YANGON INSTITUTE OF ECONOMICS

Ph.D. PROGRAMME

BOOTSTRAPPING CONFIDENCE INTERVALS: AN APPLICATION TO FORECASTING THE PRODUCTIVITY OF SPIRULINA

WIN TUN

NOVEMBER, 2011

YANGON INSTITUTE OF ECONOMICS

Ph.D. PROGRAMME

BOOTSTRAPPING CONFIDENCE INTERVALS:
AN APPLICATION TO FORECASTING THE PRODUCTIVITY OF SPIRULINA

WIN TUN

NOVEMBER, 2011

YANGON INSTITUTE OF ECONOMICS

Ph.D. PROGRAMME

BOOTSTRAPPING CONFIDENCE INTERVALS: AN APPLICATION TO FORECASTING THE PRODUCTIVITY OF SPIRULINA

by

Win Tun

4 Ph.D. Res-Ah-1

NOVEMBER, 2011

BOOTSTRAPPING CONFIDENCE INTERVALS: AN APPLICATION TO FORECASTING THE PRODUCTIVITY OF SPIRULINA

Partial fulfillment of the requirement for the degree of

Ph.D.

of the Department of Statistics

Yangon Institute of Economics

Submitted by

Win Tun

November, 2011

Certification

I hereby certify that the content of this thesis is wholly my own work unless otherwise referenced or acknowledged.

Win Tun

4 Ph.D. Res-Ah-1

BOOTSTRAPPING CONFIDENCE INTERVALS: AN APPLICATION TO FORECASTING THE PRODUCTIVITY OF SPIRULINA

Board of Examiners

Prof. Dr. Daw Khin San Yee
(Chairman)
Rector
Yangon Institute of Economics

Prof. Dr. Daw Than Toe
(External Examiner)
Professor / Head (Retd.)
Department of Statistics
Yangon Institute of Economics

Prof. Dr. Thet Lwin
(Referee)
Professor / Head (Retd.)
Department of Statistics
Yangon Institute of Economics

ဗွယ်ဗွန်

U Myint Swe
(Supervisor)
Associate Professor (Retd.)
Department of Statistics
Yangon Institute of Economics

Prof. Dr. Lay Kyi
(Supervisor)
Professor / Head
Department of Statistics
Yangon Institute of Economics

3/12/11

Prof. Dr. Soe Win
(Member)
Director General
Department of Higher Education
(Upper Myanmar)

Prof. U Kyaw Min Htun
(Member)
Pro-Rector (Retd.)
Yangon Institute of Economics

Prof. Daw Khin San Myint
(Member)
Professor / Head (Retd.)
Department of Statistics
Yangon Institute of Economics

Prof. Dr. Daw San Kyi
(Member)
Professor (Retd.)
Department of Statistics
Yangon Institute of Economics

Prof. Dr. Daw Khin May Than
(Internal Examiner)
Professor / Head
Department of Economics
Yangon University of Distance Education

Prof. Dr. Thaung Htay
(Internal Examiner)
Professor / Head
Department of Statistics
Monywa Institute of Economics

ABSTRACT

This study aims to develop a relatively new bootstrap method in which a dataset applied to regression analysis is contaminated with outliers and to apply it in the construction of prediction intervals for Spirulina productivity; the productivity is measured in terms of the optical density of Spirulina. The importance of this study lies not only in the simplicity of the proposed method but also in the accuracy of prediction intervals for the optical density of Spirulina in culturing ponds at Myanmar Spirulina Factory.

In analyzing a dataset contaminated with outliers in linear regression model, a relatively new bootstrap method was proposed to find reliable distributions of the regression estimates. Simulation results have shown that the proposed bootstrap method outperforms the residual resampling bootstrapping.

The residual resampling bootstrapping and proposed bootstrap method were applied to the construction of prediction intervals for the optical density of Spirulina. At first, 95% one-day-ahead bootstrap prediction intervals were computed on the basis of 364 daily recorded cases on the optical density of Spirulina and related variables of 2007 using regression model with dynamic behavior (dynamic regression model) and the residual resampling bootstrapping. It was found that 94.0% of actual readings of the optical density of Spirulina fell within the computed prediction intervals. Next, using linear regression model and the proposed bootstrap method, 95% prediction intervals were computed based on only 226 cases left after removing the cases in which the optical density of Spirulina in a day was less than that in the previous day. 96.4% of actual readings of the optical density of Spirulina were found to fall within the computed prediction intervals. According to the comparative results between the widths of respective prediction intervals obtained from both regression models, it was concluded that prediction intervals obtained from the linear regression model were more precise than those obtained from the dynamic regression model and desirable to be applied in making forecasts.

ACKNOWLEDGEMENTS

I would like to express my gratitude to Prof. Dr. Kan Zaw, Rector (Retd.), Yangon Institute of Economics, for his kind permission to conduct this piece of work. I am very grateful to Prof. Dr. Daw Khin San Yee, Rector, Yangon Institute of Economics, for her guidance in submitting this thesis.

This thesis could not have been completed without consistent guidance with valuable suggestions and comments throughout the study from my late supervisor, U Myint Swe, Associate Professor (Retd.), Department of Statistics, Yangon Institute of Economics. I also would like to acknowledge the valuable supervision and suggestions given by my supervisor Prof. Dr. Lay Kyi, Head, Department of Statistics, Yangon Institute of Economics.

I am very indebted to Prof. U Kyaw Min Htun, Pro-Rector (Retd.), Yangon Institute of Economics, Prof. Dr. Thet Lwin, Head (Retd.), Department of Statistics, Yangon Institute of Economics, Prof. Daw Khin San Myint, Head (Retd.), Department of Statistics, Yangon Institute of Economics, Prof. Dr. Daw Than Toe, Head (Retd.), Department of Statistics, Yangon Institute of Economics, and Prof. Dr. Daw San Kyi, Professor (Retd.), Department of Statistics, Yangon Institute of Economics, who posed many penetrating and insightful questions, and contributed many valuable suggestions and comments towards my thesis.

A special thank is due to Prof. Dr. Min Thein, General Manager (Retd.), Myanma Spirulina Factory, Ministry of Industry (1), for providing relevant data on the productivity of Spirulina.

Last but not the least, I thank my family for their patience during the hard period this thesis absorbed my time and energy.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AR(1)	First-Order Autoregressive
ARMA	Autoregressive Moving Average
BG test	Breusch-Godfrey test
BOLS	Bootstrap Based on OLS Estimators
CDF	Cumulative Distribution Function
EDF	Empirical Distribution Function
FGLS	Feasible Generalized Least Squares
GLS	Generalized Least Squares
JB test	Jarque-Bera test
LTS	Least Trimmed Squares
MSF	Myanma Spirulina Factory
OLS	Ordinary Least Squares
OLS-LTS	OLS Based on LTS Weights
PDF	Probability Density Function
RBOLS	Resistant Bootstrap Based on OLS Estimators
rmse	Estimated Root Mean Squared Error
$rmse^*$	Bootstrap Root Mean Squared Error
WLS	Weighted Least Squares

CHAPTER I
INTRODUCTION

1.1 Rationale of the Study

In Myanmar, several Spirulina based drugs and consumer goods are manufactured by MSF which is situated in Sagaing Township, Sagaing Region under the Ministry of Industry (1). Raw Spirulina is obtained from four natural lakes in Sagaing Region, namely, Lake Twyn Taung in Budalin Township, Lake Ye Kharr in Sagaing Township, Lake Taung Pyauk in Kani Township and Lake Twyn Ma in Kani Township. Moreover, it is also obtained from 20 culturing ponds of Spirulina constructed within the compound of MSF.

A good quality of Spirulina can be produced by artificial culture method. Spirulina strain from Lake Twyn Taung is chosen for artificial culture. MSF media is prepared with Ayeyarwady river water and some nutrients. The selected Spirulina strain is added to MSF media with aeration at one over night. Then, it is transferred to a 400 liter volume photobioreactor (Plate A1, Appendix A). When initial optical density of Spirulina of 0.3 (in 680 nanometer) reaches 1.2 (in 680 nanometer) in the photobioreactor, it is transferred to an inoculation pond (Plate A2) with 30,000 gallons of MSF media. After one week, 30 percent of media with Spirulina from the inoculation pond is put in a culturing pond (Plate A3) with 120,000 gallons of MSF media. This culturing pond is initiated at the optical density of 0.3 (in 680 nanometer), pH value of 8.5 and salinity of 3 to 4 (in part per thousand). Spirulina in the culturing pond is harvested approximately every four to five days, depending on the rate of Spirulina growth. About 75 percent is harvested, and 25 percent is transferred back to the culturing pond. Harvesting is done through the cascade filter (Plate A4) (May Yu Khaing, 2007).

The officials of MSF want to understand some explanatory variables which have significant effect on the Spirulina productivity in culturing ponds and to obtain forecasts of Spirulina productivity at different levels of these explanatory variables. Therefore, an attempt has been made in this study to

construct forecasting models and prediction intervals for Spirulina productivity in culturing ponds at MSF.

It is interesting to note that, at MSF, measurement of the optical density on a particular *day (t)* is taken to be Spirulina productivity of that day. Therefore, in what follows, it is to be understood that Spirulina productivity on a day is expressed in terms of the optical density of Spirulina, but not in terms of weight or volume of Spirulina harvested, if any, on that day.

Some studies on the optical density of Spirulina have shown, as a hypothesis, that the optical density of Spirulina on a particular day might mainly depend on the previous day's (i) optical density of Spirulina, (ii) salinity of water, (iii) pH value of water, (iv) air temperature, (v) light, (vi) season, and (vii) condition that Spirulina was harvested or not. Therefore, a culturing pond is randomly chosen out of 20 culturing ponds to take data on the above said variables at ten o'clock everyday in 2007. In the dataset collected by MSF at the chosen pond, the optical density of Spirulina is measured in *680 nanometer*, salinity of the water is measured in *part per thousand*, air temperature is measured in *centigrade* and light is measured in *watt per (meter)*^{2}.

Based on the dataset over a period of 365 days for the year 2007, a dynamic regression model of the optical density of Spirulina that is supposed to be consistent with the above said hypothesis is fitted by the method of Ordinary Least Square (OLS). In this case, the disturbances are found to be autocorrelated. Therefore, the model is fitted again by the method of Feasible Generalized Least Squares (FGLS). In the model, since the disturbances are not normally distributed, the residual resampling bootstrapping is used to generate the distributions of the FGLS estimates and then to construct the bootstrap prediction intervals for the optical density of Spirulina.

When the original dataset does not contain any outliers, bootstrap distribution of the FGLS estimate is desirable. However, when the dataset is contaminated with outliers, bootstrap distribution is a very poor estimator of the distribution of the FGLS estimate. In such a situation, in order to obtain reliable distribution of the regression estimate an alternative bootstrap method

which is not only computationally simple but also resistant to the effects of outliers is proposed.

In some cases of the dataset used for fitting the dynamic regression model, the optical density of Spirulina on *day (t+1)* is lower than that on *day (t)* because Spirulina was harvested on *day (t)*. These cases are removed from the dataset in undertaking the analysis and a linear regression model of the optical density of Spirulina is fitted based on the rest of the original dataset. In this case, the disturbances do not follow the normal distribution. Therefore, the proposed bootstrap method is applied to construct the bootstrap prediction intervals for the optical density of Spirulina.

1.2 Objectives of the Study

The objectives of the study are as follows:

(i) To develop an alternative bootstrap method, that provides reliable bootstrap distributions of the regression estimates in linear regression model, whenever the dataset is contaminated with outliers

(ii) To illustrate an empirical computation of bootstrap confidence intervals for the regression parameters and bootstrap prediction intervals for Spirulina productivity in culturing ponds at MSF using (i) the residual resampling bootstrapping in dynamic regression model, and (ii) the proposed bootstrap method in linear regression model.

1.3 Scope and Limitation of the Study

In the dataset collected by the MSF from a randomly chosen culturing pond, data were recorded daily only on such variables as (i) optical density of Spirulina, (ii) salinity of water, (iii) pH value of water, (iv) air temperature, (v) light, and (vi) season. Upon inspection of all available datasets recorded by MSF for some years, the year 2007 was found to be the only year in which all 365 data points on the above said variables had been recorded on a daily basis. Therefore, the recorded dataset for 2007 was chosen and collected from MSF.

In the dataset recorded and compiled on a daily basis, it was found that the amount of Spirulina harvested was not reported nor recorded on the day on which other variables were observed. Therefore, the dataset lacked the amount of Spirulina harvested on each day. Instead, the state of Spirulina harvest by a day was available, if Spirulina had been harvested on that day.

1.4 Method of the Study

The method of the study is an analytical one, in which an alternative bootstrap method in linear regression model is proposed. Based on daily data on Spirulina productivity and related variables in a randomly chosen culturing pond at MSF, an illustration of fitting a dynamic regression model and a linear regression model for Spirulina productivity is carried out. To construct bootstrap confidence intervals for the regression parameters and prediction intervals for Spirulina productivity, the residual resampling bootstrapping is used in dynamic regression model, and the proposed bootstrap method is applied in linear regression model.

1.5 Organization of the Study

This study is divided into six chapters. Chapter I is concerned with introduction. It presents rationale, objectives, scope and limitation, method, and organization of the study. Chapter II deals with a literature review on bootstrap methods. Chapter III presents methods concerning with linear regression model and dynamic regression model. In the same chapter the basics of bootstrap and some bootstrap methods involving regression models including specific algorithms for bootstrapping in regression models are also presented.

Chapter IV provides an alternative bootstrap method that provides reliable bootstrap distributions of the regression estimates in linear regression model whenever the dataset is contaminated with outliers.

Chapter V presents an algorithm for the residual resampling bootstrapping in dynamic regression model with First-Order Autoregressive

(AR(1)) scheme of disturbances. In the same chapter an illustration of practical application of bootstrap methods is provided to Spirulina productivity in culturing ponds at MSF. Based on collected data on Spirulina productivity and related variables for the year 2007, a dynamic regression model and a linear regression model for Spirulina productivity are fitted. Using the residual resampling bootstrapping in dynamic regression model and the proposed bootstrap method in linear regression model, the prediction intervals for Spirulina productivity are computed.

Based on the results and findings in Chapter IV and Chapter V, conclusion on performance of the proposed bootstrap method and suggestions on prediction intervals for Spirulina productivity, together with recommendations, are presented in Chapter VI.

CHAPTER II
LITERATURE REVIEW ON BOOTSTRAP METHODS

In this Chapter II, a literature review on basic notions about bootstrap methods, specifically on robust bootstrap methods in regression analysis is presented. The purpose of this review is to present the process of development of bootstrap methods, especially in the field of robust bootstrap methods in regression analysis.

2.1 Basic Bootstrap Methods

Monte Carlo methods of statistical inference had already been used for many years when Efron (1979) made the connection to standard methods of parametric inference, drew the attention of statisticians to their potential for nonparametric inference, and introduced the term 'bootstrap'. This work made strong connections with the jackknife, which had been introduced by Quenouille and Tukey. The jackknife is an interesting nonparametric method for estimating the bias and variance of a statistic of interest, and also for testing the null hypothesis that the distribution of a statistic is centered at some prespecified point. Efron reported that bootstrap method generally works more satisfactorily than jackknife on a variety of estimation problems. The jackknife was shown to be a particular linear approximation method for the bootstrap. Efron proceeded the exposition by a series of examples: variance of the sample median, error rate in a linear discriminant analysis, ratio estimation and estimating regression parameters, among many other examples.

Bickel and Freedman (1981) were among the first to discuss the conditions under which the bootstrap is consistent. In a latter paper by Efron and Gong (1983), they could provide a general review of the bootstrap method. The essential idea for bootstrap method is to offer a computer-intensive method of generating reasonable and reliable probability distributions in circumstances in which precise mathematical reasoning is intractable. Athreya (1987) showed

that the bootstrap can fail for long-tailed distributions. Shao and Tu (1995) gave an extensive theoretical overview of the bootstrap.

Bootstrap confidence intervals were introduced in the original bootstrap paper by Efron (1979); bias adjustment and studentizing were discussed by Efron (1981). Hall (1986) analyzed the effect of discreteness on confidence intervals. The adjusted percentile method was developed by Efron (1987). Efron (1987) investigated construction of better bootstrap confidence intervals for a single parameter in a multiparameter family. According to Efron, the standard approximate intervals based on maximum likelihood theory can be quite misleading. In practice, tricks based on transformation, bias correction and so forth, are often used to improve their accuracy. The bootstrap confidence intervals proposed by Efron automatically incorporate such tricks without requiring the situations to think them through for each new application, at the price of a considerable increase in computational effort. The new intervals by Efron were found to incorporate an improvement over previously suggested classical methods. In addition to parameteric families, Efron developed better bootstrap intervals for nonparametric situations. Hall (1988) strongly advocated the use of studentized bootstrap statistics for confidence intervals and significance tests. An earlier review of bootstrap confidence intervals, with discussion, was given by DiCiccio and Romano (1988). Geisser (1993) surveyed several approaches to calculating prediction intervals.

2.2 Bootstrap Methods in Regression Analysis

The use of bootstrap methods in regression was initiated by Efron (1979). Important early work on the theory of resampling for linear regression was due to Freedman (1981). Freedman (1981) showed that with independent and identically distributed disturbances, the bootstrap approximation to the distribution of the OLS estimator is valid. That is, as both the sample size n and the number of bootstrap samples B increase, the bootstrap distribution converges to the true distribution of the OLS estimator.

Freedman and Peters (1984) applied the bootstrap method to a seemingly unrelated regression model of the demand for energy by industry. The model was estimated by the method of Generalized Least Squares (GLS). The two alternative ways to obtain standard errors for the GLS parameter estimates were the conventional asymptotic formulas and the bootstrap method. Monte Carlo simulations were also used by Freedman and Peters to compare the performance, in a finite-size sample of these two alternatives. The conventional asymptotic estimates were found to be far too optimistic, though the bootstrapped standard errors were only slightly optimistic. Based on their findings, Freeman and Peters concluded that the bootstrap method provides much more realistic standard errors than the conventional asymptotic theory.

In fact, the bootstrap relies on resampling from an independent and identical distribution. Time series data, therefore, present obvious problems as the result of dependence error terms in the model. The bootstrap is straightforward in the linear model with an Autoregressive Moving Average (ARMA) error structure and resampling the underlying white noise error. Freedman (1984) first introduced bootstrap method for a dynamic linear simultaneous equations regression model estimated by the method of two-stage least squares in the context of linear models. This method assumes the underlying error is independently and identically distributed. For general dependent data without ARMA specification, for example, nonstationary data, the moving blocks bootstrap method can be used satisfactorily.

Use of the bootstrap for calculating prediction intervals for regression model was discussed by Stine (1985).

Bernard and Veall (1987) employed the bootstrapping technique to estimate the probability distribution of future electricity demand for Hydro Quebec of Canada. Their application followed the regression approach of Freedman and Peters (1984) but also allowed for serially correlated disturbances and uncertainty in the independent variable forecasts. The article of Bernard and Veall (1987) illustrated the case of bootstrapping to estimate the probability distribution of future peak demand for Hydro Quebec, conditional

on current information. The results illustrated that reasonable estimates of demand uncertainty are typically very large and hence such measures should likely play an important role in the planning process. If anything, these large uncertainty estimates from the nonparametric bootstrap are likely to be conservative because parametric bootstrap estimates are even larger. Moreover, they employed deterministic time trends as opposed to more volatile autoregressive integrated moving average processes, and it has been assumed throughout that there will be no structural shifts before the target period of the forecast.

Prescott and Stengos (1987) made an effort to demonstrate how the distribution-free method of bootstrapping can be applied to the construction of confidence intervals for forecasts generated by a dynamic econometric model. In their paper, because the exogenous variables must be forecast, the forecast of the dependent variable were taken to be the functions of stochastic forecast-period exogenous variables. Using a simple autoregressive model of U.S. pork supply, they illustrated how the bootstrap method can account for the sources of randomness in forecast errors, including the errors due to the use of estimated structural parameters, the lack of independence of forecasts produced by an autoregressive model, and the stochastic nature of forecast-period exogenous variables. According to Prescott and Stengos, the flexibility of the bootstrapping approach to constructing forecast intervals and the lack of robust alternatives were strong motivations for further research and software development in the research area of bootstrap methods.

Hall (1989) showed that bootstrap methods can provide unusually accurate confidence intervals in regression problems. Olshen *et al.* (1989) described an interesting application to a complicated prediction problem in the context of regression analysis.

Kim (2005) proposed an improved bootstrap procedure when statistical inference is conducted for the regression model with AR(1) disturbances. It is distinct from the past studies on the following points. First, bias-correction is conducted in two stages of the bootstrap. That is, pseudo-datasets of the

bootstrap are generated using a bias-corrected estimator for the AR(1) coefficient, and then bias-correction is again given to the AR(1) coefficient estimate obtained from the pseudo-datasets. For this purpose, the bias-corrected estimators based on the bootstrap and jackknife methods are used. Secondly, the FGLS estimates for regression coefficients are re-calculated using the bias-corrected estimate for the AR(1) coefficient, again in two stages of the bootstrap. As a result, bias-correction is implemented to estimation of the regression coefficients as well as to the AR(1) coefficient. The third point is related to the way in which bootstrap inference is carried out.

In his article, Kim (2005) reported that the bias-corrected bootstrap substantially improves size distortions of the statistical test in the regression model with autocorrelated errors. The bias-corrected bootstrap based on the test statistic approach was found to provide superior size properties to that based on the confidence region approach, especially when the sample size is small. Both the bootstrap and jackknife were found to be effective for bias-correction, but the results suggested that bootstrap be preferred as a means of bias-correction when the sample size is more than moderately large.

2.3 Robust Bootstrap Methods in Regression Analysis

De Angelis, Hall and Young (1993) gave a detailed theoretical analysis of residual resampling bootstrapping in L_1 estimation.

Salibian-Barrera and Zamar (2002) proposed an alternative bootstrap method, which is called fast bootstrap, to estimate the distribution and standard error of robust regression MM-estimates calculated with an initial S-estimate. Their method is to bootstrap a re-weighted representation of the estimate. For each bootstrap sample, only a weighted average to recalculate the scale estimate as well as a weighted least squares estimate have to be calculated to obtain the bootstrap regression estimate.

Willems and Aelst (2004) proposed a simple approximating bootstrap method for Least Trimmed Squares (LTS), which is called fast and robust bootstrap for LTS. Their method is to draw bootstrap samples, but instead of

recalculating the LTS estimates in each bootstrap sample; an approximation is computed using information gathered from the LTS solution of the original dataset.

Midi *et al.* (2009) proposed a bootstrap algorithm based on LTS estimator. This method used the residual bootstrap with LTS estimator, instead of OLS estimator, for the original sample as well as bootstrap samples. In the bootstrap algorithm of Midi *et al.* (2009), any bootstrap sample, which has percentage of outlying residuals larger than the breakdown point which is equal to $[\{(n-p)/2+1\}/n]$, is omitted and is replaced with a new sample.

CHAPTER III
STATISTICAL METHODS USED IN BOOTSTRAPPING

In the preceding Chapter II, a review of literature on bootstrap methods was presented. In this Chapter III, statistical methods applied as the major tools of bootstrapping in this study are presented. In the earlier sections, a brief description of the linear regression model, Jarque-Bera test for testing the normality of disturbances and White test for testing heteroscedasticity of disturbances, robust regression, statistical methods for dynamic regression model with AR(1) scheme of disturbances, and one-step-ahead forecasting are provided. In the later sections, the basics of bootstrap and some bootstrap methods involving regression models are presented. Moreover, two algorithms for bootstrapping in dynamic regression model and linear regression model with AR(1) disturbances are also presented.

3.1 Linear Regression Model

One of the most important, frequent and widely used types of statistical analysis in practice is regression analysis, in which one studies the effects of explanatory variables or covariates on a response variable of interest. The major objective of traditional regression analysis is to estimate and/or predict the mean or average value of the response variable on the basis of the known or fixed values of the explanatory variables.

The most commonly used regression model is a linear regression model that is considered as

$$Y_i = \beta_0 + \beta_1 X_{i1} + \ldots + \beta_p X_{ip} + u_i \qquad \text{for } i=1,\ldots,n, \tag{3.1}$$

where $\beta_0, \beta_1, \ldots, \beta_p$ are the unknown parameters of interest, Y_i stands for the response variable, and $X_{i1}, \ldots, X_{ip}$ are the explanatory variables. Classical theory assumes the disturbance term u_i to have a normal distribution with mean 0 and constant variance σ^2.

The OLS regression method minimizes the sum of squares of the residuals e_i where $e_i = Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_{i1} + ... + \hat{\beta}_p X_{ip})$. Formally, this can be written as

$$\underset{(\hat{\beta}_0,...,\hat{\beta}_p)}{Minimize} \sum_{i=1}^{n} e_i^2 . \tag{3.2}$$

The basic idea is to make the sum of all the squared residuals as small as possible.

In a more convenient vector form, the model in Equation (3.1) can be expressed as

$$Y_i = x_i' \beta + u_i , \tag{3.3}$$

with $x_i' = (1, X_{i1},..., X_{ip})$. The combined matrix representation for all response $y' = (Y_1,..., Y_n)$ is

$$y = X\beta + u , \tag{3.4}$$

with $X' = (x_1,..., x_n)$ and $u' = (u_1,..., u_n)$.

The OLS estimates of β for Equation (3.4) based on observed response vector y are

$$\hat{\beta} = (X'X)^{-1} X'y ,$$

and corresponding fitted values are

$$\hat{Y} = Hy ,$$

where $H = X(X'X)^{-1} X'$ is the matrix, whose diagonal elements h_{ii} — denoted by h_i for simplicity. The residuals are

$$e = (I - H)y . \tag{3.5}$$

Under homoscedasticity of disturbances the standard formula for the estimated variance of $\hat{\beta}$ is

$$\hat{V}(\hat{\beta}) \approx s^2 (X'X)^{-1} ,$$

with s^2 equal to the residual mean square $\dfrac{1}{(n-p-1)} e'e$.

3.2 Testing for Normality and Heteroscedasticity of Disturbances

If Equation (3.1) holds with homoscedastic random disturbances u_i and if those random disturbances are normally distributed, or if the dataset is large, then standard distributional results will be adequate for drawing inferences with the least squares estimates. If non-normality or heteroscedasticity appears to be present, then robust regression estimates may be considered in place of the least squares estimates. In this section, Jarque-Bera test for testing the normality of disturbances and White test for testing heteroscedasticity of disturbances are briefly explained.

3.2.1 Jarque-Bera Test of Normality

Several tests of normality are discussed in the literature. One of the widely used tests of normality is the Jarque-Bera (JB) test. It is a large-sample test. It is based on the OLS residuals. This test first computes the skewness and kurtosis measures of the OLS residuals and uses the following test statistic;

$$JB = n\left[\frac{S^2}{6} + \frac{(K-3)^2}{24}\right], \tag{3.6}$$

where S represents skewness and K represents kurtosis.

Under the null hypothesis that the disturbances are normally distributed, Jarque and Bera (1987) showed that asymptotically the JB statistic given in Equation (3.6) follows the chi-square distribution with 2 degrees of freedom.

If the p-value of the computed chi-square statistic in an application is sufficiently low, one can reject the null hypothesis that the disturbances are normally distributed. But if the p-value is reasonably high, one does not reject the null hypothesis of normality assumption.

3.2.2 White's General Heteroscedasticity Test

The general test of heteroscedasticity proposed by White (1980) is easy to apply. As an illustration of the basic idea, consider the linear regression model in Equation (3.1) with p = 2:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + u_i . \tag{3.7}$$

The White test proceeds as follows:

Step 1. Given the data, fit Equation (3.7) by OLS method and obtain the residuals, $\hat{u}_i$.

Step 2. Run the following (auxiliary) regression:

$$\hat{u}_i^2 = \alpha_0 + \alpha_1 X_{i1} + \alpha_2 X_{i2} + \alpha_3 X_{i1}^2 + \alpha_4 X_{i2}^2 + \alpha_5 X_{i1} X_{i2} + v_i .$$

That is, the squared residuals from the original regression are regressed on the original X variables or regressors, their squared values, and the cross product(s) of the regressors. Obtain the R^2 from this (auxiliary) regression.

Step 3. Under the null hypothesis that there is no heteroscedasticity, it can be shown that sample size (n) times the R^2 obtained from the auxiliary regression asymptotically follows the chi-square distribution with degrees of freedom equal to the number of regressors (excluding the constant term) in the auxiliary regression. That is,

$$nR^2 \underset{asy}{\sim} \chi_{df}^2 ,$$

where *df* stands for degrees of freedom. In this example, there are 5 degrees of freedom since there are 5 regressors in the auxiliary regression. If, in an application, the p-value of the computed test statistic nR^2 is sufficiently low, the conclusion is that there is heteroscedasticity.

The test uses many degrees of freedom for models with just a moderate number of regressors. It is possible to obtain a test that is easier to apply than the White test and more conserving on degrees of freedom. The fitted values are defined, for each observation i, by

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \ldots + \hat{\beta}_p X_{ip} .$$

These $\hat{y}_i$s are just linear functions of the regressors. If the fitted values are squared, a particular function of all the squares and cross products of the

regressors is obtained. This suggests testing for heteroscedasticity by fitting the equation

$$\hat{u}_i^2 = \delta_0 + \delta_1 \hat{y}_i + \delta_2 \hat{y}_i^2 + v_i . \tag{3.8}$$

The nR^2 statistic for the null hypothesis that there is no heteroscedasticity can be used. The test from Equation (3.8) can be viewed as a special case of the White test.

A Special Case of the White Test for Heteroscedasticity

Step 1. Fit the Equation (3.1) by OLS method. Obtain OLS residuals $\hat{u}_i$ and the fitted values $\hat{y}_i$. Compute the squared OLS residuals $\hat{u}_i^2$ and the squared fitted values $\hat{y}_i^2$.

Step 2. Run the regression in Equation (3.8). Keep the value of R^2 from this regression.

Step 3. From the test statistic nR^2, compute the p-value using the χ_2^2 distribution.

3.3 Robust Regression

Use of least squares regression estimates is preferred when errors are near-normal in distribution and homoscedastic. However, the estimates are very sensitive to outliers; those are cases which deviate strongly from the general relationship. Any regression analysis should, therefore, include appropriate inspection of diagnostics based on residuals to detect outliers. Depending on the general pattern of residuals, one may feel confident in fitting by least squares, or a more robust regression method, that can resist several outliers, may be chosen to be safe.

A number of robust regression methods that provide stable results in the presence of outliers have been investigated by many researchers in the regression literature. In the case of statistical application of robust regression, the methods most commonly used are found to be M estimation, high breakdown value estimation, and combinations of these two methods.

Specifically, four such methods are M estimation, LTS estimation, S estimation, and MM estimation.

In this section, the breakdown value, which is a rough but useful measure of robustness of an estimator, and the LTS estimation, which is the one of the preferred robust regression methods, are briefly explained.

3.3.1 The Breakdown Value

Consider a dataset $Z = \{(x_{i1},...,x_{ip}, y_i); i = 1,...,n\}$ and a regression estimator T. Applying T to Z yields a vector $(\hat{\beta}_0,...,\hat{\beta}_p)$ of regression coefficients. Now consider all possible contaminated data sets Z' obtained by replacing any m of the original observations by arbitrary values of $X_1,...,X_p,Y$. This yields the maximum bias

$$Maxbias(m;T,Z) := \underset{Z'}{Max} \|T(Z') - T(Z)\|, \tag{3.9}$$

where $\|..\|$ is the Euclidean norm. If m outliers can have an arbitrarily large effect on T, it follows that $Maxbias(m;T,Z) = \infty$; hence $T(Z')$ becomes useless. Therefore the breakdown value of the estimator T at the dataset Z is defined as

$$\varepsilon_n^*(T,Z) := Min\{m/n : Maxbias(m;T,Z) = \infty\}. \tag{3.10}$$

In other words, it is the smallest fraction of contamination that can cause the regression estimator T to run away arbitrarily far from $T(Z)$. For many estimators $\varepsilon_n^*(T,Z)$ varies only slightly with Z and n, so that its limiting value (for $n \to \infty$) is denoted by $\varepsilon^*(T)$.

For OLS, one outlier may be sufficient to destroy T. Its breakdown value is thus $\varepsilon_n^*(T,Z) = \frac{1}{n}$, hence $\varepsilon^*(T) = 0$. The following concerns estimators T with $\varepsilon^*(T)>0$, which will be called positive-breakdown estimators. Estimator T with $\varepsilon^*(T) = 50\%$ will be called high-breakdown estimators (Rousseeuw, 2006).

3.3.2 Least Trimmed Squares

A number of robust regression methods that provide resistant results in the presence of outliers have been investigated by many researchers in the regression literature. For datasets with possibly multiple outliers, diagnosis is aided by initial use of fitted method that is highly resistant to the effects of outliers. One preferred resistant method is the LTS estimation, given by

$$\underset{(\hat{\beta}_0,\ldots,\hat{\beta}_p)}{Minimize} \sum_{i=1}^{h} (e^2)_i , \tag{3.11}$$

where $(e^2)_1 \le \ldots \le (e^2)_n$ are the ordered squared residuals and h is defined in the range $\frac{n}{2}+1 \le h \le \frac{3n+p+1}{4}$. Equation (3.11) resembles method of OLS but does not count the largest squared residuals, thereby allowing the LTS fit to steer clear of outliers.

The LTS method is a high breakdown method. For the default setting $h \approx \frac{n}{2}$, $\varepsilon^* = 50\%$, whereas for larger h, $\varepsilon^* \approx \frac{(n-h)}{n}$.

Residuals from LTS fit should clearly identify outliers. The fit itself is not very efficient, and should best be thought of as an initial step in a more efficient analysis (Davison and Hinkley, 1997).

3.4 Statistical Methods Used in Dynamic Regression Model with AR(1) Disturbances

In this section, some salient statistical methods used in dynamic regression model with AR(1) scheme of disturbances are presented. In Subsection 3.4.1, dynamic regression model with AR(1) disturbances is presented for the purpose of familiarity with the notations to be used throughout the study. Detection of autocorrelation in the dynamic regression model, as a test of first-order autocorrelation (Durbin h-test) and a general test of autocorrelation (Breusch-Godfrey test) are briefly explained in Subsection 3.4.2. In Subsection 3.4.3, a remedial measure for autocorrelation (method of

GLS) is presented. In Subsection 3.4.4, estimation of first-order autocorrelation in the model is dealt with, following the Cochrane-Orcutt iterative method which has become quite popular in practice.

3.4.1 Dynamic Regression Model with AR(1) Disturbances

In regression analysis involving time series data, if the model includes one or more lagged values of the dependent variable among its explanatory variables, it is called a dynamic model (Gujarati and Sangeetha, 2007). The common form of a dynamic regression model is described as

$$Y_t = \beta_0 + \beta_1 X_t + \beta_2 Y_{t-1} + u_t, \tag{3.12}$$

where $\beta_0, \beta_1, ..., \beta_p$ are the unknown parameters of interest, Y_t stands for the dependent variable at time t, X_t is the explanatory variables at time t, Y_{t-1} is the one lagged dependent variable, and u_t is the disturbance term at time t. The OLS method may not be directly applicable to such models. The reason is probably the existence of autocorrelation in the disturbances. Since u_t are unobservable, the nature of autocorrelation is often a matter of speculation. In practice, it is usually assumed that the u_t follows the AR(1) scheme, namely,

$$u_t = \rho u_{t-1} + \eta_t, \tag{3.13}$$

where $|\rho| < 1$ and where the η_t follows the classical OLS assumptions of zero expected value, constant variance, and nonautocorrelation (Gujarati, 1995).

If an explanatory variable in a regression model is correlated with the stochastic disturbance term, OLS estimators of such model are not only biased but also inconsistent; that is, even if the sample size is increased indefinitely, the estimators do not approximate their true population values satisfactorily.

3.4.2 Detection of Autocorrelation

In this Subsection 3.4.2, two tests for autocorrelation in the dynamic regression model: (i) Durbin *h* test and (ii) Breusch-Godfrey test are briefly explained.

