varepsilon_i \]

for \( i = 1, \ldots, n \), where \( \varepsilon_1, \ldots, \varepsilon_n \) are independent and identically distributed random variables with \( E(\varepsilon_i) = 0, \text{Var}(\varepsilon_i) = \sigma^2 \), following the distribution \( F \).

The error term \( \varepsilon \) was generated by five different distributions: normal, skewed unimodal, strongly skewed, kurtotic unimodal, and outlier. The shapes of these distributions are shown in Figure 5.1 and the densities of each of these distributions are given in Table 5.1. These functions are from Marron and Wand (1992).

This study consisted of generating 500 samples of size 6, 11, 26 and 51. The design points \( x_i \) ranged from 0 to 1 by \( 1/(n - 1) \) for \( i = 1, \ldots, n \), \( n = 6, 11, 26, 51 \). The error \( \varepsilon \) was generated from each of the five distributions in Table 5.1 and \( Y_i \) was calculated using a predefined slope parameter \( \beta_1 = 0.25, 0.5, 1, 1.5 \), such that

\[ Y_i = \beta_1 x_i + \varepsilon_i \]

for \( i = 1, \ldots, n \) and \( n = 6, 11, 26, 51 \). Using \( Y_i \) and \( x_i \), the least squares estimate
Figure 5.1: The distribution shapes.
Table 5.1: Distributions for residuals $\varepsilon_i$.

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal</td>
<td>$N(0,1)$</td>
</tr>
<tr>
<td>Skewed unimodal</td>
<td>$\frac{1}{5}N(0, 1) + \frac{1}{5}N(\frac{1}{2}, \frac{2}{3}) + \frac{3}{5}N(\frac{13}{12}, \frac{5}{9})$</td>
</tr>
<tr>
<td>Strongly skewed</td>
<td>$\sum_{i=1}^{7} \frac{1}{8}N(3{(\frac{2}{3})^i - 1}, (\frac{2}{3})^i)$</td>
</tr>
<tr>
<td>Kurtotic unimodal</td>
<td>$\frac{2}{3}N(0, 1) + \frac{1}{3}N(0, \frac{1}{10})$</td>
</tr>
<tr>
<td>Outlier</td>
<td>$\frac{1}{10}N(0, 1) + \frac{9}{10}N(0, \frac{1}{10})$</td>
</tr>
</tbody>
</table>

$\hat{\beta}_1$ was calculated. From this the observed confidence levels, $\hat{\alpha}_{\text{ord}}$, $\alpha_{\text{stud}}$, $\hat{\alpha}_{\text{hyb}}$, and $\hat{\alpha}_{\text{back}}$ were calculated for the desired interval $(0,1)$. Where $\hat{\alpha}_{\text{ord}}$ was calculated using the normal approximations, $\hat{\alpha}_{\text{stud}}$, $\hat{\alpha}_{\text{hyb}}$ and $\hat{\alpha}_{\text{back}}$ were calculated using the bootstrap estimates based on $b = 1000$ resamples of the residual estimate $\hat{\varepsilon}$. These simulations were calculated using R. The code to calculate these observed confidence levels is in the Appendix.

The results are given in Tables 5.2, 5.3, 5.4, 5.5, and 5.6. Also, Figures 5.2, 5.3, 5.4, 5.5, and 5.6 show the difference of the observed confidence levels of the different distributions based on the sample size of 51 for each distribution.

Overall, the bootstrap backwards $\hat{\alpha}_{\text{back}}$ and the normal approximation $\hat{\alpha}_{\text{ord}}$ tend to overestimate the confidence level when $\beta_1$ is in the desired interval. When $\beta_1 = 1.5$, which is not in the desired interval, $\hat{\alpha}_{\text{ord}}$ and $\hat{\alpha}_{\text{back}}$ underestimate the observed confidence level. The bootstrap Studentized $\hat{\alpha}_{\text{stud}}$ and the bootstrap hybrid $\hat{\alpha}_{\text{hyb}}$ were similar and close to $\alpha_{\text{stud}}$. This is evident by the accuracy of these observed confidence levels. Notice that $\hat{\alpha}_{\text{back}}$ and $\hat{\alpha}_{\text{ord}}$ are first-order accurate and that $\hat{\alpha}_{\text{hyb}}$, $\hat{\alpha}_{\text{stud}}$, $\hat{\alpha}_{\text{hyb}}$ were second-order accurate. Thus, the hybrid observed confidence level is

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Table 5.2: Observed confidence levels using the standard normal distribution.

