The Bootstrap for Regression and Time-Series Models

Matthew Pollard[1]

12th April 2005

[1]Department of Mathematics, The Australian National University. Email: matthewcpollard@berkeley.edu

Abstract

This report serves as an introduction to the bootstrap principle and its wide-ranging statistical applications. In particular, modification of Efron's bootstrap for time series data is treated in detail, and the problem of block length selection in block resampling is approached using asymptotic and empirical methods. The main results are: proving the optimal block length for n-length time series is given by $\widehat{l} = C_0 n^{1/d}$ where d is either 3, 4 or 5, and C_0 is a constant proportional to the covariance structure of the data; and determining optimal block lengths the variance of the sample mean and bias of sample variance using AR(1) data.

Aknowledgements

I would like to express thanks to my supervisors, Prof. Peter Hall and Dr. Michael Martin. It has been a privilege to have them as supervisors. Secondly, I would like to thank Lukas Sigut and Robin Pollard for their support. Finally, I must give praise for the modern personal computer. Over 1 billion bootstrap simulations were performed for the purpose of this report.

Contents

1 Introduction **5**

1.1 The Main Principle . . . 6

1.2 Examples . . . 7

1.2.1 Sample mean . . . 8

1.2.2 Sample variance . . . 8

1.3 Simulation Error . . . 10

1.4 Bias Correction . . . 11

1.4.1 Example: estimating σ^2 . . . 13

2 Application in Regression **15**

2.1 Introduction . . . 15

2.2 Model-based resampling . . . 16

2.2.1 Case study: variance of β_1 and β_0 . . . 16

2.3 Point-based Resampling . . . 21

2.3.1 Case study: variance of β_1 and β_0 . . . 24

3 Application in Time Series **29**

3.1 Introduction . . . 29

3.1.1 SABL decomposition . . . 30

3.1.2 Box-Jenkins decomposition . . . 30

3.1.3 Resampling methods . . . 33

3.2 Model-based resampling . . . 33

3.3 Block resampling . . . 34

3.3.1 Selecting an optimal block length . . . 35

3.3.2 Empirical method for block choice . . . 43

3.3.3 Case study: variance of sample mean . . . 44

3.3.4 Case study: bias of sample variance . . . 45

3.3.5 Results . . . 46

4 Appendix **53**

4.1 Combinatorial aspects of the bootstrap . . . 53

4.2 R code for bootstrapping . . . 54

4.2.1 boot . . . 54

4.2.2 modelboot . . . 54

4.2.3 pointboot 55
4.2.4 tsmodelboot 56
4.2.5 tsbooty 57
4.2.6 bootlength 57
4.2.7 blocklength 58

Chapter 1

Introduction

Efron's bootstrap (1979) has become an important tool for analysing the distribution of complicated statistics in a wide variety of settings. While the fundamental idea behind the bootstrap, what Efron calls the "plug-in principle," had existed many years before, its marriage with Monte Carlo approximation makes it tremendously useful.

The bootstrap is a resampling technique. Data is repeatedly sampled with replacement from the sample to make same-sized new data sets, called *resamples*. Using relatively simple theory, Efron (1979) showed that the bootstrap can give accurate non-parametric estimates of

1. standard errors of estimators,
2. confidence intervals for unknown parameters, and
3. p values for test statistics under a null hypothesis.

Consider that a statistical model is essentially a set of probability distributions that attempts to describe the true state of nature. The goal of statistical inference is to choose one of these probability distributions and give some notion of the uncertainty of that choice (usually through means of 1, 2 or 3 above).

The bootstrap works by treating inference of the true probability distribution F, given the original data, being analogous to inference of the empirical distribution $\widehat{F}$, given resampled data. The accuracy of inferences regarding $\widehat{F}$ using the resampled data can be assessed because we *know* $\widehat{F}$. If $\widehat{F}$ is a reasonable approximation to F, then the accuracy of inference on F can in turn be inferred.

One area where Efron's bootstrap fails is where data are dependent, such as in time series data (Singh, 1981). Dependency structure is usually what is of interest in time series, and any such structure is destroyed by resampling individual observations independently with replacement. There are two main methods to avoid this: model based (Freedman, 1984) and block based (Carlstein, 1986 and Künsch, 1989). The former removes dependence structure through fitting a model, leaving independent errors that are sampled. The latter divides the observed series into contiguous data "blocks" which are used to capture the dependence in the original series.

This report discusses a variety of bootstrap applications with particular focus on time series. The problem of choosing optimal block length using Künsch's overlapping block rule is addressed

through both asymptotic arguments and empirical methods. Simulations using AR(1) data are used to empirically determine optimal block lengths for two simple estimators: the variance of the sample mean and the bias of sample variance.

The report is organized as follows. Chapter 1 serves as an introduction to bootstrap theory. Section 1.1 formally introduces the plug-in principle and using Monte Carlo simulation to give bootstrap bias and variance estimates. Section 1.2 derives bootstrap estimates of bias and variance for several simple statistics and compares these to traditional exact estimates. Section 1.3 looks at the accuracy of Monte Carlo simulation and section 1.4 treats bootstrap bias correction.

Chapter 2 deals with bootstrap methodology applied to regression. Section 2.2 introduces model-based resampling and derives bootstrap estimates of variance and bias of least-square linear regression coefficients. These are compared to traditional variance estimates of linear regression coefficients and the Monte Carlo asymptotics of the bootstrap estimates are empirically analysed. Section 2.3 introduces point-based resampling and the two resampling techniques are compared in the context of linear regression through two data-driven experiments. The Monte Carlo asymptotics of point-based are empirically determined and the efficiency of the two estimators are compared.

Chapter 3 surveys the bootstrap applied to time series. Model-based and block resampling techniques are discussed in sections 3.2 and 3.3. The problem of selecting optimal block lengths using non-overlapping (Carlstein's rule) and overlapping (Künsch's rule) is addressed and section 3.3.3 presents an asymptotic argument to estimate mean square error for estimators as a function of block length. A formula for mean square error minimising block length is derived. Section 3.3.4 discusses the Hall, Horowitz and Jing (1995) empirical method of determining optimal block length and presents an alternative method to accurately empirically determine the optimum length for simple estimators. Section 3.3.5 and section 3.3.6 use this alternative method to derive the optimum block length for estimating $Var(\overline{X})$, the variance of the sample mean, and $E\left\{n^{-1}\sum_{i=1}^{n}(X_i-\overline{X})^2\right\}-Var(X)$, the bias of sample variance. The data used is generated from AR(1) processes using various autoregressive coefficients.

1.1 The Main Principle

The basic idea behind the bootstrap is a deceptively simple one, the "plug-in principle". Suppose we wish to estimate a function of an unknown population distribution F, such as the mean

$$\mu = \int x dF(x)$$

The bootstrap estimate of the function employs the empirical distribution $\widehat{F}$ in place of F, which in this example gives us the sample mean

$$\overline{X} = \int x d\widehat{F}(x)$$

In general, for any parameter $\theta = t(F)$ where $t(\cdot)$ is a continuous function, $\widehat{\theta} = t\left(\widehat{F}\right)$ is the plug-in estimate. The empirical distribution function $\widehat{F}$ is defined by

$$\widehat{F} = \frac{\#\{x_i \leq x\}}{n} = \frac{1}{n}\sum_{i=1}^{n} H(y - y_i)$$

where $H(y)$ is the Heaviside function $H(y) = \begin{cases} 0, & y < 0 \\ 1, & y \geq 0 \end{cases}$

Suppose we are interested in the bias β and the variance v of an estimator $\widehat{\theta}$, that is

$$\beta = E(\widehat{\theta}|F) - t(F), \;\; v = Var(\widehat{\theta}|F)$$

The bootstrap estimates of β and v are

$$\widehat{\beta} = E(\widehat{\theta}|\widehat{F}) - t(\widehat{F}), \;\; \widehat{v} = Var(\widehat{\theta}|\widehat{F}) \tag{1.1}$$

In simple cases, exact expressions for $\widehat{\beta}$ and $\widehat{v}$ can be determined symbolically. In most cases however, some form of Monte Carlo simulation is needed.

Monte Carlo bootstrap methods depend on the notion of a *bootstrap resample.* Call the original sample $\chi = (X_1, ..., X_n)$. The bootstrap resample $\chi^* = (X_1^*, ..., X_n^*)$ is a random sample of size n drawn from $\widehat{F}$. That is, each X_i is drawn independently with replacement and has an equal $\frac{1}{n}$ probability of appearing in χ^*. Corresponding to a bootstrap resample χ^* is a bootstrap replication of $\widehat{\theta} = \theta[\chi]$,

$$\widehat{\theta}^* = \theta[\chi^*]$$

The statistic $\widehat{\theta}^*$ is repeatedly calculated B times by randomly drawing new resamples[1]. Properties of $\widehat{\theta} - \theta$ are then estimated from $\widehat{\theta}_1^* = \theta[\chi_1^*], ..., \widehat{\theta}_B^* = \theta[\chi_B^*]$. For example, the bias estimate $\widehat{\beta} = E(\widehat{\theta}|\widehat{F}) - t(\widehat{F})$ is estimated by

$$\widehat{\beta}^* = \frac{\sum_{b=1}^{B} \widehat{\theta}_b^*}{B} - \widehat{\theta} \tag{1.2}$$

and the variance estimate $\widehat{v}$ is estimated by

$$\widehat{v}^* = \frac{1}{B-1}\sum_{b=1}^{B}\left(\widehat{\theta}_b^* - \overline{\widehat{\theta}}^*\right)^2 \tag{1.3}$$

These estimators are justified by the law of large numbers. As $B \to \infty$, $\widehat{\beta}^* \to \widehat{\beta}$ and $\widehat{v}^* \to \widehat{v}$. This is formally shown in section 1.3.

1.2 Examples

This section considers two simple examples of applying the bootstrap: estimating mean, μ, and variance, σ^2. The plug-in principle is used derive estimators for μ and σ^2, and bootstrap esti-

[1] There are in fact $\binom{2n-1}{n}$ different possible resamples that can be drawn from n original data. This is shown in the appendix.

mates of variance and bias of these estimators are determined. These are compared to traditional estimators of mean and variance.

1.2.1 Sample mean

Suppose we draw a random sample $\chi = \{X_1, ..., X_n\}$ of size n from an unknown probability distribution F. The plug-in estimate of the population mean $E[X]$ is

$$\begin{aligned}\widehat{\theta} &= E_{\widehat{F}}[X] = \int x d\widehat{F}(x) = \int x d\left\{\frac{1}{n}\sum_{i=1}^{n} H(x - x_i)\right\} \\ &= \frac{1}{n}\sum_{i=1}^{n}\int x dH(x - x_i) = \frac{\sum_{i=1}^{n} X_i}{n}\end{aligned}$$

the sample mean. The exact bias of $\overline{X}$ is given by

$$\beta(\overline{X}) = E\left\{\frac{\sum_{i=1}^{n} X_i}{n}\right\} - E(X) = \frac{nE(X)}{n} - E(X) = 0$$

and the variance of $\overline{X}$ is given by

$$Var(\overline{X}) = \frac{Var\left(\sum_{i=1}^{n} X_i\right)}{n^2} = \frac{nVar(X)}{n^2} = \frac{\sigma^2}{n}$$

Since σ^2 is generally unknown, the unbiased estimate $S^2 = (n-1)^{-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$ is used, giving the estimate $\widehat{v}(\overline{X})$ where

$$\widehat{v}(\overline{X}) = \frac{1}{n(n-1)}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$$

The bootstrap estimator of bias and variance, given by (1.2) and (1.3), are

$$\widehat{\beta}^*(\overline{X}) = E^*\left\{\sum_{b=1}^{B}\overline{X}_b^*\right\} - \overline{X} = E^*\left\{\overline{X}^*\right\} - \overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} - \overline{X} = 0$$

$$\widehat{v}^*(\overline{X}) = Var^*(\overline{X}^*) = \frac{1}{n}E^*\left\{X^* - E^*(X^*)\right\}^2 = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{n}\left(X_i - \overline{X}\right)^2 = \frac{n-1}{n}\widehat{v}$$

where $E^*(\cdot) = E(\cdot|\widehat{F})$ and $Var^*(\cdot) = Var(\cdot|\widehat{F})$.

The bootstrap estimate of bias is identical to the exact bias of zero, however the bootstrap variance estimate is slightly biased by $-n^{-1}\widehat{v}$.

1.2.2 Sample variance

Suppose we draw a random sample $\chi = \{X_1, ..., X_n\}$ of size n from an unknown probability distribution F. The plug-in estimate $\widehat{v}$ of the population variance $\sigma^2 = Var(X) = \int x^2 dF(x) -$

$\left\{\int xdF(x)\right\}^2$ is

$$\begin{aligned}
\widehat{v} &= \int x^2 d\widehat{F}(x) - \left\{\int xd\widehat{F}(x)\right\}^2 \\
&= \int x^2 d\frac{1}{n}\sum_{i=1}^{n} H(x-x_i) - \left\{\int xd\frac{1}{n}\sum_{i=1}^{n} H(x-x_i)\right\}^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\int x^2 dH(x-x_i) - \left\{\frac{1}{n}\sum_{i=1}^{n}\int xdH(x-x_i)\right\}^2 \\
&= \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \left\{\frac{1}{n}\sum_{i=1}^{n} X_i\right\}^2 \\
&= \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 \qquad (1.4)
\end{aligned}$$

The bias of $\widehat{v}$ is

$$\begin{aligned}
\beta(\widehat{v}) &= E\left\{\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2\right\} - \sigma^2 \\
&= E(X^2) - E(\overline{X}^2) - \sigma^2 \qquad (1.5) \\
&= \sigma^2 + \mu^2 - n^{-1}\sigma^2 - \mu^2 - \sigma^2 \\
&= -n^{-1}\sigma^2
\end{aligned}$$

Thus $\widehat{v}$ is downward biased.

The variance of $\widehat{v}$ (Kenny and Keeping, 1951)is

$$Var(\widehat{v}) = \frac{(n-1)[(n-1)\mu_4 - (n-3)\sigma^4]}{n^3} \qquad (1.6)$$

where $\mu_4 = \int (x-\mu)^4 dF$, the forth central moment.

The bootstrap estimators of bias and variance of $\widehat{v}$, given by (1.2) and (1.3), are

$$\begin{aligned}
\widehat{\beta}^*(\widehat{v}) &= E^*\left\{\frac{1}{n}\sum_{i=1}^{n}\left(X_i^* - \overline{X}^*\right)^2\right\} - \widehat{v} \\
&= E^*\left\{\left(X_i^* - \overline{X}^*\right)^2\right\} - \widehat{v} \\
&= E^*[X_i^{*2}] - 2E[X_i^*\overline{X}^*] + E[\overline{X}^{*2}] - \widehat{v} \\
&= \widehat{v} + \overline{X}^2 - n^{-1}\widehat{v} - \overline{X}^2 - \widehat{v} \\
&= -n^{-1}\widehat{v} \qquad (1.7)
\end{aligned}$$

and

$$\widehat{v}^*(\widehat{v}) = \frac{(n-1)[(n-1)\widehat{\mu}_4 - (n-3)\widehat{v}^2]}{n^3} \qquad (1.8)$$

The bootstrap estimates of bias and variance of $\widehat{v}$ are slightly biased since $\widehat{v}$ is substituted in place of σ^2. From (1.5), we know that $\widehat{v}$ is itself a biased estimator of σ^2.

1.3 Simulation Error

The example of bootstrapping the sample mean and sample variance are special cases where bootstrap estimates of $\widehat{v}$ and $\widehat{\beta}$ can be obtained exactly. In most other cases Monte Carlo simulation is necessary, however this introduces *simulation error* into our estimates. Consider the bootstrap bias estimator $\widehat{\beta}^*$. The Monte Carlo variability in $B^{-1}\sum_{b=1}^{B}\widehat{\theta}_b^*$ can only be removed by infinite simulation, which is impossible. Fortunately, we can gain an idea of what an appropriate value for B is in general by considering the bias and variance of bootstrap bias and variance estimators.

First consider $\widehat{\beta}_B^*(\widehat{\theta}) = B^{-1}\sum_{b=1}^{B}\widehat{\theta}_b^* - \widehat{\theta}$. Let $E^*(\cdot) = E(\cdot|\widehat{F})$ and $Var^*(\cdot) = Var(\cdot|\widehat{F})$. The expectation of this estimator given $\widehat{F}$ is

$$E^*[\widehat{\beta}_B^*(\widehat{\theta})] = E^*\left\{B^{-1}\sum_{b=1}^{B}\widehat{\theta}_b^* - \widehat{\theta}\right\} = E^*[\widehat{\theta}^*] - \widehat{\theta} = 0$$

since $E^*[\widehat{\theta}^*] = \widehat{\theta}$, and the variance of the estimator is

$$Var^*[\widehat{\beta}_B^*(\widehat{\theta})] = Var^*\left\{B^{-1}\sum_{b=1}^{B}\widehat{\theta}_b^* - \widehat{\theta}\right\} = \frac{BVar^*(\widehat{\theta}^*)}{B^2} = \frac{\widehat{v}(\widehat{\theta})}{B} \tag{1.9}$$

where $\widehat{v}(\widehat{\theta})$ is the plug-in estimate for $Var(\widehat{\theta})$. For example, where $\widehat{\theta} = \overline{X}$ we have the mean of the bias equal to zero, thus the bias of the bias estimator is $0 - \big(E(\overline{X}) - E(X)\big) = 0$. The variance of bias is equal to $\frac{\widehat{v}}{nB}$ where $\widehat{v}$ is $\frac{n-1}{n}S^2$.

Now consider $\widehat{v}_B^*(\widehat{\theta}) = \frac{1}{B-1}\sum_{b=1}^{B}\left(\widehat{\theta}_b^* - \overline{\widehat{\theta}^*}\right)^2$. The mean of this estimator is

$$\begin{aligned}
E^*\left\{\frac{1}{B-1}\sum_{b=1}^{B}\left(\widehat{\theta}_b^* - \overline{\widehat{\theta}^*}\right)^2\right\} &= \frac{B}{B-1}E^*\left\{\left(\widehat{\theta}_b^* - \overline{\widehat{\theta}^*}\right)^2\right\} \\
&= \frac{B}{B-1}Var^*\left(\widehat{\theta}_b^* - \overline{\widehat{\theta}^*}\right) \\
&= \frac{B}{B-1}\left\{Var^*(\widehat{\theta}_b^*) - B^{-1}Var^*(\widehat{\theta}_b^*)\right\} \\
&= Var^*(\widehat{\theta}_b^*) \\
&= \widehat{v}(\widehat{\theta})
\end{aligned} \tag{1.10}$$

Therefore the bias of $\widehat{v}_B^*(\widehat{\theta})$ is zero. The variance of $\widehat{v}_B^*(\widehat{\theta})$ is given by

$$Var^*\left\{\widehat{v}_B^*(\widehat{\theta})\right\} = E^*\left\{\widehat{v}_B^*(\widehat{\theta})^2\right\} - E^*\left\{\widehat{v}_B^*(\widehat{\theta})\right\}^2$$

We know $E^*\{\widehat{v}_B^*(\widehat{\theta})\} = Var^*(\widehat{\theta}_b^*)$ from (1.10), so it remains only to find $E^*\{\widehat{v}_B^*(\widehat{\theta})^2\}$. It is worth noting that this problem is almost identical to finding $E(S^2)^2$. From section 1.2.2, equation (1.6),

$$E\left(S^2\right)^2 = E\left\{\frac{1}{(n-1)^2}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^4\right\} = \frac{(n-1)\mu_4 + (n^2-2n+3)\mu_2^2}{n(n-1)}$$

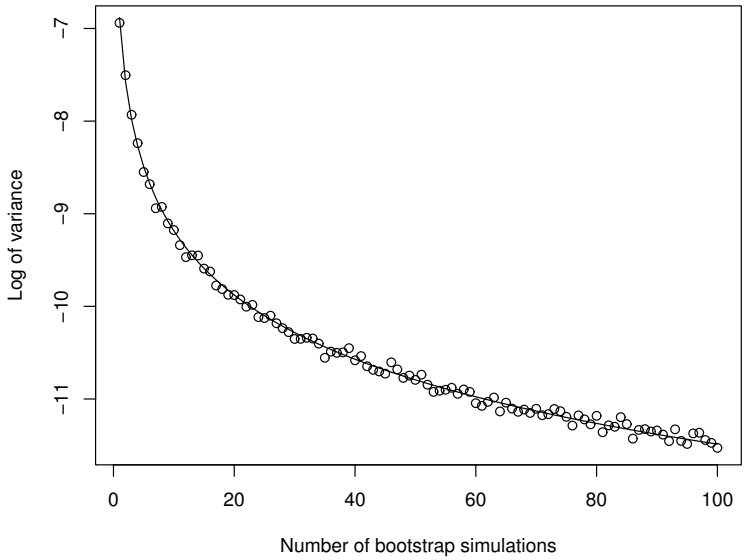

Figure 1.1: *Bootstrap estimates of variance of bias* $\widehat{\beta}(\overline{X})$ *for B from 1 to 100*

The bootstrap version is

$$E^*\left\{\widehat{v}_B^*(\widehat{\theta})^2\right\} = E^*\left\{\frac{1}{(B-1)^2}\sum_{b=1}^{B}\left(\widehat{\theta}_b^* - \overline{\widehat{\theta}^*}\right)^4\right\} = \frac{(B-1)\widehat{\mu}_4 + (B^2-2B+3)\widehat{\mu}_2^2}{B(B-1)} \tag{1.11}$$

Combining (1.10) and (1.11) gives

$$Var^*\left\{\widehat{v}_B^*(\widehat{\theta})\right\} = \frac{(B-1)\widehat{\mu}_4 + (B-3)\left\{\widehat{v}(\widehat{\theta})\right\}^2}{B(B-1)}$$

A simple simulation was carried out to confirm the result (1.9). A thousand bootstrap estimates of the bias of $\overline{X}$ in estimating μ were generated for each value of B between 1 and 100. The sample variance of the thousand estimates was recorded. By (1.9), the expected variance is given by $\frac{\widehat{v}^2}{nB}$. Figure 1.1 shows the results. The sample variance of bias estimates for each B are shown by points and the theoretical variances corresponding to different B are indicated by the line. It is clear that the expected value of $\frac{\widehat{v}}{nB}$ matches the observed results well.