Durbin h Test of First-Order Autocorrelation

Durbin (1970) has proposed a large sample test of first-order autocorrelation when some of the regressors are lagged dependent variables. This statistic, called the *h*-statistic, is as follows:

$$h = \hat{\rho}\sqrt{\frac{n}{1 - n[\text{var}(\hat{\beta}_2)]}}, \tag{3.14}$$

where n = sample size, $\text{var}(\hat{\beta}_2)$ = estimated variance of the coefficient of the lagged Y_{t-1}, and $\hat{\rho}$ = estimate of the first-order autocorrelation ρ.

For large sample sizes, Durbin has shown that if $\rho = 0$, the *h*-statistic follows the standardized normal distribution. ρ can also be approximated from the estimated Durbin-Watson *d* statistic as follows:

$$\hat{\rho} \approx 1 - \frac{1}{2}d,$$

where $d = \dfrac{\sum_{t=2}^{n}(\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=2}^{n}\hat{u}_t^{\,2}}$. Therefore, Equation (3.14) can be written as

$$h \approx (1 - \frac{1}{2}d)\sqrt{\frac{n}{1 - n[\text{var}(\hat{\beta}_2)]}}. \tag{3.15}$$

The steps involved in the application of the *h*-statistic are as follows:

Step 1. Fit Equation (3.12) by OLS method and note $\text{var}(\hat{\beta}_2)$.

Step 2. Compute *h*-statistic.

Step 3. Assuming n is large, h is asymptotically normally distributed with zero mean and unit variance. Therefore, the decision rule is

(a) if h > 1.96, reject the null hypothesis that there is no positive first-order autocorrelation, and

(b) if h < -1.96, reject the null hypothesis that there is no negative first-order autocorrelation, but

(c) if h lies between -1.96 and 1.96, do not reject the null hypothesis that there is no first-order autocorrelation.

Breusch - Godfrey Test of Higher-Order Autocorrelation

Breusch (1978) and Godfrey (1978) have developed a test of autocorrelation that is general in the sense that it allows for (i) nonstochastic regressors, such as the lagged values of the regressand; (ii) higher-order autoregressive scheme and (iii) simple or higher-order moving averages of white noise error terms. Assume that the disturbance term u_t is generated by the p^{th}-order autoregressive, AR(p), scheme as follows:

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + ... + \rho_p u_{t-p} + v_t, \tag{3.16}$$

where v_t is a random disturbance term with zero mean and constant variance. The null hypothesis H_0 to be tested is that $H_0 : \rho_1 = \rho_2 = ... = \rho_p = 0$, that is, there is no autocorrelation of any order. The Breusch - Godfrey (BG) test involves the following steps:

Step 1. Fit Equation (3.12) by OLS method and obtain the residuals, $\hat{u}_t$.

Step 2. Run the following regression:

$$\hat{u}_t = \alpha_1 + \alpha_2 x_t + \alpha_3 y_{t-1} + \hat{\rho}_1 \hat{u}_{t-1} + \hat{\rho}_2 \hat{u}_{t-2} + ... + \hat{\rho}_p \hat{u}_{t-p} + w_t, \tag{3.17}$$

where w_t is the error term at time t. Obtain the R^2 value from this auxiliary regression.

Step 3. If the sample size *n* is large, Breusch and Godfrey have shown that

$$(n-p)R^2 \underset{asy}{\sim} \chi^2_p. \tag{3.18}$$

That is, asymptotically, (*n*-*p*) times the R^2 value obtained from the auxiliary regression of Equation (3.17) follows the chi-square distribution with p degrees of freedom. If in an application, the *p*-value of the computed test statistic $(n-p)R^2$ happens to be sufficiently low, the null hypothesis is rejected, in which case at least one ρ in Equation (3.16) is significantly different from zero.

If in Equation (3.16) p = 1, meaning first-order autoregression, then the BG test is known as Durbin's M test.

3.4.3 Remedial Measure for Autocorrelation

Since the OLS estimators are inefficient in the presence of autocorrelation, it is essential to seek a remedial measure. If the validity of Equation (3.13) is assumed and if the first-order autocorrelation is known, the problem of autocorrelation can be easily solved by transforming the data into a generalized difference equation. From Equations (3.12) and (3.13), the original variables in Equation (3.12) are transformed as

$$(Y_t - \rho Y_{t-1}) = \beta_0(1-\rho) + \beta_1(X_t - \rho X_{t-1}) + \beta_2(Y_{t-1} - \rho Y_{t-2}) + \eta_t, \qquad (3.19)$$

where ρ is the first-order autocorrelation. Equation (3.19) can be expressed as

$$Y_t^T = \beta_0^T + \beta_1 X_t^T + \beta_2 Y_{t-1}^T + \eta_t, \qquad (3.20)$$

where $Y_t^T = (Y_t - \rho Y_{t-1})$, $X_t^T = (X_t - \rho X_{t-1})$, $Y_{t-1}^T = (Y_{t-1} - \rho Y_{t-2})$ and $\beta_0^T = \beta_0(1-\rho)$.

If ρ is known, since η_t satisfies all OLS assumptions, one can proceed to apply OLS to the transformed variables Y^T and X^T, and obtain estimators with all the optimum properties, namely, best linear unbiased estimator of the respective parameters. In effect, running Equation (3.20) is tantamount to using GLS. Regression in Equation (3.19) is known as the generalized, or quasi, difference equation. In this differencing procedure one observation is lost because the first observation has no antecedent. To avoid this loss of one observation, the first observation on Y and X is transformed as follows: $y_1\sqrt{(1-\rho^2)}$ and $x_1\sqrt{(1-\rho^2)}$. This transformation is known as the Prais-Winsten transformation.

3.4.4 Estimation of First-Order Autocorrelation

In practice, first-order autocorrelation, ρ, is rarely known. Therefore, an alternative method needs to be devised. Among several methods of estimating ρ, the Cochrane-Orcutt iterative method has become quite popular in practice (Gujarati and Sangeetha, 2007). It uses the estimated disturbances

$\hat{u}_t$ to obtain information about the unknown ρ. Cochrane and Orcutt (1949) recommended the following steps to estimate ρ:

Step 1. Fit Equation (3.12) by OLS method and obtain the residuals, $\hat{u}_t$.

Step 2. Using these residuals, run the following regression to obtain $\hat{\rho}$.

$$\hat{u}_t = \hat{\rho}\,\hat{u}_{t-1} + v_t, \tag{3.21}$$

where v_t is the error term at time t.

Step 3. Using $\hat{\rho}$ obtained from Equation (3.21), run the following generalized difference equation

$$(Y_t - \hat{\rho}\,Y_{t-1}) = \beta_0(1-\hat{\rho}) + \beta_1(X_t - \hat{\rho}\,X_{t-1}) + \beta_2(Y_{t-1} - \hat{\rho}\,Y_{t-2}) + (u_t - \hat{\rho}\,u_{t-1}).$$

or

$$y_t^T = \beta_0^T + \beta_1 x_t^T + \beta_2 y_{t-1}^T + \varepsilon_t, \tag{3.22}$$

where $y_t^T = (Y_t - \hat{\rho}\,Y_{t-1})$, $x_t^T = (X_t - \hat{\rho}\,X_{t-1})$, $y_{t-1}^T = (Y_{t-1} - \rho\,Y_{t-2})$,

$\beta_0^T = \beta_0(1-\hat{\rho})$ and $\varepsilon_t = (u_t - \hat{\rho}\,u_{t-1})$.

Step 4. Substitute the values of $\hat{\beta}_0^T = \hat{\beta}_0(1-\hat{\rho})$, $\hat{\beta}_1$, and $\hat{\beta}_2$ obtained from Equation (3.22) into the original regression Equation (3.12) and obtain the new residuals, $\hat{u}_t^{\bullet}$, as

$$\hat{u}_t^{\bullet} = Y_t - \hat{\beta}_0 - \hat{\beta}_1 X_t - \hat{\beta}_2 Y_{t-1}. \tag{3.23}$$

Step 5. Now fit this regression

$$\hat{u}_t^{\bullet} = \hat{\hat{\rho}}\,\hat{u}_{t-1}^{\bullet} + w_t,$$

where w_t is the error term at time t. The above equation is similar to Equation (3.21). Thus, $\hat{\hat{\rho}}$ is the second-round estimate of ρ.

Then, continue with the third-round estimate, and so on. This procedure is to stop carrying out iterations when the successive estimates of ρ differ by a very small amount by less than 0.01 or 0.005.

If an estimate of the unknown ρ is obtained in the first step and that estimate is used to transform the variables for estimating the generalized difference equation in the second step, all these methods of estimation culminating in two steps are known in the literature as FGLS.

3.5 One-Step-Ahead Forecasting

Suppose that at time t one wants to forecast the outcome of Y at time $t+1$, or y_{t+1}. The time period in practice could correspond to a year, a quarter, a month, a week, or even a day. Let I_t denote information that can be observed at time t. This information set includes y_t, earlier values of Y, and often other variables dated at time t or earlier. This information can be combined in innumerable ways in order to make one-step-ahead forecast of Y_{t+1}.

There is one best way, provided the loss associated with forecast error is specified. Let f_t denote the forecast of Y_{t+1} made at time t. The f_t is called a one-step-ahead forecast. The forecast error is $e_{t+1} = Y_{t+1} - f_t$, which is observed once the outcome on Y_{t+1} is observed. The most common measure of loss is the same one that leads to OLS estimation of a linear regression model: the squared error, e_{t+1}^2. But e_{t+1} is unknown at time t. Therefore, any useful criterion for choosing f_t must be based on what is known at time t. It is natural to choose the forecast to minimize the expected squared forecast error, given information set I_t:

$$E(e_{t+1}^2 | I_t) = E\left[(Y_{t+1} - f_t)^2 | I_t\right] . \tag{3.24}$$

A basic fact from probability is that $E(Y_{t+1}|I_t)$ minimizes Equation(3.24). In other words, if the expected squared forecast error given information at time t desires to be minimized, the forecast should be the expected value of Y_{t+1} given the variables that are known at time t (Wooldridge, 2009).

Types of Regression Models Used for Forecasting

There are many regression models that can be used to forecast future values of a time series. One of the regression models for time series data is the static model. The static model, which contains a single explanatory variable, is given by

$$Y_t = \beta_0 + \beta_1 X_t + u_t, \tag{3.25}$$

where β_0 and β_1 are the parameters, Y_t is the dependent variable at time t, X_t is the explanatory variables at time t, and u_t is the disturbance term at time t. Suppose that the parameters β_0 and β_1 are known. If X_{t+1} is known at time t, then the forecast of Y_{t+1} at time t

$$E(Y_{t+1}|I_t) = \beta_0 + \beta_1 X_{t+1},$$

where I_t contains $x_{t+1}, y_t, x_t, y_{t-1}, \ldots, y_1, x_1$. The kind of forecast is usually called a conditional forecast because it is conditional on knowing the value of X at time $t+1$.

Unfortunately, at any time, the values of the explanatory variables in future time periods are rarely known. If X_{t+1} is not known at time t, then

$$E(Y_{t+1}|I_t) = \beta_0 + \beta_1 E(X_{t+1}|I_t).$$

This means that in order to forecast Y_{t+1}, X_{t+1} must be forecast first, based on the same information set. This is usually called an unconditional forecast because knowledge of X_{t+1} at time t is not assumed.

For forecasting, besides the static model in Equation (3.25) a model that depends only on lagged values of Y and X can also be used. This saves the extra step of having a forecast for a right-hand side variable before forecasting Y. The kind of model is

$$Y_t = \delta_0 + \alpha_1 Y_{t-1} + \gamma_1 X_{t-1} + u_t, \tag{3.26}$$

where δ_0, α_1 and γ_1 are the parameters, Y_t is the dependent variable at time t, Y_{t-1} is the one lagged dependent variable, X_{t-1} is the explanatory variables at time t-1, and u_t is the disturbance term at time t. I_t contains Y and X dated at time $t-1$ and earlier. Now, the forecast of Y_{t+1} at time t is

$$f_t = \delta_0 + \alpha_1 Y_t + \gamma_1 X_t,$$

if the parameters are known, one can just plug in the values of Y_t and X_t. Especially, for forecasting one step ahead, such model can be very useful.

One-Step-Ahead Prediction Interval

Obtaining a forecast one period after the sample ends is relatively straightforward using models such as Equation (3.26). As usual, let n be the sample size. Assuming that the parameters have been estimated by OLS, the forecast of Y_{t+1}, denoted by $\hat{y}_{t+1}$, is

$$\hat{y}_{t+1} = \hat{\delta}_0 + \hat{\alpha}_1 Y_t + \hat{\gamma}_1 X_t . \tag{3.27}$$

The forecast error which we will not know until time $t+1$ is

$$\hat{e}_{t+1} = Y_{t+1} - \hat{y}_{t+1} .$$

The forecast $\hat{y}_{t+1}$ of Y_{t+1} is usually called a point forecast. A forecast interval can also be obtained. A forecast interval is essentially the same as a prediction interval. Though the model in Equation (3.27) contains lagged dependent variables, if the disturbance term u_t given I_{t-1} is normally distributed with zero mean and constant variance, the (1-2α) prediction interval is given by

$$\hat{y}_{t+1} - t_{\alpha,n-p-1}\, s_{t+1},\ \hat{y}_{t+1} + t_{\alpha,n-p-1}\, s_{t+1}, \tag{3.28}$$

where s_{t+1}, the standard error of forecast error, is given as

$$s_{t+1} = \left[s^2 + [s(\hat{y}_{t+1})]^2 \right]^{1/2} . \tag{3.29}$$

In Equation (3.29), s^2 is the residual mean square error for the linear regression and $s(\hat{y}_{t+1})$ is the standard error of the forecast which equals to the $(s^2\, x_f'(X'X)^{-1} x_f)^{1/2}$ where $x_f' = (1, y_t, x_t)$.

However, if the disturbance term u_t, given information set I_{t-1}, is not normally distributed, the prediction interval can be constructed using the method of bootstrap.

3.6 The Basics of Bootstrap

If the disturbance term u_t, which is entering the linear regression model of Equation (3.1), is normally distributed, the OLS estimators are not only the best unbiased estimators but also follow normal probability distributions. Furthermore, forecasts, based on the OLS fit and which are conditional on

fixed explanatory variables, are also normally distributed, even in small samples.

If the disturbances are not normally distributed but can only be assumed to be independent and identically distributed random disturbances, then the finite sample distributions of the OLS estimators and forecasts cannot be explicitly determined, and consequently the application of the methods and procedures developed under normality assumption for statistical inference would not be valid. In such cases, the empirical distributions of the OLS estimators and forecasts can be generated by one of the resampling methods, which is called bootstrap method.

The method of the bootstrap was first developed by Efron (1979). The key point made by Efron is that the observed data set is a random sample of size n drawn from the actual probability distribution which is generating the data. In a sense, the empirical distribution based on the data is the best estimate of the actual distribution from which the data have come. As such, the Empirical Distribution Function (EDF) is defined to be the discrete distribution that places a probability of $1/n$ on each of the observed values. A random sample of same size n is drawn with replacement from the empirical distribution. The sample so obtained is called the bootstrap sample and the statistic of interest is calculated from the simulated bootstrap sample. From replications of the bootstrap sample simulation, the empirical distribution of the statistic of interest is obtained (Enders, 2004). Based on this distribution, the properties of the statistic are then estimated.

A classic illustration of the power of the bootstrap is the computation of the standard error of the sample median. A sample of n independent observations is drawn from the unknown distribution. A consistent estimate of standard error of the sample median could be calculated if the probability distribution generating the samples was known. Usually, it is not known. Efron's suggestion was to use the sample data to generate an estimate of the distribution. The bootstrap method for calculating the standard error of the sample median would be (i) to create a bootstrap sample of size n by drawing

randomly and with replacement from the original sample, (ii) to calculate the median ($\hat{\theta}^{*b}$) for this bootstrap sample, and (iii) to repeat (i) and (ii) a large number of times, say, B. The probability distribution of the estimated median can then be approximated using the empirical distribution of the simulated medians of B bootstrap samples. The bootstrap standard error of the sample median can be calculated as

$$se^*(\hat{\theta}^*) = \sqrt{\frac{1}{B-1}\sum_{b=1}^{B}\left(\hat{\theta}^{*b} - \bar{\hat{\theta}}^*\right)^2},$$

where $$\bar{\hat{\theta}}^* = \frac{1}{B}\sum_{b=1}^{B}\hat{\theta}^{*b}.$$

In this section the basic bootstrap methods which are applicable to a single, homogeneous sample of data are presented. The sample values, denoted by $y_1,...,y_n$, are thought of as the outcomes of independent and identically distributed random variables $Y_1,...,Y_n$ whose Probability Density Function (PDF) and Cumulative Distribution Function (CDF) are denoted by f and F, respectively. The sample is to be used to make inferences about a population characteristic, generically denoted by θ, using a statistic $\hat{\theta}$. It is assumed that the choice of $\hat{\theta}$ has been made, that is, an estimate for θ. The major attention is focused on questions concerning the probability distribution of $\hat{\theta}$ such as its bias, its standard error, its quantiles or confidence limits for θ.

There are two situations to distinguish, the parametric and the nonparametric. When there is a particular probability model, with adjustable constants or parameters ψ that fully determine the PDF, such a model is called parametric model, and statistical methods based on this model are parametric methods. In this case the parameter of interest θ is a component of or function of ψ. When no such probability model is used, the statistical analysis is nonparametric, and it uses only the fact that the random variables Y_is are independent and identically distributed.

Parametric Simulation

When theoretical properties of $\hat{\theta}$ might be hard to determine with sufficient accuracy, the alternative of repeated simulation of datasets from a fitted parametric model is more practical.

Suppose that a particular parametric model for the distribution of the data $y_1,...,y_n$ is known. The CDF is denoted by $F_\psi(y)$. When parametric function ψ is estimated by $\hat{\psi}$, its substitution in the model gives the fitted model, with CDF $\hat{F}(y) = F_{\hat{\psi}}(y)$. The $\hat{F}$ can be used to calculate properties of $\hat{\theta}$. The symbol Y^* is used to denote the random variable which is distributed according to the fitted model $\hat{F}$, and the superscript * will be used with expectation (*E*), variance (*var*) and so forth when these moments are calculated according to the fitted distribution.

Moment Estimates

Suppose that theoretical calculation with the fitted model is too complex, approximations may not be satisfactorily available, or they may be untrustworthy, perhaps because the sample size is small (Davison and Hinkley, 1997). The alternative is to estimate the properties required, from simulated datasets. Such a dataset is denoted by $Y_1^*,...,Y_n^*$ where Y_i^* are independently sampled from the fitted distribution $\hat{F}$. When the statistic of interest θ is calculated from a simulated dataset, it is denoted by $\hat{\theta}^*$. From *B* replications of the data, simulation estimates $\hat{\theta}^{*1},...,\hat{\theta}^{*B}$ are obtained. Properties of $\hat{\theta}-\theta$ are then estimated from $\hat{\theta}^{*1},...,\hat{\theta}^{*B}$. For example, the estimator of the bias of $\hat{\theta}$ in estimating parameter θ is defined as

$$bias^*(\hat{\theta}^*) = E^*(\hat{\theta}^*) - \hat{\theta},$$

and this in turn is estimated by

$$bias^*(\hat{\theta}^*) = \frac{1}{B}\sum_{b=1}^{B}\hat{\theta}^{*b} - \hat{\theta} = \bar{\hat{\theta}}^* - \hat{\theta}.$$

The corresponding estimator of the variance of $\hat{\theta}$ is given by

$$var^*(\hat{\theta}^*) = \frac{1}{B-1}\sum_{b=1}^{B}(\hat{\theta}^{*b} - \bar{\hat{\theta}}^*)^2,$$

with similar estimators for other moments.

Distribution and Quantile Estimates

If, as is often the case, one is approximating the distribution of $\hat{\theta} - \theta$ by that of $\hat{\theta}^* - \hat{\theta}$, then cumulative probabilities are estimated by the EDF of the simulated values $\hat{\theta}^* - \hat{\theta}$. More formally, if CDFof $\hat{\theta} - \theta$, denoted by *G(u)*, is

$$G(u) = Pr(\hat{\theta} - \theta \le u),$$

then the simulation estimate of *G(u)* is

$$\hat{G}^*(u) = \frac{\#\{\hat{\theta}_b^* - \hat{\theta} \le u\}}{B}, \qquad (3.30)$$

where # represents the number.

Quantiles of the distribution of $\hat{\theta} - \theta$ are often estimated. These are approximated using ordered values of $\hat{\theta}^* - \hat{\theta}$. The p quantile of $\hat{\theta} - \theta$ is estimated by the $(B+1)p^{th}$ ordered value of $\hat{\theta}^* - \hat{\theta}$, that is $\hat{\theta}^*_{((B+1)p)} - \hat{\theta}$. It is assumed that B is chosen so that $(B+1)p$ is an integer.

The simulation approximation estimates $\hat{G}^*$ and the corresponding quantiles therefrom are in principle better than results obtained by normal approximation, provided that B is large enough, because they avoid the supposition that the distribution of $\hat{\theta}^* - \hat{\theta}$ has a particular form (Davison and Hinkley, 1997).

Nonparametric Simulation

Suppose that one is not considering a particular parametric model but that it is sensible to assume that $Y_1,...,Y_n$ are independent and identically distributed according to an unknown distribution function F. The

corresponding estimate of the unknown CDF F is the EDF $\hat{F}$, which is defined as the sample proportion

$$\hat{F}(y) = \frac{\#\{y_j \le y\}}{n}. \tag{3.31}$$

Simulation is applied with EDF which puts equal probabilities $1/n$ at each sample value y_i, each Y^* is independently sampled at random from the original sample values. Therefore, the simulated sample $Y_1^*,...,Y_n^*$ is a random sample taken with replacement from the original sample. This sampling procedure is called the nonparametric bootstrap (Davison and Hinkley, 1997).

Simple Bootstrap Confidence Intervals

The major application for distributions and quantiles of an estimator $\hat{\theta}$ lies in the calculation of confidence limits. There are several ways of using bootstrap simulation results in this context; two basic methods are described below:

If the bootstrap estimates of quantiles for $\theta - \theta$ are used, an equitailed (1-2α) confidence interval will have limits given by

$$2\hat{\theta} - \hat{\theta}^*_{((B+1)(1-\alpha))},\ 2\hat{\theta} - \hat{\theta}^*_{((B+1)\alpha)}. \tag{3.32}$$

This is based on the probability implication that

$$\Pr(a \le \hat{\theta} - \theta \le b) = 1 - 2\alpha \quad \Rightarrow \quad \Pr(\hat{\theta} - b \le \theta \le \hat{\theta} - a) = 1 - 2\alpha.$$

The limits in Equation (3.32) are referred as the basic bootstrap confidence limits. Their accuracy depends upon B and one would typically take $B \ge 1000$ to be safe (Davison and Hinkley, 1997).

The studentized bootstrap estimates of quantiles for $\hat{\theta} - \theta$ is defined as

$$z^* = \frac{\hat{\theta}^* - \hat{\theta}}{[var^*(\hat{\theta}^*)]^{1/2}}, \tag{3.33}$$

where $\hat{\theta}^*$ and $var^*(\hat{\theta}^*)$ are based on a simulated random sample, $Y_1^*,...,Y_n^*$. If the model is parametric, the Y_i^* are generated from the fitted parametric distribution, and if the model is nonparametric, they are generated from the

EDF $\hat{F}$. In either case, the $(B+1)\alpha^{th}$ order statistic of the simulated values $z_1^*,...,z_B^*$, namely $z^*_{((B+1)\alpha)}$, to estimate z_α, is used. Then the studentized bootstrap confidence interval for θ has limits

$$\hat{\theta} - [var(\hat{\theta})]^{1/2} z^*_{((B+1)(1-\alpha))},\ \hat{\theta} - [var(\hat{\theta})]^{1/2} z^*_{((B+1)\alpha)}. \tag{3.34}$$

This studentized bootstrap method is most likely to be of wide use in nonparametric problems (Davison and Hinkley, 1997).

3.7 Bootstrap Methods Involving Regression Models

In this section, the bootstrap methods, which are usually employed in linear regression model and dynamic regression model, are presented. In Subsection 3.7.1 and 3.7.2, an explanation of residual resampling bootstrapping for the least squares fit and weighted least squares fit is presented. In Subsections 3.7.3 and 3.7.4, bootstrap confidence intervals for regression parameters and bootstrap prediction intervals are provided.

3.7.1 Bootstrapping the Least Squares Fit

The residuals e_i's obtained from Equation (3.5) are modified as

$$r_i = \frac{e_i}{(1-h_i)^{1/2}}. \tag{3.35}$$

These r_i are referred as modified residuals. The conditional distribution of Y_i^* given x_i specified by the estimated version of Equation (3.3) is given by

$$Y_i^* = x_i'\hat{\beta} + \varepsilon_i^* \qquad \text{for } i = 1,...,n, \tag{3.36}$$

with ε_i^* randomly sampled with replacement from the $r_i - \bar{r}$, where $\bar{r}$ is the average of the r_i. The residual resampling bootstrapping for the least squares fit is also called the method of Bootstrap Based on OLS Estimators (BOLS). The algorithm to generate simulated datasets and corresponding estimates is as follows:

Algorithm for BOLS Method in Linear Regression

In the algorithm for BOLS method, a total of four steps needs to be followed. They are:

Step 1. Fit OLS regression to the original sample of observations to obtain OLS estimates $\hat{\beta}_{OLS}$, compute the estimated values $\hat{y}_i = f(X_i, \hat{\beta}_{OLS})$ for $i=1,...,n$, and the residuals $e_i = Y_i - \hat{y}_i$ for $i=1,...,n$, and obtain the modified residuals $r_i = \frac{e_i}{(1-h_i)^{1/2}}$ for $i=1,...,n$, where h_i is a diagonal element of the hat matrix $\hat{H}$.

Step 2. Draw a random sample ε_i^* of size n, with replacement, from the $r_1 - \bar{r},...,r_n - \bar{r}$, where $\bar{r}$ is the average of the r_i, and obtain new bootstrap values Y_i^* of Y_i where $Y_i^* = f(X_i, \hat{\beta}_{OLS}) + \varepsilon_i^*$ for $i=1,...,n$.

Step 3. Fit OLS regression again to the new bootstrap values Y_i^* obtained in ***Step2*** against the independent variables of X's to obtain bootstrap estimates $\hat{\beta}^{*b} = (X'X)^{-1}X'Y^{*b}$ and $s^{*b}(\hat{\beta}^*)$.

Step 4. Repeat ***Step2*** and ***Step3*** for B replications to obtain bootstrap estimates $\hat{\beta}^{*b}$ and $s^{*b}(\hat{\beta}^*)$ $(b = 1,...,B)$.

The advantage of resampling is to obtain an improved quantile estimation when normal-theory distributions of the estimators ($\hat{\beta}$) and residual mean square error for the linear regression (s^2) are not accurate and valid.

3.7.2 Bootstrapping the Weighted Least Squares Fit

The method of Weighted Least Squares (WLS) is just a special case of GLS. Suppose that variance-covariance matrix $\mathbf{var}(u) = k\,W^{-1}$ where W is the diagonal matrix of known case weights w_i. Then WLS estimates are

$$\hat{\beta} = (X'WX)^{-1}X'Wy, \tag{3.37}$$

the fitted values are

$$\hat{Y} = X\hat{\beta},$$

and the residual vector is

$$e = (I\text{-}H)y,$$

where the matrix H is defined by

$$H = X(X'WX)^{-1}X'W, \tag{3.38}$$

whose diagonal elements are h_i. The residual vector e has variance $var(e) = k\,(I - H)W^{-1}$, whose ith diagonal element is $k(1-h_i)w_i^{-1}$. So the modified residual is given by

$$r_i = \frac{Y_i - \hat{y}_i}{w_i^{-1/2}(1-h_i)^{1/2}}. \tag{3.39}$$

Residual resampling is defined by

$$Y_i^* = x_i'\hat{\beta} + w_i^{-1/2}\varepsilon_i^*, \tag{3.40}$$

where ε_i^* is randomly sampled from the $r_1 - \bar{r},...,r_n - \bar{r}$. It is not necessary to estimate k in applying this algorithm; but if an estimate is required, it would be $\hat{k} = (n-p-1)^{-1}\,y'W(I-H)y$.

3.7.3 Bootstrap Confidence Intervals for Regression Parameters

The $(1-2\alpha)$ basic bootstrap confidence limits for parameter β_j are given by

$$2\hat{\beta}_j - \hat{\beta}^*_{j((B+1)(1-\alpha))},\quad 2\hat{\beta}_j - \hat{\beta}^*_{j((B+1)\alpha)} \tag{3.41}$$

where the $\hat{\beta}^*_{j((B+1)\alpha)}$ and $\hat{\beta}^*_{j((B+1)(1-\alpha))}$ are the α and $(1-\alpha)$ empirical quantiles of the $\hat{\beta}_j^*$s, whose ordered values are denoted by $\hat{\beta}^*_{j(1)} \le ... \le \hat{\beta}^*_{j(B)}$.

A modification of this is to use the form of the normal approximation confidence limits, but it would be required to replace the standard normal approximation for $Z = \dfrac{\hat{\beta}_j - \beta_j}{s(\hat{\beta}_j)}$ by a bootstrap approximation. Each simulated sample is used to calculate $\hat{\beta}_j^*$, the standard error estimate $s^*(\hat{\beta}_j^*)$, and hence the bootstrap version $z^* = \dfrac{\hat{\beta}_j^* - \hat{\beta}_j}{s^*(\hat{\beta}_j^*)}$ of Z. The B simulated values of z^* are

ordered and the p quantile of Z is estimated by the $(B+1)p^{th}$ of these. Then the studentized bootstrap confidence limits would become

$$\hat{\beta}_j - s(\hat{\beta}_j)z^*_{((B+1)(1-\alpha))},\ \hat{\beta}_j - s(\hat{\beta}_j)z^*_{((B+1)\alpha)}. \tag{3.42}$$

In principle, this method is superior to the previous basic method. For the applying of both equations; Equation (3.41) and Equation (3.42), it is necessary that $(B+1)\alpha$ should be an integer (Davison and Hinkley, 1997).

3.7.4 Bootstrap Prediction Intervals

A fitted linear regression is often used for prediction of a new individual response Y_f when the explanatory variable vector is equal to x_f. Confidence limits for the response Y_f itself — usually called prediction limits — require additional resampling to simulate the variation of Y_f about the mean response $x'_f\beta$. The quantity to be predicted is

$$Y_f = x'_f\beta + \varepsilon_f,$$

and the point predictor is

$$\hat{y}_f = x'_f\hat{\beta}.$$

The random disturbance ε_f is assumed to be independent of the random disturbances $\varepsilon_1,...,\varepsilon_n$ in the observed responses, and for the sake of simplicity it is assumed that they all come from the same distribution: in particular, the disturbances have equal variances.

To assess the accuracy of the point predictor, the distribution of the prediction error

$$\delta = \hat{y}_f - Y_f = x'_f\hat{\beta} - (x'_f\beta + \varepsilon_f)$$

can be estimated by the distribution of

$$\delta^* = x'_f\hat{\beta}^* - (x'_f\hat{\beta} + \varepsilon^*_f), \tag{3.43}$$

where ε^*_f is sampled from $r_1 - \bar{r},...,r_n - \bar{r}$ and $\hat{\beta}^*$ is a simulated vector of estimates from the residual resampling algorithm. This assumes homoscedasticity of random disturbance. Unconditional properties of the

prediction error correspond to averaging over the distributions of both ε_f and the estimate $\hat{\beta}$, which are done in the simulation by repeating Equation (3.43) for each set of values of $\hat{\beta}^*$. Having obtained the modified residuals r_i from the data fit, the algorithm to generate B sets; each with only one step prediction is as follows:

Algorithm for Prediction in Linear Regression

In the algorithm for prediction in linear regression, a total of five steps needs to be followed. They are:

Step 1. Fit OLS regression to the original sample of observations to obtain OLS estimates $\hat{\beta}_{OLS}$, compute the estimated values $\hat{y}_i = f(X_i, \hat{\beta}_{OLS})$ for $i=1,...,n$, and the residuals $e_i = Y_i - \hat{y}_i$ for $i=1,...,n$, and obtain the modified residuals $r_i = \dfrac{e_i}{(1-h_i)^{1/2}}$ for $i=1,...,n$, where h_i is a diagonal element of the hat matrix $\hat{H}$.

Step 2. Draw a random sample ε_i^* of size n, with replacement, from the $r_1 - \bar{r},...,r_n - \bar{r}$, where $\bar{r}$ is the average of the r_i, and obtain new bootstrap values Y_i^* of Y_i where $Y_i^* = f(X_i, \hat{\beta}_{OLS}) + \varepsilon_i^*$ for $i=1,...,n$.

Step 3. Fit OLS regression again to the new bootstrap values Y_i^* obtained in ***Step2*** against the independent variables of X's to obtain bootstrap estimates $\hat{\beta}^{*b} = (X'X)^{-1}X'Y^{*b}$ and s^{*b}.

Step 4. Draw ε_f^* randomly from $r_1 - \bar{r},...,r_n - \bar{r}$, and compute prediction error $\delta^{*b} = x_f'\hat{\beta}^{*b} - (x_f'\hat{\beta}_{OLS} + \varepsilon_f^{*b})$.

Step 5. Repeat ***Step2*** to ***Step4*** for B replications to obtain δ^{*b} $(b = 1,...,B)$.

If predictions at several values of x_f are required, then only ***Step4*** of the algorithm needs to be repeated for each x_f.

The $(1-2\alpha)$ basic bootstrap prediction limits for Y_f are

$$\hat{y}_f - \delta^*_{((B+1)(1-\alpha))},\ \hat{y}_f - \delta^*_{((B+1)\alpha)} \tag{3.44}$$

where $\hat{y}_f = x'_f\hat{\beta}_{OLS}$, and $\delta^*_{((B+1)\alpha)}$ and $\delta^*_{((B+1)(1-\alpha))}$ are the α and $(1-\alpha)$ empirical quantiles of the δ^*s, whose ordered values are denoted by $\delta^*_{(1)} \le \ldots \le \delta^*_{(B)}$. This is analogous to the basic bootstrap method for confidence intervals.

A somewhat better approach, which is analogous to the standard normal-theory analysis, is to work with studentized prediction error

$$Z = \frac{\hat{y}_f - Y_f}{s},$$

where s is the square root of residual mean square error for the linear regression. The corresponding simulated values are given by

$$z^{*b} = \frac{\delta^{*b}}{s^{*b}},$$

with s^{*b} calculated in *Step3* of the algorithm. The α and $(1-\alpha)$ quantiles of Z are estimated by $z^*_{((B+1)\alpha)}$ and $z^*_{((B+1)(1-\alpha))}$ respectively, where $z^*_{(1)} \le \ldots \le z^*_{(B)}$ are the ordered value of all B z^*s. Then the studentized bootstrap prediction interval for Y_f is given by

$$\hat{y}_f - s_f z^*_{((B+1)(1-\alpha))},\ \hat{y}_f - s_f z^*_{((B+1)\alpha)}. \tag{3.45}$$

3.8 Specific Algorithms for Bootstrapping in Regression Models

Some major steps to be followed in the proposed residual resampling bootstrapping in dynamic regression model are based on some of the steps in two algorithms for residual resampling bootstrapping applied by Prescott and Stengos (1987) and Bernard and Veall (1987). Therefore, algorithm for bootstrapping in dynamic regression model demonstrated by Precott and Stengos (1987), and algorithm for bootstrapping in linear regression model with AR(1) disturbances as proposed by Bernard and Veall (1987) are presented in this Section 3.8.

Prescott and Stengos (1987) demonstrated how the bootstrap method can be applied to the construction of confidence intervals for the forecasts of pork production generated by a dynamic econometric model of pork supply when exogenous variables are stochastic during the period chosen for making the forecasts. They presented the following two equations:

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 X_{t-2} + u_t, \tag{3.46}$$

where β_0, β_1 and β_2 are the parameters, Y_t is the dependent variable at time t, Y_{t-1} is the one lagged dependent variable, X_{t-2} is the explanatory variables at time t-2, and u_t is the disturbance term at time t.

$$X_t = \gamma_0 + \gamma_1 X_{t-1} + v_t, \tag{3.47}$$

where γ_0 and γ_1 are the parameters, X_t is the dependent variable at time t, X_{t-1} is the one lagged dependent variable, and v_t is the disturbance term at time t. One of the interesting point in the work done by Prescott and Stengos (1987) was the construction of confidence interval for Y_f, a forecast of future pork production. To this end, it is assumed that initial values y_0, x_0 and x_{-1} are available.