<table>
<thead>
<tr>
<th>n</th>
<th>$\beta_1$</th>
<th>$\hat{\alpha}_{ord}$</th>
<th>$\hat{\alpha}_{stud}$</th>
<th>$\hat{\alpha}_{stud}^*$</th>
<th>$\hat{\alpha}_{hyb}^*$</th>
<th>$\hat{\alpha}_{back}^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>0.25</td>
<td>0.2566</td>
<td>0.1466</td>
<td>0.1232</td>
<td>0.1497</td>
<td>0.4270</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.2591</td>
<td>0.1526</td>
<td>0.1324</td>
<td>0.1546</td>
<td>0.4478</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.2474</td>
<td>0.1388</td>
<td>0.1262</td>
<td>0.1668</td>
<td>0.3592</td>
</tr>
<tr>
<td></td>
<td>1.5</td>
<td>0.2119</td>
<td>0.1329</td>
<td>0.1212</td>
<td>0.1422</td>
<td>0.2140</td>
</tr>
<tr>
<td>11</td>
<td>0.25</td>
<td>0.3030</td>
<td>0.1822</td>
<td>0.1708</td>
<td>0.1945</td>
<td>0.4501</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.3074</td>
<td>0.1804</td>
<td>0.1656</td>
<td>0.1849</td>
<td>0.4598</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.2821</td>
<td>0.1829</td>
<td>0.1603</td>
<td>0.1746</td>
<td>0.3810</td>
</tr>
<tr>
<td></td>
<td>1.5</td>
<td>0.2244</td>
<td>0.1649</td>
<td>0.1547</td>
<td>0.1677</td>
<td>0.2268</td>
</tr>
<tr>
<td>26</td>
<td>0.25</td>
<td>0.4019</td>
<td>0.2723</td>
<td>0.2590</td>
<td>0.2537</td>
<td>0.5455</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.4211</td>
<td>0.2771</td>
<td>0.2591</td>
<td>0.2588</td>
<td>0.5859</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.3696</td>
<td>0.2647</td>
<td>0.2397</td>
<td>0.2670</td>
<td>0.4466</td>
</tr>
<tr>
<td></td>
<td>1.5</td>
<td>0.2440</td>
<td>0.2021</td>
<td>0.2175</td>
<td>0.2059</td>
<td>0.1974</td>
</tr>
<tr>
<td>51</td>
<td>0.25</td>
<td>0.5132</td>
<td>0.3347</td>
<td>0.3328</td>
<td>0.3451</td>
<td>0.6555</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.5489</td>
<td>0.3695</td>
<td>0.3515</td>
<td>0.3627</td>
<td>0.7225</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.4378</td>
<td>0.3231</td>
<td>0.3391</td>
<td>0.3207</td>
<td>0.4832</td>
</tr>
<tr>
<td></td>
<td>1.5</td>
<td>0.2178</td>
<td>0.2360</td>
<td>0.2312</td>
<td>0.2274</td>
<td>0.1387</td>
</tr>
</tbody>
</table>

Table 5.3: Observed confidence levels using the skewed unimodal distribution.

<table>
<thead>
<tr>
<th>n</th>
<th>$\beta_1$</th>
<th>$\hat{\alpha}_{ord}$</th>
<th>$\hat{\alpha}_{stud}$</th>
<th>$\hat{\alpha}_{stud}^*$</th>
<th>$\hat{\alpha}_{hyb}^*$</th>
<th>$\hat{\alpha}_{back}^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>0.25</td>
<td>0.3538</td>
<td>0.3475</td>
<td>0.3366</td>
<td>0.3572</td>
<td>0.7264</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.3623</td>
<td>0.3561</td>
<td>0.3585</td>
<td>0.3723</td>
<td>0.8140</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.3339</td>
<td>0.3173</td>
<td>0.3410</td>
<td>0.3423</td>
<td>0.4877</td>
</tr>
<tr>
<td></td>
<td>1.5</td>
<td>0.2448</td>
<td>0.2430</td>
<td>0.2361</td>
<td>0.2335</td>
<td>0.0878</td>
</tr>
<tr>
<td>11</td>
<td>0.25</td>
<td>0.5071</td>
<td>0.4220</td>
<td>0.4296</td>
<td>0.4384</td>
<td>0.7569</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.5407</td>
<td>0.4240</td>
<td>0.4409</td>
<td>0.4300</td>
<td>0.8528</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.4313</td>
<td>0.3918</td>
<td>0.3797</td>
<td>0.3714</td>
<td>0.4960</td>
</tr>
<tr>
<td></td>
<td>1.5</td>
<td>0.2168</td>
<td>0.2307</td>
<td>0.2735</td>
<td>0.2262</td>
<td>0.0790</td>
</tr>
<tr>
<td>26</td>
<td>0.25</td>
<td>0.7418</td>
<td>0.5223</td>
<td>0.5557</td>
<td>0.5389</td>
<td>0.8402</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.8365</td>
<td>0.5922</td>
<td>0.6100</td>
<td>0.6038</td>
<td>0.9435</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.5053</td>
<td>0.4761</td>
<td>0.4424</td>
<td>0.4352</td>
<td>0.4992</td>
</tr>
<tr>
<td></td>
<td>1.5</td>
<td>0.0818</td>
<td>0.1892</td>
<td>0.2090</td>
<td>0.2183</td>
<td>0.0278</td>
</tr>
<tr>
<td>51</td>
<td>0.25</td>
<td>0.9547</td>
<td>0.6823</td>
<td>0.6774</td>
<td>0.6499</td>
<td>0.9063</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.9681</td>
<td>0.7491</td>
<td>0.7650</td>
<td>0.7379</td>
<td>0.9895</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.5226</td>
<td>0.4949</td>
<td>0.5085</td>
<td>0.5342</td>
<td>0.4998</td>
</tr>
<tr>
<td></td>
<td>1.5</td>
<td>0.0172</td>
<td>0.1282</td>
<td>0.1214</td>
<td>0.1334</td>
<td>0.0050</td>
</tr>
</tbody>
</table>
Table 5.4: Observed confidence levels using strongly skewed distribution.