1.4 Bias Correction

This section shows how the bootstrap can be used to correct bias in estimators. Prediction error estimation is one important problem in which bias correction is useful.

Consider a biased estimator $\widehat{\theta}$ for statistic θ with bias $\beta = E(\widehat{\theta}) - \theta$. Suppose we wish to correct the bias of $\widehat{\theta}$ by subtracting a constant t, making the bias-corrected estimator

$$\widehat{\theta}_{bc} = \widehat{\theta} - t$$

For $E[\widehat{\theta}_{bc}] = \theta$, we require $t = E(\widehat{\theta}) - \theta$, thus $t = \beta$ and $\widehat{\theta}_{bc} = 2\theta - E(\widehat{\theta})$. A bootstrap estimate of β and $\widehat{\theta}_{bc}$ is

$$\begin{aligned} \widehat{\beta} &= E^*(\widehat{\theta}^*) - \widehat{\theta} \\ \widehat{\theta}_{bc} &= 2\widehat{\theta} - E^*(\widehat{\theta}^*) \end{aligned}$$

The estimator $\widehat{\theta}_{bc}$ will be itself biased since $\widehat{\beta}$ is biased - it plugs in the biased $\widehat{\theta}$ in place of θ. We may *iterate* bias correction so that each iteration corrects the bias in the previous correction. Let the pth corrected estimator be $\widehat{\theta}_{bc}^{(p)}$ and $E^{**}(\widehat{\theta}^{**})$ denote the expectation of $\widehat{\theta}$ over the distribution $\widehat{F}^{**}$, the EDF of a sample drawn from $\widehat{F}^*$. The second iterated bias correction is given by

$$\begin{aligned} \widehat{\theta}_{bc}^{(2)} &= \widehat{\theta} - (\widehat{\beta} - \widehat{\beta}[\widehat{\beta}]) = \widehat{\theta} - \widehat{\beta} + \widehat{\beta}[\widehat{\beta}] \\ &= 2\widehat{\theta} - E^*\left(\widehat{\theta}^*\right) + E^*\left(\widehat{\beta}^*\right) - \widehat{\beta} \\ &= 2\widehat{\theta} - E^*\left(\widehat{\theta}^*\right) + E^*\left[E^{**}\left(\widehat{\theta}^{**}\right) - \widehat{\theta}^*\right] - E^*\left(\widehat{\theta}^*\right) + \widehat{\theta} \\ &= 3\widehat{\theta} - 3E^*\left(\widehat{\theta}^*\right) + E^*\left[E^{**}\left(\widehat{\theta}^{**}\right)\right] \end{aligned}$$

and the third bias correction by

$$\begin{aligned} \widehat{\theta}_{bc}^{(3)} &= \widehat{\theta} - (\widehat{\beta} - (\widehat{\beta}[\widehat{\beta}] - \widehat{\beta}\{\widehat{\beta}[\widehat{\beta}]\})) \\ &= \widehat{\theta} - \widehat{\beta} + \widehat{\beta}[\widehat{\beta}] - \widehat{\beta}\{\widehat{\beta}[\widehat{\beta}]\} \\ &= 3\widehat{\theta} - 3E^*\left(\widehat{\theta}^*\right) + E^*\left[E^{**}\left(\widehat{\theta}^{**}\right)\right] - E^*\left[\widehat{\theta}^* - 2E^{**}\left(\widehat{\theta}^{**}\right) + \right. \\ &\quad \left. ...E^{**}\left\{E^{***}\left(\widehat{\theta}^{***}\right)\right\}\right] - \widehat{\theta} + 2E^*\left(\widehat{\theta}^*\right) - E^*\left\{E^{**}\left(\widehat{\theta}^{**}\right)\right\} \\ &= 3\widehat{\theta} - 3E^*\left(\widehat{\theta}^*\right) + E^*\left[E^{**}\left(\widehat{\theta}^{**}\right)\right] - \left(3E^*\left(\widehat{\theta}^*\right) - 3E^*\left\{E^{**}\left(\widehat{\theta}^{**}\right)\right\} + \right. \\ &\quad \left. E^*\left[E^{**}\left\{E^{***}\left(\widehat{\theta}^{***}\right)\right\}\right] - \widehat{\theta}\right) \\ &= 4\widehat{\theta} - 6E^*\left(\widehat{\theta}^*\right) + 4E^*\left[E^{**}\left(\widehat{\theta}^{**}\right)\right] - E^*\left[E^{**}\left\{E^{***}\left(\widehat{\theta}^{***}\right)\right\}\right] \end{aligned}$$

An inductive argument may be used to show that the pth iterated estimate is given by, (Hall, 2004)

$$\widehat{\theta}_{bc}^{(p)} = \sum_{i=1}^{p+1} \begin{pmatrix} p+1 \\ i \end{pmatrix} (-1)^{i+1} E\{\theta(F_i)|F_1\} \tag{1.12}$$

where $E\{\theta(F_1)|F_1\} = \widehat{\theta}$, $E\{\theta(F_2)|F_1\} = E^*(\widehat{\theta}^*)$, $E\{\theta(F_3)|F_1\} = E^*[E^{**}(\widehat{\theta}^{**})]$ and so on. In general, $\widehat{\theta}_{bc}^{(0)} = \widehat{\theta}$ has bias of order n^{-a} and $\widehat{\theta}_{bc}^{(p)}$ has bias of order $n^{-(p+a)}$, where a is typically $\frac{1}{2}$, 1 or $\frac{3}{2}$ (Davidson & Hinkley, 1997).

There is a penalty to be paid for iterated bias reduction, that is, variance usually increases. If bootstrap biases are obtained through simulation, each estimate of $\widehat{\theta}_{bc}^{(p)}$ is determined by estimating $\beta[\widehat{\theta}_{bc}^{(p-1)}]$, which is in turn determined by estimating $\beta[\widehat{\theta}_{bc}^{(p-2)}]$ and so on. One problem with this is that additive simulation errors are likely to be accumulated by nesting bootstrap simulations. The other is that after a certain number of iterations, the resamples will converge to lists of repeats of a single observation which are of little use to estimate bias. It can be shown that this rate

of convergence is asymptotic to e^{-1}. This is particularly a problem with small data sets where convergence to a constant resample may happen with relatively few iterations.

1.4.1 Example: estimating σ^2

Suppose σ^2 is estimated by the bootstrap estimator $\widehat{v} = n^{-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$. Section 1.2.2 shows that the bias of this estimator is $\beta = -n^{-1}\sigma^2$, which the bootstrap estimates as $\widehat{\beta} = t = -n^{-1}\widehat{v}$. The bias-corrected bootstrap estimate is therefore

$$\begin{aligned} \widehat{v}_{bc} &= \widehat{v} + n^{-1}\widehat{v} \\ &= \frac{n+1}{n}\widehat{v} \end{aligned} \tag{1.13}$$

(1.13) is not the usual unbiased estimate $S^2 = \frac{n}{n-1}\widehat{v}$ since the bootstrap bias estimate is itself biased. It plugs in the biased $\widehat{v}$ for σ^2 in $\widehat{\beta}$. The bias of $\widehat{\beta}$, $\beta[\widehat{\beta}]$, is given by

$$\begin{aligned} E[\widehat{\beta}] - \beta &= -n^{-2}(n-1)\sigma^2 + n^{-1}\sigma^2 \\ &= n^{-2}\sigma^2 \end{aligned}$$

The bootstrap estimates this as $\widehat{\beta}[\widehat{\beta}] = n^{-2}\widehat{v}$, and the bias-corrector t can be adjusted such that

$$\begin{aligned} t^{(2)} &= \widehat{\beta} - \widehat{\beta}[\widehat{\beta}] \\ &= -n^{-1}\widehat{v} - n^{-2}\widehat{v} \end{aligned}$$

This is an improvement since

$$E(t) = -n^{-2}(n+1)\sigma^2 = -(1+n^{-1})n^{-1}\sigma^2 = \left(1+n^{-1}\right)\beta$$

$$E(t^{(2)}) = -n^{-2}(n+1)\sigma^2 - n^{-3}(n+1)\sigma^2 = -(1+n^{-2})n^{-1}\sigma^2 = \left(1+n^{-2}\right)\beta$$

In general, the *pth* corrected t is given by

$$t^{(p)} = \sum_{i=1}^{p} -n^{-i}\widehat{v}$$

This may be shown from (1.12), however I will prove it directly by induction. Let the inductive hypothesis be $t^{(p-1)} = \sum_{i=1}^{p-1} -n^{-i}\widehat{v}$. Then by definition of $t^{(p)}$ it follows that

$$\begin{aligned} t^{(p)} &= t^{(p-1)} - \widehat{\beta}\left(t^{(p-1)}\right) \\ &= t^{(p-1)} - \left[E^*\left(t^{(p-1)}|\widehat{F}\right) - \widehat{\beta}\left(\widehat{v}\right)\right] \\ &= t^{(p-1)} - n^{-1}(n-1)t^{(p-1)} - n^{-1}\widehat{v} \\ &= \frac{t^{(p-1)}}{n} - n^{-1}\widehat{v} \\ &= \frac{\sum_{i=1}^{p-1} -n^{-i}\widehat{v}}{n} - n^{-1}\widehat{v} \end{aligned}$$

$$
\begin{aligned}
&= \sum_{i=2}^{p} -n^{-i}\widehat{v} - n^{-1}\widehat{v} \\
&= \sum_{i=1}^{p} -n^{-i}\widehat{v}
\end{aligned}
$$

■

Using this result, we can write the pth corrected estimator $\widehat{v}^{(p)}$ as

$$
\begin{aligned}
\widehat{v}^{(p)} &= \widehat{v} + \sum_{i=1}^{p} n^{-i}\widehat{v} \\
&= \sum_{i=0}^{p} n^{-i}\widehat{v}
\end{aligned}
$$

As $p \to \infty$, $\sum_{i=0}^{p} n^{-i} \to \frac{1}{1-n^{-1}} = \frac{n}{n-1}$. Therefore

$$
\widehat{v}^{(p)} \to \frac{n}{n-1}\widehat{v} = S^2
$$

It is also easy to show that

$$
\begin{aligned}
E[\widehat{v}_{bc}^{(p)}] &= E[\widehat{v} - t^{(p)}] \\
&= \sigma^2 + \beta(\widehat{v}) - \left(1 + n^{-p}\right)\beta(\widehat{v}) \\
&= \sigma^2 + n^{-p}\beta(\widehat{v}) \\
&= \sigma^2 - n^{-(p+1)}\sigma^2
\end{aligned}
$$

hence the bias of $\widehat{v}_{bc}^{(p)}$ is of order $n^{-(p+1)}$ and the bias of $\widehat{v}$ is of order n^{-1}.

Chapter 2

Application in Regression

2.1 Introduction

Regression analysis is one of the most important and frequently used types of statistical analysis. Regression seeks to model the covariance between explanatory and response variables. The use of the bootstrap in regression was first proposed by Efron (1979).

Regression probability models belong in the class of *semiparametric* models. This is where some aspects of the data distribution are specified in terms of parameters, but other aspects are left arbitrary, such as the distribution of disturbances. A regression model consists of a parametric structural component and a non-parametric random variation component. For example, a simple linear model relating the response Y_i to an explanatory X_i is

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

where ε_i are sampled $i.i.d$ from an unknown distribution F. The sample data are a collection of pairs $(X_1, Y_1), ..., (X_n, Y_n)$.

It is often assumed that ε_i are independent with zero mean and constant variance. Where ε_i are normally distributed and the two assumptions are justified, least-squares regression analysis can provide exact and relatively simple methods for inference. However, like in many other cases of statistical modelling, believing these assumptions requires great optimism. For example, the $i.i.d$ assumption is often violated in practice (R. Y. Lau, 1988).

The bootstrap has much to offer in these situations. However, it requires modification. The bootstrap, as presented in section 1, works well as a statistical technique when the data are sampled $i.i.d$. It does not work well, however, when data are structurally dependent, such as in regression. Random resampling from the sample destroys any dependency present and the resampled data are unlikely to model the sampled data appropriately. Modification is necessary in the way resampling is carried out.

Two important methods for resampling regression data have been proposed: *model-based* resampling and *point-based* resampling (Efron 1979, Freedman 1981). Under ideal resampling conditions, both essentially reproduce the exact theoretical analysis, however they also work well in non-ideal cases such as data with heteroscedastic ε_i. Model-based resampling is a parametric

method that is accurate and efficient if the structural model is known and the X_i are considered fixed, whereas point-based resampling is a non-parametric method that is particularly appropriate where X_i are considered as random variables and errors are heteroscedastic.

2.2 Model-based resampling

Model-based resampling uses estimated residuals to form resamples. Suppose that we have a regression model $Y_i = g(X_i) + \varepsilon_i$ where the X_i are fixed and ε_i are sampled from a common error distribution F. The estimated raw residuals are defined as

$$\widehat{\varepsilon}_i = Y_i - \widehat{g}(X_i)$$

If the $\widehat{\varepsilon}_i$ are independent, homoscedastic and have mean of zero (according to the assumptions in 2.1), they can be directly used to construct the empirical error distribution $\widehat{F}$. However in practice $\widehat{\varepsilon}_i$ are usually modified before resampling. For example, a useful variance stabilizing modification in linear regression is

$$r_i = \frac{\widehat{\varepsilon}_i}{(1-h_i)^{1/2}}$$

where h_i is the ith diagonal element of the hat matrix $H = X(X^TX)^{-1}X^T$. To ensure that $\widehat{F}$ has mean zero, the modified residuals are often re-centered by setting $r_i^c = r_i - \overline{r}$. $\widehat{F}$ is then constructed using r_i^c and the resampled $\widehat{\varepsilon}_i^*$ are drawn from $\widehat{F}$.

The resampling algorithm is given as follows,

1. for $b = 1, ..., B$
   (a) for $i = 1, ..., n$,
      i. set $x_i = x_i^*$
      ii. randomly resample $\widehat{\varepsilon}_i^*$ from $r_1^c, ..., r_n^c$
      iii. define $Y_{i,b}^* = \widehat{g}(X_i) + \widehat{\varepsilon}_i^*$
   (b) fit $\widehat{g}_b^*(X_i)$ using $(X_{1,b}^*, Y_{1,b}^*), ..., (X_{n,b}^*, Y_{n,b}^*)$
2. use$\{\widehat{g}_1^*(X_i), ..., \widehat{g}_B^*(X_i)\}$ to estimate the statistic of interest

Model-based resampling gives reliable estimation when the proposed structural model is correct. Since the estimated residuals are highly reliant on the choice of fitted model, great care must be taken in selecting an appropriate model.

2.2.1 Case study: variance of β_1 and β_0

In this section, a simple linear regression example is considered. Suppose we wish to estimate the variance of estimates of β_1 and β_0 in the model

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

where $\varepsilon_i \sim F$ with mean zero and variance σ^2. The least square estimates of the coefficients are given by

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^n (X_i - \overline{X})Y_i}{S_{xx}}, \;\; \widehat{\beta}_0 = \overline{Y} - \widehat{\beta}_1 \overline{X}$$

where $S_{xx} = \sum_{i=1}^n (X_i - \overline{X})^2$. It is easy to show that that the exact variances of β_1 and β_0 are

$$Var(\beta_1) = \frac{\sigma^2}{S_{xx}}, \;\; Var(\beta_0) = \sigma^2 \left(\frac{1}{n} + \frac{\overline{X}^2}{S_{xx}} \right)$$

and that unbiased estimates of these variances are obtained by substituting $S^2 = \frac{1}{n-2} \sum \widehat{\varepsilon}_i^2$

$$\widehat{v}(\beta_1) = \frac{S^2}{S_{xx}} \tag{2.1}$$

$$\widehat{v}(\beta_0) = S^2 \left(\frac{1}{n} + \frac{\overline{X}^2}{S_{xx}} \right) \tag{2.2}$$

Model-based bootstrapping estimates $Var(\beta)$ by $Var^*(\widehat{\beta}^*)$ where $\widehat{\beta}_1^*$ and $\widehat{\beta}_0^*$ are given by

$$\widehat{\beta}_1^* = \frac{\sum_{i=1}^n (X_i - \overline{X})Y_i^*}{S_{xx}} = \widehat{\beta}_1 + \frac{\sum_{i=1}^n (X_i - \overline{X})\widehat{\varepsilon}_i^*}{S_{xx}}$$

$$\begin{aligned} \widehat{\beta}_0^* = \overline{Y}^* - \widehat{\beta}_1^* \overline{X} &= n^{-1} \sum_{i=1}^n \left(\widehat{\beta}_0 + \widehat{\beta}_1 X_i + \widehat{\varepsilon}_i^* \right) \\ &\quad - \widehat{\beta}_1 \overline{X} - \frac{\sum_{i=1}^n (X_i - \overline{X})\overline{X}\widehat{\varepsilon}_i^*}{S_{xx}} \\ &= \widehat{\beta}_0 + n^{-1} \sum_{i=1}^n (\widehat{\varepsilon}_i^*) - \frac{\sum_{i=1}^n (X_i - \overline{X})\overline{X}\widehat{\varepsilon}_i^*}{S_{xx}} \end{aligned}$$

Since $E^*(\widehat{\varepsilon}_i^*) = n^{-1} \sum_{i=1}^n (r_i - \overline{r}) = 0$ and $Var^*(\widehat{\varepsilon}_i^*) = n^{-1} \sum_{i=1}^n (r_i - \overline{r})^2$

$$\begin{aligned} E^*(\widehat{\beta}_1^*) &= \widehat{\beta}_1 + \frac{\sum_{i=1}^n (X_i - \overline{X}) E^*(\widehat{\varepsilon}_i^*)}{S_{xx}} \\ &= \widehat{\beta}_1 \end{aligned}$$

$$\begin{aligned} E^*(\widehat{\beta}_0^*) &= \widehat{\beta}_0 + n^{-1} \sum_{i=1}^n (E^*(\widehat{\varepsilon}_i^*)) - \frac{\sum_{i=1}^n (X_i - \overline{X})\overline{X} E^*(\widehat{\varepsilon}_i^*)}{S_{xx}} \\ &= \widehat{\beta}_0 \end{aligned}$$