In the context of Prescott and Stengos (1987), a forecast of pork production at some time in the future f can be specified and determined as follows:

Step 1. Estimate β_0, β_1 and β_2 by OLS and compute OLS estimates of γ_0 and γ_1.

Step 2. Predict X_f as

$$\hat{x}_f = \hat{\gamma}_0 + \hat{\gamma}_1 \hat{x}_{f-1}.$$

Forecast y_f with the equation

$$\hat{y}_f = \hat{\beta}_0 + \hat{\beta}_1 \hat{y}_{f-1} + \hat{\beta}_2 \hat{x}_{f-2}.$$

The residual resampling bootstrapping is implemented on this process in the following algorithm:

Algorithm for Bootstrapping in Dynamic Regression Model

Step 1. Compute estimates $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\gamma}_0$ and $\hat{\gamma}_1$.

Step 2. Construct *B* bootstrap samples:

(a) Draw a random sample of size $n+m$ with replacement from the set of residuals $\hat{v}_t$. Let the resampled residuals be $\hat{v}_t^*$.

(b) The artificial sample X_t^* is then developed as

$$X_t^* = \hat{\gamma}_0 + \hat{\gamma}_1 X_{t-1} + v_t^* \qquad \text{for } t = 2,\ldots,n,$$

where $X_1^* = \hat{\gamma}_0 + \hat{\gamma}_1 x_0 + v_1^*$ and

$$X_f^* = \hat{\gamma}_0 + \hat{\gamma}_1 X_{f-1}^* + v_f^* \qquad \text{for } f = n+2,\ldots,n+m,$$

where $X_{n+1}^* = \hat{\gamma}_0 + \hat{\gamma}_1 x_n + v_{n+1}^*$.

(c) Draw a random sample of size $n+m$ with replacement from the set of residuals $\hat{u}_t$. Let the resampled residuals be $\hat{u}_t^*$.

(d) The artificial sample Y_t^* is then developed as

$$Y_t^* = \hat{\beta}_0 + \hat{\beta}_1 Y_{t-1} + \hat{\beta}_2 X_{t-2} + \hat{u}_t^* \qquad \text{for } t = 2,\ldots,n,$$

where $Y_1^* = \hat{\beta}_0 + \hat{\beta}_1 y_0 + \hat{\beta}_2 x_{-1} + \hat{u}_1^*$ and

$$Y_f^* = \hat{\beta}_0 + \hat{\beta}_1 Y_{f-1}^* + \hat{\beta}_2 X_{f-2}^* + \hat{u}_f^* \qquad \text{for } f = n+2,\ldots,n+m,$$

where $Y_{n+1}^* = \hat{\beta}_0 + \hat{\beta}_1 y_n + \hat{\beta}_2 x_{n-1} + \hat{u}_{n+1}^*$.

Step 3. Using the artificial data (periods $t=1,\ldots,n$), $\hat{\gamma}_0^*$, $\hat{\gamma}_1^*$, $\hat{\beta}_0^*$, $\hat{\beta}_1^*$, and $\hat{\beta}_2^*$ are reestimated by OLS and sets of forecasts for $\hat{x}_f^b$ and $\hat{y}_f^b$ are recalculated.

$$\hat{x}_f^b = \hat{\gamma}_0^{*b} + \hat{\gamma}_1^{*b} \hat{x}_{f-1}^b \qquad \text{for } f = n+2,\ldots,n+m-2 \text{ and } b = 1,\ldots,B,$$

where $\hat{x}_{n+1}^b = \hat{\gamma}_0^{*b} + \hat{\gamma}_1^{*b} x_n$.

$$\hat{y}_f^b = \hat{\beta}_0^{*b} + \hat{\beta}_1^{*b} \hat{y}_{f-1}^b + \hat{\beta}_2^{*b} \hat{x}_{f-2}^b \quad \text{for } f = n+2,\ldots,n+m \text{ and } b=1,\ldots,B,$$

where $\hat{y}_{n+1}^b = \hat{\beta}_0^{*b} + \hat{\beta}_1^{*b} y_n + \hat{\beta}_2^{*b} x_{n-1}$.

The superscript *b* denotes that the estimate is taken from the b^{th} bootstrap sample.

Step 4. For period f, the simulated forecast error is computed as

$$Y_f^{*b} - \hat{y}_f^b \qquad \text{for } f = n+1,\ldots,n+m \text{ and } b=1,\ldots,B.$$

Bernard and Veall (1987) exploited the usefulness of the bootstrap for time series applications in which the disturbance term has a (known) time series correlation (Johnston and DiNardo, 1997). They applied the residual resampling bootstrapping to the construction of a confidence interval for a forecast of electricity demand Y in Quebec Province of Canada at some future point in time. They settled on the following two-equation system:

$$Y_t = \beta_0 + \beta_1 X_t + u_t, \tag{3.48}$$

where β_0 and β_1 are the parameters, Y_t is the dependent variable at time t, X_t is the explanatory variables at time t, and u_t is the disturbance term at time t.

$$X_t = \gamma Z_t + v_t, \tag{3.49}$$

where γ is the parameter, X_t is the dependent variable at time t, Z_t is the explanatory variables at time t, and v_t follows a AR(1) process given by

$$v_t = \rho v_{t-1} + \eta_t. \tag{3.50}$$

Bernard and Veall (1987) were interested in constructing confidence intervals for Y_f, a forecast of future electricity demand. For this purpose it is assumed that future values of Z are known.

In the context of electricity demand by Bernard and Veall (1987), a forecast of electricity demand at some time in the future f can be determined as follows:

Step 1. Estimate β_0 and β_1 by OLS and compute an OLS estimate of γ. Use the residuals from OLS estimation of Equation (3.49) to calculate $\hat{\rho}$. Calculate $\hat{\gamma}_{GLS}$ using the estimate $\hat{\rho}$.

Step 2. Predict X_f as

$$\hat{x}_f = \hat{\gamma}_{GLS} Z_f,$$

for a given set of values of Z_f. Forecast $\hat{y}_f$ with the equation

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

The residual resampling bootstrapping is implemented on this process in the following algorithm:

Algorithm for Bootstrapping in Linear Regression Model with AR(1) Disturbances

Step 1. Compute estimates $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\rho}$ and $\hat{\gamma}_{GLS}$.

Step 2. Construct B bootstrap samples:

(a) Draw a random sample $\hat{\eta}_t^*$ of size n with replacement from the set of residuals $\hat{\eta}$. Take the first element of $\hat{\eta}^*$ and divide by $\sqrt{1-\hat{\rho}^2}$ to yield $\hat{v}_1^*$. Construct the remaining elements by:

$$\hat{v}_t^* = \hat{\rho}\hat{v}_{t-1}^* + \hat{\eta}_t^* \qquad \text{for } t = 2,\ldots,n.$$

(b) The artificial sample of X_t is then developed as

$$X_t^* = \hat{\gamma}_{GLS} Z_t + \hat{v}_t^* \qquad \text{for } t = 1,\ldots,n,$$

where $\hat{\gamma}_{GLS}$ is the original GLS estimate.

(c) Draw a random sample $\hat{u}_t^*$ of size n with replacement from the set of residuals $\hat{u}$.

(d) The bootstrap sample is then completed by using

$$Y_t^* = \hat{\beta}_0 + \hat{\beta}_1 X_t^* + \hat{u}_t^* \qquad \text{for } t = 1,\ldots,n,$$

where $\hat{\beta}_0$, and $\hat{\beta}_1$ are the original OLS estimates.

Step 3. With the bootstrap samples in hand, $\hat{\gamma}_{GLS}^{*b}$, $\hat{\beta}_0^*$ and $\hat{\beta}_1^*$ are reestimated. One can calculate $\hat{x}_f^b$:

$$\hat{x}_f^b = \hat{\gamma}_{GLS}^{*b} Z_f \qquad \text{for } f = n+1,\ n+2,\ldots \text{ and } b=1,\ldots,B,$$

where the superscript b denotes that the estimate is taken from the b^{th} bootstrap sample. This b^{th} bootstrap estimate, $\hat{x}_f^b$, along with the corresponding bootstrap estimates $\hat{\beta}_0^*$ and $\hat{\beta}_1^*$, can be used to construct an estimate of $\hat{y}_f^b$:

$$\hat{y}_f^b = \hat{\beta}_0^{*b} + \hat{\beta}_1^{*b}\hat{x}_f^b \qquad \text{for } f = n+1,\ n+2,\ldots \text{ and } b=1,\ldots,B.$$

CHAPTER IV
RESISTANT BOOTSTRAP BASED ON OLS ESTIMATORS

In this Chapter IV, an alternative bootstrap method, that provides reliable bootstrap distributions of the regression estimates in linear regression model whenever the dataset is contaminated with outliers, is developed and introduced. The performance of the proposed bootstrap method is evaluated through simulations.

4.1 Resistant Bootstrap Based on OLS Estimators in Linear Regression

One of the conventional assumptions usually adopted and introduced in a linear regression model is that disturbances are independently and identically distributed normal variates with mean zero and finite variance. In such a case, the OLS estimators of regression parameters are also normally distributed. If the disturbances are not normally distributed, but assumed to be independently and identically distributed random disturbances, the OLS estimators are asymptotically normal. However, in the case of finite samples, the sampling distributions of the OLS estimators, under nonnormality assumption of disturbances, cannot be explicitly determined. An alternative approach is to use bootstrap method instead of formal and standard methods under the assumption of normally distributed disturbances.

However, the bootstrap distribution is a very poor estimator of the distribution of the OLS estimate when a dataset is contaminated with outliers because the OLS estimates are very sensitive to outliers. For analyzing datasets that are contaminated with outliers, a robust regression method can be used to obtain stable regression estimates. The standard error of the robust regression estimate can be estimated using its asymptotic variance. The asymptotic distribution of the robust regression estimate has been mainly investigated under the normal model, which does not hold in most practical situations; in such cases of non-normality, robust methods are to be highly recommended and inevitably resorted to. Since many datasets with outliers do

not satisfy the symmetry assumption, the calculation of the asymptotic distribution of the robust regression estimate for asymmetric disturbances becomes involved. The sampling distribution of a particular robust regression estimate and its standard error can also be estimated using the bootstrap method.

However, two problems usually arise when bootstrap method is used to estimate the distribution of the robust regression estimate. Firstly, the bootstrap distribution might be a very poor estimator of the distribution of the robust regression estimate because the proportion of outliers in the bootstrap samples can be higher than that in the original dataset. This problem arises because outlying and non-outlying observations have the same chance of being present in the bootstrap samples. In particular, a certain proportion of the re-calculated values of robust regression estimates may be heavily influenced by the outlier in the data. Thus, the outliers can heavily affect the tails of the bootstrap distribution, which are of much concern and interest when building confidence intervals for the unknown regression parameters in practice.

And then, secondly, it may not be practically feasible to obtain a few thousand re-calculated robust regression estimates for moderately high dimensional problems in which there are many independent variables in the regression model. The number of bootstrap samples, which are needed to obtain reliable distribution estimates, grows with the dimension of the statistic, and it makes the problem even more computationally intensive and complicated to solve. Therefore, robust bootstrap methods in linear regression, that could overcome the above two problems, have been investigated by many researchers.

Perhaps, the most important point is that gross outliers should be removed before undertaking final regression analysis, including resampling. There are two reasons for this. The first reason is that methods that are resistant to outliers are usually not very efficient, and they may behave badly under resampling. The second reason is that outliers can be disruptive to resampling analysis based on the methods that are not resistant to outliers. For

residual resampling, the distribution of disturbances will be contaminated and the outliers can then occur in resampling at any covariates values (Davison and Hinkley, 1997).

For datasets with multiple outliers, diagnosis is done by a robust regression method that is highly resistant to the effects of outliers. One preferred resistant method is LTS estimation method. The fit itself is not very efficient, and should best be thought of as an initial step in a more efficient analysis (Davison and Hinkley, 1997).

In this Section 4.1 an alternative bootstrap method which is not only computationally simple but also resistant to the effects of outliers in linear regression is proposed. This method provides reliable distributions of the regression estimates for the linear regression model in Equation (3.1):

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_p X_{ip} + u_i \qquad \text{for } i = 1,\ldots,n.$$

The idea for obtaining reliable bootstrap distributions is based on the technique of finding OLS estimates which are not affected by outliers not only in the original sample but also in the bootstrap replications. From the original sample of cases, LTS weight for each case is computed by taking h, which is defined in Section 3.3.2 of Chapter III, $\frac{3n+p+1}{4}$. The cases having zero LTS weights as outlying cases are removed. Then, BOLS method is used for the rest of the cases. This method hence forth will be called Resistant Bootstrap Based on OLS Estimators (RBOLS). The algorithm for RBOLS method to generate simulated datasets and corresponding estimates is as follows:

Algorithm for RBOLS Method in Linear Regression

In the algorithm for RBOLS method, a total of five steps needs to be followed. They are:

Step 1: Fit LTS regression to the original sample of cases to obtain LTS weight for each of the original cases, and remove the cases having zero LTS weights from the dataset.

Step 2: Fit OLS regression to the rest of the cases (say n') left from ***Step1*** to obtain regression estimates. These estimates will be called OLS Based on LTS Weights (OLS-LTS) estimates, denoted by $\hat{\beta}_{OLS\text{-}LTS}$, from now on. Then, compute the estimated values $\hat{y}_i$ given by

$$\hat{y}_i = f(X_i, \hat{\beta}_{OLS\text{-}LTS}) \qquad \text{for } i = 1,\ldots, n',$$

and obtain the residuals r_i given by

$$r_i = Y_i - \hat{y}_i \qquad \text{for } i = 1,\ldots, n'.$$

Step 3: Draw a random sample ε_i^* of size n' from the values of r_i from ***Step2*** with replacement and obtain new bootstrap values Y_i^* of Y_i where

$$Y_i^* = f(X_i, \hat{\beta}_{OLS\text{-}LTS}) + \varepsilon_i^* \quad \text{for } i = 1,\ldots, n'.$$

Step 4: Fit OLS regression again to the new bootstrap values Y_i^* obtained in ***Step3*** on the independent variables of X's to obtain the bootstrap estimate $\hat{\beta}^{*b}$, $s^{*b}(\hat{\beta}^*)$ and s^{*b}, respectively.

Step 5: Repeat ***Step3*** and ***Step4*** for B times of bootstrap replications to obtain bootstrap estimates $\hat{\beta}^{*b}$, $s^{*b}(\hat{\beta}^*)$ and s^{*b} $(b = 1,\ldots,B)$.

4.2 Assessment of the Bootstrap Methods

The performances of the bootstrap methods are evaluated based on the bootstrap root mean squared error ($rmse^*$). The smaller the $rmse^*$, the better the method. The $rmse^*$ can be computed using the following statistics and formulae.

The mean of the bootstrap distribution, $\bar{\hat{\beta}}^*$, in this study is given by

$$\bar{\hat{\beta}}^* = \frac{1}{B}\sum_{b=1}^{B} \hat{\beta}^{*b}, \tag{4.1}$$

and the bootstrap estimate of bias of $\hat{\beta}^*$, $bias^*(\hat{\beta}^*)$, is estimated by

$$bias^*(\hat{\beta}^*) = \bar{\hat{\beta}}^* - \hat{\beta}, \tag{4.2}$$

where $\hat{\beta}$ is an estimate of the parameter β obtained by a specific estimation method using the original dataset. The bootstrap estimated standard error of $\hat{\beta}^*$, $se^*(\hat{\beta}^*)$, is given by

$$se^*(\hat{\beta}^*) = \sqrt{\frac{1}{B-1}\sum_{b=1}^{B}\left(\hat{\beta}^{*b} - \bar{\hat{\beta}}^*\right)^2}. \tag{4.3}$$

The bootstrap root mean squared error of $\hat{\beta}^*$, $rmse^*(\hat{\beta}^*)$, is given by

$$rmse^*(\hat{\beta}^*) = \sqrt{\left(bias^*(\hat{\beta}^*)\right)^2 + \left(se^*(\hat{\beta}^*)\right)^2}. \tag{4.4}$$

4.3 Simulation Results for a Three-Variable Regression Model

A simulation study is carried out to illustrate the performance of RBOLS method compared with that of BOLS method. Computations are done by using S-PLUS software. In the simulation study, observations on the dependent variable Y are generated by the same linear regression model used by Riadth *et al.* (2002) in their simulation study given below:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + u, \tag{4.5}$$

where $\beta_0 = 2$, $\beta_1 = 0.7$, and $\beta_2 = 0.5$.

For the above linear regression model, the distribution used for generating data for the independent variable X_1 is $X_1 \sim N(0.6,25)$; for X_2 is $X_2 \sim N(-0.1,0.81)$; and for the disturbance term u is $u \sim N(0,0.04)$. Outliers are introduced into generated datasets as contaminants. Some good disturbances are deliberately deleted, and they are replaced with bad disturbances. The contaminated bad disturbances are generated using the distribution $N(10,9)$, that is, $u_{bad} \sim N(10,9)$, as was also used by Raidth *et al.* (2002).

Discussion on OLS Estimates and OLS-LTS Estimates

The estimates of the parameters β_0, β_1 and β_2 in Equation (4.5), their estimated root mean squared errors (*rmses*) and coefficient of determination (R^2) are computed using OLS method and OLS-LTS method for various

sample sizes having different percentages of outliers. The simulation results are presented in Table (4.1) — Table (4.4).

Table (4.1)

Estimates, *rmse*s and R^2 by OLS Method and OLS-LTS Method for Three-Variable Regression Model When n=30

Percentage of Outliers	Estimation Method	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	*rmse* $(\hat{\beta}_0)$	*rmse* $(\hat{\beta}_1)$	*rmse* $(\hat{\beta}_2)$	R^2
0%	OLS	1.9885**	0.6928**	0.4673**	0.0379	0.0105	0.0514	0.9968
	OLS-LTS	1.9565**	0.6958**	0.4749**	0.0542	0.0079	0.0427	0.9978
5%	OLS	2.8868**	0.5866**	1.2317	1.0832	0.1739	1.0018	0.4696
	OLS-LTS	1.9966**	0.6918**	0.4745**	0.0386	0.0115	0.0500	0.9968
10%	OLS	3.2951**	0.5354**	1.3337	1.4779	0.2233	1.1439	0.3742
	OLS-LTS	1.9796**	0.6940**	0.4686**	0.0421	0.0097	0.0511	0.9973
15%	OLS	3.7192**	0.6021**	1.7595	1.8892	0.1928	1.5261	0.4016
	OLS-LTS	1.9900**	0.6916**	0.4635**	0.0408	0.0126	0.0582	0.9960
20%	OLS	3.9466**	0.6382**	1.7789	2.1094	0.1830	1.5603	0.4064
	OLS-LTS	1.9973**	0.6932**	0.4668**	0.0395	0.0115	0.0557	0.9963

** significant at 1% level

Table (4.2)

Estimates, *rmse*s and R^2 by OLS Method and OLS-LTS Method for Three-Variable Regression Model When n=60

Percentage of Outliers	Estimation Method	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	*rmse* $(\hat{\beta}_0)$	*rmse* $(\hat{\beta}_1)$	*rmse* $(\hat{\beta}_2)$	R^2
0%	OLS	2.0114**	0.6976**	0.4687**	0.0262	0.0058	0.0425	0.9972
	OLS-LTS	2.0200**	0.6974**	0.4722**	0.0300	0.0055	0.0387	0.9976
5%	OLS	2.6041**	0.6303**	1.0688*	0.6965	0.1041	0.7086	0.6127
	OLS-LTS	2.0182**	0.6976**	0.4735**	0.0288	0.0055	0.0380	0.9978
10%	OLS	3.0058**	0.6927**	1.2747*	1.0844	0.0906	0.9189	0.5889
	OLS-LTS	2.0287**	0.6988**	0.4766**	0.0365	0.0052	0.0361	0.9976
15%	OLS	3.2662**	0.7167**	1.1508*	1.3368	0.0969	0.8347	0.5686
	OLS-LTS	2.0235**	0.6980**	0.4780**	0.0330	0.0057	0.0353	0.9976
20%	OLS	3.8677**	0.7662**	1.1868	1.9430	0.1365	0.9479	0.4885
	OLS-LTS	2.0245**	0.6975**	0.4796**	0.0345	0.0060	0.0348	0.9976

** significant at 1% level * significant at 5% level

Table (4.3)

Estimates, *rmse*s and R^2 by OLS Method and OLS-LTS Method for Three-Variable Regression Model When n=100

Percentage of Outliers	Estimation Method	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	*rmse* $(\hat{\beta}_0)$	*rmse* $(\hat{\beta}_1)$	*rmse* $(\hat{\beta}_2)$	R^2
0%	OLS	2.0119**	0.6946**	0.4793**	0.0226	0.0068	0.0291	0.9966
	OLS-LTS	2.0177**	0.6943**	0.4821**	0.0257	0.0070	0.0267	0.9969
5%	OLS	2.6238**	0.6782**	0.9922**	0.6688	0.0573	0.5558	0.6585
	OLS-LTS	2.0293**	0.6922**	0.5035**	0.0343	0.0087	0.0195	0.9973
10%	OLS	2.9871**	0.7004**	0.9754**	1.0326	0.0666	0.5755	0.5634
	OLS-LTS	2.0231**	0.6931**	0.4980**	0.0296	0.0080	0.0196	0.9972
15%	OLS	3.4475**	0.7635**	1.1443**	1.4918	0.1016	0.7513	0.5229
	OLS-LTS	2.0289**	0.6938**	0.4999**	0.0346	0.0075	0.0196	0.9973
20%	OLS	4.0206**	0.7296**	1.6446**	2.0599	0.0928	1.2222	0.4779
	OLS-LTS	2.0426**	0.6929**	0.5064**	0.0472	0.0083	0.0219	0.9971

** significant at 1% level

Table (4.4)

Estimates, *rmse*s and R^2 by OLS Method and OLS-LTS Method for Three-Variable Regression Model When n=200

Percentage of Outliers	Estimation Method	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	*rmse* $(\hat{\beta}_0)$	*rmse* $(\hat{\beta}_1)$	*rmse* $(\hat{\beta}_2)$	R^2
0%	OLS	2.0081**	0.7018**	0.5037**	0.0169	0.0036	0.0170	0.9963
	OLS-LTS	2.0059**	0.7010**	0.5058**	0.0157	0.0032	0.0173	0.9964
5%	OLS	2.4811**	0.6578**	0.7279**	0.5018	0.0518	0.2783	0.7265
	OLS-LTS	2.0087**	0.6997**	0.5103**	0.0175	0.0032	0.0197	0.9964
10%	OLS	2.9570**	0.6467**	0.8281**	0.9785	0.0685	0.4001	0.5604
	OLS-LTS	2.0076**	0.7008**	0.5114**	0.0175	0.0034	0.0209	0.9964
15%	OLS	3.4708**	0.6187**	0.6558*	1.4925	0.0973	0.3241	0.4251
	OLS-LTS	2.0040**	0.7018**	0.5142**	0.0166	0.0038	0.0227	0.9963
20%	OLS	3.9563**	0.6497**	1.0091**	1.9770	0.0784	0.6012	0.4055
	OLS-LTS	2.0065**	0.7013**	0.5110**	0.0180	0.0037	0.0216	0.9964

** significant at 1% level * significant at 5% level

In Table (4.1) for n = 30, at 0% level of outlier percentage, OLS estimates of parameters β_0, β_1 and β_2 are 1.9885, 0.6928 and 0.4673, respectively, while the corresponding OLS-LTS estimates are 1.9565, 0.6958 and 0.4749, respectively. The estimates provided by both estimation methods are not significantly different from each other. Comparing the *rmse*s of the estimates under the two methods of estimation, neither method is found to provide smaller *rmse*s of the estimates. The *rmse*s of OLS estimates are 0.0379, 0.0105 and 0.0514 while the *rmse*s of OLS-LTS estimates are 0.0542, 0.0079 and 0.0427. On comparison of R^2 under the two methods of estimation, the values of R^2 provided by both estimation methods are also found to be almost the same. For these reasons, it can be concluded that OLS-LTS method provides almost equally good estimates of parameters β_0, β_1 and β_2, so does OLS method in the case of no outliers in the dataset.

At 5% level of outlier percentage, OLS estimate of constant term β_0 is found to be 2.8868 while OLS-LTS estimate of β_0 is 1.9966. Comparing the two estimates, OLS estimate of 2.8868 is conspicuously different from the actual parameter value of β_0, which is 2.0; OLS-LTS estimate of 1.9966 is found to be very near to actual value of β_0. When it comes to the estimates of regression coefficient β_1, OLS method provides the estimate of 0.5866 while OLS-LTS method gives the estimate of 0.6918, which is a little more closer to the actual value 0.7 of β_1 than OLS estimate. It is the same situation as in the above case of the two estimates of constant term β_0. When moving on to the estimates of regression coefficient β_2 through the two methods of estimation, OLS estimate of β_2 is found to be 1.2317, which diverges more conspicuously from the actual value 0.5 of β_2. However, OLS-LTS estimate of β_2 is 0.4745, which is more or less closer to the actual value 0.5 of β_2 than OLS estimate of 1.2317. These results indicate that the biases of OLS-LTS estimates are smaller than those of OLS estimates; that is, OLS-LTS estimates are more desirable than OLS estimates from the standpoint of biases of the estimates.

Comparing the *rmse*s of the estimates of constant term β_0 under the two methods of estimation at the same 5% level of outlier percentage, the *rmse* of OLS estimate is found to be 1.0832 while the *rmse* of OLS-LTS estimate is 0.0386, which is much smaller than that of OLS estimate. The same situations are found to be the case when it comes to the *rmse*s of the estimates of two regression coefficients β_1 and β_2 under the two methods of estimation. On comparison of R^2 under both estimation methods, R^2 value of 0.9966 provided by OLS-LTS method is much larger than R^2 value of 0.4696 provided by OLS method. These results lead to the conclusion that OLS-LTS estimates are far more efficient and desirable than OLS estimates if the data are contaminated with outliers.

At 10% level of outlier percentage, OLS estimate of β_0 becomes 3.2951. However, OLS-LTS estimate of β_0 turns out to be 1.9796 which is much closer to the actual value 2.0 of β_0. Similarly, OLS-LTS estimate 0.6940 of β_1 is also very near to the actual value 0.7 of β_1 than OLS estimate 0.5354 of β_1. As regards the estimation of β_2, OLS-LTS estimate of 0.4686 is also found to be closer to the actual value 0.5 of β_2 than OLS estimate of 1.3337. When confining one's attention to the *rmse*s of the estimates at 10% level of outlier percentage, it is fortunately and satisfactorily found that the *rmse*s of OLS-LTS estimates of β_0, β_1 and β_2 are always much smaller than those of OLS estimates. When evaluating the value of R^2 under both estimation methods, R^2 provided by OLS-LTS method is found to be much larger than R^2 provided by OLS method. From the above results, it can be concluded that OLS-LTS method is more efficient than and preferable to OLS method at 10% level of outlier percentage.

In the case of 15% and 20% levels of outlier percentage, the same scenario of performance and behaviour of OLS estimates and OLS-LTS estimates of constant term β_0 and other regression coefficients β_1 and β_2 is captured from the standpoints of the bias and *rmse* of the corresponding estimates and R^2. It leads to the conclusion that OLS-LTS method, in the

presence of outliers in a given dataset, is satisfactorily more efficient than OLS method in the sense that the biases as well as *rmse*s of the corresponding OLS-LTS estimates of the parameters are almost always smaller than those of the corresponding OLS estimates and R^2 provided by OLS-LTS method is always larger than that provided by OLS method at 15% and 20% levels of outlier percentage, respectively.

Turning one's attention to the results in Table (4.2) through Table (4.4), at 0% level of outlier percentage for various sample sizes, OLS-LTS estimates of parameters β_0, β_1 and β_2 are found to be significantly no different from OLS estimates. Comparing the *rmse*s of the estimate under both estimation methods, it is found that there is no estimation method which always provides the smaller *rmse*s of the estimates. On comparison of R^2 under the two methods of estimation, the values of R^2 provided by both estimation methods are also found to be almost the same, indicating almost the same goodness of fit of estimation methods. For these findings, it can be concluded that OLS-LTS method can be used instead of OLS method for the dataset with no outliers. However, OLS-LTS method is desirably more efficient than OLS method on the existence of outliers in a given dataset at different percentages of outliers apart from 0% under study.

Finally, based on the results in Table (4.1) — Table (4.4), it is concluded that, if there is no outlier in a dataset, OLS-LTS estimates are as desirable as OLS estimates. It is also concluded that, when there exists outliers in a dataset, OLS-LTS estimates are always much better than OLS estimates. It indicates, in case of outliers in a dataset, that (i) OLS-LTS method is more appropriate and preferable than OLS method and (ii) one should resort to OLS-LTS method rather than OLS method so as to obtain less biased and efficient regression estimates.

Comparison of RBOLS Estimates with BOLS Estimates

In most practical situations for which residuals are not normally distributed, the sampling distributions of the regression estimates and their

standard errors can be estimated using the bootstrap method. By the use of BOLS method and RBOLS method, the distributions of regression estimates are estimated based on 1,000 bootstrap replications at different percentages of outliers for various sample sizes. The means and the $rmse^{\bullet}$s computed from the bootstrap distributions so obtained are presented in Table (4.5) — Table(4.8).

Table (4.5)

Estimates and $rmse^{\bullet}$s by BOLS Method and RBOLS Method for Three-Variable Regression Model When n=30

Percentage of Outliers	Bootstrap Method	$\bar{\hat{\beta}}_0^{\bullet}$	$\bar{\hat{\beta}}_1^{\bullet}$	$\bar{\hat{\beta}}_2^{\bullet}$	$rmse^{\bullet}(\hat{\beta}_0^{\bullet})$	$rmse^{\bullet}(\hat{\beta}_1^{\bullet})$	$rmse^{\bullet}(\hat{\beta}_2^{\bullet})$
0%	BOLS	1.9889	0.6926	0.4684	0.0343	0.0074	0.0375
	RBOLS	1.9567	0.6955	0.4756	0.0298	0.0063	0.0330
5%	BOLS	2.8837	0.5946	1.2435	0.5829	0.1294	0.6384
	RBOLS	1.9965	0.6920	0.4753	0.0356	0.0078	0.0401
10%	BOLS	3.2514	0.5311	1.3656	0.6818	0.1388	0.7247
	RBOLS	1.9795	0.6939	0.4683	0.0346	0.0071	0.0376
15%	BOLS	3.7141	0.6052	1.7417	0.7515	0.1535	0.8169
	RBOLS	1.9900	0.6917	0.4636	0.0380	0.0088	0.0417
20%	BOLS	3.9294	0.6332	1.7699	0.7451	0.1595	0.8453
	RBOLS	1.9981	0.6934	0.4688	0.0357	0.0088	0.0422

Table (4.6)

Estimates and $rmse^{\bullet}$s by BOLS Method and RBOLS Method for Three-Variable Regression Model When n=60

Percentage of Outliers	Bootstrap Method	$\bar{\hat{\beta}}_0^{\bullet}$	$\bar{\hat{\beta}}_1^{\bullet}$	$\bar{\hat{\beta}}_2^{\bullet}$	$rmse^{\bullet}(\hat{\beta}_0^{\bullet})$	$rmse^{\bullet}(\hat{\beta}_1^{\bullet})$	$rmse^{\bullet}(\hat{\beta}_2^{\bullet})$
0%	BOLS	2.0109	0.6978	0.4703	0.0226	0.0051	0.0287
	RBOLS	2.0204	0.6975	0.4738	0.0215	0.0049	0.0261
5%	BOLS	2.5796	0.6279	1.0965	0.3245	0.0745	0.4231
	RBOLS	2.0180	0.6975	0.4738	0.0218	0.0046	0.0263
10%	BOLS	3.0070	0.6860	1.2700	0.3957	0.0854	0.4677
	RBOLS	2.0285	0.6987	0.4764	0.0213	0.0051	0.0270
15%	BOLS	3.2826	0.7167	1.1728	0.4213	0.0927	0.5168
	RBOLS	2.0223	0.6980	0.4789	0.0223	0.0053	0.0272
20%	BOLS	3.8949	0.7659	1.1731	0.5253	0.1170	0.6462
	RBOLS	2.0243	0.6973	0.4808	0.0235	0.0053	0.0263

Table (4.7)

Estimates and $rmse^{\bullet}$ s by BOLS Method and RBOLS Method for Three-Variable Regression Model When n=100

Percentage of Outliers	Bootstrap Method	$\bar{\hat{\beta}}_0^{\bullet}$	$\bar{\hat{\beta}}_1^{\bullet}$	$\bar{\hat{\beta}}_2^{\bullet}$	$rmse^{\bullet}(\hat{\beta}_0^{\bullet})$	$rmse^{\bullet}(\hat{\beta}_1^{\bullet})$	$rmse^{\bullet}(\hat{\beta}_2^{\bullet})$
0%	BOLS	2.0114	0.6946	0.4786	0.0191	0.0041	0.0212
	RBOLS	2.0178	0.6942	0.4825	0.0180	0.0040	0.0197
5%	BOLS	2.6138	0.6782	1.0004	0.2425	0.0520	0.2494
	RBOLS	2.0283	0.6923	0.5031	0.0174	0.0039	0.0182
10%	BOLS	2.9892	0.6976	0.9754	0.3023	0.0673	0.3355
	RBOLS	2.0228	0.6932	0.4978	0.0176	0.0039	0.0197
15%	BOLS	3.4611	0.7633	1.1546	0.3648	0.0757	0.3822
	RBOLS	2.0300	0.6941	0.4995	0.0192	0.0042	0.0192
20%	BOLS	4.0249	0.7257	1.6385	0.4009	0.0895	0.4225
	RBOLS	2.0431	0.6930	0.5058	0.0203	0.0044	0.0195

Table (4.8)

Estimates and $rmse^{\bullet}$ s by BOLS Method and RBOLS Method for Three-Variable Regression Model When n=200

Percentage of Outliers	Bootstrap Method	$\bar{\hat{\beta}}_0^{\bullet}$	$\bar{\hat{\beta}}_1^{\bullet}$	$\bar{\hat{\beta}}_2^{\bullet}$	$rmse^{\bullet}(\hat{\beta}_0^{\bullet})$	$rmse^{\bullet}(\hat{\beta}_1^{\bullet})$	$rmse^{\bullet}(\hat{\beta}_2^{\bullet})$
0%	BOLS	2.0073	0.7018	0.5032	0.0143	0.0029	0.0162
	RBOLS	2.0058	0.7011	0.5053	0.0140	0.0031	0.0160
5%	BOLS	2.4762	0.6576	0.7249	0.1432	0.0297	0.1612
	RBOLS	2.0090	0.6997	0.5108	0.0150	0.0031	0.0169
10%	BOLS	2.9618	0.6472	0.8214	0.2072	0.0436	0.2273
	RBOLS	2.0078	0.7007	0.5118	0.0157	0.0033	0.0178
15%	BOLS	3.4731	0.6189	0.6479	0.2482	0.0529	0.2768
	RBOLS	2.0036	0.7019	0.5140	0.0160	0.0033	0.0179
20%	BOLS	3.9515	0.6494	1.0104	0.2778	0.0571	0.3142
	RBOLS	2.0067	0.7013	0.5113	0.0173	0.0035	0.0184

Confining one's attention to the results given in Table (4.5) through Table (4.8), almost all *rmse*$^{\bullet}$s of RBOLS estimates are found to be smaller than those of BOLS estimates at 0% outliers in the dataset. Moreover, the *rmse*$^{\bullet}$s of RBOLS estimates in all the above tables are distinctively less than those of BOLS estimates for datasets having 5% to 20% outliers. As percentage of outliers increases from 5% to 20% the *rmse*$^{\bullet}$s of BOLS estimates become gradually larger and larger. Nevertheless, the *rmse*$^{\bullet}$s of RBOLS estimates become almost stable. These facts support the desirability of the proposed RBOLS method over BOLS method whenever a dataset is contaminated with outliers.