<table>
<thead>
<tr>
<th>n</th>
<th>$\beta_1$</th>
<th>$\hat{\alpha}_{\text{ord}}$</th>
<th>$\hat{\alpha}_{\text{stud}}$</th>
<th>$\hat{\alpha}_{\text{stud}}^*$</th>
<th>$\hat{\alpha}_{\text{hyb}}^*$</th>
<th>$\hat{\alpha}_{\text{back}}^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>0.25</td>
<td>0.3876</td>
<td>0.6558</td>
<td>0.5894</td>
<td>0.6843</td>
<td>0.9392</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.3998</td>
<td>0.7178</td>
<td>0.6452</td>
<td>0.7478</td>
<td>0.9926</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.3545</td>
<td>0.4695</td>
<td>0.4630</td>
<td>0.4526</td>
<td>0.4987</td>
</tr>
<tr>
<td></td>
<td>1.5</td>
<td>0.2415</td>
<td>0.1342</td>
<td>0.1485</td>
<td>0.1319</td>
<td>0.0029</td>
</tr>
<tr>
<td>11</td>
<td>0.25</td>
<td>0.5731</td>
<td>0.7245</td>
<td>0.7185</td>
<td>0.7285</td>
<td>0.9567</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.6182</td>
<td>0.8303</td>
<td>0.7987</td>
<td>0.8532</td>
<td>0.9984</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.4529</td>
<td>0.5157</td>
<td>0.5206</td>
<td>0.5268</td>
<td>0.5003</td>
</tr>
<tr>
<td></td>
<td>1.5</td>
<td>0.1860</td>
<td>0.0835</td>
<td>0.0882</td>
<td>0.0616</td>
<td>0.0011</td>
</tr>
<tr>
<td>26</td>
<td>0.25</td>
<td>0.8191</td>
<td>0.8512</td>
<td>0.8288</td>
<td>0.8443</td>
<td>0.9902</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.9369</td>
<td>0.9445</td>
<td>0.9506</td>
<td>0.9631</td>
<td>1.0000</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.5062</td>
<td>0.4813</td>
<td>0.5176</td>
<td>0.4869</td>
<td>0.5007</td>
</tr>
<tr>
<td></td>
<td>1.5</td>
<td>0.0303</td>
<td>0.0274</td>
<td>0.0251</td>
<td>0.0252</td>
<td>0.0000</td>
</tr>
<tr>
<td>51</td>
<td>0.25</td>
<td>0.9547</td>
<td>0.9214</td>
<td>0.9133</td>
<td>0.9156</td>
<td>0.9991</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.9990</td>
<td>0.9923</td>
<td>0.9946</td>
<td>0.9956</td>
<td>1.0000</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.4983</td>
<td>0.5174</td>
<td>0.5018</td>
<td>0.4835</td>
<td>0.5012</td>
</tr>
<tr>
<td></td>
<td>1.5</td>
<td>0.0005</td>
<td>0.0012</td>
<td>0.0042</td>
<td>0.0051</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Table 5.5: Observed confidence levels using the kurtotic unimodal distribution.

<table>
<thead>
<tr>
<th>n</th>
<th>$\beta_1$</th>
<th>$\hat{\alpha}_{\text{ord}}$</th>
<th>$\hat{\alpha}_{\text{stud}}$</th>
<th>$\hat{\alpha}_{\text{stud}}^*$</th>
<th>$\hat{\alpha}_{\text{hyb}}^*$</th>
<th>$\hat{\alpha}_{\text{back}}^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>0.25</td>
<td>0.3572</td>
<td>0.2179</td>
<td>0.2209</td>
<td>0.2338</td>
<td>0.5670</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.3781</td>
<td>0.2248</td>
<td>0.2349</td>
<td>0.2267</td>
<td>0.6041</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.3356</td>
<td>0.4538</td>
<td>0.2377</td>
<td>0.2294</td>
<td>0.4372</td>
</tr>
<tr>
<td></td>
<td>1.5</td>
<td>0.2309</td>
<td>0.1906</td>
<td>0.1888</td>
<td>0.1714</td>
<td>0.1766</td>
</tr>
<tr>
<td>11</td>
<td>0.25</td>
<td>0.4226</td>
<td>0.2609</td>
<td>0.2800</td>
<td>0.2649</td>
<td>0.5812</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.4470</td>
<td>0.2759</td>
<td>0.2976</td>
<td>0.2815</td>
<td>0.6401</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.3749</td>
<td>0.2628</td>
<td>0.2742</td>
<td>0.2728</td>
<td>0.4564</td>
</tr>
<tr>
<td></td>
<td>1.5</td>
<td>0.2369</td>
<td>0.2210</td>
<td>0.2220</td>
<td>0.1938</td>
<td>0.1762</td>
</tr>
<tr>
<td>26</td>
<td>0.25</td>
<td>0.5559</td>
<td>0.3923</td>
<td>0.3678</td>
<td>0.3851</td>
<td>0.6961</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.5912</td>
<td>0.3803</td>
<td>0.4019</td>
<td>0.4013</td>
<td>0.7697</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.4582</td>
<td>0.3290</td>
<td>0.3522</td>
<td>0.3474</td>
<td>0.4911</td>
</tr>
<tr>
<td></td>
<td>1.5</td>
<td>0.1946</td>
<td>0.2308</td>
<td>0.2557</td>
<td>0.2388</td>
<td>0.1125</td>
</tr>
<tr>
<td>51</td>
<td>0.25</td>
<td>0.6666</td>
<td>0.4935</td>
<td>0.4848</td>
<td>0.4944</td>
<td>0.7846</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.7363</td>
<td>0.4927</td>
<td>0.5013</td>
<td>0.5240</td>
<td>0.8905</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.5004</td>
<td>0.4206</td>
<td>0.4200</td>
<td>0.4527</td>
<td>0.4992</td>
</tr>
<tr>
<td></td>
<td>1.5</td>
<td>0.1146</td>
<td>0.2336</td>
<td>0.2256</td>
<td>0.2451</td>
<td>0.0545</td>
</tr>
</tbody>
</table>
Table 5.6: Observed confidence levels using the outlier distribution.