Therefore our bootstrap values of $\widehat{\beta}_1^*$ and $\widehat{\beta}_0^*$ are unbiased. The variance of $\widehat{\beta}_1^*$and $\widehat{\beta}_0^*$ is given by

$$Var^*\left(\widehat{\beta}_1^*\right) = \frac{\sum_{i=1}^n (X_i - \overline{X})^2 Var^*(\widehat{\varepsilon}_i^*)}{S_{xx}^2} = \frac{n^{-1} \sum_{i=1}^n (r_i - \overline{r})^2}{S_{xx}} \tag{2.3}$$

$$
\begin{aligned}
Var^*\left(\widehat{\beta}_0^*\right) &= n^{-1}Var\left(\widehat{\varepsilon}_i^*\right) + \frac{\sum_{i=1}^n (X_i-\overline{X})^2\overline{X}^2 Var\left(\widehat{\varepsilon}_i^*\right)}{S_{xx}^2} - 2Cov\left(n^{-1}\sum_{i=1}^n\left(\widehat{\varepsilon}_i^*\right), \frac{\sum_{i=1}^n (X_i-\overline{X})\overline{X}\widehat{\varepsilon}_i^*}{S_{xx}}\right) \\
&= n^{-2}\sum_{i=1}^n (r_i-\overline{r})^2 + \frac{\overline{X}^2 n^{-1}\sum_{i=1}^n (r_i-\overline{r})^2}{S_{xx}} - 2n^{-1}\overline{X}\sum_{i=1}^n\sum_{j=1}^n \frac{(X_i-\overline{X})\widehat{\varepsilon}_j^*}{S_{xx}} Cov(\widehat{\varepsilon}_i^*, \widehat{\varepsilon}_j^*) \\
&= n^{-1}\sum_{i=1}^n (r_i-\overline{r})^2\left(\frac{1}{n}+\frac{\overline{X}^2}{S_{xx}}\right) - 2n^{-1}\overline{X}\sum_{i=1}^n\sum_{j=1}^n \frac{(X_i-\overline{X})}{S_{xx}} Var^*\left(\widehat{\varepsilon}_i^*\right) \\
&= n^{-1}\sum_{i=1}^n (r_i-\overline{r})^2\left(\frac{1}{n}+\frac{\overline{X}^2}{S_{xx}}\right) - 2n^{-2}\overline{X}\sum_{k=1}^n (r_k-\overline{r})^2\sum_{i=1}^n \frac{(X_i-\overline{X})}{S_{xx}} \\
&= n^{-1}\sum_{i=1}^n (r_i-\overline{r})^2\left(\frac{1}{n}+\frac{\overline{X}^2}{S_{xx}}\right) \qquad (2.4)
\end{aligned}
$$

While (2.3) and (2.4) may not be (2.1) and (2.2) exactly, they are in fact very close. To show this, note that $n^{-1}\sum_{i=1}^n h_i = 2n^{-1}$. Expanding $Var^*\left(\widehat{\varepsilon}_i^*\right)$ gives

$$
\begin{aligned}
n^{-1}\sum_{i=1}^n (r_i-\overline{r})^2 &= n^{-1}\sum_{i=1}^n\left(\widehat{\varepsilon}_i\left(1-h_i\right)^{-1/2} - n^{-1}\sum_{i=1}^n \widehat{\varepsilon}_i\left(1-h_i\right)^{-1/2}\right)^2 \\
&\approx n^{-1}\sum_{i=1}^n\left(\widehat{\varepsilon}_i\left(1-\overline{h}\right)^{-1/2}\right)^2 \\
&= n^{-1}\left(1-\overline{h}\right)^{-1}\sum_{i=1}^n\left(\widehat{\varepsilon}_i^2\right) \\
&= n^{-1}\left(1-2n^{-1}\right)^{-1}\sum_{i=1}^n\left(\widehat{\varepsilon}_i^2\right) \\
&= \frac{1}{n-2}\sum_{i=1}^n\left(\widehat{\varepsilon}_i^2\right) = S^2
\end{aligned}
$$

To test the accuracy of bootstrap variance estimation against theoretical results, Monte Carlo bootstrap simulation was performed on two data sets: one simulated and one real-world. The bootstrap estimate of the mean square error (MSE) of $\widehat{\beta}$ is compared to the exact $MSE = Var(\widehat{\beta})$.

The first data set is simulated according to the model

$$Y_i = 3 + 5X_i + \varepsilon_i$$

where $\{X_1, X_2, ..., X_{50}\} = \{1, 2, ..., 50\}$ and $\varepsilon_i \sim i.i.d\ N(0, 40^2)$. The errors are independent, heteroscedastic and have zero mean. The least-squares fit is given by

$$\widehat{Y}_i = -10.2236 + 5.6185X_i$$

The function `modelboot` was used to run the model-based resampling algorithm with $B = 5000$. A plot of the simulated data, original least-squares fit (solid line) and 5000 bootstrap least-squares fits (broken line) is shown in figure 2.1.

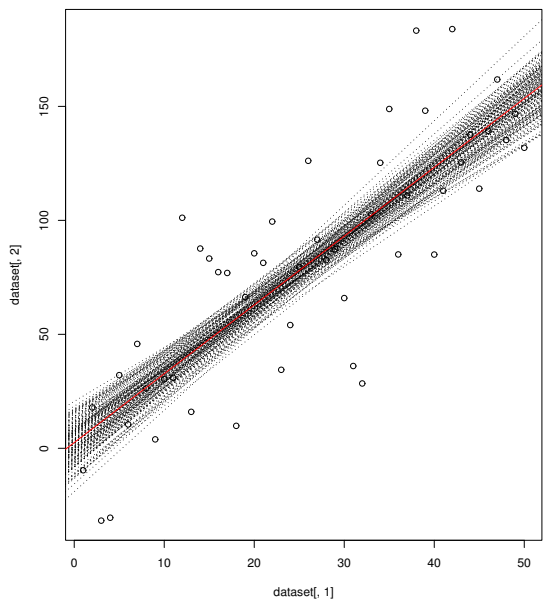

Figure 2.1: *Simulated data with least-squares fit (red) and bootstrap fits (dotted lines).*

		Theoretical	Model-based
$\widehat{\beta}_1^*$	bias	0	-0.01514
	standard error	0.3808	0.379885
	mean square error	0.14501	0.144542
$\widehat{\beta}_0^*$	bias	0	0.001443
	standard error	11.1582	11.15763
	mean square error	124.5054	124.4927

Table 2.1: *Model-based estimates of bias and standard error of $\widehat{\beta}^*$ using simulated data.*

The bootstrap standard errors and bias are compared to theoretical standard errors and bias in table 2.2. It is clear that the bootstrap estimates of standard error are very close to those obtained by theory. In fact, the bootstrap estimates display relative errors of -0.20% for $\widehat{\beta}_1^*$ and -0.010% for $\widehat{\beta}_0^*$ based on the difference of MSE. Notice, however, that I employed a reasonably high number of bootstrap simulations ($B = 5000$). While this is practicable with small data sets, such as this one ($n = 50$), being able to employ smaller values of B without significantly increasing simulation error is attractive.

To investigate how smaller B affects estimation, 100 trials of estimating the MSE of $\widehat{\beta}$ were performed with $B = \{50, 100, ..., 4950, 5000\}$. Bootstrap estimates of the MSE of $\widehat{\beta}_1$ and $\widehat{\beta}_0$ are plotted in figures (2.2) and (2.3). Both plots clearly show that bootstrap estimates converge to their exact values with increasing B, as expected. There seems not to be much difference between the shape of plots of $\widehat{\beta}_1$ or $\widehat{\beta}_0$ and both converge at similar rates. A B value exceeding 1000 seems sufficient in both cases for the estimates to lie within 5% of the exact MSE.

The second data set relates waiting time (T) and eruption duration (D) of the Old Faithful

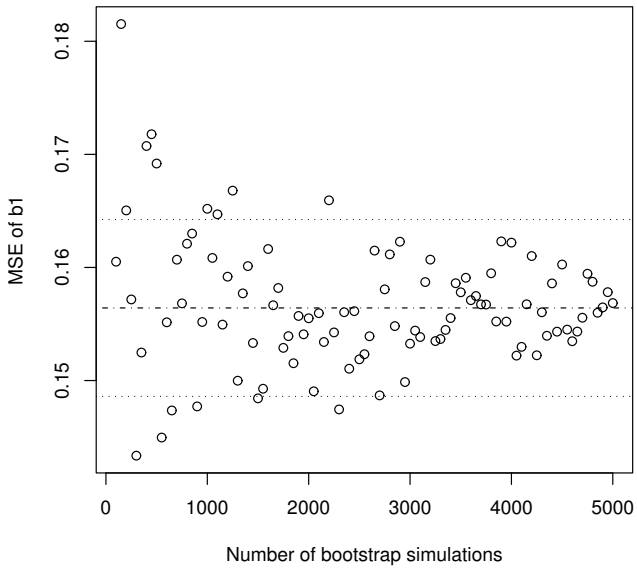

Figure 2.2: *MSE of* $\widehat{\beta}_1$ *with with various* B *using simulated data. The central line is the exact MSE and the outer dotted lines are 5% relative error bands.*

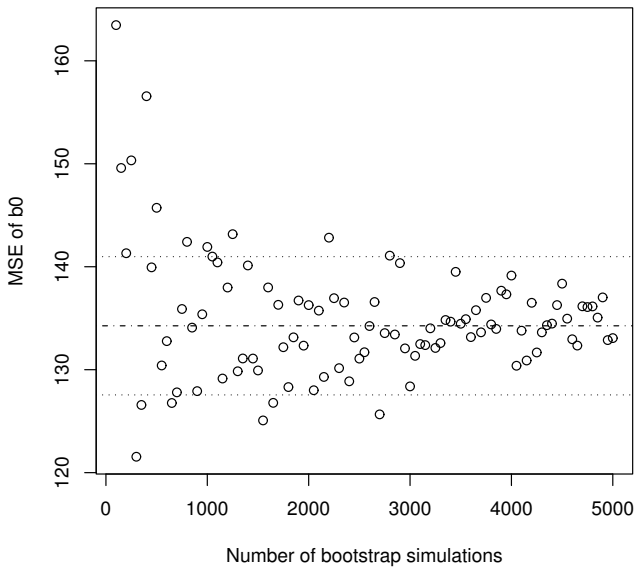

Figure 2.3: *MSE of* $\widehat{\beta}_0$ *with with various* B *using simulated data. The central line is the exact MSE and the outer dotted lines are 5% relative error bands.*

geyser in Yellowstone National Park (Azzalini & Bowman, 1990). The data is plotted in figure 2.4 with line of best fit and 5000 bootstrap refits drawn. A linear model seems appropriate, however its worth noting that measurements of the eruption duration (the explanatory variable) seem to be clustered into two groups. This indicates that the explanatory values are not fixed constants, which is assumed in model-based bootstrapping. The least-squares fit is given by

$$\widehat{T}_i = 33.47 + 10.73 D_i$$

Plots of studentised residuals and the residual quantiles against normal quantiles are shown in figure 2.5. The residuals show no sign of heteroscedasity and appear normal.

The bootstrap standard error and bias estimates are compared to theoretical standard error and bias in table 2.2. The bootstrap estimates of $\widehat{\beta}_1^*$ and $\widehat{\beta}_0^*$ have relative errors of -1.1% and -2.6% based on the difference of mean square errors. Again, we see that bootstrap estimation is accurate at estimating the true variance of $\widehat{\beta}^*$.

The effects of varying B on bootstrap estimates was measured through running 100 trials with $B = \{50, 100, ..., 4950, 5000\}$. The results are shown in figures 2.6 and 2.7.

Both plots exhibit a slower convergence to the exact MSE than in the simulated data example. For $B > 200$ the variance in the scatter of points is reasonably constant, whereas with the the simulated data this scatter variance reduced approximately uniformly with B. A reason for this may be due to the fact that the bootstrap refits are much closer to each other than in the simulated data example (see figure 2.4). Values of B exceeding 1000 give estimates reasonably well contained within the 5% error margins.

In summary, both the simulated and real data support that $B \approx 1000$ is sufficient for bootstrap estimates to be within 5% of the actual value.

2.3 Point-based Resampling

Point-based resampling uses the "points" (X_i, Y_i) to construct resamples. It treats the sample as a bivariate distribution F of X and Y, and accordingly resamples *both* X_i and Y_i randomly as pairs of observations. The resampling algorithm is given as follows,

For $b = 1, 2, ..., B$

1. sample with replacement $i_1^*, i_2^* ..., i_n^*$ from $\{1, 2, ..., n\}$
2. for j in 1 to n, set $(X_j^*, Y_j^*) = (X_{i_j^*}, Y_{i_j^*})$
3. fit $\widehat{g}_b^*(X_i)$ using $\{(X_1^*, Y_1^*), ..., (X_n^*, Y_n^*)\}$
4. use $\{\widehat{g}_1^*(X_i), ..., \widehat{g}_B^*(X_i)\}$ to estimate the statistic of interest

The most important feature of point-based resampling is that it is non-parametric. The resampling procedure is model independent unlike model-based resampling. Consequently, no assumptions of variance homogeneity are necessary. This gives the approach some robustness when dealing with heteroscedastic data. Conversely, where a constant variance model is correct, it can be shown that point-based resampling is less efficient than model-based, that is, the variance of the model-based estimator is less than the variance of the point-based estimator.

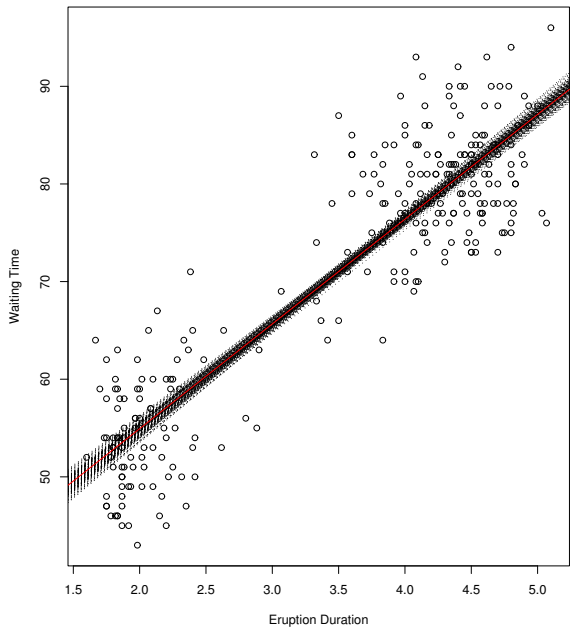

Figure 2.4: *Old Faithful data with least-squares fit (red) and bootstrap fits (dotted lines).*

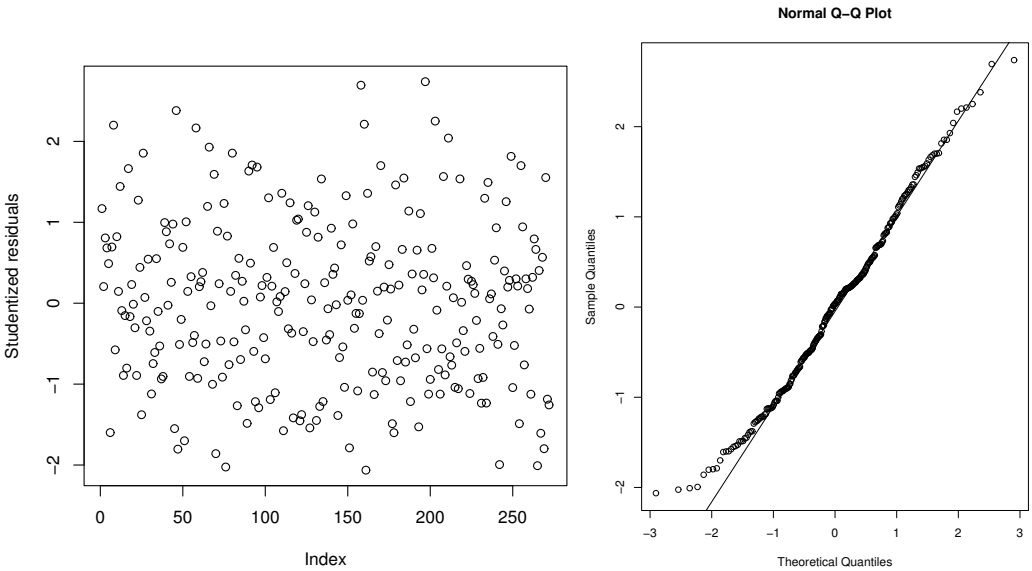

Figure 2.5: *Studentised residuals (left) and quantile-quantile plot (right) of the Old Faithful least-squares fit.*

		Theoretical	Model-based
$\widehat{\beta}_1^*$	bias	0	-0.000735
	standard error	0.31482	0.312476
	mean square error	0.098710	0.097642
$\widehat{\beta}_0^*$	bias	0	0.011637
	standard error	1.15494	1.139582
	mean square error	1.33389	1.298781

Table 2.2: *Bootstrap estimates of bias and standard error of $\widehat{\beta}$ using Old Faithful data.*

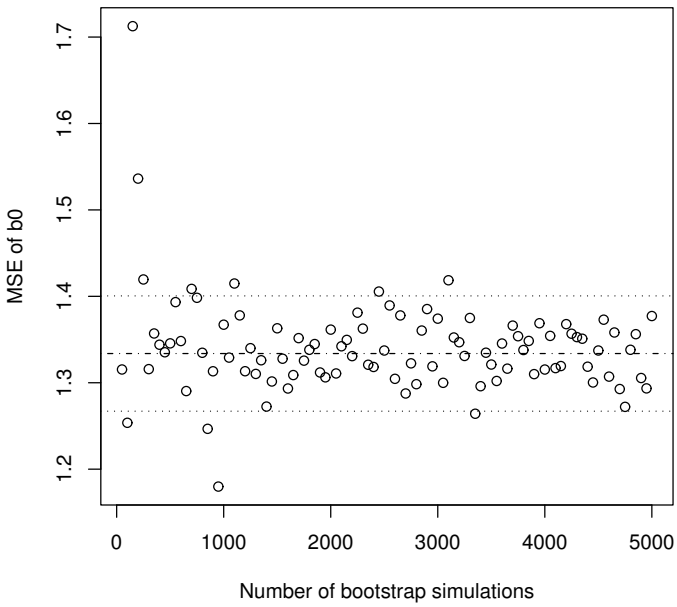

Figure 2.6: *Model-based MSE of $\widehat{\beta}_1$ with with various B using Old Faithful data. The central line is the exact MSE and the outer dotted lines are 5% relative error bands.*

The second important difference is in the design of the resampled sets. The set $\{X_1^*, ..., X_n^*\}$ is a random sample and subset of $\{X_1, ..X_n\}$. Regression inference should regard the information contained in the data. Variation in $\{X_1^*, ..., X_n^*\}$ causes variation in the information used in resamples and will inflate variance estimates. This problem, however, is negligible with reasonably large data sets.

2.3.1 Case study: variance of β_1 and β_0

To compare point-based resampling to model-based resampling, the two data sets used before were employed. The R function `pointboot` was used to run the point-based resampling algorithm with $B = 5000$ for the two data sets. The results are summarised in tables 3 and 4. In addition to the exact theoretical bias and standard error, robust (variance stabilizing) theoretical values are shown. The robust variance estimates of $\widehat{\beta}_1$ and $\widehat{\beta}_0$ are given by

$$Var_R(\widehat{\beta}_1) = \frac{\sum_{i=1}^n (X_1 - \overline{X})^2 r_i^2}{S_{xx}^2}$$

$$Var_R(\widehat{\beta}_0) = \frac{\sum_{i=1}^n r_i^2}{n^2} + \frac{\overline{X}^2 \sum_{i=1}^n (X_1 - \overline{X})^2 r_i^2}{S_{xx}^2}$$

where r_i are variance stabilised residuals defined in Section 2.2. These variance estimates are similar, but not identical to those of $Var^*(\widehat{\beta})$ using model-based resampling.

Point-based estimates of variance in the simulated data set were uniformly and significantly higher than the exact and model-based values - for $\widehat{\beta}_1$, MSE is overestimated by 17% and for $\widehat{\beta}_0$ by 15%. Surprisingly, the opposite happened using Old Faithful - the estimates of variance were uniformly lower: for $\widehat{\beta}_1$ it underestimated by 3.7% and for $\widehat{\beta}_0$ by 3.6%. Conversely, the point-based estimates are *very* close to the robust estimates: the difference using simulated data is less than 0.0015% for $\widehat{\beta}_1$ and 0.09% for $\widehat{\beta}_0$; using Old Faithful data the difference is less than 0.54% for $\widehat{\beta}_1$ and 0.24% for $\widehat{\beta}_0$.

To observe the convergence properties of point-based estimates, plots of estimated MSE for $\widehat{\beta}_1$ and $\widehat{\beta}_0$ with $B = \{50, 100, ..., 5000\}$ using the two data sets were generated. Figures 2.8 and 2.9 use the simulated data, and figures 2.10 and 2.11 use the Old Faithful data.