In order to clearly illustrate superiority of RBOLS method over BOLS method in case of existence of outliers in a dataset, the graphs drawn for the *rmse*$^{\bullet}$s of BOLS and RBOLS estimates at different percentages of outliers for various sample sizes are presented in Figure B1(a) — Figure B4(c) of Appendix B.

4.4 Simulation Results for a Four-Variable Regression Model

In another simulation study, observations on the dependent variable Y are generated by the following linear regression model consisting of three independent variables:

$$Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_3 D + u \qquad (4.6)$$

with presumed values of the regression coefficients that are set at $\alpha_0 = 2$, $\alpha_1 = 0.7$, $\alpha_2 = 0.5$ and $\alpha_3 = 0.2$.

For the above linear regression model, the distribution used for generating data for the independent variables X_1 and X_2 are $X_1 \sim N(0.6,25)$ and $X_2 \sim N(-0.1,0.81)$; and for the disturbance term u is $u \sim N(0,0.04)$. The variable D is the dummy variable.

Outliers are introduced into generated datasets as contaminants. Some good disturbances are deliberately deleted, and they are replaced with bad disturbances. The contaminated bad disturbances are generated using the

distribution $N(10,9)$, that is, $u_{bad} \sim N(10,9)$. This simulation study is the same as the simulation study presented in Section 4.3 except that the dummy variable D is added to the three-variable regression model in Equation (4.5).

Discussion on OLS Estimates and OLS-LTS Estimates

The estimates of parameters α_0, α_1, α_2 and α_3 in Equation (4.6), their *rmse*s and R^2 are computed using OLS method and OLS-LTS method for various sample sizes having different percentages of outliers. The simulation results are presented in Table (4.9) — Table (4.12).

Table (4.9)

Estimates, *rmse*s and R^2 by OLS Method and OLS-LTS Method for Four-Variable Regression Model When n=30

Percentage of Outliers	Estimation Method	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$rmse(\hat{\alpha}_0)$	$rmse(\hat{\alpha}_1)$	$rmse(\hat{\alpha}_2)$	$rmse(\hat{\alpha}_3)$	R^2
0%	OLS	1.9912**	0.6932**	0.4667**	0.1917*	0.0457	0.0108	0.0527	0.0795	0.9969
	OLS-LTS	1.9894**	0.6960**	0.4758**	0.1461*	0.0433	0.0090	0.0455	0.0940	0.9974
5%	OLS	3.2105**	0.6263**	1.1621	-0.7985	1.4317	0.1612	0.9609	1.6772	0.4833
	OLS-LTS	2.0026**	0.6925**	0.4733**	0.1825*	0.0483	0.0116	0.0516	0.0843	0.9969
10%	OLS	3.7585**	0.5923**	1.2342	-1.2292	1.9618	0.1955	1.0802	2.0960	0.3957
	OLS-LTS	1.9795**	0.6940**	0.4686**	0.2003*	0.0509	0.0103	0.0520	0.0778	0.9973
15%	OLS	3.6537**	0.5941**	1.7736	0.4018	1.9185	0.2109	1.5514	1.7262	0.4063
	OLS-LTS	1.9808**	0.6902**	0.4654**	0.2306*	0.0516	0.0142	0.0580	0.0917	0.9961
20%	OLS	3.6398**	0.6005**	1.8448	1.1462	1.9225	0.2129	1.6262	2.0064	0.4183
	OLS-LTS	1.9800**	0.6906**	0.4712**	0.2615**	0.0509	0.0138	0.0540	0.1068	0.9964

** significant at 1% level * significant at 5% level

Table (4.10)

Estimates, *rmse*s and R^2 by OLS Method and OLS-LTS Method for Four-Variable Regression Model When n=60

Percentage of Outliers	Estimation Method	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$rmse(\hat{\alpha}_0)$	$rmse(\hat{\alpha}_1)$	$rmse(\hat{\alpha}_2)$	$rmse(\hat{\alpha}_3)$	R^2
0%	OLS	2.0160**	0.6977**	0.4672**	0.1875**	0.0343	0.0058	0.0443	0.0520	0.9972
	OLS-LTS	2.0297**	0.6975**	0.4691**	0.1741**	0.0413	0.0055	0.0415	0.0539	0.9976
5%	OLS	2.8950**	0.6358**	0.9719*	-0.5855	0.9974	0.1005	0.6395	1.0750	0.6186
	OLS-LTS	2.0273**	0.6978**	0.4709**	0.1769**	0.0397	0.0054	0.0404	0.0521	0.9978
10%	OLS	2.8282**	0.6893**	1.3339*	0.6795	0.9771	0.0917	0.9767	0.9884	0.5915
	OLS-LTS	2.0262**	0.6988**	0.4776**	0.2072**	0.0387	0.0053	0.0363	0.0493	0.9976
15%	OLS	3.2650**	0.7167**	1.1512*	0.2032	1.3793	0.0980	0.8455	0.9167	0.5682
	OLS-LTS	2.0172**	0.6980**	0.4804**	0.2167**	0.0346	0.0057	0.0348	0.0524	0.9976
20%	OLS	4.0349**	0.7694**	1.1311	-0.2514	2.1475	0.1392	0.9227	1.2302	0.4990
	OLS-LTS	2.0231**	0.6975**	0.4802**	0.2036**	0.0391	0.0061	0.0355	0.0524	0.9976

** significant at 1% level * significant at 5% level

Table (4.11)

Estimates, *rmse*s and R^2 by OLS Method and OLS-LTS Method for Four-Variable Regression Model When n=100

Percentage of Outliers	Estimation Method	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$rmse(\hat{\alpha}_0)$	$rmse(\hat{\alpha}_1)$	$rmse(\hat{\alpha}_2)$	$rmse(\hat{\alpha}_3)$	R^2
0%	OLS	2.0110**	0.6946**	0.4795**	0.2021**	0.0274	0.0069	0.0292	0.0394	0.9966
	OLS-LTS	2.0143**	0.6934**	0.4945**	0.2136**	0.0272	0.0076	0.0200	0.0387	0.9972
5%	OLS	2.6388**	0.6775**	0.9903**	0.1634	0.7123	0.0587	0.5553	0.4959	0.6557
	OLS-LTS	2.0219**	0.6932**	0.4972**	0.2028**	0.0321	0.0079	0.0200	0.0366	0.9971
10%	OLS	2.8446**	0.7076**	0.9935**	0.5497	0.9326	0.0684	0.5921	0.7121	0.5627
	OLS-LTS	2.0220**	0.6914**	0.5098**	0.2011**	0.0320	0.0095	0.0225	0.0369	0.9973
15%	OLS	3.5865**	0.7566**	1.1266**	-0.1413	1.6551	0.0988	0.7380	0.8144	0.5207
	OLS-LTS	2.0352**	0.6936**	0.4992**	0.1847**	0.0430	0.0076	0.0198	0.0410	0.9972
20%	OLS	4.2186**	0.7197**	1.6195**	-0.2860	2.2793	0.0920	1.2000	0.9528	0.4763
	OLS-LTS	2.0557**	0.6933**	0.4948**	0.1466**	0.0619	0.0081	0.0220	0.0674	0.9969

** significant at 1% level

Table (4.12)

Estimates, *rmse*s and R^2 by OLS Method and OLS-LTS Method for Four-Variable Regression Model When n=200

Percentage of Outliers	Estimation Method	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\alpha}_3$	$rmse(\hat{\alpha}_0)$	$rmse(\hat{\alpha}_1)$	$rmse(\hat{\alpha}_2)$	$rmse(\hat{\alpha}_3)$	R^2
0%	OLS	2.0397**	0.7014**	0.5088**	0.1358**	0.0448	0.0034	0.0187	0.0705	0.9964
	OLS-LTS#	2.0397**	0.7014**	0.5088**	0.1358**	0.0448	0.0034	0.0187	0.0705	0.9964
5%	OLS	2.5479**	0.6569**	0.7387**	0.0648	0.5835	0.0526	0.2883	0.3165	0.7265
	OLS-LTS	2.0511**	0.6995**	0.5135**	0.1254**	0.0553	0.0032	0.0216	0.0804	0.9965
10%	OLS	3.0041**	0.6461**	0.8357**	0.1045	1.0445	0.0691	0.4079	0.4212	0.5604
	OLS-LTS	2.0401**	0.7002**	0.5169**	0.1352**	0.0457	0.0033	0.0244	0.0718	0.9965
15%	OLS	3.7427**	0.6152**	0.6994*	-0.3508	1.7787	0.1003	0.3494	0.7491	0.4271
	OLS-LTS	2.0445**	0.7020**	0.5175**	0.1257**	0.0501	0.0040	0.0251	0.0809	0.9963
20%	OLS	4.1290**	0.6474**	1.0368**	-0.1499	2.1666	0.0800	0.6267	0.6709	0.4060
	OLS-LTS	2.0488**	0.7009**	0.5147**	0.1286**	0.0543	0.0036	0.0238	0.0788	0.9965

** significant at 1% level

All the LTS weights for the original dataset are one's which leads to same results as OLS.

Based on the results in Table (4.9) — Table (4.12), at 0% level of outlier percentage for various sample sizes, neither method is found to provide smaller biases and *rmse*s of the estimates. On comparison of R^2 under the two methods, the values of R^2 provided by both estimation methods are also found to be almost the same, indicating almost the same goodness of fit for both estimation methods. However, at 5% to 20% levels of outlier percentage for various sample sizes, it is also found that the biases and *rmse*s of the corresponding OLS-LTS estimates are always smaller than those of OLS estimates, and R^2 provided by OLS-LTS method is always larger than that provided by OLS method. It can be concluded that OLS-LTS method is preferable to OLS method whenever a dataset is contaminated with outliers.

Comparison of RBOLS Estimates with BOLS Estimates

BOLS method and RBOLS method are used to estimate the distributions of regression estimates on the basis of 1,000 bootstrap replications at different percentages of outliers for various sample sizes. The means and the *rmse**s computed from the bootstrap distributions of the regression estimates are presented in Table (4.13) — Table (4.16).

Table (4.13)

Estimates and $rmse^{\bullet}$ s by BOLS Method and RBOLS Method for Four-Variable Regression Model When n=30

Percentage of Outliers	Bootstrap Method	$\overline{\hat{\alpha}}_0^{\bullet}$	$\overline{\hat{\alpha}}_1^{\bullet}$	$\overline{\hat{\alpha}}_2^{\bullet}$	$\overline{\hat{\alpha}}_3^{\bullet}$	$rmse^{\bullet}(\hat{\alpha}_0^{\bullet})$	$rmse^{\bullet}(\hat{\alpha}_1^{\bullet})$	$rmse^{\bullet}(\hat{\alpha}_2^{\bullet})$	$rmse^{\bullet}(\hat{\alpha}_3^{\bullet})$
0%	BOLS	1.9900	0.6925	0.4681	0.1961	0.0428	0.0078	0.0374	0.0764
	RBOLS	1.9906	0.6958	0.4783	0.1468	0.0390	0.0073	0.0381	0.0693
5%	BOLS	3.2026	0.6195	1.1628	-0.7934	0.7022	0.1328	0.6446	1.2139
	RBOLS	2.0010	0.6927	0.4700	0.1855	0.0432	0.0083	0.0407	0.0758
10%	BOLS	3.7461	0.5949	1.2426	-1.1850	0.7942	0.1426	0.7172	1.3762
	RBOLS	1.9823	0.6940	0.4691	0.1978	0.0430	0.0078	0.0384	0.0727
15%	BOLS	3.6152	0.5992	1.8146	0.3790	0.9268	0.1717	0.8155	1.5726
	RBOLS	1.9834	0.6901	0.4664	0.2285	0.0438	0.0093	0.0436	0.0766
20%	BOLS	3.6415	0.6123	1.8224	1.1596	0.9263	0.1739	0.8509	1.6048
	RBOLS	1.9795	0.6905	0.4718	0.2619	0.0425	0.0092	0.0424	0.0775

Table (4.14)

Estimates and $rmse^*$ s by BOLS Method and RBOLS Method for Four-Variable Regression Model When n=60

Percentage of Outliers	Bootstrap Method	$\bar{\hat{\alpha}}_0^*$	$\bar{\hat{\alpha}}_1^*$	$\bar{\hat{\alpha}}_2^*$	$\bar{\hat{\alpha}}_3^*$	$rmse^*(\hat{\alpha}_0^*)$	$rmse^*(\hat{\alpha}_1^*)$	$rmse^*(\hat{\alpha}_2^*)$	$rmse^*(\hat{\alpha}_3^*)$
0%	BOLS	2.0157	0.6978	0.4671	0.1869	0.0293	0.0052	0.0288	0.0495
	RBOLS	2.0313	0.6975	0.4697	0.1746	0.0278	0.0049	0.0262	0.0456
5%	BOLS	2.9118	0.6359	0.9504	-0.6102	0.4361	0.0736	0.4235	0.7202
	RBOLS	2.0277	0.6978	0.4714	0.1766	0.0283	0.0047	0.0263	0.0464
10%	BOLS	2.8068	0.6885	1.3328	0.7329	0.5047	0.0882	0.5158	0.8219
	RBOLS	2.0253	0.6990	0.4791	0.2092	0.0278	0.0050	0.0274	0.0459
15%	BOLS	3.2333	0.7161	1.1532	0.2239	0.5240	0.0890	0.5266	0.8881
	RBOLS	2.0164	0.6980	0.4802	0.2185	0.0279	0.0052	0.0276	0.0472
20%	BOLS	3.9976	0.7756	1.1069	-0.2184	0.6637	0.1179	0.6770	1.0974
	RBOLS	2.0209	0.6975	0.4800	0.2051	0.0307	0.0054	0.0287	0.0496

Table (4.15)

Estimates and $rmse^*$ s by BOLS Method and RBOLS Method for Four-Variable Regression Model When n=100

Percentage of Outliers	Bootstrap Method	$\bar{\hat{\alpha}}_0^*$	$\bar{\hat{\alpha}}_1^*$	$\bar{\hat{\alpha}}_2^*$	$\bar{\hat{\alpha}}_3^*$	$rmse^*(\hat{\alpha}_0^*)$	$rmse^*(\hat{\alpha}_1^*)$	$rmse^*(\hat{\alpha}_2^*)$	$rmse^*(\hat{\alpha}_3^*)$
0%	BOLS	2.0105	0.6948	0.4794	0.2033	0.0246	0.0042	0.0207	0.0376
	RBOLS	2.0138	0.6936	0.4941	0.2141	0.0221	0.0036	0.0191	0.0354
5%	BOLS	2.6487	0.6753	1.0050	0.1689	0.3180	0.0533	0.2495	0.5106
	RBOLS	2.0226	0.6932	0.4970	0.2024	0.0225	0.0041	0.0192	0.0341
10%	BOLS	2.8346	0.7088	0.9811	0.5616	0.3816	0.0668	0.3282	0.6236
	RBOLS	2.0225	0.6915	0.5103	0.2014	0.0226	0.0041	0.0203	0.0360
15%	BOLS	3.5796	0.7595	1.1306	-0.1194	0.4615	0.0797	0.3710	0.7469
	RBOLS	2.0351	0.6936	0.4988	0.1847	0.0242	0.0039	0.0195	0.0365
20%	BOLS	4.2133	0.7241	1.6224	-0.2646	0.5154	0.0875	0.4261	0.8050
	RBOLS	2.0547	0.6934	0.4940	0.1468	0.0256	0.0044	0.0203	0.0397

Table (4.16)

Estimates and $rmse^{\bullet}$ s by BOLS Method and RBOLS Method for Four-Variable Regression Model When n=200

Percentage of Outliers	Bootstrap Method	$\overline{\hat{\alpha}}_0^{\bullet}$	$\overline{\hat{\alpha}}_1^{\bullet}$	$\overline{\hat{\alpha}}_2^{\bullet}$	$\overline{\hat{\alpha}}_3^{\bullet}$	$rmse^{\bullet}(\hat{\alpha}_0^{\bullet})$	$rmse^{\bullet}(\hat{\alpha}_1^{\bullet})$	$rmse^{\bullet}(\hat{\alpha}_2^{\bullet})$	$rmse^{\bullet}(\hat{\alpha}_3^{\bullet})$
0%	BOLS	2.0388	0.7014	0.5093	0.1363	0.0204	0.0030	0.0167	0.0288
	RBOLS #	2.0388	0.7014	0.5093	0.1363	0.0204	0.0030	0.0167	0.0288
5%	BOLS	2.5520	0.6568	0.7435	0.0661	0.2053	0.0316	0.1599	0.2938
	RBOLS	2.0515	0.6996	0.5134	0.1260	0.0207	0.0031	0.0162	0.0296
10%	BOLS	3.0080	0.6432	0.8323	0.0989	0.2886	0.0415	0.2355	0.4136
	RBOLS	2.0405	0.7001	0.5165	0.1340	0.0214	0.0031	0.0174	0.0306
15%	BOLS	3.7569	0.6153	0.7060	-0.3706	0.3515	0.0544	0.2848	0.5120
	RBOLS	2.0442	0.7018	0.5180	0.1263	0.0230	0.0034	0.0176	0.0320
20%	BOLS	4.1197	0.6476	1.0382	-0.1142	0.3898	0.0591	0.3179	0.5808
	RBOLS	2.0480	0.7009	0.5138	0.1297	0.0237	0.0034	0.0183	0.0329

All the LTS weights for the original dataset are one's which leads to same results as BOLS.

Confining one's attention to the results given in Table (4.13) through Table (4.16), almost all $rmse^{\bullet}$s of RBOLS estimates are found to be smaller than those of BOLS estimates in the absence of outliers, i.e., at 0% outliers in the dataset. Moreover, the $rmse^{\bullet}$s of RBOLS estimates in all the above tables are distinctively less than those of BOLS estimates for datasets having 5% to 20% outliers. These facts support the desirability of the proposed RBOLS method over BOLS method whenever the dataset is contaminated with outliers.

The superiority of RBOLS method over BOLS method can be seen in the graphs drawn for the $rmse^{\bullet}$s of BOLS and RBOLS estimates at different percentages of outliers for various sample sizes as shown in Figure B5(a) — Figure B8(d) of Appendix B.

Based on the findings from the simulation results conducted for two different linear regression models considered in this chapter, a conclusion can be drawn that the proposed RBOLS method has proved to be better than BOLS method whenever the dataset applied to linear regression model has been contaminated with outliers.

CHAPTER V
BOOTSTRAP PREDICTION INTERVALS FOR SPIRULINA PRODUCTIVITY

In this Chapter V, one-day-ahead bootstrap prediction intervals for Spirulina productivity in culturing ponds at MSF are constructed. At first, a dynamic regression model of optical density of Spirulina is fitted. Using the fitted model, the prediction intervals are constructed by the method of residual resampling bootstrapping. Then, a linear regression model of the optical density of Spirulina is fitted based on the dataset in which outliers had been removed. Using the fitted model, the prediction intervals are constructed by applying the proposed RBOLS method developed in Chapter IV.

5.1 Bootstrap Prediction Intervals in Dynamic Regression Model

Based on the two algorithms presented in Section 3.8, an algorithm for the residual resampling bootstrapping associated with a dynamic regression model in which the disturbance term follows the AR(1) scheme is suggested. A dynamic regression model is described as:

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 X_{t-1,1} + \beta_3 X_{t-1,2} + u_t \quad \text{for } t = 1,...,n, \qquad (5.1)$$

where $\beta_0,...,\beta_3$ are the unknown parameters of interest, Y_t is the dependent variable at time t, Y_{t-1} is the one lagged dependent variable, $X_{t-1,1}$ and $X_{t-1,2}$ are the explanatory variables at time t-1, and u_t is the disturbance term at time t. It is assumed that u_t follows the AR(1) scheme in Equation (3.13), namely,

$$u_t = \rho u_{t-1} + v_t, \qquad -1 < \rho < 1.$$

Algorithm for Prediction Intervals in Dynamic Regression Model

In the algorithm for one-step-ahead bootstrap prediction intervals in a dynamic regression model of Equation (4.1), a total of six steps needs to be followed. They are:

Step1: Fit the OLS regression to the original sample of observations and use Cochrane-Orcutt iterative method to obtain the estimated first-order correlation coefficient $\hat{\rho}$, and transform the original observations as follows:

$$y_t^T = Y_t - \hat{\rho} Y_{t-1} \qquad \text{for } t = 2,...,n,$$

$$x_{t,1}^T = X_{t,1} - \hat{\rho} X_{t-1,1} \qquad \text{for } t = 2,...,n,$$

$$x_{t,2}^T = X_{t,2} - \hat{\rho} X_{t-1,2} \qquad \text{for } t = 2,...,n.$$

Step2: Fit again the OLS regression to the transformed observations to obtain FGLS estimates $\hat{\beta}_{FGLS}$, then compute the estimated values $\hat{y}_t^T$ given by

$$\hat{y}_t^T = f(y_{t-1}^T, x_{t-1,1}^T, x_{t-1,2}^T, \hat{\beta}_{FGLS}) \qquad \text{for } t = 3,...,n,$$

and obtain the residuals r_t's given by

$$r_t = y_t^T - \hat{y}_t^T \qquad \text{for } t = 3,...,n.$$

Step3: Draw a random sample ε_t^* of size $(n-2)$ from the values of r_t with replacement and obtain new bootstrap values Y_t^* of y_t^T where

$$Y_t^* = \hat{\beta}_0 + \hat{\beta}_1 Y_{t-1}^* + \hat{\beta}_2 x_{t-1,1}^T + \hat{\beta}_3 x_{t-1,2}^T + \varepsilon_t^* \qquad \text{for } t = 4,...,n,$$

in which

$$Y_3^* = \hat{\beta}_0 + \hat{\beta}_1 y_2^T + \hat{\beta}_2 x_{2,1}^T + \hat{\beta}_3 x_{2,2}^T + \varepsilon_3^*.$$

Step4: Fit the OLS regression to the new bootstrap values Y_t^* obtained in ***Step3*** on the $(Y_{t-1}^*, x_{t-1,1}^T, x_{t-1,2}^T)$ to obtain bootstrap estimates $\hat{\beta}_i^{*b}$, $s^*(\hat{\beta}_i^{*b})$ and s^{*b}, respectively.

Step5: Repeat ***Step3*** and ***Step4*** for B replications to obtain the bootstrap distributions of $\hat{\beta}_i^{*b}$, $s^*(\hat{\beta}_i^{*b})$ and s^{*b} $(b = 1,...,B)$.

Step6: (a) Obtain the bootstrap distribution of the forecast $\hat{y}_{t+1}^{*b}$ given by

$$\hat{y}_{t+1}^{*b} = \hat{\beta}_0^{**b} + \hat{\beta}_1^{*b} Y_t + \hat{\beta}_2^{*b} X_{t,1} + \hat{\beta}_3^{*b} X_{t,2} \qquad \text{for } b = 1,...,B,$$

where $\hat{\beta}_0^{**b} = \hat{\beta}_0^{*b} / (1 - \hat{\rho})$.

(b) Obtain the bootstrap studentized prediction error z^{*b} given by

$$z^{*b} = \frac{\hat{y}^{*b}_{t+1} - (\hat{y}_{t+1} + \varepsilon^{*b}_{t+1})}{s^{*b}}, \qquad \text{for } b = 1,...,B,$$

where $\hat{y}_{t+1} = \hat{\beta}_{0(GLS)} + \hat{\beta}_{1(GLS)} Y_t + \hat{\beta}_{2(GLS)} X_{t,1} + \hat{\beta}_{3(GLS)} X_{t,2}$.

(c) Construct the $(1\text{-}2\alpha)$ one-day-ahead prediction interval for Y_{t+1} given by

$$[\hat{y}_{t+1} - s_f \; z^{*}_{((B+1)(1-\alpha))}, \hat{y}_{t+1} - s_f \; z^{*}_{((B+1)\alpha)}],$$

where $s_f = s(1 + x'_f (X'X)^{-1} x_f)^{1/2}$,

in which $x'_f = (1, y_t, x_{t,1}, x_{t,2})$.

5.2 Bootstrap Prediction Intervals for Spirulina Productivity in Dynamic Regression Model

In order to compute prediction intervals for Spirulina productivity (the optical density of Spirulina) in culturing ponds at MSF, a dynamic regression model of the optical density of Spirulina is first specified on the basis of the dataset on optical density of Spirulina and related variables over a period of 365 days for the year 2007 presented in Table C1, Appendix C. Computations are done by using S-PLUS software and Microsoft Office Excel 2007 with Pop Tool 3.1.0 package add-in.

5.2.1 A Dynamic Regression Model of the Optical Density of Spirulina

Some studies on the optical density of Spirulina have shown, as a hypothesis, that the optical density of Spirulina on a particular day might mainly depend on the previous day's (i) optical density of Spirulina, (ii) salinity of water, (iii) pH value of water, (iv) air temperature, (v) light, (vi) season, and (vii) condition that Spirulina was harvested or not. Therefore, a dynamic regression model of the optical density of Spirulina that is supposed to be consistent with the above said hypothesis is postulated as follows:

$$OD_t = \beta_0 + \beta_1 OD_{t-1} + \beta_2 SAL_{t-1} + \beta_3 DW_{t-1} + \beta_4 DHAR_{t-1} + \beta_5 DS_{t-1} + \beta_6 PH_{t-1} + \beta_7 TEM_{t-1} + \beta_8 LIG_{t-1} + u_t, \quad (5.2)$$

where OD_t is the optical density of Spirulina on *day (t)*,

SAL_{t-1} is the salinity of water on *day (t-1)*(in parts per ten thousand),

DW_{t-1} is the dummy variable that takes value 1 if the season happens to be winter (cold season) on *day (t-1)* or value 0 otherwise,

$DHAR_{t-1}$ is the dummy variable that takes value 1 if Spirulina was harvested on *day (t-1)* or value 0 otherwise,

DS_{t-1} is the dummy variable that takes value 1 if the season happens to be summer (hot season) on *day (t-1)* or value 0 otherwise,

PH_{t-1} is the pH value of water on *day (t-1)*,

TEM_{t-1} is the air temperature on *day (t-1)*, and

LIG_{t-1} is the light on *day (t-1)*.

The final term u_t is the random disturbance term at *day (t)*. It is assumed that u_t follows the *AR(1)* process in Equation (3.13), namely,

$$u_t = \rho\, u_{t-1} + v_t. \qquad -1 < \rho < 1$$

The disturbance term v_t is assumed to be independent and identically distributed with a zero mean and a finite variance with no autocorrelation at all.

Equation (5.2) is fitted using OLS method and applying the stepwise variable selection procedure. Of the variables considered, DS_{t-1}, PH_{t-1}, TEM_{t-1} and LIG_{t-1} are found not to be significant. The fitted model obtained is as follows:

$$\widehat{OD}_t = 0.090 + 0.777\, OD_{t-1} + 0.040\, SAL_{t-1} + 0.023\, DW_{t-1} - 0.104\, DHAR_{t-1}. \quad (5.3)$$

The specific OLS results of the above fitted model are reported in Table (5.1).

All of the regression coefficients in Table (5.1) are statistically significant and different from zero. Nevertheless, Durbin *h*-statistic in Equation (3.15), which is an appropriate test statistic when the lagged dependent variable appears on the right-hand size of Equation (5.3), takes the value of -4.271, which leads to the conclusion that there exists first-order autocorrelation in the fitted model (*p*-value = 0.000).

Table (5.1)
OLS Estimates of the Dynamic Regression Model

Variable	Estimated coefficient	Standard error	t-statistic	Sig-t
(Constant)	0.090	0.009	9.682	0.000
OD_{t-1}	0.777	0.032	24.121	0.000
SAL_{t-1}	0.040	0.020	2.047	0.041
DW_{t-1}	0.023	0.007	3.283	0.001
$DHAR_{t-1}$	-0.104	0.006	-16.865	0.000
R^2 = 0.733, Adjusted R^2 = 0.730, Standard error of the estimate = 0.054 F(4, 359) = 246.974, Sig. F = 0.000				

BG test statistic in Equation (3.18) for higher orders of autocorrelation and autoregressive structure in the disturbances is found to take the value of 22.063. It indicates that the null hypothesis of no autocorrelation is rejected against the alternative hypothesis of autocorrelation up to order three (p-value = 0.000). In light of results from Durbin h test and BG test, one can conclude that the disturbances are not independent of each other; that is, they are autocorrelated to each other.

For that reason, Equation (5.2) is fitted again using the method of FGLS as discussed in Section 3.4 of Chapter III. The first-order autocorrelation in Equation (3.13) is estimated by the Cochrane-Orcutt iterative method. The estimated value of -0.264 is used to transform the variable in Equation (5.3). The dynamic regression model of transformed optical density of Spirulina OD_t^T fitted by the method of FGLS is finally obtained as follows:

$$\widehat{OD}_t^T = 0.091 + 0.857\ OD_{t-1}^T + 0.021\ SAL_{t-1}^T + 0.015\ DW_{t-1}^T - 0.101\ DHAR_{t-1}^T, \tag{5.4}$$

where $OD_t^T = (OD_t - \hat{\rho}\, OD_{t-1})$, $SAL_t^T = (SAL_t - \hat{\rho}\, SAL_{t-1})$, $DW_t^T = (DW_t - \hat{\rho}\, DW_{t-1})$ and $DHAR_t^T = (DHAR_t - \hat{\rho}\, DHAR_{t-1})$. The FGLS results for the fitted dynamic regression model in Equation (5.4) are presented in the following Table (5.2).

Table (5.2)
FGLS Estimates of the Dynamic Regression Model

Variable	Estimated coefficient	Standard error	t-statistic	Sig-t
(Constant)	0.091	0.009	9.596	0.000
OD_{t-1}^{T}	0.857	0.026	32.533	0.000
SAL_{t-1}^{T}	0.021	0.010	2.100	0.036
DW_{t-1}^{T}	0.015	0.006	2.765	0.006
$DHAR_{t-1}^{T}$	-0.101	0.006	-17.199	0.000
R^2 = 0.824, Adjusted R^2 = 0.822 Standard error of the estimate = 0.053 F(4, 358) = 418.946, Sig. F = 0.000				

All of the regression coefficients are statistically significant and different from zero. Durbin *h*-statistic takes the value of -0.672, which is consistent with independent disturbances (*p*-value = 0.501). To check further, BG test is also carried out. BG test statistic turns out to take the value of 4.536 indicating that the null hypothesis of no autocorrelation is not rejected against the alternative hypothesis of autocorrelation up to order three (*p*-value = 0.209). In light of these tests, one may be confident that the disturbances are independent of each other.

All the correlation coefficients between each explanatory variable and residual $\hat{v}_t$ are 0.000. The null hypothesis of no correlation between each explanatory variable and the disturbance term v_t is not rejected (*p*-value = 1.000). This means that all explanatory variables and the disturbance term v_t are uncorrelated.

To verify whether there exists heteroscedasticity in the model, special case of White test in Subsection 3.2.2 is also used. Based on White test statistic value of 3.495, one cannot reject the null hypothesis of no heteroscedasticity (*p*-value = 0.174); that is, there exists no heteroscedasticity in the model.

And then, to ensure whether the disturbances follow the normal distribution, JB test of normality in Subsection 3.2.1 is conducted. Since the

value of JB statistic turns out to be 304.658, the hypothesis that the disturbances are normally distributed is rejected (p-value = 0.000); that is, one concludes that the disturbances are not normally distributed.

Now, to apply the model in Equation (5.4) for forecasting one-day-ahead optical density of Spirulina, i.e., on *day (t+1)*, Equation (5.4) can be re-expressed as follows:

$$\widehat{OD}_t = 0.072 + 0.857\, OD_{t-1} + 0.021\, SAL_{t-1} + 0.015\, DW_{t-1} - 0.101\, DHAR_{t-1}. \quad (5.5)$$

Using Equation (5.5), the optical density of Spirulina on *day (t+1)* can be predicted by the information on the optical density of Spirulina, salinity of water, season and harvest on *day (t)*.

5.2.2 Confidence Intervals for the Parameters of the Dynamic Regression Model

Based on JB test of normality, it has been found that the disturbances in the fitted regression model are not normally distributed. In this subsection, the bootstrap distributions of the FGLS estimates and the corresponding confidence intervals for the respective parameters are presented. The bootstrap distributions of the FGLS estimates are based on 1,000 bootstrap replications, that is, B = 1,000. These distributions are presented in Figure (5.1) — Figure(5.5).

