REGRESSION MODEL WITH CENSORED OBSERVATIONS

By

SUPOL DURONGWATANA

Bachelor of Arts
Chulalongkorn University
Bangkok, Thailand
1979

Master of Science
Oklahoma State University
Stillwater, Oklahoma
1983

Submitted to the Faculty of the Graduate College
of the Oklahoma State University
in partial fulfillment of the requirements
for the Degree of
DOCTOR OF PHILOSOPHY
December, 1985

REGRESSION MODEL WITH CENSORED OBSERVATIONS

Thesis Approved:

W. M. Warde
Thesis Adviser

Nunnkhopadhy

Ronald W. McNew

Glenn J. Knowles

Norman N. Durham
Dean of the Graduate College

ACKNOWLEDGMENTS

I wish to express my deep appreciation to both the former and the present members of my dissertation committee.

In writing this dissertation, I am greatly indebted to Dr. William H. Stewart, my first major advisor, for not only providing inspiration and encouragement at every stage of this study but also giving me indispensable assistance and proofreading my dissertation.

I would like to thank Dr. Donald Holbert for his valuable guidance and support while serving as my second major advisor. A sincere appreciation goes to my present major advisor, Dr. William D. Warde, who took the trouble of reading and giving detailed comments which I greatly appreciate. The inspiration of his encouragement and his exceptional insight have been of special value to me.

I wish to acknowledge my two committee members--Dr. Ronald W. McNew, Dr. Nitis Mukhopadhyay--and my outside committee member, Dr. Glenn J. Knowles, for their advice.

I also take this opportunity to express my gratitude to Dr. John L. Folks for granting me the assistantship during my M.S. and Ph.D. program. Special thanks go to Mrs. Janet Sallee, for her excellent typing skills and to Ms. Naengnoi Promsuwansiri and Ms. Phanida Suthamchai for their help in proofreading. I also thank my Thai friends and colleagues at OSU whose friendly support encouraged me to pursue the Ph.D. program.

Finally, I would like to express my deepest gratitude to my parents and my two brothers for their love, care, and moral support which have

been a continual source of inspiration for me through the course of this study and throughout my education. I dedicate this dissertation to them.

TABLE OF CONTENTS

Chapter		Page
I.	INTRODUCTION.	1
	Model.	6
II.	REVIEW OF LITERATURE.	11
III.	REGRESSION METHODS FOR CENSORED OBSERVATIONS.	21
	Adjusted Method.	21
	3.1 Introduction	21
	3.2 Kaplan-Meier Estimation.	24
	Kaplan-Meier Estimation of Distribution Function Censored Observations	26
	3.3 Adjustment of Regression Model With Censored Observations.	28
	3.4 Diagram of the Adjusted Method for Censored Observations.	33
	Bootstrapping Method	35
	3.5 Bootstrapping for Censored Data.	35
	3.6 The Bootstrap Estimate of Bias	38
	3.7 Bootstrapping Regression Model With Censored Observations.	39
	3.8 Diagram of Bootstrapping Simple Linear Regression for Censored Observations.	42
	3.9 Generalization of Multiple Linear Regression Methods for Censored Observations Under the Same Assumptions as the Previous Sections	45
	Algorithm for the Adjusted Method.	45
	Algorithm for the Bootstrapping Method	48
IV.	COMPUTER RESULTS.	53
	4.1 Design of the Simulation Study.	53
	Assessing Performance	54
	4.2 Results of the Simulation Study	55
	4.3 Heart Transplant Data	110
V.	SUMMARY AND CONCLUSIONS	115
	5.1 Application	115
	5.2 Result Conclusions.	116
	5.3 Further Work.	117

Chapter | Page

A SELECTED BIBLIOGRAPHY. 118

APPENDIX A - A PROGRAM FOR THE ADJUSTED METHOD 122

APPENDIX B - A PROGRAM FOR THE BOOTSTRAPPING METHOD. 125

APPENDIX C - A PROGRAM FOR BUCKLEY AND JAMES' METHOD 128

APPENDIX D - A PROGRAM FOR MILLER'S METHOD 132

APPENDIX E - STANFORD HEART TRANSPLANT DATA. 135

LIST OF TABLES

Table		Page
I.	Simulations Calculating the Estimates of α and β Based on 100 Replications ($\alpha = 1$, $\beta = 0.2$, $C_i \sim U(10 + \alpha, 50 + \alpha)$, $X_i = 2i$, $\varepsilon_i \sim N(0,1)$ and Sample Size = 10) . .	58
II.	Simulations Calculating the Estimates of α and β Based on 100 Replications ($\alpha = 1$, $\beta = 0.2$, $C_i \sim U(10 + \alpha, 50 + \alpha)$, $X_i = 2i$, $\varepsilon_i \sim N(0,1)$ and Sample Size = 20) . .	59
III.	Simulations Calculating the Estimates of α and β Based on 100 Replications ($\alpha = 1$, $\beta = 0.2$, $C_i \sim U(10 + \alpha, 50 + \alpha)$, $X_i = 2i$, $\varepsilon_i \sim N(0,1)$ and Sample Size = 30) . .	60
IV.	Simulations Calculating the Estimates of α and β Based on 100 Replications ($\alpha = 1$, $\beta = 0.2$, $C_i = 1.5X_i - 0.015\, X_i^2 + \alpha$, $X_i = 2i$, $\varepsilon_i \sim N(0,1)$ and Sample Size = 10)	62
V.	Simulations Calculating the Estimates of α and β Based on 100 Replications ($\alpha = 1$, $\beta = 0.2$, $C_i = 1.5X_i - 0.015X_i^2 + \alpha$, $X_i = 2i$, $\varepsilon_i \sim N(0,1)$ and Sample Size = 20)	63
VI.	Simulations Calculating the Estimates of α and β Based on 100 Replications ($\alpha = 1$, $\beta = 0.2$, $C_i = 1.5X_i - 0.015X_i^2 + \alpha$, $X_i = 2i$, $\varepsilon_i \sim N(0,1)$ and Sample Size = 30)	64
VII.	Simulations Calculating the Estimates of α and β Based on 100 Replications ($\alpha = 1$, $\beta = 0.2$, $C_i = 31$, $X_i = 2i$, $\varepsilon_i \sim N(0,1)$ and Sample Size = 10)	66
VIII.	Simulations Calculating the Estimates of α and β Based on 100 Replications ($\alpha = 1$, $\beta = 0.2$, $C_i = 31$, $X_i = 2i$, $\varepsilon_i \sim N(0,1)$ and Sample Size = 20)	67
IX.	Simulations Calculating the Estimates of α and β Based on 100 Replications ($\alpha = 1$, $\beta = 0.2$, $C_i = 31$, $X_i = 2i$, $\varepsilon_i \sim N(0,1)$ and Sample Size = 30)	68
X.	Simulations Calculating the Estimates of α and β Based on 100 Replications ($\alpha = 1$, $\beta = 0.2$, $C_i \sim U(\alpha + \beta X_i, \alpha + \beta X_i + 20)$, $X_i = 2i$, $\varepsilon_i \sim N(0,1)$ and Sample Size = 10) .	70

Table		Page
XI.	Simulations Calculating the Estimates of α and β Based on 100 Replications ($\alpha = 1$, $\beta = 0.2$, $C_i \sim U(\alpha + \beta X_i, \alpha + \beta X_i + 20)$, $X_i = 2i$, $\varepsilon_i \sim N(0,1)$ and Sample Size = 20)	71
XII.	Simulations Calculating the Estimates of α and β Based on 100 Replications ($\alpha = 1$, $\beta = 0.2$, $C_i \sim U(\alpha + \beta X_i, \alpha + \beta X_i + 20)$, $X_i = 2i$, $\varepsilon_i \sim N(0,1)$ and Sample Size = 30)	72
XIII.	Simulations Calculating the Estimates of α and β Based on 100 Replications ($\alpha = 1$, $\beta = 0.2$, $C_i \sim U(0,50)$, $X_i = 2i$, $\varepsilon_i \sim N(0,1)$ and Sample Size = 10)	74
XIV.	Simulations Calculating the Estimates of α and β Based on 100 Replications ($\alpha = 1$, $\beta = 0.2$, $C_i \sim U(0,50)$, $X_i = 2i$, $\varepsilon_i \sim N(0,1)$ and Sample Size = 20)	75
XV.	Simulations Calculating the Estimates of α and β Based on 100 Replications ($\alpha = 1$, $\beta = 0.2$, $C_i \sim U(0,50)$, $X_i = 2i$, $\varepsilon_i \sim N(0,1)$ and Sample Size = 30)	76
XVI.	Simulations Calculating the Estimates of α and β Based on 100 Replications ($\alpha = 1$, $\beta = 0.4$, $C_i \sim N(\alpha + 40, 16)$, $X_i = 2i$, $\varepsilon_i \sim N(0,1)$ and Sample Size = 10)	78
XVII.	Simulations Calculating the Estimates of α and β Based on 100 Replications ($\alpha = 1$, $\beta = 0.4$, $C_i \sim N(\alpha + 40, 16)$, $X_i = 2i$, $\varepsilon_i \sim N(0,1)$ and Sample Size = 20)	79
XVIII.	Simulations Calculating the Estimates of α and β Based on 100 Replications ($\alpha = 1$, $\beta = 0.4$, $C_i \sim N(\alpha + 40, 16)$, $X_i = 2i$, $\varepsilon_i \sim N(0,1)$ and Sample Size = 30)	80
XIX.	Simulations Calculating the Estimates of α and β Based on 100 Replications ($\alpha = 10$, $\beta = 0.2$, $C_i \sim U(\alpha + \beta X_i, \alpha + \beta X_i + 40)$, $X_i = 2i$, $\varepsilon_i \sim N(0,100)$ and Sample Size = 25)	82
XX.	Simulations Calculating the Estimates of α and β Based on 100 Replications ($\alpha = 10$, $\beta = 0.2$, $C_i \sim U(\alpha + \beta X_i, \alpha + \beta X_i + 40)$, $X_i = 2i$, $\varepsilon_i \sim N(0,100)$ and Sample Size = 50)	83
XXI.	Simulations Calculating the Estimates of α and β Based on 100 Replications ($\alpha = 10$, $\beta = 0.2$, $C_i \sim U(\alpha + \beta X_i, \alpha + \beta X_i + 40)$, $X_i = 2i$, $\varepsilon_i \sim N(0,100)$ and Sample Size = 75)	84
XXII.	Simulations Calculating the Estimates of α and β Based on 100 Replications ($\alpha = 10$, $\beta = 0.2$, $C_i = 30$, $X_i = 2i$, $\varepsilon_i \sim N(0,100)$ and Sample Size = 25)	86

Table		Page

XXIII. Simulations Calculating the Estimates of α and β Based on 100 Replications ($\alpha = 10$, $\beta = 0.2$, $C_i = 30$, $X_i = 2i$, $\varepsilon_i \sim N(0,100)$ and Sample Size = 50) 87

XXIV. Simulations Calculating the Estimates of α and β Based on 100 Replications ($\alpha = 10$, $\beta = 0.2$, $C_i = 30$, $X_i = 2i$, $\varepsilon_i \sim N(0,100)$ and Sample Size = 75) 88

XXV. Simulations Calculating the Estimates of α and β Based on 100 Replications ($\alpha = 10$, $\beta = 0.2$, $C_i = 1.5X_i - 0.015X_i^2 + \alpha$, $X_i = 2i$, $\varepsilon_i \sim N(0,100)$ and Sample Size = 25) 90

XXVI. Simulations Calculating the Estimates of α and β Based on 100 Replications ($\alpha = 10$, $\beta = 0.2$, $C_i = 1.5X_i - 0.015X_i^2 + \alpha$, $X_i = 2i$, $\varepsilon_i \sim N(0,100)$ and Sample Size = 50) 91

XXVII. Simulations Calculating the Estimates of α and β Based on 100 Replications ($\alpha = 10$, $\beta = 0.2$, $C_i = 1.5X_i - 0.015X_i^2 + \alpha$, $X_i = 2i$, $\varepsilon_i \sim N(0,100)$ and Sample Size = 75) 92

XXVIII. Simulations Calculating the Estimates of α and β Based on 100 Replications ($\alpha = 10$, $\beta = 0.2$, $X_i \sim U(0,100)$, $\varepsilon_i \sim N(0,100)$ and Sample Size = 25) 94

XXIX. Simulations Calculating the Estimates of α and β Based on 100 Replications ($\alpha = 10$, $\beta = 0.2$, $X_i \sim U(0,100)$, $\varepsilon_i \sim N(0,100)$ and Sample Size = 50) 95

XXX. Simulations Calculating the Estimates of α and β Based on 100 Replications ($\alpha = 10$, $\beta = 0.2$, $X_i \sim U(0,100)$, $\varepsilon_i \sim N(0,100)$ and Sample Size = 75) 96

XXXI. Simulations Calculating the Estimates of α and β Based on 100 Replications ($\alpha = 10$, $\beta = 0.2$, $X_i \sim U(0,100)$, $\varepsilon_i \sim N(0,100)$, Sample Size = 25 and 25% Censoring) . 98

XXXII. Simulations Calculating the Estimates of α and β Based on 100 Replications ($\alpha = 10$, $\beta = 0.2$, $X_i \sim U(0,100)$, $\varepsilon_i \sim N(0,100)$, Sample Size = 25 and 50% Censoring) . 99

XXXIII. Simulations Calculating the Estimates of α and β Based on 100 Replications ($\alpha = 10$, $\beta = 0.2$, $X_i \sim U(1,100)$, $\varepsilon_i \sim N(0,100)$, Sample Size = 25 and 75% Censoring) . 100

XXXIV. Simulations Calculating the Estimates of α and β Based on 100 Replications ($\alpha = 10$, $\beta = 0.2$, $X_i \sim U(0,100)$, $\varepsilon_i \sim N(0,100)$, Sample Size = 50 and 25% Censoring) . 102

Table		Page
XXXV.	Simulations Calculating the Estimates of α and β Based on 100 Replications ($\alpha = 10$, $\beta = 0.2$, $X_i \sim U(0,100)$, $\varepsilon_i \sim N(0,100)$, Sample Size = 50 and 50% Censoring) .	103
XXXVI.	Simulations Calculating the Estimates of α and β Based on 100 Replications ($\alpha = 10$, $\beta = 0.2$, $X_i \sim U(0,100)$, $\varepsilon_i \sim N(0,100)$, Sample Size = 50 and 75% Censoring) .	104
XXXVII.	Simulations Calculating the Estimates of α and β Based on 100 Replications ($\alpha = 10$, $\beta = 0.2$, $X_i \sim U(0,100)$, $\varepsilon_i \sim N(0,100)$, Sample Size = 75 and 25% Censoring) .	106
XXXVIII.	Simulations Calculating the Estimates of α and β Based on 100 Replications ($\alpha = 10$, $\beta = 0.2$, $X_i \sim U(0,100)$, $\varepsilon_i \sim N(0,100)$, Sample Size = 75 and 50% Censoring) .	107
XXXIX.	Simulations Calculating the Estimates of α and β Based on 100 Replications ($\alpha = 10$, $\beta = 0.2$, $X_i \sim U(0,100)$, $\varepsilon_i \sim N(0,100)$, Sample Size = 75 and 75% Censoring) .	108
XXXX.	Regression Estimates and Standard Deviations for Log_{10} of Time to Death Versus Age at Transplant With n = 157 Stanford Heart Transplant Patients	111
XXXXI.	Regression Estimates and Standard Deviations for Log_{10} of Time to Death Versus T5 Mismatch Scores With n = 157 Stanford Heart Transplant Patients	112

LIST OF FIGURES

Figure | Page

1. MSE of the Estimates of α and β Based on 100 Replications ($\alpha = 1$, $\beta = 0.2$, $C_i \sim U(10 + \alpha, 50 + \alpha)$, $X_i = 2i$, $\varepsilon_i \sim N(0,1)$ and Sample Size = 10) 61

2. MSE of the Estimates of α and β Based on 100 Replications ($\alpha = 1$, $\beta = 0.2$, $C_i = 1.5X_i - 0.015X_i^2 + \alpha$, $X_i = 2i$ and $\varepsilon_i \sim N(0,1)$) 65

3. MSE of the Estimates of α and β Based on 100 Replications ($\alpha = 1$, $\beta = 0.2$, $C_i = 31$, $X_i = 2i$ and $\varepsilon_i \sim N(0,1)$) 69

4. MSE of the Estimates of α and β Based on 100 Replications ($\alpha = 1$, $\beta = 0.2$, $C_i \sim U(\alpha + \beta X_i, \alpha + \beta X_i + 20)$, $X_i = 2i$ and $\varepsilon_i \sim N(0,1)$) 73

5. MSE of the Estimates of α and β Based on 100 Replications ($\alpha = 1$, $\beta = 0.2$, $C_i \sim U(0,50)$, $X_i = 2i$ and $\varepsilon_i \sim N(0,1)$) . . 77

6. MSE of the Estimates of α and β Based on 100 Replications ($\alpha = 1$, $\beta = 0.2$, $C_i \sim N(\alpha + 40, 16)$, $X_i = 2i$ and $\varepsilon_i \sim N(0,1)$) 81

7. MSE of the Estimates of α and β Based on 100 Replications ($\alpha = 1$, $\beta = 0.2$, $C_i \sim U(\alpha + \beta X_i, \alpha + \beta X_i + 40)$, $X_i = 2i$ and $\varepsilon_i \sim N(0,100)$) 85

8. MSE of the Estimates of α and β Based on 100 Replications ($\alpha = 10$, $\beta = 0.2$, $C_i = 30$, $X_i = 2i$ and $\varepsilon_i \sim N(0,100)$) . . . 89

9. MSE of the Estimates of α and β Based on 100 Replications ($\alpha = 10$, $\beta = 0.2$, $C_i = 1.5X_i - 0.015X_i^2 + \alpha$, $X_i = 2i$ and $\varepsilon_i \sim N(0,100)$) 93

10. MSE of the Estimates of α and β Based on 100 Replications ($\alpha = 10$, $\beta = 0.2$), $C_i \sim U(0,50)$, $X_i \sim U(0,100)$ and $\varepsilon_i \sim N(0,100)$ 97

11. MSE of the Estimates of α and β Based on 100 Replications ($\alpha = 10$, $\beta = 0.2$, $X_i \sim U(0,100)$, $\varepsilon_i \sim N(0,100)$ and Sample Size = 25) 101

Figure | Page

12. MSE of the Estimates of α and β Based on 100 Replications ($\alpha = 10$, $\beta = 0.2$, $X_i \sim U(0,100)$, $\varepsilon_i \sim N(0,100)$ and Sample Size = 50) 105

13. MSE of the Estimates of α and β Based on 100 Replications ($\alpha = 10$, $\beta = 0.2$, $X_i \sim U(0,100)$, $\varepsilon_i \sim N(0,100)$ and Sample Size = 75) 109

14. Scatterplot of Log_{10} Survival Time (in Days) Versus Age at Transplant (in Years) for 157 Stanford Heart Transplant Patients. Patients Denoted by "1" are Deceased and Those by "0" Were Still Alive as of February 1980 113

15. Scatterplot of Log_{10} Survival Time (in Days) Versus T5 Mismatch Score for 157 Stanford Heart Transplant Patients. Patients Denoted by "1" are Deceased and Those by "0" Were Still Alive as of February 1980. 114

CHAPTER I

INTRODUCTION

A random variable of interest in many situations is the time from an event defining the start of observation to the occurrence of another well-defined event which terminates the natural observation period. In clinical medicine, one may wish to investigate the survival experience after different treatments. The waiting time between arrest and initiation of the trial proceedings is another example. In demography, one may wish to describe and compare the risks of death, divorce or migration.

Examples of random variables in most applications (time variables)

Starting Time (Primary Event)	Concluding Time (Secondary Event)
Medicine:	
Heart transplant	Death
Cancer treatment	Death
Treatment of a chronic disease	Remission of symptoms
Application of carcinogen on a mouse	Appearance of tumor
Health Administration:	
Admission to institution	Discharge
Enrollment in health maintenance organization	Withdrawal
Appointment to job class	Promotion out of job class
Purchase of insurance	Claim filed
Report of child abuse	Investigation of report

Demography:

Birth	Death
Marriage	Divorce
Establishment of residence in a community	Move out of a community
Birth of the first child	Birth of the second child

Industry (Reliability of Tested Materials):

Starting time of exposure to stress	Time of breaking up

The observed data are frequently incomplete because the occurrence of the secondary event may be interrupted by some other events. If the secondary event, when an interrupting event takes place, is a random variable, then the random censorship model is said to hold. Such an observation measuring from starting event to interrupting event is referred to as a censored observation.

When the random censoring occurs, an incomplete observation of occurrence times due to random censorship creates difficulties in drawing statistical inferences about the random variable of interest (time of occurrence). Such a phenomenon can occur, for instance, in a clinical trial, during which patients may be treated with one of several possible therapies each time they enter the study. Instead of observing their life-times, experimenters get randomly censored observations which can occur due to the removal of patients from the study for an unrelated reason. Examples of this are: lost to follow-up, dropping out, or having observation time terminated by the study after random entry into the study.

The time of occurrence in medical study is usually called life-time data or survival time. An example of survival data is reported by

Freireich (taken from Gehan (1965)). The survival times of 21 leukemia patients were as follows:

Survival Times (in Weeks)

1, 1, 2, 2, 3, 4, 4, 5, 5, 8, 8, 8, 8, 11, 11, 12, 12, 15, 17, 22, 23

If all survival data were as complete as the above leukemia data, then survival analysis would not require its own statistical techniques. The special feature of most survival studies is that exact survival times cannot always be ascertained. One major concern in a medical study is the need of doing a statistical analysis before all the patients have died. For a patient who has not died at the termination of the experiment, one can only record a censoring time, given by the time elapsed between entry into the study and termination of the study. Patients may also be removed from the study for an unrelated reason such as being lost to follow-up or dropping out.

Freireich was concerned with survival under treatment with the drug 6-mercapotopurine (6-MP). The survival data given above was for the group administered a placebo while the survival times for the 21 patients treated with 6-MP were:

Survival Times (in Weeks)

6, 6, 6, 7, 10, 13, 16, 22, 23

6+, 9+, 10+, 11+, 17+, 19+, 20+, 25+, 32+, 32+, 34+, 35+

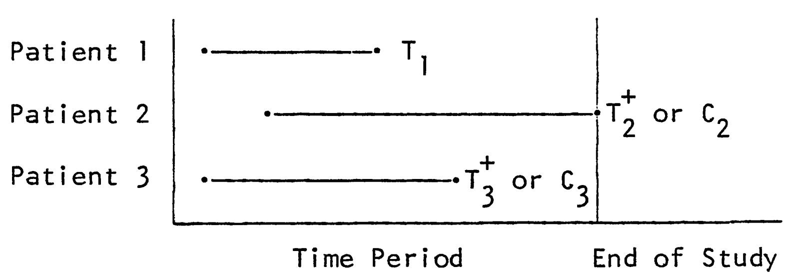
By convention, the censored survival times are indicated by a plus sign. For the treatment group, the longer survival times appear among the censored observations. The true survival times for these individuals are even greater. Any technique that does not capitalize on the special nature of the censored observations may be misleading.

Censored data appears in other settings. A standard industrial example is the study of lifetimes of light bulbs or tubes. For such studies, one can easily start all the light bulbs at the same time and let the experiment continue for a fixed duration. If there are bulbs still burning at the end of the experiment, then all of these have a censored survival time equal to the length of the experiment. For studies of survival times of laboratory animals, the experiments will usually initiate observation of all animals at the same time and then observe them for a fixed duration. Animals alive at the end of the study have the same censored survival time; there are no other censored observations. The medical experiments will seldom have as much control. Patients enter the hospital at different times and not in large groups. Also patients are lost or withdrawn during the experiment. Thus survival analysis must allow for variable censoring. In the statistical literature, there are papers restricting the censoring to a fixed time. These are not general enough for medical applications.

One can consider the regression problem of survival time regressed against covariates. The Stanford Heart Transplantation Program provides the application of the regression problem to survival data. Miller (1976) reports the survival times for 69 patients given heart transplants at Stanford between October 1, 1967 and April 1, 1974. The covariates reported are age at transplant and mismatch score. Miller describes the mismatch score as a measure of dissimilarity between the donor and the recipient tissue; higher scores represent worse matches. He also records whether or not the cause of death was due to rejection of the donor's heart. For the analysis of survival times with mismatch score, Miller treats nonrejection death as censored observations since those patients would hypothetically have died later from rejection. Thus his analysis was performed separately for the regression of survival time with each of the two covariates.

Another example of a regression model for censored data is given by Prentice (1973). He reports survival data from the Veterans Administrative Lung Cancer Study Group. There are 4 covariates; a general measure of medical status, time from diagnosis to entry into the study, age, and being or not being in any previous therapy.

Finally, Dyer (1973) discusses the study of the Chicago People Gas Company. The study followed 1,233 white males between the age of 40-59 who are free of coronary heart disease at entry. At the end of 14 years, there were 246 observed deaths. For each patient, 3 covariates were chosen: systolic blood pressure, serum cholestrol, and cigarette smoking. Dyer (1973) considers regression models of survival time against the above 3 covariates. These covariates are considered risk factors for coronary heart disease (CHD) and cardiovascular renal disease (CVR). The models considered survival times for CVR deaths, CHD deaths, and deaths from other causes as well as censored observations. Those observations can be measured as the following diagram.



Censored Sample

When covariates affect the time of occurrence, the models which incorporate the effect of the covariates must be developed. For example, age of a patient and severity of disease will affect length of stay in an institution. In such a case, one may be interested in studying the administrative implications of a policy change involving the covariate structure of a patient population.