It is clear from these four plots that point-based bootstrap estimates *do not* converge to the theoretical MSE. Instead they seem to converge to the robust estimates. Figures 2.10 and 2.11 show this particularly well.

A disadvantage of point-based resampling is its relative inefficiency compared to model-based resampling when a constant variance linear model is correct. Davidson and Hinkley (1997, pp. 265) performed a similar comparison between model-based and point-based methods and found that the loss of efficiency is approximately 40%. To test this finding, the Old Faithful data set was used. The variance of $v_B^*(\beta_1)$ was determined for $B = \{10, 20, ..., 1000\}$ using model- and point-based resampling. For each B, 75 repeated simulations were performed and the sample variance taken. If point-based resampling was around 40% less efficient than model-based, we expect to see that the variance of $v_B^*(\beta_1)$ using point-based resampling to be consistently higher (by 40%). The results are plotted in figure 2.12. The light dotted line is shows the point-based variances and the darker dotted line show the model-based variances.

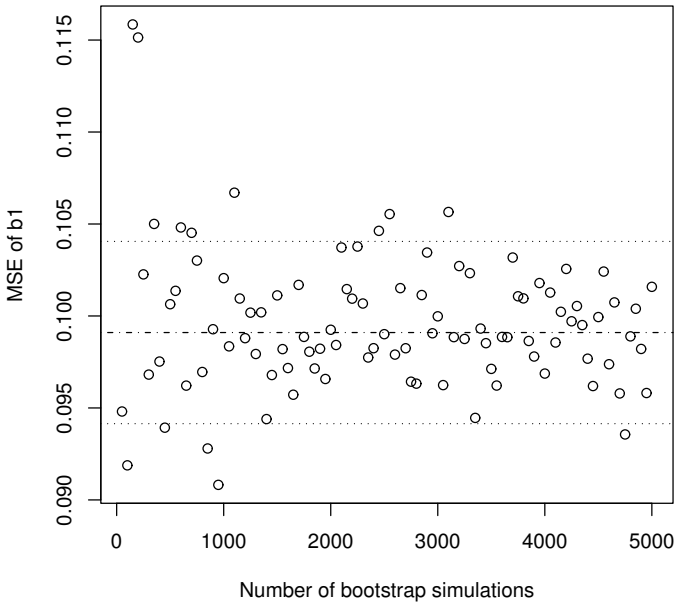

Figure 2.7: *Model-based MSE of $\widehat{\beta}_0$ with various B using Old Faithful data. The central line is the exact MSE and the outer dotted lines are 5% relative error bands.*

		Theoretical	Robust	Model-based	Point-based
$\widehat{\beta}_1^*$	bias	0	≈ 0	-0.01514	-0.004671
	standard error	0.3808	0.41714	0.379885	0.412336
	mean square error	0.14501	0.17004	0.144542	0.1700425
$\widehat{\beta}_0^*$	bias	0	≈ 0	0.001443	0.289857
	standard error	11.1582	11.9781	11.15763	11.98016
	mean square error	124.5054	143.475	124.4927	143.6083

Table 2.3: *Estimated bias and standard error of $\widehat{\beta}^*$ using simulated data*

		Theoretical	Robust	Model-based	Point-based
$\widehat{\beta}_1^*$	bias	0	≈ 0	-0.000735	-0.002338
	standard error	0.31482	0.30107	0.312476	0.301876
	mean square error	0.098710	0.09064	0.097642	0.091133
$\widehat{\beta}_0^*$	bias	0	≈ 0	0.011637	0.009557
	standard error	1.15494	1.10960	1.139582	1.110883
	mean square error	1.33389	1.23121	1.29878	1.234152

Table 2.4: *Estimated bias and standard error of $\widehat{\beta}$ using Old Faithful data.*

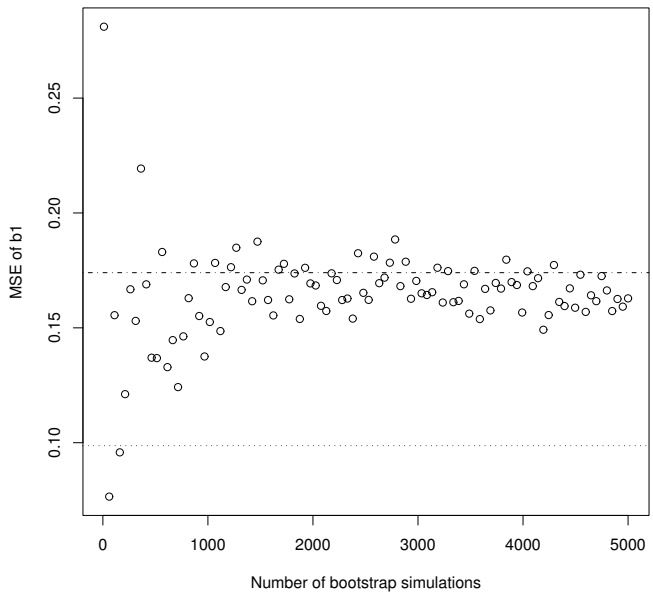

Figure 2.8: *Point-based MSE of* $\widehat{\beta}_1$ *with various* B *using simulated data. The upper line is the robust theoretical MSE; the lower is the exact MSE.*

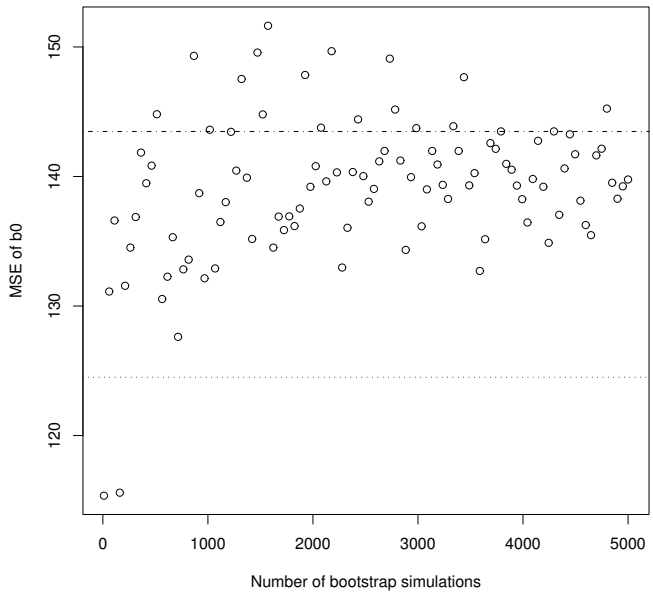

Figure 2.9: *Point-based MSE of* $\widehat{\beta}_0$ *with various* B *using simulated data. The upper line is the robust theoretical MSE; the lower is the exact MSE.*

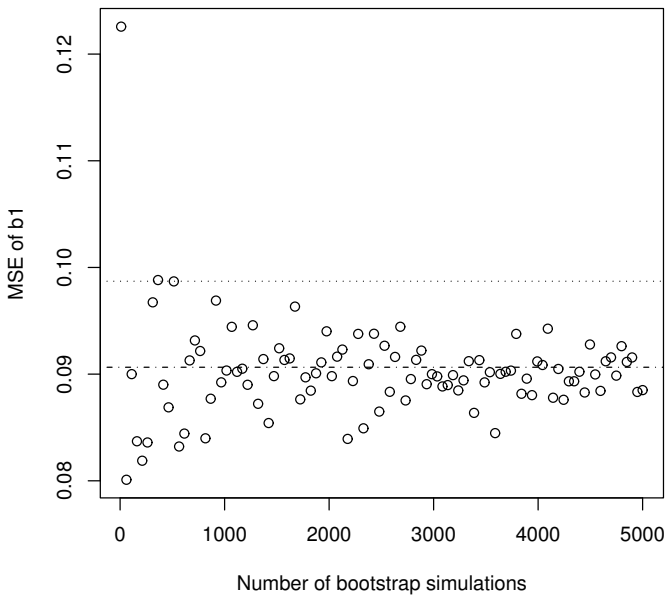

Figure 2.10: *Point-based MSE of $\widehat{\beta}_1$ with various B using Old Faithful data. The lower line is the robust theoretical MSE; the upper is the exact MSE.*

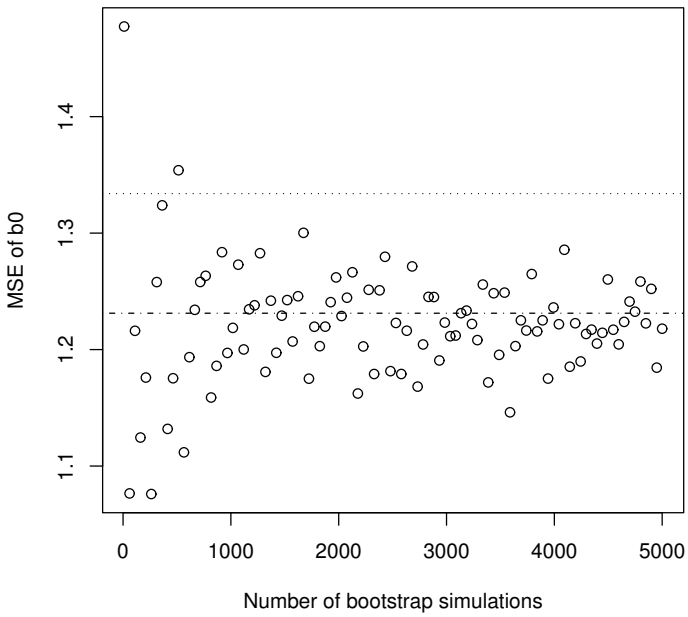

Figure 2.11: *Point-based MSE of $\widehat{\beta}_0$ with various B using Old Faithful data. The lower line is the robust theoretical MSE; the upper is the exact MSE.*

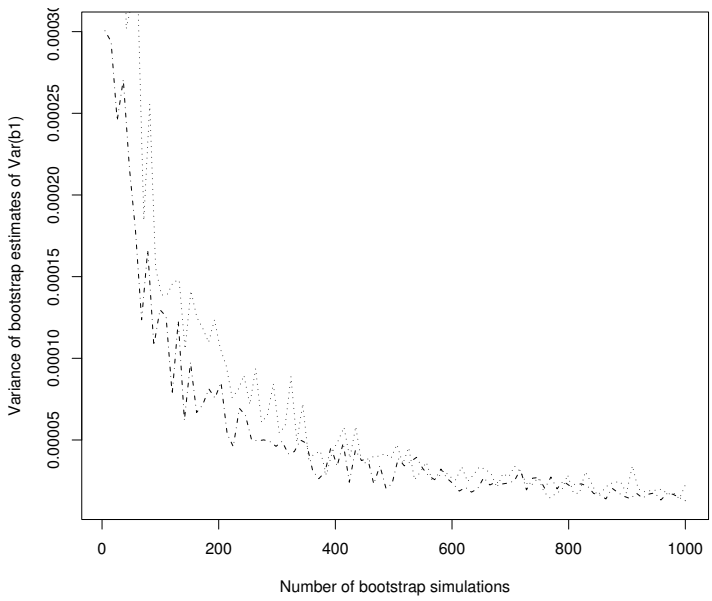

Figure 2.12: *Variance of $v_B^*(\beta_1)$ of the Old Faithful data using model-based (broken line) and point-based (dotted line) for $10 \leq B \leq 1000$.*

There is some difference between the variances of model-based and point-based resampling for each B. The plot shows that point-based variances are higher with low B, however for $B > 500$, there is little difference. The estimated inefficiency of point-based based on all the variance values is approximately 25%, however if we only consider points with $B > 500$, the inefficiency drops to only 3%.

In summary, point-based estimates appear to converge to robust theoretical estimates. This is especially clear with estimating β_1 and β_0 using the Old Faithful data. When data are homoscedastic, the empirical inefficiency of point-based resampling compared to model-based is low - only 3% using the Old Faithful data for $B > 500$.

Chapter 3

Application in Time Series

3.1 Introduction

Analysing time series data is a tremendously important in numerous areas, particularly in financial and applied statistical research. The subject of time series analysis is a relatively new area of statistics and its theoretical groundwork is less mature than that of regression. However, as we have seen with applying the bootstrap in regression, accurate results are easily obtainable using Monte Carlo simulation, and importantly, in areas where classical theory is held back by limits of mathematical tractability. The bootstrap applied to the time domain was first proposed by Freedman (1984) and has since experienced a great deal of research interest. See Li and Maddala (1996) for a detailed survey.

A time series is a sequence of successive observations measured over a set of times. The measured times are typically discrete and equally spaced. Time series are characterised by time dependency between neighbouring observations. Often this dependency structure is very complex. For example, the observation X_i of a time series $X = \{X_0, .., X_n\}$ may be serially correlated to X_{i-1} according to

$$X_i = \alpha X_{i-1} + \varepsilon_{i-1}\beta + \varepsilon_i$$

for $i \geq 1$, where α is a constant and$\{\varepsilon_i\}$is a white noise process of errors with $E(\varepsilon_i) = 0$ and $Var(\varepsilon_i) = \sigma^2$. This particular relation is called a ARMA(1,1) process.

Most models for time series assume that the observations are stationary. This means the joint distribution of any subset of observations depends only on the relative times of occurrence to each other, and not the absolute position in the series. Formally, a time series $\{X_0, ..., X_n\}$ is stationary if

1. $E[X_i]^2 < \infty \ \forall \ i \in \{0, ..., n\}$
2. $E[X_i] = \mu \ \forall \ i \in \{0, ..., n\}$
3. $Cov(X_r, X_s) = Cov(X_{r+i}, X_{s+i}) \ \forall \ (r+i), (s+i) \in \{0, ..., n\}$

Observations from a stationary time series vary around a common mean and are contained within a finite bound, within which the series does not consistently drift upwards or downwards. Standard

statistical properties such as mean and variance are constant over time. This property is an important initial step for meaningful statistical analysis to be performed. However, raw collection data are rarely stationary, and the first steps in analysing data in these cases is to "decompose" or extract elements from the data that *are* stationary.

The classical "decomposition" model considers the realizations X_i as the combination

$$X_i = m_i + s_i + Y_i \tag{3.1}$$

where m_i is a slowly changing function known as a "trend component", s_i is a function with known period d referred to as a "seasonal component", and Y_i is a "random noise" or "residual" component which is stationary. Two famous and widely-used techniques that perform classical model decomposition are *SABL* and *Box-Jenkins.*

3.1.1 SABL decomposition

SABL (Seasonal Adjusted Bell Laboratories) uses *lowess*, a robust non-linear regression filter, to decompose the time series into a m_i component and an s_i component. The m_i is estimated by fitting a smoothed lowess regression through the original data. The s_i is estimated by applying a differencing operator with lag d and fitting another lowess regression. The desired component Y_i is then estimated by

$$\widehat{Y}_i = X_i - \widehat{m}_i - \widehat{s}_i$$

SABL was designed especially to handle data plagued with outliers and is attractively robust. See Cleveland et. al. (1981) for a detailed discussion.

An example is shown in figure (3.1) using monthly atmospheric CO_2 concentration data from 1959 to 1997 (Keeling and Whorf, 1997). An AR(8) model for the residuals is selected using Akaike's information criterion (AIC).

3.1.2 Box-Jenkins decomposition

An alternative approach, developed extensively by Box and Jenkins (1970), is to apply a difference operator repeatedly to data $\{X_i\}$ until the differenced observations resemble a realisation of some stationary process $\{Y_i\}$. To see how this works, let us define the first difference operator ∇_1 by

$$\nabla_1 X_i = X_i - X_{i-1}$$

and the kth difference operator ∇_k by

$$\nabla_k X_i = X_i - X_{i-k}$$

If the operator ∇_1 is applied to a linear trend function $m_t = at + b$, then we obtain the constant function $\nabla_1 m_t = a$. Similarly, a k degree polynomial can be reduced to a constant by applying ∇_1^k.

Now suppose we have data distributed according to (3.1) where $\{s_i\}$ has period d. Applying

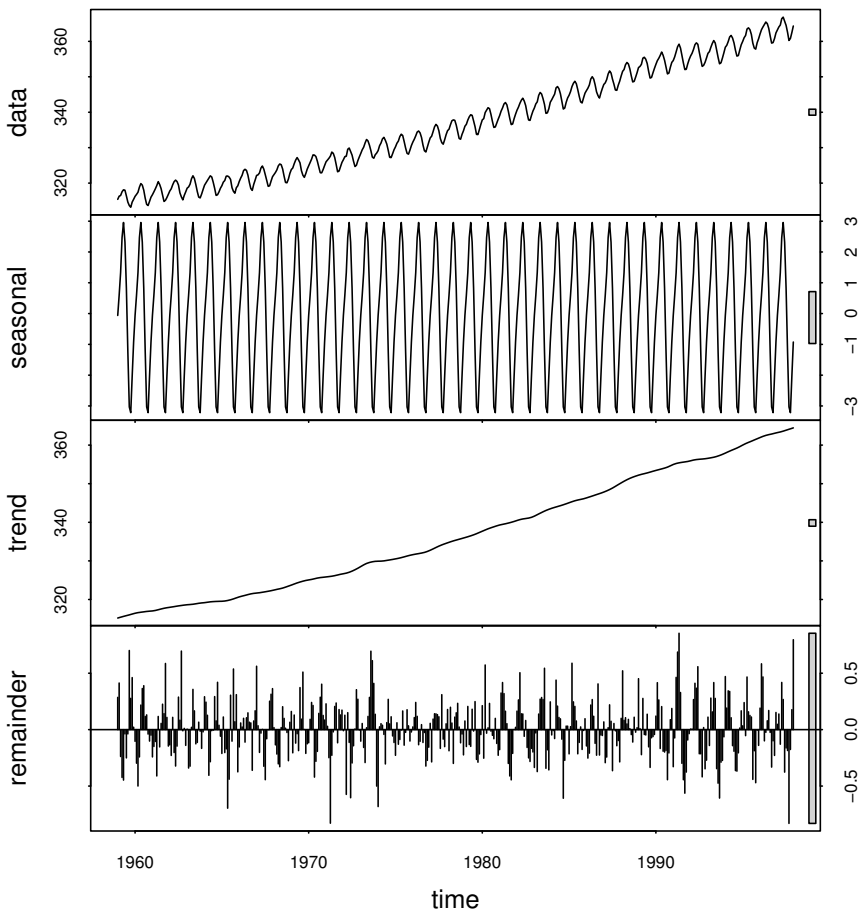

Figure 3.1: *SABL decomposition of monthly CO_2 levels into seasonal, trend and residuals*

∇_d, we have

$$\nabla_d X_i = m_i - m_{i-d} + Y_i - Y_{i-d}$$

which eliminates the seasonal component. The trend $m_i - m_{i-d}$ can then be eliminated by applying some power of the operator ∇_1. If $m_i - m_{i-d}$ is a polynomial, then it will be completely removed; otherwise it will only be partially removed. It is often found in practice that the power required is quite small, frequently one or two.

As an example, consider the monthly CO_2 data shown in figure (3.1). By inspection, the seasonal period is 12 months and the underlying trend is either quadratic or linear. Applying the differencing filters ∇_{12}^1 and ∇_{12}^2 yields the residual plots shown in figures 3.2, 3.3. The residuals from applying ∇_{12}^1 clearly have an upward linear trend present and AIC suggests that a fitted AR model should have order 26, which is very high. Conversely, the residuals from ∇_{12}^2 do not seem to have any trend, and are well modeled by an AR(4) model.