Figure (5.1) Bootstrap Distribution of $\hat{\beta}_0^*$

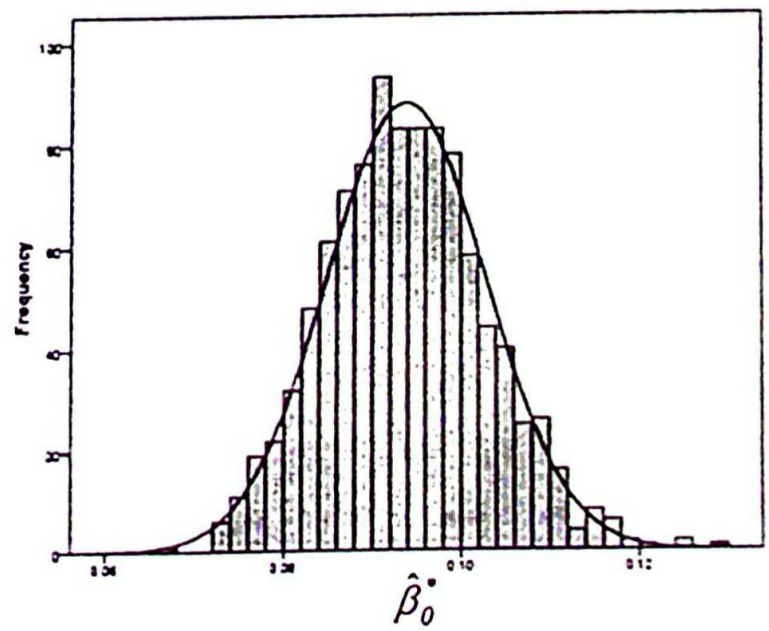


Figure (5.2) Bootstrap Distribution of $\hat{\beta}_1^*$

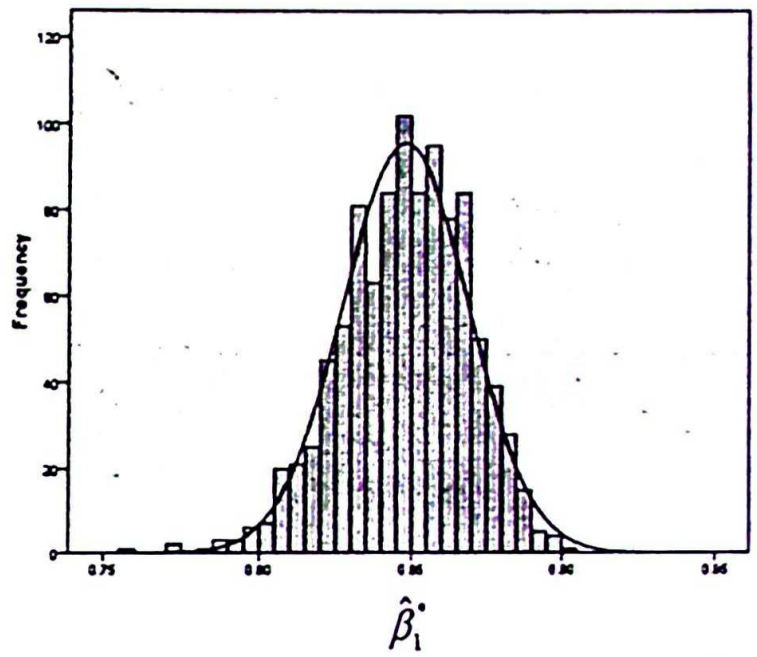


Figure (5.3) Bootstrap Distribution of $\hat{\beta}_2^*$

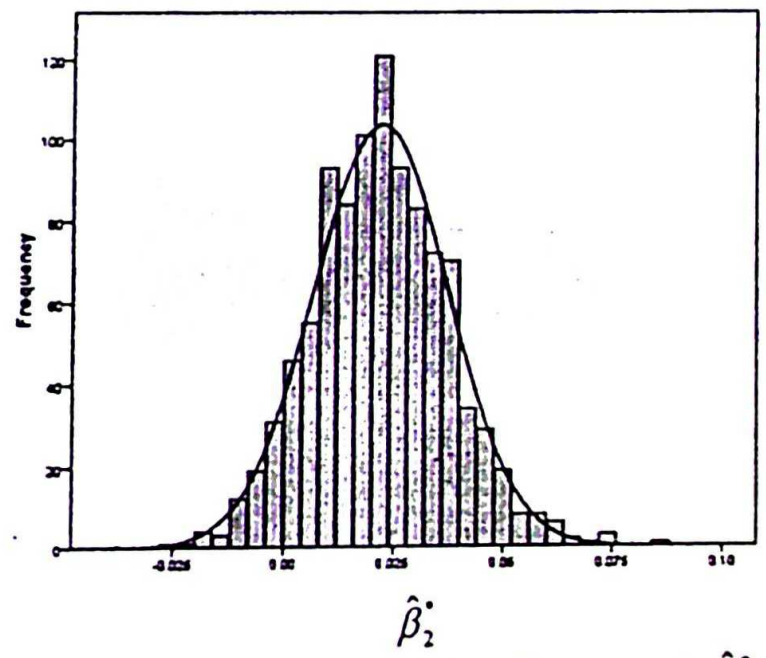


Figure (5.4) Bootstrap Distribution of $\hat{\beta}_3^*$

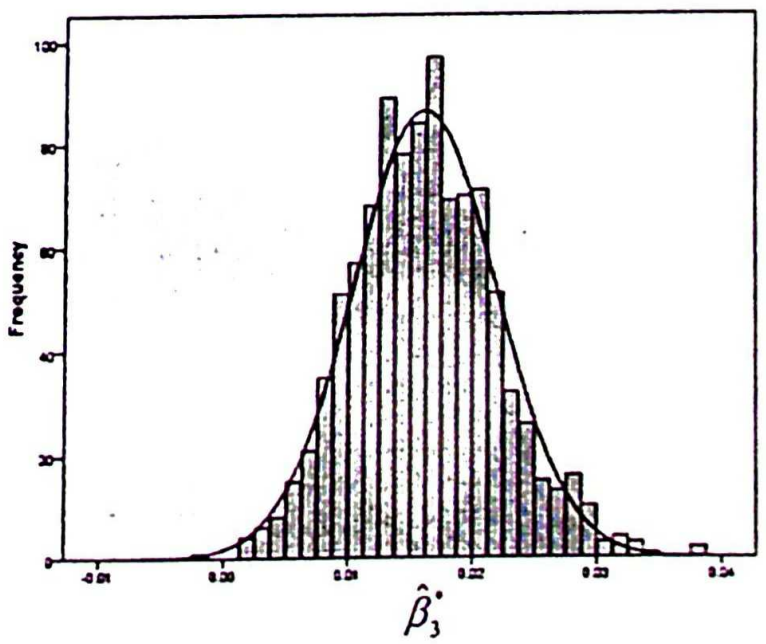
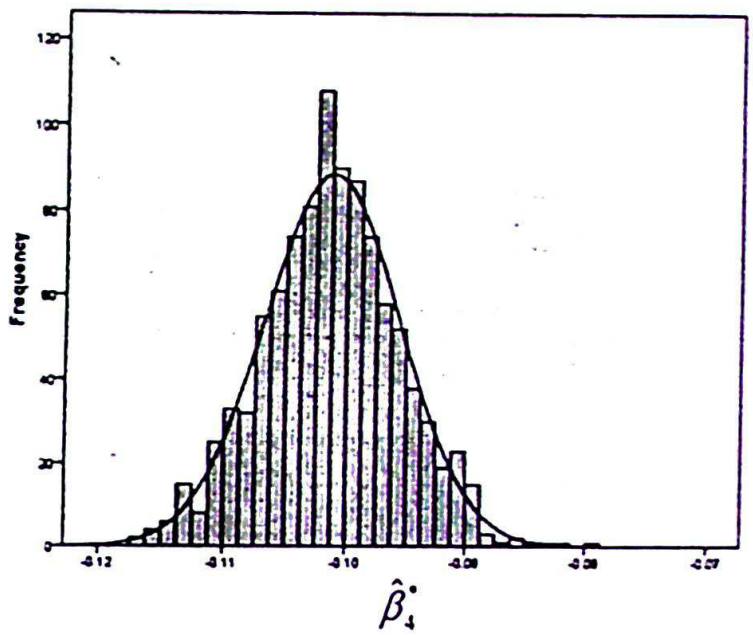


Figure (5.5) Bootstrap Distribution of $\hat{\beta}_4^*$



After obtaining the bootstrap distributions of $\hat{\beta}_0^*$, $\hat{\beta}_1^*$, $\hat{\beta}_2^*$, $\hat{\beta}_3^*$ and $\hat{\beta}_4^*$, JB test is then conducted to examine their shapes. The values of skewness for the distributions of $\hat{\beta}_0^*$, $\hat{\beta}_1^*$, $\hat{\beta}_2^*$, $\hat{\beta}_3^*$ and $\hat{\beta}_4^*$, are 0.253, -0.391, 0.234, 0.100 and 0.306, respectively. The results from JB test indicate that all the above bootstrap distributions are not normally distributed except the bootstrap distribution of $\hat{\beta}_3^*$. The means and standard errors of the distributions of $\hat{\beta}^*$'s are computed. They are compared with the FGLS estimates in the following Table(5.3).

Table (5.3)
FGLS Estimates and Bootstrap Estimates
of the Dynamic Regression Model

Variable	FGLS Estimate		Bootstrap Estimate	
	Coefficient	Standard error	Mean	Standard error
(Constant)	0.091	0.009	0.094	0.009
OD_{t-1}^T	0.857	0.026	0.848	0.021
SAL_{t-1}^T	0.021	0.010	0.022	0.010
DW_{t-1}^T	0.015	0.006	0.016	0.006
DHR_{t-1}^T	-0.101	0.006	-0.101	0.006

It can be found from the above Table (5.3) that the bootstrap estimates are very close to the FGLS estimates. Moreover, the bootstrap estimated standard errors are found to be approximately equal to those of the corresponding FGLS estimates.

In order to construct the studentized bootstrap confidence intervals for the parameters, the bootstrap distributions of the FGLS estimates are first transformed into the bootstrap distributions of z_j^*'s: $z_j^{*b} = \dfrac{\hat{\beta}_j^{*b} - \hat{\beta}_j}{s^*(\hat{\beta}_j^*)}$ (j=0,1,2,3,4). The bootstrap distributions of z_j^*'s are presented in Figure (5.6) — Figure (5.10).

Figure (5.6) Bootstrap Distribution of z_0^* for $\hat{\beta}_0^*$

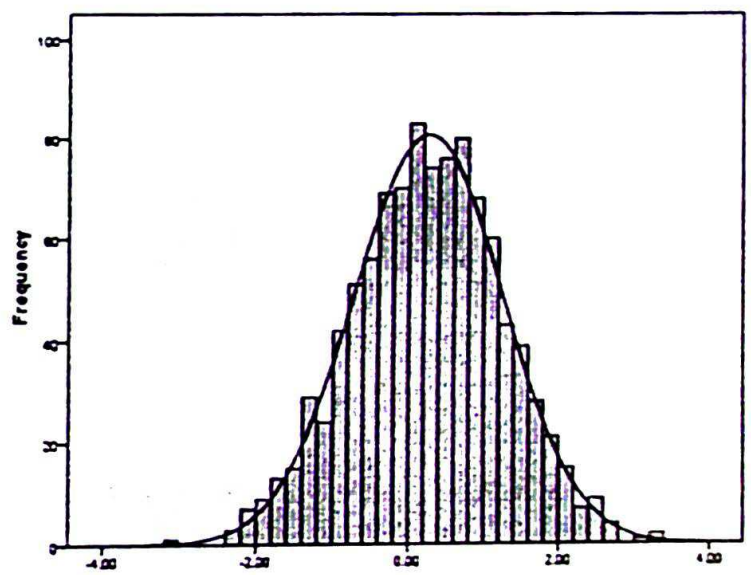


Figure (5.7) Bootstrap Distribution of z_1^* for $\hat{\beta}_1^*$

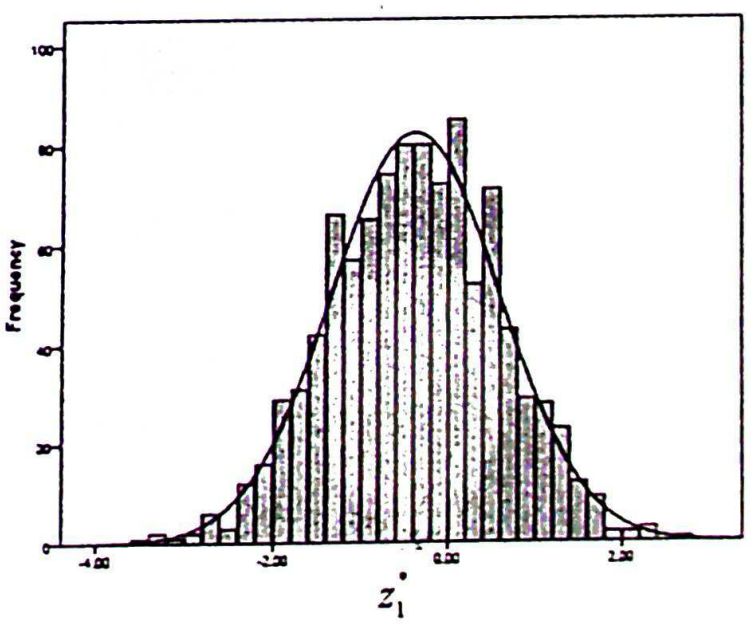


Figure (5.8) Bootstrap Distribution of z_2^* for $\hat{\beta}_2^*$

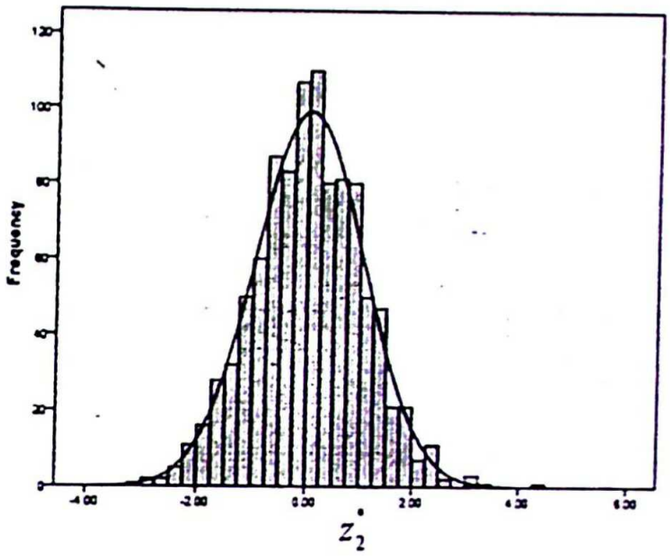


Figure (5.9) Bootstrap Distribution of z_3^* for $\hat{\beta}_3^*$

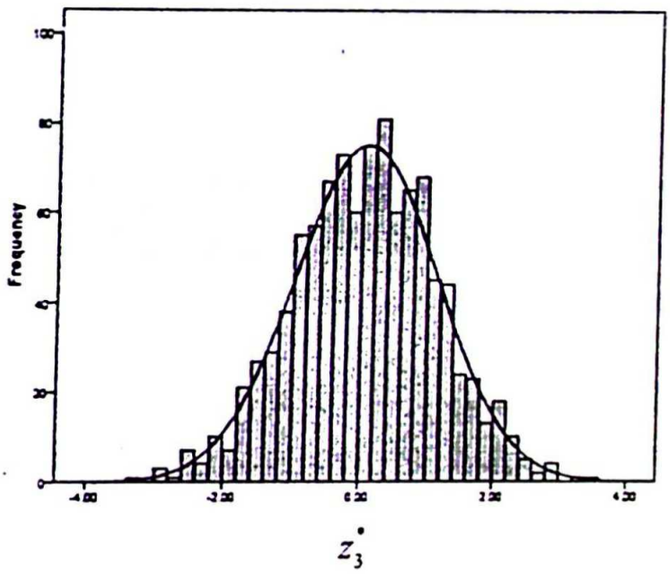
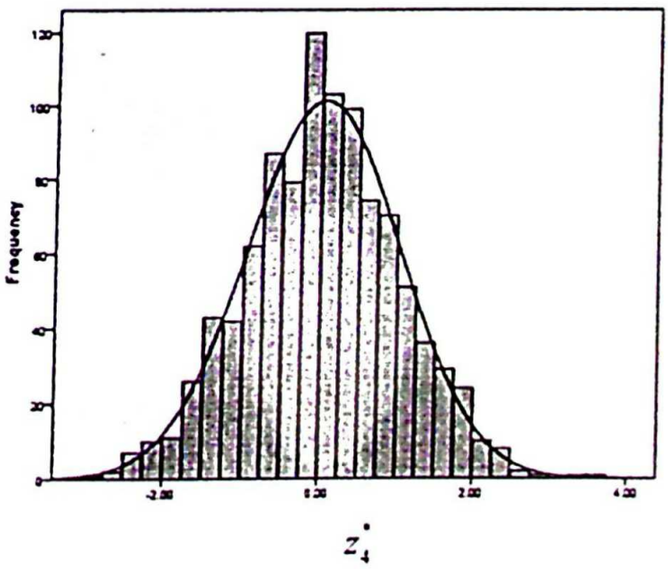


Figure (5.10) Bootstrap Distribution of z_4^* for $\hat{\beta}_4^*$



Unlike the distributions of $\hat{\beta}_j^*$, all the above distributions of z_j^*'s that are used to get studentized bootstrap confidence intervals are found to be normally distributed indicated by JB test results. The values of skewness for the distributions of z_0^*, z_1^*, z_2^*, z_3^* and z_4^* are -0.039, -0.035, 0.109, 0.073 and -0.001, respectively. Since all the distributions of z_j^*'s are found to be more or less symmetrical, all studentized bootstrap confidence intervals for the parameters are found to be approximately symmetrical. The 95% studentized bootstrap confidence intervals for the respective parameters are presented in Table (5.4).

Table (5.4)
Bootstrap Confidence Intervals for the Parameters of the Dynamic Regression Model

Variable	95 % confidence interval		Variable	95 % confidence interval	
	Lower limit	Upper limit		Lower limit	Upper limit
(Constant)	0.070	0.107	(Constant)	0.060	0.097
OD_{t-1}^T	0.819	0.916	OD_{t-1}	0.819	0.916
SAL_{t-1}^T	-0.011	0.051	SAL_{t-1}	-0.011	0.051
DW_{t-1}^T	0.003	0.027	DW_{t-1}	0.003	0.027
$DHAR_{t-1}^T$	-0.113	-0.090	$DHAR_{t-1}$	-0.113	-0.090

Attempts have been made to construct prediction intervals for the optical density of Spirulina. Some of these prediction intervals are discussed in the following Subsection 5.2.3.

5.2.3 Prediction Intervals for the Optical Density of Spirulina Using the Dynamic Regression Model

The construction of 95% one-day-ahead bootstrap prediction intervals for the optical density of Spirulina is carried out. The median value of the optical density of Spirulina in the recorded dataset is found to be 0.29 (in 680 nanometer). Taking the median value of the optical density of 0.29 (in 680 nanometer), and different values of salinity of water, season, and state of being harvested as the data on *day (t)*, the prediction intervals for the optical density

of Spirulina on *day (t+1)* are computed and presented in Table (5.5). In Table(5.5), the prediction intervals are found to cover almost all the observed (actual) optical density of Spirulina, except Case No. (2).

Table (5.5)
One-Day-Ahead Prediction Intervals for the Optical Density of Spirulina Using the Dynamic Regression Model

Case No.	OD_t	SAL_t	DW_t	$DHAR_t$	Forecast OD_{t+1}	95% prediction interval		Observed OD_{t+1}
						Lower limit	Upper limit	
1	0.29	0.1	0	0	0.323	0.230	0.428	0.30
2	0.29	0.1	0	0	0.323	0.230	0.428	0.48
3	0.29	0.1	1	0	0.338	0.246	0.445	0.36
4	0.29	0.2	0	0	0.325	0.232	0.429	0.32
5	0.29	0.2	0	1	0.223	0.128	0.325	0.25
6	0.29	0.2	1	0	0.340	0.247	0.446	0.29
7	0.29	0.2	1	0	0.340	0.247	0.446	0.29
8	0.29	0.2	1	1	0.239	0.143	0.342	0.28
9	0.29	0.3	0	0	0.327	0.234	0.431	0.29
10	0.29	0.3	0	0	0.327	0.234	0.431	0.30
11	0.29	0.3	0	1	0.225	0.130	0.328	0.24
12	0.29	0.3	0	1	0.225	0.130	0.328	0.26
13	0.29	0.3	0	1	0.225	0.130	0.328	0.28
14	0.29	0.3	1	0	0.342	0.249	0.448	0.29
15	0.29	0.3	1	0	0.342	0.249	0.448	0.33
16	0.29	0.3	1	0	0.342	0.249	0.448	0.39
17	0.29	0.4	0	1	0.228	0.131	0.330	0.22
18	0.29	0.4	0	1	0.228	0.131	0.330	0.23
19	0.29	0.5	0	1	0.230	0.132	0.334	0.27

For different values of the optical density of Spirulina, salinity of water, season and state of being harvested in 2007 taken as the observed values on *day (t)*, 95% one-day-ahead bootstrap prediction intervals for the optical density of Spirulina are computed and presented in Table C2, Appendix C. Out of 364 prediction intervals computed, it is found that 342 actual readings of the optical density of Spirulina on a day (94.0 percent) fall within the intervals.

5.3 Bootstrap Prediction Intervals for Spirulina Productivity in Linear Regression Model

In 138 cases of the dataset used for fitting the dynamic regression model, the optical density of Spirulina on *day (t+1)* is lower than that on *day (t)* because Spirulina was harvested on *day (t)*. Under the natural situation, one can seldom find that the optical density of Spirulina on *day (t+1)* is unusually lower than that on *day (t)*. Therefore, these cases are removed from the dataset in undertaking the analysis and a linear regression model of the optical density of Spirulina is fitted based on the rest of the dataset containing 226 cases which are approximately 62 percent of the original dataset, and prediction intervals for the optical density of Spirulina are determined.

5.3.1 A Linear Regression Model of the Optical Density of Spirulina

A linear regression model of the optical density of Spirulina is formulated as follows:

$$OD_t = \alpha_0 + \alpha_1 OD_{t-1} + \alpha_2 SAL_{t-1} + +\alpha_3 DW_{t-1} + \alpha_4 DS_{t-1} + \alpha_5 PH_{t-1} + \alpha_6 TEM_{t-1} + \alpha_7 LIG_{t-1} + u_t, \quad (5.6)$$

where OD_t is the optical density of Spirulina on *day (t)*,

SAL_{t-1} is the salinity of water on *day (t-1)* (in parts per ten thousand),

DW_{t-1} is the dummy variable that takes value 1 if the season happens to be winter (cold season) on *day (t-1)* or value 0 otherwise,

DS_{t-1} is the dummy variable that takes value 1 if the season happens to be summer (hot season) on *day (t-1)* or value 0 otherwise,

PH_{t-1} is the pH value of water on *day (t-1)*,

TEM_{t-1} is the air temperature on *day (t-1)*, and

LIG_{t-1} is the light on *day (t-1)*.

The final term u_t is the random disturbance at *day (t)*, which is assumed to be independent and identically distributed with a zero mean and a finite variance.

Equation (5.6) is fitted (i) using OLS estimation method and applying the stepwise procedure for variable selection and (ii) using OLS-LTS estimation method. The fitted model so obtained is as follows:

$$\widehat{OD}_t = 0.036 + 0.934\ OD_{t-1} + 0.030\ SAL_{t-1} + 0.009\ DW_{t-1}. \qquad (5.7)$$

The OLS-LTS results of the fitted model in Equation (5.7) are reported in Table(5.6).

Table (5.6)
OLS-LTS Estimates of the Linear Regression Model

Variable	Estimated coefficient	Standard error	t-statistic	Sig-t
(Constant)	0.036	0.006	6.402	0.000
OD_{t-1}	0.934	0.021	43.753	0.000
SAL_{t-1}	0.030	0.012	2.560	0.011
DW_{t-1}	0.009	0.005	2.031	0.044
$R^2 = 0.940$, Adjusted $R^2 = 0.939$ Standard error of the estimate = 0.024 F(3, 190) = 994.332, Sig. F = 0.000				

All of the regression coefficients have the expected signs, and they are statistically significant and different from zero. Based on the significance of individual regression coefficients and overall significance of regression model itself, the fitted regression model in Equation (5.7) is able to explain the optical density of Spirulina well. To check for higher orders of autoregressive structure in the disturbances, BG test is carried out and test statistic takes the value of 4.996 indicating that the null hypothesis of no autocorrelation is not rejected against the alternative hypothesis of autocorrelation up to order three (p-value = 0.172). In the light of BG test, one may be confident that the disturbances are independent of each other.

All the correlation coefficients between each explanatory variable and residual $\hat{u}_t$ are 0.000. The null hypothesis of no correlation between each explanatory variable and the disturbance term u_t is not rejected (p-value = 1.000). This means that all explanatory variables on the right hand side of Equation (5.7) and the disturbance term u_t are uncorrelated.

To verify that whether there is heteroscedasticity in the disturbance in the fitted model, special case of White test is then carried out. It is found that the test statistic value of 0.417 is not able to reject the null hypothesis of no heteroscedasticity (p-value = 0.812); that is, there exists homoscedasticity in the disturbances.

Further, to ensure whether the disturbances follow the normal distribution, JB test is also conducted. Since the value of JB statistic turns out to be 14.444, the hypothesis that the disturbances are normally distributed can be rejected (p-value = 0.001).

Using the fitted model, the optical density of Spirulina on *day (t+1)* can be forecasted by the information on the optical density of Spirulina, salinity of water, season on *day (t)*.

On comparing the FGLS estimates in Table (5.2) with the OLS-LTS estimates in Table (5.6), it is found that adjusted R^2 and computed F-value for the linear regression model are substantially larger than those for the dynamic regression model and standard error of the estimate of linear regression model is noticeably smaller than that of dynamic regression model.

5.3.2 Confidence Intervals for the Parameters of the Linear Regression Model

In the simulation study, regressors are assumed as stochastic and the estimators belong to desirable asymptotic properties. Since bootstrapping is considered in large sample case, the estimators can be safely applied to actual data. Based on 1,000 bootstrap replications, the bootstrap distributions of the OLS-LTS estimates obtained through RBOLS method as proposed in Chapter III are presented in Figure (5.11) — Figure (5.14).

Figure (5.11) Bootstrap Distribution of $\hat{\alpha}_0^*$

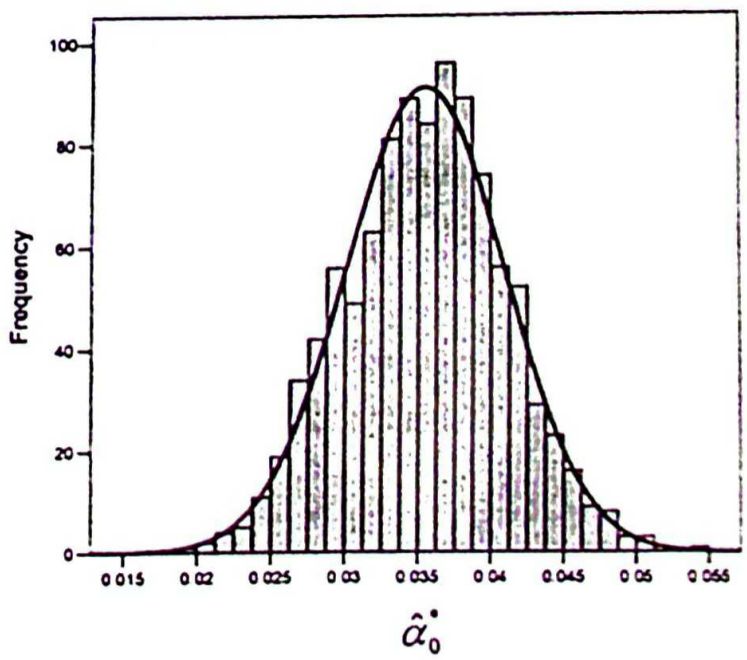


Figure (5.12) Bootstrap Distribution of $\hat{\alpha}_1^*$

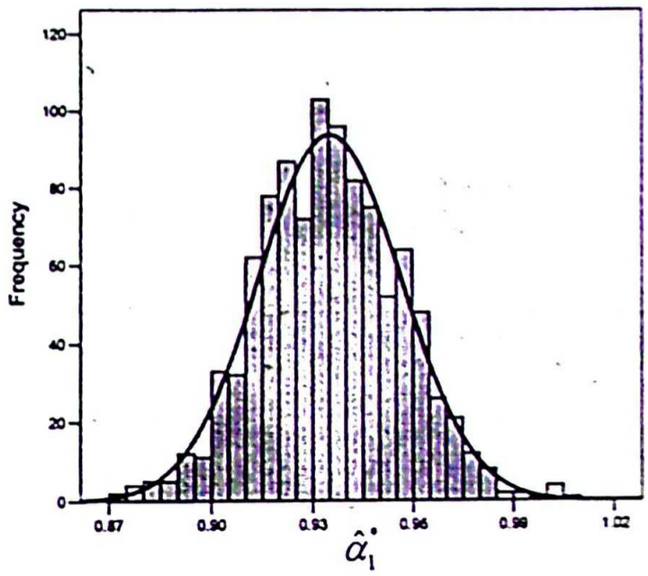


Figure (5.13) Bootstrap Distribution of $\hat{\alpha}_2^*$

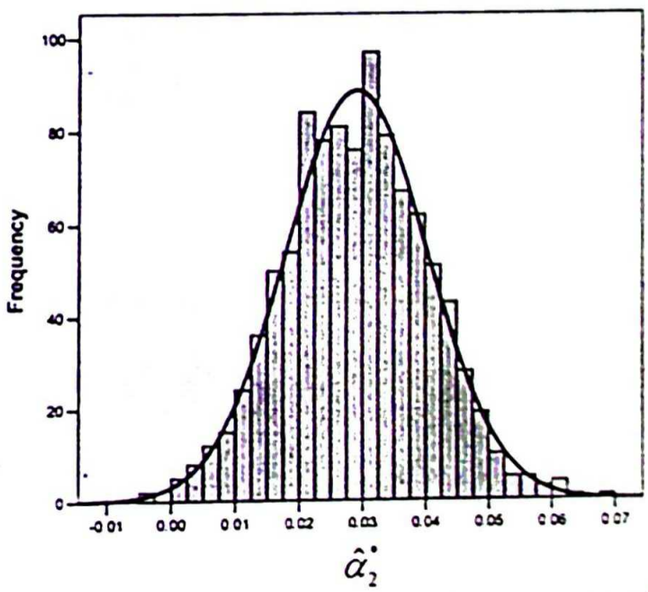
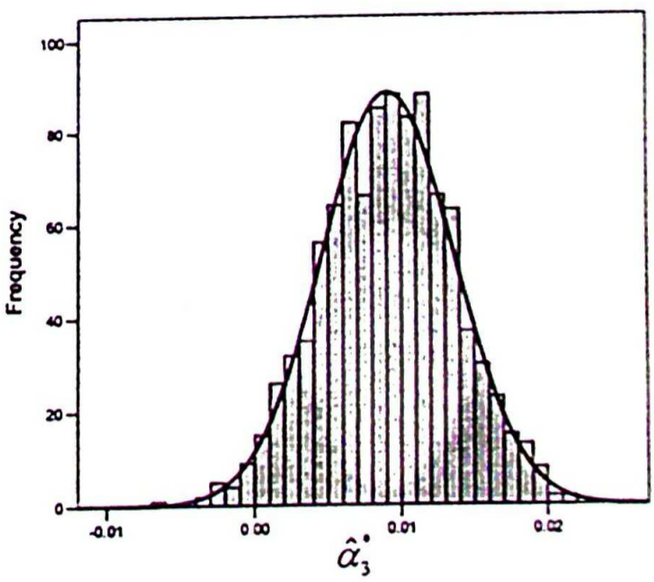


Figure (5.14) Bootstrap Distribution of $\hat{\alpha}_3^*$



The values of skewness for the distributions of $\hat{\alpha}_0^*$, $\hat{\alpha}_1^*$, $\hat{\alpha}_2^*$ and $\hat{\alpha}_3^*$ are 0.039, 0.113, 0.084 and -0.060, respectively and JB test indicates that all the above distributions are concluded to be normally distributed. The means and standard errors of the distributions of $\hat{\alpha}_j^*$ are computed. These results are then compared with the OLS-LTS estimates in the following Table (5.7).

Table (5.7)
OLS-LTS Estimates and Bootstrap Estimates of the Linear Regression Model

Variable	OLS-LTS Estimates		Bootstrap Estimates	
	Coefficient	Standard error	Mean	Standard error
(Constant)	0.036	0.006	0.035	0.005
OD_{t-1}	0.934	0.021	0.935	0.021
SAL_{t-1}	0.030	0.012	0.029	0.011
DW_{t-1}	0.009	0.005	0.009	0.005

It can be seen from the above Table (5.7) that the bootstrap estimates and bootstrap estimated standard errors are found to be almost the same as the corresponding OLS-LTS coefficients and their estimated standard errors. In order to construct the bootstrap confidence intervals for the parameters, the bootstrap distributions of the OLS-LTS estimates are then transformed into the bootstrap distributions of z_j^*'s *(j = 0,1,2,3)*. The bootstrap distributions of z_j^*'s are presented in Figure (5.15) — Figure (5.18).

Figure (5.15) Bootstrap Distribution of z_0^* for $\hat{\alpha}_0^*$

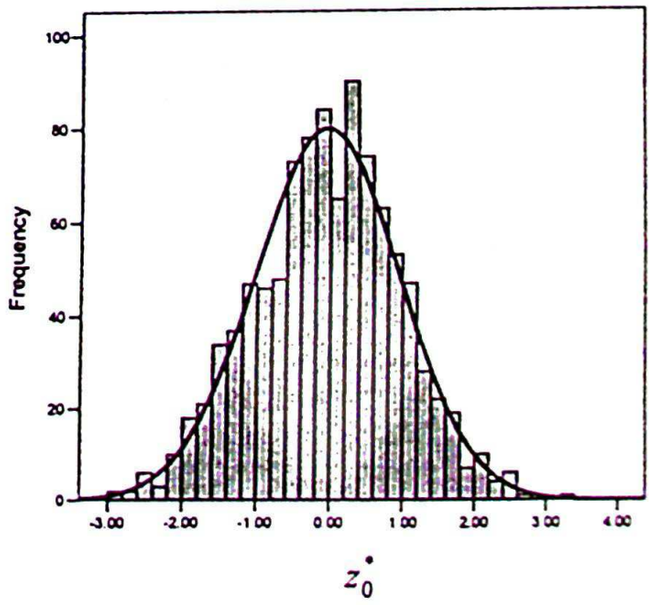


Figure (5.16) Bootstrap Distribution of z_1^* for $\hat{\alpha}_1^*$

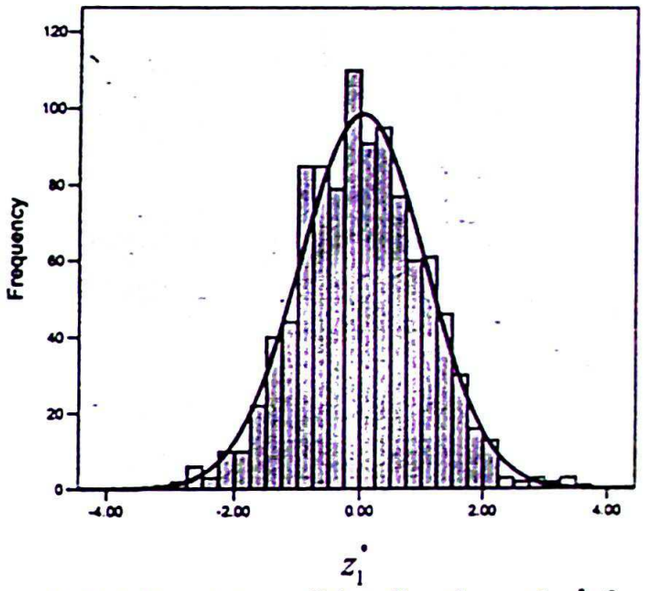


Figure (5.17) Bootstrap Distribution of z_2^* for $\hat{\alpha}_2^*$

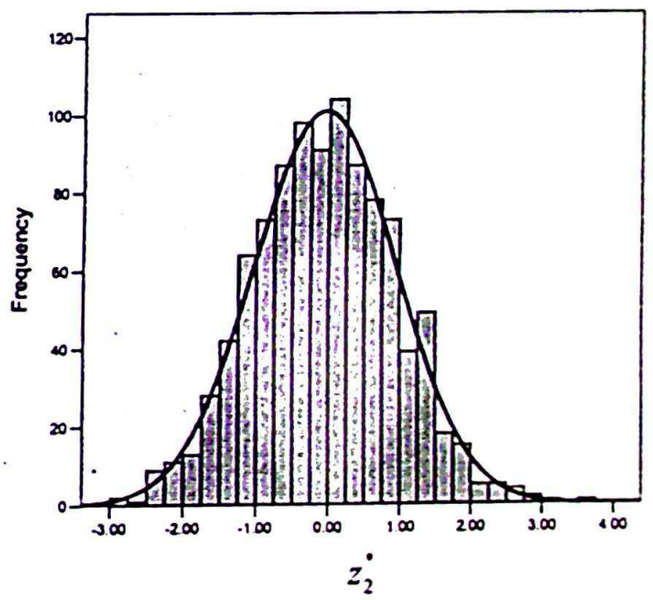
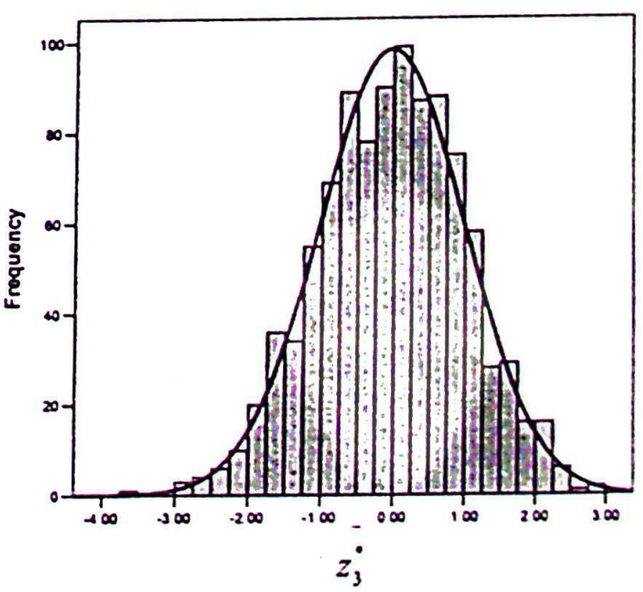


Figure (5.18) Bootstrap Distribution of z_3^* for $\hat{\alpha}_3^*$



The values of skewness for the distributions of z_0^*, z_1^*, z_2^* and z_3^* are -0.022, 0.120, 0.109 and -0.066, respectively. JB test indicates that all the above distributions of z_j^*'s are normally distributed. The 95% studentized bootstrap confidence intervals for the parameters, that are computed based on the OLS-LTS estimates, their standard errors and the distributions of z_j^*'s, are presented in Table (5.8).

Table (5.8)
Bootstrap Confidence Intervals for the Parameters of the Linear Regression Model

Variable	95 % confidence interval	
	Lower limit	Upper limit
(Constant)	0.025	0.047
OD_{t-1}	0.891	0.975
SAL_{t-1}	0.008	0.052
DW_{t-1}	0.000	0.018

To sum up and to end with, the prediction intervals for the optical density of Spirulina are constructed.