<table>
<thead>
<tr>
<th>n</th>
<th>$\beta$</th>
<th>$\hat{\alpha}_{\text{ord}}$</th>
<th>$\hat{\alpha}_{\text{stud}}$</th>
<th>$\hat{\alpha}_{\text{hyb}}^*$</th>
<th>$\hat{\alpha}_{\text{back}}^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>0.25</td>
<td>0.8848</td>
<td>0.7392</td>
<td>0.6895</td>
<td>0.7536</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.9804</td>
<td>0.8310</td>
<td>0.7445</td>
<td>0.8408</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.4999</td>
<td>0.5037</td>
<td>0.4749</td>
<td>0.4994</td>
</tr>
<tr>
<td></td>
<td>1.5</td>
<td>0.0104</td>
<td>0.0930</td>
<td>0.1088</td>
<td>0.0493</td>
</tr>
<tr>
<td>11</td>
<td>0.25</td>
<td>0.9274</td>
<td>0.8057</td>
<td>0.7750</td>
<td>0.8116</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.9947</td>
<td>0.9147</td>
<td>0.8844</td>
<td>0.9175</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.5074</td>
<td>0.5107</td>
<td>0.4795</td>
<td>0.4936</td>
</tr>
<tr>
<td></td>
<td>1.5</td>
<td>0.0024</td>
<td>0.0548</td>
<td>0.0613</td>
<td>0.0333</td>
</tr>
<tr>
<td>26</td>
<td>0.25</td>
<td>0.9833</td>
<td>0.8911</td>
<td>0.8933</td>
<td>0.9015</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.9999</td>
<td>0.9914</td>
<td>0.9901</td>
<td>0.9891</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.4846</td>
<td>0.4717</td>
<td>0.5055</td>
<td>0.5134</td>
</tr>
<tr>
<td></td>
<td>1.5</td>
<td>0.0000</td>
<td>0.0075</td>
<td>0.0094</td>
<td>0.0057</td>
</tr>
<tr>
<td>51</td>
<td>0.25</td>
<td>0.9979</td>
<td>0.9544</td>
<td>0.9613</td>
<td>0.9536</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>1.0000</td>
<td>0.9988</td>
<td>0.9990</td>
<td>0.9998</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.5054</td>
<td>0.5223</td>
<td>0.4960</td>
<td>0.4951</td>
</tr>
<tr>
<td></td>
<td>1.5</td>
<td>0.0000</td>
<td>0.0004</td>
<td>0.0002</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Figure 5.2: Normal distribution.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Figure 5.3: Skewed unimodal distribution.

Figure 5.4: Strongly skewed distribution.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
Figure 5.5: Kurtotic unimodal distribution.

Figure 5.6: Outlier distribution.
as good of an estimate for the observed confidence level as the bootstrap Studentized estimate of the observed confidence level for $\beta_1$. 
CHAPTER 6

FUTURE WORK

This work demonstrates the observed confidence levels for parameters for simple linear regression models and multiple regression models. These observed confidence levels were computed using the ordinary, Studentized, hybrid, backwards (percentile), bias-corrected and accelerated bias-corrected critical points. The future work for this project includes:

1. Computing observed confidence levels for multivariate, multiple regression parameters.

2. Computing observed confidence levels for nonlinear regression parameters.

3. Creating simulations of the observed confidence levels based on the accelerated bias-corrected critical point.
BIBLIOGRAPHY


APPENDIX

R CODE

The calculations and simulation that were performed in this paper were calculated using the R statistical computing environment.

Compost Data

This section demonstrates the R code that was used in the example in Section 3.4. To calculate the estimates of the regression parameter, the function in R is called lm for linear models. The following code calculates the estimates for the model

\[ EGI = \beta_0 + GI \beta_1. \]

\[ GI.fit<-lm(EGI~GI) \]

To calculate the observed confidence levels, a function was created for each \( \hat{\alpha}_{hyb}^*, \hat{\alpha}_{back}^*, \) and \( \hat{\alpha}_{stud}^* \). The following is the function to calculate \( \hat{\alpha}_{hyb}^* \).

\[
giboothyb<-function(data,i,tval,fit){
  Ystar<-fit$coefficients[1]+fit$coefficients[2]*data+fit$residuals[i]
  fitstar<-lm(Ystar~data)
    delta<-1
  else delta <-0
  return(delta)
}

86

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
This function is used with the bootstrap library. The library boot must be initiated first. Then the command

\[
\text{sum(boot(GL, statistic=giboothyb, R=10000, tval=1, fit=lm.egi)$t)/10000}
\]

will return \( \hat{\alpha}_{\text{hyb}}^* \) for \( \Theta = \beta < 1 \).