Covariates are commonly incorporated into censored models in either of the following ways. First, the proportional hazard model described by Cox (1972) assumes the covariates act multiplicatively on the hazard function, which is the instantaneous rate of occurrence at a given time, conditional upon no occurrence up to that time. Kalbfleisch and Prentice (1980) have discussed this model in their book. The other way is to assume that the expected occurrence time (or a transform) is a linear combination of the covariates. This dissertation will study only linear regression model.

Model

The random variables and observations will be denoted as follows:

Let the random Y_i be the time of occurrence, or a transform of the time, for the i^{th} subject, with distribution F_{Y_i}.

Let the random variable C_i be the time to censoring of the i^{th} subject with distribution G_{C_i}.

Assume Y_i and C_i are independent.

Let $\underset{\sim}{X}_i$ be a $(p+1)\times 1$ vector for the i^{th} subject, the first term of which is a constant 1, the remaining terms of which are p covariates.

Assume C_i and $\underset{\sim}{X}_i$ are independent.

Define the random variable $T_i = \mathrm{Min}(Y_i, C_i)$ and the indicator random variable for the i^{th} subject by

$$\delta_i = \begin{cases} 1 & \text{if } Y_i \le C_i \\ 0 & \text{if } Y_i > C_i \end{cases} \tag{1.1}$$

Hence, an observation on the i^{th} subject from a sample of size n will consist of $(t_i, \delta_i, \underset{\sim}{x}_i)$, $i = 1, 2, \ldots, n$

The general least squares model is

$$\underset{\sim}{Y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon},\ E(\underset{\sim}{\varepsilon}) = \underset{\sim}{0},\ V(\underset{\sim}{\varepsilon}) = \sigma^2 V \tag{1.2}$$

where x is a covariate matrix and V is a known positive definite matrix or

$$\underset{\sim}{Y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon},\ E(\underset{\sim}{\varepsilon}) = \underset{\sim}{0},\ V(\underset{\sim}{\varepsilon}) = \sigma^2 V_{diag} \tag{1.3}$$

where V_{diag} is a known diagonal matrix

or

$$\underset{\sim}{Y} = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon},\ E(\underset{\sim}{\varepsilon}) = \underset{\sim}{0},\ V(\underset{\sim}{\varepsilon}) = \sigma^2 I \tag{1.4}$$

where I is the identity matrix.

In all cases,

$$\underset{\sim}{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \qquad X = \begin{bmatrix} 1 & X_{11} & X_{21} \cdots X_{(p)1} \\ 2 & X_{12} & X_{22} \cdots X_{(p)2} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & X_{1n} & X_{2n} & X_{(p)n} \end{bmatrix}$$

and

$$\underset{\sim}{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} , \qquad \underset{\sim}{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} .$$

If all error variances are assumed equal, the last of these models is suitable for almost all of the cases. It will henceforth be referred to as the uncensored model.

This model states that ε_i's are uncorrelated with common mean and variance. Suppose that F is the common distribution, then the relation between F and F_i (or F_{Y_i}) under linear regression $Y_i = \underset{\sim}{X}_i^T \underset{\sim}{\beta} + \varepsilon_i$ is

$$\begin{aligned} F_i(t) &= P(Y_i \le t) \\ &= P(Y_i - X_i\underset{\sim}{\beta} \le t - X_i\underset{\sim}{\beta}) \\ &= P(\varepsilon_i \le t - X_i\underset{\sim}{\beta}) \\ &= F(t - X_i\underset{\sim}{\beta}) \text{ for all } i . \end{aligned} \tag{1.5}$$

The random variables δ_i of (1.1) are independent but not identically distributed unless

$$\delta_i = \begin{cases} 1 & \text{if } \ Y_i - X_i\underset{\sim}{\beta} \le C_i - X_i\underset{\sim}{\beta} \\ 0 & \text{if } \ Y_i - X_i\underset{\sim}{\beta} > C_i - X_i\underset{\sim}{\beta} \end{cases} \quad \text{for all } i \tag{1.6}$$

that is, if $G_i(t) = G(t - X_i\underset{\sim}{\beta})$. If $\underset{\sim}{\beta} = 0$, there is no regression effect. Then both $F_i(t) = F(t)$ and $G_i(t) = G(t)$.

The least squares objective of fitting uncensored model is to obtain the estimate of $\underset{\sim}{\beta}$ which minimizes the sum of squared residuals $(\underset{\sim}{Y} - X\underset{\sim}{\beta})^T(\underset{\sim}{Y} - X\underset{\sim}{\beta})$. Hence, the least squares solution is

$$\hat{\underset{\sim}{\beta}} = (X^TX)^{-1} X^T\underset{\sim}{Y}$$

which has the properties of

$$E(\hat{\underset{\sim}{\beta}}) = \underset{\sim}{\beta}, \quad V(\hat{\underset{\sim}{\beta}}) = \sigma^2(X^TX)^{-1}$$

and an unbiased estimate of σ^2 is given by

$$\hat{\sigma}^2 = (\underset{\sim}{Y} - X\hat{\underset{\sim}{\beta}})^T(\underset{\sim}{Y} - X\hat{\underset{\sim}{\beta}})/(n-k) .$$

With censoring, the objective of estimating $\underset{\sim}{\beta}$ is complicated by the fact that Y_i is sometimes unobservable. When this happens, many methods will substitute Y_i^* for the unobservable random variable Y_i.

Hence, we will have the model

$$\underset{\sim}{Y}^* = X\underset{\sim}{\beta} + \underset{\sim}{\varepsilon}^*, \quad E(\underset{\sim}{\varepsilon}^*) = \underset{\sim}{0}, \quad V(\underset{\sim}{\varepsilon}^*) = \sigma^{*2}V^* .$$

This will be called the censored model. Under the model, the least squares solution to minimizing

$$(\underset{\sim}{Y}^* - X\underset{\sim}{\beta})^T \; V^{*-1}(\underset{\sim}{Y}^* - X\underset{\sim}{\beta})$$

is

$$\hat{\underset{\sim}{\beta}}^* = (X^TV^{*-1}X)^{-1} X^TV^{*-1}\underset{\sim}{Y}^*$$

and if $\underset{\sim}{Y}^*$ and V^* were known, an unbiased estimate of σ^{*2} would be given by

$$\sigma^{*2} = (\underset{\sim}{Y}^* - X\hat{\underset{\sim}{\beta}}^*)^T \, V^{*-1}(Y^* - X\hat{\underset{\sim}{\beta}}^*)/(n-k) .$$

Note that σ^{*2} of the censored model is not the same as σ^2 of the uncensored model except in the special case of no censoring. If there is no censoring, $\underset{\sim}{Y}^* = \underset{\sim}{Y}$, $\hat{\underset{\sim}{\beta}}^* = \hat{\underset{\sim}{\beta}}$, and the censored model reduces to uncensored model. The hope is that in the presence of censoring, $\underset{\sim}{Y}^*$ is a good substitute for $\underset{\sim}{Y}$.

If $\underset{\sim}{Y}^*$ were known, then the least squares estimate for $\underset{\sim}{\beta}$ under the censored model could be defined in terms of $\underset{\sim}{Y}^*$. However, $\underset{\sim}{Y}^*$ in general is not fully known but has to be estimated by a quantity which can be called $\hat{\underset{\sim}{Y}}^*$. The corresponding least squares estimates of $\underset{\sim}{\beta}$ will therefore be defined in terms of $\hat{\underset{\sim}{Y}}^*$.

CHAPTER II

REVIEW OF LITERATURE

A model of the survival time which incorporates the effects of the covariates has been developed by Cox (1972). He assumes that the covariates act multiplicatively on the hazard function, which is the instantaneous rate of surviving at a given time, by conditioning upon no occurrence up to that time.

If F(y;x) is the underlying distribution function for the survival time Y when the covariates are X, and f(y;x) is the corresponding density function, the proportional hazards model assumes that the hazard rate

$$\lambda(y;\underset{\sim}{x}) \;=\; f(y;\underset{\sim}{x})/(1 - F(y;\underset{\sim}{x})) \quad \text{where} \quad o \le r \le 1$$

is given by

$$\lambda(y;\underset{\sim}{x}) \;=\; \lambda_0(y)\ \exp(\underset{\sim}{x}^T\underset{\sim}{\beta}) \ ,$$

where $\underset{\sim}{\beta}$ is the vector of regression coefficients and $\lambda_0(y)$ is the hazard rate when $\underset{\sim}{x} = \underset{\sim}{0}$. He proposed a partial likelihood approach to estimate $\underset{\sim}{\beta}$ since the function $\lambda_0(y)$ being unknown prevents a full likelihood analysis. The patients in the risk set $R(y)$ are those still alive and in the study at time y-. If it is known that a patient dies at time y, then the conditional probability that it is patient i among those at risk is

$$\exp(\underset{\sim}{x}_i^T\underset{\sim}{\beta}) / \sum_{j \varepsilon R(y)} \exp(\underset{\sim}{x}_j^T\underset{\sim}{\beta}) \ .$$

If $t_{(1)} \leq t_{(2)} \leq \cdots \leq t_{(n)}$ are the ordered observations of the survival time; censored or uncensored, then the partial likelihood is

$$L_e = \prod_{i=1}^{n} \left[\frac{\exp(\underset{\sim}{x}_{(i)}^T \underset{\sim}{\beta})}{\sum_{j\varepsilon(t_{(i)})} \exp(\underset{\sim}{x}_j^T \underset{\sim}{\beta})} \right]^{\delta_{(i)}} , \tag{2.1}$$

where $\underset{\sim}{x}_{(i)}$ and $\underset{\sim}{\delta}_{(i)}$ are associated with $t_{(i)}$. The value of $\underset{\sim}{\beta}$ maximizing (2.1) is obtained by solving for the root of

$$\sum_{i=1}^{n} \delta_{(i)} \left\{ \underset{\sim}{x}_{(i)}^T \underset{\sim}{\beta} - \log \sum_{j\varepsilon(t_{(i)})} \exp(\underset{\sim}{x}_j^T \underset{\sim}{\beta}) \right\} = 0 . \tag{2.2}$$

Other methods developed by Miller (1976), Buckley and James (1979), and Koul, Susarla, and Van Ryzin (1981) are based on the standard linear model with

$$E(Y|\underset{\sim}{x}) = \underset{\sim}{x}^T \underset{\sim}{\beta} \tag{2.3}$$

where $\underset{\sim}{\beta}$ is the vector of regression coefficients for the covariates X. If Y is measured on a log scale so that Y = log U where U is the actual survival time, then (2.3) corresponds to an accelerated time model.

The first least squares type estimator for censored data was published by Miller (1976). It assumes that $F(y;\underset{\sim}{x}) = F(y - \underset{\sim}{x}^T \underset{\sim}{\beta})$ where F has zero expectation. This gives the expectation (2.3) and homogeneous variance along the regression line.

Miller proposed using an iterative sequence to calculate the estimate of the regression coefficient vector $\underset{\sim}{\beta}$:

$$\hat{\underset{\sim}{\beta}}_{p+1} = (X^T W(\hat{\underset{\sim}{\beta}}_p) X)^{-1} X^T W(\hat{\underset{\sim}{\beta}}_p) \underset{\sim}{t} \tag{2.4}$$

where

$$\underset{\sim}{t} = (t_1, t_2, \ldots, t_n)^T,$$

$$X = \text{matrix } (x_{ij}), \text{ and}$$

$$W(\hat{\underset{\sim}{\beta}}_p) = \text{diagonal matrix } (w_i(\hat{\underset{\sim}{\beta}}_p)) \ . \qquad (2.5)$$

The limit of the sequence $\hat{\underset{\sim}{\beta}}_p$, $p = 0, 1, 2, \ldots$, is the estimate of $\underset{\sim}{\beta}$.

The weight $w_i(\hat{\underset{\sim}{\beta}}_p)$ in (2.4) - (2.5) is the size of the jump assigned to $\hat{\varepsilon}_i = \hat{\varepsilon}_i(\hat{\underset{\sim}{\beta}}_p) = t_i - \underset{\sim}{x}_i^T\hat{\underset{\sim}{\beta}}_p$ by the Kaplan-Meier estimator applied to $\hat{\varepsilon}_1, \hat{\varepsilon}_2, \ldots, \hat{\varepsilon}_n$; i.e.,

$$w_i(\hat{\underset{\sim}{\beta}}_p) = \hat{F}(\hat{\varepsilon}_i;\hat{\underset{\sim}{\beta}}_p) - \hat{F}(\hat{\varepsilon}_i - ;\hat{\underset{\sim}{\beta}}_p) \qquad (2.6)$$

Only the uncensored t_i actually appear in (2.4) since the weight assigned to any censored observation is zero. For this reason, it makes sense to use as a starting value $\hat{\underset{\sim}{\beta}}_0$ the ordinary (unweighted) least squares estimator applied to only the uncensored data. It becomes

$$\hat{\underset{\sim}{\beta}}_{p+1} = (X_{un}^T \ W^*(\hat{\underset{\sim}{\beta}}_p)X_{un})^{-1} \ X_{un}^T \ W^*(\hat{\underset{\sim}{\beta}}_p)\underset{\sim}{t}_{un} \qquad (2.7)$$

where

$\underset{\sim}{t}_{un}$ = vector of uncensored survival observations

X_{un} = matrix (x_{ij}) of associated uncensored covariates

$W^*(\hat{\underset{\sim}{\beta}}_p)$ = diagonal matrix $(w_i(\hat{\underset{\sim}{\beta}}_p))$ excluding 0 diagonal terms.

Buckley and James (1979) do not assume random censorship. They consider the censoring variables as fixed and given values. They define the random variable

$$Y_i^* = T_i\delta_i + E(Y_i|Y_i > C_i)(1 - \delta_i), \text{ for } i = 1, 2, \ldots, n$$

where

δ_i = the indicator variable

$$\delta_i = \begin{cases} 1 & \text{if } Y_i \le C_i \\ 0 & \text{if } Y_i > C_i \end{cases} .$$

They obtain the least squares solution

$$\hat{\underset{\sim}{\beta}} = (X^TX)^{-1} X^T \underset{\sim}{t}$$

where

$$\underset{\sim}{t} = (t_1, t_2, \ldots, t_n)^T .$$

The idea is to replace each censored observation by $E(Y_i|Y_i > C_i)$. Since $E(Y_i|Y_i > C_i)$ is unknown, Buckley and James estimate it from the Kaplan-Meier estimator for the residuals. Specifically, if $\delta_i = 1$, let $\hat{t}_i(\hat{\underset{\sim}{\beta}}_p) = t_i$, but if $\delta_i = 0$, let

$$\hat{t}_i(\hat{\underset{\sim}{\beta}}_p) = \underset{\sim}{x}_i^T \hat{\underset{\sim}{\beta}}_p + \frac{\sum_{\hat{\varepsilon}_j > \hat{\varepsilon}_i} w_j(\hat{\underset{\sim}{\beta}}_p)\hat{\underset{\sim}{\varepsilon}}_j}{1 - \hat{F}(\hat{\varepsilon}_i;\hat{\underset{\sim}{\beta}}_p)} , \tag{2.8}$$

where $\hat{\varepsilon}_j = t_j - \underset{\sim}{x}_j^T \hat{\underset{\sim}{\beta}}_p$, $\hat{F}$ is defined as:

In the case of no tied uncensored observations

$$1 - \hat{F}(\hat{\varepsilon}_i;\hat{\underset{\sim}{\beta}}_p) = \prod_{\hat{\varepsilon}_{(j)} \le \hat{\varepsilon}_i} \left(1 - \frac{1}{n-j+1}\right)^{\delta_{(j)}}, \tag{2.9}$$

where $\hat{\varepsilon}_{(1)} \le \hat{\varepsilon}_{(2)} \le \dots \le \hat{\varepsilon}_{(n)}$ and $\delta_{(j)}$ is associated with $\hat{\varepsilon}_{(j)}$. With tied uncensored observations

$$1 - \hat{F}(\hat{\varepsilon}_i;\hat{\beta}_p) = \prod_{\hat{\varepsilon}'_{(j)} \le \hat{\varepsilon}_i} \left(1 - \frac{d'_{(j)}}{n'_{(j)}}\right)^{\delta'_{(j)}} \tag{2.10}$$

where $\hat{\varepsilon}'_{(1)} < \hat{\varepsilon}'_{(2)} < \dots$ are the ordered distinct values of $\hat{\varepsilon}_j$, $n'_{(j)}$ is the number at risk at $\hat{\varepsilon}'_{(j)}-$, $d'_{(j)}$ is the number dying at $\hat{\varepsilon}'_{(j)}$, and $\delta'_{(j)} = 1$ if $d'_{(j)} > 0$, $= 0$ otherwise. $w_j(\hat{\underset{\sim}{\beta}}_p)$ is defined by (2.6). The summation in (2.8) is overall $\hat{\varepsilon}_j = t_j - \underset{\sim}{x}_j^T\hat{\underset{\sim}{\beta}}_p$ greater than $\hat{\varepsilon}_i = t_i - \underset{\sim}{x}_i^T\hat{\beta}_p$. The regression estimator $\hat{\underset{\sim}{\beta}}_{p+1}$ at the $(p+1)^{st}$ step is the usual least squares estimator

$$\hat{\underset{\sim}{\beta}}_{p+1} = (X^TX)^{-1}\, X^T\hat{\underset{\sim}{t}}(\hat{\underset{\sim}{\beta}}_p), \tag{2.11}$$

where

$$\hat{\underset{\sim}{t}}(\hat{\underset{\sim}{\beta}}_p) = (\hat{t}_1(\hat{\underset{\sim}{\beta}}_p),\ \hat{t}_2(\hat{\underset{\sim}{\beta}}_p),\ \dots,\ \hat{t}_n(\hat{\underset{\sim}{\beta}}_p))^T, \text{ and}$$

$$X = \text{matrix } (x_{ij}).$$

The iteration is continued until $\hat{\underset{\sim}{\beta}}_p$ converges to a limiting value $\hat{\underset{\sim}{\beta}}$ or becomes trapped in a loop like the Miller estimator.

Since the estimator (2.11) uses a value for the dependent variable at every $\underset{\sim}{x}_i$, it seems sensible to take for the starting $\hat{\underset{\sim}{\beta}}_0$ the least squares estimator $(X^TX)^{-1}\, X^T\underset{\sim}{t}$ which treats all the observations as uncensored whether they are uncensored or not.

The Buckley-James estimator exploits the following linear relationship:

$$E(\delta_i T_i + (1-\delta_i)\ E(Y_i|Y_i > T_i)|\underset{\sim}{x}_i)$$

$$= \int_{-\infty}^{\infty} y(1-G(y;\underset{\sim}{x}_i))dF(y;\underset{\sim}{x}_i) + \int_{-\infty}^{\infty} \left[\frac{\int_y^{\infty} u\ dF(u,\underset{\sim}{x}_i)}{1 - F(y;\underset{\sim}{x})}\right] (1-F(y;\underset{\sim}{x}_i))dG(y;x)$$

$$= \int_{-\infty}^{\infty} y\ dF(y;\underset{\sim}{x}_i),$$

$$= \underset{\sim}{x}_i^T \underset{\sim}{\beta}\ . \tag{2.12}$$

An estimate of the conditional expectation based on the Kaplan-Meier estimator is substituted in the variable $\hat{t}_i = \delta_i t_i + (1-\delta_i)\ \hat{E}(Y_i|Y_i > t_i)$ and then the usual least squares normal equations are solved.

For the Koul-Susarla-Van Ryzin (1981) estimator, a different linear relationship forms a basis. Assume that the censoring distributions are independent of $\underset{\sim}{x}_i$, i.e., $G(y;\underset{\sim}{x}_i) \equiv G(y)$. Then,

$$E(\delta_i T_i (1-G(T_i))^{-1}|\underset{\sim}{x}_i)$$

$$= \int_{-\infty}^{\infty} y(1-G(y))^{-1}\ (1-G(y))\ dF(y;\underset{\sim}{x}_i),$$

$$= \underset{\sim}{x}_i^T \underset{\sim}{\beta}\ . \tag{2.13}$$

In the Koul-Susarla-Van Ryzin estimator, an estimate for $G(y)$ is substituted in the variable $\hat{t}_i = \delta_i t_i (1-\hat{G}(t_i))^{-1}$ and then the usual least squares normal equations are solved. One could have allowed $G(y)$ to depend on $\underset{\sim}{x}_i$ in (2.13), but there would be no way of estimating each $G(y;\underset{\sim}{x}_i)$ from the data without imposing assumption on $G(y;\underset{\sim}{x})$ as a function of $\underset{\sim}{x}$.

The Kaplan-Meier estimator with the roles of y_i and e_i reversed could be used to estimate the common censoring distribution $G(y)$. The great advantage of the Koul-Susarla-Van Ryzin estimator is that no iteration is required in the computation of the estimate. Specifically,

$$\hat{\hat{\underset{\sim}{\beta}}} = (X^T X)^{-1} X^T \hat{\underset{\sim}{t}} ,$$

where X is defined in (2.5) and $\hat{\underset{\sim}{t}} = (\hat{t}_1, \hat{t}_2, \ldots, \hat{t}_n)^T$ where $\hat{t}_i$ for $i = 1, 2, \ldots, n$ are computed as mentioned.

Schmee and Hahn (1979) define a random variable

$$Y_i^* = \delta_i T_i + (1-\delta_i)\ E(Y_i | Y_i > C_i),\ \text{for } i = 1, 2, \ldots, n$$

where δ_i is the indicator variable. $E(Y_i|Y_i > C_i)$ is computed by using the additional assumption of normal errors for survival time distribution. Their estimates are

$$\hat{\hat{\underset{\sim}{\beta}}} = (X^T X)^{-1} X^T \hat{\underset{\sim}{y}}^*$$

where $\hat{\underset{\sim}{y}}^* = (\hat{y}_1^*, \hat{y}_2^*, \ldots, \hat{y}_n^*)^T$ and $\hat{y}_i^*$, for $i = 1, 2, \ldots, n$ can be estimated assuming $\hat{y}$ has a normal distribution. The method is also iterative.

Following the idea of Buckley and James (1979), Koul, Susarla, and Van Ryzin (1981) define a random variable.

$$Y_i^* = \delta_i T_i + (1-\delta_i)\ E(Y_i | Y_i > C_i),\ \text{for } i = 1, 2, \ldots, n .$$

That is, when the survival time is censored, the mean lifetime given censoring of Y_i at C_i should be used. This idea is the same as that of Buckley and James but has a different approach, depending on the mathematical form of this quantity under the assumptions.

Friedman and Stuetzle (1981) define a random variable

$$Y_i^* = \delta_i T_i + (1-\delta_i)T_i^o, \text{ for } i = 1, 2, \ldots, n$$

where T_i^o is given by the censoring time C_i if it exceeds the predicted value of survival time, i.e., if $C_i > \underset{\sim}{x}_i^T \hat{\underset{\sim}{\beta}}$ and by zero if the predicted value exceeds the censoring time. Their least squares solution to minimizing

$$(\underset{\sim}{Y}^* - X\underset{\sim}{\beta})^T(\underset{\sim}{Y}^* - X\underset{\sim}{\beta})$$

is

$$\hat{\underset{\sim}{\beta}}^* = (X^TX)^{-1} X^T\underset{\sim}{y}^*$$

where

$$\underset{\sim}{y}^* = (y_1^*, y_2^*, \ldots, y_n^*)^T .$$

Instead of getting $\hat{\underset{\sim}{\beta}}^*$, they have

$$\hat{\hat{\underset{\sim}{\beta}}} = (X^TX)^{-1} X^T\hat{\underset{\sim}{y}}^*$$

where

$$\hat{\underset{\sim}{y}}^* = (\hat{y}_1^*, \hat{y}_2^*, \ldots, \hat{y}_n^*)^T .$$

The method is iterative by using an initial value from the least squares estimate based only on uncensored observations. They say that their algorithm always produces a unique solution.

Chatterjee and Meleisk (1981) define a random variable

$$Y_i^* = \delta_i T_i + (1-\delta_i) T_i^* ,$$

where T_i^* is $E(Y_i | Y_i > C_i)$, again assuming normal errors. The estimate of $E(Y_i | Y_i > C_i)$ is

$$\hat{E}(Y_i | Y_i > C_i) = \underset{\sim}{x}_i^T \hat{\underset{\sim}{\beta}} + \frac{\sigma_{p-1,i} \, \phi(c_{p-1,i})}{1 - \Phi(c_{p-1,i})} - (\underset{\sim}{\beta} - \hat{\underset{\sim}{\beta}}_{p-1})^T H(\hat{\hat{\underset{\sim}{\beta}}}_{p-1}) c_{pi}$$

where

$$c_{pi} = (t_i - \underset{\sim}{x}_i^T \hat{\hat{\underset{\sim}{\beta}}}_p)/\sigma_{pi}$$

$$H(a) = \frac{d}{dx} \left[\frac{\phi(x)}{1 - \Phi(X)}\right]_{X=a}$$

at the p^{th} iteration. Their estimate of $\underset{\sim}{\beta}$ is

$$\hat{\hat{\underset{\sim}{\beta}}} = (X^T X)^{-1} X^T \hat{\underset{\sim}{y}}^*$$

where

$$\hat{\underset{\sim}{y}}^* = (\hat{\underset{\sim}{y}}_1^*, \hat{\underset{\sim}{y}}_2^*, \ldots, \hat{\underset{\sim}{y}}_n^*)^T .$$

An initial estimate of $\underset{\sim}{\beta}$ is needed to evaluate $\hat{\underset{\sim}{y}}^*$, and consequently the method is iterative.