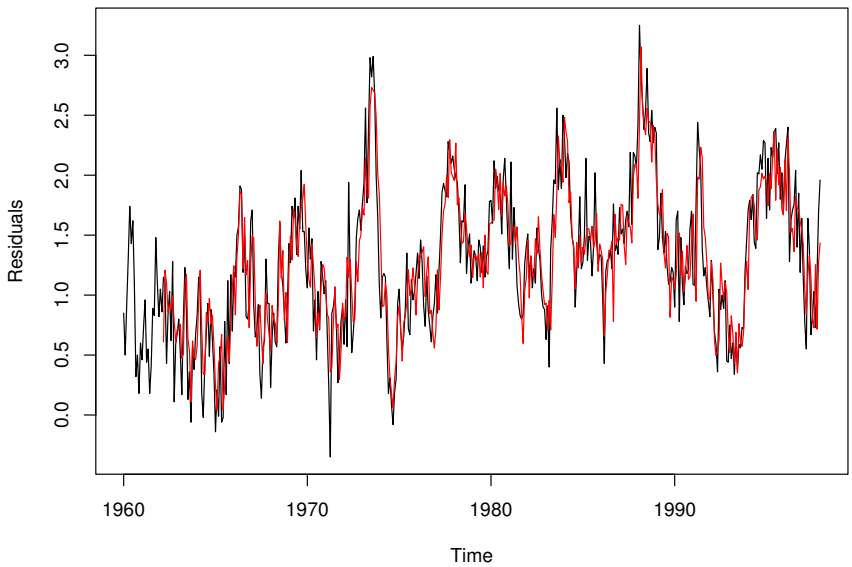

Figure 3.2: *Box-Jenkins decomposition of monthly CO_2 levels by ∇_{12}^1 filter. The black line shows the residuals; the red line shows an AR(26) fit.*

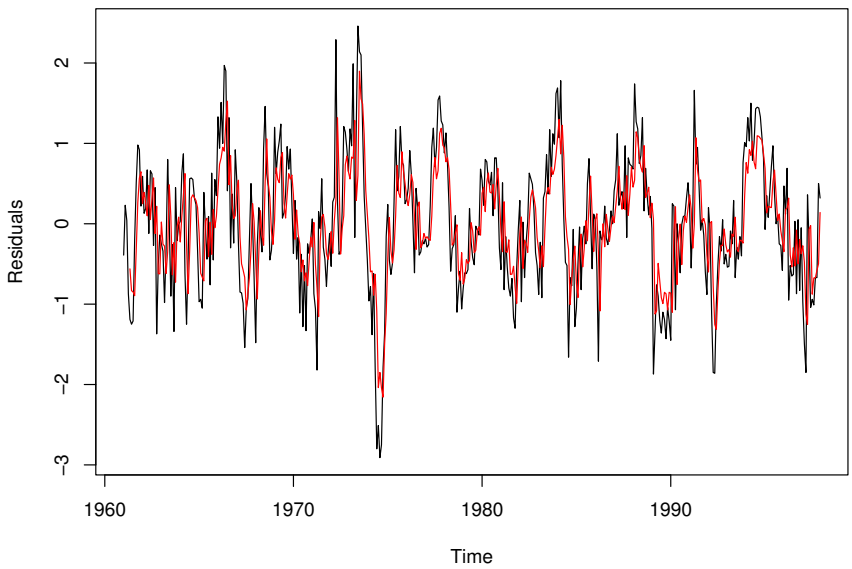

Figure 3.3: *Box-Jenkins decomposition of monthly CO_2 levels by ∇_{12}^2 filter. The black line shows the residuals; the red line shows an AR(4) fit.*

3.1.3 Resampling methods

As seen with the case of regression, the bootstrap's resampling mechanism must be modified to allow for dependency between observations. There are two approaches to resampling in the time domain: *model-based* resampling (Freedman, 1984) and *block* resampling (Hall, 1995; Carlstein 1986).

Model-based resampling in time-series is strongly analogous to model-based resampling in regression: a suitable model is fitted to the data and the estimated residuals (called "innovations" in the time series domain) are resampled to give time series resamples. It is a parametric method that is efficient when a suitable model can be selected with some degree of confidence. The idea of model-based resampling is attributed to Efron (1979), where he applied it in the context of regression. Freedman (1984), however, was the first to apply the method strictly to time series, specifically finite order autoregressive models.

Block bootstrapping samples contiguous "blocks" of ordered observations and constructs resamples by joining these blocks together. The blocks may be non-overlapping (Hall 1985, Carlstein 1986) or overlapping (Hall 1985, Künsch 1989). The underlying idea is if the blocks are long enough (i.e. there is sufficient lag between most of the observations *between* blocks) then much of the original dependency structure will remain in the resamples. The approach is non-parametric and the idea is similar to point-based resampling in regression.

3.2 Model-based resampling

Model-based resampling uses estimated residuals to form resamples. Suppose that we have a time series $\{y_0, y_1, ...y_n\}$ that is serially correlated according to an ARMA(p,0) (an *autoregressive order* p) model

$$\begin{aligned} y_i &= \alpha_1 y_{i-1} + \alpha_2 y_{i-2} + ... + \alpha_p y_{i-p} + \varepsilon_i \\ &= \sum_{j=1}^{p} \alpha_j y_{i-j} + \varepsilon_i \end{aligned}$$

where $i = p, ..., n$ and $Var(\varepsilon_i) = \sigma^2$. Fitting an ARMA$(p, 0)$ model to the data gives estimated coefficients $\{\widehat{\alpha}_1, ..., \widehat{\alpha}_p\}$. The residuals can then be estimated by

$$\widehat{\varepsilon}_i = y_i - \sum_{j=1}^{p} \widehat{\alpha}_j y_{i-j}$$

for $i = p, ..., n$. Model-based resampling then proceeds by equi-probable sampling with replacement from the centred residuals $\widehat{\varepsilon}_p - \overline{\widehat{\varepsilon}}, ..., \widehat{\varepsilon}_n - \overline{\widehat{\varepsilon}}$ to obtain $\varepsilon_0^*, ..., \varepsilon_n^*$ and then setting $y_0^* = \varepsilon_0^*$ and

$$y_i^* = \sum_{j=1}^{p} \alpha_j y_{i-j}^* + \varepsilon_i^*, \quad i = 1, ..., n$$

While the sequence $(y_0, ..., y_n)$ is strictly stationary, the sequence $(y_0^*, ..., y_n^*)$ typically is not. In practice, a longer resample sequence is typically generated and initial observations are discarded

to ensure stationarity (this is called "burning-in" the process).

As was the case in regression, model-based resampling is a attractive method where we strong confidence in the structure and order of the fitted model. If the chosen structure is incorrect, the resamples will have different statistical properties than the original data, resulting misleading inference.

3.3 Block resampling

A non-parametric alternative to resampling using residuals is to resample using blocks. Unlike the previous section where estimated residuals are used to construct resampled series, block resampling treats contiguous blocks of data as interchangeable units to construct resampled series.

The first and simplest approach was introduced by Hall (1985) and Carlstein (1986) whereby data are divided into b non-overlapping blocks of length l such that $n = bl$. Hall (1995) calls this blocking scheme "Carlstein's rule". For a stationary time series $\{x_i\} = (x_1, ..., x_n)$, we set $z_1 = (x_1, ..., x_l)$, $z_2 = (x_{l+1}, ..., x_{2l})$ and so on, producing blocks $z_1, ..., z_b$. New series are simulated by randomly sampling with replacement from z_i with equal probability of b^{-1} and binding the selected blocks end-to-end to give $\{x_i^*\}$. A simple amendment that avoids the strict $n = bl$ requirement in Carlstein's rule is to wrap data around in a circle, however this introduces a dependency structure "discontinuity" in blocks that span the ends of the time series.

The idea behind block resampling is that, given the blocks are long enough, sufficient dependency structure present in the original data will be preserved in the collection of resample blocks, and that these blocks will be approximately independent. This allows equi-probable resampling of blocks to form new time-series resamples.

Clearly whether blocks display these two desirable properties is dependent on block length. The longer the blocks, the less the original dependency structure is "whitened" by being systematically divided into discrete blocks of structure. This effect can easily be seen by comparing a plot of the original time series to plots of resamples with various block lengths. Figure 3.4 shows an example using a shortened version Tong's Sunspots data (Tong, 1990). Also, for stationary processes with $Cov(x_i, x_{i+h}) \to 0$ as $h \to \infty$, the larger the lag h is between observations, the more independent a cluster of observations separated by h will become. However, longer block sizes mean that resampled data will tend toward looking similar to the original data and that fewer unique bootstrap resamples are possible. This suggests that a trade-off exists. This trade off is formally quantified using asymptotic theory in section 3.3.1.

There are several variants on Carlstein's resampling plan. Hall (1985) and Künsch (1989) have suggested overlapping blocks with lag 1 between each other. Hall (1995) calls this "Künsch's rule". Here we set $z_1 = (x_1, ..., x_l)$, $z_2 = (x_{2,}..., z_{l+1})$ and so on, giving blocks $z_1, ..., z_{n-l+1}$. This incurs end-effects since the first and last $l-1$ observations are less likely to appear in resamples, however wrapping the data in a circle fixes this. Clearly, for a given n, the number of possible overlapping blocks b' is much larger than non-overlapping blocks b, and this suggests that the optimum block length under Künsch's block rule is longer than that of Carlstein's (Hall, 1995). This resampling technique is implemented in the function `tsbooty` and its code is in the appendix.

Several more complicated variants have been suggested. *Post-blackening* uses a two step process to increase the robustness of block resampling. Data is "pre-whiten" by fitting a simple model,

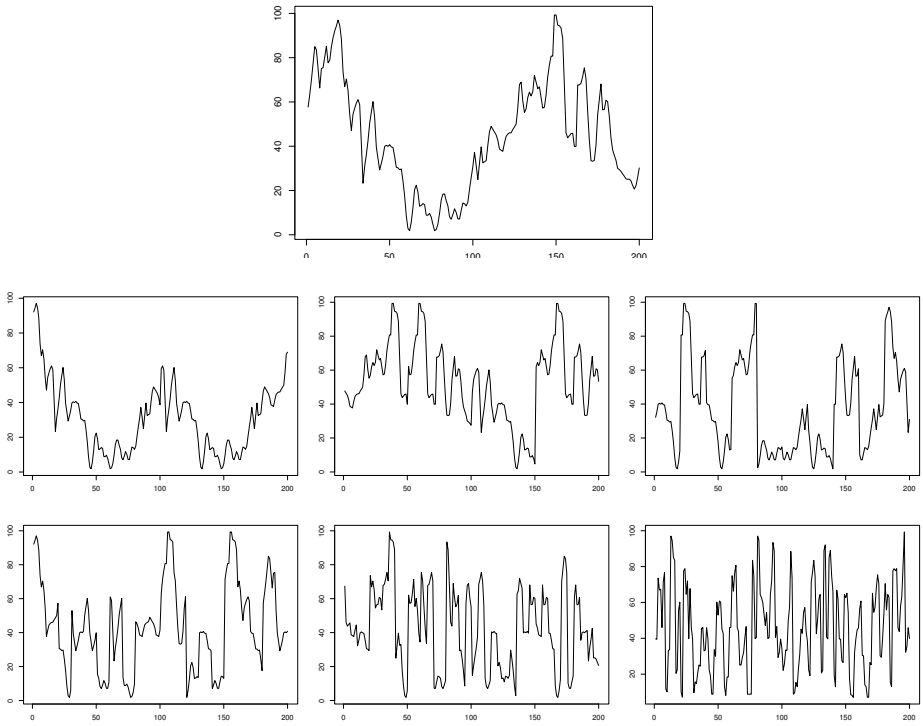

Figure 3.4: *Sunspot data (top) and Künsch's rule block resamples with block length* $b = \{100, 50, 20, 10, 5, 2\}$ *displayed from top-left to bottom-right.*

usually AR(1), to remove gross dependency structure. A series of residual resamples is then generated through block resampling and "post-blackened" by applying the estimated model to the residuals. Post-blackening attempts to reduce the whitening effect incurred with both Carlstein's or Künsch's block resampling. The *stationary bootstrap* (Politis and Romano, 1994) concerns another problem: if the original data is stationary, the resampled data is not. This is because the joint distribution of resampled observations between blocks differs from that in the centre of the block. The stationary bootstrap overcomes this by randomly selecting a block length l according the geometric distribution density $Pr(l = i) = (1 - p)^{i-1}p$ where $i = \{1, 2, ...\}$. The expected block length under this scheme is $l = p^{-1}$.

3.3.1 Selecting an optimal block length

Selecting an appropriate block length involves a trade-off between two effects. Suppose we are interested in some statistic t with distribution T. For small b, the "whitening" of resamples shown in 3.4 distorts inference about T by reducing dependence structure in T^*. For large b, we have fewer possible bootstrap resamples to calculate t^* hence reducing the accuracy of T^* in approximating T. Clearly there exists an optimal balance of the two effects.

The importance of selecting an appropriate block length can be demonstrated though example. Suppose we wish to estimate the variance of the sample mean of an $ARMA(1, 1)$ process with unknown coefficients α, β and length $n = 1000$, and use non-overlapping blocks with length l. Figure (3.5) plots the estimated variance as a function of l, $1 < l < 667$. Its clear that both

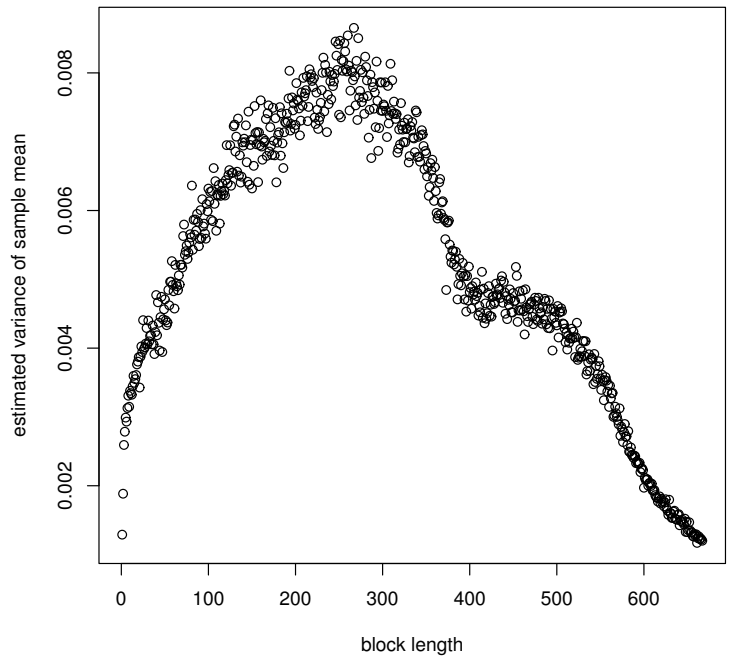

Figure 3.5: *Estimate of variance of sample mean of an ARMA(1,1) process with* $n = 1000$ *for block lengths 1 to 667. Two thousand bootstrap simulations were used for each variance estimate.*

very large block lengths (say $l > 500$) and small ($l < 100$) block lengths give significantly lower estimates than for l between 150 and 350. The difference is approximately a factor of 3! Clearly the estimate is highly sensitive to choice of l. This is an important problem of both overlapping and non-overlapping block resampling methods.

There is a significant amount of literature dedicated entirely to block length selection. Carlstein (1986) and Künsch (1989) both give asymptotic formulae for the optimal block length under Carlstein's rule and Künsch's rule respectively. Both papers conclude that optimal l is proportional to $n^{1/3}$ in estimating bias or variances, however they are uncertain of the proportional coefficient. Hall, Horowitz and Jing (1995) prove that optimal l is higher using Carlstein's rule than with Künsch's rule, give asymptotic formulae for MSE of bootstrap estimates as a function of l, and propose an iterated algorithm for selecting l that minimises the empirical MSE of estimate. More recently, Politis and White have proposed an automatic selection method based on the notion of spectral estimation using flat-top lag-windows.

An important result presented by Hall, Horowitz and Jing (1995) is that the mean square error of an estimate $\widehat{\kappa}(n, l)$ where $\kappa = h(\mu)$ is proportional to

$$\frac{1}{n^d}\left(\frac{C_1}{l^2} + \frac{C_2 l^c}{n}\right) \tag{3.2}$$

where C_1 and C_2 depend on κ and the covariance structure, and $d = 2,\ c = 1$ if κ is a bias or variance, $d = 1,\ c = 2$ if κ is a one-sided significance probability, and $d = 2,\ c = 3$ if κ is a two-sided significance probability. However, the paper only sketches the proof of this result. The

remainder of this section provides a full proof for the class of $\widehat{\kappa}$ that estimate bias or variance. The proof has two sections: the first derives asymptotic relationships for bias and variance of the estimate; the second uses these to derive (3.2) and find the mean square error minimising block length.

Suppose $\kappa = h(\mu)$ where h is twice differentiable. Let $\gamma_h = Cov(x_i, x_{i+h})$, the autocovariance function. The bootstrap estimate of κ is $\widehat{\kappa} = h(\overline{X})$ whose bias and variance are

$$\beta = E\{h(\overline{Y}) - h(\mu)\} \tag{3.3}$$

$$v = Var\{h(\overline{X})\} \tag{3.4}$$

Let us find a delta method approximation to (3.3). Applying a Taylor expansion on $h(\overline{Y})$ centred at μ gives

$$\begin{aligned} h(\overline{Y}) &= \sum_{k=0}^{\infty} \frac{h^{(k)}(\mu)(\overline{X}-\mu)^k}{k!} \\ &= h(\mu) + h'(\mu)(\overline{X}-\mu) + \frac{1}{2}h''(\mu)(\overline{X}-\mu)^2 + ... \end{aligned}$$

Hence the bias is

$$\begin{aligned} \beta &= E\{h(\mu) + h'(\mu)(\overline{X}-\mu) + \frac{1}{2}h''(\mu)(\overline{X}-\mu)^2 + ...\} - h(\mu) \\ &= E\left\{\frac{1}{2}h''(\mu)(\overline{X}-\mu)^2\right\} + O(n^{-2}) \text{ given } \sup|h^{(4)}| < \infty \\ &= \frac{1}{2}h''(\mu)Var(\overline{X}) + O(n^{-2}) \end{aligned}$$

Therefore $\beta \sim \frac{1}{2}h''(\mu)Var(\overline{X})$. Similarly, applying a Taylor expansion on (3.4) gives

$$\begin{aligned} v &= Var\{h(\mu) + h'(\mu)(\overline{X}-\mu) + ...\} \\ &= h'(\mu)Var(\overline{X}) + O(n^{-2}) \end{aligned}$$

hence $v \sim h'(\mu)Var(\overline{X})$ for large n. The bootstrap estimates of the bias and variance of T are given by

$$\begin{aligned} \widehat{\beta} &= E^*\{h(\overline{X}^*) - h(\overline{X})\} \\ &= E^*\{h(\overline{X}) + h'(\overline{X})(\overline{X}^* - \overline{X}) + \frac{1}{2}h''(\overline{X})(\overline{X}^* - \overline{X})^2 + ... - h(\overline{X})\} \\ &\doteq h'(\overline{X})E^*(\overline{X}^* - \overline{X}) + \frac{1}{2}h''(\overline{X})E^*(\overline{X}^* - \overline{X})^2 \end{aligned} \tag{3.5}$$

$$\begin{aligned} \widehat{v} &= Var^*\{h(\overline{X}^*)\} \\ &= Var^*\{h(\overline{X}) + h'(\overline{X})(\overline{X}^* - \overline{X}) + ...\} \\ &\doteq h'(\overline{X})^2 Var^*(\overline{X}^*) \end{aligned} \tag{3.6}$$

Suppose that we estimate β and v by non-overlapping blocks of length l such that $n = bl$ and denote S_i as the sample average of the ith block, $S_i = l^{-1}\sum_{i=1}^{b} X_{(i-1)l+1}$, for $i = 1, ..., b$. Then

$\overline{X} = \overline{S}$ and $\overline{X}^* = b^{-1}\sum_{i=1}^{b} S_i^*$ where S_i^* are resamples from the set $\{S_1, ..., S_b\}$. Also,

$$E^*(\overline{X}^*) = b^{-1}\sum_{i=1}^{b} E(S_i^*) = \overline{S} = \overline{X}$$

$$Var^*(\overline{X}^*) = b^{-1}Var^*(S_i^*) = b^{-2}\sum_{i=1}^{b}(S_i - \overline{S})^2 \tag{3.7}$$

Using (3.7), we can now calculate the MSE of bootstrap estimates $\widehat{\beta}$ and $\widehat{v}$.