To compute \( \hat{\alpha}_{\text{back}}^* \), the function

\[
gibootperc<-\text{function}(data, i, tval, fit) \{
Ystar<-\text{fit$coeff}[1]+\text{fit$coeff}[2]*data+\text{fit$residuals}[i]
fitstar<-\text{lm}(Ystar~data)
if(fitstar$coeff[2] < tval) delta=1 else delta=0
return(delta)}
\]

is used. Then

\[
\text{sum(boot(GL, statistic=gibootperc, R=10000, tval=1, fit=lm.egi)$t)/10000}
\]

will result in \( \hat{\alpha}_{\text{back}}^* \). To compute \( \hat{\alpha}_{\text{stud}}^* \), \( \sigma^* \) has to be calculated. Thus the function is

\[
gibootstud<-\text{function}(data, i, tval, fit) \{
Ystar<-\text{fit$coeff}[1]+\text{fit$coeff}[2]*data+\text{fit$residuals}[i]
fitstar<-\text{lm}(Ystar~data)
shat <- \text{sum(fit$residuals^2)}/\text{length(fit$residuals)}
shatstar<-\text{sum(fitstar$residuals^2)}/\text{length(fitstar$residuals)}
if(fit$coeff[2]-shat*(fitstar$coeff[2]-fit$coeff[2])/shatstar<tval)
delta=1 else delta=0
return(delta)}
\]
Then

\[ \text{sum(boot(GI, statistic=giboothyb, R=10000, tval=1, fit=lm.egi)$t)/10000} \]

will result in \( \hat{\alpha}_{\text{stud}} \).

**Tree Data**

The following code calculates the regression model for each tree type from the example in Section 4.4.

```r
lm.maple <- lm(time.maple.sim ~ diam.maple.sim)
lm.oak <- lm(time.oak.sim ~ diam.oak.sim)
lm.birch <- lm(time.birch.sim ~ diam.birch.sim)
that <- matrix(c(lm.maple$coefficients[2], lm.oak$coefficients[2],
                 lm.birch$coefficients[2]), 3, 1)
```

Recall that the variance for \( n^{-1} \hat{\beta} \) was

\[
\hat{\Sigma}_T = \begin{bmatrix}
\hat{\sigma}_1^2/\sigma_{x,1}^2 & 0 & 0 \\
0 & \hat{\sigma}_2^2/\sigma_{x,2}^2 & 0 \\
0 & 0 & \hat{\sigma}_3^2/\sigma_{x,3}^2
\end{bmatrix}.
\]

This is the code to calculate \( \hat{\Sigma}_T \).

```r
sigma.maple <- sum(lm.maple$residuals^2)/length(diam.maple.sim)
sigma.oak <- sum(lm.oak$residuals^2)/length(diam.oak.sim)
sigma.birch <- sum(lm.birch$residuals^2)/length(diam.birch.sim)
sdx.maple <- sd(diam.maple.sim)*sqrt((length(diam.maple.sim)-1)/
                 length(diam.maple.sim))
sdx.oak <- sd(diam.oak.sim)*sqrt((length(diam.oak.sim)-1))
```

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
\[ \text{length(diam.oak.sim))} \]
\[ \text{sdx.birch<-sd(diam.birch.sim)*sqrt((length(diam.birch.sim)-1)/length(diam.birch.sim))} \]
\[ \text{Sx.tree<-matrix(0,3,3)} \]
\[ \text{Sx.tree[1,1]<-sigma.maple/sdx.maple^2} \]
\[ \text{Sx.tree[2,2]<-sigma.oak/sdx.oak^2} \]
\[ \text{Sx.tree[3,3]<-sigma.birch/sdx.birch^2} \]

The singular-value decomposition of \( \hat{\Sigma}_T \) is calculated to get \( \hat{\Sigma}_T^{-1/2} \).

\[ \text{svdSx.tree<-svd(Sx.tree)} \]
\[ \text{srSx.tree<-svdSx.tree$u*sqrt(diag(svdSx.tree$d))%*%t(svdSx.tree$u)} \]

The normal approximation of \( \hat{\sigma}_{\text{stud}} \) can be calculated using numerical integration based on the multivariate standard normal density. This can also be calculated as follows: Let \( \mathbf{Z}_i^*, \ldots, \mathbf{Z}_b^* \) be a set of independent and identically distributed random variables from a multivariate normal density with mean vector \( 0_3 \) and covariance matrix \( \mathbf{I}_3 \). Then for this study

\[
\hat{\sigma}_{\text{stud}}(\Theta_k) = \int_{n^{1/2}\hat{\Sigma}_T^{-1/2}(\hat{\beta} - \Theta_k)} \phi(t)dt \\
\approx b^{-1} \sum_{i=1}^{b} I[\mathbf{Z}_i^* \in n^{1/2}\hat{\Sigma}_T^{-1/2}(\hat{\beta} - \Theta_k)] \\
= b^{-1} \sum_{i=1}^{b} I[\hat{\beta} - n^{-1/2}\hat{\Sigma}_T^{1/2}\mathbf{Z}_i^* \in \Theta_k]
\]