Durongwatana (1983) performed some simulations for estimating regression coefficients by using only uncensored observations. The comparisons between his estimators and those of Miller, Buckley and James, and Koul-Susarla-Van Ryzin were made. The results show that his estimates have lower mean square error than the others do.

In this dissertation an attempt is made to avoid iterative procedures which have a disadvantage in case of divergence. It adjusts for bias when using only uncensored observations. Furthermore, the quality of those estimators will be shown.

CHAPTER III

REGRESSION METHODS FOR CENSORED OBSERVATIONS

Adjusted Method

3.1 Introduction

We consider the usual linear regression situation with the following model

$$Y_i = \alpha + \beta X_i + \varepsilon_i, \; i = 1, 2, \ldots, n \qquad (3.1)$$

where X_i are known constant covariates, α and β are unknown regression coefficients to be estimated and ε_i are the independent random errors with common distribution F such that

$$E(\varepsilon_i) = 0,$$

$$V(\varepsilon_i) = \sigma^2,$$

and $$\mathrm{Cov}(\varepsilon_i, \varepsilon_j) = 0, \; i \neq j, \; i, j = 1, 2, \ldots, n .$$

Let $C_1, C_2, \ldots, C_n$ be independent censoring random variables with distribution G; C_i is censoring time associated with Y_i. Assume that C_i is independent of Y_i and X_i for $i = 1, 2, \ldots, n$.

F and G are unknown.

We observe,

$$T_i = \mathrm{Min}(Y_i, C_i),$$

and

$$\delta_i = \begin{cases} 1 & \text{when } Y_i \le C_i \\ 0 & \text{when } Y_i > C_i \ . \end{cases} \tag{3.2}$$

Suppose f and g are the survival-time density function and the censoring-time density function respectively, then

$$\begin{aligned} P(Y \le y_o, Y \le C) &= \int_{-\infty}^{y_o} \int_{y}^{\infty} f(y)\ g(c)\ dc\ dy \\ &= \int_{-\infty}^{y_o} f(y)\ (1-G(y))dy, \end{aligned}$$

and

$$\begin{aligned} P(Y \le C) &= \int_{-\infty}^{\infty} \int_{y}^{\infty} f(y)\ g(c)\ dc\ dy \\ &= \int_{-\infty}^{\infty} f(y)\ (1-G(y))dy \\ &= E_Y(1-G(y)) \end{aligned}$$

Hence,

$$P(Y \le y_o | Y \le C) = \frac{\int_{-\infty}^{y_o} f(y)(1-G(y))dy}{E_Y(1-G(y))},$$

and

$$f(Y|Y \le C) = \frac{f(y)(1-G(y))}{E_Y(1-G(y))},$$

then

$$E(Y_i|Y_i \le C_i, X_i = x_i) = \frac{\int_{-\infty}^{\infty} y_i f(y_i)(1-G(y_i))dy_i}{E(1-G(y_i))}$$

$$= \frac{(\int_{-\infty}^{\infty} Y_i f(y_i)dy_i) - (\int_{-\infty}^{\infty} y_i G(y_i)f(y_i)dy_i)}{E(1-G(Y_i))}$$

$$= \frac{E(Y_i) - E(Y_i G(Y_i))}{E(1-G(Y_i))}$$

$$= \frac{E(Y_i) - E((\alpha + \beta x_i)G(Y_i)) - E(\varepsilon_i G(Y_i))}{E(1-G(Y_i))}$$

$$= \frac{(\alpha + \beta x_i)E(1-G(Y_i)) - E(\varepsilon_i G(Y_i))}{E(1-G(Y_i))}$$

$$= (\alpha + \beta x_i) - \frac{E(\varepsilon_i G(Y_i))}{E(1-G(Y_i))}$$

then,

$$E(Y_i|Y_i \le C_i, X_i = x_i) = E(Y_i) - \frac{E(\varepsilon_i G(Y_i))}{E(1-G(Y_i))} \tag{3.2}$$

The idea is that if we estimate the α and β from the model only from the uncensored observations, ignoring the censored observations, the estimators would be biased estimators for α and β. This method proposes the way to adjust for the biases. The difficulties are the results of lack of knowledge about the specific forms of both F and G. With a nonparametric method, the Kaplan-Meier product limit estimator of distribution function involving censored observations, the biases can be estimated.

3.2 Kaplan-Meier Estimation

An important part of the adjusted method is the product limit estimator introduced by Kaplan and Meier (1958). Consider the case in which all individuals or animals are observed to die so that the survival times can be exact and known (no censoring).

Let $y_1, y_2, \ldots, y_n$ be the exact survival times (occurrence times) of the n individuals. An estimator of the survival function S(y) is the estimated proportion of individuals in the sample who survive longer than y, that is,

$$S(y) = \frac{\text{number of individuals in the sample who survive longer than } y}{\text{total number of individuals in the sample}}$$

If relabeling of n survival times $y_1, y_2, \ldots, y_n$ in ascending order is done, they become

$$y_{(1)} \leq y_{(2)} \leq y_{(3)} \leq \ldots\ldots \leq y_{(n)} .$$

Therefore, the survival function at $y_{(i)}$ can be estimated as

$$\hat{S}(y)_{(i)} = \frac{n-i}{n}$$

where (n-i) is the number of individuals in the sample surviving longer than $y_{(i)}$. If two or more $y_{(i)}$ are equal (tied observations), the largest (i) value is used. For example, if $y_{(2)} = y_{(3)} = y_{(4)}$, then $\hat{S}(y_{(2)}) = \hat{S}(y_{(3)}) = \hat{S}(y_{(4)}) = \frac{n-4}{n}$.

This method can only be applied if all the individuals are followed until death (uncensored). If some are still alive at the end of the study, a modified method of estimating S(y) is required. Kaplan and Meier developed a method based on a censored sample to estimate the distribution function. For example, suppose 10 patients joined a clinical study at the beginning of 1983. During that year 6 patients died and 4 survived. At the end of the year, 20 additional patients joined the study. In 1984, three patients who entered in the beginning of 1983 and 15 patients who entered later died, leaving 1 and 5 survivors respectively. The study terminated at the end of 1984. We want to estimate the proportion of patients in the population surviving for 2 years or more, i.e. S(2).

The first group of patients in this example is followed for 2 years while the second group is followed only for one year. Patients who survived two years may be considered as surviving the first year and then surviving one more year. Thus, the probability of surviving for 2 years or more is equal to the probability of surviving the first year and then surviving one more year. That is

S(2) = P(surviving the first year and then surviving one more year)

which can be written as

S(2) = P(surviving two years given that patient has survived the first year) x P(surviving the first year).

The Kaplan-Meier estimate of S(2) is

S(2) = (Proportion of patients surviving two years given that they survive for 1 year) x (Proportion of patients surviving 1 year).

This simple rule may be generalized as follows: The probability of surviving $k(\geq 2)$ or more years from the beginning of the study is product of k observed survival rates;

$$\hat{S}(k) = P_1 \times P_2 \times \ldots\ldots \times P_n$$

where

P_1 denotes the proportion of patients surviving at least one year after the beginning of the study,

P_2 denotes the proportion of patients surviving the second year after they have survived one year from the beginning of the study, etc., and,

P_k denotes the proportion of patients surviving the k^{th} year after they have survived (k-1) year from the beginning of the study.

Therefore, the product-limit estimate of the probability of surviving any particular number of years from the beginning of the study is the product of the same estimate up to the previous year and the observed conditional survival rate for the particular year.

Kaplan-Meier Estimation of Distribution Function

Censored Observations

1. Order all the survival times, both censored and uncensored, from smallest to largest, $t_{(1)} \leq t_{(2)} \leq \ldots\ldots \leq t_{(n)}$. There are $\delta_{(1)}$, $\delta_{(2)}$, ..., $\delta_{(n)}$ corresponding to $t_{(1)}$, $t_{(2)}$, ..., $t_{(n)}$. If a censored observation has the same value as an uncensored, the former should appear first.

2. Label each ordered observation in 1) with the rank i, i = 1, 2, ..., n. In case, for example, there are ties among rank p, p+1, p+2,

use rank p for all three observations. The next rank will be p+3.

3. Compute $(n-i)/(n-i+1)$ for every observation $t_{(i)}$ where i is the rank for $t_{(i)}$ assigned in step 2. This will give the proportion of patients or animals surviving up to and then through $t_{(i)}$.

4. Compute $((n-i)/(n-i+1))^{1-\delta_{(i)}}$ for every $t_{(i)}$.

5. $\hat{S}(t)$ is the product of all values of $(n-i)/(n-i+1)$ up to and including t.

6. If some censored observations are ties, the smallest $\hat{S}(t)$ would be used.

Hence, we have

$$\hat{S}(t) = \prod_{t_{(i)}<t} \left(\frac{n-i}{n-i+1}\right)^{1-\delta_{(i)}} .$$

Using this method, the estimation $\hat{G}(t) = 1 - \hat{S}(t)$ can be made. For example, consider 12 observations

t_i	δ_i	Step 1	Step 2	Step 3	Step 4	Step 5	Step 6	$\hat{G}(t)$
1	1	1	1	11/12	1	1	1	0
1	1	1	1	11/12	1	1	1	0
2.5	0	2	3	9/10	1	1	1	0
3	1	2	3	9/10	1	1	1	0
2	1	2.5+	5	7/8	7/8	7/8	$(7/8)^4$	$1-(7/8)^4$
2	1	2.5+	5	7/8	7/8	$(7/8)^2$	$(7/8)^4$	$1-(7/8)^4$
2.5	0	2.5+	5	7/8	7/8	$(7/8)^3$	$(7/8)^4$	$1-(7/8)^4$
3.5	0	2.5+	5	7/8	7/8	$(7/8)^4$	$(7/8)^4$	$1-(7/8)^4$

t_1	δ_1	Step 1	Step 2	Step 3	Step 4	Step 5	Step 6	$\hat{G}(t)$
2.5	0	3	9	3/4	1	$(7/8)^4$	$(7/8)^4$	$1-(7/8)^4$
3	1	3	9	3/4	1	$(7/8)^4$	$(7/8)^4$	$1-(7/8)^4$
2.5	0	3.5+	11	1/2	1/2	$(7/8)^4(1/2)$	0	1
4.0	1	4.0	12	0	0	0	0	1

NOTE: + means censored observation.

3.3 Adjustment of Regression Model With Censored Observations

After G has been estimated by the empirical distribution $\hat{G}$, the algorithm for estimating α and β can be done as follows:

Step 1. Take all uncensored observations together with their covariates and use the least squares method to get initial estimates of α and β. Hence, we will have

$$\hat{\underset{\sim}{\beta}}_{un} = (X_{un}^T X_{un})^{-1} X_{un}^T \underset{\sim}{Y}_{un}$$

where

$$\hat{\underset{\sim}{\beta}}_{un} = \begin{bmatrix} \hat{\alpha}_o \\ \hat{\beta}_o \end{bmatrix},$$

n_{un} = the number of uncensored observations,

X_{un} = the covariates for associated uncensored observations,

and $\underset{\sim}{Y}_{un}$ = the uncensored observations.

Step 2. Calculate prediction value $\hat{Y}_i$ and residuals $\hat{\varepsilon}_i$ from step 1 for $i = 1, 2, \ldots, n_{un}$ where n_{un} is the number of uncensored observations.

$$\begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_{n_{un}} \end{bmatrix}, \begin{bmatrix} \hat{\varepsilon}_1 \\ \hat{\varepsilon}_2 \\ \vdots \\ \hat{\varepsilon}_{n_{un}} \end{bmatrix}$$

where

$$\hat{y}_i = \hat{\alpha}_o + \hat{\beta}_o x_i ,$$

and

$$\hat{\varepsilon}_i = y_i - (\hat{\alpha}_o + \hat{\beta}_o x_i) \text{ for } i = 1, 2, \ldots, n_{un} .$$

Step 3. For each $X_i = x_i$ from uncensored observations, there will be n_{un} residuals from step 2 and corresponding $\hat{y}_i$. Calculate y_i corresponding to $X_i = x_i$ as follows:

for $X_1 = x_1$

$$\begin{bmatrix} \hat{y}_1 \\ \hat{y}_1 \\ \vdots \\ \hat{y}_1 \end{bmatrix}, \begin{bmatrix} \hat{\varepsilon}_1 \\ \hat{\varepsilon}_2 \\ \vdots \\ \hat{\varepsilon}_{n_{un}} \end{bmatrix}, \begin{bmatrix} \hat{y}_1 + \hat{\varepsilon}_1 & = y_{11} \\ \hat{y}_1 + \hat{\varepsilon}_2 & = y_{12} \\ \vdots & \\ \hat{y}_1 + \hat{\varepsilon}_{n_{un}} & = y_{1n_{un}} \end{bmatrix},$$

for $X_2 = x_2$

$$\begin{bmatrix} \hat{y}_2 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_2 \end{bmatrix}, \begin{bmatrix} \hat{\varepsilon}_1 \\ \hat{\varepsilon}_2 \\ \vdots \\ \hat{\varepsilon}_{n_{un}} \end{bmatrix}, \begin{bmatrix} \hat{y}_2 + \hat{\varepsilon}_1 = y_{21} \\ \hat{y}_2 + \hat{\varepsilon}_2 = y_{22} \\ \vdots \\ \hat{y}_2 + \hat{\varepsilon}_{n_{un}} = y_{2n_{un}} \end{bmatrix},$$

$$\vdots \qquad\qquad \vdots$$

for $X_{n_{un}} = x_{n_{un}}$

$$\begin{bmatrix} \hat{y}_{n_{un}} \\ \hat{y}_{n_{un}} \\ \vdots \\ \hat{y}_{n_{un}} \end{bmatrix}, \begin{bmatrix} \hat{\varepsilon}_1 \\ \hat{\varepsilon}_2 \\ \vdots \\ \hat{\varepsilon}_{n_{un}} \end{bmatrix}, \begin{bmatrix} \hat{y}_{n_{un}} + \hat{\varepsilon}_1 = y_{n_{un}1} \\ \hat{y}_{n_{un}} + \hat{\varepsilon}_2 = y_{n_{un}2} \\ \vdots \\ \hat{y}_{n_{un}} + \hat{\varepsilon}_{n_{un}} = y_{n_{un}n_{un}} \end{bmatrix}.$$

Step 4. For each $X_i = x_i$, the corresponding $y_{i1}, y_{i2}, \ldots, y_{in_{un}}$ are calculated. Figure out $\hat{G}(y_{i1})$, $\hat{G}(y_{i2})$, ..., and $\hat{G}(y_{in_{un}})$ by evaluating from the empirical censoring distribution function calculated in Section 3.2.

Step 5. Compute the estimates of bias for given $X_i = x_i$ by the formula below:

For given $X_i = x_i$

$$\hat{\text{Bias}}\ (Y_i \mid Y_i \le C_i,\ X_i = x_i) = \frac{\hat{E}(\varepsilon_j G(Y_{ij}) \mid X_i = x_i)}{\hat{E}(1-G(Y_{ij}) \mid X_i = x_i)}$$

$$= \frac{\hat{E}(\hat{\varepsilon}_j \hat{G}(y_{ij}) \mid X_i = x_i)}{\hat{E}(1-\hat{G}(y_{ij}) \mid X_i = x_i)}$$

$$= \frac{\sum_{j=1}^{n_{un}} \hat{\varepsilon}_j \hat{G}(y_{ij})/n_{un}}{\sum_{j=1}^{n_{un}} (1-\hat{G}(y_{ij}))/n_{un}}$$

thus,

$$\hat{\text{Bias}}\ (Y_i \mid Y_i \le C_i,\ X_i = x_i) = \frac{\sum_{j=1}^{n_{un}} \hat{\varepsilon}_j \hat{G}(y_{ij})}{n_{un} - \sum_{j=1}^{n_{un}} \hat{G}(y_{ij})} ,$$

for $i = 1, 2, \ldots, n_{un}$ and $j = 1, 2, \ldots, n_{un}$.

Step 6. Perform the calculation as follows, for simplicity, assuming that the original uncensored observations are

$$\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ \vdots & \vdots \\ 1 & x_{n_{un}} \end{bmatrix} \text{ and } \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_{n_{un}} \end{bmatrix}$$

the adjusted uncensored observations with their associated estimates of the biases are as follows:

$$\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ \vdots & \vdots \\ 1 & x_{n_{un}} \end{bmatrix} , \begin{bmatrix} y_1^* & = y_1 & + \hat{Bias}\ (Y_1 | Y_1 \le C_1,\ X_1 = x_1) \\ y_2^* & = y_2 & + \hat{Bias}\ (Y_2 | Y_2 \le C_2,\ X_2 = x_2) \\ y_3^* & = y_3 & + \hat{Bias}\ (Y_3 | Y_3 \le C_3,\ X_3 = x_3) \\ \vdots & \vdots & \vdots \\ y_{n_{un}}^* & = y_{n_{un}} & + \hat{Bias}\ (Y_{n_{un}} | Y_{n_{un}} \le C_{n_{un}},\ X_{n_{un}} = x_{n_{un}}) \end{bmatrix} .$$

Step 7. Calculate the estimates of α and β by least squares method from the observations in step 6.

Let

$$X_{un} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ \vdots & \vdots \\ 1 & x_{n_{un}} \end{bmatrix} , \quad \hat{\underset{\sim}{Y}}^* = \begin{bmatrix} y_1^* \\ y_2^* \\ y_3^* \\ \vdots \\ y_{n_{un}}^* \end{bmatrix} ,$$

then,

$$\hat{\hat{\underset{\sim}{\beta}}} = (X_{un}^T X_{un})^{-1} X_{un}^T \hat{\underset{\sim}{Y}}^* ,$$

and $\hat{V}(\hat{\underset{\sim}{\beta}}) = \hat{\sigma}^2 (X_{un}^T X_{un})^{-1}$ used as an approximation of $V(\hat{\underset{\sim}{\beta}})$

where $$\hat{\sigma}^2 = \frac{(\underset{\sim}{Y}_{un} - X_{un}\hat{\beta})^T(\underset{\sim}{Y}_{un} - X_{un}\hat{\underset{\sim}{\beta}})}{n_{un} - 2}, \quad n_{un} > 2$$

and $$\underset{\sim}{Y}_{un} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ \vdots \\ y_{n_{un}} \end{bmatrix}$$

3.4 Diagram of the Adjusted Method for Censored Observations

Collected data

$(x_1, t_1, \delta_1), (x_2, t_2, \delta_2), \ldots, (x_n, t_n, \delta_n)$.

↓

Estimate α, β by ordinary least squares method from only uncensored observations

$\hat{y}_i = \hat{\alpha}_o + \hat{\beta}_o x_i, \quad i = 1, 2, \ldots, n_{un}$

↓

Calculate the residuals for all uncensored observations

$$\begin{array}{lcll} \hat{\varepsilon}_1 & = & y_1 & - (\hat{\alpha}_o + \hat{\beta}_o x_1) \\ \hat{\varepsilon}_2 & = & y_2 & - (\hat{\alpha}_o + \hat{\beta}_o x_2) \\ \vdots & & \vdots & \quad\quad \vdots \\ \hat{\varepsilon}_{n_{un}} & & y_{n_{un}} & - (\hat{\alpha}_o + \hat{\beta}_o x_{n_{un}}) \end{array}$$

↓

Compute empirical censoring distribution function by Kaplan-Meier Estimation Procedure.

For $X_i = x_i$,

$$\hat{\varepsilon}_{ij} = \hat{\varepsilon}_j, \; j = 1, 2, \ldots, n_{un}$$

$$\hat{y}_{ij} = \hat{y}_i, \; j = 1, 2, \ldots, n_{un}$$

and $y_{ij} = \hat{y}_{ij} + \hat{\varepsilon}_{ij}$ (or $\hat{\varepsilon}_j$), $j = 1, 2, \ldots, n_{un}$
$j = 1, 2, \ldots, n_{un}$.

Figure out the distribution function of

$\hat{G}(\hat{y}_{ij} + \hat{\varepsilon}_j)$ for a given $X_i = x_i$, $j = 1, 2, \ldots, n_{un}$
$j = 1, 2, \ldots, n_{un}$.

Compute the estimates of bias and variance for each $X_i = x_i$ using

$$\hat{\text{Bias}}\,(Y_i | Y_i \le C_i,\, X_i = x_i) = \frac{\sum_{j=1}^{n_{un}} \hat{\varepsilon}_j \hat{G}(y_{ij})}{n_{un} - \sum_{j=1}^{n_{un}} \hat{G}(y_{ij})} \quad \text{for } j = 1, 2, \ldots, n_{un}, \; i = 1, 2, \ldots, n_{un}$$

Having x_i, y_i $\hat{\text{Bias}}_i$ $i = 1, 2, \ldots, n_{un}$,

calculate $\hat{y}_i^* = y_i + \hat{\text{Bias}}_i$

Estimate the regression coefficients by least squares estimation procedure

$$\hat{\hat{\beta}} = (X_{un}^T X_{un})^{-1} X_{un}^T \hat{Y}^*$$

where $X_{un} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \\ 1 & x_{n_{un}} \end{bmatrix}$, $\hat{Y}^* = \begin{bmatrix} y_1^* \\ y_2^* \\ \vdots \\ y_{n_{un}}^* \end{bmatrix}$

and $\hat{V}(\hat{\hat{\beta}}) = \hat{\hat{\sigma}}^2 (X_{un}^T X_{un})^{-1}$

where $\hat{\hat{\sigma}}^2 = \dfrac{(Y_{un} - X_{un}\hat{\hat{\beta}})^T (Y_{un} - X_{un}\hat{\hat{\beta}})}{n_{un} - 2}$, $n_{un} > 2$

$$Y_{un} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{n_{un}} \end{bmatrix}$$

Bootstrapping Method

3.5 Bootstrapping for Censored Data

Suppose we have a real-valued statistic $\hat{\theta}(X_1, X_2, \ldots, X_n)$ where X_i are independent and identically distributed with some unknown probability distribution

$$X_1, X_2, \ldots, X_n \underset{\sim}{\text{iid}} F .$$

Having observed

$$X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n ,$$

we wish to estimate a given functional $\theta(F)$, perhaps the mean, median, correlation, etc., and we agree to use the estimate $\hat{\theta} = \theta(\hat{F})$ where $\hat{F}$ is the empirical distribution function obtained by putting mass $\frac{1}{n}$ at each observed value x_i. We wish to assign some measure of accuracy to $\hat{\theta}$.

Let $\sigma(F)$ be some measure of accuracy that we would use if F were known, for example $\sigma(F) = SD_F(\hat{\theta})$, the standard deviation of $\hat{\theta}$ when $X_1, X_2, \ldots, X_n \underset{\sim}{\text{iid}} F$. The bootstrap estimate of accuracy is $\sigma_{BOOT} = \sigma(\hat{F})$. In other words, $\hat{\sigma}_{BOOT}$ is the measure of accuracy we would obtain if the true F equaled the nonparametric $\hat{F}$. This has been shown by Efron (1979).

In order to calculate $\hat{\sigma}_{BOOT}$, it is usually necessary to employ computer simulation methods.

(i) A "bootstrap sample" $X_1^*, X_2^*, \ldots, X_n^*$ is drawn from $\hat{F}$, in which each X_i^* independently takes value x_j with probability $\frac{1}{n}$, $j = 1, 2, \ldots, n$. In other words, $X_1^*, X_2^*, \ldots, X_n^*$ is an independent sample of size n drawn with replacement from the set of observations $\{x_1, x_2, \ldots, x_n\}$.

(ii) Step (i) gives a bootstrap empirical distribution function $\hat{F}^*$, the empirical distribution of the n values $X_1^*, X_2^*, \ldots X_n^*$, and a corresponding bootstrap value $\hat{\theta}^* = \theta(\hat{F}^*)$.

(iii) Steps (i) and (ii) are independently repeated a large number of times, say N, giving bootstrap values

$$\hat{\theta}^{*1},\ \hat{\theta}^{*2},\ \ldots,\ \hat{\theta}^{*N}\ .$$

(iv) The value of $\hat{\sigma}_{BOOT}$ is approximated, in the case when $\sigma(F)$ is the standard deviation, by the sample standard deviation of the $\hat{\theta}^{*}$ values,

$$\hat{\sigma}_{BOOT} = \sqrt{\frac{\sum_{j=1}^{N} (\hat{\theta}^{*j})^2 - (\sum_{j=1}^{N} \hat{\theta}^{*j})^2/N}{N-1}} \tag{3.3}$$

Right censored data is of the form $\{(x_1,\delta_1),\ (x_2,\delta_2),\ \ldots,\ (x_n,\delta_n)\}$ where x_j is the j^{th} ordered observation, censored or not, and

$$\delta_j = \begin{cases} 1 & \text{if } x_j \text{ is uncensored} \\ 0 & \text{if } x_j \text{ is censored} \end{cases} \quad . \tag{3.4}$$

We have some estimated functional $\hat{\theta} = \theta(\text{data})$ based on $\{(x_1,\delta_1),$ $(x_2,\delta_2),\ \ldots,\ (x_n,\delta_n)\}$. $\hat{\sigma}_{BOOT}$ in the censored case is the same as in the uncensored case. This has been evaluated by Efron (1967) and Gilbert (1962). They showed that the simple method of bootstrap sampling for censored data described later is the same as the one given at the beginning of this paragraph, except that the individual data points are now the pairs (x_j,δ_j).

(i) We draw a bootstrap sample $(X_1^*,\delta_1^*),\ (X_2^*,\delta_2^*),\ \ldots,\ (X_n^*,\delta_n^*)$ by independent sampling n times with replacement from the set of pairs $\{(x_1,\delta_1),\ (x_2,\delta_2),\ \ldots,\ (x_n,\delta_n)\}$.

(ii) Letting data* represent this artificial data set, we calculate $\hat{\theta}^* = \theta(\text{data})$.