Bias of bias estimator From (3.5) we have

$$\begin{aligned}
E(\widehat{\beta}) - \beta &\sim E\{h'(\overline{X})E^*(\overline{X}^* - \overline{X})\} + E\{\frac{1}{2}h''(\overline{X})E^*\{(\overline{X}^* - \overline{X})^2\}\} - \frac{1}{2}h''(\mu)Var(\overline{X}) \\
&= E\{\frac{1}{2}h''(\overline{X})Var^*(\overline{X}^*)\} - \frac{1}{2}h''(\mu)Var(\overline{X}) \quad \text{since } E^*(\overline{X}^*) = \overline{X} \\
&= E\{\frac{1}{2}[h''(\mu) + (\overline{X} - \mu)h'''(\mu) + ...]Var^*(\overline{X}^*)\} - \frac{1}{2}h''(\mu)Var(\overline{X}) \\
&\doteq \frac{1}{2}h''(\mu)E\{Var^*(\overline{X}^*)\} - \frac{1}{2}h''(\mu)Var(\overline{X}) \\
&= \frac{1}{2}h''(\mu)[E\{Var^*(\overline{X}^*)\} - Var(\overline{X})] \qquad (3.8)
\end{aligned}$$

Note that

$$\begin{aligned}
Var^*(\overline{X}^*) &= b^{-2}\sum_{i=1}^{b}(S_i - \overline{S})^2 \\
&= b^{-2}\left\{\sum_{i=1}^{b}(S_i - \mu)^2 - b(\overline{S} - \mu)^2\right\}
\end{aligned}$$

hence

$$\begin{aligned}
E\{Var^*(\overline{X}^*)\} &= \frac{1}{b}E\{(S_1 - \mu)^2\} - \frac{1}{b}E\{(\overline{S} - \mu)^2\} \\
&= \frac{1}{b}\{Var(S_1) - Var(\overline{X})\}
\end{aligned}$$

A well known result[1] is that $Var(\overline{X})$ can be expressed as a sum of the covariances γ_i,

$$Var(\overline{X}) = n^{-1}\left\{\gamma_0 + 2\sum_{i=1}^{n-1}(1 - i/n)\gamma_i\right\}$$

As $n \to \infty$, $n^{-1}\left\{\gamma_0 + 2\sum_{i=1}^{n-1}(1 - i/n)\gamma_i\right\} \to n^{-1}\sum_{i=-\infty}^{\infty}\gamma_i = n^{-1}\zeta$, say. Also,

$$\begin{aligned}
Var(S_1) &= Var(X_1 + ... + X_l) \\
&= l^{-1}\left\{\gamma_0 + 2\sum_{i=1}^{l-1}(1 - i/l)\gamma_i\right\}
\end{aligned}$$

[1] See section 3.3.3 for a proof.

So

$$E\{Var^*(\overline{X}^*)\} = \frac{1}{b}\left\{l^{-1}\gamma_0 + 2l^{-1}\sum_{i=1}^{l-1}(1-i/l)\gamma_i - n^{-1}\gamma_0 - 2n^{-1}\sum_{i=1}^{n-1}(1-i/n)\gamma_i\right\} \quad (3.9)$$

Plugging this into (3.8) gives

$$\begin{aligned} E(\widehat{\beta}) - \beta &\sim \frac{1}{2}h''(\mu)\left\{\frac{1}{bl}\gamma_0 + \frac{2}{bl}\sum_{i=1}^{l-1}(1-i/l)\gamma_i - \frac{1}{nb}\gamma_0 - \frac{2}{nb}\sum_{i=1}^{n-1}(1-i/n)\gamma_i - E(\overline{X}-\mu)^2\right\} \\ &= \frac{1}{2}h''(\mu)\left\{\frac{1}{n}\gamma_0 + \frac{2}{n}\sum_{i=1}^{l-1}(1-i/l)\gamma_i - \frac{1}{nb}\gamma_0 - \frac{2}{nb}\sum_{i=1}^{n-1}(1-i/n)\gamma_i - \frac{1}{n}\gamma_0 \right. \\ &\quad \left. -\frac{2}{n}\sum_{i=1}^{n-1}(1-i/n)\gamma_i\right\} \end{aligned}$$

since $\frac{2}{n}\sum_{i=1}^{l-1}(1-i/l)\gamma_i$ can be expanded as

$$\frac{2}{n}\sum_{i=1}^{l-1}(1-i/l)\gamma_i = \frac{2}{nl}\sum_{i=1}^{l-1}(l-i)\gamma_i = \frac{-2}{nl}\sum_{i=1}^{l-1}i\gamma_i + O(n^{-1})$$

and similarly for the two others, we have

$$E(\widehat{\beta}) - \beta \sim \frac{1}{2}h''(\mu)\left\{\frac{-2}{n}\sum_{i=l}^{n-1}i\gamma_i - \frac{2}{nl}\sum_{i=1}^{l-1}i\gamma_i + \frac{2}{n^2}\sum_{i=l}^{n-1}i\gamma_i\right\} + O\{(nb)^{-1}\}$$

Let $\lim_{l\to\infty}\sum_{i=1}^{l-1}i\gamma_i = \sum_{i=1}^{\infty}i\gamma_i \sim \frac{1}{2}\tau$, say. This gives

$$E(\widehat{\beta}) - \beta \sim -\frac{1}{2}h''(\mu)\frac{\tau}{ln} + O\left\{n^{-1}\sum_{i=l}^{\infty}i|\gamma_i| + 1/(nb)\right\}$$

Given $\sum_{i=l}^{\infty}i^2|\gamma_i| < \infty$, $\sum_{i=l}^{\infty}i|\gamma_i| = o(l^{-1})$ and $n^{-1}\sum_{i=l}^{\infty}i|\gamma_i| = o\{(nl)^{-1}\}$, so

$$E(\widehat{\beta}) - \beta = -\frac{1}{2}h''(\mu)\frac{\tau}{ln} + o\{(nl)^{-1}\} + O\{(nb)^{-1}\}$$

hence

$$E(\widehat{\beta}) - \beta \sim \frac{1}{2}h''(\mu)\frac{\tau}{ln}$$

Variance of bias estimator From (3.5), we have

$$\begin{aligned} Var(\widehat{\beta}) &= Var\left\{\frac{1}{2}h''(\overline{X})E^*(\overline{X}^* - \overline{X})^2\right\} \\ &\doteq \left\{\frac{1}{2}h''(\mu)\right\}^2 Var\left\{Var^*(\overline{X}^*)\right\} \\ &= \left\{\frac{1}{2}h''(\mu)\right\}^2 Var\left\{b^{-2}\sum_{i=1}^{b}(S_i - \overline{S})^2\right\} \end{aligned}$$

Calculation of $Var\left\{\sum_{i=1}^{b}(S_i-\overline{S})^2\right\}$ is messy, however can be cleaned up under a few simplifying assumptions. First, assume that $\{X_i\}$ is a $m-$dependent process, that is $\gamma_{m+1}=\gamma_{m+2}=...=0$ and central moments $\mu_3=\mu_4=...=0$. Also, suppose that $m<l$. Then we have

$$Var\left\{\sum_{i=1}^{b}(S_i-\overline{S})^2\right\} \doteq bVar\{(S_1-\mu)^2\}+2(b-1)Cov\{(S_1-\mu)^2,(S_2-\mu)^2\}$$

For normal data, we have

$$\begin{aligned} Var\{(S_1-\mu)^2\} &= 2\{Var(S_1-\mu)\}^2 \\ Cov\{(S_1-\mu)^2,(S_2-\mu)^2\} &= 2\{Cov(S_1-\mu,S_2-\mu)\}^2 \end{aligned}$$

so

$$\begin{aligned} Var\left\{\sum_{i=1}^{b}(S_i-\overline{S})^2\right\} &\doteq 2b\{Var(S_1-\mu)\}^2+4(b-1)\{Cov(S_1-\mu,S_2-\mu)\}^2 \\ &= 2b\left\{l^{-1}\gamma_0+2l^{-1}\sum_{i=1}^{l-1}(1-i/l)\gamma_i\right\}^2+4(b-1)Cov\left(\sum_{i=1}^{l}X_i,\sum_{j=l+1}^{2l}X_j\right)^2 \end{aligned}$$

Let $l^{-1}c_0^{(l)}=\gamma_0+2l^{-1}\sum_{i=1}^{l-1}(1-i/l)\gamma_i$ and $c_1^{(l)}=\sum_{i=1}^{l}i\gamma_i$. Then

$$\begin{aligned} Var\left\{\sum_{i=1}^{b}(S_i-\overline{S})^2\right\} &= 2b(l^{-2}c_0^{(l)})^2+4(b-1)\left\{\sum_{i=1}^{l}\sum_{j=l+1}^{2l}Cov(X_i,X_j)\right\}^2 \\ &\doteq 2b(l^{-2}c_0^{(l)})^2+4b(l^{-2}c_1^{(l)})^2 \end{aligned} \tag{3.10}$$

Substituting (3.10) into the expression for $Var(\widehat{\beta})$ gives

$$Var(\widehat{\beta}) \sim \left\{\frac{1}{2}h''(\overline{X})\right\}^2 2b^{-3}l^{-4}\left\{(c_0^{(l)})^2+2(c_1^{(l)})^2\right\}$$

Under the simplifying condition that $l\to\infty$ as $n\to\infty$ such that $l/n\to 0$, we have $l^{-1}c_0^{(l)}\to\sum_{i=1}^{\infty}i\gamma_i=\xi$. Thus,

$$(c_0^{(l)})^2+2(c_1^{(l)})^2\sim l^2\xi^2$$

which gives us

$$\begin{aligned} Var(\widehat{\beta}) &\sim \left\{\frac{1}{2}h''(\mu)\right\}^2\frac{2l^2\xi^2}{b^3l^4} \\ &= \left\{\frac{1}{2}h''(\mu)\right\}^2 2ln^{-3}\xi^2 \end{aligned}$$

Bias of variance estimator From (3.6), we have

$$\begin{aligned} E(\widehat{v})-v &= E\{h'(\overline{X})^2Var^*(\overline{X}^*)\}-h'(\mu)Var(\overline{X}) \\ &\doteq h'(\mu)^2[E\{Var^*(\overline{X}^*)-Var(\overline{X})\} \end{aligned}$$

and (3.9) from

$$\begin{aligned} E\{Var^*(\overline{X}^*) - Var(\overline{X})\} &\sim \frac{-2}{n}\sum_{i=l}^{n-1} i\gamma_i - \frac{2}{ln}\sum_{i=l}^{n-1} i\gamma_i + \frac{2}{n^2}\sum_{i=l}^{n-1} i\gamma_i + O\{(nb)^{-1}\} \\ &= \frac{-\tau}{ln} + o\{(nl)^{-1}\} + O\{(nb)^{-1}\} \end{aligned}$$

hence it follows that

$$E(\widehat{v}) - v \sim -h'(\mu)^2 n^{-1} l^{-1}\tau$$

Variance of variance estimator From (3.6),

$$\begin{aligned} Var(v) &= Var\{h'(\overline{X})^2 Var^*(\overline{X}^*)\} \\ &\doteq h'(\mu)^4 Var\{Var^*(\overline{X}^*)\} \end{aligned}$$

From consideration of the variance of bias estimate, we know

$$Var\{Var^*(\overline{X}^*)\} \sim 2ln^{-3}\xi^2$$

so

$$Var(v) \sim h'(\mu)^4 2ln^{-3}\xi^2$$

Summary The following asymptotic relations have been shown

$$E[\widehat{\beta}] - \beta \sim -\frac{1}{2}h''(\mu)n^{-1}l^{-1}\tau, \quad Var(\widehat{\beta}) \sim \left\{\frac{1}{2}h''(\mu)\right\}^2 2ln^{-3}\xi^2 \tag{3.11}$$

$$E[\widehat{v}] - v \sim -h'(\mu)n^{-1}l^{-1}\tau, \quad Var(\widehat{v}) \sim h'(\mu)^4 2ln^{-3}\xi^2 \tag{3.12}$$

This allows us to prove the following proposition.

Proposition If $\widehat{\kappa}(n,l)$ estimates bias or variance, then

$$MSE\{\widehat{\kappa}(n,l)\} \sim \frac{1}{n^2}\left(C_1\frac{1}{l^2} + \frac{C_2 l}{n}\right) \ \textit{and} \ \widehat{l} = \arg\min_l \{MSE(\widehat{\kappa}(n,l))\} = C_3 n^{1/3}$$

Proof Using $MSE(\widehat{\theta}) = \left(E\{\widehat{\theta}\} - \theta\right)^2 + Var\left(\widehat{\theta}\right)$ and (3.11), (3.12) it follows that

$$\begin{aligned} MSE(\widehat{\kappa}(n,l)) &\sim \begin{cases} \frac{1}{4}\{h''(\mu)\}^2 n^{-2}l^{-2}\tau^2 + \frac{1}{2}\{h''(\mu)\}^2 n^{-3}l\xi^2 & \text{bias case} \\ \{h'(\mu)\}^2 n^{-2}l^{-2}\tau^2 + 2\{h'(\mu)\}^4 n^{-3}l\xi^2 & \text{variance case} \end{cases} \\ &= \begin{cases} \frac{1}{n^2}\left(\frac{\{2^{-1}h''(\mu)\tau\}^2}{l^2} + \frac{\{2^{-1/2}h''(\mu)\xi\}^2 l}{n}\right) \\ \frac{1}{n^2}\left(\frac{\{h'(\mu)\tau\}^2}{l^2} + \frac{\{2^{1/2}h'(\mu)^2\xi\}^2 l}{n}\right) \end{cases} \\ &= \frac{1}{n^2}\left(\frac{C_1}{l^2} + \frac{C_2 l}{n}\right) \end{aligned}$$

where $C_1 = \left\{2^{-1}h''(\mu)\tau\right\}^2$, $C_2 = \left\{2^{-1/2}h''(\mu)\xi\right\}^2$ for bias estimation and $C_1 = \{h'(\mu)\tau\}^2$ $C_2 = \left\{2^{1/2}h'(\mu)^2\xi\right\}^2$ for variance estimation. Note that C_1 and C_2 depend only on ξ and τ, which measure the strength covariance structure.

Let $\widehat{l} = \arg\min_l \{MSE(\widehat{\kappa}(n,l))\}$ be the optimum block length.

$$\left.\frac{\partial MSE}{\partial l}\right|_{l=\widehat{l}} = \frac{1}{n^2}\left(-2C_1\frac{1}{\widehat{l}^3} + \frac{C_2}{n}\right) = 0$$

$$\therefore \widehat{l} = \left(\frac{2C_1}{C_2}n\right)^{1/3} \tag{3.13}$$

■

By similar arguments, it is can be shown that if $\widehat{\kappa}$ estimates a one-sided distribution

$$MSE(\widehat{\kappa}(n,l)) \sim \frac{1}{n^2}\left(\frac{C_1}{l^2} + \frac{C_2 l^2}{n}\right)$$

$$\widehat{l} = \left(\frac{C_1}{C_2}n\right)^{1/4}$$

and if $\widehat{\kappa}$ estimates a two-sided distribution

$$MSE(\widehat{\kappa}(n,l)) \sim \frac{1}{n^2}\left(\frac{C_1}{l^2} + \frac{C_2 l^3}{n}\right)$$

$$\widehat{l} = \left(\frac{2C_1}{3C_2}n\right)^{1/5}$$

When resampling overlapping blocks, a very similar argument applies if data is wrapped in a circle for resampling[2]. In this case the optimum block length is still given by $C_0 n^{1/k}$ where $k = 3,\ 4$ or 5 depending on what statistic κ is. In fact C_1 is still the same, although C_2 now is given by (Hall, 1995):

$$C_2^k = \frac{2}{3}\left\{2^{-1/2}h''(\mu)\xi\right\}^2 \text{ if } \widehat{\kappa} \text{ estimates bias}$$

$$C_2^k = \frac{2}{3}\left\{2^{1/2}h'(\mu)^2\xi\right\}^2 \text{ if } \widehat{\kappa} \text{ estimates variance}$$

Thus the optimum block length under Künsch's rule, $\widehat{l}_k$, is related to $\widehat{l}$ by

$$\widehat{l}_k = \left(\frac{3}{2}\right)^{1/3}\widehat{l}$$

where κ is a bias or variance. That is, $\widehat{l}_k$ is 14.4% larger than $\widehat{l}$.

[2]if data is not wrapped in a circle, technical issues arise. See Davidson and Hinkley (1997), page 407.

3.3.2 Empirical method for block choice

Section 3.3.1 showed that the optimal block length equals $(2C_1/C_2n)^{1/3}$ for estimating variance of bias of some estimator $\widehat{\kappa} = h(\overline{X})$. However, the coefficients C_1 and C_2 are unlikely to be optimal under non-asymptotic conditions, that is, when n is relatively small. Also, calculation of C_1 and C_2 requires accurate estimation of γ_h, which may be a harder problem than the one at hand. This presents us with a problem - without knowing the constant C_0 in $\widehat{l} = C_0 n^{1/3}$, we still do not have much idea of how to select an appropriate block length.

A solution proposed by Hall, Horowitz and Jing (1995) is to empirically estimate $MSE(\widehat{\kappa})$ as a function of l. An initial guess l_0 is made and an estimate $\widehat{\kappa}(n, l_0)$ is found. The series is then split into contiguous subseries of length $m < n$, and for each $j < l_0$, $\widehat{\kappa}(m, j)$ is calculated from $\{x_i, ..., x_{i+m-1}\}$ for $i = 1, ..., n - m + 1$. Mean square error is estimated by

$$MSE(m, k) = \frac{1}{n - m + 1} \sum_{i=1}^{n-m+1} \{\widehat{\kappa}_i(m, j) - \widehat{\kappa}(n, l_0)^2\} \tag{3.14}$$

A block length $\widehat{k}$ that minimises (3.14) is selected. Then an estimate of $\widehat{l}$ is given by

$$\widehat{l} = \widehat{k} \times (n/m)^{1/(c+2)}$$

where $c = 1$ if bias or variance is estimated and $c = 2$ (3) if a one-sided (or two-sided) significance probability is estimated. This procedure can be iterated by setting $l_0 = \widehat{l}$ and repeating. The method is implemented in function `bootlength` and the code is in the appendix. Experiments using this method on AR(1) and AR(2) data suggest that estimates $\widehat{l}$ are highly sensitive to choice of m. Unfortunately, selecting m is not discussed in Hall, Horowitz and Jing's paper.

Another, far simpler, empirical method is to estimate $\widehat{l}$ in cases where κ can be derived exactly. The optimum $\widehat{l}$ that minimises mean square error is

$$\widehat{l} = \arg\min_l \{|E[\widehat{\kappa}(n, l)] - \kappa|^2 + Var(\widehat{\kappa}(n, l)\}$$

This can be estimated by

$$\widetilde{l} = \arg\min_l \{|\overline{\kappa}(n, l) - \kappa|^2 + s_l^2\}$$

where $\overline{\kappa} = a^{-1} \sum_{i=1}^{a} \widehat{\kappa}_i$ and $s_l^2 = (a-1)^{-1} \sum_{i=1}^{a} (\widehat{\kappa}_i - \overline{\kappa})^2$.

Of course we would not need to apply the bootstrap to estimate κ if it is known exactly. However, having the exact result for comparison allows for accurate computation of $\widehat{l}$ in important simple cases. Two statistics are considered: $\kappa_1 = Var(\overline{X})$, the variance of the sample mean; and $\kappa_2 = E(\widehat{v}) - \sigma^2$ where $\widehat{v} = n^{-1} \sum_{i=1}^{n} (X_i - \overline{X})^2$, the bias of the bootstrap estimate of variance. The data used comes from AR(1) processes with $\alpha = \{0.1, 0.2, ..., 0.9\}$. Recall that

$$C_0 \sim \frac{A\tau^2}{B\xi^2}$$

where A and B are positive constants and where $\tau = 2\sum_{i=1}^{\infty} i\gamma_i = \frac{2}{1-\alpha^2} \sum_{i=1}^{\infty} i\alpha^i$ and $\xi = \sum_{i=-\infty}^{\infty} \gamma_i = \frac{1}{1-\alpha^2} \left(1 + 2\sum_{i=1}^{\infty} \alpha^i\right)$. Therefore $C_0(\alpha)$ should be an increasing and convex function.

Overlapping (Künsch) block-based bootstrapping is employed throughout. The method used

was as follows:

1. Generate AR(1) data, S_α, for each $\alpha = \{0.1, ..., 0.9\}$ with $n = 10000$. Discard the first 1000 points to ensure stationarity of the series.

2. Split S_α into 20 contiguous subseries, $M_{1,\alpha}, ..., M_{20,\alpha}$ of length 400.

3. for $l = \{1, ..., 100\}$

   (a) for $m = \{1, ..., 20\}$

      i. calculate $\widehat{\kappa}_m(400, l)$ using $M_{m,\alpha}$ and $B = 2499$ bootstrap simulations

      ii. calculate $\overline{\kappa}_l = 20^{-1} \sum_{m=1}^{20} \widehat{\kappa}_m =$ and $s_l^2 = 20^{-1} \sum_{m=1}^{20} (\widehat{\kappa}_m - \overline{\kappa})^2$

   (b) Let $MSE(l) = \{|\overline{\kappa}_l - \kappa|^2 + s_l^2\}$

4. Set $\widehat{l}_\alpha = \arg\min_l MSE(l)$

The two statistics are studied separately in sections 3.3.3 and 3.3.4.