5.3.3 Prediction Intervals for the Optical Density of Spirulina Using the Linear Regression Model

The construction of 95% one-day-ahead bootstrap prediction intervals for the optical density of Spirulina is carried out. Taking the optical density of 0.28 (in 680 nanometer) that is the median value as well as the modal value, and different values of salinity of water and season taken as data on *day(t)*, the prediction intervals for the optical density of Spirulina on *day (t+1)* are computed, and they are presented in Table (5.9). It is found from the Table(5.9) that the prediction intervals cover all the observed (actual) optical density of Spirulina.

Table (5.9)
One-Day-Ahead Prediction Intervals for the Optical Density of Spirulina Using the Linear Regression Model

Case. No.	OD_t	SAL_t	DW_t	Forecast	95% prediction interval		Observed
				OD_{t+1}	Lower limit	Upper limit	OD_{t+1}
1	0.28	0.1	0	0.300	0.266	0.358	0.33
2	0.28	0.2	0	0.303	0.270	0.361	0.28
3	0.28	0.2	0	0.303	0.270	0.361	0.28
4	0.28	0.2	0	0.303	0.270	0.361	0.30
5	0.28	0.2	0	0.303	0.270	0.361	0.32
6	0.28	0.2	0	0.303	0.270	0.361	0.35
7	0.28	0.2	1	0.312	0.278	0.370	0.29
8	0.28	0.2	1	0.312	0.278	0.370	0.30
9	0.28	0.3	0	0.306	0.273	0.363	0.29
10	0.28	0.4	0	0.309	0.276	0.367	0.28
11	0.28	0.4	0	0.309	0.276	0.367	0.34
12	0.28	0.5	0	0.312	0.278	0.369	0.29

For different values on the optical density of Spirulina, salinity of water and season in 2007 taken as the observed values on *day (t)*, 95% one-day-ahead bootstrap prediction intervals for the optical density of Spirulina are computed and these intervals can be obtained in Table C3, Appendix C. Out of 194 prediction intervals computed, 187 actual readings of the optical density of Spirulina on a day (96.4 percent) are found to fall within the intervals.

The widths of prediction intervals obtained from both regression models are computed and compared. It is found that the maximum value of the widths of prediction intervals obtained from linear regression model is less than the minimum value of the widths of prediction intervals obtained from dynamic regression model. Therefore, it is concluded that all the widths of prediction intervals obtained from linear regression model are narrower than those obtained from the dynamic regression model and the prediction intervals obtained from the linear regression model are more precise than those obtained from the dynamic regression model.

Based on the findings from the results of bootstrap prediction intervals for Spirulina productivity computed in this chapter, the following suggestions can be drawn.

The prediction intervals for the optical density of Spirulina on *day (t+1)* obtained from the linear regression model can be applied for the cases in which Spirulina was not harvested on *day (t)*. However, for the cases in which Spirulina was harvested on *day (t)*, the prediction intervals for the optical density of Spirulina on *day (t+1)* obtained from the dynamic regression model should be used. In order to apply the prediction intervals for the optical density of Spirulina on *day (t+1)* by using the linear regression model when Spirulina was harvested on *day (t)*, one needs to measure the optical density of Spirulina after harvesting and filling media in the culturing pond.

CHAPTER VI
CONCLUSION

In this final Chapter VI, based on the results and findings in this study, conclusion on performance of the proposed RBOLS method compared to BOLS method and some suggestions on prediction intervals for Spirulina productivity in culturing ponds at MSF, together with recommendations are presented.

6.1 Conclusion on Performance of the Proposed RBOLS Method

If the disturbances in a linear regression model do not follow normal distribution, the finite sample distributions of the OLS estimates would no longer follow normal distributions. At this juncture, to estimate the distributions of the OLS estimates, BOLS method can be employed. Bootstrap distributions of the OLS estimates are desirable whenever the original dataset does not contain any outliers. However, the bootstrap distribution is a very poor estimator of the distribution of the OLS estimate when the dataset is contaminated with outliers because the OLS estimates are very sensitive to outliers. In such a situation, in order to obtain reliable distributions of the regression estimates, robust bootstrap methods should be resorted to, instead of BOLS method. An alternative bootstrap method called RBOLS which is not only computationally simple but also resistant to the effect of outliers has been introduced in this study. Based on the findings from the simulation results, it was found that bootstrap distributions of the regression estimates, which were obtained from RBOLS method, were better than the bootstrap distributions of the corresponding regression estimates which were obtained from BOLS method. Achieving almost always smaller $rmse^{\bullet}$s was the criterion in selecting the most plausible method to be put into use. It could be concluded that the proposed RBOLS method has proved to be better than BOLS method whenever the dataset applied to linear regression model has been contaminated with outliers.

6.2 Suggestions on Prediction Intervals for Spirulina Productivity

In order to empirically compute one-day-ahead bootstrap prediction intervals for Spirulina productivity (the optical density of Spirulina) in culturing ponds at MSF, the daily recorded data on the optical density of Spirulina, salinity of water, pH value of water, air temperature, light, season and state of being harvested for the year 2007 at a randomly chosen culturing pond were collected. A dynamic regression model of optical density of Spirulina was fitted by the method of FGLS. Using the residual resampling bootstrapping, confidence intervals for the respective parameters and 95% one-day-ahead bootstrap prediction intervals for the optical density of Spirulina were computed. Out of 364 prediction intervals computed, it was found that 342 actual readings of the optical density of Spirulina on a day (94.0 percent) fell within the intervals.

In a total of 364 cases of the dataset used for fitting the dynamic regression model, it was found that there were 138 cases in which the optical density of Spirulina on *day (t+1)* was less than the optical density of Spirulina observed on *day (t)*. These 138 cases were excluded from the dataset, and a linear regression model of the optical density of Spirulina was fitted based on the rest of the original dataset. The proposed RBOLS method was applied to compute confidence intervals for the regression parameters and 95% one-day-ahead bootstrap prediction intervals for the optical density of Spirulina. Out of 194 prediction intervals computed, 187 actual readings of the optical density of Spirulina on a day (96.4 percent) were found to fall within the intervals.

Comparing between the widths of respective prediction intervals obtained from both regression models, it was found that all the widths of prediction intervals obtained from the linear regression model were narrower than those obtained from the dynamic regression model. Therefore, it was concluded that the prediction intervals obtained from the linear regression model were more precise than those obtained from the dynamic regression model and desirable to be applied in making forecasts.

However, only for the cases in which Spirulina was not harvested on *day(t)*, the prediction intervals for the optical density of Spirulina on *day (t+1)* obtained from the linear regression model can be applied. For the cases in which Spirulina was harvested on *day(t)*, the prediction intervals for the optical density of Spirulina on *day (t+1)* obtained from the dynamic regression model should be used. In order to apply the prediction intervals for the optical density of Spirulina on *day (t+1)* by using the linear regression model when Spirulina was harvested on *day (t)*, one needs to measure the optical density of Spirulina after harvesting and filling media in the culturing pond.

6.3 Recommendations

A few areas of research work concerned with the application of robust bootstrap methods in linear regression are given below:

In the proposed RBOLS method, outliers are removed by using the LTS weights. Other methods that deal with outliers not by using the LTS weights should be explored in order to compare the efficiency of the bootstrap distributions of regression estimates and select the more efficient method.

If the original dataset is contaminated with outliers, better bootstrap methods based on robust regression methods, that are capable of resisting outliers in both original dataset as well as bootstrap samples, can be developed. However, achievement of such bootstrap methods would depend upon the availability of ready-made computer software and researcher's capacity to write the required computer programs.

REFERENCES

1. Athreya, K. B. (1987), Bootstrap of the Mean in the Infinite Variance Case, *Annals of Statistics*, vol. 15, 724-731.

2. Bernard, J. T. and M. Veall (1987), The Probability Distributions of Future Demand: The Case of Hydro Quebac, *Journal of Business and Economic Statistics*, vol. 5, 417-424.

3. Bickel, P. J. and D. A. Freedman (1981), Some Asymptotic Theory for the Bootstrap, *Annals of Statistics*, vol. 9, 1196-1217.

4. Breusch, T. S. (1978), Testing for Autocorrelation in Dynamic Linear Models, *Australian Economic Papers*, vol. 17, 334-355.

5. Cochrane, D. and G. H. Orcutt (1949), Application of Least Squares Regressions to Relationships Containing Autocorrelated Error Terms, *Journal of the American Statistical Association*, vol. 44, 32-61.

6. Davison, A. C. and D. V. Hinkley (1997), *Bootstrap Methods and Their Applications*, Cambridge University Press, New York.

7. De Angelis, D., P. Hall and G. A. Young (1993), Analytical and Bootstrap Approximations to Estimator Distributions in L_1 Regression, *Journal of the American Statistical Association*, vol. 88, 1310-1316.

8. DiCiccio, T. J. and J. P. Romano (1988), A Review of Bootstrap Confidence Intervals (with Discussions), *Journal of the Royal Statistical Society series B*, vol. 50, 338-370.

9. Durbin, J. (1970), Testing for Serial Correlation in Least-Squares Regression When Some of the Regressors Are Lagged Dependent Variables, *Econometrica*, vol. 38, 410-421.

10. Efron, B. (1979), Bootstrap Methods: Another Look at the Jackknife, *Annals of Statiatics*, vol. 7, 1-26.

11. Efron, B. (1981), Censored Data and the Bootstrap, *Journal of the American Statistical Association*, vol. 76, 312-319.

12. Efron, B. (1987), Better Bootstrap Confidence Intervals (with Discussion), *Journal of the American Statistical Association*, vol. 82, 171-200.

13. Efron, B. and G. Gong (1983), A Leisurely Look at the Bootstrap, the Jackknife, and Cross-Validation, *American Statistician*, vol. 37, 36-48.

14. Enders, W. (2004), *Applied Econometric Time Series*, John Wiley & Sons, Inc, New York.

15. Freedman, D. A. (1981), Bootstrapping Regression Models, *Annals of Statistics*, vol. 9, 1218-1228.

16. Freedman, D. A. (1984), On Bootstrapping Two-Stage Least-Squares Estimates in Stationary Linear Models, *Annals of Statistics*, vol. 12, 827-842.

17. Freedman, D. A. and S. C. Peters (1984), Bootstrapping a Regression Equation: Some Empirical Results, *Journal of the American Statistical Association*, vol. 79, 97-106.

18. Geisser, S. (1993), *Predicted Inference: An Introduction*, Chapman & Hall, London.

19. Godfrey, L. G. (1978), Testing Against General Autoregressive and Moving Average Error Models When the Regressors Include Lagged Dependent Variables, *Econometrica*, vol. 46, 1293-1302.

20. Gujarati, D. N. (1995), *Basic Econometrics* (3rd ed.), McGraw-Hill, Singapore.

21. Gujarati, D. N. and Sangeetha (2007), *Basic Econometrics* (4th ed.), Tata McGraw-Hill, New Delhi.

22. Hall, P. (1986), On the Bootstrap and Confidence Interval, *Annals of Statistics*, vol. 14, 1431-1452.

23. Hall, P. (1988), Theoretical Comparison of Bootstrap Confidence Intervals (with Discussion), *Annals of Statistics*, vol. 16, 927-985.

24. Hall, P. (1989), Unusual Properties of Bootstrap Confidence Intervals in Regression Problems, *Probability Theory and Related Fields*, vol. 81, 247-273.

25. Jarque, C. M. and A. K. Bera (1987), A Test for Normality of Observations and Regression Residuals, *International Statistical Review*, vol. 55, 163-172.

26. Johnston, J. and J. DiNardo (1997), *Econometric Methods*, McGraw-Hill, Singapore.

27. Kim, J. (2005), Bias-Corrected Bootstrap Inference for Regression Models with Autocorrelated Error, *Economics Bulletin*, vol. 3, no. 44, 1-8.

28. May Yu Khaing (2007), Productivity and Quality Control of Spirulina Platensis Biomass Culture on Commercial Scale in Myanmar, Unpublished Ph.D. Thesis, University of Mandalay, Myanmar.

29. Midi, H., H. S. Uraibi and B. A. Talib (2009), Dynamic Robust Bootstrap Method Based on LTS Estimators, *European Journal of Scientific Research*, vol. 32, no. 3, 277-287.

30. Olshen, R. A., E. N. Biden, M. P. Wyatt and D. H. Sutherland (1989), Gait Analysis and the Bootstrap, *Annals of Statistics*, vol. 17, 1419-1440.

31. Prescott, D. M. and T. Stengos (1987), Bootstrapping Confidence Intervals: An Application to Forecasting the Supply of Pork, *American Journal of Agricultural Economics*, vol. 69, 266-273.

32. Riadth, K., L. Cottrelt and V. Vigneron (2002), Bootstrap for Neural Model Selection, *Neurocomputing Journal*, vol. 48, 175-183.

33. Rousseeuw, P. J. (2006), Robust Regression: Positive Breakdown in, In *Encyclopedia of Statistical Science* (2nd ed.), vol. 11, 7316-7330, John Wiley & Sons, Inc, New Jersey.

34. Salibian-Barrera, M., and R. H. Zamar (2002), Bootstrapping Robust Estimates of Regression, *Annals of Statistics*, vol. 30, no. 2, 556-582.

35. Shao, J. and D. Tu (1995), *The Jackknife and Bootstrap*, Springer, New York.

36. Stine, R. A. (1985), Bootstrap Prediction Intervals for Regression, *Journal of the American Statistical Association*, vol. 80, 1026-1031.

37. White, H. (1980), A Heteroscedasticity Consistent Covariance Matrix Estimator and a Direct Test of Heteroscedasticity, *Econometrica*, vol. 48, 817-818.

38. Willems, G. and S. V. Aelst (2004), Fast and Robust Bootstrap for LTS, *Computational Statistics & Data Analysis*, vol. 48, 703-715.

39. Wooldridge, J. M. (2009), *Introductory Econometrics: A Modern Approach* (4th ed.), South-Western Cengage Learning, Canada.

APPENDICES

APPENDIX A

Plate A1 Photobioreactor for Pre-Culture of Selected Spirulina Strain

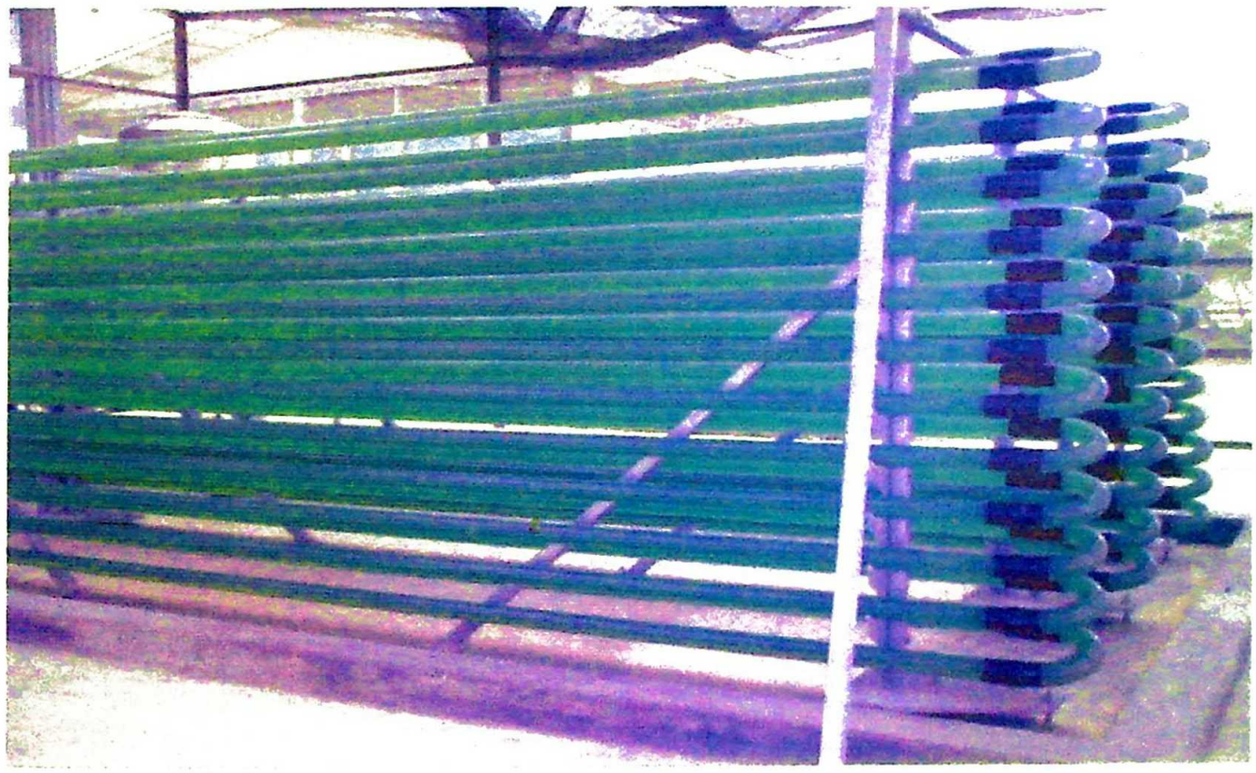
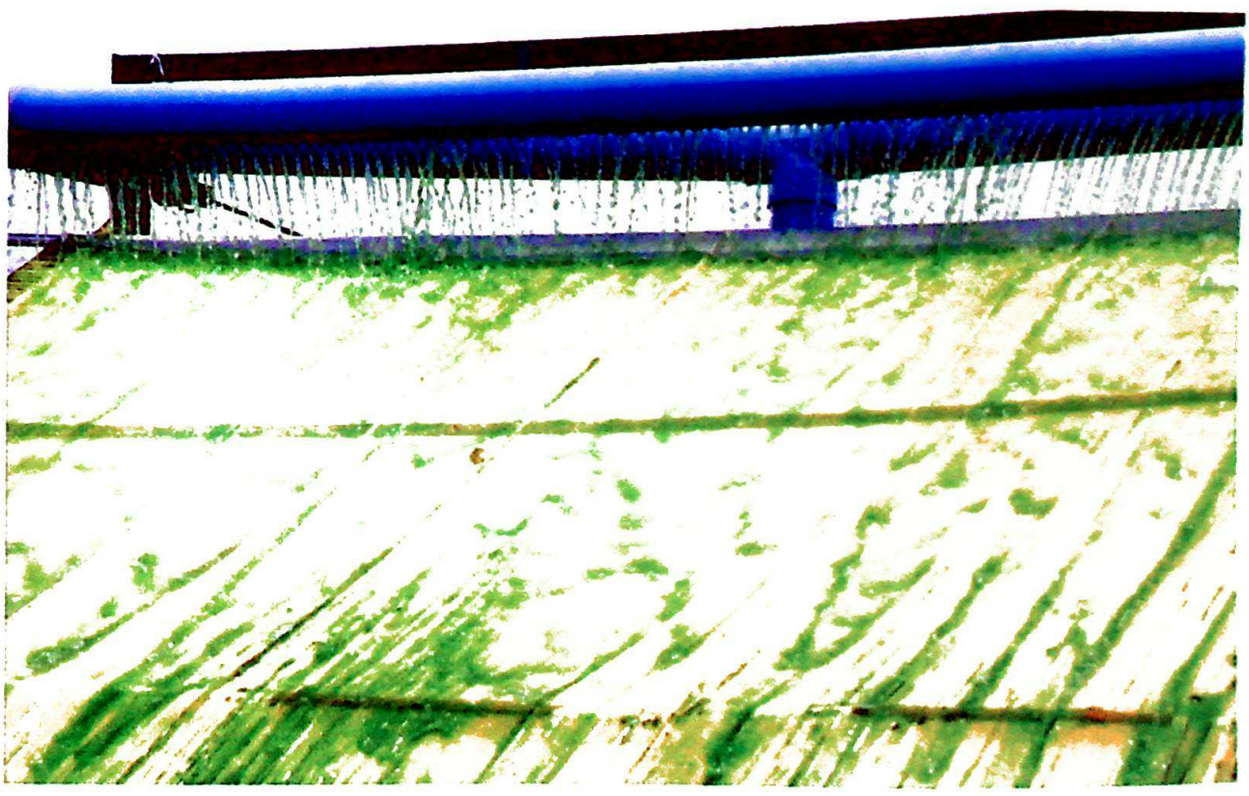


Plate A2 Inoculation Ponds (110′ x 55′ x 1.5′)

Plate A3 A Culturing Pond (420′ x 55′ x 1.5′)

Plate A4 Cascade Filter for Harvesting



APPENDIX B

Figure B1(a)

The Graph of the $rmse^*$s of BOLS and RBOLS Estimates of β_0 at Different Percentages of Outliers When n = 30

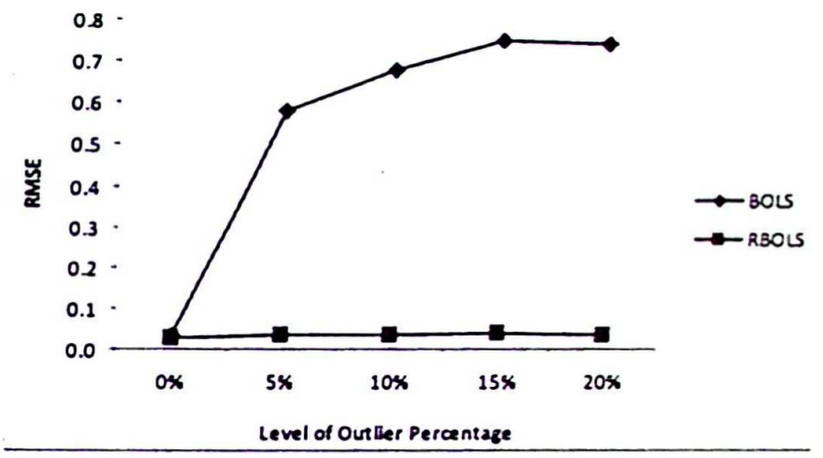


Figure B1(b)

The Graph of the $rmse^*$s of BOLS and RBOLS Estimates of β_1 at Different Percentages of Outliers When n = 30

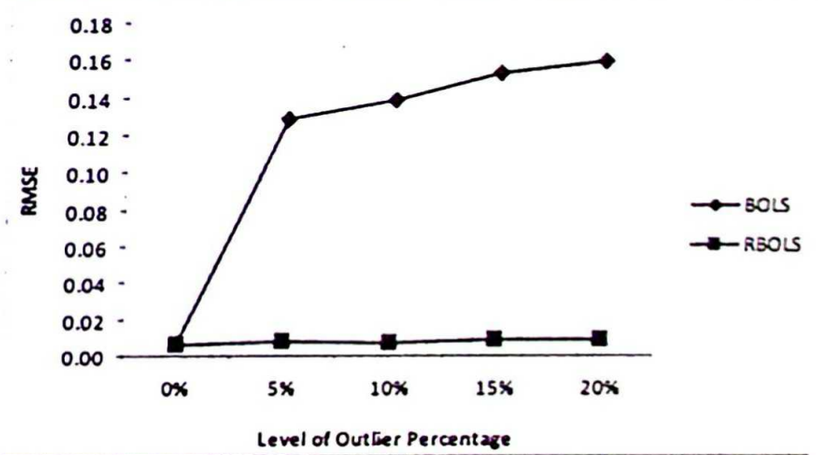


Figure B1(c)

The Graph of the $rmse^*$s of BOLS and RBOLS Estimates of β_2 at Different Percentages of Outliers When n = 30

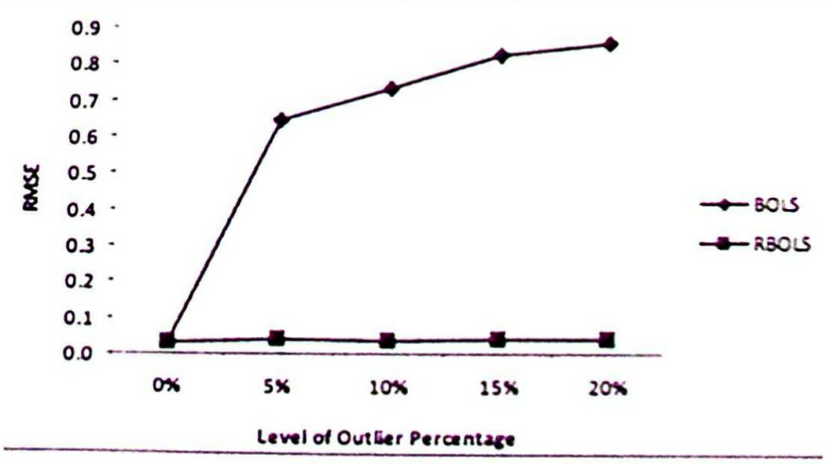


Figure B2(a)

The Graph of the $rmse^*$s of BOLS and RBOLS Estimates of β_0 at Different Percentages of Outliers When n = 60

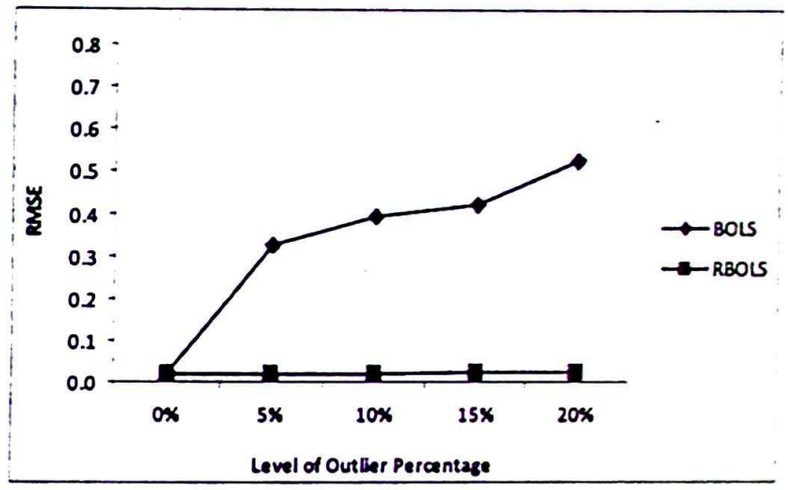


Figure B2(b)

The Graph of the $rmse^*$s of BOLS and RBOLS Estimates of β_1 at Different Percentages of Outliers When n = 60

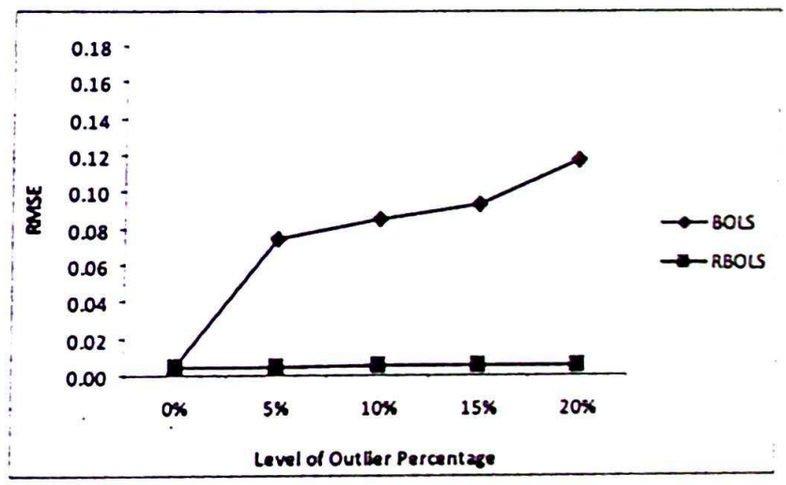


Figure B2(c)

The Graph of the $rmse^*$s of BOLS and RBOLS Estimates of β_2 at Different Percentages of Outliers When n = 60

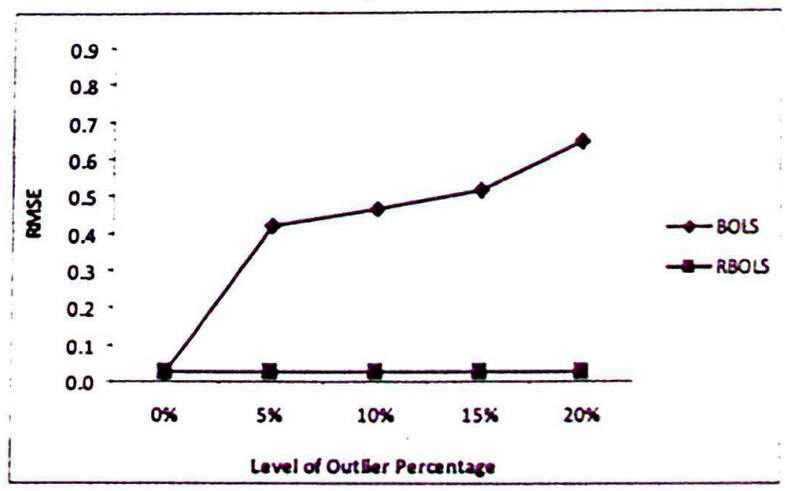


Figure B3(a)

The Graph of the $rmse^*$s of BOLS and RBOLS Estimates of β_0 at Different Percentages of Outliers When n = 100

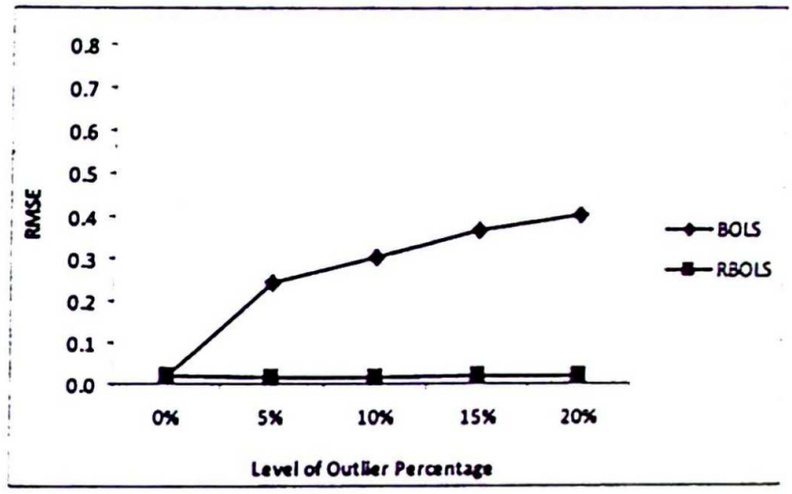


Figure B3(b)

The Graph of the $rmse^*$s of BOLS and RBOLS Estimates of β_1 at Different Percentages of Outliers When n = 100

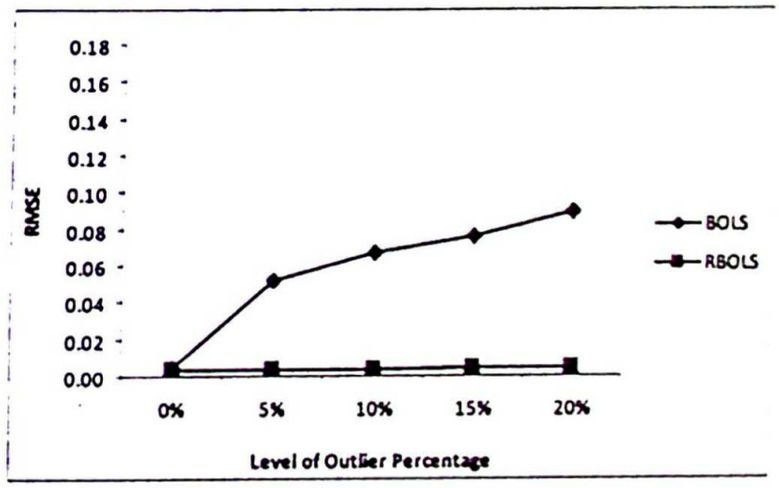


Figure B3(c)

The Graph of the $rmse^*$s of BOLS and RBOLS Estimates of β_2 at Different Percentages of Outliers When n = 100

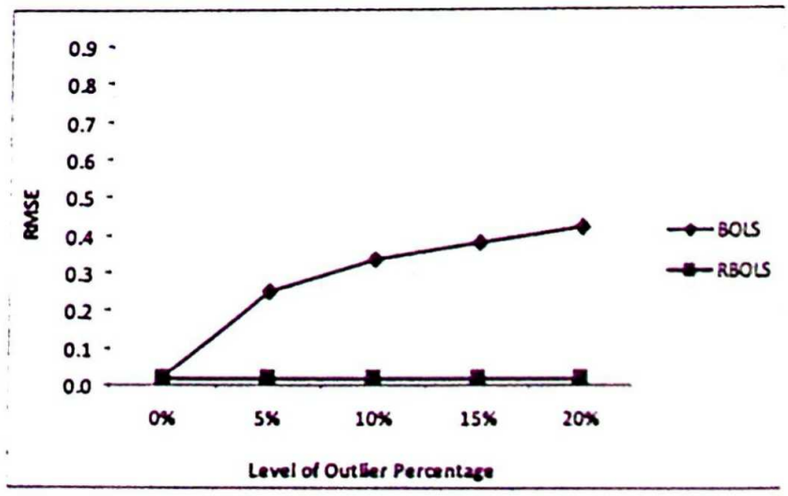


Figure B4(a)

The Graph of the $rmse^{*}$s of BOLS and RBOLS Estimates of β_0 at Different Percentages of Outliers When n = 200

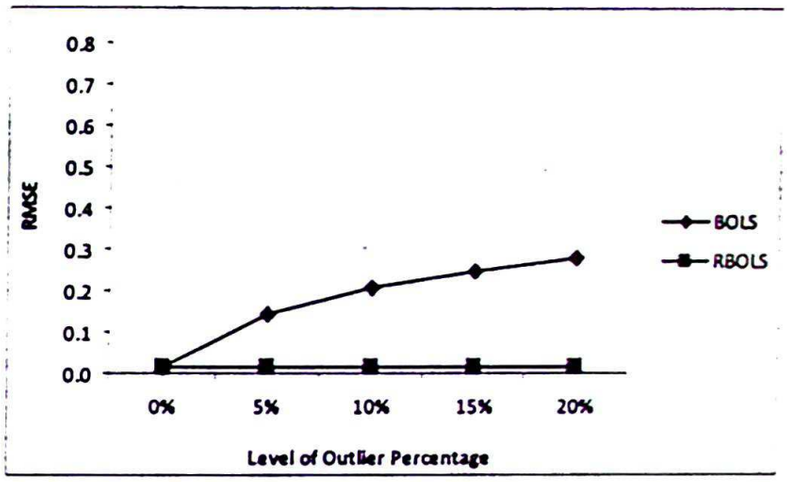


Figure B4(b)

The Graph of the $rmse^{*}$s of BOLS and RBOLS Estimates of β_1 at Different Percentages of Outliers When n = 200

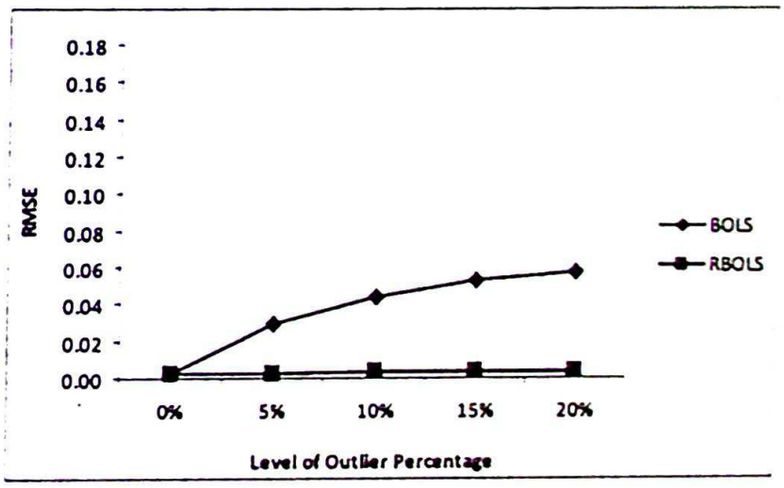


Figure B4(c)

The Graph of the $rmse^{*}$s of BOLS and RBOLS Estimates of β_2 at Different Percentages of Outliers When n = 200