(iii) We independently repeat step (i) and (ii) N times, obtaining $\theta^{*1}, \theta^{*2}, \theta^{*3}, \ldots, \theta^{*N}$.

(iv) Calculate $\hat{\sigma}_{BOOT}$ by

$$\hat{\sigma}_{BOOT} = \sqrt{\frac{\sum_{j=1}^{N} (\hat{\theta}^{*j})^2 - (\sum_{j=1}^{N} \hat{\theta}^{*j})^2/N}{N-1}}$$

3.6 The Bootstrap Estimate of Bias

The idea originally was introduced by Quenouille (1949) as a means of reducing the bias in an estimator (see Miller (1974)). We wish to estimate the bias of a statistic $\hat{\theta} = \theta(\hat{F}_n)$; then the bias is defined.

$$\text{bias} = E\{\theta(\hat{F}_n) - \theta(F)\} .$$

The bootstrap estimate of bias is defined as

$$\text{bias}_{BOOT} = E^*\{\theta(\hat{F}^*) - \theta(\hat{F}_n)\}$$

where E^* and $\hat{F}^*$ denote expectation in terms of bootstrap sampling and the bootstrap empirical probability distribution respectively. In practice, the bootstrap estimate of bias is approximated by computer simulation methods. The steps (i), (ii), and (iii) are the same as those in Section 1. At step (iv), we calculate

$$\widehat{\text{Bias}}_{BOOT} = \frac{1}{N} \sum_{j=1}^{N} \hat{\theta}^{*j} - \hat{\theta} .$$

We would use this result to correct the estimator for bias in the following way:

$$\hat{\theta}_{CORRECTED} = \hat{\theta} - \widehat{Bias}_{BOOT} .$$

3.7 Bootstrapping Regression Model With Censored Observations

This section is concerned with the presentation of the bootstrapping for linear regression model with censored data.

Consider the usual linear regression model

$$Y_i = \alpha + \beta X_i + \varepsilon_i, \ i = 1, 2, \ldots, n ,$$

where X_i are known constant covariates, α and β are unknown regression coefficients to be estimated, and ε_i are the independent random errors with unknown common distribution F such that

$$E(\varepsilon_i) = 0 ,$$

$$V(\varepsilon_i) = \sigma^2,$$

and $$\mathrm{Cov}(\varepsilon_i,\varepsilon_j) = 0, \ i \neq j, \ i,j = 1, 2, \ldots, n .$$

Let C_i, $i = 1, 2, \ldots, n$ be independent censoring random variable with unknown distribution G. Assume that C_i is independent of Y_i and X_i, for $i = 1, 2, \ldots, n$.

We observe

$$T_i = \mathrm{Min}(Y_i, C_i) ,$$

and

$$\delta_i = \begin{cases} 1 & \text{when } Y_i \le C_i \\ 0 & \text{when } Y_i > C_i \end{cases} .$$

The bootstrapping used in this method is done by resampling the residuals calculated from least-square estimator of α and β using all n observations, censored or uncensored. Then, we calculate the estimates of biases of those estimators as mentioned in the previous section. Finally, we correct the estimators using the estimates of biases.

Now, consider the following observed data

$$(t_i, x_i, \delta_i),$$

where

t_i = observed survival time, censored or uncensored,

x_i = observed covariate

δ_i = observed indicator, 1 or 0.

The regression coefficients are estimated as $\hat{\alpha}$ and $\hat{\beta}$ usually by the least squares estimation procedure. After α and β are estimated residuals are calculated as

$$\hat{\varepsilon}_i = t_i(\alpha,\beta) - \hat{t}_i(\hat{\alpha},\hat{\beta}), \; i = 1, 2, \ldots, n$$

i.e., the difference between the actual observations and the predicted observations. Let F_n be the empirical distribution function of the residuals, putting mass $\frac{1}{n}$ on each of $\hat{\varepsilon}_i$, i = 1, 2, ..., n,

$$\hat{F}: \; \text{mass } \frac{1}{n} \text{ at } \hat{\varepsilon}_i = t_i - \hat{t}_i(\alpha,\beta) \; .$$

Draw a bootstrap data set

$$t_i^* = \hat{\alpha} + \hat{\beta}x_i + \varepsilon_i^*, \; i = 1, 2, \ldots, n \; ,$$

where ε_i^* are independent bootstrap samples from $\hat{F}$.

Then use the least squares estimation procedure and compute the bootstrap estimates $\hat{\alpha}^*$ and $\hat{\beta}^*$ from bootstrap data. Independently repeat N times, obtaining bootstrap replications $\hat{\alpha}^{*1}, \hat{\alpha}^{*2}, \ldots, \hat{\alpha}^{*N}$ and $\hat{\beta}^{*1}, \hat{\beta}^{*2}, \ldots, \hat{\beta}^{*N}$.

Finally, the estimate of bias for $\hat{\alpha}$ is calculated by

$$\widehat{\text{Bias}}_{\text{BOOT}}(\hat{\alpha}) = \frac{1}{N}\sum_{j=1}^{N} \hat{\alpha}^{*j} - \hat{\alpha}, \text{ and}$$

the estimate of bias for $\hat{\beta}$ is

$$\widehat{\text{Bias}}_{\text{BOOT}}(\hat{\beta}) = \frac{1}{N}\sum_{j=1}^{N} \hat{\beta}^{*j} - \hat{\beta} \ .$$

The bootstrapping estimators of α and β for censored observations are computed by

$$\hat{\alpha}^*_{\text{CORRECTED}} = \hat{\alpha} - \widehat{\text{Bias}}_{\text{BOOT}}(\hat{\alpha});$$

$$\hat{\beta}^*_{\text{CORRECTED}} = \hat{\beta} - \widehat{\text{Bias}}_{\text{BOOT}}(\hat{\beta}), \text{ and use}$$

$$\hat{V}(\hat{\underset{\sim}{\beta}}^*_{\text{CORRECTED}}) = \hat{\sigma}^2 (X^T \Delta X)^{-1} \text{ as an approximation of } V(\hat{\underset{\sim}{\beta}}_{\text{CORRECTED}}),$$

where $\Delta = \{\delta_i\}$ the diagonal matrix. The i^{th} diagonal element δ_i is the indicator observation defined in the previous sections. The estimate of σ^2 is computed as

$$\hat{\sigma}^2 = \frac{(\underset{\sim}{t} - X\hat{\underset{\sim}{\beta}}^*_{\text{CORRECTED}})^T \Delta (\underset{\sim}{t} - X\hat{\underset{\sim}{\beta}}^*_{\text{CORRECTED}})}{\text{trace }(\Delta) - 2}, \text{ trace } (\Delta) > 2$$

where $$\underset{\sim}{t} = (t_1, t_2, \ldots, t_n)^T, \text{ and}$$

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} .$$

3.8 Diagram of Bootstrapping Simple Linear Regression for Censored Observations

Collect data

$(x_1,t_1,\delta_1),\ (x_2,t_2,\delta_2),\ \ldots,\ (x_n,t_n,\delta_n)$.

Estimate α and β by ordinary least squares estimation procedure, and predicted values are

$$\begin{aligned} \hat{t}_1 &= \hat{\alpha} + \hat{\beta}x_1 \\ \hat{t}_2 &= \hat{\alpha} + \hat{\beta}x_2 \\ &\ \ \vdots \\ \hat{t}_n &= \hat{\alpha} + \hat{\beta}x_n \ . \end{aligned}$$

Calculate the residuals

$$\begin{aligned} \hat{\varepsilon}_1 &= t_1 - (\hat{\alpha} + \hat{\beta}x_1) \\ \hat{\varepsilon}_2 &= t_2 - (\hat{\alpha} + \hat{\beta}x_2) \\ &\ \ \vdots \\ \hat{\varepsilon}_n &= t_n - (\hat{\alpha} + \hat{\beta}x_n) \ . \end{aligned}$$

Resample $\hat{\varepsilon}_1, \hat{\varepsilon}_2, \ldots, \hat{\varepsilon}_n$ by computer random number generator (copying $\hat{\varepsilon}_1, \hat{\varepsilon}_2, \ldots, \hat{\varepsilon}_n$ in memory, selecting $\hat{\varepsilon}_1, \hat{\varepsilon}_2, \ldots, \hat{\varepsilon}_n$ randomly with replacement by matching from uniform random generator).

Bootstrap Sample 1

$$\begin{bmatrix} \varepsilon_1^{*1} \\ \varepsilon_2^{*1} \\ \vdots \\ \varepsilon_n^{*1} \end{bmatrix} = \begin{bmatrix} \hat{\varepsilon}_7 \\ \hat{\varepsilon}_{20} \\ \vdots \\ \hat{\varepsilon}_3 \end{bmatrix} ; \begin{bmatrix} t_1^{*1} = \hat{\alpha}+\hat{\beta}x_1+\varepsilon_1^* \\ t_2^{*1} = \hat{\alpha}+\hat{\beta}x_2+\varepsilon_2^* \\ \vdots \\ t_n^{*1} = \hat{\alpha}+\hat{\beta}x_n+\varepsilon_n^* \end{bmatrix} ;$$

Bootstrap Sample 2

$$\begin{bmatrix} \varepsilon_1^{*2} \\ \varepsilon_2^{*2} \\ \vdots \\ \varepsilon_n^{*2} \end{bmatrix} = \begin{bmatrix} \hat{\varepsilon}_n \\ \hat{\varepsilon}_{20}^* \\ \vdots \\ \hat{\varepsilon}_n \end{bmatrix} ; \begin{bmatrix} t_1^{*2} = \hat{\alpha}+\hat{\beta}x_1+\varepsilon_1^{*2} \\ t_2^{*2} = \hat{\alpha}+\hat{\beta}x_1+\varepsilon_2^{*2} \\ \vdots \\ t_n^{*2} = \hat{\alpha}+\hat{\beta}x_n+\varepsilon_n^{*2} \end{bmatrix}$$

Bootstrap Sample N

$$\begin{bmatrix} \varepsilon_1^{*N} \\ \varepsilon_2^{*N} \\ \vdots \\ \varepsilon_n^{*N} \end{bmatrix} = \begin{bmatrix} \hat{\varepsilon}_{10} \\ \hat{\varepsilon}_{17} \\ \vdots \\ \hat{\varepsilon}_3 \end{bmatrix} ; \begin{bmatrix} t_1^{*N} = \hat{\alpha}+\hat{\beta}x_1+\varepsilon_1^{*N} \\ t_2^{*N} = \hat{\alpha}+\hat{\beta}x_2+\varepsilon_2^{*N} \\ \vdots \\ t_n^{*N} = \hat{\alpha}+\hat{\beta}x_n+\varepsilon_n^{*N} \end{bmatrix} .$$

↓

Calculate estimates $\hat{\alpha}^{*j}$ and $\hat{\beta}^{*j}$ using least squares estimation procedure for bootstrap regression observation j, j = 1, 2, ..., N. We will have

$$(\hat{\alpha}^{*1}, \hat{\beta}^{*1}), (\hat{\alpha}^{*2}, \hat{\beta}^{*2}), \ldots, (\hat{\alpha}^{*N}, \hat{\beta}^{*N}) .$$

↓

Find the estimates of biases using the calculation as follows:

$$\hat{\text{Bias}}_{\text{BOOT}}(\hat{\alpha}) = \frac{1}{N} \sum_{j=1}^{N} \hat{\alpha}^{*j} - \hat{\alpha}, \text{ and}$$

$$\hat{\text{Bias}}_{\text{BOOT}}(\hat{\beta}) = \frac{1}{N} \sum_{j=1}^{N} \hat{\beta}^{*j} - \hat{\beta} .$$

↓

Correcting the biases using corrected estimators calculated by

$$\hat{\alpha}^*_{CORRECTED} = \hat{\alpha} - \widehat{Bias}_{BOOT}(\hat{\alpha}), \text{ and}$$

$$\hat{\beta}^*_{CORRECTED} = \hat{\beta} - \widehat{Bias}_{BOOT}(\hat{\beta}) .$$

↓

The estimates of variances of parameters estimated are calculated by

$$\hat{V}(\hat{\underset{\sim}{\beta}}^*_{CORRECTED}) = \hat{\sigma}^2(X^T \Delta X)^{-1} ,$$

where

$$\hat{\underset{\sim}{\beta}}^*_{CORRECTED} = \begin{bmatrix} \hat{\alpha}^*_{CORRECTED} \\ \hat{\beta}^*_{CORRECTED} \end{bmatrix}$$

Δ = the diagonal matrix where the i^{th} diagonal element δ_i is the indicator observation, and

$$\hat{\sigma}^2 = \frac{(\underset{\sim}{t} - x\hat{\underset{\sim}{\beta}}^*_{CORRECTED})^T \Delta (\underset{\sim}{t} - x\hat{\underset{\sim}{\beta}}_{CORRECTED})}{\text{Trace }(\Delta) - 2} , \text{ trace }(\Delta) > 2,$$

$$\underset{\sim}{t} = (t_1, t_2, \ldots, t_n)^T .$$

3.9 Generalization of Multiple Linear Regression Methods for Censored Observations Under the Same Assumptions as the Previous Sections

Algorithm for the Adjusted Method

Collected data

$((x_{11}, x_{12}, \ldots, x_{1p}, t_1, \delta_1), (x_{21}, x_{22}, \ldots, x_{2p}, t_1, \delta_2), \ldots, (x_{n1}, x_{n2}, \ldots, x_{np}, t_n, \delta_n))^T$.

↓

Estimate $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_p$ by least squares estimation procedure using only uncensored observations;

$$\hat{\underset{\sim}{\beta}} = (X^T \Delta X)^{-1} X^T \Delta \underset{\sim}{t}$$

where

$$\hat{\underset{\sim}{\beta}} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_p)^T$$

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & & \vdots & \ddots & \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}$$

$$\underset{\sim}{t} = (t_1, t_2, \ldots, t_n)^T, \text{ and}$$

$$\Delta = \begin{bmatrix} \delta_1 & 0 & 0 & \cdots & 0 \\ 0 & \delta_2 & 0 & \cdots & 0 \\ 0 & 0 & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & \cdots & \cdots & \cdots & \delta_n \end{bmatrix} \quad \text{where trace } (\Delta) > p + 1 .$$

↓

Calculate the residuals; some are censored and some are not,

$$\hat{\varepsilon}_1 = t_1 - (\hat{\beta}_0 + \hat{\beta}_1 x_{11} + \hat{\beta}_2 x_{12} + \cdots + \hat{\beta}_p x_{1p})$$

$$\hat{\varepsilon}_2 = t_2 - (\hat{\beta}_0 + \hat{\beta}_1 x_{21} + \hat{\beta}_2 x_{22} + \cdots + \hat{\beta}_p x_{2p})$$

$$\vdots$$

$$\hat{\varepsilon}_n = t_n - (\hat{\beta}_0 + \hat{\beta}_1 x_{n1} + \hat{\beta}_2 x_{n2} + \cdots + \hat{\beta}_p x_{np})$$

↓

Compute empirical censoring distribution function by Kaplan-Meier Estimation Procedure.

For each $x_i = (1\ x_{1i}\ x_{2i}, \ldots, x_{pi})$

$\hat{\underset{\sim}{\varepsilon}}^*_i = (\hat{\varepsilon}_1\ \hat{\varepsilon}_2\ \hat{\varepsilon}_3, \ldots, \hat{\varepsilon}_n)^T$, we calculate

$\hat{t}_i = x_i \hat{\underset{\sim}{\beta}}$, and

$t_i = x_i \hat{\underset{\sim}{\beta}} + \hat{\varepsilon}_i$ for $i = 1, 2, \ldots, n$.

↓

Calculate the empirical censoring distribution function for each x_i, $i = 1, 2, \ldots, n$ as mentioned in Section 3.2.

For given $\underset{\sim}{x}_i = (1\ x_{1i}\ x_{2i}, \ldots, x_{pi})$

$$\hat{\text{Bias}}_i = \frac{\hat{E}(\varepsilon_j G(Y_{ij}) \mid \underset{\sim}{x}_i = (1\ x_{1i}\ x_{2i}, \ldots, x_{pi}))}{\hat{E}(1-G(Y_{ij}) \mid \underset{\sim}{x}_i = (1\ x_{1i}\ x_{21}, \ldots, x_{pi}))}$$

$$= \hat{\underset{\sim}{\varepsilon}}^{*T} \Delta \underset{\sim}{G}_i / \underset{\sim}{J}^T \Delta(\underset{\sim}{J} - \underset{\sim}{G}_i),$$

where $\underset{\sim}{G}_i$ = the jumps for each element of $\underset{\sim}{t}_i$ for $i = 1, 2, \ldots, n$,

$\underset{\sim}{J} = (1, 1, \ldots, 1)^T$, 1xn vector.

↓

Then, we have the original observations

$$\begin{bmatrix} t_1, & x_{11}, & x_{12}, & \ldots, & x_{1p}, & \delta_1 \\ t_2, & x_{21}, & x_{22}, & \ldots, & x_{2p}, & \delta_2 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ t_n, & x_{n1}, & x_{n2}, & \ldots, & x_{np}, & \delta_n \end{bmatrix}$$

the adjusted observations with their associated estimates of biases

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}, \quad \Delta = \begin{bmatrix} \delta_1 & & & & 0 \\ & \delta_2 & & & \\ & & \delta_3 & & \\ 0 & & & \ddots & \\ & & & & \delta_n \end{bmatrix}, \text{ and}$$

$$\underset{\sim}{t}_{adj} = \begin{bmatrix} t_1 + \hat{Bias}_1 \\ t_2 + \hat{Bias}_2 \\ t_3 + \hat{Bias}_3 \\ \vdots \\ t_n + \hat{Bias}_n \end{bmatrix} .$$

↓

Calculate the estimates of $\underset{\sim}{\beta}$ by the following method.

$$\hat{\hat{\underset{\sim}{\beta}}} = (X^T \Delta X)^{-1} X^T \underset{\sim}{t}_{adj} \quad \text{and approximation}$$

$$\hat{V}(\hat{\hat{\beta}}) = \hat{\hat{\sigma}}^2 (X^T \Delta X)^{-1}, \text{ where}$$

$$\hat{\hat{\sigma}}^2 = (\underset{\sim}{t} - X\hat{\hat{\underset{\sim}{\beta}}})^T \Delta (\underset{\sim}{t} - X\hat{\hat{\underset{\sim}{\beta}}}) / \text{Trace } (\Delta) - (p + 1), \text{ trace } (\Delta) > p + 1$$

Algorithm for the Bootstrapping Method

Collected data

$((x_{11}, x_{12}, \ldots, x_{1p}, t_1, \delta_1), (x_{21}, x_{22}, \ldots, x_{2p}, t_2, \delta_2), \ldots., (x_{n1}, x_{n2}, \ldots, x_{np}, t_n, \delta_n))^T$.

↓

Estimate $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_p$ by least squares estimation procedure

$$\hat{\underset{\sim}{\beta}} = (X^T X)^{-1} X^T \underset{\sim}{t} ,$$

where

$$\underset{\sim}{\hat{\beta}} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_p)^T ,$$

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & & & & \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} , \text{ and}$$

$$\underset{\sim}{t} = (t_1, t_2, \ldots, t_n)^T .$$

Calculate the residuals; some are censored and some are not.

$$\hat{\varepsilon}_1 = t_1 - (\hat{\beta}_0 + \hat{\beta}_1 x_{11} + \hat{\beta}_2 x_{12} + \ldots + \hat{\beta}_p x_{1p})$$

$$\hat{\varepsilon}_2 = t_2 - (\hat{\beta}_0 + \hat{\beta}_1 x_{21} + \hat{\beta}_2 x_{22} + \ldots + \hat{\beta}_p x_{2p})$$

$$\vdots$$

$$\hat{\varepsilon}_n = t_n - (\hat{\beta}_0 + \hat{\beta}_1 x_{n1} + \hat{\beta}_2 x_{n2} + \ldots + \hat{\beta}_p x_{np}) .$$

Resample $\hat{\varepsilon}_1, \hat{\varepsilon}_2, \ldots, \hat{\varepsilon}_n$ by copying $\hat{\varepsilon}_1, \hat{\varepsilon}_2, \ldots, \hat{\varepsilon}_n$ in memory; select them randomly with replacement by matching from uniform random generator. We will have

Bootstrap Sample 1

$$\begin{bmatrix} \varepsilon_1^{*1} \\ \varepsilon_2^{*1} \\ \varepsilon_3^{*1} \\ \vdots \\ \varepsilon_n^{*1} \end{bmatrix} = \begin{bmatrix} \hat{\varepsilon}_3 \\ \hat{\varepsilon}_{27} \\ \hat{\varepsilon}_n \\ \vdots \\ \hat{\varepsilon}_3 \end{bmatrix} ; \begin{bmatrix} t_1^{*1} = \hat{\beta}_0 + \hat{\beta}_1 x_{11} + \hat{\beta}_2 x_{12} + \cdots + \hat{\beta}_p x_{1p} + \varepsilon_1^{*1} \\ t_2^{*1} = \hat{\beta}_0 + \hat{\beta}_1 x_{21} + \hat{\beta}_2 x_{22} + \cdots + \hat{\beta}_p x_{2p} + \varepsilon_2^{*1} \\ t_3^{*1} = \hat{\beta}_0 + \hat{\beta}_1 x_{31} + \hat{\beta}_2 x_{32} + \cdots + \hat{\beta}_p x_{3p} + \varepsilon_3^{*1} \\ \vdots \\ t_n^{*1} = \hat{\beta}_0 + \hat{\beta}_1 x_{n1} + \hat{\beta}_2 x_{n2} + \cdots + \hat{\beta}_p x_{np} + \varepsilon_n^{*1} \end{bmatrix},$$

Bootstrap Sample 2

$$\begin{bmatrix} \varepsilon_1^{*2} \\ \varepsilon_2^{*2} \\ \varepsilon_3^{*2} \\ \vdots \\ \varepsilon_n^{*2} \end{bmatrix} = \begin{bmatrix} \hat{\varepsilon}_{100} \\ \hat{\varepsilon}_{16} \\ \hat{\varepsilon}_{13} \\ \vdots \\ \hat{\varepsilon}_8 \end{bmatrix} ; \begin{bmatrix} t_1^{*2} = \hat{\beta}_0 + \hat{\beta}_1 x_{11} + \hat{\beta}_2 x_{12} + \cdots + \hat{\beta}_p x_{1p} + \varepsilon_1^{*2} \\ t_2^{*2} = \hat{\beta}_0 + \hat{\beta}_1 x_{21} + \hat{\beta}_2 x_{22} + \cdots + \hat{\beta}_p x_{2p} + \varepsilon_2^{*2} \\ t_3^{*2} = \hat{\beta}_0 + \hat{\beta}_1 x_{31} + \hat{\beta}_2 x_{32} + \cdots + \hat{\beta}_p x_{3p} + \varepsilon_3^{*2} \\ \vdots \\ t_n^{*2} = \hat{\beta}_0 + \hat{\beta}_1 x_{n1} + \hat{\beta}_2 x_{n2} + \cdots + \hat{\beta}_p x_{np} + \varepsilon_n^{*2} \end{bmatrix},$$

$$\vdots$$

Bootstrap Sample N

$$\begin{bmatrix} \varepsilon_1^{*N} \\ \varepsilon_2^{*N} \\ \varepsilon_3^{*N} \\ \vdots \\ \varepsilon_n^{*N} \end{bmatrix} = \begin{bmatrix} \hat{\varepsilon}_1 \\ \hat{\varepsilon}_1 \\ \hat{\varepsilon}_{116} \\ \vdots \\ \hat{\varepsilon}_{87} \end{bmatrix} ; \begin{bmatrix} t_1^{*N} = \hat{\beta}_0 + \hat{\beta}_1 x_{11} + \hat{\beta}_2 x_{12} + \cdots + \hat{\beta}_p x_{1p} + \varepsilon_1^{*N} \\ t_2^{*N} = \hat{\beta}_0 + \hat{\beta}_1 x_{21} + \hat{\beta}_2 x_{22} + \cdots + \hat{\beta}_p x_{2p} + \varepsilon_2^{*N} \\ t_3^{*N} = \hat{\beta}_0 + \hat{\beta}_1 x_{31} + \hat{\beta}_2 x_{32} + \cdots + \hat{\beta}_p x_{3p} + \varepsilon_3^{*N} \\ \vdots \\ t_n^{*N} = \hat{\beta}_0 + \hat{\beta}_1 x_{n1} + \hat{\beta}_2 x_{n2} + \cdots + \hat{\beta}_p x_{np} + \varepsilon_n^{*N} \end{bmatrix}.$$

Calculate estimates $\hat{\underset{\sim}{\beta}}^{*j}$ using least squares estimation procedure. We will have

$$\hat{\underset{\sim}{\beta}}^{*j} = (X^T X)^{-1} X^T \underset{\sim}{t}^{*j} \quad \text{for } j = 1, 2, \ldots, N,$$

where

$$\hat{\underset{\sim}{\beta}}^{*j} = (\hat{\beta}_0^{*j}, \hat{\beta}_1^{*j}, \hat{\beta}_2^{*j}, \ldots, \hat{\beta}_p^{*j})^T, \text{ and}$$

$$\underset{\sim}{t}^{*j} = (t_1^{*j}, t_2^{*j}, t_3^{*j}, \ldots, t_n^{*j})^T .$$

↓

The estimates of biases are calculated as follows:

$$\widehat{\text{Bias}}_{\text{BOOT}}(\hat{\underset{\sim}{\beta}}) = \Big(\frac{1}{N}\sum_{j=1}^{N} \hat{\beta}_0^{*j} - \hat{\beta}_0, \frac{1}{N}\sum_{j=1}^{N} \hat{\beta}_1^{*j} - \hat{\beta}_1, \frac{1}{N}\sum_{j=1}^{N} \hat{\beta}_2^{*j} - \hat{\beta}_2, \ldots\ldots, \frac{1}{N}\sum_{j=1}^{N} \hat{\beta}_p^{*j} - \hat{\beta}_p\Big)^T .$$

↓

Then the corrected estimators of $\underset{\sim}{\beta}$ are

$$\hat{\underset{\sim}{\beta}}_{\text{CORRECTED}} = \Big(\hat{\beta}_0 - \big(\frac{1}{N}\sum_{j=1}^{N} \hat{\beta}_0^{*j} - \hat{\beta}_0\big), \hat{\beta}_1 - \big(\frac{1}{N}\sum_{j=1}^{N} \hat{\beta}_1^{*j} - \hat{\beta}_1\big), \ldots\ldots, \hat{\beta}_p - \big(\frac{1}{N}\sum_{j=1}^{N} \hat{\beta}_p^{*j} - \hat{\beta}_p\big)\Big)^T.$$

↓

The estimates of variances are approximated by

$$\hat{V}(\hat{\beta}^{*}_{\text{CORRECTED}}) = \hat{\sigma}^2 (X^T \Delta X)^{-1},$$

where

$$\Delta = \begin{bmatrix} \delta_1 & & & & \\ & \delta_2 & & 0 & \\ & & \delta_3 & & \\ & 0 & & \ddots & \\ & & & & \delta_n \end{bmatrix}, \text{ and}$$

$$\hat{\sigma}^2 = \frac{(\underset{\sim}{t} - X\hat{\underset{\sim}{\beta}}_{\text{CORRECTED}})^T \Delta (\underset{\sim}{t} - X\hat{\underset{\sim}{\beta}}_{\text{CORRECTED}})}{\text{trace}(\Delta) - (p+1)}, \text{ trace}(\Delta) > p + 1$$

$$\underset{\sim}{t} = (t_1, t_2, \ldots, t_n)^T .$$

CHAPTER IV

COMPUTER RESULTS

4.1 Design of the Simulation Study

In this section, we examine how censoring mechanism, amount of censoring, and sample size affect the performance of the estimators from all four methods.