3.3.3 Case study: variance of sample mean

We wish to find the optimum block length $\widehat{l}$ for $\kappa = Var(\overline{X})$. To apply the empirical method outlined 3.3.2, we need an exact expression for κ.

Consider the AR(1) series $X_{\alpha,n} = \{X_1, ...X_n\}$ where

$$X_i = \alpha X_{i-1} + \varepsilon_i, \ \varepsilon_i \sim N(0, 1) \tag{3.15}$$

and whose autovariances are $\gamma_h = Cov(X_i, X_{i+h})$ for $h = \{1, 2, ...\}$. The sample mean $\overline{X}$ is an unbiased estimator of μ in the time series domain since

$$E(\overline{X}_n) = n^{-1}(EX_1 + ... + EX_n) = \mu$$

The variance of the sample mean is

$$\begin{aligned} Var(\overline{X}_n) &= n^{-2} Var(X_1 + ... + X_n) \\ &= n^{-2} \sum_{i=1}^{n} \sum_{j=1}^{n} Cov(X_i, X_j) \\ &= n^{-2} \sum_{i-j=-n}^{n} (n - |i - j|)\gamma_{i-j} \\ &= n^{-1} \sum_{h=-n}^{n} \left(1 - \frac{|h|}{n}\right) \gamma_h \\ &= n^{-1} \left\{ \gamma_0 + 2 \sum_{h=1}^{n-1} \left(1 - \frac{h}{n}\right) \gamma_h \right\} \end{aligned} \tag{3.16}$$

Let us now determine an expression for γ_h of the series given by (3.15). The autocovariance at lag zero is

$$\gamma_0 = E[\alpha X_{i-1} + \varepsilon_i]^2 = \alpha^2\gamma_0 + 1$$

and hence

$$\gamma_0 = \frac{1}{1-\alpha^2}$$

Now $\gamma(h) = E[X_i, X_{i+h}] = E[(\alpha X_{i-h-1} + \varepsilon_{i+h})X_i] = \alpha\gamma(h-1)$ assuming $E[X_i\varepsilon_{i+h}] = 0$ for $h > 0$. Iterating this we get

$$\gamma(h) = \alpha^h\gamma(0) = \frac{\alpha^h}{1-\alpha^2}$$

Now that an expression for $\kappa = Var(\overline{X})$ has been derived, we can compare κ to overlapping-block bootstrap estimates for various l. Data was generated according to (3.15) with $\alpha = \{0.1, ..., 0.9\}$ with $n = 400$. Each series was burnt-in to ensure stationarity. The function `tsbooty` was used to generate $\widehat{\kappa}(400, l)$ for $1 \leq l \leq 100$ using $B = 1499$ simulations. Figure (3.6) shows plots of $\widehat{\kappa}$ compared to κ for each α.

The plots show point-estimates of the bias of $\widehat{\kappa}(l)$ and display several interesting things. First, the $\widehat{\kappa}(l)$ estimates lie on curves that are highly non-linear. They generally are reach a global maximum with $l < 100$ and have an approximate concave shape. From repeated simulation, it seems that bumps and wiggles in the plots are artifacts from outliers in the generated data. Secondly, for $\alpha = \{0.1, 0.3, 0.7\}$, the curve of $\widehat{\kappa}(l)$ intersects the exact κ twice. For all three, the first band occurs with $l < 60$ and the second for $l > 120$. Recall that the data length is 400: block lengths greater than 100 are suspiciously high. Thirdly, we see that for $\alpha = \{0.6, 0.8, 0.9\}$, the curves of $\widehat{\kappa}(l)$ do not intersect κ: the estimates are all too low. These three plots nicely illustrate that as covariance structure becomes stronger, block-based resampling tends to underestimate $Var(\overline{X})$ no matter how large l is.

The mean square error of $\widehat{\kappa}(l)$ was calculated according to the method given in section 3.3.2. Figure 3.7 shows the plots. They exhibit clear global minima for reasonably small l and with approximately linear increase in MSE as l increases thereafter. The minimum gradually drifts to the right for larger α, as predicted by theory. The mean square minimising $\widehat{l}$ are given in table 3.1. The estimates for C_0 are increasing with α (with the exception of $\alpha = 0.3$ which is probably an outlier).

3.3.4 Case study: bias of sample variance

Consider the biased sample variance estimator for $\gamma_0 = Var(X_i)$

$$\widehat{v} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

The bias is given by

$$\begin{aligned} E(\widehat{v}) - \gamma_0 &= E(X_1 - \overline{X})^2 - \gamma_0 \\ &= Var(X_1) + Var(\overline{X}) - 2Cov(X_1, \overline{X}) - \gamma_0 \end{aligned}$$

$$
\begin{aligned}
&= Var(\overline{X}) - \frac{2}{n} Cov(X_1, \sum_{j=1}^{n} X_j) \\
&= Var(\overline{X}) - \frac{2}{n} \sum_{j=1}^{n} Cov(X_1, X_j) \\
&= -n^{-1} \left\{ \gamma_0 + 2n^{-1} \sum_{j=1}^{n} j\gamma_j \right\}
\end{aligned}
$$

Figure 3.8 shows point-estimates of $\widehat{\kappa}(l)$ for each α. Again we see that $\widehat{\kappa}(l)$ follow highly non-linear curves and that as α grows, the estimates become worse: the bias is over estimated. Generally for low l, the distance $|\widehat{\kappa}(l) - \kappa|$ is minimised. Plots of the estimated $MSE\{\widehat{\kappa}(l)\}$ are shown in figure 3.9. The plots show that MSE reaches its minimum rapidly between $1 < l < 11$ and gradually rises thereafter. The minimum point drifts to the right as α increases. Table 3.1 gives MSE-minimising $\widehat{l}$.

3.3.5 Results

α	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
$\widehat{l}$ for $\kappa_1 = Var(\overline{X})$	4	9	7	11	14	14	17	18	34
C_0 in $\widehat{l} = C_0 n^{1/3}$	0.54	1.22	0.95	1.49	1.90	1.90	2.31	2.44	4.61
$\widehat{l}$ for $\kappa_2 = \beta(\widehat{v})$	1	2	2	4	6	6	7	10	11
C_0 in $\widehat{l} = C_0 n^{1/3}$	0.14	0.27	0.27	0.54	0.81	0.81	0.95	1.36	1.49

Table 3.1: *Optimum block lengths for estimating $Var(\overline{X})$ and $\beta(\widehat{v})$ for various α.*

From table 3.1, we can see that optimal $C_0(\alpha)$ is an increasing and approximately linear function. C_0 in variance estimation of the sample mean is around 3 times greater than C_0 in bias estimation.

I was curious to see how the asymptotic estimate $C_0 = (2C_1/C_2)^{1/3}$ compared to these empirical results. Figure (3.10) shows $(400 \times 2C_1/C_2)^{1/3}$ plotted against the empirical results for $\kappa = Var(\overline{X})$. The asymptotic estimates are surprisingly close to the empirical results. Unfortunately, for $\kappa = \beta(\widehat{v})$, the $\widehat{l}$ results plotted against their asymptotic estimates are less surprising - the asymptotic estimates are much higher.

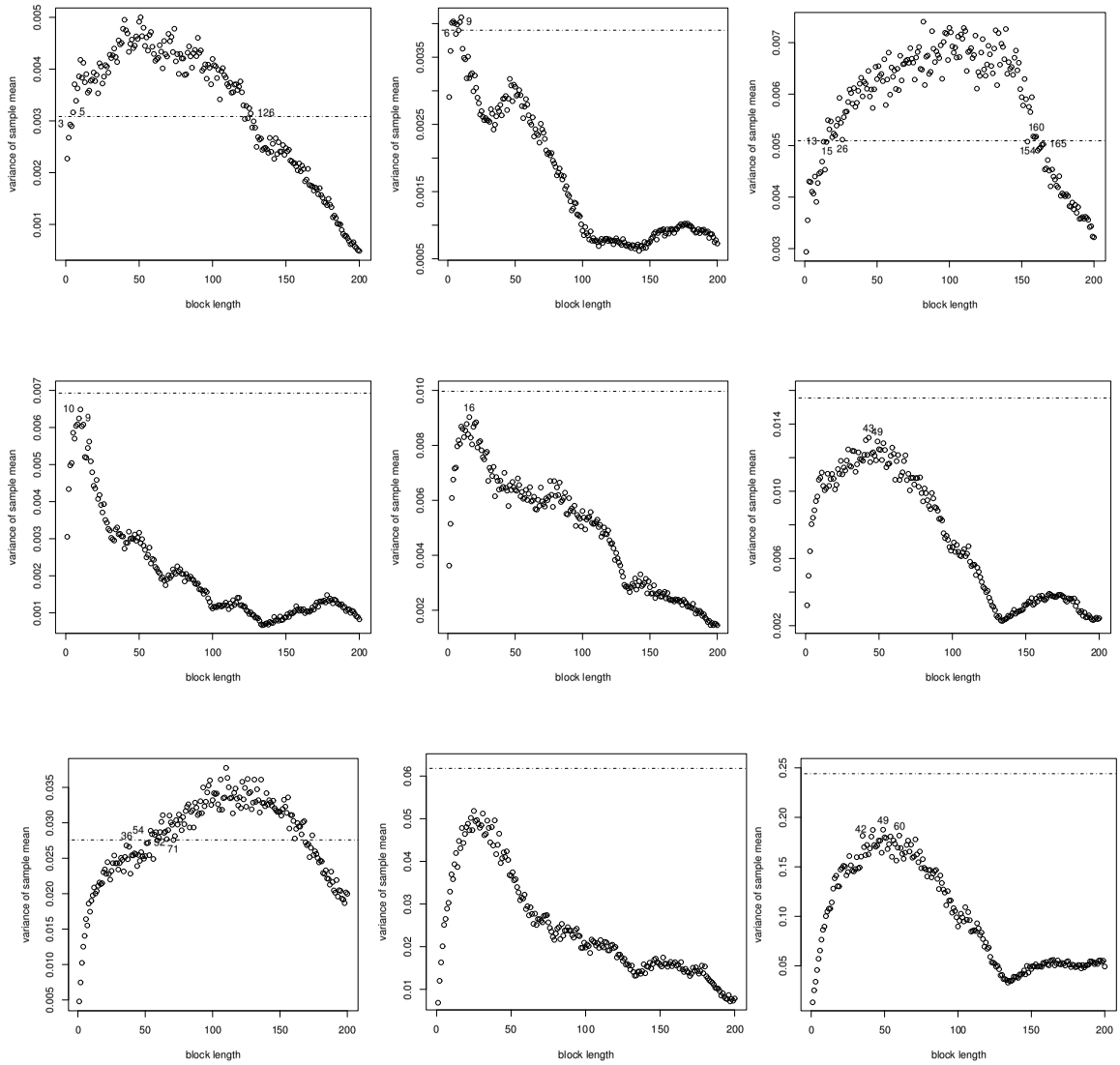

Figure 3.6: *Overlapping block bootstrap estimates of* $Var(\overline{X})$ *using AR(1) data with* $n = 400$ *and coefficients* $\alpha = \{0.1, 0.2, ..., 0.9\}$ *displayed in order from left-right, top-bottom. The horizontal axes are block lengths* $b = \{1, ..., 200\}$. *Each estimate is generated using* $B = 1499$ *bootstrap simulations. The dotted line is the value* $Var(\overline{X})$.

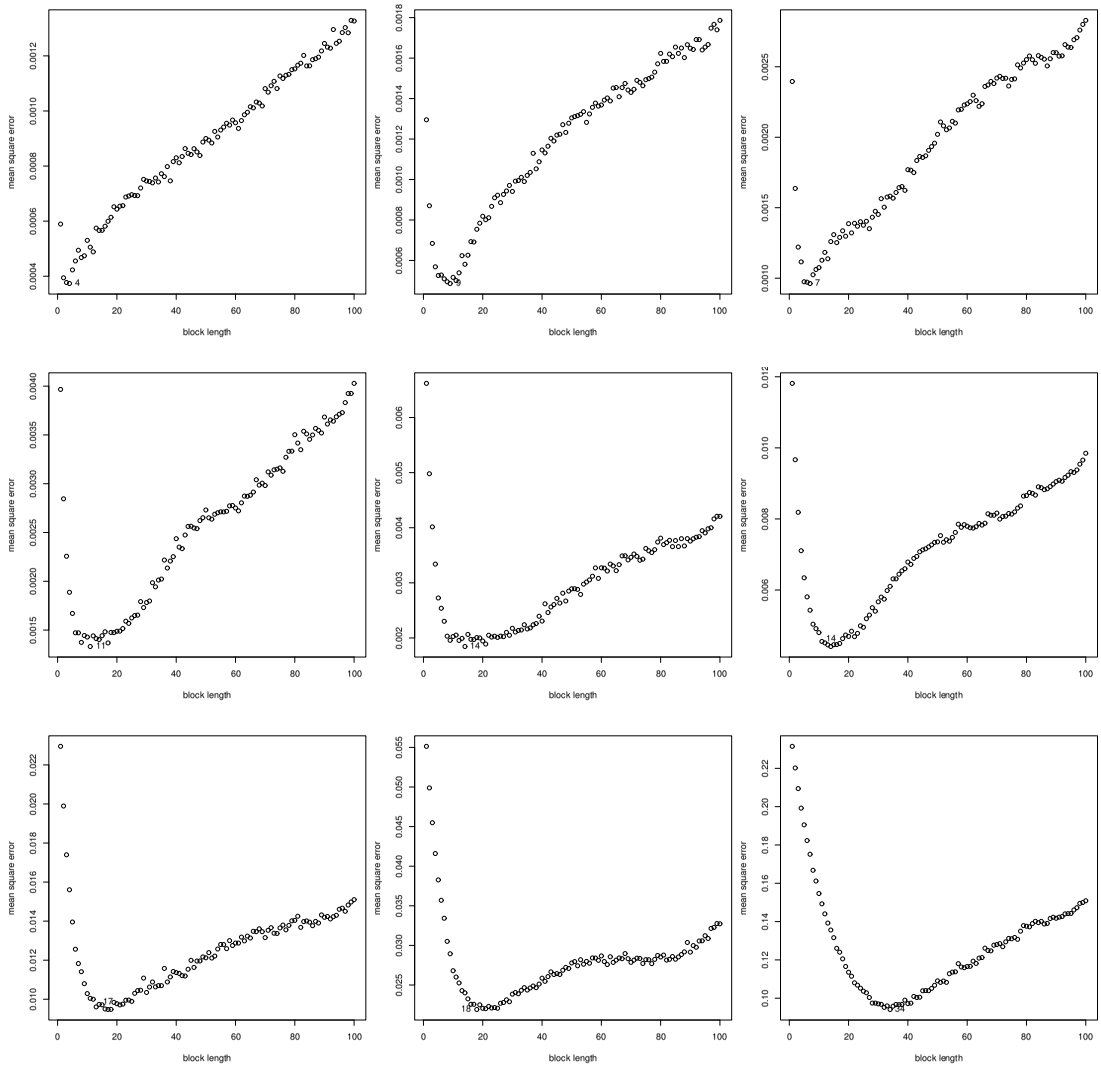

Figure 3.7: *Estimated mean square error of* $\widehat{\kappa}(l)$, $\kappa = Var(\overline{X})$, *using AR(1) data with* $n = 400$ *and coefficients* $\alpha = \{0.1, 0.2, ..., 0.9\}$ *displayed in order from left-right, top-bottom. The horizontal axes are block lengths* $b = \{1, ..., 200\}$. *Each estimate is generated from 20 repeated simulations each using* $B = 2499$.

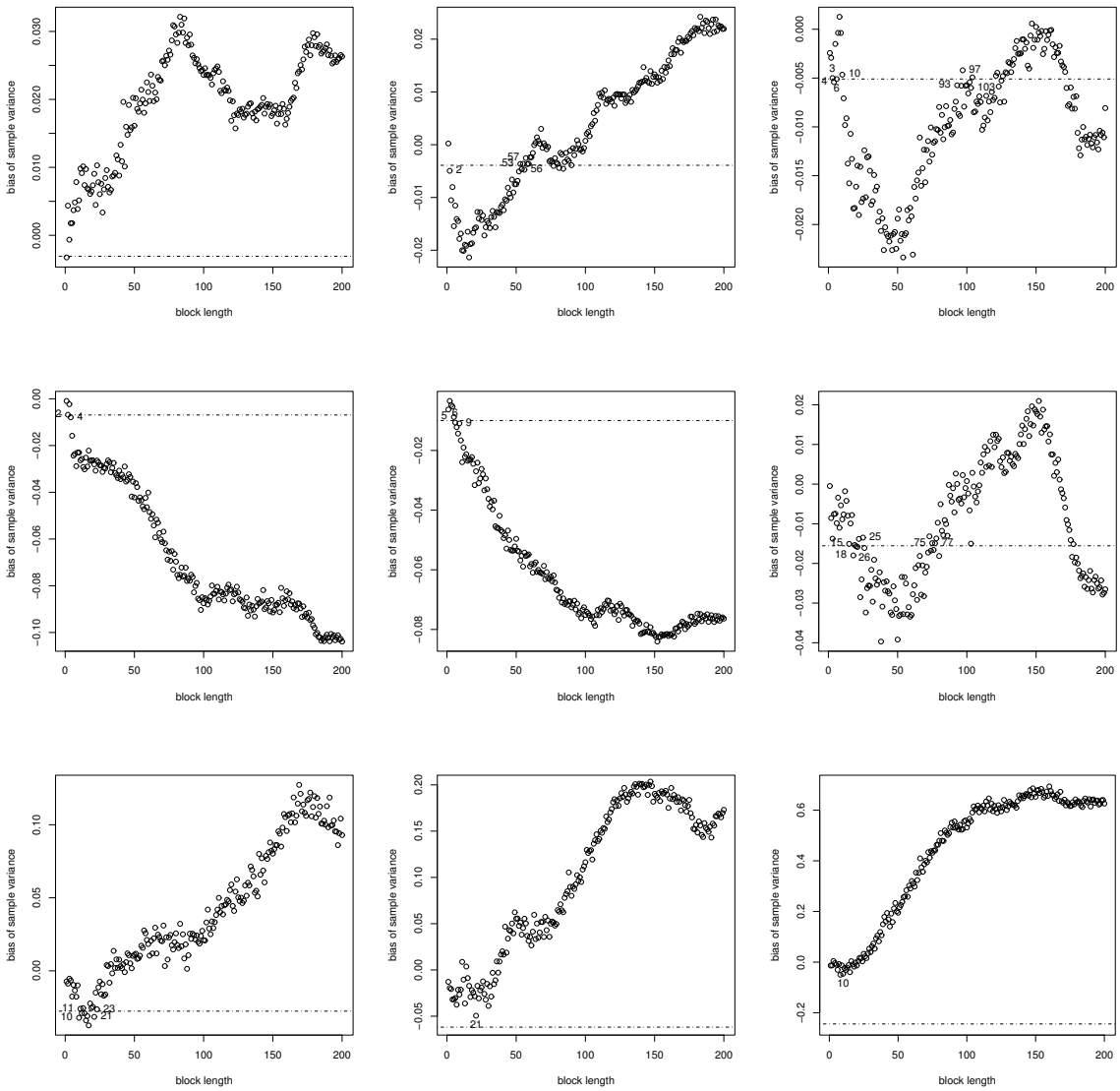

Figure 3.8: *Overlapping block bootstrap estimates of* $\beta(\widehat{v})$ *using AR(1) data with* $n = 400$ *and coefficients* $\alpha = \{0.1, 0.2, ..., 0.9\}$ *displayed in order from left-right, top-bottom. The horizontal axes are block lengths* $b = \{1, ..., 200\}$. *Each estimate is generated using* $B = 1499$ *bootstrap simulations. The dotted line is the true bias.*