where \( I \) is the indicator function. This calculation uses the following code:

\[ \text{b<-1000} \]
n<-length(diam.maple.sim)
kount<-matrix(0,1,6)
Z<-matrix(rnorm(3*b),3,b)
Zt<-matrix(0,3,b)
for (i in 1:b)
{
  Zt[,i]<-that-(srSx.tree%*%Z[,i])/sqrt(n)
  if( (Zt[1,i]<Zt[2,i])&& (Zt[2,i]<Zt[3,i]) ) kount[1,1]<- kount[1,1] +1
  if( (Zt[1,i]<Zt[3,i])&& (Zt[3,i]<Zt[2,i]) ) kount[1,2]<- kount[1,2] +1
  if( (Zt[2,i]<Zt[3,i])&& (Zt[3,i]<Zt[1,i]) ) kount[1,3]<- kount[1,3] +1
  if( (Zt[2,i]<Zt[1,i])&& (Zt[1,i]<Zt[3,i]) ) kount[1,4]<- kount[1,4] +1
  if( (Zt[3,i]<Zt[1,i])&& (Zt[1,i]<Zt[2,i]) ) kount[1,5]<- kount[1,5] +1
  if( (Zt[3,i]<Zt[2,i])&& (Zt[2,i]<Zt[1,i]) ) kount[1,6]<- kount[1,6] +1
}

where each column of kount represents one of the six regions. Then \( \hat{\alpha}_{\text{stud}} \) is calculated by

alpha.stud<-kount/b.
To calculate $\hat{\sigma}_{\text{stud}}$, the data was combined into a matrix $[W : X]$ where $W$ and $X$ are defined in Equations (4.4.1) and (4.4.1) respectively. Then

$$
\hat{\sigma}_{\text{stud}}(\Theta_k) = \int_{n^{1/2} \hat{\Sigma}_{i}^{-1/2}(\hat{\beta} - \Theta_k)}^{b} \hat{h}_n(t) dt
$$

$$
\approx b^{-1} \sum_{i=1}^{b} I[n^{1/2} \hat{\Sigma}_{T}^{-1/2}(\hat{\beta}_i^* - \hat{\beta}) \in n^{1/2} \hat{\Sigma}_{T}^{-1/2}(\hat{\beta} - \Theta_k)]
$$

$$
= b^{-1} \sum_{i=1}^{b} I[\hat{\beta} - \hat{\Sigma}_{T}^{-1/2}(\hat{\beta}_i^* - \hat{\beta}) \in \Theta_k].
$$

The following is the bootstrap function that calculates this approximation.

```r
bootstud.tree<-function(data,i,fit)
{
  tstar<-matrix(c(fitstar$coefficients[4],fitstar$coefficients[5],fitstar$coefficients[6]),3,1)
  res.maple<-matrix(0,20,1)
  res.oak<-matrix(0,20,1)
  res.birch<-matrix(0,20,1)
  for (j in 1:20)
    {
      tstar<-tstar+tstar
      res.maple[j]<-tstar[1]
    }
}
```

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
\{
res.maple[j,1]<-fitstar$residuals[j]
res.oak[j,1]<-fitstar$residuals[j+20]
res.birch[j,1]<-fitstar$residuals[j+40]
\}

shat.maple<-sum(res.maple^2)/20
shat.oak<-sum(res.oak^2)/20
shat.birch<-sum(res.birch^2)/20

Sxstar<-matrix(0,3,3)
Sxstar[1,1]<-shat.maple/sdx.maple^2
Sxstar[2,2]<-shat.oak/sdx.oak^2
Sxstar[3,3]<-shat.birch/sdx.birch^2
svdSxstar<-svd(Sxstar)
srSxstar<-svdSxstar$u%*%sqrt(diag(svdSxstar$d))%*%t(svdSxstar$u)
ttheta<-that-srSx%*%solve(srSxstar)%*%(tstar-that)
delta<-0
if((ttheta[1,]<ttheta[2,])&&(ttheta[2,]<ttheta[3,])) delta<-1
if((ttheta[1,]<ttheta[3,])&&(ttheta[3,]<ttheta[2,])) delta<-2
if((ttheta[2,]<ttheta[1,])&&(ttheta[1,]<ttheta[3,])) delta<-3
if((ttheta[2,]<ttheta[3,])&&(ttheta[3,]<ttheta[1,])) delta<-4
if((ttheta[3,]<ttheta[1,])&&(ttheta[1,]<ttheta[2,])) delta<-5
if((ttheta[3,]<ttheta[2,])&&(ttheta[2,]<ttheta[1,])) delta<-6

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
return(delta)

Then the command

table(boot(tree, statistic=bootstud.tree, R=1000, fit=lm.tree2)$t)/1000

will return a table for the observed confidence level for each of the six regions.