1. The different levels of the survival time distribution factor corresponding to covariate x_i, ε_i, and $(\alpha,\beta)^T$. In this study, x_i and ε_i have two possible conditions: $x_i = 2i$ and $x_i \sim U(0,100)$ where U refers to the uniform distribution whereas $\varepsilon_i \sim N(0,1)$ and $\varepsilon_i \sim N(0,100)$ where N refers to the normal distribution. $(\alpha,\beta)^T$ are fixed as (1,0.2), (10,0.2), and (1,-0.4). The errors (ε_i) were generated by drawing pseudo-random variates from the normal distribution. The covariates (x_i) when $x_i \sim U(0,100)$ were generated by drawing pseudo-random variates from the uniform distribution. Then we have $y_i = \alpha + \beta x_i + \varepsilon_i$.

2. The three levels of the censoring factor correspond to random, fixed, and fractional censoring mechanisms. For the random censorship model, the censoring times (c_i) were obtained by different pseudo-random variates independent from pseudo-random variates in 1. For the fixed censoring mechanism, the c_i's were assigned a prespecified fixed value. For the fractional censoring mechanism, the y_i's were first generated from step 1 and at the same time the pseudo-random variates from uniform distribution (0,1) were generated. A cutoff value (e.g., .25)

corresponding to the desired censoring level (in this case, 25 percent) was used to determine which individuals were to be censored (< cutoff) or uncensored (≥ cutoff). If an individual were to be censored, another random U(0,1) number would be generated and the y_i multiplied by the random number. The observed data were then obtained by

$$t_i = \min(y_i, c_i)$$

$$\delta_i = 1 \text{ if } y_i \le c_i$$

$$= 0 \text{ if } y_i > c_i .$$

3. The three levels of amount of censoring correspond to 25%, 50%, and 75% censoring.

4. The different levels of sample size are n = 10, 20, 25, 30, 50 and 75.

Assessing Performance

The performance of the four methods is to be assessed on the basis of MSE, the mean square error, computed by MSE = $(\text{bias})^2$ + variance of an estimate. A bias is calculated by using the value of a parameter estimated subtracted from the average of all estimates over 100 trials, i.e., bias of $\hat{\alpha}$ is calculated by $\frac{1}{100}\sum_{s=1}^{100} \hat{\alpha}_s - \alpha$ and bias of $\hat{\beta}$ is calculated by $\frac{1}{100}\sum_{s=1}^{100} \hat{\beta}_s - \beta$. The variance is then calculated by the formula

$$\sum_{s=1}^{100} (\hat{\alpha}_s - \sum_{s=1}^{100} \hat{\alpha}_s/100)/100 \text{ for } \hat{\alpha} \text{ and } \sum_{s=1}^{100} (\hat{\beta}_s - \sum_{s=1}^{100} \hat{\beta}_s/100)/100 \text{ for } \hat{\beta} .$$

Therefore, we have

$$\mathrm{MSE}(\hat{\alpha}) \;=\; (\frac{1}{100}\sum_{s=1}^{100} \hat{\alpha}_s - \alpha)^2 + \sum_{s=1}^{100} (\hat{\alpha}_s - \sum_{s=1}^{100} \hat{\alpha}_s/100)/100, \text{ and}$$

$$\mathrm{MSE}(\hat{\beta}) \;=\; (\frac{1}{100}\sum_{s=1}^{100} \hat{\beta}_s - \beta)^2 + \sum_{s=1}^{100} (\hat{\beta}_s - \sum_{s=1}^{100} \hat{\beta}_s/100)\,100 \;.$$

Since both Buckley and James' method and Miller's method are iterative methods, they require starting values of the estimates. Only the uncensored observation y_i receives nonzero weight. For this reason, it makes sense to use as starting values $(\hat{\alpha}_o, \hat{\beta}_o)^T$ the ordinary unweighted least squares estimator applied to only the uncensored observations (Miller and Halpern, 1982) for Miller's method. For Buckley and James' method, since the estimators use values for dependent variable at every x_i, it seems sensible to take for the starting values $(\hat{\alpha}_o, \hat{\beta}_o)^T$ the least squares estimators treating all the observations as uncensored whether they are uncensored or not (Miller and Halpern, 1982). All computations were performed using SAS packages and FORTRAN programs.

4.2 Results of the Simulation Study

In this section, we discuss the performance of the estimators from all four methods. We do not intend to argue that the estimators from the adjusted method and those from the bootstrapping method will be able to replace Buckley and James' method and Miller's method in all experiments. Rather, we wish to evaluate both proposed methods in light of the performance of Buckley and James' method and Miller's method at different settings. A reasonable overall performance would suggest that the proposed methods may be of use when one is not prepared to adopt Buckley and James' method and Miller's method.

In Table I - Table III, under the uniform random censoring variable (c_i) from 10 + α to 50 + α where α = 1, $\varepsilon_i \sim N(0,1)$, β = 0.2, and x_i = 2i we compute the estimates of α and β, their biases, and their variances for all four methods. The sample size is increased from 10 to 30. MSE performance is shown in Figure 1. The results show that increasing the sample size under the conditions mentioned would reduce the MSE of the estimates. The estimates from both proposed methods which provide indistinguishable MSE give nearly the same MSE as the MSE from the Buckley and James' method in both estimators. The MSE of the estimates from Miller's method are much higher than the others. Miller's method are remarkably biased.

Table IV - Table VI show that if the censoring variable (C_i) has the form $C_i = 1.5x_i - 0.015x_i^2 + \alpha$ where α, β, and ε_i are the same as case 1 with the sample size increases from 10 to 30, the result becomes almost the same as in case 1. The MSE from both of the proposed methods and from Buckley and James' method are remarkably indistinguishable. However, using both proposed methods provide better results than using any of the other methods. Miller's method provides the worst MSE in this case. Moreover, it shows strong bias for $\hat{\alpha}$. The MSE for all methods are decreased as the sample size is increased as shown in Figure 2.

If C_i is changed to be fixed value, C_i = 31 keeping α, β, ε_i the same as in case 1 and case 2. There are some differences between the MSE from the proposed methods and the USE from Buckley and James' method. Miller's method gives the worst MSE. Even between the adjusted method and the bootstrapping method there are different MSE. The adjusted method will be reasonably used in this case. However, the bootstrapping method could be a good substitute for Buckley and James' method and

Miller's. There are no biases shown up except the bias of $\hat{\alpha}$ from Miller's method. The results are shown in Table VII - Table IX and Figure 3.

In Table X - Table XII, $c_i \sim U(\alpha + \beta x_i, \alpha + \beta x_i + 20)$ where $\alpha = 1$, $\beta = 0.2$, $x_i = 2i$. Both adjusted method and bootstrapping method provide little better results than Buckley and James' method. Miller's method again provides the worst result. It shows the bias of $\hat{\alpha}$ as well. Figure 4 has shown the comparison of MSE among these four methods as the sample size is increased from 10 to 30.

In Table XIII - Table XV and Figure 5, c_i is generated as $U(0,50)$. The other parameters are the same as in the previous cases. In Table XVI- Table XVIII and Figure 6, c_i is generated as $N(40 + \alpha, 16)$ and $\beta = -0.4$, $\alpha = 1$. The other random generatings are the same as the previous cases. The biases are remarkably shown up among all four methods. However, the adjusted method and the bootstrapping method are still the best candidates.

In Table XIX - Table XXI and Figure 7, $x_i = 2i$, $\alpha = 10$, $\beta = 0.2$, C_i is generated as $U(\alpha + \beta x_i, \alpha + \beta x_i + 40)$, and $\varepsilon_i \sim N(0,100)$. The sample sizes considered are 25, 50, and 75. In Table XXII - Table XXIV and Figure 8, $x_i = 2i$, $\alpha = 10$, $\beta = 0.2$, $c_i = 30$ (fixed), and $\varepsilon_i \sim N(0,100)$. In Table XXV - Table XXVII and Figure 9, $x_i = 2i$, $\alpha = 10$, $\beta = 0.2$, $c_i = 1.5x_i - 0.015x_i^2$, and $\varepsilon_i \sim N(0,100)$. In Table XXVIII - Table XXX and Figure 10, $x_i \sim U(0,100)$, $\alpha = 10$, $\beta = 0.2$, $c_i \sim U(0,50)$, and $\varepsilon_i \sim N(0,100)$. From most of the cases the results show that the adjusted method and the bootstrapping method provide the MSE of the estimates better than Buckley and James' method. Miller's method always provides the worst results. However, most of the cases shows biases. The MSE of all methods decreased

TABLE I

SIMULATIONS CALCULATING THE ESTIMATES OF α AND β BASED ON 100 REPLICATIONS ($\alpha = 1$, $\beta = 0.2$, $C_i \sim U(10 + \alpha, 50 + \alpha)$, $X_i = 2i$, $\varepsilon_i \sim N(0,1)$ AND SAMPLE SIZE = 10)

Method / Estimator	α				
	$\hat{\alpha}$ (1)	$Var(\hat{\alpha})$ (2)	Bias (3)	MSE (4)	Z-Value (5)
Adjusted Method	0.999447	0.439986	-0.000553	0.4399863	-0.0083369
Bootstrapping	1.01129	0.443872	0.01129	0.443994	0.1694589
Buckley and James	0.999447	0.439986	-0.000553	0.4399863	-0.0083369
Miller	0.678699	0.708981	-0.321301	0.8122153	-3.815881

Method / Estimator	β				
	$\hat{\beta}$ (6)	$Var(\hat{\beta})$ (7)	Bias (8)	MSE (9)	Z-Value (10)
Adjusted Method	0.200219	0.00336882	0.000219	0.00336884	0.0377316
Bootstrapping	0.199681	0.00342705	-0.000319	0.00342715	-0.0054491
Buckley and James	0.200219	0.00336882	0.000219	0.00336844	0.0377316
Miller	0.19071	0.00584167	-0.00929	0.0059279	-1.2154786

TABLE II

SIMULATIONS CALCULATING THE ESTIMATES OF α AND β BASED ON 100 REPLICATIONS ($\alpha = 1$, $\beta = 0.2$, $C_i \sim U(10 + \alpha, 50 + \alpha)$, $X_i = 2i$, $\varepsilon_i \sim N(0,1)$ AND SAMPLE SIZE = 20)

Method / Estimator	α				
	$\hat{\alpha}$ (1)	Var($\hat{\alpha}$) (2)	Bias (3)	MSE (4)	Z-Value (5)
Adjusted Method	0.957667	0.240447	-0.042333	0.242239	-0.863315
Bootstrapping	0.993426	0.241537	-0.006574	0.2415802	-0.1337635
Buckley and James	0.957667	0.240447	-0.042333	0.242239	-0.863315
Miller	0.400642	0.366883	-0.599358	0.726113	-9.8951477

Method / Estimator	β				
	$\hat{\beta}$ (6)	Var($\hat{\beta}$) (7)	Bias (8)	MSE (9)	Z-Value (10)
Adjusted Method	0.201577	0.00040056	0.001577	0.00040085	0.787523
Bootstrapping	0.199835	0.00040582	-0.000165	0.00040583	-0.0819083
Buckley and James	0.201577	0.00040056	0.001577	0.00040085	0.7879523
Miller	0.199335	0.000772238	-0.000665	0.000972642	-0.2393016

TABLE III

SIMULATIONS CALCULATING THE ESTIMATES OF α AND β BASED ON 100 REPLICATIONS
($\alpha = 1$, $\beta = 0.2$, $C_i \sim U(10 + \alpha, 50 + \alpha)$, $X_i = 2i$, $\varepsilon_i \sim N(0,1)$
AND SAMPLE SIZE = 30)

Method \ Estimator	α				
	$\hat{\alpha}$ (1)	Var($\hat{\alpha}$) (2)	Bias (3)	MSE (4)	Z-Value (5)
Adjusted Method	0.970852	0.104528	-0.029148	0.1053776	-0.9015554
Bootstrapping	0.988504	0.117337	-0.11496	0.1174691	-0.3356056
Buckley and James	0.971011	0.140484	-0.028989	0.1413243	-0.7734278
Miller	0.705478	0.181211	-0.294522	0.2679542	-6.9187153

Method \ Estimator	β				
	$\hat{\beta}$ (6)	Var($\hat{\beta}$) (7)	Bias (8)	MSE (9)	Z-Value (10)
Adjusted Method	0.200583	0.0001334	0.000583	0.000133739	0.5047666
Bootstrapping	0.201337	0.0001856	0.001337	0.000187387	0.9813917
Buckley and James	0.202154	0.0002103	0.002154	0.000214939	1.4853406
Miller	0.211448	0.0004247	0.011448	0.000555756	5.5550566

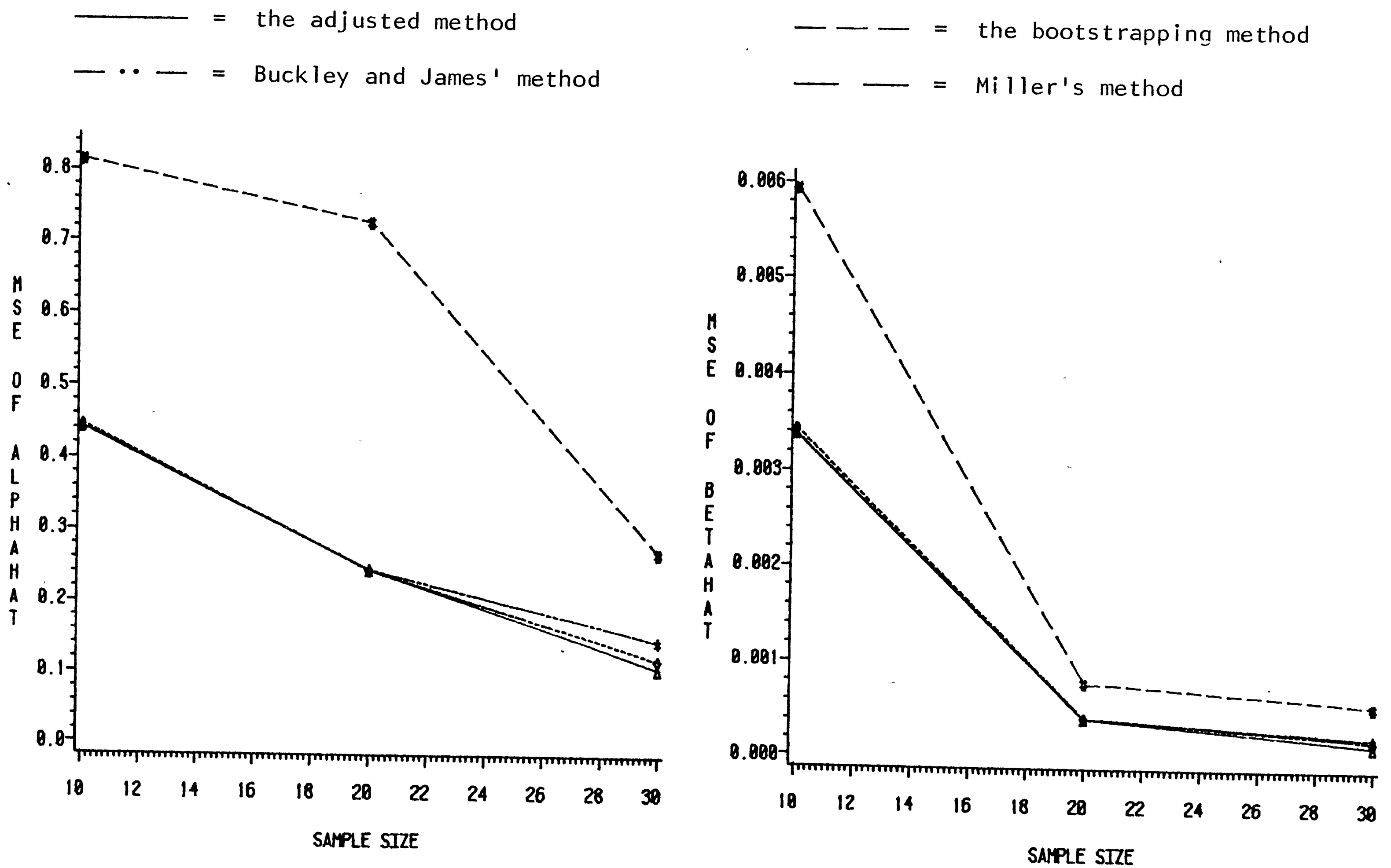


Figure 1. MSE of the Estimates of α and β Based on 100 Replications (α = 1, β = 0.2, $C_i \sim U(10 + \alpha, 50 + \alpha)$, $X_i = 2i$, $\varepsilon_i \sim N(0,1)$ and Sample Size = 10)

TABLE IV

SIMULATIONS CALCULATING THE ESTIMATES OF α AND β BASED ON 100 REPLICATIONS
($\alpha = 1$, $\beta = 0.2$, $C_i = 1.5X_i - 0.015\, X_i^2 + \alpha$, $X_i = 2i$, $\varepsilon_i \sim N(0,1)$
AND SAMPLE SIZE = 10)

Method \ Estimator	$\hat{\alpha}$ (1)	$\mathrm{Var}(\hat{\alpha})$ (2)	Bias (3)	MSE (4)	Z-Value (5)
	α				
Adjusted Method	0.991633	0.419551	-0.008367	0.419621	-0.1291747
Bootstrapping	1.01056	0.441765	0.01056	0.4418765	0.1588796
Buckley and James	0.991683	0.419651	-0.008367	0.419621	-0.1291747
Miller	0.677735	0.708401	-0.322265	0.8122557	-3.8288941

Method \ Estimator	$\hat{\beta}$ (6)	$\mathrm{Var}(\hat{\beta})$ (7)	Bias (8)	MSE (9)	Z-Value (10)
	β				
Adjusted Method	0.200752	0.00329412	0.000752	0.00324466	0.1310232
Bootstrapping	0.19973	0.00341924	-0.00027	0.00341927	-0.0461741
Buckley and James	0.200452	0.00329412	0.000752	0.00329466	0.1310232
Miller	0.190774	0.00584298	-0.009226	0.00592801	-1.2069692

TABLE V

SIMULATIONS CALCULATING THE ESTIMATES OF α AND β BASED ON 100 REPLICATIONS ($\alpha = 1$, $\beta = 0.2$, $C_i = 1.5X_i - 0.015X_i^2 + \alpha$, $X_i = 2i$, $\varepsilon_i \sim N(0,1)$ AND SAMPLE SIZE = 20)

Method \ Estimator	α				
	$\hat{\alpha}$ (1)	Var($\hat{\alpha}$) (2)	Bias (3)	MSE (4)	Z-Value (5)
Adjusted Method	0.953523	0.232007	-0.046477	0.2341671	-0.9649117
Bootstrapping	0.9846213	0.240059	-0.0153787	0.2402955	-0.3138778
Buckley and James	0.95323	0.232007	-0.046977	0.2341671	-0.9649117
Miller	0.400547	0.366831	-0.599453	0.7261749	-9.8974187

Method \ Estimator	β				
	$\hat{\beta}$ (6)	Var($\hat{\beta}$) (7)	Bias (8)	MSE (9)	Z-Value (10)
Adjusted Method	0.201725	0.000388848	0.001725	0.000391775	0.8747819
Bootstrapping	0.198544	0.00038141	0.001456	0.000383519	0.7455401
Buckley and James	0.201725	0.000388848	0.001725	0.000391775	0.8747519
Miller	0.199339	0.000772252	-0.000661	0.000772736	-0.2378604

TABLE VI

SIMULATIONS CALCULATING THE ESTIMATES OF α AND β BASED ON 100 REPLICATIONS
($\alpha = 1$, $\beta = 0.2$, $C_i = 1.5X_i - 0.015X_i^2 + \alpha$, $X_i = 2i$, $\varepsilon_i \sim N(0,1)$
AND SAMPLE SIZE = 30)

Method \ Estimator	α				
	$\hat{\alpha}$ (1)	$Var(\hat{\alpha})$ (2)	Bias (3)	MSE (4)	Z-Value (5)
Adjusted Method	0.983471	0.121452	-0.016529	0.1217252	-0.4742903
Bootstrapping	0.990632	0.124427	-0.009378	0.1245149	-0.2658599
Buckley and James	0.979423	0.124863	-0.020577	0.1252864	-0.5823246
Miller	0.789327	0.188424	-0.210673	0.2328071	-4.8533406

Method \ Estimator	β				
	$\hat{\beta}$ (6)	$Var(\hat{\beta})$ (7)	Bias (8)	MSE (9)	Z-Value (10)
Adjusted Method	0.201013	0.00014273	0.001013	0.000143726	0.8480031
Bootstrapping	0.212008	0.00018854	0.012008	0.000332692	0.7461113
Buckley and James	0.205598	0.00018881	0.005598	0.00022014	4.074102
Miller	0.200101	0.00040231	0.000101	0.00040231	0.0503554

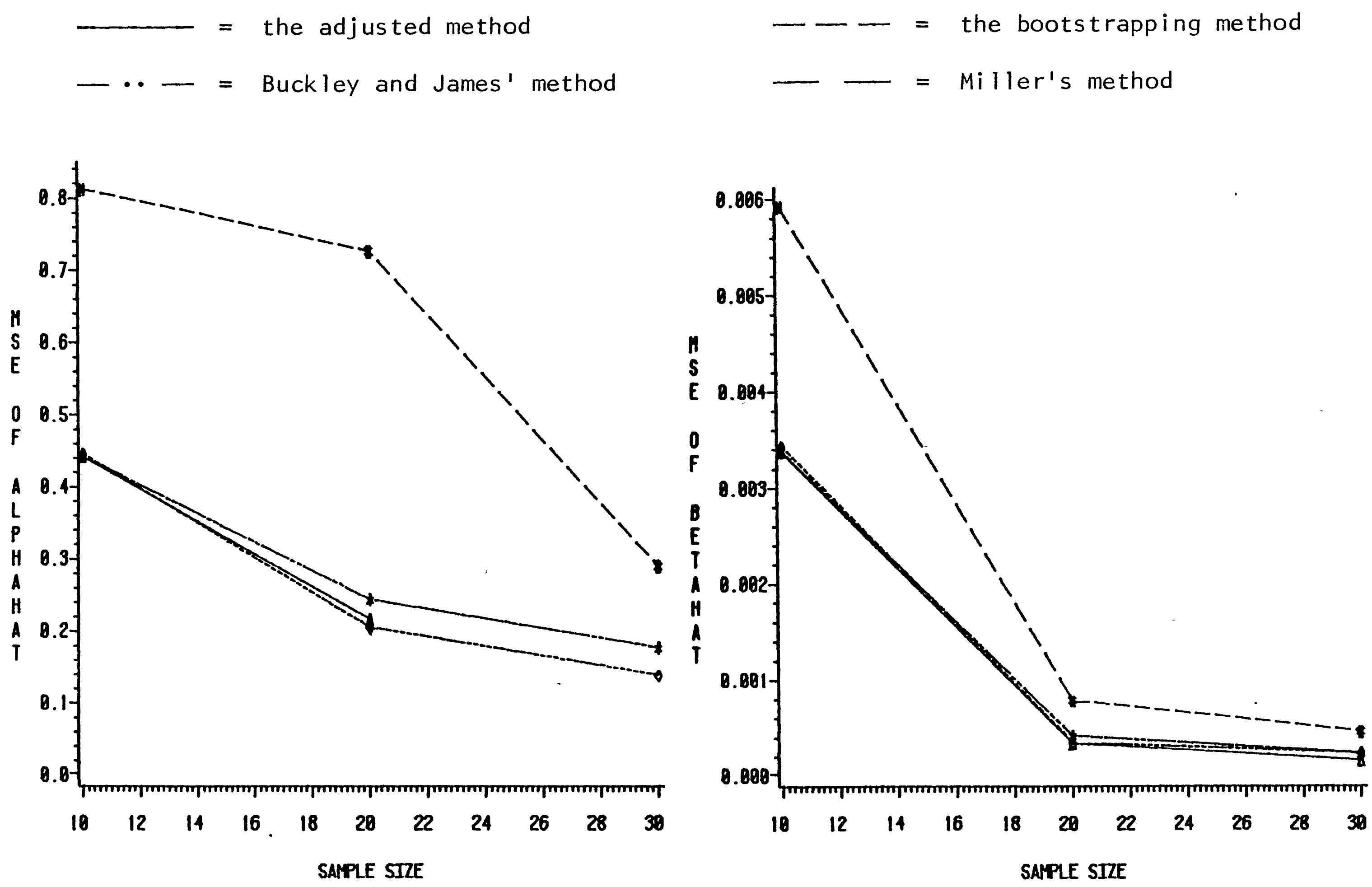


Figure 2. MSE of the Estimates of α and β Based on 100 Replications ($\alpha = 1$, $\beta = 0.2$, $C_i = 1.5X_i - 0.015X_i^2 + \alpha$, $X_i = 2i$ and $\varepsilon_i \sim N(0,1)$)