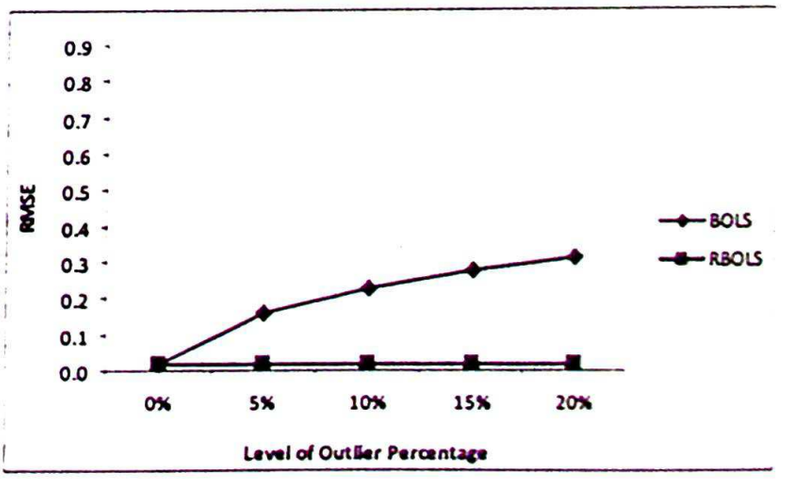


Figure B5(a)

The Graph of the $rmse^*$s of BOLS and RBOLS Estimates of α_0 at Different Percentages of Outliers When n = 30

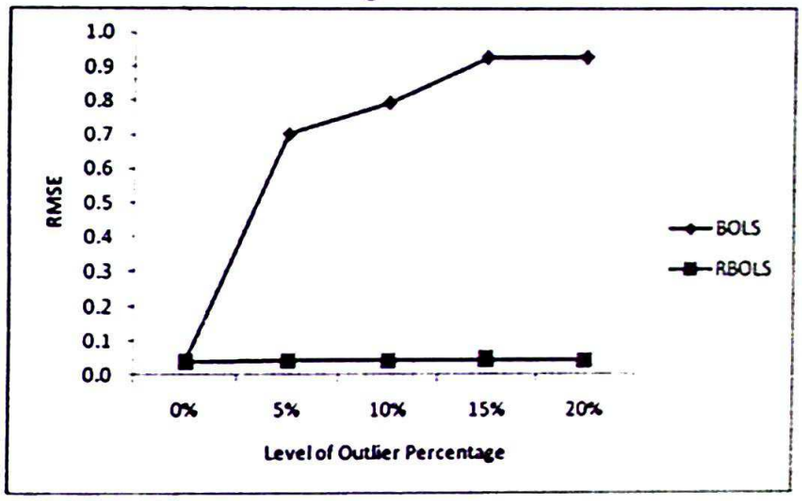


Figure B5(b)

The Graph of the $rmse^*$s of BOLS and RBOLS Estimates of α_1 at Different Percentages of Outliers When n = 30

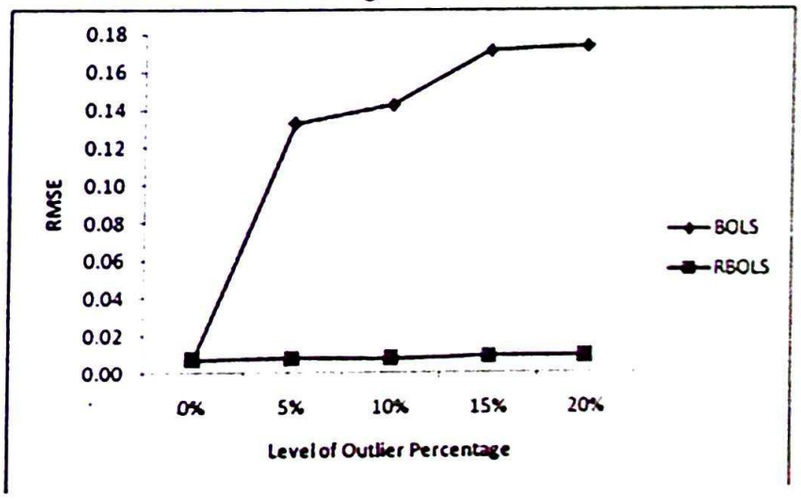


Figure B5(c)

The Graph of the $rmse^*$s of BOLS and RBOLS Estimates of α_2 at Different Percentages of Outliers When n = 30

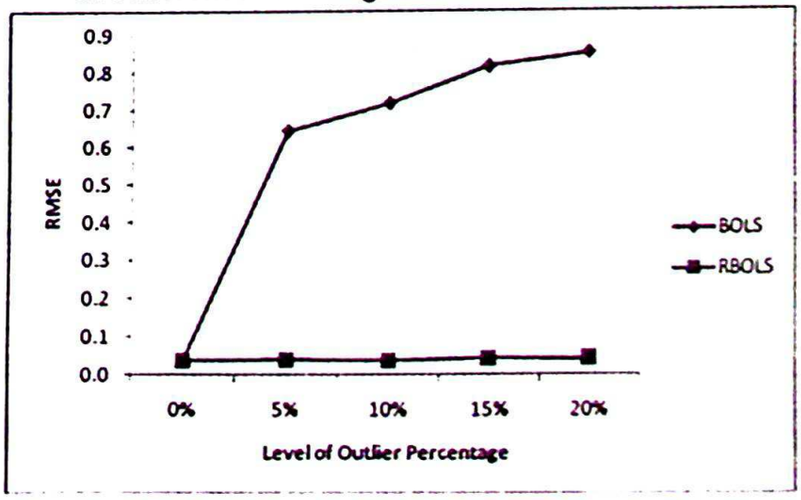


Figure B5(d)

The Graph of the $rmse^*$s of BOLS and RBOLS Estimates of α_3 at Different Percentages of Outliers When n = 30

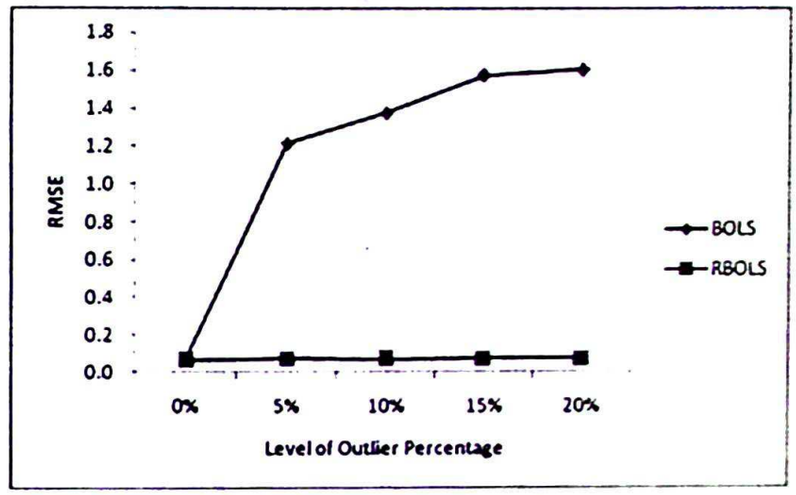


Figure B6(a)

The Graph of the $rmse^*$s of BOLS and RBOLS Estimates of α_0 at Different Percentages of Outliers When n = 60

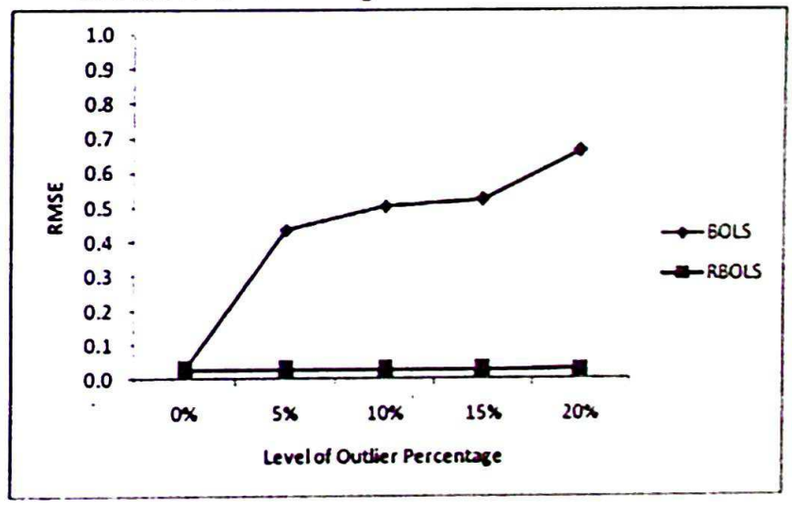


Figure B6(b)

The Graph of the $rmse^*$s of BOLS and RBOLS Estimates of α_1 at Different Percentages of Outliers When n = 60

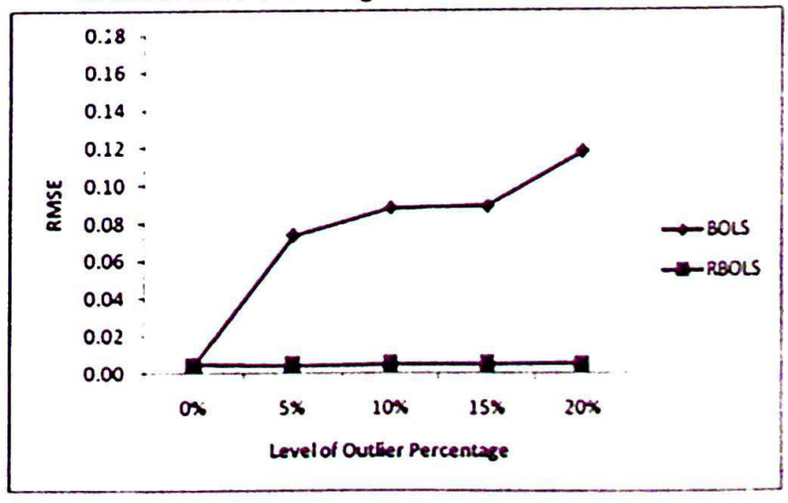


Figure B6(c)

The Graph of the $rmse^*$s of BOLS and RBOLS Estimates of α_2 at Different Percentages of Outliers When n = 60

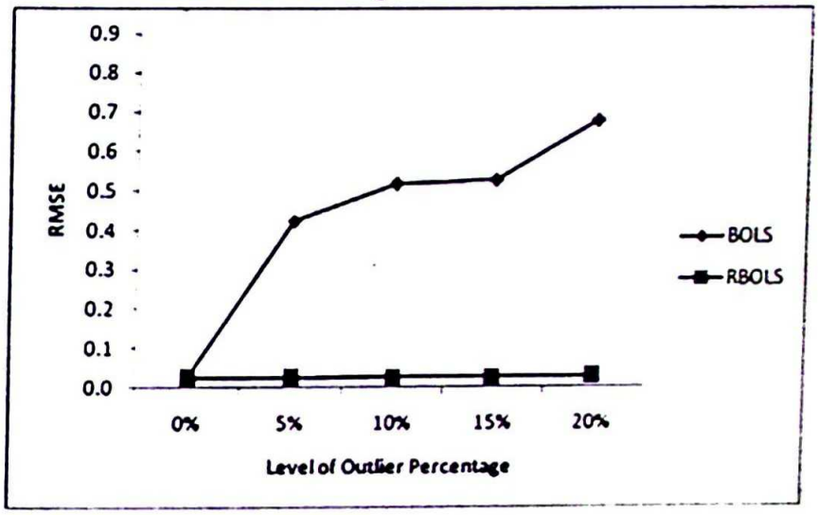


Figure B6(d)

The Graph of the $rmse^*$s of BOLS and RBOLS Estimates of α_3 at Different Percentages of Outliers When n = 60

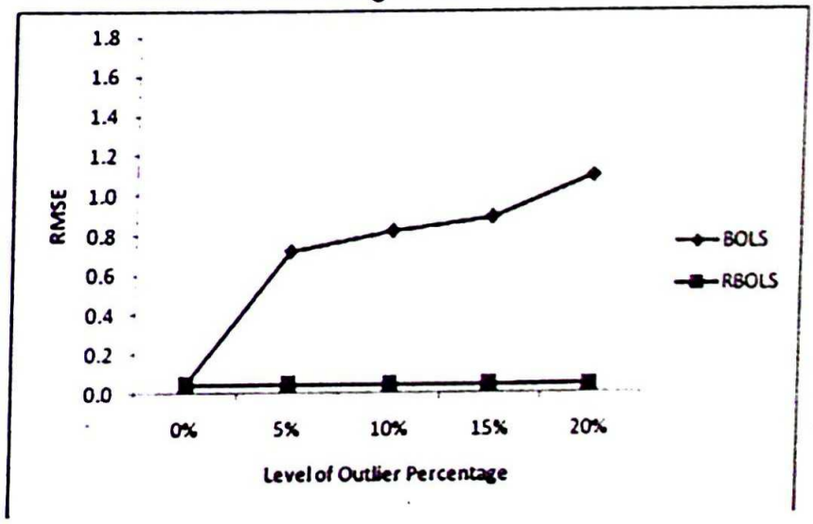


Figure B7(a)

The Graph of the $rmse^*$s of BOLS and RBOLS Estimates of α_0 at Different Percentages of Outliers When n = 100

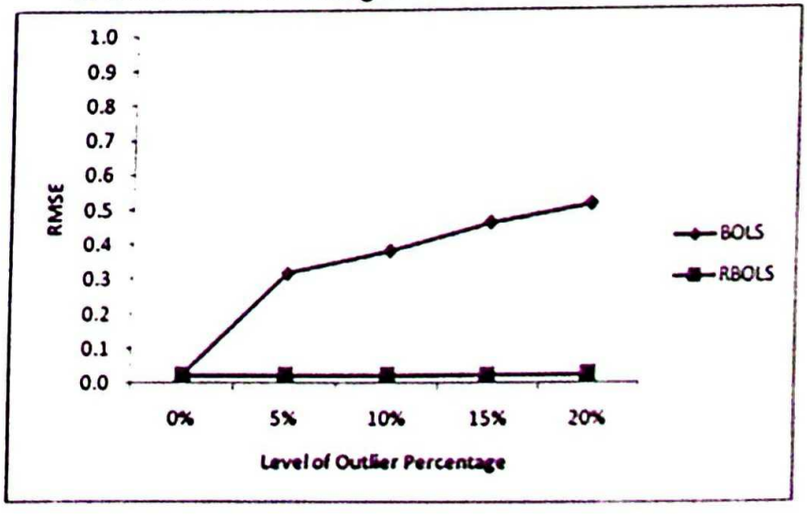


Figure B7(b)

The Graph of the $rmse^*$s of BOLS and RBOLS Estimates of α_1 at Different Percentages of Outliers When n = 100

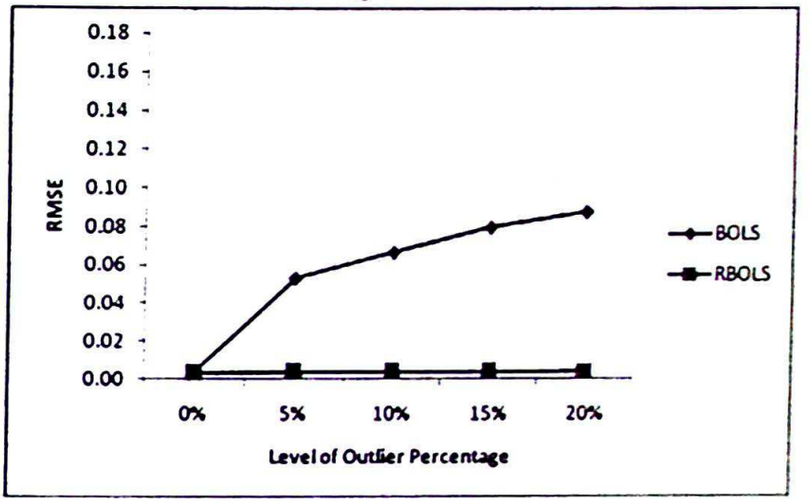


Figure B7(c)

The Graph of the $rmse^*$s of BOLS and RBOLS Estimates of α_2 at Different Percentages of Outliers When n = 100

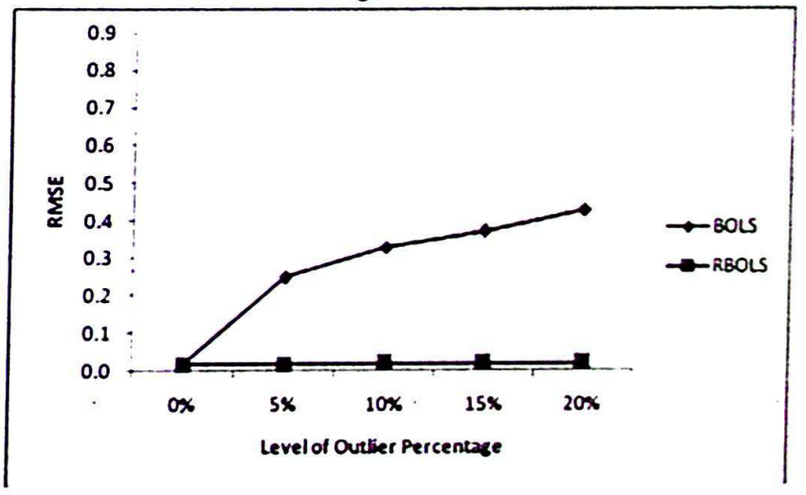


Figure B7(d)

The Graph of the $rmse^*$s of BOLS and RBOLS Estimates of α_3 at Different Percentages of Outliers When n = 100

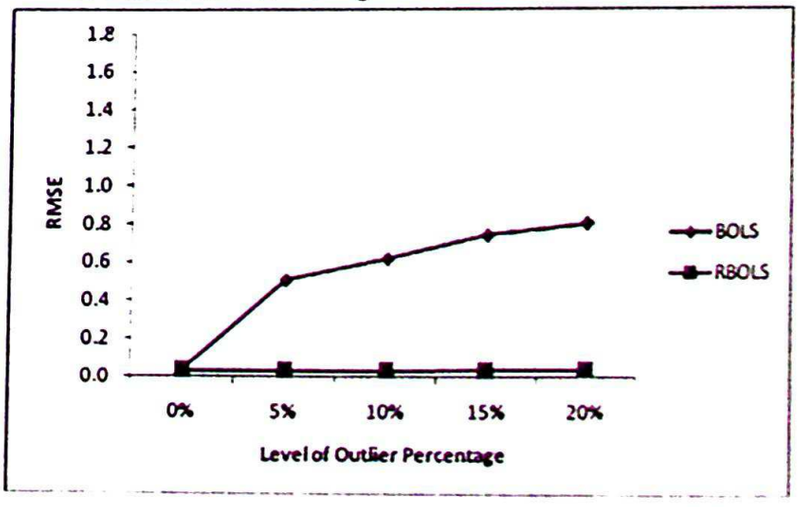


Figure B8(a)

The Graph of the $rmse^*$s of BOLS and RBOLS Estimates of α_0 at Different Percentages of Outliers When n = 200

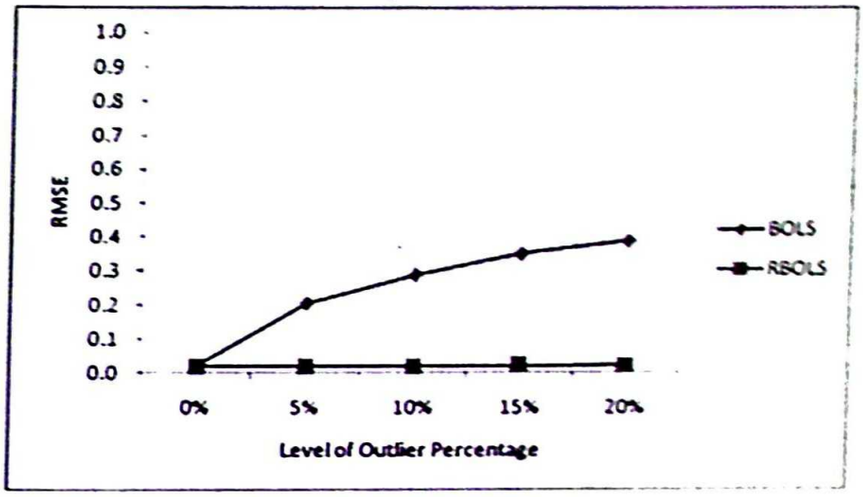


Figure B8(b)

The Graph of the $rmse^*$s of BOLS and RBOLS Estimates of α_1 at Different Percentages of Outliers When n = 200

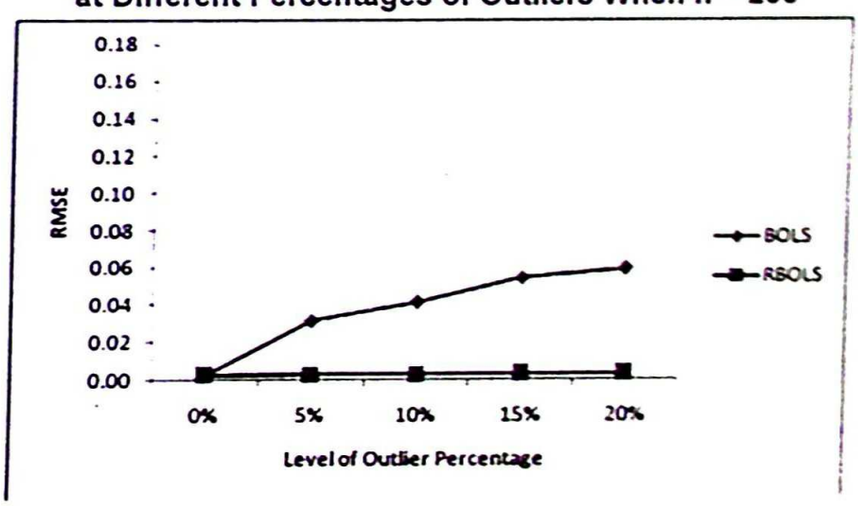


Figure B8(c)

The Graph of the $rmse^*$s of BOLS and RBOLS Estimates of α_2 at Different Percentages of Outliers When n = 200

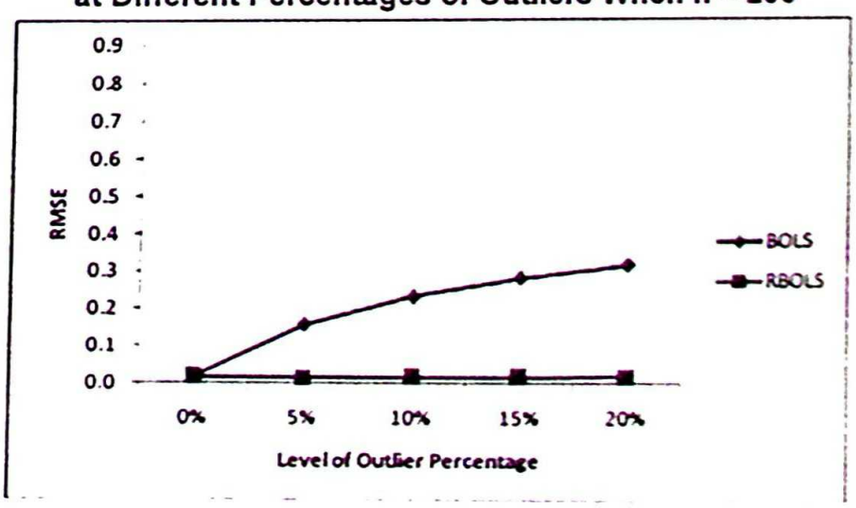
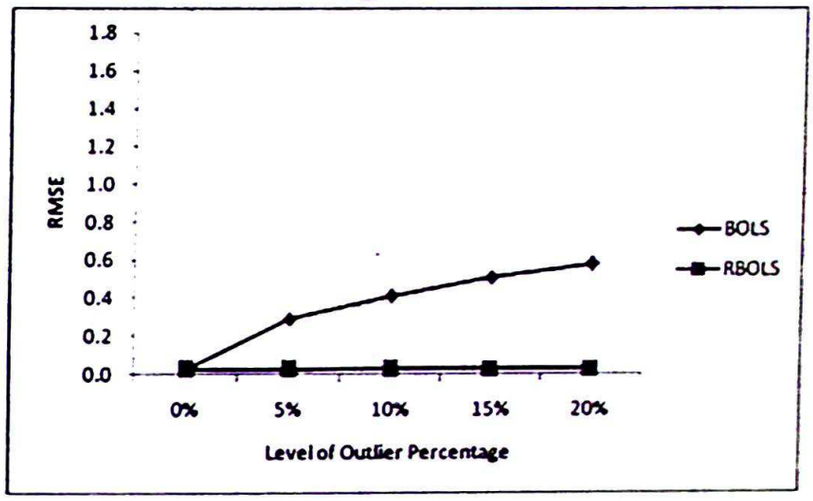


Figure B8(d)

The Graph of the $rmse^{*}$s of BOLS and RBOLS Estimates of α_3 at Different Percentages of Outliers When n = 200



APPENDIX C

Table C1

Data on the Optical Density of Spirulina and Related Variables at a Culturing Pond in 2007

Day	Optical Density *(OD)* *(680 nanometer)*	pH *(PH)*	Salinity *(SAL)* *(part : per thousand)*	Air Temperature *(TEM)* *(centigrade)*	Light *(LIG)* *(watt per (meter)^2)*	Harvest *(DHAR)* (1=yes,0=no)
1	0.19	9.9	2	19	80.2	1
2	0.18	9.8	2	22	88.2	0
3	0.49	10.3	2	18	92.7	1
4	0.30	10.3	2	20	93.5	1
5	0.29	10.4	2	24	113.2	1
6	0.28	10.0	2	24	87.5	0
7	0.30	10.3	2	22	103.3	0
8	0.36	10.4	2	22	107.2	0
9	0.36	10.6	2	21	90.7	1
10	0.22	10.9	2	22	106.3	0
11	0.24	10.7	2	23	115.3	0
12	0.26	10.8	2	24	118.3	0
13	0.39	10.1	2	24	116.8	1
14	0.32	10.1	2	26	119.5	0
15	0.33	10.7	4	25	100.8	0
16	0.43	10.8	3	26	106.3	0
17	0.46	10.9	3	26	98.7	0
18	0.48	11.0	3	26	95.4	1
19	0.28	11.0	2	21	96.1	0
20	0.29	10.9	3	21	99.4	0
21	0.39	10.9	3	20	95.2	1
22	0.35	11.1	2	20	100.7	1
23	0.34	10.8	2	20	99.5	0
24	0.36	10.5	1	20	90.1	0
25	0.36	10.4	2	21	90.9	1
26	0.34	10.5	1	21	102.1	0
27	0.41	11.1	2	22	106.7	0
28	0.42	11.2	2	22	96.8	0
29	0.46	11.5	1	22	85.7	1
30	0.33	11.6	2	22	56.2	1
31	0.23	11.3	1	22	92.4	0
32	0.23	9.0	1	18	35.5	0
33	0.37	9.9	1	18	37.6	0
34	0.39	10.1	1	22	95.5	1
35	0.31	10.0	1	20	71.4	1
36	0.26	10.0	1	20	66.3	0
37	0.31	9.7	2	24	79.3	0
38	0.48	10.0	3	21	54.0	0
39	0.54	10.1	3	22	89.6	0
40	0.65	10.4	3	20	46.5	1

Table C1 Contd.
Data on the Optical Density of Spirulina and Related Variables at a Culturing Pond in 2007

Day	Optical Density *(OD)* *(680 nanometer)*	pH *(PH)*	Salinity *(SAL)* *(part per thousand)*	Air Temperature *(TEM)* *(centigrade)*	Light *(LIG)* *(watt per (meter)^2)*	Harvest *(DHAR)* (1=yes,0=no)
41	0.33	10.3	3	24	114.7	1
42	0.30	10.2	3	24	105.5	1
43	0.25	10.1	3	24	100.2	1
44	0.15	11.2	3	24	105.6	0
45	0.29	10.0	3	23	99.9	0
46	0.29	10.0	3	22	105.2	0
47	0.33	9.8	2	22	77.6	1
48	0.23	9.7	2	23	96.3	0
49	0.27	9.6	2	24	79.5	0
50	0.33	9.8	2	26	63.6	1
51	0.23	10.0	2	24	79.5	0
52	0.27	9.9	2	22	85.7	0
53	0.31	9.8	2	24	58.6	0
54	0.31	9.7	2	24	83.2	0
55	0.40	9.7	3	24	65.4	0
56	0.40	9.7	3	24	75.5	0
57	0.41	9.9	3	24	91.2	1
58	0.39	10.1	3	26	95.4	0
59	0.44	10.0	4	26	88.2	1
60	0.36	9.7	4	28	82.6	1
61	0.31	9.7	4	26	75.4	1
62	0.25	9.7	4	27	74.5	1
63	0.21	9.7	4	26	81.6	0
64	0.25	9.6	2	28	79.3	1
65	0.21	9.8	3	25	70.4	0
66	0.28	9.8	2	23	80.8	1
67	0.21	9.5	2	23	84.2	0
68	0.32	9.4	2	24	89.5	1
69	0.17	9.8	2	26	86.3	0
70	0.22	10.0	2	27	80.3	0
71	0.25	10.0	2	28	73.5	0
72	0.29	10.0	2	29	73.7	1
73	0.25	10.2	2	26	69.1	0
74	0.28	9.8	2	28	92.4	0
75	0.28	9.8	4	29	95.6	0
76	0.28	9.5	4	28	90.3	0
77	0.34	9.9	2	27	87.5	1
78	0.18	10.0	2	26	61.5	0
79	0.30	10.5	2	26	71.6	0
80	0.35	10.1	2	29	84.7	0
81	0.35	10.1	2	30	112.6	1

Table C1 Contd.
Data on the Optical Density of Spirulina and Related Variables at a Culturing Pond in 2007

Day	Optical Density *(OD)* *(680 nanometer)*	pH *(PH)*	Salinity *(SAL)* *(part · per thousand)*	Air Temperature *(TEM)* *(centigrade)*	Light *(LIG)* *(watt per (meter)^2)*	Harvest *(DHAR)* (1=yes,0=no)
82	0.13	9.9	2	28	94.2	0
83	0.40	9.9	2	28	97.5	0
84	0.40	10.0	2	30	98.6	0
85	0.40	10.3	3	32	103.8	1
86	0.35	10.3	3	31	92.5	1
87	0.25	10.2	5	30	83.3	1
88	0.14	10.0	1	30	68.4	0
89	0.14	10.0	1	30	89.4	0
90	0.14	10.0	1	32	92.2	0
91	0.20	10.5	4	30	70.3	0
92	0.20	10.5	4	30	62.1	1
93	0.07	10.5	4	30	82.3	0
94	0.21	10.5	3	34	81.3	0
95	0.27	10.3	3	31	87.3	1
96	0.25	10.3	3	31	87.6	0
97	0.27	10.5	3	32	92.0	0
98	0.28	10.4	3	32	99.1	0
99	0.29	10.5	3	30	94.3	1
100	0.28	10.5	3	22	32.2	1
101	0.20	10.2	1	30	98.3	0
102	0.25	10.2	2	31	100.6	0
103	0.28	10.3	2	31	102.2	0
104	0.35	10.5	2	31	91.5	0
105	0.38	10.3	2	32	115.6	1
106	0.30	10.3	2	30	101.4	0
107	0.32	10.4	2	29	92.5	1
108	0.31	10.4	2	28	83.3	0
109	0.33	10.4	2	28	98.4	1
110	0.28	10.5	2	28	100.2	0
111	0.30	10.2	2	30	112.4	0
112	0.31	10.1	3	29	88.6	1
113	0.18	10.4	5	30	106.5	0
114	0.18	10.2	5	28	108.5	0
115	0.36	10.2	7	37	112.3	0
116	0.39	10.3	7	34	107.1	1
117	0.37	10.0	7	33	111.8	0
118	0.37	10.0	10	30	74.3	0
119	0.40	10.1	7	31	90.8	0
120	0.42	10.2	8	31	99.2	1
121	0.39	10.1	10	35	100.1	0
122	0.44	10.2	10	37	106.5	0

Table C1 Contd.
Data on the Optical Density of Spirulina and Related Variables at a Culturing Pond in 2007

Day	Optical Density *(OD)* *(680 nanometer)*	pH *(PH)*	Salinity *(SAL)* *(part · per thousand)*	Air Temperature *(TEM)* *(centigrade)*	Light *(LIG)* *(watt per (meter)^2)*	Harvest *(DHAR)* (1=yes,0=no)
123	0.45	10.2	10	34	104.5	0
124	0.52	10.0	7	31	79.0	0
125	0.56	9.9	7	37	118.9	1
126	0.48	9.9	8	35	120.6	1
127	0.23	9.8	10	32	119.1	0
128	0.30	10.0	5	30	54.9	0
129	0.34	10.0	5	27	56.9	0
130	0.46	10.1	4	29	97.0	1
131	0.36	10.2	5	31	68.8	0
132	0.37	10.3	4	30	38.7	1
133	0.35	10.1	4	33	101.4	0
134	0.38	10.1	4	34	105.2	1
135	0.30	10.2	4	31	84.4	0
136	0.33	10.1	4	28	22.9	1
137	0.25	10.2	4	32	92.4	1
138	0.17	10.3	3	28	18.5	0
139	0.32	10.4	4	25	24.3	1
140	0.28	10.4	2	26	24.5	1
141	0.12	10.4	1	27	28.7	0
142	0.13	10.3	1	25	9.8	0
143	0.23	10.1	2	29	58.0	1
144	0.15	10.1	2	31	113.9	0
145	0.21	10.2	2	29	38.5	0
146	0.24	10.3	3	32	107.9	1
147	0.20	10.3	5	33	110.3	1
148	0.17	10.5	4	34	108.0	0
149	0.27	10.6	4	33	115.7	1
150	0.25	10.4	4	33	113.3	0
151	0.32	10.2	4	33	110.4	1
152	0.3	10.1	4	34	123.7	0
153	0.32	10.1	4	34	113.8	1
154	0.3	10.2	4	34	112.5	1
155	0.21	10.4	5	31	67.3	0
156	0.37	9.9	6	33	109.4	1
157	0.15	10.1	3	33	121.1	0
158	0.22	10.2	2	32	81.3	0
159	0.24	10.0	2	32	59.9	1
160	0.23	10.3	2	34	62.4	0
161	0.24	10.2	2	32	60.4	0
162	0.25	10.1	5	32	234.8	1
163	0.23	9.8	5	28	83.0	0

Table C1 Contd.
Data on the Optical Density of Spirulina and Related Variables at a Culturing Pond in 2007

Day	Optical Density *(OD)* *(680 nanometer)*	pH *(PH)*	Salinity *(SAL)* *(part per thousand)*	Air Temperature *(TEM)* *(centigrade)*	Light *(LIG)* *(watt per (meter)^2)*	Harvest *(DHAR)* (1=yes,0=no)
164	0.23	9.8	5	26	21.7	1
165	0.17	10.0	2	32	159.4	0
166	0.21	9.9	3	31	146.8	0
167	0.21	10.0	1	32	172.3	1
168	0.13	10.1	2	32	158.3	0
169	0.13	10.1	2	33	220.2	1
170	0.1	10.0	3	33	215.8	0
171	0.17	10.1	5	36	258.6	0
172	0.26	10.0	5	34	255.1	0
173	0.29	10.0	4	34	290.7	1
174	0.23	9.8	3	35	308.3	0
175	0.23	9.8	3	34	290.7	0
176	0.24	9.8	3	35	265.2	1
177	0.19	9.9	3	34	253.8	1
178	0.17	9.9	3	33	257.3	0
179	0.25	9.9	2	32	216.6	0
180	0.27	10.0	5	34	325.5	0
181	0.59	9.9	4	32	276.9	1
182	0.18	9.8	5	33	265.8	0
183	0.36	9.6	4	34	231.9	1
184	0.23	9.7	3	31	159.6	0
185	0.30	9.7	4	31	99.5	0
186	0.30	9.7	2	32	280.2	0
187	0.42	9.6	2	29	75.7	0
188	0.42	9.6	2	30	76.3	1
189	0.40	9.7	2	30	75.3	1
190	0.39	9.7	2	31	133.6	0
191	0.42	10.0	3	34	245.9	1
192	0.34	10.1	3	33	248.7	1
193	0.23	10.3	2	32	123.4	0
194	0.24	10.2	3	34	248.5	0
195	0.28	10.5	2	35	275.5	0
196	0.28	10.5	2	33	252.3	0
197	0.32	10.5	2	32	177.0	1
198	0.26	10.2	3	32	249.6	0
199	0.31	10.0	3	33	206.7	0
200	0.31	10.0	3	30	223.1	0
201	0.31	10.0	2	27	36.1	1
202	0.27	10.0	2	26	49.2	0
203	0.27	10.0	3	30	215.2	1
204	0.20	10.0	2	31	262.2	0