To calculate $\hat{\alpha}_{hyb}$ we use

$$\hat{\alpha}_{hyb}(\Theta_k) = \int_{n^{1/2} \Sigma_T^{-1/2} (\beta - \Theta_k)} \hat{h}_n(t) dt$$

$$\approx b^{-1} \sum_{i=1}^{b} I[n^{1/2} \Sigma_T^{-1/2} (\hat{\beta}^*_i - \hat{\beta}) \in n^{1/2} \Sigma_T^{-1/2} (\hat{\beta} - \Theta_k)]$$

$$= b^{-1} \sum_{i=1}^{b} I[2\hat{\beta}^*_i - \hat{\beta} \in \Theta_k].$$

The following bootstrap function calculates this observed confidence interval.

boothyb.tree<-function(data,i,fit)
{
  Ystar<-fit$coefficients[1]*data[,1]+fit$coefficients[2]*data[,2]+
          fit$coefficients[3]*data[,3]+fit$coefficients[4]*data[,4]+
          fit$coefficients[5]*data[,5]+fit$coefficients[6]*data[,6]+
          fit$residuals[i]
              data[,6])
  tstar<-matrix(c(fitstar$coefficients[4],fitstar$coefficients[5],
                  fitstar$coefficients[6]),3,1)
  ttheta<-2*that-tstar
delta<-0
if((ttheta[1,<ttheta[2,,]&&(ttheta[2,<ttheta[3,,])) delta<-1
if((ttheta[1,<ttheta[3,,]&&(ttheta[3,<ttheta[2,,])) delta<-2
if((ttheta[2,<ttheta[1,,]&&(ttheta[1,<ttheta[3,,])) delta<-3
if((ttheta[2,<ttheta[3,,]&&(ttheta[3,<ttheta[1,,])) delta<-4
if((ttheta[3,<ttheta[1,,]&&(ttheta[1,<ttheta[2,,])) delta<-5
if((ttheta[3,<ttheta[2,,]&&(ttheta[2,<ttheta[1,,])) delta<-6
return(delta)
}

Then the command

table(boot(tree,statistic=boothyb.tree,R=1000,fit=lm.tree2)$t)/1000

will give a table of the observed confidence levels for each region.

The bootstrap approximation of the percentile method can be calculated by

\[ \hat{\alpha}_{perc}^* \approx b^{-1} \sum_{i=1}^{b} I(\hat{\beta}_i^* \in \Theta_k). \]

The following function was used to calculate this observed confidence level:

\[
\text{bootperc.tree<-function(data,i,fit)}
\{
\text{Ystar<-fit$coefficients[1]*data[,1]+fit$coefficients[2]*data[,2]+}
\text{fit$coefficients[3]*data[,3]+fit$coefficients[4]*data[,4]+}
\text{fit$coefficients[5]*data[,5]+fit$coefficients[6]*data[,6]+}
\text{fit$residuals[i]}
\text{fitstar<-lm(Ystar~0+data[,1]+data[,2]+data[,3]+data[,4]+data[,5]+}
\]

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
data[,6])
tstar<-matrix(c(fitstar$coefficients[4],fitstar$coefficients[5],
fitstar$coefficients[6]),3,1)

delta<-0
if((tstar[1,]<tstar[2,])&&(tstar[2,]<tstar[3,])) delta<-1
if((tstar[1,]<tstar[3,])&&(tstar[3,]<tstar[2,])) delta<-2
if((tstar[2,]<tstar[1,])&&(tstar[1,]<tstar[3,])) delta<-3
if((tstar[2,]<tstar[3,])&&(tstar[3,]<tstar[1,])) delta<-4
if((tstar[3,]<tstar[1,])&&(tstar[1,]<tstar[2,])) delta<-5
if((tstar[3,]<tstar[2,])&&(tstar[2,]<tstar[1,])) delta<-6
return(delta)
}

Then the command

table(boot(tree,statistic=bootperc.tree,R=1000,fit=lm.tree2)$t)/1000

will give a table of observed confidence levels.

**R Code for Simulations**

The observed confidence levels that were created from the simulations in Chapter 5 utilized R. This appendix demonstrates some of the code that was used to calculate the observed confidence levels. The distributions of the random errors for the regression models come from Table 5.1. The code that is shown here will use the outlier distribution. These functions change depending on the distribution. The
observed confidence levels are calculated using $(0, 1)$ as the confidence interval.

The following code calculates $\hat{\alpha}_{ord}$ for the sample size 6 and for $\beta = 0.25$.

c<-1000
tl<-0
tu<-1
nm1<-matrix(0,c,1)
number<-5
x<-seq(0,1,1/number)
sx<-sigmax(x,number)
beta<-.25
for (j in 1:c)
{
e<-(1/10)*rnorm(number+1)+(9/10)*rnorm(number+1,0,(1/10))
e<-e-mean(e)
y<-beta*x+e
int<-bhat(x,y)
r<-y-mean(y)-(x-mean(x))*int[2]
sigma<-shat(r,number)
U1<-sx*(int[2]-tl)*sqrt(length(x))/sigma
U2<-sx*(int[2]-tu)*sqrt(length(x))/sigma
nm1[j]<-pnorm(U1)-pnorm(U2)
}
mmnm1<-mean(nm1)
The following code is to calculate $\alpha_{\text{stud}}$ for sample size $n = 11$ and $\beta = 1$. 

```r
c<-500
stud2<-matrix(0,c,1)
tl<-0
tu<-1
b<-1000
l<-matrix(0,b,2)
t<-c(1:b)
z<-c(1:b)
a<-c(1:b)
number<-10
x<-seq(0,1,1/number)
sx<-sigmax(x,number)

beta<-0.5
for (j in 1:c)
{
e<-(1/10)*rnorm(number+1)+(9/10)*rnorm(number+1,0,(1/10))
e<-e-mean(e)
y<-beta*x+e
y.lm<-lm(y~x)
r<-y.lm$residuals
betahat<-y.lm$coefficients[2]
sigma<-shat(r,number)
```