TABLE VII

SIMULATIONS CALCULATING THE ESTIMATES OF α AND β BASED ON 100 REPLICATIONS
($\alpha = 1$, $\beta = 0.2$, $C_i = 31$, $X_i = 2i$, $\varepsilon_i \sim N(0,1)$ AND SAMPLE SIZE = 10)

Method / Estimator	α				
	$\hat{\alpha}$ (1)	Var($\hat{\alpha}$) (2)	Bias (3)	MSE (4)	Z-Value (5)
Adjusted Method	0.999447	0.439986	-0.000553	0.4399863	-0.0083369
Bootstrapping	1.01129	0.443872	0.01129	0.443994	0.1694589
Buckley and James	0.999447	0.439986	-0.000553	0.4399863	-0.0083369
Miller	0.678066	0.707619	-0.321934	0.8112605	-3.8270758

Method / Estimator	β				
	$\hat{\beta}$ (6)	Var(β) (7)	Bias (8)	MSE (9)	Z-Value (10)
Adjusted Method	0.200219	0.00336882	0.000219	0.00336884	0.0377316
Bootstrapping	0.199681	0.00342705	-0.000319	0.00342715	-0.0054491
Buckley and James	0.200219	0.00336882	0.000219	0.00336884	0.0377316
Miller	0.190767	0.0058302	0.009233	0.00591544	-1.209209

TABLE VIII

SIMULATIONS CALCULATING THE ESTIMATES OF α AND β BASED ON 100 REPLICATIONS
($\alpha = 1$, $\beta = 0.2$, $C_i = 31$, $X_i = 2i$, $\varepsilon_i \sim N(0,1)$ AND SAMPLE SIZE = 20)

Method \ Estimator	α				
	$\hat{\alpha}$ (1)	Var($\hat{\alpha}$) (2)	Bias (3)	MSE (4)	Z-Value (5)
Adjusted Method	0.974762	0.2148561	-0.025238	0.215493	-0.544478
Bootstrapping	1.002549	0.203784	0.002549	0.2037905	0.0504657
Buckley and James	0.957667	0.240447	-0.042333	0.242239	-0.863315
Miller	0.400642	0.366883	-0.599358	0.726113	-9.8951477

Method \ Estimator	β				
	$\hat{\beta}$ (6)	Var($\hat{\beta}$) (7)	Bias (8)	MSE (9)	Z-Value (10)
Adjusted Method	0.203285	0.000311547	0.003285	0.00032229	1.861257
Bootstrapping	0.2011137	0.00032973	0.0011137	0.00033044	0.6133505
Buckley and James	0.201577	0.000400557	0.001577	0.00003086	0.7879523
Miller	0.199335	0.000778238	0.000665	0.00077264	-0.2393016

TABLE IX

SIMULATIONS CALCULATING THE ESTIMATES OF α AND β BASED ON 100 REPLICATIONS ($\alpha = 1$, $\beta = 0.2$, $C_i = 31$, $X_i = 2i$, $\varepsilon_i \sim N(0,1)$ AND SAMPLE SIZE = 30)

Method \ Estimator	α				
	$\hat{\alpha}$ (1)	Var($\hat{\alpha}$) (2)	Bias (3)	MAE (4)	Z-Value (5)
Adjusted Method	0.983644	0.123743	-0.016356	0.1240105	-0.4649612
Bootstrapping	1.010311	0.126647	0.010311	0.1372786	0.2897365
Buckley and James	0.988640	0.175303	-0.01136	0.1754320	-0.2713211
Miller	0.707826	0.203514	-0.292174	0.2888796	-6.4765604

Method \ Estimator	β				
	$\hat{\beta}$ (6)	Var($\hat{\beta}$) (7)	Bias (8)	MAE (9)	Z-Value (10)
Adjusted Method	0.201036	0.00014201	0.0001036	0.000143073	0.8693918
Bootstrapping	0.206076	0.0001832	0.006076	0.000220117	4.4890561
Buckley and James	0.201191	0.0002101	0.001191	0.000211518	0.8216724
Miller	0.200764	0.0004412	0.000764	0.000441783	0.363727

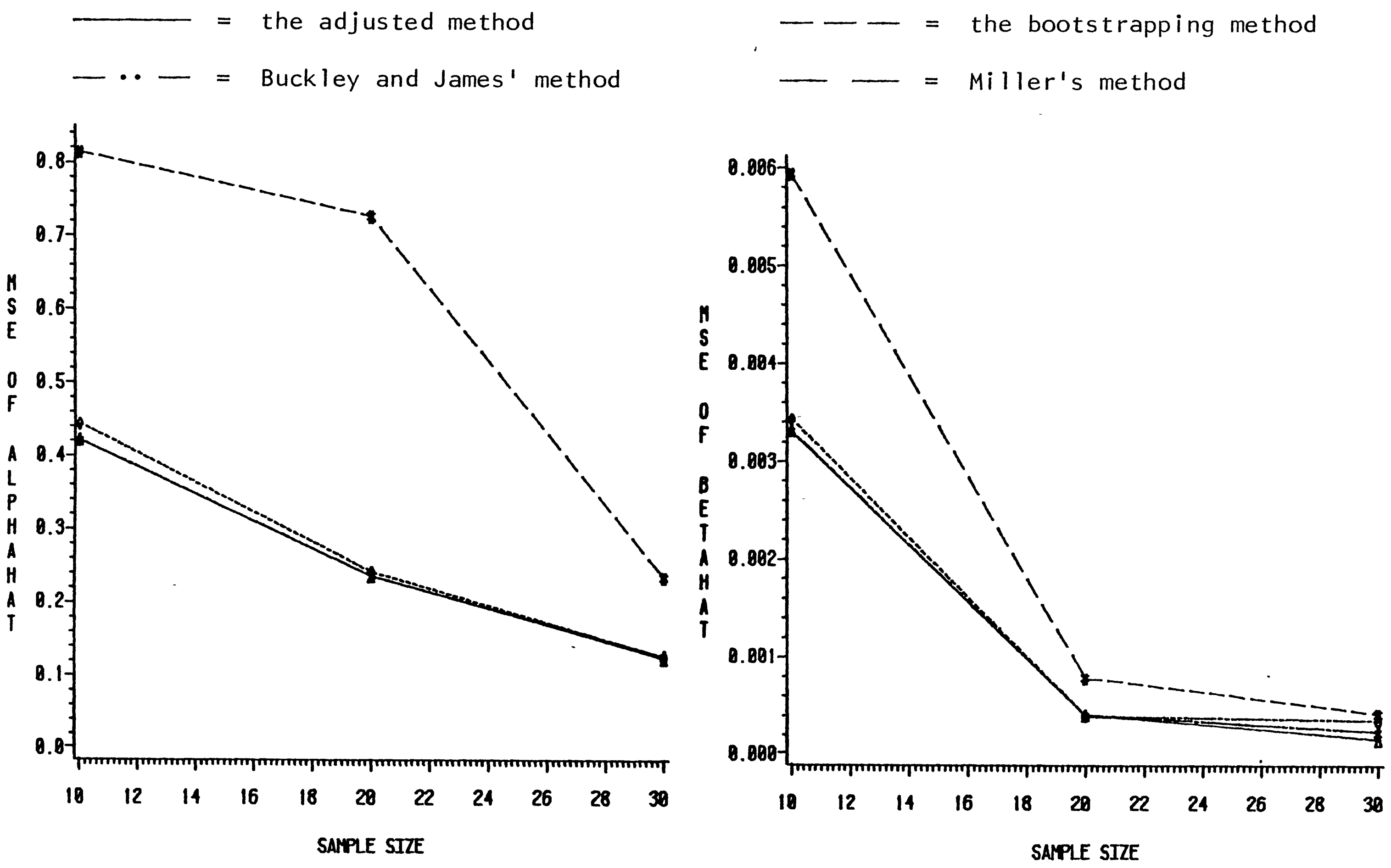


Figure 3. MSE of the Estimates of α and β Based on 100 Replications ($\alpha = 1$, $\beta = 0.2$, $C_i = 31$, $X_i = 2i$ and $\varepsilon_i \sim N(0,1)$)

TABLE X

SIMULATIONS CALCULATING THE ESTIMATES OF α AND β BASED ON 100 REPLICATIONS
($\alpha = 1$, $\beta = 0.2$, $C_i \sim U(\alpha + \beta X_i, \alpha + \beta X_i + 20)$, $X_i = 2i$, $\varepsilon_i \sim N(0,1)$
AND SAMPLE SIZE = 10)

Method / Estimator	α				
	$\hat{\alpha}$ (1)	Var($\hat{\alpha}$) (2)	Bias (3)	MSE (4)	Z-Value (5)
Adjusted Method	1.06369	0.422522	0.006369	0.426578	0.979821
Bootstrapping	1.00321	0.439734	0.00321	0.439744	0.0484071
Buckley and James	1.00586	0.442809	0.00586	0.4428433	0.0880621
Miller	0.628391	0.70174	-0.371609	0.8398332	-0.4360683

Method / Estimator	β				
	$\hat{\beta}$ (6)	Var($\hat{\beta}$) (7)	Bias (8)	MSE (9)	Z-Value (10)
Adjusted Method	0.200405	0.0033168	0.000405	0.00331696	0.0703223
Bootstrapping	0.199255	0.0033849	-0.000745	0.0033855	-0.1280511
Buckley and James	0.19985	0.00359402	-0.00015	0.0035941	-0.0250808
Miller	0.194254	0.00539906	-0.005746	0.0053201	-0.7819995

TABLE XI

SIMULATIONS CALCULATING THE ESTIMATES OF α AND β BASED ON 100 REPLICATIONS ($\alpha = 1$, $\beta = 0.2$, $C_i \sim U(\alpha + \beta X_i, \alpha + \beta X_i + 20)$, $X_i = 2i$, $\varepsilon_i \sim N(0,1)$ AND SAMPLE SIZE = 20)

Method \ Estimator	α				
	$\hat{\alpha}$ (1)	Var($\hat{\alpha}$) (2)	Bias (3)	MSE (4)	Z-Value (5)
Adjusted Method	1.08646	0.268572	0.08646	0.2760473	1.668339
Bootstrapping	0.9476311	0.2511732	-0.0523689	0.2539157	-1.094929
Buckley and James	0.966401	0.247792	-0.033599	0.2489208	-0.6749674
Miller	0.373579	0.401944	-0.626421	0.7943472	-9.8805977

Method \ Estimator	β				
	$\hat{\beta}$ (6)	Var($\hat{\beta}$) (7)	Bias (8)	MSE (9)	Z-Value (10)
Adjusted Method	0.201312	0.00037903	0.001312	0.00038072	0.6739022
Bootstrapping	0.200846	0.00032572	0.000846	0.00032645	0.4687718
Buckley and James	0.201395	0.000408781	0.001395	0.00041095	0.6899689
Miller	0.200227	0.000836164	0.000227	0.000836171	0.0785018

TABLE XII

SIMULATIONS CALCULATING THE ESTIMATES OF α AND β BASED ON 100 REPLICATIONS ($\alpha = 1$, $\beta = 0.2$, $C_i \sim U(\alpha + \beta X_i, \alpha + \beta X_i + 20)$, $X_i = 2i$, $\varepsilon_i \sim N(0,1)$ AND SAMPLE SIZE = 30)

Estimator / Method	α				
	$\hat{\alpha}$ (1)	Var($\hat{\alpha}$) (2)	Bias (3)	MSE (4)	Z-Value (5)
Adjusted Method	1.001563	0.102247	0.001563	0.1022494	0.0488802
Bootstrapping	0.963944	0.101785	-0.03656	0.1030850	-1.1301489
Buckley and James	0.964113	0.121174	-0.035887	0.1224618	-1.0309377
Miller	0.728201	0.184003	-0.271799	0.2578777	-6.3362952

Estimator / Method	β				
	$\hat{\beta}$ (6)	Var($\hat{\beta}$) (7)	Bias (8)	MSE (9)	Z-Value (10)
Adjusted Method	0.200473	0.0001254	0.000473	0.000125623	0.4223887
Bootstrapping	0.201132	0.0001247	0.001132	0.000125981	1.0137088
Buckley and James	0.201882	0.0001993	0.001882	0.000202841	1.3331099
Miller	0.200570	0.0004228	0.000570	0.000423124	0.277209

———— = the adjusted method

— — — = the bootstrapping method

— ·· — = Buckley and James' method

—— —— = Miller's method

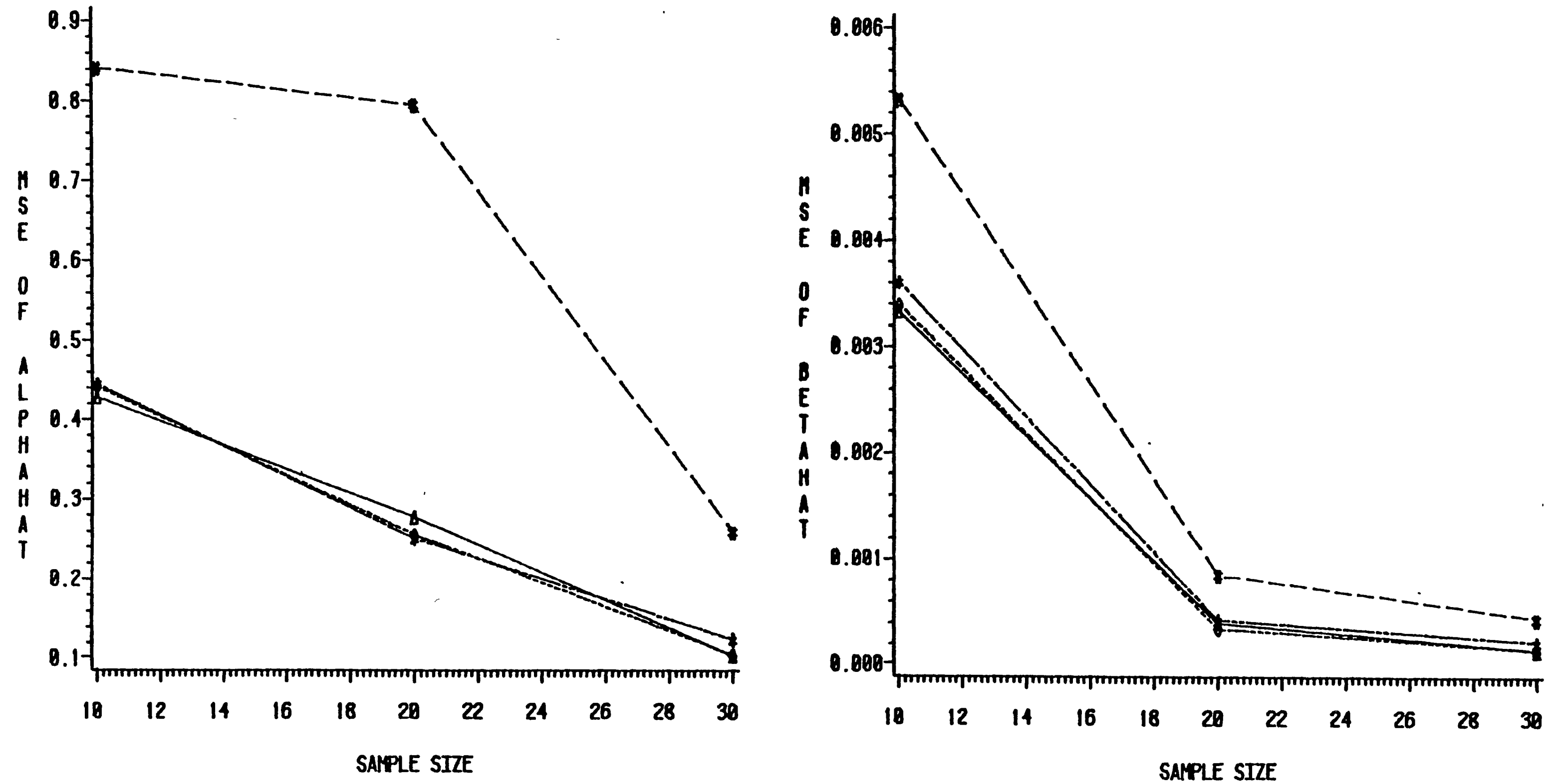


Figure 4. MSE of the Estimates of α and β Based on 100 Replications ($\alpha = 1$, $\beta = 0.2$, $C_i \sim U(\alpha + \beta X_i, \alpha + \beta X_i + 20)$, $X_i = 2i$ and $\varepsilon_i \sim N(0,1)$)

TABLE XIII

SIMULATIONS CALCULATING THE ESTIMATES OF α AND β BASED ON 100 REPLICATIONS (α = 1, β = 0.2, $C_i \sim U(0,50)$, $X_i = 2i$, $\varepsilon_i \sim N(0,1)$ AND SAMPLE SIZE = 10)

Method \ Estimator	α				
	$\hat{\alpha}$ (1)	Var($\hat{\alpha}$) (2)	Bias (3)	MSE (4)	Z-Value (5)
Adjusted Method	1.31853	0.746214	0.31853	0.8476753	3.6873855
Bootstrapping	1.03839	0.516911	0.03839	0.533962	0.533962
Buckley and James	0.987708	0.51912	-0.012292	0.519271	-0.1706037
Miller	0.47353	0.801317	-0.526457	1.078474	-5.85113

Method \ Estimator	β				
	$\hat{\beta}$ (6)	Var($\hat{\beta}$) (7)	Bias (8)	MSE (9)	Z-Value (10)
Adjusted Method	0.199891	0.00353426	-0.000109	0.00353428	-0.0183348
Bootstrapping	0.183269	0.00444387	-0.016731	0.00472372	-2.5098106
Buckley and James	0.207916	0.00400861	0.007916	0.00407126	1.2502843
Miller	0.197988	0.00584785	-0.002012	0.00585194	-0.2631055

TABLE XIV

SIMULATIONS CALCULATING THE ESTIMATES OF α AND β BASED ON 100 REPLICATIONS ($\alpha = 1$, $\beta = 0.2$, $C_i \sim U(0,50)$, $X_i = 2i$, $\varepsilon_i \sim N(0,1)$ AND SAMPLE SIZE = 20)

Estimator / Method	α				
	$\hat{\alpha}$ (1)	$Var(\hat{\alpha})$ (2)	Bias (3)	MSE (4)	Z-Value (5)
Adjusted Method	0.99598	0.199526	-0.00402	0.1995421	- 0.0899967
Bootstrapping	1.035246	0.1754321	0.035246	0.1766743	0.8415023
Buckley and James	0.99598	0.199526	-0.00402	0.1995421	- 0.0899967
Miller	0.336194	0.439232	-0.663806	0.6333942	-10.015994

Estimator / Method	β				
	$\hat{\beta}$ (6)	$Var(\hat{\beta})$ (7)	Bias (8)	MSE (9)	Z-Value (10)
Adjusted Method	-0.399174	0.0035598	0.000826	0.00035668	-0.4377944
Bootstrapping	-0.3954611	0.00033423	0.0045389	0.000354801	2.4828323
Buckley and James	-0.399174	0.00035398	0.000826	0.00035668	-0.7377944
Miller	-0.396483	0.000947709	0.003517	0.00096007	-1.1424432

TABLE XV

SIMULATIONS CALCULATING THE ESTIMATES OF α AND β BASED ON 100 REPLICATIONS ($\alpha = 1$, $\beta = 0.2$, $C_i \sim U(0,50)$, $X_i = 2i$, $\varepsilon_i \sim N(0,1)$ AND SAMPLE SIZE = 30)

Method \ Estimator	α				
	$\hat{\alpha}$ (1)	Var($\hat{\alpha}$) (2)	Bias (3)	MSE (4)	Z-Value (5)
Adjusted Method	1.13857	0.143662	0.13857	0.1628636	3.6559333
Bootstrapping	1.094714	0.173511	0.094714	0.1824817	2.2737917
Buckley and James	0.971187	0.179336	-0.028813	0.1801661	-0.680385
Miller	0.797424	0.223641	-0.202576	0.264678	-4.2836332

Method \ Estimator	β				
	$\hat{\beta}$ (6)	Var($\hat{\beta}$) (7)	Bias (8)	MSE (9)	Z-Value (10)
Adjusted Method	0.200372	0.0001217	0.000372	0.000121838	0.3372078
Bootstrapping	0.201528	0.0001959	0.001528	0.000198234	1.0917071
Buckley and James	0.210338	0.0002473	0.010338	0.000354174	6.5739208
Miller	0.211459	0.0006137	0.011459	0.000745008	4.6256062

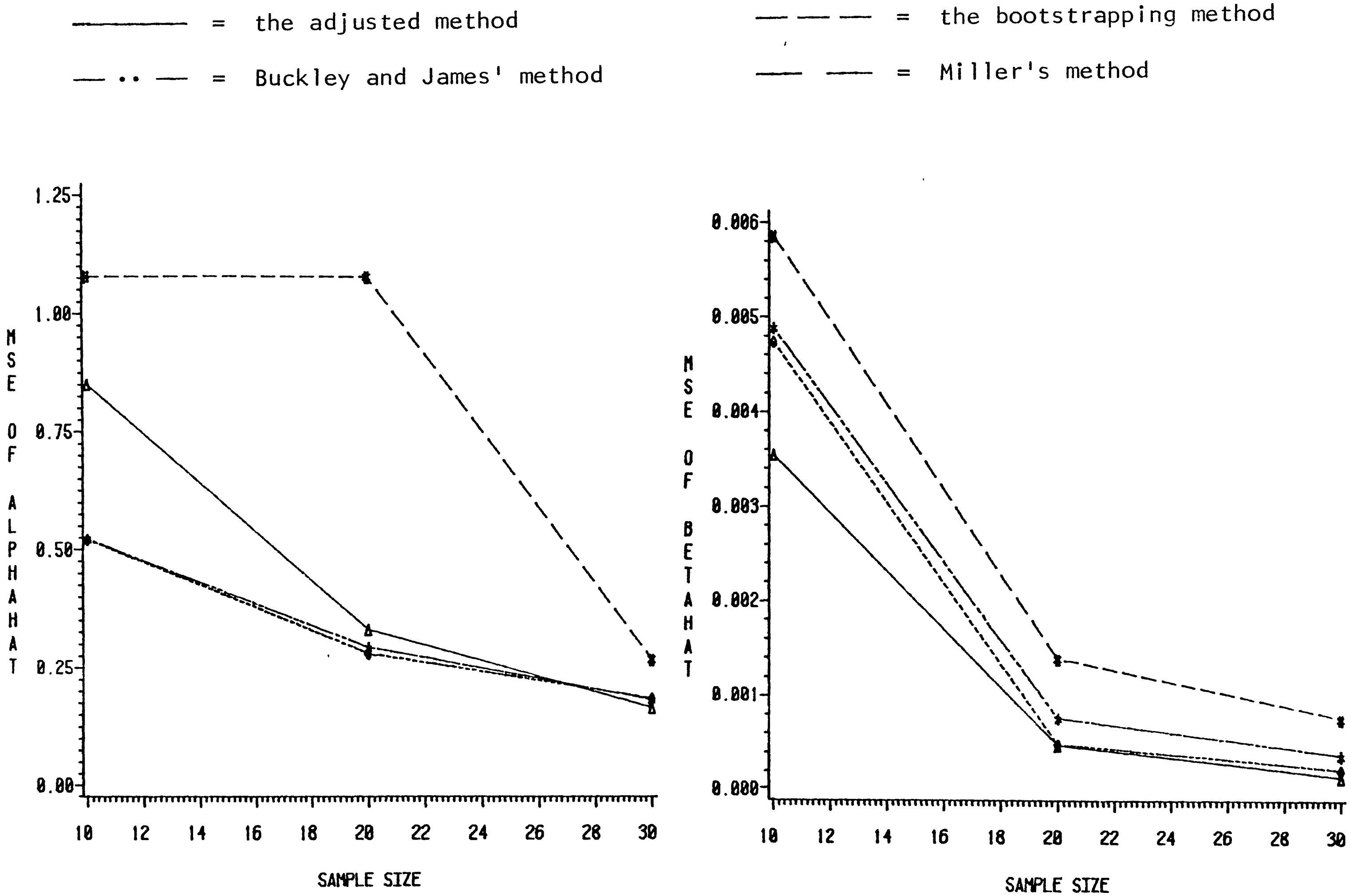


Figure 5. MSE of the Estimates of α and β Based on 100 Replications ($\alpha = 1$, $\beta = 0.2$, $C_i \sim U(0,50)$, $X_i = 2i$ and $\varepsilon_i \sim N(0,1)$)