Table C1 Contd.
Data on the Optical Density of Spirulina and Related Variables at a Culturing Pond in 2007

Day	Optical Density *(OD)* *(680 nanometer)*	pH *(PH)*	Salinity *(SAL)* *(part per thousand)*	Air Temperature *(TEM)* *(centigrade)*	Light *(LIG)* *(watt per (meter)^2)*	Harvest *(DHAR)* (1=yes,0=no)
205	0.29	9.9	4	28	59.7	1
206	0.22	9.8	3	30	182.8	1
207	0.18	9.9	3	30	160.1	0
208	0.20	9.8	2	26	37.8	0
209	0.20	9.8	2	30	129.5	1
210	0.16	9.9	3	30	130.3	0
211	0.16	9.9	3	32	192.1	0
212	0.19	9.5	3	31	142.8	0
213	0.24	10.1	2	33	260.0	0
214	0.29	9.9	3	32	190.9	1
215	0.24	10.1	2	29	141.9	0
216	0.29	9.9	3	32	205.7	0
217	0.3	10.0	2	32	230.2	0
218	0.31	10.5	1	33	260.7	0
219	0.31	10.2	2	35	260.1	1
220	0.22	10.2	2	33	99.8	0
221	0.22	10.3	2	35	260.3	0
222	0.27	10.5	3	33	240.9	0
223	0.29	10.1	2	31	261.8	0
224	0.32	9.9	2	30	223.3	0
225	0.36	9.7	2	30	81.0	1
226	0.29	9.8	3	30	101.5	1
227	0.26	9.3	2	31	169.9	0
228	0.34	10.8	4	31	118.4	0
229	0.35	9.9	5	29	74.6	1
230	0.33	9.9	3	31	125.2	0
231	0.35	10.2	3	32	203.5	0
232	0.37	10.2	4	33	286.5	0
233	0.45	10.1	5	32	248.9	0
234	0.53	9.9	4	34	249.5	1
235	0.33	9.8	5	30	101.3	0
236	0.36	9.9	5	30	149.2	0
237	0.4	9.8	4	30	108.2	1
238	0.37	9.7	3	31	201.1	0
239	0.38	9.7	3	32	249.6	1
240	0.21	9.8	3	34	281.6	0
241	0.28	9.7	5	29	94.1	0
242	0.29	9.8	3	33	278.1	0
243	0.29	9.8	5	30	108.5	1
244	0.27	9.5	2	33	252.3	0
245	0.27	9.8	3	32	175.5	0

Table C1 Contd.
Data on the Optical Density of Spirulina and Related Variables at a Culturing Pond in 2007

Day	Optical Density *(OD)* *(680 nanometer)*	pH *(PH)*	Salinity *(SAL)* *(part per thousand)*	Air Temperature *(TEM)* *(centigrade)*	Light *(LIG)* *(watt per (meter)^2)*	Harvest *(DHAR)* (1=yes,0=no)
246	0.39	9.8	3	33	79.8	0
247	0.45	9.7	3	33	248.0	0
248	0.61	9.8	2	32	164.7	1
249	0.37	9.8	4	32	122.4	0
250	0.37	9.8	3	29	145.0	0
251	0.37	9.8	3	32	148.7	1
252	0.33	9.8	3	33	180.2	1
253	0.32	9.8	3	33	193.8	1
254	0.26	9.8	2	34	279.6	1
255	0.22	9.8	3	32	126.0	0
256	0.22	9.8	1	34	235.6	0
257	0.31	9.6	1	32	195.1	1
258	0.25	9.4	1	28	52.6	1
259	0.23	9.4	1	30	120.7	1
260	0.10	8.6	1	33	250.1	0
261	0.12	8.5	1	32	121.9	1
262	0.10	8.9	1	36	260.6	0
263	0.15	9.0	1	34	235.4	1
264	0.13	8.9	1	34	119.8	0
265	0.14	9.2	1	33	190.5	0
266	0.15	9.3	1	32	188.5	0
267	0.17	8.4	1	29	64.0	0
268	0.19	8.8	1	34	192.3	0
269	0.20	8.7	1	34	213.8	0
270	0.25	9.1	1	33	162.9	0
271	0.30	9.4	1	33	247.8	1
272	0.25	9.2	1	31	116.9	0
273	0.27	9.5	1	31	159.4	0
274	0.29	9.8	1	33	276.5	0
275	0.48	9.8	1	33	284.6	1
276	0.28	9.7	1	38	242.8	0
277	0.33	9.7	1	36	235.9	1
278	0.23	9.7	1	35	140.5	0
279	0.39	9.6	1	32	138.6	0
280	0.39	9.6	1	32	125.6	1
281	0.38	9.9	1	32	114.0	1
282	0.34	9.9	1	33	235.3	0
283	0.35	9.6	1	34	233.0	0
284	0.35	9.5	1	32	184.5	0
285	0.41	9.5	1	31	152.6	1
286	0.33	9.6	1	32	144.7	1

Table C1 Contd.
Data on the Optical Density of Spirulina and Related Variables at a Culturing Pond in 2007

Day	Optical Density *(OD) (680 nanometer)*	pH *(PH)*	Salinity *(SAL) (part per thousand)*	Air Temperature *(TEM) (centigrade)*	Light *(LIG) (watt per (meter)^2)*	Harvest *(DHAR)* (1=yes,0=no)
287	0.30	9.6	1	32	160.0	1
288	0.29	9.5	1	31	100.3	0
289	0.30	9.5	1	30	8.4	1
290	0.22	9.4	1	30	12.6	0
291	0.23	9.6	1	31	12.4	0
292	0.35	9.3	1	32	99.8	1
293	0.24	9.3	1	31	105.0	1
294	0.23	9.3	1	31	110.0	1
295	0.21	9.4	1	30	101.4	1
296	0.17	9.4	1	30	100.1	1
297	0.15	9.3	1	30	100.7	0
298	0.18	9.3	1	30	99.8	0
299	0.18	9.3	1	31	131.4	0
300	0.18	9.3	1	31	129.0	1
301	0.12	9.3	2	31	118.9	0
302	0.13	9.1	2	31	121.0	1
303	0.06	9.2	2	31	113.3	0
304	0.09	9.2	2	30	128.9	0
305	0.26	9.9	2	31	130.9	1
306	0.15	9.5	1	30	131.0	0
307	0.20	9.7	1	29	128.9	0
308	0.21	9.9	1	29	130.6	1
309	0.17	10.1	1	28	108.0	0
310	0.25	10.0	2	27	50.3	0
311	0.36	9.9	1	29	72.8	1
312	0.21	9.9	1	31	112.3	0
313	0.25	9.9	1	31	102.3	0
314	0.36	9.9	4	31	100.7	1
315	0.21	9.9	1	31	95.4	0
316	0.25	9.9	1	31	84.0	0
317	0.26	9.8	1	30	76.3	0
318	0.28	9.8	1	29	39.3	1
319	0.21	9.8	1	23	22.2	0
320	0.24	9.5	2	25	28.4	0
321	0.30	9.2	1	25	14.8	1
322	0.25	9.9	1	26	30.0	1
323	0.20	9.6	1	27	34.0	0
324	0.25	9.0	1	29	105.1	1
325	0.23	9.8	1	29	101.6	0
326	0.40	10.3	2	29	99.7	1
327	0.29	10.2	2	28	75.2	0

Table C1 Contd.
Data on the Optical Density of Spirulina and Related Variables at a Culturing Pond in 2007

Day	Optical Density *(OD)* *(680 nanometer)*	pH *(PH)*	Salinity *(SAL)* *(part per thousand)*	Air Temperature *(TEM)* *(centigrade)*	Light *(LIG)* *(watt per (meter)^2)*	Harvest *(DHAR)* (1=yes,0=no)
328	0.29	10.2	2	27	62.5	0
329	0.29	9.5	1	26	59.6	0
330	0.36	9.2	1	27	49.0	1
331	0.32	9.3	1	28	59.7	0
332	0.38	9.3	1	27	47.9	0
333	0.39	9.5	1	20	8.0	0
334	0.46	9.5	1	26	67.6	1
335	0.42	9.5	2	26	78.6	1
336	0.40	9.5	1	26	70.4	0
337	0.44	9.3	1	27	78.4	0
338	0.44	9.3	2	27	80.4	0
339	0.44	9.3	2	28	83.1	0
340	0.59	9.4	2	25	87.7	1
341	0.50	9.9	1	27	94.3	1
342	0.46	9.4	2	26	92.4	0
343	0.46	9.7	2	26	88.4	0
344	0.49	9.5	2	26	93.0	0
345	0.54	9.5	2	25	89.9	1
346	0.49	9.6	2	25	85.1	0
347	0.54	9.3	2	26	101.3	1
348	0.51	9.3	2	22	103.0	0
349	0.53	9.3	2	25	108.3	1
350	0.51	9.3	2	23	100.4	0
351	0.51	9.5	2	24	111.6	0
352	0.51	9.5	2	20	16.4	1
353	0.47	9.5	2	27	131.75	1
354	0.44	9.5	3	25	108.04	0
355	0.46	9.3	3	26	112.36	0
356	0.46	9.3	3	22	121.84	0
357	0.47	9.2	3	21	109.35	0
358	0.47	9.3	3	22	115.65	0
359	0.47	9.4	3	21	109.65	0
360	0.47	9.0	3	21	98.71	1
361	0.45	9.2	2	22	129.2	1
362	0.42	9.2	2	22	108.32	1
363	0.37	9.4	2	22	104.12	0
364	0.41	9.4	2	17	88.44	0
365	0.46	9.4	2	20	95.65	1

Source: Myanma Spirulina Factory.

Table C2
95% One-Day-Ahead Bootstrap Prediction Intervals
for the Optical Density of Spirulina Using the Dynamic Regression Model

Sr. No.	OD_t	SAL_t	DW_t	$DHAR_t$	Forecast OD_{t+1}	95% prediction interval	
						Lower limit	Upper limit
1	0.06	0.2	0	0	0.128	0.033	0.235
2	0.07	0.4	0	0	0.140	0.044	0.246
3	0.09	0.2	0	0	0.153	0.058	0.260
4	0.10	0.1	0	0	0.160	0.063	0.267
5	0.10	0.3	0	0	0.164	0.070	0.270
6	0.12	0.1	0	0	0.177	0.081	0.284
7	0.12	0.1	0	1	0.075	0.000	0.178
8	0.12	0.2	0	0	0.179	0.084	0.285
9	0.13	0.1	0	0	0.185	0.090	0.292
10	0.13	0.2	0	0	0.188	0.092	0.294
11	0.13	0.2	0	1	0.086	0.000	0.188
12	0.14	0.1	0	0	0.194	0.099	0.301
13	0.15	0.1	0	0	0.203	0.107	0.309
14	0.15	0.1	0	1	0.101	0.007	0.203
15	0.15	0.1	1	0	0.218	0.124	0.326
16	0.15	0.2	0	0	0.205	0.109	0.311
17	0.15	0.3	0	0	0.207	0.113	0.312
18	0.15	0.3	1	0	0.222	0.128	0.329
19	0.16	0.3	0	0	0.215	0.121	0.321
20	0.17	0.1	0	0	0.220	0.125	0.326
21	0.17	0.1	0	1	0.118	0.023	0.220
22	0.17	0.1	1	0	0.235	0.142	0.343
23	0.17	0.2	0	0	0.222	0.126	0.327
24	0.17	0.3	0	0	0.224	0.130	0.329
25	0.17	0.4	0	0	0.226	0.133	0.332
26	0.17	0.5	0	0	0.228	0.132	0.335
27	0.18	0.1	0	0	0.228	0.134	0.335
28	0.18	0.1	0	1	0.127	0.032	0.228
29	0.18	0.2	0	0	0.230	0.135	0.336
30	0.18	0.2	1	0	0.246	0.153	0.354
31	0.18	0.3	0	0	0.233	0.138	0.337
32	0.18	0.5	0	0	0.237	0.141	0.344
33	0.19	0.1	0	0	0.237	0.143	0.343
34	0.19	0.2	1	1	0.153	0.055	0.257
35	0.19	0.3	0	0	0.241	0.147	0.346
36	0.19	0.3	0	1	0.140	0.043	0.242
37	0.20	0.1	0	0	0.245	0.152	0.352
38	0.20	0.1	1	0	0.261	0.167	0.368
39	0.20	0.2	0	0	0.248	0.153	0.353
40	0.20	0.2	0	1	0.146	0.051	0.248
41	0.20	0.4	0	0	0.252	0.158	0.358
42	0.20	0.4	0	1	0.150	0.054	0.254
43	0.20	0.5	0	1	0.152	0.055	0.257

Table C2 Contd.
95% One-Day-Ahead Bootstrap Prediction Intervals for the Optical Density of Spirulina Using the Dynamic Regression Model

Sr. No.	OD_t	SAL_t	DW_t	$DHAR_t$	Forecast OD_{t+1}	95% prediction interval	
						Lower limit	Upper limit
44	0.21	0.1	0	1	0.153	0.058	0.253
45	0.21	0.1	1	0	0.269	0.176	0.377
46	0.21	0.1	1	1	0.168	0.071	0.273
47	0.21	0.2	0	0	0.256	0.162	0.361
48	0.21	0.3	0	0	0.258	0.164	0.363
49	0.21	0.4	0	0	0.260	0.167	0.366
50	0.21	0.5	0	0	0.262	0.167	0.369
51	0.22	0.1	0	0	0.263	0.169	0.369
52	0.22	0.2	0	0	0.265	0.171	0.370
53	0.22	0.2	1	0	0.280	0.188	0.387
54	0.22	0.3	0	0	0.267	0.173	0.371
55	0.22	0.3	0	1	0.165	0.069	0.268
56	0.23	0.1	0	0	0.271	0.178	0.377
57	0.23	0.1	0	1	0.170	0.075	0.270
58	0.23	0.1	1	0	0.287	0.193	0.394
59	0.23	0.2	0	0	0.273	0.179	0.379
60	0.23	0.2	0	1	0.172	0.077	0.274
61	0.23	0.2	1	0	0.289	0.196	0.395
62	0.23	0.3	0	0	0.275	0.181	0.379
63	0.23	0.5	0	0	0.280	0.185	0.386
64	0.23	0.5	0	1	0.178	0.081	0.283
65	0.23	1.0	0	0	0.290	0.188	0.400
66	0.24	0.1	0	1	0.178	0.084	0.279
67	0.24	0.2	0	0	0.282	0.188	0.387
68	0.24	0.2	0	1	0.180	0.086	0.282
69	0.24	0.2	1	0	0.297	0.205	0.404
70	0.24	0.3	0	0	0.284	0.190	0.388
71	0.24	0.3	0	1	0.183	0.087	0.285
72	0.25	0.1	0	0	0.288	0.195	0.394
73	0.25	0.1	0	1	0.187	0.093	0.287
74	0.25	0.1	1	0	0.304	0.210	0.411
75	0.25	0.1	1	1	0.202	0.106	0.307
76	0.25	0.2	0	0	0.290	0.197	0.396
77	0.25	0.2	0	1	0.189	0.095	0.291
78	0.25	0.2	1	0	0.306	0.213	0.412
79	0.25	0.3	0	0	0.293	0.199	0.397
80	0.25	0.3	1	1	0.207	0.110	0.310
81	0.25	0.4	0	0	0.295	0.201	0.401
82	0.25	0.4	0	1	0.193	0.096	0.297
83	0.25	0.5	0	1	0.195	0.098	0.300
84	0.26	0.1	1	0	0.312	0.219	0.420
85	0.26	0.2	0	0	0.299	0.206	0.404
86	0.26	0.2	0	1	0.198	0.103	0.299

Table C2 Contd.
95% One-Day-Ahead Bootstrap Prediction Intervals
for the Optical Density of Spirulina Using the Dynamic Regression Model

Sr. No.	OD_t	SAL_t	DW_t	$DHAR_t$	Forecast	95% prediction interval	
					OD_{t+1}	Lower limit	Upper limit
87	0.26	0.2	1	0	0.314	0.222	0.420
88	0.26	0.2	1	1	0.213	0.117	0.316
89	0.26	0.3	0	0	0.301	0.208	0.405
90	0.26	0.5	0	0	0.305	0.211	0.412
91	0.27	0.1	0	0	0.306	0.213	0.411
92	0.27	0.2	0	0	0.308	0.215	0.413
93	0.27	0.2	1	0	0.323	0.230	0.429
94	0.27	0.3	0	0	0.310	0.217	0.414
95	0.27	0.3	0	1	0.208	0.113	0.311
96	0.27	0.4	0	1	0.210	0.113	0.314
97	0.27	0.5	0	0	0.314	0.219	0.420
98	0.28	0.1	0	0	0.314	0.221	0.420
99	0.28	0.1	1	1	0.228	0.133	0.332
100	0.28	0.2	0	0	0.316	0.224	0.421
101	0.28	0.2	0	1	0.215	0.120	0.317
102	0.28	0.2	1	0	0.332	0.239	0.437
103	0.28	0.3	0	0	0.318	0.226	0.422
104	0.28	0.3	0	1	0.217	0.121	0.319
105	0.28	0.4	0	0	0.320	0.227	0.426
106	0.28	0.5	0	0	0.322	0.228	0.429
107	0.29	0.1	0	0	0.323	0.230	0.428
108	0.29	0.1	1	0	0.338	0.246	0.445
109	0.29	0.2	0	0	0.325	0.232	0.429
110	0.29	0.2	0	1	0.223	0.128	0.325
111	0.29	0.2	1	0	0.340	0.247	0.446
112	0.29	0.2	1	1	0.239	0.143	0.342
113	0.29	0.3	0	0	0.327	0.234	0.431
114	0.29	0.3	0	1	0.225	0.130	0.328
115	0.29	0.3	1	0	0.342	0.249	0.448
116	0.29	0.4	0	1	0.228	0.131	0.330
117	0.29	0.5	0	1	0.230	0.132	0.334
118	0.30	0.1	0	1	0.230	0.135	0.330
119	0.30	0.1	1	1	0.245	0.150	0.349
120	0.30	0.2	0	0	0.333	0.241	0.438
121	0.30	0.2	1	0	0.349	0.256	0.454
122	0.30	0.2	1	1	0.247	0.152	0.350
123	0.30	0.3	1	1	0.249	0.154	0.353
124	0.30	0.4	0	0	0.338	0.244	0.444
125	0.30	0.4	0	1	0.236	0.139	0.339
126	0.30	0.5	0	0	0.340	0.246	0.446
127	0.31	0.1	0	0	0.340	0.248	0.445
128	0.31	0.1	0	1	0.238	0.144	0.339
129	0.31	0.1	1	1	0.254	0.159	0.357

Table C2 Contd.
95% One-Day-Ahead Bootstrap Prediction Intervals for the Optical Density of Spirulina Using the Dynamic Regression Model

Sr. No.	OD_t	SAL_t	DW_t	$DHAR_t$	Forecast OD_{t+1}	95% prediction interval	
						Lower limit	Upper limit
130	0.31	0.2	0	0	0.342	0.250	0.447
131	0.31	0.2	0	1	0.240	0.145	0.342
132	0.31	0.2	1	0	0.357	0.264	0.462
133	0.31	0.3	0	0	0.344	0.251	0.448
134	0.31	0.3	0	1	0.243	0.147	0.345
135	0.31	0.4	0	1	0.245	0.148	0.347
136	0.32	0.1	1	0	0.364	0.272	0.470
137	0.32	0.2	0	0	0.350	0.259	0.456
138	0.32	0.2	0	1	0.249	0.154	0.351
139	0.32	0.2	1	0	0.366	0.273	0.471
140	0.32	0.3	0	1	0.251	0.156	0.353
141	0.32	0.4	0	1	0.253	0.157	0.356
142	0.33	0.1	0	1	0.256	0.161	0.356
143	0.33	0.2	0	1	0.258	0.162	0.360
144	0.33	0.2	1	1	0.273	0.178	0.376
145	0.33	0.3	0	0	0.361	0.269	0.466
146	0.33	0.3	0	1	0.260	0.165	0.362
147	0.33	0.3	1	1	0.275	0.180	0.378
148	0.33	0.4	0	1	0.262	0.165	0.364
149	0.33	0.4	1	0	0.379	0.286	0.486
150	0.33	0.5	0	0	0.365	0.272	0.471
151	0.34	0.1	0	0	0.366	0.274	0.471
152	0.34	0.1	1	0	0.381	0.289	0.487
153	0.34	0.2	0	1	0.266	0.171	0.368
154	0.34	0.2	1	0	0.383	0.291	0.488
155	0.34	0.3	0	1	0.268	0.173	0.370
156	0.34	0.4	0	0	0.372	0.280	0.478
157	0.34	0.5	0	0	0.374	0.280	0.480
158	0.35	0.1	0	0	0.374	0.283	0.480
159	0.35	0.1	0	1	0.273	0.179	0.373
160	0.35	0.2	0	0	0.376	0.284	0.482
161	0.35	0.2	0	1	0.275	0.180	0.377
162	0.35	0.2	1	1	0.290	0.196	0.393
163	0.35	0.3	0	0	0.378	0.286	0.483
164	0.35	0.3	0	1	0.277	0.182	0.379
165	0.35	0.4	0	0	0.380	0.288	0.487
166	0.35	0.5	0	1	0.281	0.183	0.386
167	0.36	0.1	1	0	0.398	0.307	0.504
168	0.36	0.1	1	1	0.297	0.203	0.400
169	0.36	0.2	0	1	0.283	0.188	0.386
170	0.36	0.2	1	0	0.400	0.309	0.505
171	0.36	0.2	1	1	0.299	0.205	0.401
172	0.36	0.4	0	1	0.288	0.191	0.390

Table C2 Contd.
95% One-Day Ahead Bootstrap Prediction Intervals for the Optical Density of Spirulina Using the Dynamic Regression Model

Sr. No.	OD_t	SAL_t	DW_t	$DHAR_t$	Forecast	95% prediction interval	
					OD_{t+1}	Lower limit	Upper limit
173	0.36	0.4	1	1	0.303	0.207	0.407
174	0.36	0.5	0	0	0.391	0.297	0.497
175	0.36	0.7	0	0	0.395	0.301	0.504
176	0.37	0.1	1	0	0.407	0.316	0.512
177	0.37	0.2	1	0	0.409	0.318	0.514
178	0.37	0.3	0	0	0.395	0.304	0.500
179	0.37	0.3	0	1	0.294	0.199	0.396
180	0.37	0.4	0	0	0.398	0.306	0.504
181	0.37	0.4	0	1	0.296	0.199	0.399
182	0.37	0.6	0	1	0.300	0.201	0.407
183	0.37	0.7	0	0	0.404	0.310	0.512
184	0.37	1.0	0	0	0.410	0.310	0.520
185	0.38	0.1	0	1	0.298	0.205	0.399
186	0.38	0.1	1	0	0.415	0.325	0.521
187	0.38	0.2	0	1	0.300	0.206	0.403
188	0.38	0.3	0	1	0.303	0.207	0.404
189	0.38	0.4	0	1	0.305	0.208	0.407
190	0.39	0.1	0	0	0.408	0.318	0.514
191	0.39	0.1	0	1	0.307	0.214	0.408
192	0.39	0.1	1	0	0.424	0.334	0.529
193	0.39	0.1	1	1	0.322	0.229	0.425
194	0.39	0.2	0	0	0.410	0.320	0.516
195	0.39	0.2	1	1	0.324	0.231	0.427
196	0.39	0.3	0	0	0.413	0.322	0.518
197	0.39	0.3	1	0	0.428	0.336	0.534
198	0.39	0.3	1	1	0.327	0.232	0.429
199	0.39	0.7	0	1	0.320	0.220	0.428
200	0.39	1.0	0	0	0.427	0.328	0.537
201	0.40	0.1	1	0	0.432	0.342	0.538
202	0.40	0.2	0	0	0.419	0.328	0.525
203	0.40	0.2	0	1	0.318	0.223	0.420
204	0.40	0.2	1	1	0.333	0.239	0.435
205	0.40	0.3	0	1	0.320	0.224	0.421
206	0.40	0.3	1	0	0.437	0.345	0.543
207	0.40	0.4	0	1	0.322	0.225	0.425
208	0.40	0.7	0	0	0.430	0.335	0.538
209	0.41	0.1	0	1	0.324	0.231	0.425
210	0.41	0.2	1	0	0.443	0.353	0.549
211	0.41	0.3	1	1	0.344	0.249	0.447
212	0.42	0.2	0	0	0.436	0.346	0.542
213	0.42	0.2	0	1	0.335	0.240	0.437
214	0.42	0.2	1	0	0.452	0.361	0.557
215	0.42	0.2	1	1	0.350	0.256	0.452

Table C2 Contd.
95% One-Day Ahead Bootstrap Prediction Intervals for the Optical Density of Spirulina Using the Dynamic Regression Model

Sr. No.	OD_t	SAL_t	DW_t	$DHAR_t$	Forecast OD_{t+1}	95% prediction interval	
						Lower limit	Upper limit
216	0.42	0.3	0	1	0.337	0.241	0.438
217	0.42	0.8	0	1	0.347	0.246	0.456
218	0.43	0.3	1	0	0.462	0.371	0.569
219	0.44	0.1	1	0	0.467	0.375	0.572
220	0.44	0.2	1	0	0.469	0.378	0.575
221	0.44	0.3	1	0	0.471	0.380	0.578
222	0.44	0.4	1	1	0.372	0.276	0.476
223	0.44	1.0	0	0	0.470	0.370	0.579
224	0.45	0.2	1	1	0.376	0.282	0.478
225	0.45	0.3	0	0	0.464	0.372	0.570
226	0.45	0.5	0	0	0.468	0.375	0.574
227	0.45	1.0	0	0	0.479	0.379	0.587
228	0.46	0.1	1	1	0.382	0.289	0.485
229	0.46	0.2	1	0	0.486	0.395	0.592
230	0.46	0.3	1	0	0.488	0.397	0.595
231	0.46	0.4	0	1	0.373	0.278	0.476
232	0.47	0.2	1	1	0.393	0.300	0.495
233	0.47	0.3	1	0	0.497	0.405	0.604
234	0.47	0.3	1	1	0.395	0.300	0.499
235	0.48	0.1	0	1	0.384	0.290	0.486
236	0.48	0.3	1	1	0.404	0.309	0.507
237	0.48	0.3	1	0	0.505	0.414	0.612
238	0.48	0.8	0	1	0.399	0.299	0.509
239	0.49	0.2	1	0	0.512	0.420	0.618
240	0.49	0.2	1	1	0.410	0.317	0.512
241	0.50	0.1	1	1	0.417	0.323	0.519
242	0.51	0.2	1	0	0.529	0.437	0.636
243	0.51	0.2	1	1	0.427	0.334	0.529
244	0.52	0.7	0	0	0.532	0.437	0.641
245	0.53	0.2	1	1	0.444	0.350	0.547
246	0.53	0.4	0	1	0.433	0.337	0.536
247	0.54	0.2	1	1	0.453	0.358	0.555
248	0.54	0.3	1	0	0.557	0.464	0.664
249	0.56	0.7	0	1	0.465	0.368	0.574
250	0.59	0.2	1	1	0.496	0.401	0.598
251	0.59	0.4	0	1	0.485	0.389	0.588
252	0.61	0.2	0	1	0.498	0.402	0.601
253	0.65	0.3	1	1	0.549	0.454	0.654

- Prediction intervals are shown in ascending order of the values of OD_t, SAL_t, DW_t and $DHAR_t$.
- For same values of OD_t, SAL_t, DW_t and $DHAR_t$, prediction interval is only shown once.

Table C3
95% One-Day-Ahead Bootstrap Prediction Intervals for the Optical Density of Spirulina Using the Linear Regression Model

Sr. No.	OD_t	SAL_t	DW_t	Forecast OD_{t+1}	95% prediction interval	
					Lower limit	Upper limit
1	0.06	0.2	0	0.098	0.062	0.157
2	0.10	0.1	0	0.132	0.097	0.190
3	0.10	0.3	0	0.138	0.104	0.196
4	0.12	0.1	0	0.151	0.116	0.209
5	0.12	0.2	0	0.154	0.119	0.212
6	0.13	0.1	0	0.160	0.125	0.218
7	0.13	0.2	0	0.163	0.129	0.221
8	0.14	0.1	0	0.169	0.135	0.227
9	0.15	0.1	0	0.179	0.144	0.236
10	0.15	0.1	1	0.188	0.152	0.245
11	0.15	0.2	0	0.182	0.147	0.240
12	0.15	0.3	0	0.185	0.151	0.243
13	0.16	0.3	0	0.194	0.160	0.252
14	0.17	0.1	0	0.197	0.163	0.255
15	0.17	0.1	1	0.207	0.171	0.265
16	0.17	0.2	0	0.200	0.167	0.258
17	0.17	0.3	0	0.203	0.169	0.261
18	0.17	0.5	0	0.209	0.175	0.267
19	0.18	0.1	0	0.207	0.173	0.264
20	0.18	0.3	0	0.213	0.179	0.270
21	0.18	0.5	0	0.219	0.184	0.276
22	0.19	0.1	0	0.216	0.182	0.274
23	0.19	0.3	0	0.222	0.188	0.279
24	0.20	0.1	0	0.225	0.192	0.283
25	0.20	0.1	1	0.235	0.199	0.292
26	0.20	0.2	0	0.228	0.195	0.286
27	0.20	0.4	0	0.234	0.201	0.292
28	0.21	0.1	1	0.244	0.209	0.302
29	0.21	0.2	0	0.238	0.204	0.296
30	0.21	0.3	0	0.241	0.208	0.298
31	0.21	0.4	0	0.244	0.210	0.301
32	0.22	0.1	0	0.244	0.210	0.302
33	0.22	0.2	0	0.247	0.214	0.305
34	0.22	0.2	1	0.256	0.221	0.314
35	0.22	0.3	0	0.250	0.217	0.307
36	0.23	0.1	1	0.263	0.228	0.320
37	0.23	0.2	0	0.256	0.223	0.314
38	0.23	0.2	1	0.266	0.230	0.323
39	0.23	0.3	0	0.259	0.226	0.317
40	0.23	0.5	0	0.265	0.231	0.322
41	0.23	1.0	0	0.280	0.239	0.343
42	0.24	0.2	0	0.266	0.232	0.323
43	0.24	0.2	1	0.275	0.240	0.333

Table C3 Contd.
95% One-Day-Ahead Bootstrap Prediction Intervals for the Optical Density of Spirulina Using the Linear Regression Model

Sr. No.	OD_t	SAL_t	DW_t	Forecast OD_{t+1}	95% prediction interval	
					Lower limit	Upper limit
44	0.24	0.3	0	0.269	0.236	0.327
45	0.25	0.1	0	0.272	0.238	0.330
46	0.25	0.1	1	0.281	0.246	0.338
47	0.25	0.2	0	0.275	0.241	0.332
48	0.25	0.3	0	0.278	0.245	0.336
49	0.25	0.4	0	0.281	0.248	0.339
50	0.26	0.1	1	0.291	0.256	0.347
51	0.26	0.2	0	0.284	0.251	0.342
52	0.26	0.3	0	0.287	0.254	0.345
53	0.26	0.5	0	0.293	0.260	0.350
54	0.27	0.1	0	0.291	0.257	0.349
55	0.27	0.2	0	0.294	0.260	0.351
56	0.27	0.2	1	0.303	0.269	0.361
57	0.27	0.3	0	0.297	0.263	0.354
58	0.28	0.1	0	0.300	0.266	0.358
59	0.28	0.2	0	0.303	0.270	0.361
60	0.28	0.2	1	0.312	0.278	0.370
61	0.28	0.3	0	0.306	0.273	0.363
62	0.28	0.4	0	0.309	0.276	0.367
63	0.28	0.5	0	0.312	0.278	0.369
64	0.29	0.1	0	0.310	0.275	0.367
65	0.29	0.1	1	0.319	0.284	0.376
66	0.29	0.2	0	0.313	0.279	0.370
67	0.29	0.2	1	0.322	0.287	0.379
68	0.29	0.3	0	0.315	0.282	0.373
69	0.29	0.3	1	0.325	0.290	0.382
70	0.30	0.2	0	0.322	0.288	0.380
71	0.30	0.2	1	0.331	0.296	0.389
72	0.30	0.4	0	0.328	0.294	0.385
73	0.30	0.5	0	0.331	0.297	0.388
74	0.31	0.1	0	0.328	0.294	0.386
75	0.31	0.2	0	0.331	0.298	0.389
76	0.31	0.2	1	0.340	0.306	0.398
77	0.31	0.3	0	0.334	0.301	0.392
78	0.32	0.1	1	0.347	0.312	0.404
79	0.32	0.2	0	0.341	0.307	0.399
80	0.32	0.2	1	0.350	0.315	0.407
81	0.33	0.3	0	0.353	0.320	0.410
82	0.33	0.5	0	0.359	0.325	0.416
83	0.34	0.1	0	0.356	0.322	0.415
84	0.34	0.1	1	0.365	0.331	0.423
85	0.34	0.2	1	0.368	0.334	0.426
86	0.34	0.4	0	0.365	0.332	0.423

Table C3 Contd.

95% One-Day-Ahead Bootstrap Prediction Intervals for the Optical Density of Spirulina Using the Linear Regression Model

Sr. No.	OD_t	SAL_t	DW_t	Forecast OD_{t+1}	95% prediction interval	
					Lower limit	Upper limit
87	0.35	0.1	0	0.366	0.331	0.424
88	0.35	0.2	0	0.369	0.335	0.427
89	0.35	0.3	0	0.372	0.339	0.429
90	0.35	0.4	0	0.374	0.341	0.432
91	0.36	0.1	1	0.384	0.349	0.442
92	0.36	0.2	1	0.387	0.353	0.445
93	0.36	0.5	0	0.387	0.353	0.444
94	0.36	0.7	0	0.393	0.357	0.451
95	0.37	0.1	1	0.393	0.359	0.451
96	0.37	0.2	1	0.396	0.362	0.454
97	0.37	0.3	0	0.390	0.357	0.448
98	0.37	0.4	0	0.393	0.360	0.450
99	0.37	0.7	0	0.402	0.367	0.461
100	0.37	1.0	0	0.411	0.371	0.473
101	0.38	0.1	1	0.403	0.368	0.461
102	0.39	0.1	0	0.403	0.367	0.461
103	0.39	0.1	1	0.412	0.377	0.470
104	0.39	0.2	0	0.406	0.371	0.464
105	0.39	0.3	0	0.409	0.376	0.467
106	0.39	0.3	1	0.418	0.384	0.476
107	0.39	1.0	0	0.430	0.390	0.492
108	0.40	0.1	1	0.421	0.386	0.479
109	0.40	0.2	0	0.415	0.381	0.473
110	0.40	0.3	1	0.427	0.393	0.485
111	0.40	0.7	0	0.430	0.394	0.489
112	0.41	0.2	1	0.434	0.399	0.491
113	0.42	0.2	0	0.434	0.399	0.492
114	0.42	0.2	1	0.443	0.408	0.501
115	0.43	0.3	1	0.455	0.421	0.513
116	0.44	0.1	1	0.459	0.423	0.517
117	0.44	0.2	1	0.462	0.426	0.519
118	0.44	0.3	1	0.465	0.430	0.522
119	0.44	1.0	0	0.476	0.436	0.538
120	0.45	0.5	0	0.471	0.436	0.529
121	0.45	1.0	0	0.486	0.446	0.548
122	0.46	0.2	1	0.480	0.445	0.538
123	0.46	0.3	1	0.483	0.448	0.541
124	0.47	0.3	1	0.493	0.457	0.550
125	0.48	0.3	1	0.502	0.466	0.559
126	0.49	0.2	1	0.509	0.472	0.566
127	0.51	0.2	1	0.527	0.491	0.585
128	0.52	0.7	0	0.542	0.504	0.602

- Prediction intervals are shown in ascending order of the values of OD_t, SAL_t and DW_t.
- For same values of OD_t, SAL_t and DW_t, prediction interval is only shown once.