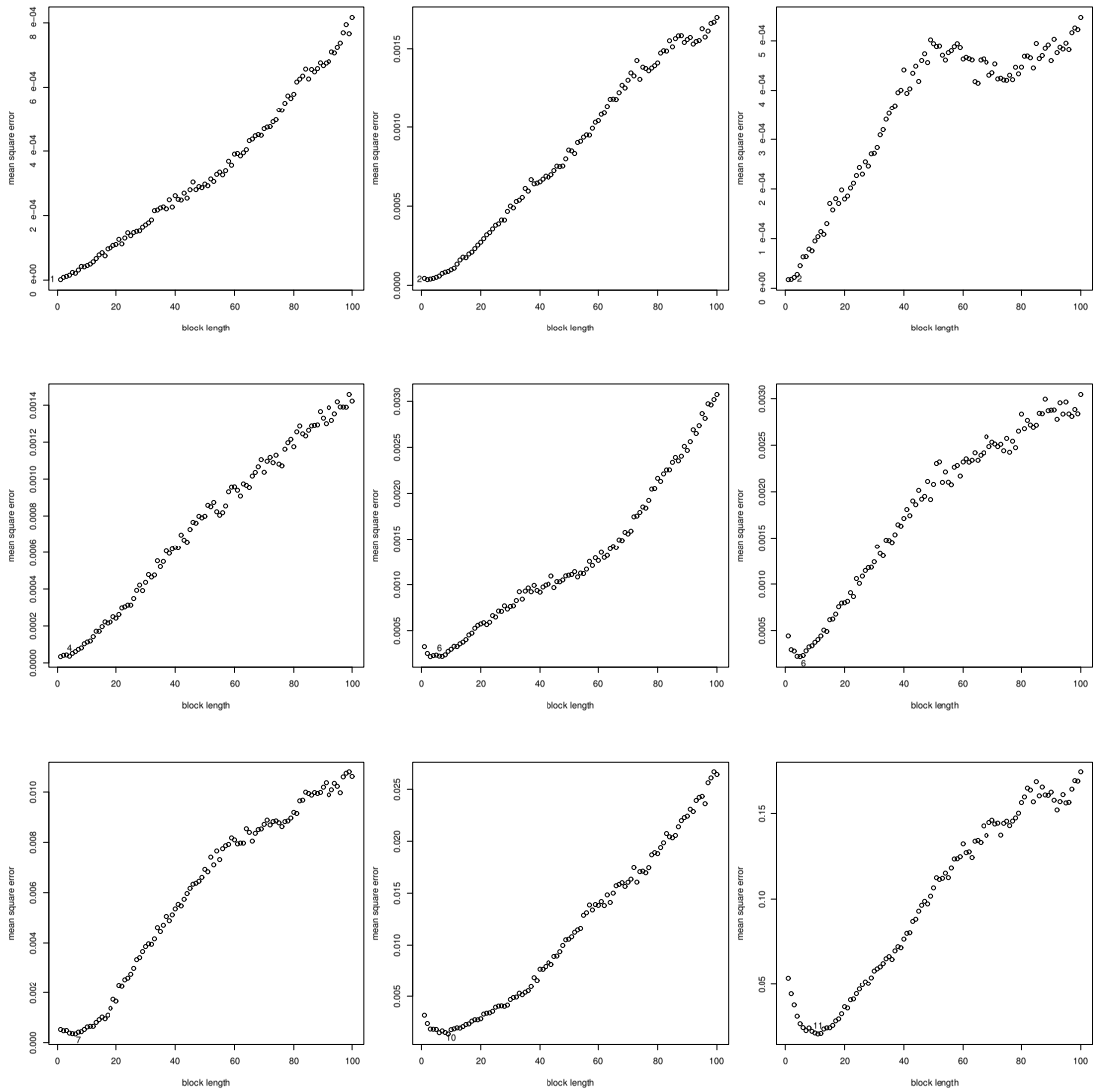

Figure 3.9: *Estimated mean square error of* $\widehat{\kappa}(l)$, $\kappa = \beta(\widehat{v})$, *using AR(1) data with* $n = 400$ *and coefficients* $\alpha = \{0.1, 0.2, ..., 0.9\}$ *displayed in order from left-right, top-bottom. The horizontal axes are block lengths* $b = \{1, ..., 200\}$. *Each estimate is generated from 20 repeated simulations each using* $B = 2499$.

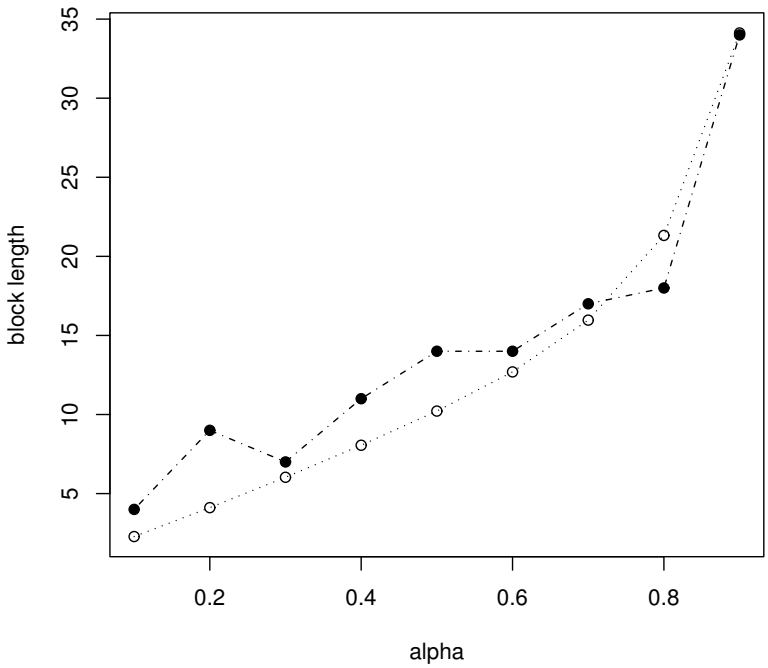

Figure 3.10: *Asymptotic estimates (dotted line) and empirical estimates (broken line) of* $\widehat{l}$ *for* $\alpha = \{0.1, ..., 0.9\}$ *and* $\kappa = Var(\overline{X})$.

Conclusion

This report serves as a broad introduction to the bootstrap principle and its wide-ranging statistical applications. In particular, modification of Efron's bootstrap for time series data is the central topic, and solving the problem of block length selection in block resampling is its centre piece.

The two main results regard choosing the optimal block length, $\widehat{l}$. The first is a proof that the optimal block length is given by $\widehat{l} = C_0 n^{1/d}$ where d is either 3, 4 or 5 and C_0 is a constant depending on the estimator and strength of covariance structure. The second is determining optimal block lengths for two simple estimators, variance of the sample mean and bias of sample variance, using AR(1) data. Specifically, the coefficient C_0 in $\widehat{l} = C_0 n^{1/3}$ was shown to be approximately linear and increasing with the autoregressive coefficient α.

Other results of the report include: deriving plug-in estimators of mean and variance and deriving bootstrap bias and variance estimates of these estimators; determining formulae for the simulation error introduced in Monte Carlo estimation; providing an example of iterated bootstrap bias correction using $\widehat{v} = n^{-1} \sum (X_i - \overline{X})^2$; proving model-based bootstrap estimates of $Var(\beta_1)$ and $Var(\beta_0)$ converge to exact variance of β_1 and β_0 in linear regression; showing that point-based bootstrap estimates of $Var(\beta_1)$ and $Var(\beta_0)$ converge to robust variance estimates of β_1 and β_0; finding the empirical inefficiency of point-based variance estimates compared to model-based variance estimates; deriving a formula to compare $\widehat{l}$ under overlapping and non-overlapping block resampling; finding the number of possible unique bootstrap resamples for a given sample size; and writing original R code to implement nearly all of the techniques and applications presented.

Chapter 4

Appendix

4.1 Combinatorial aspects of the bootstrap

Proposition *A sample* $\chi = \{X_1, ..., X_n\}$, *where every element* X_i *is distinct, has* $N(n) = \begin{pmatrix} 2n-1 \\ n \end{pmatrix}$ *unique bootstrap resamples.*

Proof $N(n)$ equals the number of ways of placing n indistinguishable objects into n numbered boxes (box i representing X_i). The boxes are allowed to contain any number of objects, and number of objects in box i, represents the number of times X_i appears in the the sample.

Representing objects by $\bullet$ and separating boxes by $|$, the first $|$ denoting the beginning and another $|$ denoting the end, we can represent the arrangements as n $\bullet$ inserted between $n+1$ vertical lines $|$. For example $|\ |\bullet\bullet|\bullet\ \bullet\ \bullet|\ |\ |\bullet|$ means there are six boxes and six balls, and the number balls assigned to boxes one through six are {0,2,3,0,0,1} respectively. The problem can be though of as starting with $n + (n+1)$ $|$ and choosing n $|$ to replace with $\bullet$. Notice that the first and last $|$ cannot be chosen, hence we have $n + n - 1$ objects to choose from, and the number of configurations is $N(n) = \begin{pmatrix} 2n-1 \\ n \end{pmatrix}$.

■

$N(n)$ grows exponentially with n. Some values of $N(n)$ are shown in table (4.1). An asymptotic

n	$N(n)$
2	3
3	10
5	126
10	92378
20	6.9×10^{10}

Table 4.1: *N(n) for various n*

relation for $N(n)$ as $n \to \infty$ using Stirling's formula $n! \approx \sqrt{2\pi n} n^n e^{-n}$ is given by

$$\binom{2n-1}{n} \sim \frac{4^n}{2\sqrt{\pi n}}$$

4.2 R code for bootstrapping

This section presents all the R-based functions written for the purposes of this report. The two main functions are `boot` and `tsbooty` which implement univariate and time series bootstrapping respectively. While both are slower than pre-existing functions, namely 'boot' and 'tsboot' from the CRAN boot library, they have improved input/output calls and are simpler to use. In short, better.

4.2.1 boot

Performs univariate bootstrap estimation of bias and variance for an arbitrary statistic, *theta*. It generates R bootstrap replicates of *theta* applied to data. The kernel of the code comes from Efron and Tibshirani's (1993) 'bootstrap' function.

```
boot<-function(x,theta,R=600){
data<-matrix(sample(x,size=length(x)*R,replace=T),nrow=R)
thetastar<-apply(data,1,theta)
se<-(var(thetastar))^(1/2)
bias<-theta(x)-mean(thetastar)
original<-theta(x)
result<-cbind(original,bias,se)
return(result)}
```

4.2.2 modelboot

Performs model-based bootstrapping with R bootstrap resamples. Dataset must be a $n \times 2$ matrix with the explanatory variable in the first column and the response variable in the second column. By default (diag=1), the function draws the resample fits.

```
modelboot<-function(dataset,R,diag=1,lty=3){
lmfunc<-function(data){
model<-lm(data~dataset[,1])
coef<-model$coef
return(coef)}
plot(dataset[,1],dataset[,2])
model<-lm(dataset[,2]~dataset[,1])
rawreg<-model$resid
rreg<-rawreg/(1-lm.influence(model)$hat)^(1/2)
rregstar<-matrix(sample(rreg,size=length(rreg)*R,replace=T),nrow=R)
ystar<-matrix(nrow=R,ncol=length(rreg))
```

```
for (i in 1:R){
     ystar[i,]<-(model$coef[1]+model$coef[2]*t(dataset[,1])+rregstar[i,])}
     coef<-t(apply(ystar,1,lmfunc))
if (diag==1){
for (i in 1:R)
abline(coef[i,],lty=lty)}
original.b0<-model$coef[1]
original.b1<-model$coef[2]
var.b0<-var(coef[,1])
var.b1<-var(coef[,2])
bias.b0<-(model$coef[1]-mean(coef[,1]))
bias.b1<-(model$coef[2]-mean(coef[,2]))
return(cbind(original.b0,original.b1,bias.b0,bias.b1,var.b0,var.b1))}
```

4.2.3 pointboot

Performs model-based bootstrapping with R bootstrap resamples. Dataset must be a $n \times 2$ matrix with the explanatory variable in the first column and the response variable in the second column. By default (diag=1), the function draws the resample fits.

```
pointboot<-function(dataset,R,diag=1,lty=3){
lmfunc<-function(dataset,index){
model<-lm(dataset[,2*index]~dataset[,2*index-1])
coef<-model$coef
return(coef)}
coefstar<-matrix(nrow=R,ncol=2)
coef<-lm(dataset[,2]~dataset[,1])$coef
plot(dataset[,1],dataset[,2])
resample<-matrix(nrow=l,ncol=(2*R))
model<-lm(dataset[,2]~dataset[,1])
l<-length(dataset[,1])
for (i in 1:R){
     r<-sample(1:l,l,replace=T)
     for(p in 1:l){
        resample[p,(2*i-1):(2*i)]<-dataset[r[p],]}
     }
for (i in 1:R){
     coefstar[i,]<-t(lmfunc(resample,i))}
if (diag==1){
    for (i in 1:R)
    abline(coefstar[i,],lty=lty)
    abline(coef,col="red")
    }
```

```
original.b0<-coef[1]
original.b1<-coef[2]
var.b0<-var(coefstar[,1])
var.b1<-var(coefstar[,2])
bias.b0<-(coef[1]-mean(coefstar[,1]))
bias.b1<-(coef[2]-mean(coefstar[,2]))
return(cbind(original.b0,original.b1,bias.b0,bias.b1,var.b0,var.b1))}
```

4.2.4 tsmodelboot

Generates R bootstrap replicates of a statistic applied to a time series using model-based resampling. Reports variance and bias of statistic and model coefficients. The order of model may be specified, otherwise the default ARMA(1,1) model is used. If diag=1, the function displays the first resampled time series.

```
tsmodelboot<-function(data,stat,R,order,auto=F,diag=1){
if (auto==T)
   order<-c(1,1)
model<-arma(data,order)
coef<-model$coef
ar1<-coef[1:order[1]]
ma1<-coef[(order[1]+1):(order[1]+order[2])]
plot(data)
n<-length(data)
rawresid<-c(model$resid[(n+1-max(order)):n],na.omit(model$resid))
resid<-(rawresid-mean(rawresid))
residstar<-matrix(sample(resid,size=length(resid)*R,replace=T),nrow=R)
ystar<-matrix(nrow=R,ncol=length(resid))
residfun<-function(n){return(residstar[i,1:n])}
armamod<-function(data){return(arma(data,order)$coef)}
for (i in 1:R){
     ystar[i,]<-arima.sim(list(ar=ar1,ma=ma1),rand.gen=residfun,n=(length(data)))}
coefstar<-t(apply(ystar,1,armamod))
if (diag==1)
   lines(ystar[1,],lty=3)
bias.coef<-1:(sum(order)+1)
for (i in 1:(sum(order)+1))
bias.coef[i]<-(coef[i]-mean(coefstar[,i]))
original.coef<-c(ar1,ma1,coef[order[1]+order[2]+1])
var.coef<-diag(var(coefstar))
statstar<-apply(ystar,1,stat)
var.stat<-var(statstar)
bias.stat<-(stat(data)-mean(statstar))
```

```
return(list(cbind(original.coef,var.coef,bias.coef),cbind(var.stat,bias.stat)))}
```

4.2.5 tsbooty

Generates R bootstrap replicates of a statistic applied to a time series. The replicate time series are constructed according to Künsch's overlapping block rule (Künsch, 1989) with block length b. If diag=1, the function displays the first resampled time series. This function has similar speed to 'tsboot' from the 'boot' library, however is around 10-20 times slower than 'tsbootstrap' in the 'tseries' library.

```
tsbooty<-function(x,stat,R,b,diag=0){
n<-length(x)
k<-round(n/b)
resample<-1:(k*b)
data<-matrix(ncol=k*b,nrow=R)
random<-matrix(ncol=k,nrow=R)
for(i in 1:R){
    random[i,]<-sample(1:(n-b),k,replace=T)}
for(p in 1:R){
    for(i in 1:k){
        resample[(b*(i-1)+1):(i*b)]<-x[random[p,i]:(random[p,i]+b-1)]}
    data[p,]<-t(resample) }
statstar<-apply(data,1,stat)
se<-(var(statstar))^(1/2)
bias<-mean(statstar)-stat(x)
original<-stat(x)
result<-cbind(original,bias,se)
if (diag==0) return(result)
if (diag==1){ plot(ts(data[1,]))
return(list(result,data[1,])) } }
```

4.2.6 bootlength

Performs the block length selection algorithm proposed by Hall, Horowitz and Jing (1995) using 1 iteration. The argument "statistic" can either be a bias or variance estimate. R is the number of bootstrap simulations used; b is an initial guess for optimum block length with default $b = 3.15n^{1/3}$. Returns optimum block length "l". The method can be iterated by feeding "l" back into argument b. Calls `tsbootstrap` function in the tseries library.

```
bootlength<-function(data,statistic,R,m,b=NULL)
{
if (is.null(b))
b<-3.15*n^(1/3)
#constant of 3.15 is optimal for estimating var/bias in an AR(1) process
```

```
n<-length(data)
q<-(tsbootstrap(data,statistic=statistic,type="block",nb=R,b=b)$se^2)
x<-1:(n-m)
bias<-1:b
mse<-1:b
for (k in 1:b) {
     for (i in 1:(n-m)) {
     x[i]<-(tsbootstrap(data[i:(i+m-1)],statistic=statistic,
            type="block",nb=R,b=k)$se)^2}
mse[k]<-mean((x-q)^2)
bias[k]<-(mean(x-q))}
l<-(which.min(mse)*(n/m)^(1/3))
return(x,l,mse,bias)}
```

4.2.7 blocklength

Generates R bootstrap replicates of a statistic applied to a time series using `tsbooty` with block length b scanning from 1 to half the data length. Returns bias and variance of the statistic. Used to create figures 3.6 and 3.8.

```
blocklength<-function(data,stat,R){
n<-length(data)
mse<-1:(round(n/2))
bias<-1:(round(n/2))
variance<-1:(round(n/2))
for(l in 1:(round(n/2))) {
    boot<-tsbooty(data,stat=stat,R,B=l)
    bias[l]<-boot[2]
    variance[l]<-(boot[3])^2 }
mse<-bias^2+variance
return(bias,variance,mse) }
```

Bibliography

[1] Azzalini, A. & Bowman A. W. (1990), "A look at some data on the Old Faithful geyser," *Applied Statistics*, 39 pp. 357-365.

[2] Carlstein, E. (1986), "The use of subseries methods for estimating the variance of a general statistic from a stationary time series," *Annals of Statistics*, 14 pp. 1171-9.

[3] Clevelabnd, W. S., Devlin, S. J. & Terpenning, I. J. (1981), "The details of the SABL transformation, decomposition and calendar methods," *Computing Information Library, Bell Labs.*

[4] Davidson, A. C. & Hinkley, D. V. (1997), "Bootstrap Methods and their Application," *Cambridge Publishing* pp. 105-6.

[5] Efron, B & Tibshirani, R. J. (1987), "Bootstrap methods for standard errors, confidence intervals and other measurements of statistical accuracy," *Statistic Science*, 1 pp. 54-77.

[6] Efron, B & Tibshirani, R. J. (1993), "An Introduction to the Bootstrap," *Chapman and Hall* pp. 137.

[7] Freedman, D. A. (1984), "On bootstrapping two-stage least squares estimates in stationary linear models," *Annals of Statistics*, 12 pp. 827-842.

[8] Hall, P. (1985), "Resampling a coverage process," *Stochastic Process Applications*, 19 pp. 259-269.

[9] Hall, P., Horowitz, J. L. & Jing, B. Y. (1995), "On blocking rules for the bootstrap with dependent data," *Biometrika*, 82 pp. 561-74.

[10] Hall, P. (1997), "The Bootstrap and Edgeworth Expansion," *Springer Publishing*, pp. 35-36.

[11] Hall, P. (2004), "Methodology and Theory for the Bootstrap," lecture 2, slide 16-17.

[12] Keeling, C. D. and Whorf, T. P. (1997), Scripps Institution of Oceanography, University of California, La Jolla, California USA.

[13] Kenney, J. F. & Keeping, E. S. (1951), "Mathematics of Statistics," 2nd ed. *Princeton Publishing* pp. 164.

[14] Künsch, H. R. (1989), "The Jackknife and the bootstrap for general stationary observations," *Annals of Statistics*, 17 pp. 1217-41.

[15] Li, H & Maddala, G. (1996), "Bootstrapping time series models," *Econometric Reviews*, 15 pp. 115-158.

[16] Liu, R. Y. (1988), "Bootstrap procedures under some non-i.i.d models," *Annals of Statistics*, 16 pp. 1696-1708.

[17] Politis, D. N. & Romano, J. P. (1994), " The Stationary Bootstrap," *Journal of the American Statistical Association*, 82 pp. 1303-1313.