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
# Resampling

for (i in 1:b)
{

estar<-(1/10)*rnorm(number+1)+(9/10)*rnorm(number+1,0,(1/10))

estar<-estar-mean(estar)

y<-beta*x+estar

1[i,]<-bhat(x,y)

rstar<-y-mean(y)-(x-mean(x))*1[i,2]

sigmahat<-shat(rstar,number)

t[i]<-betahat-sigma/sigmahat*(1[i,2]-betahat)

if (t[i]>=tl) z[i]<-1 else z[i]<-0

if (t[i]>=tu) a[i]<-1 else a[i]<-0

}

stud2[j]<-sum(z)/b-sum(a)/b

}

mstud2<-mean(stud2)

The following code is to calculate $\hat{\alpha}_{\text{stud}}^*$ for sample size $n = 26$ and $\beta = 1$.

c<-500

b<-1000

l<-matrix(0,b,2)

t<-c(1:b)

tl<-0

tu<-1

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.
z<-c(1:b)
a<-c(1:b)
studboot3<-matrix(0,c,1)
number<-25
x<-seq(0,1,1/number)
sx<-sigmax(x,number)
beta<-1.0
for (j in 1:c)
{
e<-(1/10)*rnorm(number+1)+(9/10)*rnorm(number+1,0,(1/10))
e<-e-mean(e)
y<-beta*x+e
y.lm<-lm(y~x)
r<-y.lm$residuals
betahat<-y.lm$coefficients[2]
sigma<-shat(r,number)
#Resampling for Bootstrap Studentized
for (i in 1:b)
{
estar<-sample(r,replace=T)
estar<-estar-mean(estar)
y<-beta*x+estar
1[i,]<-bhat(x,y)
rstar<-y-mean(y)-(x-mean(x))*1[i,2]
sigmahat<-shat(rstar,number)
t[i]<-betahat-sigma/sigmahat*(1[i,2]-betahat)
if (t[i]>=tl) z[i]<-1 else z[i]<-0
if (t[i]>=tu) a[i]<-1 else a[i]<-0
}
studboot3[j]<-sum(z)/b-sum(a)/b
}
mstudboot3<-mean(studboot3)

The following code calculates $\hat{\alpha}_{hyb}$ for sample size $n = 51$ and $\beta = 1.5$.

c<-500
b<-1000
tl<-0
tu<-1
z<-c(1:b)
a<-c(1:b)
l<-matrix(0,b,2)
t<-c(1:b)
hyb4<-matrix(0,c,1)
number<-50
x<-seq(0,1,1/number)
sx<-sigmax(x,number)
beta<-1.5
for (j in 1:c)
{ 
    e <- (1/10)*rnorm(number+1)+(9/10)*rnorm(number+1,0,(1/10))
    e <- e - mean(e)
    y <- beta*x + e
    y.lm <- lm(y~x)
    r <- y.lm$residuals
    betahat <- y.lm$coefficients[2]
    sigma <- shat(r, number)
    # Resampling
    for (i in 1:b)
    {
        estar <- sample(r, replace=T)
        estar <-estar - mean(estar)
        y <- beta*x + estar
        l[i,] <- bhat(x, y)
        if (2*betahat-l[i,2]>=tl) z[i]<-1 else z[i]<-0
        if (2*betahat-l[i,2]>=tu) a[i]<-1 else a[i]<-0
    }
    hyb4[j]<-sum(z)/b-sum(a)/b
}
mhyb4<-mean(hyb4)

The following code calculates $\hat{\alpha}_{\text{back}}$ for sample size $n = 6$ and $\beta = 0.25$.

c<-500
b<-1000
t1<-0
tu<-1
z<-c(1:b)
a<-c(1:b)
l<-matrix(0,b,2)
back1<-matrix(0,c,1)
number<-5
x<-seq(0,1,1/number)
sx<-sigmax(x,number)
beta<-0.25
for (j in 1:c)
{
  e<-(l/10)*rnorm(number+l)+(9/10)*rnorm(number+l,0,(1/10))
e<-e-mean(e)
y<-beta*x+e
y.lm<-lm(y~x)
r<-y.lm$residuals
#observed confidence level for backward
for (i in 1:b)
{
estar<-sample(r,replace=T)
estar<-estar-mean(estar)
y<-beta*x+estar
\[ l[i,] <- bhat(x, y) \]
\[ \text{if (tu} \geq l[i,2]) z[i] <- 1 \text{ else } z[i] <- 0 \]
\[ \text{if (tl} \geq l[i,2]) a[i] <- 1 \text{ else } a[i] <- 0 \]
\}
\[ \text{back1}[j] <- \frac{\text{sum}(z)}{b} - \frac{\text{sum}(a)}{b} \]
\}
\[ \text{mback1} <- \text{mean}(\text{back1}) \]