TABLE XVI

SIMULATIONS CALCULATING THE ESTIMATES OF α AND β BASED ON 100 REPLICATIONS
($\alpha = 1$, $\beta = 0.4$, $C_i \sim N(\alpha + 40, 16)$, $X_i = 2i$, $\varepsilon_i \sim N(0,1)$
AND SAMPLE SIZE = 10)

Method \ Estimator	α				
	$\hat{\alpha}$ (1)	Var($\hat{\alpha}$) (2)	Bias (3)	MSE (4)	Z-Value (5)
Adjusted Method	0.912284	0.487624	-0.087716	0.4947181	-1.2569085
Bootstrapping	0.921749	0.487849	-0.078251	0.4939722	-1.120333
Buckley and James	0.912284	0.487024	-0.087716	0.4947181	-1.2569085
Miller	0.466033	0.686484	-0.533967	0.9716047	-6.4446504

Method \ Estimator	β				
	$\hat{\beta}$ (6)	Var($\hat{\beta}$) (7)	Bias (8)	MSE (9)	Z-Value (10)
Adjusted Method	-0.387856	0.00291047	0.012144	0.0030579	2.2510232
Bootstrapping	-0.388176	0.00292271	0.011824	0.0030625	2.1871144
Buckley and James	-0.387856	0.00291047	0.012144	0.0030579	2.2610232
Miller	-0.386718	0.00435181	0.013282	0.00452821	2.0133913

TABLE XVII

SIMULATIONS CALCULATING THE ESTIMATES OF α AND β BASED ON 100 REPLICATIONS ($\alpha = 1$, $\beta = 0.4$, $C_i \sim N(\alpha + 40, 16)$, $X_i = 2i$, $\varepsilon_i \sim N(0,1)$ AND SAMPLE SIZE = 20)

Method / Estimator	α				
	$\hat{\alpha}$ (1)	Var($\hat{\alpha}$) (2)	Bias (3)	MSE (4)	Z-Value (5)
Adjusted Method	1.25493	0.263415	0.25493	0.3284043	2.9177005
Bootstrapping	0.891136	0.264445	-0.108864	0.2762963	-2.1169793
Buckley and James	0.89818	0.279295	-0.10182	0.2896623	-1.926643
Miller	0.288365	0.569783	-0.711635	1.0762074	-9.427627

Method / Estimator	β				
	$\hat{\beta}$ (6)	Var($\hat{\beta}$) (7)	Bias (8)	MSE (9)	Z-Value (10)
Adjusted Method	0.201296	0.000457581	0.001296	0.0004502	0.6058594
Bootstrapping	0.2011317	0.00046842	0.0011317	0.00046968	0.522905
Buckley and James	0.211972	0.0006016	0.011972	0.000745	4.8810718
Miller	0.194833	0.00134695	-0.005167	0.00137359	-1.4078712

TABLE XVIII

SIMULATIONS CALCULATING THE ESTIMATES OF α AND β BASED ON 100 REPLICATIONS ($\alpha = 1$, $\beta = 0.4$, $C_i \sim N(\alpha + 40, 16)$, $X_i = 2i$, $\varepsilon_i \sim N(0,1)$ AND SAMPLE SIZE = 30)

Method \ Estimator	α				
	$\hat{\alpha}$ (1)	Var($\hat{\alpha}$) (2)	Bias (3)	MSE (4)	Z-Value (5)
Adjusted Method	0.994892	0.048621	-0.005108	0.048647	-0.2316535
Bootstrapping	1.014475	0.048884	0.014475	0.0490935	0.6546892
Buckley and James	0.985532	0.062375	-0.014468	0.0625843	-0.5792995
Miller	0.689999	0.1165841	-0.310001	0.2126847	-9.0791132

Method \ Estimator	β				
	$\hat{\beta}$ (6)	Var($\hat{\beta}$) (7)	Bias (8)	MSE (9)	Z-Value (10)
Adjusted Method	-0.396433	0.0001253	0.003567	0.000138023	3.1866002
Bootstrapping	-0.392615	0.0001647	0.007385	0.000219238	5.7544484
Buckley and James	-0.394662	0.0002841	0.005338	0.000312594	3.1669627
Miller	-0.394472	0.0004833	0.005528	0.000513858	2.5145463

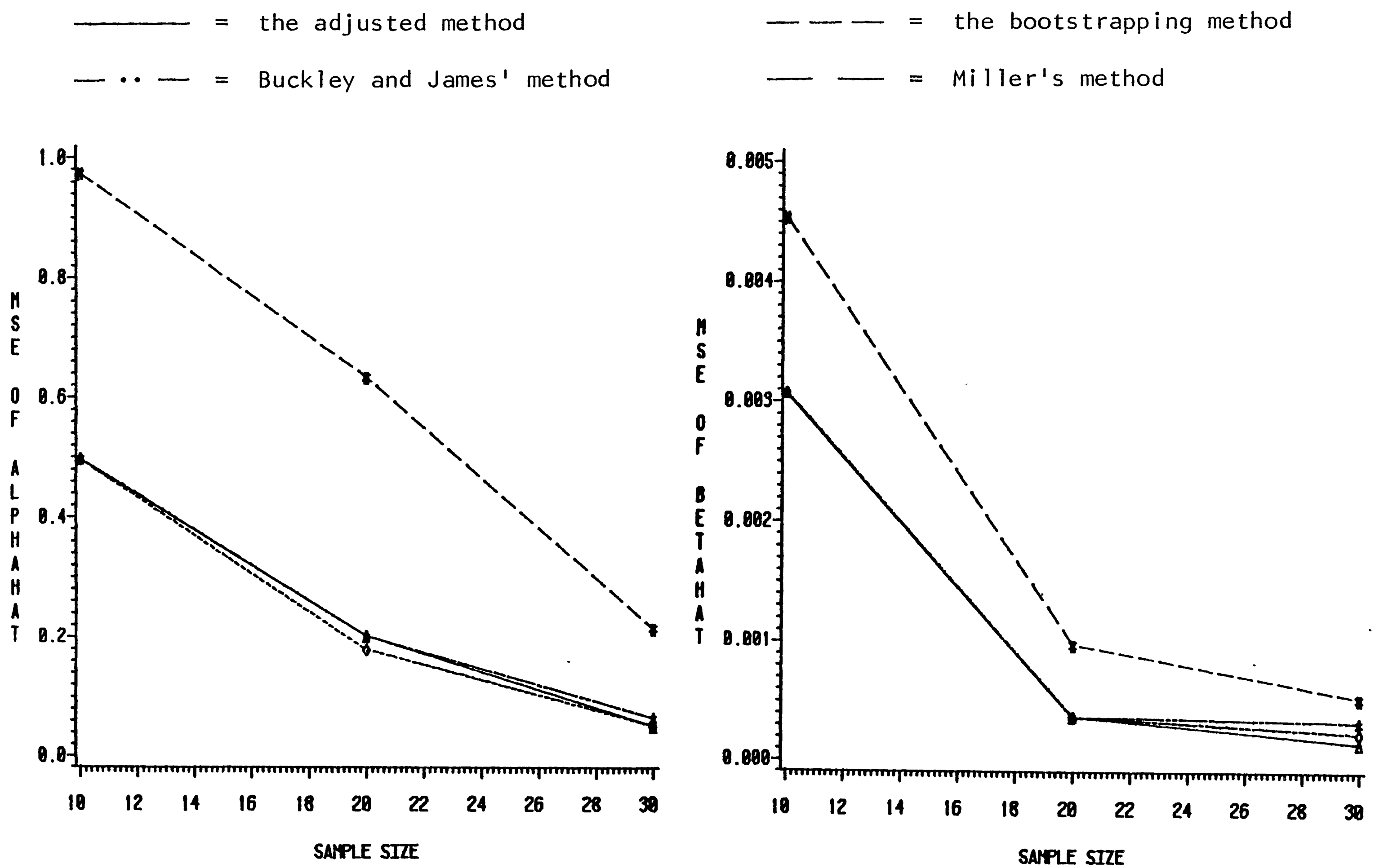


Figure 6. MSE of the Estimates of α and β Based on 100 Replications ($\alpha = 1$, $\beta = 0.2$, $C_i \sim N(\alpha + 40, 16)$, $X_i = 2i$ and $\varepsilon_i \sim N(0,1)$)

TABLE XIX

SIMULATIONS CALCULATING THE ESTIMATES OF α AND β BASED ON 100 REPLICATIONS (α = 10, β = 0.2, $C_i \sim U(\alpha + \beta X_i, \alpha + \beta X_i + 40)$, $X_i = 2i$, $\varepsilon_i \sim N(0,100)$ AND SAMPLE SIZE = 25)

Method \ Estimator	α				
	$\hat{\alpha}$ (1)	Var($\hat{\alpha}$) (2)	Bias (3)	MSE (4)	Z-Value (5)
Adjusted Method	7.5142	18.0890	-2.4858	24.268202	-5.8446553
Bootstrapping	7.2517	20.7735	-2.7483	28.326653	-6.0298888
Buckley and James	6.9941	21.2146	-3.0059	30.250035	-6.5261508
Miller	6.4454	24.1363	-3.5546	36.771481	-7.2352808

Method \ Estimator	β				
	$\hat{\beta}$ (6)	Var($\hat{\beta}$) (7)	Bias (8)	MSE (9)	Z-Value (10)
Adjusted Method	0.1950	0.0221	-0.005	0.022125	-0.3363364
Bootstrapping	0.2141	0.0383	0.0141	0.0384988	0.7204763
Buckley and James	0.2104	0.0388	0.0104	0.0389081	0.52798
Miller	0.1753	0.0642	-0.0247	0.06481	-0.9748312

TABLE XX

SIMULATIONS CALCULATING THE ESTIMATES OF α AND β BASED ON 100 REPLICATIONS ($\alpha = 10$, $\beta = 0.2$, $C_i \sim U(\alpha + \beta X_i, \alpha + \beta X_i + 40)$, $X_i = 2i$, $\varepsilon_i \sim N(0,100)$ AND SAMPLE SIZE = 50)

Method \ Estimator	α				
	$\hat{\alpha}$ (1)	Var($\hat{\alpha}$) (2)	Bias (3)	MSE (4)	Z-Value (5)
Adjusted Method	7.3476	8.2451	-2.6524	15.280326	-9.2372151
Bootstrapping	7.2714	8.7149	-2.7286	16.160158	-9.2429089
Buckley and James	7.9348	10.5727	-2.0652	14.837751	-6.3513456
Miller	6.5531	12.1117	-3.4469	23.99282	-9.9043535

Method \ Estimator	β				
	$\hat{\beta}$ (6)	Var($\hat{\beta}$) (7)	Bias (8)	MSE (9)	Z-Value (10)
Adjusted Method	0.1981	0.0024	-0.0019	0.00240361	-0.3878358
Bootstrapping	0.1702	0.0041	-0.0298	0.00498804	-4.6539781
Buckley and James	0.1893	0.0046	-0.0107	0.00471449	-1.5776289
Miller	0.1692	0.0153	-0.0308	0.0162436	-2.4900324

TABLE XXI

SIMULATIONS CALCULATING THE ESTIMATES OF α AND β BASED ON 100 REPLICATIONS ($\alpha = 10$, $\beta = 0.2$, $C_i \sim U(\alpha + \beta X_i, \alpha + \beta X_i + 40)$, $X_i = 2i$, $\varepsilon_i \sim N(0,100)$ AND SAMPLE SIZE = 75)

Method / Estimator	α				
	$\hat{\alpha}$ (1)	Var($\hat{\alpha}$) (2)	Bias (3)	MSE (4)	Z-Value (5)
Adjusted Method	7.5686	5.8707	-2.4364	11.806745	-10.055499
Bootstrapping	7.6222	6.4546	-2.3778	12.108533	- 9.3592415
Buckley and James	7.4529	6.9434	-2.5471	13.431118	- 9.6662918
Miller	6.9347	7.2749	-3.0653	16.670964	-11.364739

Method / Estimator	β				
	$\hat{\beta}$ (6)	Var($\hat{\beta}$) (7)	Bias (8)	MSE (9)	Z-Value (10)
Adjusted Method	0.1987	0.0006	-0.0013	0.00060169	-0.5307227
Bootstrapping	0.1894	0.0009	-0.0106	0.00101236	-3.5333333
Buckley and James	0.1921	0.0012	-0.0079	0.00126241	-2.2805331
Miller	0.1821	0.0124	-0.0179	0.0127204	-1.6074675

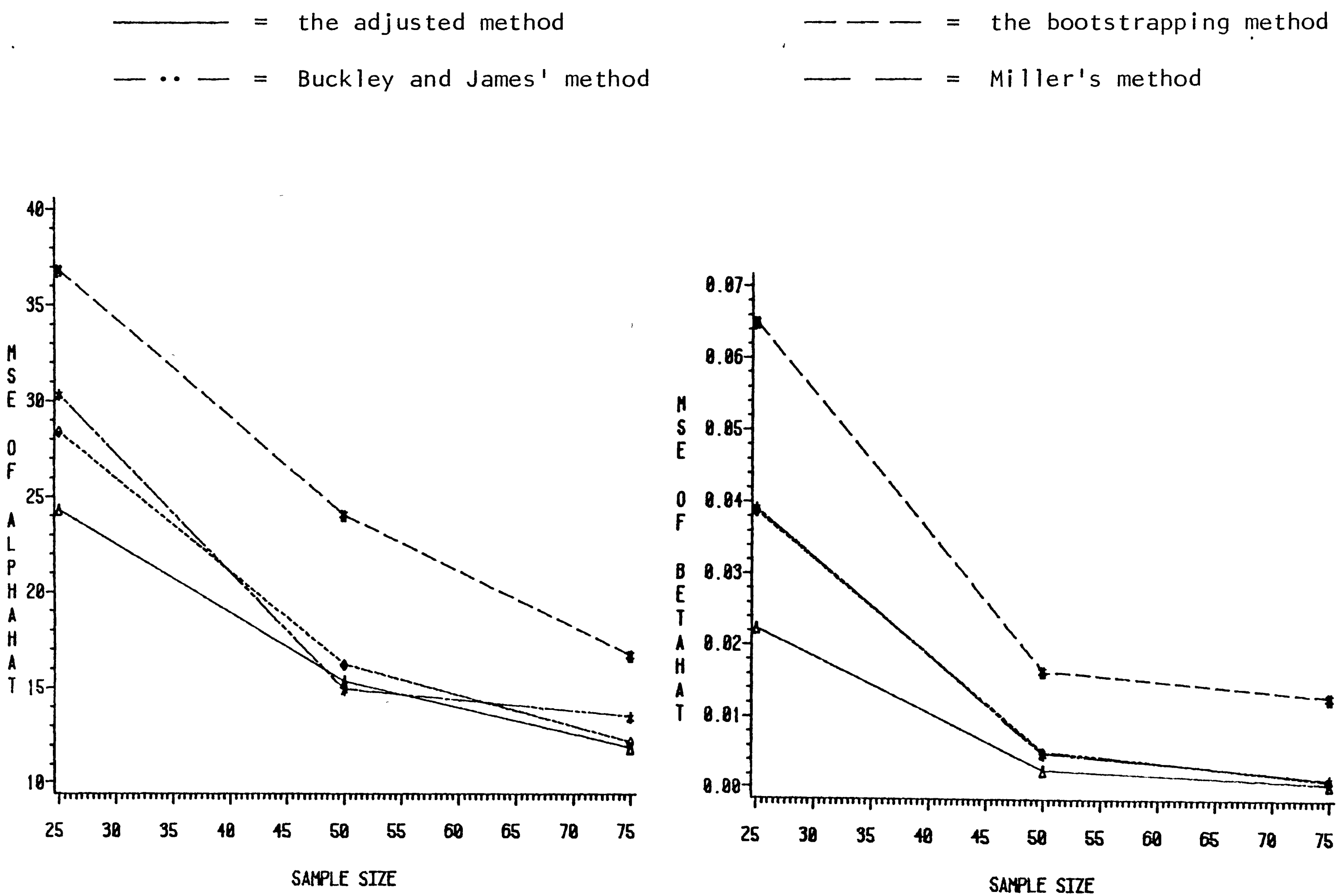


Figure 7. MSE of the Estimates of α and β Based on 100 Replications ($\alpha = 1$, $\beta = 0.2$, $C_i \sim U(\alpha + \beta X_i, \alpha + \beta X_i + 40)$, $X_i = 2i$ and $\varepsilon_i \sim N(0,100)$)

TABLE XXII

SIMULATIONS CALCULATING THE ESTIMATES OF α AND β BASED ON 100 REPLICATIONS
(α = 10, β = 0.2, C_i = 30, X_i = 2i, $\varepsilon_i \sim$ N(0,100) AND SAMPLE SIZE = 25)

Method \ Estimator	α				
	$\hat{\alpha}$ (1)	Var($\hat{\alpha}$) (2)	Bias (3)	MSE (4)	Z-Value (5)
Adjusted Method	9.9335	13.4649	-0.0605	13.469322	-0.1812258
Bootstrapping	9.8843	14.3227	-0.1157	14.336086	-0.3057179
Buckley and James	9.3762	14.9916	-0.6238	15.380726	-1.6110958
Miller	8.6330	17.1118	-1.3670	18.980489	-3.3046135

Method \ Estimator	β				
	$\hat{\beta}$	Var($\hat{\beta}$)	Bias	MSE	Z-Value
Adjusted Method	0.1734	0.0140	-0.0266	0.0147075	-2.2481103
Bootstrapping	0.1714	0.0167	-0.0286	0.0175179	-2.2131344
Buckley and James	0.1632	0.0211	-0.0368	0.0224542	-2.5334165
Miller	0.1602	0.0453	-0.0398	0.046884	-1.8699671

TABLE XXIII

SIMULATIONS CALCULATING THE ESTIMATES OF α AND β BASED ON 100 REPLICATIONS (α = 10, β = 0.2, C_i = 30, X_i = 2i, $\varepsilon_i \sim$ N(0,100) AND SAMPLE SIZE = 50)

Method / Estimator	α				
	$\hat{\alpha}$ (1)	Var($\hat{\alpha}$) (2)	Bias (3)	MSE (4)	Z-Value (5)
Adjusted Method	10.4330	7.3768	0.4330	7.564289	-1.5942407
Bootstrapping	10.2527	8.1014	0.2527	8.1652573	0.8878205
Buckley and James	9.4269	7.9918	-0.5704	8.3171562	-2.0177029
Miller	9.3688	8.7391	-0.6312	9.1375134	-2.1351758

Method / Estimator	β				
	$\hat{\beta}$ (6)	Var($\hat{\beta}$) (7)	Bias (8)	MSE (9)	Z-Value (10)
Adjusted Method	0.1245	0.0023	-0.0755	0.0080025	-15.742838
Bootstrapping	0.1093	0.0019	-0.0907	0.0101264	-20.808007
Buckley and James	0.1055	0.0035	-0.0945	0.0124302	-15.973415
Miller	0.1022	0.0186	-0.0978	0.0281648	-7.1710439

TABLE XXIV

SIMULATIONS CALCULATING THE ESTIMATES OF α AND β BASED ON 100 REPLICATIONS
(α = 10, β = 0.2, C_i = 30, X_i = 2i, ε_i ~ N(0,100) AND SAMPLE SIZE = 75)

Method \ Estimator	α				
	$\hat{\alpha}$ (1)	Var($\hat{\alpha}$) (2)	Bias (3)	MSE (4)	Z-Value (5)
Adjusted Method	11.0108	4.9272	1.0108	5.9489166	4.5537076
Bootstrapping	10.9399	4.1216	0.9399	5.005012	4.6296561
Buckley and James	9.3226	5.5447	-0.6774	6.003508	-2.8767768
Miller	8.9339	7.5129	-1.0661	8.6494692	-3.8895033

Method \ Estimator	β				
	$\hat{\beta}$ (6)	Var($\hat{\beta}$) (7)	Bias (8)	MSE (9)	Z-Value (10)
Adjusted Method	0.1086	0.0009	-0.0914	0.00925396	-30.466667
Bootstrapping	0.1133	0.0009	-0.0867	0.00841689	-28.90000
Buckley and James	0.1053	0.0012	-0.0947	0.010168	-27.337535
Miller	0.1266	0.0107	-0.0734	0.0160875	- 7.0958458

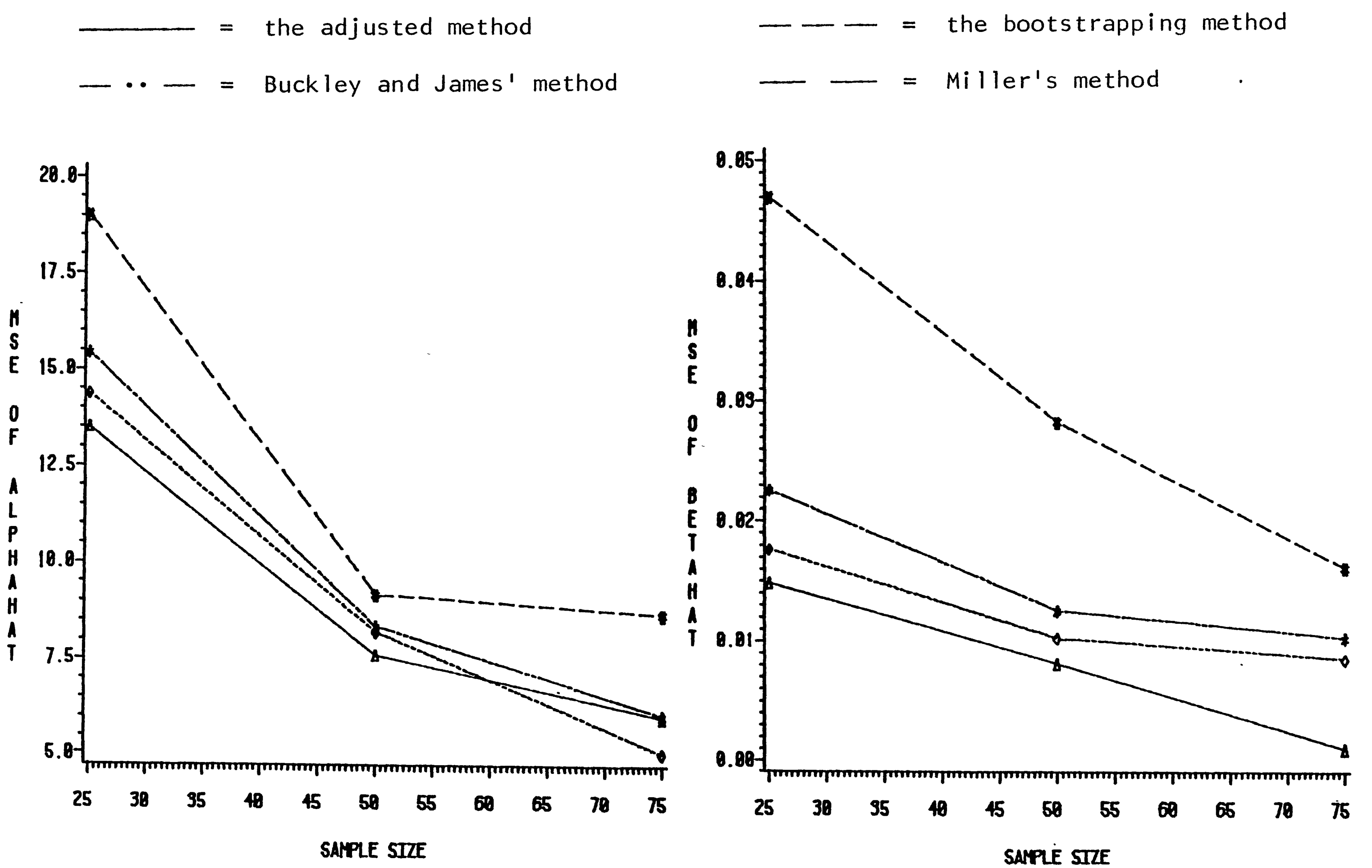


Figure 8. MSE of the Estimates of α and β Based on 100 Replications ($\alpha = 10$, $\beta = 0.2$, $C_i = 30$, $X_i = 2i$ and $\varepsilon_i \sim N(0,100)$)

TABLE XXV

SIMULATIONS CALCULATING THE ESTIMATES OF α AND β BASED ON 100 REPLICATIONS
($\alpha = 10$, $\beta = 0.2$, $C_i = 1.5X_i - 0.015X_i^2 + \alpha$, $X_i = 2i$, $\varepsilon_i \sim N(0,100)$
AND SAMPLE SIZE = 25)

Method \ Estimator	α				
	$\hat{\alpha}$ (1)	Var($\hat{\alpha}$) (2)	Bias (3)	MSE (4)	Z-Value (5)
Adjusted Method	11.7817	20.7231	1.7817	23.897555	3.9138774
Bootstrapping	12.3117	22.1573	2.3117	27.501257	4.9110353
Buckley and James	8.9462	23.7333	-1.0538	24.843794	-2.1631126
Miller	7.8540	27.4868	-2.146	32.092116	-4.0932439

Method \ Estimator	β				
	$\hat{\beta}$ (6)	Var($\hat{\beta}$) (7)	Bias (8)	MSE (9)	Z-Value (10)
Adjusted Method	0.2588	0.0195	0.0588	0.0229574	4.2107555
Bootstrapping	0.2479	0.2479	0.0479	0.0199944	3.6003845
Buckley and James	0.2759	0.0361	0.0759	0.0418608	3.9947368
Miller	0.1256	0.0583	-0.0744	0.0638353	-3.0813322

TABLE XXVI

SIMULATIONS CALCULATING THE ESTIMATES OF α AND β BASED ON 100 REPLICATIONS ($\alpha = 10$, $\beta = 0.2$, $C_i = 1.5X_i - 0.015X_i^2 + \alpha$, $X_i = 2i$, $\varepsilon_i \sim N(0,100)$ AND SAMPLE SIZE = 50)

Method \ Estimator	α: $\hat{\alpha}$ (1)	Var($\hat{\alpha}$) (2)	Bias (3)	MSE (4)	Z-Vàlue (5)
Adjusted Method	7.5489	12.3848	-2.4511	18.392691	-6.949264
Bootstrapping	7.2441	12.1121	-2.7559	19.707085	-7.9186971
Buckley and James	6.9537	14.7563	-3.0463	24.036244	-7.9301962
Miller	6.7491	17.1417	-3.2509	27.710051	-7.8519345

Method \ Estimator	β: $\hat{\beta}$ (6)	Var($\hat{\beta}$) (7)	Bias (8)	MSE (9)	Z-Value (10)
Adjusted Method	0.1861	0.0054	-0.0139	0.00559321	-1.8915504
Bootstrapping	0.1797	0.0067	-0.0203	0.00711209	-2.4800397
Buckley and James	0.1791	0.0114	-0.0209	0.0118368	-1.9574643
Miller	0.1863	0.0172	-0.0137	0.0173876	-1.0446152

TABLE XXVII

SIMULATIONS CALCULATING THE ESTIMATES OF α AND β BASED ON 100 REPLICATIONS
($\alpha = 10$, $\beta = 0.2$, $C_i = 1.5X_i - 0.015X_i^2 + \alpha$, $X_i = 2i$, $\varepsilon_i \sim N(0,100)$
AND SAMPLE SIZE = 75)

Method \ Estimator	$\hat{\alpha}$ (1)	Var($\hat{\alpha}$) (2)	Bias (3)	MSE (4)	Z-Value (5)
	α				
Adjusted Method	7.5689	12.3360	-2.4311	18.246247	-6.9217458
Bootstrapping	7.8144	13.8621	-2.1856	18.638947	-5.58702442
Buckley and James	7.0048	13.9441	-2.9952	22.915323	-8.0210382
Miller	7.5497	16.7726	-2.4503	22.77657	-5.983009

Method \ Estimator	$\hat{\beta}$ (6)	Var($\hat{\beta}$) (7)	Bias (8)	MSE (9)	Z-Value (10)
	β				
Adjusted Method	0.1858	0.0054	-0.0142	0.00560164	-1.9323752
Bootstrapping	0.1872	0.0063	-0.0128	0.00646384	-1.6126484
Buckley and James	0.2212	0.0068	0.0212	0.00724944	2.5708776
Miller	0.1813	0.0113	-0.0187	0.0116496	-1.759148

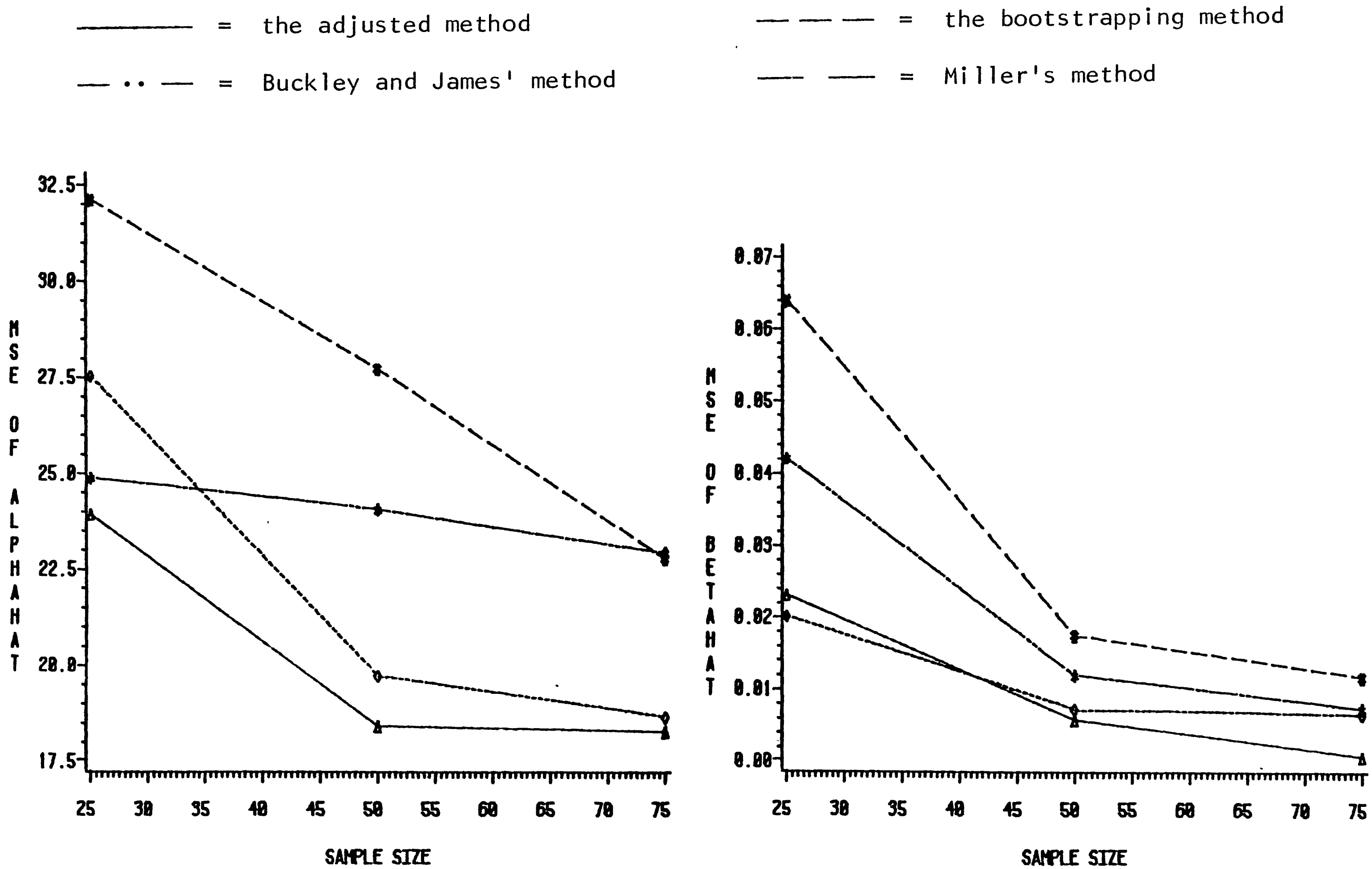


Figure 9. MSE of the Estimates of α and β Based on 100 Replications ($\alpha = 10$, $\beta = 0.2$, $C_i = 1.5X_i - 0.015X_i^2 + \alpha$, $X_i = 2i$ and $\varepsilon_i \sim N(0,100)$)

TABLE XXVIII

SIMULATIONS CALCULATING THE ESTIMATES OF α AND β BASED ON 100 REPLICATIONS (α = 10, β = 0.2, $X_i \sim U(0,100)$, $\varepsilon_i \sim N(0,100)$ AND SAMPLE SIZE = 25)

Method \ Estimator	α				
	$\hat{\alpha}$ (1)	Var($\hat{\alpha}$) (2)	Bias (3)	MSE (4)	Z-Value (5)
Adjusted Method	7.8771	34.1472	-2.1229	38.653904	-3.6328878
Bootstrapping	7.2461	33.6246	-2.7539	41.208565	-4.7491906
Buckley and James	7.2488	36.5339	-2.7512	44.103001	-4.5517054
Miller	7.2511	39.3568	-2.7489	46.913251	-4.3817647

Method \ Estimator	β				
	$\hat{\beta}$ (6)	Var($\hat{\beta}$) (7)	Bias (8)	MSE (9)	Z-Value (10)
Adjusted Method	0.1786	0.0119	-0.0214	0.0123579	-1.9617348
Bootstrapping	0.1780	0.0191	-0.022	0.019584	-1.5918641
Buckley and James	0.1696	0.0203	-0.0304	0.0212241	-2.1336617
Miller	0.1751	0.0553	-0.0249	0.05592	-1.0588557

TABLE XXIX

SIMULATIONS CALCULATING THE ESTIMATES OF α AND β BASED ON 100 REPLICATIONS
(α = 10, β = 0.2, $X_i \sim U(0,100)$, $\varepsilon_i \sim N(0,100)$ AND SAMPLE SIZE = 50)

Method \ Estimator	α				
	$\hat{\alpha}$ (1)	Var($\hat{\alpha}$) (2)	Bias (3)	MSE (4)	Z-Value (5)
Adjusted Method	8.0041	15.6963	-1.9959	19.179917	-5.0377908
Bootstrapping	7.6352	16.6511	-2.3648	22.243379	-5.7952604
Buckley and James	8.0558	18.9444	-1.9442	22.724314	-4.4668412
Miller	7.9142	22.6639	-2.0858	27.014462	-4.3813237
Method \ Estimator	β				
	$\hat{\beta}$ (6)	Var($\hat{\beta}$) (7)	Bias (8)	MSE (9)	Z-Value (10)
Adjusted Method	0.1744	0.0052	-0.0256	0.00585536	-3.5500813
Bootstrapping	0.1751	0.0058	-0.0249	0.00642001	-3.2695302
Buckley and James	0.1724	0.0063	-0.0276	0.00706176	-3.4772731
Miller	0.1684	0.0101	-0.0316	0.0110985	-3.1443175

TABLE XXX

SIMULATIONS CALCULATING THE ESTIMATES OF α AND β BASED ON 100 REPLICATIONS
(α = 10, β = 0.2, $X_i \sim U(0,100)$, $\varepsilon_i \sim N(0,100)$ AND SAMPLE SIZE = 75)

Method \ Estimator	α				
	$\hat{\alpha}$ (1)	Var($\hat{\alpha}$) (2)	Bias (3)	MSE (4)	Z-Value (5)
Adjusted Method	8.4788	7.7041	-1.5212	10.018149	-5.4805654
Bootstrapping	7.9116	8.5226	-2.0884	12.844015	-7.1536491
Buckley and James	7.3629	8.9430	-2.6371	15.897296	-8.8783023
Miller	7.2549	12.5214	-2.7451	20.056974	-7.756776

Method \ Estimator	β				
	$\hat{\beta}$ (6)	Var($\hat{\beta}$) (7)	Bias (8)	MSE (9)	Z-Value (10)
Adjusted Method	0.1844	0.0026	-0.0156	0.00284336	-3.0594117
Bootstrapping	0.1793	0.0039	-0.0207	0.00432849	-3.3146528
Buckley and James	0.1826	0.0052	-0.0174	0.0055027	-2.4129459
Miller	0.1800	0.0102	-0.0200	0.0106	-1.9802951

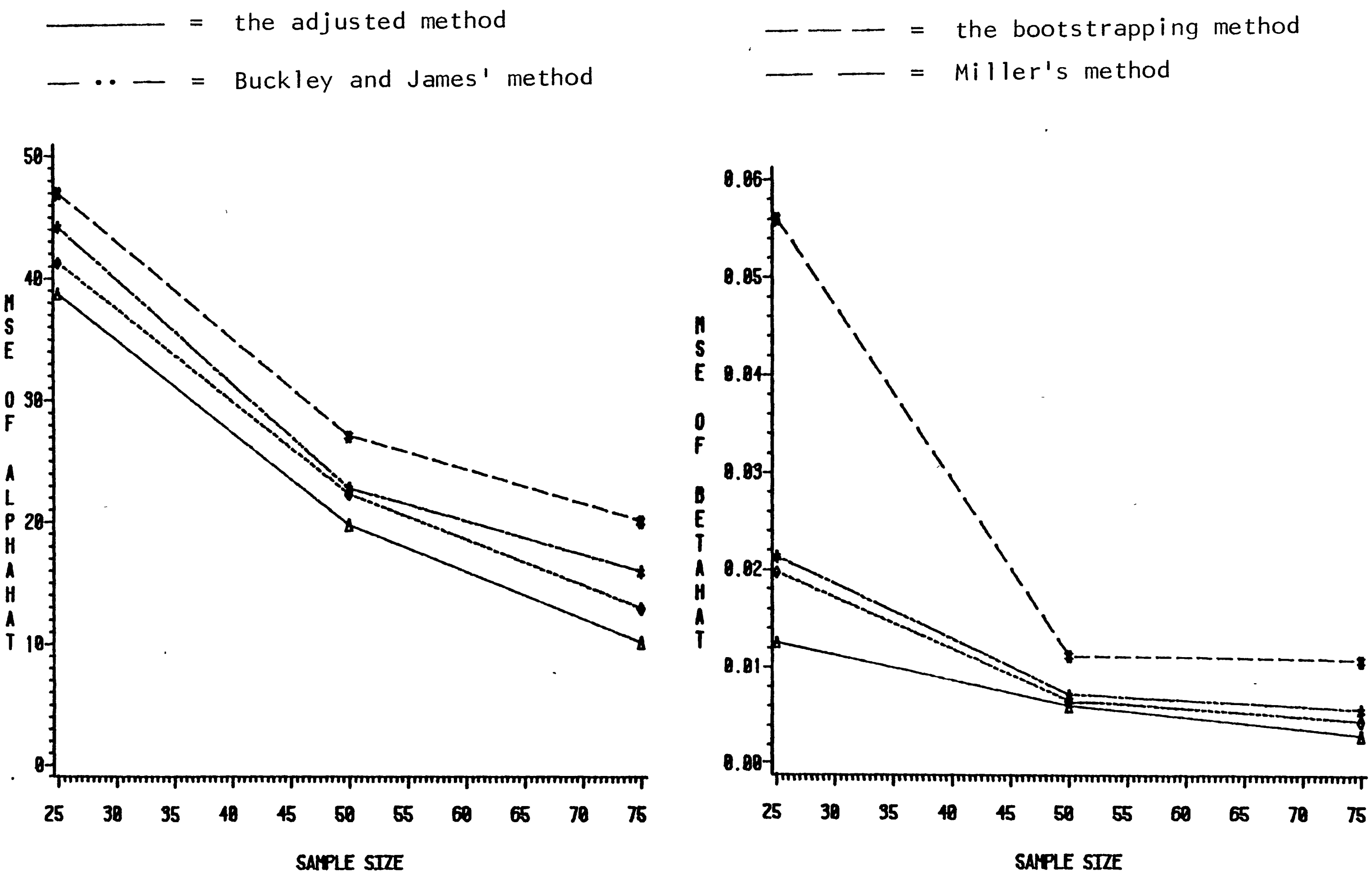


Figure 10. MSE of the Estimates of α and β Based on 100 Replications ($\alpha = 10$, $\beta = 0.2$), $C_i \sim U(0,50)$, $X_i \sim U(0,100)$ and $\varepsilon_i \sim N(0,100)$

TABLE XXXI

SIMULATIONS CALCULATING THE ESTIMATES OF α AND β BASED ON 100 REPLICATIONS (α = 10, β = 0.2, $X_i \sim U(0,100)$, $\varepsilon_i \sim N(0,100)$, SAMPLE SIZE = 25 AND 25% CENSORING)

Method / Estimator	α				
	$\hat{\alpha}$ (1)	Var($\hat{\alpha}$) (2)	Bias (3)	MSE (4)	Z-Value (5)
Adjusted Method	9.7142	31.6553	-0.2858	31.736982	-0.5079711
Bootstrapping	9.7033	31.7128	-0.2967	31.800831	-0.5268661
Buckley and James	9.6924	31.8846	-0.3076	31.979218	-0.5447482
Miller	9.7348	32.0627	-0.2652	32.133031	-0.4683531
	β				
Method / Estimator	$\hat{\beta}$ (6)	Var($\hat{\beta}$) (7)	Bias (8)	MSE (9)	Z-Value (10)
Adjusted Method	0.2084	0.0101	0.0084	0.0101705	0.8358312
Bootstrapping	0.2046	0.0153	0.0046	0.0153211	0.3718879
Buckley and James	0.2102	0.0207	0.0102	0.020804	0.708949
Miller	0.2110	0.0271	0.0110	0.027221	0.6682024

TABLE XXXII

SIMULATIONS CALCULATING THE ESTIMATES OF α AND β BASED ON 100 REPLICATIONS (α = 10, β = 0.2, $X_i \sim U(0,100)$, $\varepsilon_i \sim N(0,100)$, SAMPLE SIZE = 25 AND 50% CENSORING)

Method / Estimator	α				
	$\hat{\alpha}$ (1)	Var($\hat{\alpha}$) (2)	Bias (3)	MSE (4)	Z-Value (5)
Adjusted Method	10.0625	52.4497	0.0625	52.453606	0.0862995
Bootstrapping	9.9571	53.0047	-0.0429	53.00654	-0.058925
Buckley and James	9.8946	53.1053	-0.1054	53.116409	-0.1446344
Miller	9.8774	53.6226	-0.1226	53.637631	-0.1674235
Method / Estimator	**β**				
	$\hat{\beta}$ (6)	Var($\hat{\beta}$) (7)	Bias (8)	MSE (9)	Z-Value (10)
Adjusted Method	0.1951	0.0153	-0.0049	0.015324	-0.3961413
Bootstrapping	0.2046	0.0158	0.0046	0.0158211	0.365963
Buckley and James	0.2077	0.0213	0.0077	0.0213592	0.5275953
Miller	0.2052	0.0279	0.0052	0.0279270	0.3113158

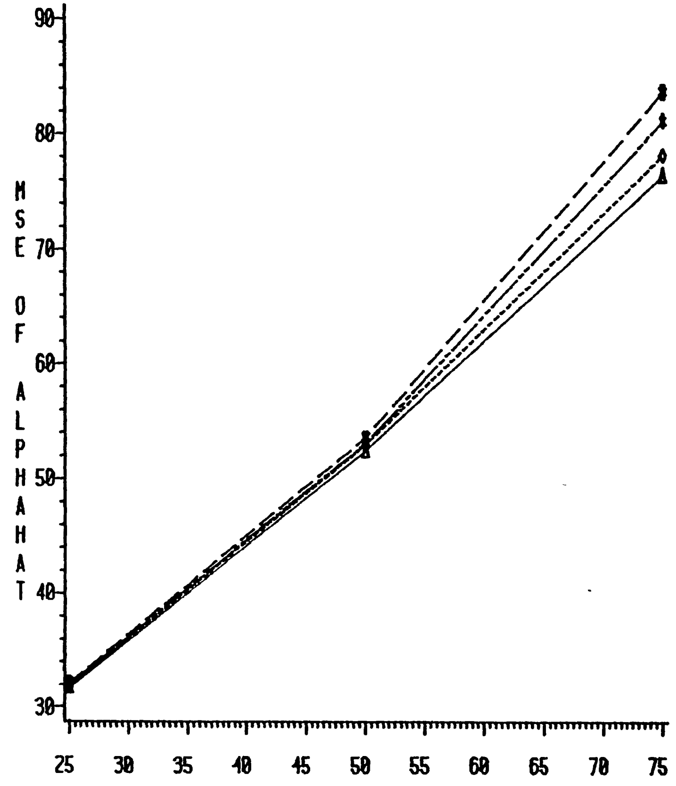
TABLE XXXIII

SIMULATIONS CALCULATING THE ESTIMATES OF α AND β BASED ON 100 REPLICATIONS (α = 10, β = 0.2, $X_i \sim U(1,100)$, $\varepsilon_i \sim N(0,100)$, SAMPLE SIZE = 25 AND 75% CENSORING)

Method \ Estimator	$\hat{\alpha}$ (1)	Var($\hat{\alpha}$) (2)	Bias (3)	MSE (4)	Z-Value (5)
Adjusted Method	9.8527	76.5433	-0.1473	76.569997	-0.1683639
Bootstrapping	9.7113	78.2145	-0.2887	78.297848	-0.3264398
Buckley and James	9.8116	81.2462	-0.1884	81.281695	-0.2090159
Miller	9.7762	83.7666	-0.2238	83.816746	-0.2445257

Method \ Estimator	$\hat{\beta}$ (6)	Var($\hat{\beta}$) (7)	Bias (8)	MSE (9)	Z-Value (10)
Adjusted Method	0.1894	0.1091	-0.0106	0.1092123	-0.3209175
Bootstrapping	0.2230	0.1102	0.0230	0.110729	0.6928465
Buckley and James	0.2197	0.1463	0.0197	0.196688	0.5150436
Miller	0.2283	0.1644	0.0283	0.1652008	0.697968

———— = the adjusted method
— • • — = Buckley and James' method
— — — = the bootstrapping method
—— —— = Miller's method



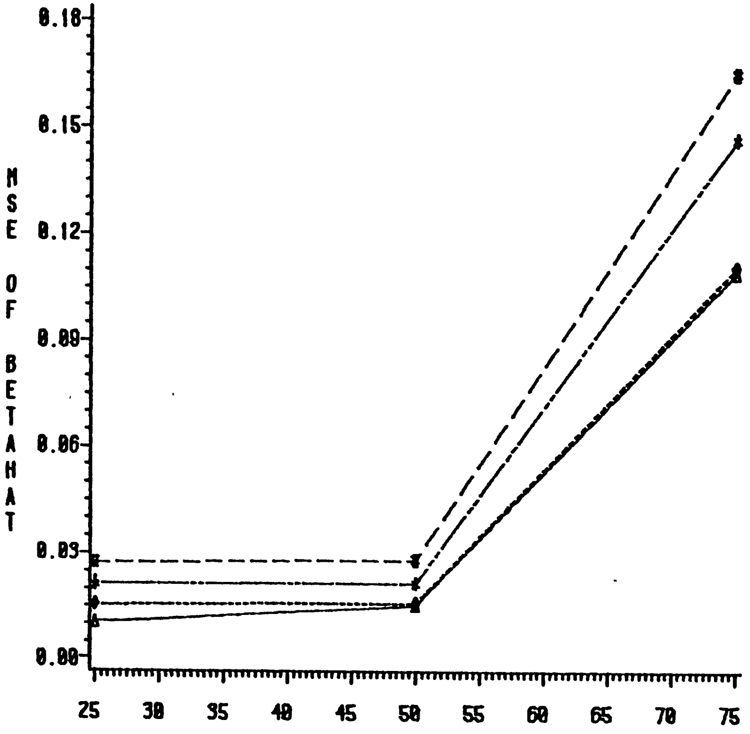


Figure 11. MSE of the Estimates of α and β Based on 100 Replications (α = 10, β = 0.2, $X_i \sim U(0,100)$, $\varepsilon_i \sim N(0,100)$ and Sample Size = 25)

TABLE XXXIV

SIMULATIONS CALCULATING THE ESTIMATES OF α AND β BASED ON 100 REPLICATIONS (α = 10, β = 0.2, $X_i \sim U(0,100)$, $\varepsilon_i \sim N(0,100)$, SAMPLE SIZE = 50 AND 25% CENSORING)

Method \ Estimator	α				
	$\hat{\alpha}$ (1)	Var($\hat{\alpha}$) (2)	Bias (3)	MSE (4)	Z-Value (5)
Adjusted Method	9.9161	14.1548	-0.0839	14.161639	-0.2230026
Bootstrapping	9.8936	14.3774	-0.1064	14.388721	-0.2806089
Buckley and James	9.8116	14.3920	-0.1884	14.427495	-0.4966155
Miller	9.6531	15.7103	-0.3469	15.83064	-0.8752095

Method \ Estimator	β				
	$\hat{\beta}$ (6)	Var($\hat{\beta}$) (7)	Bias (8)	MSE (9)	Z-Value (10)
Adjusted Method	0.1990	0.0029	-0.0010	0.002901	-0.1856953
Bootstrapping	0.1982	0.0032	-0.0018	0.00320324	-0.318198
Buckley and James	0.1969	0.0032	-0.0031	0.00320961	-0.5480077
Miller	0.1924	0.0046	-0.0076	0.00465776	-1.1205589

TABLE XXXV

SIMULATIONS CALCULATING THE ESTIMATES OF α AND β BASED ON 100 REPLICATIONS (α = 10, β = 0.2, $X_i \sim U(0,100)$, $\varepsilon_i \sim N(0,100)$, SAMPLE SIZE = 50 AND 50% CENSORING)

Method \ Estimator	$\hat{\alpha}$ (1)	Var($\hat{\alpha}$) (2)	Bias (3)	MSE (4)	Z-Value (5)
	α				
Adjusted Method	9.8633	21.1661	-0.1367	21.184787	-0.297131
Bootstrapping	9.9963	21.3277	-0.0837	21.334706	-0.1812397
Buckley and James	9.7624	22.0853	-0.2376	22.141754	-0.5055857
Miller	9.7101	22.1538	-0.2899	22.237842	-0.6159197

Method \ Estimator	$\hat{\beta}$ (6)	Var($\hat{\beta}$) (7)	Bias (8)	MSE (9)	Z-Value (10)
	β				
Adjusted Method	0.2023	0.0049	0.0023	0.00490529	0.3285714
Bootstrapping	0.2011	0.0048	0.0011	0.00480121	0.1587713
Buckley and James	0.2157	0.0049	0.0157	0.00514649	2.2428571
Miller	0.2171	0.0057	0.0171	0.00599241	2.2649503

TABLE XXXVI

SIMULATIONS CALCULATING THE ESTIMATES OF α AND β BASED ON 100 REPLICATIONS (α = 10, β = 0.2, $X_i \sim U(0,100)$, $\varepsilon_i \sim N(0,100)$, SAMPLE SIZE = 50 AND 75% CENSORING)

Method \ Estimator	$\hat{\alpha}$ (1)	Var($\hat{\alpha}$) (2)	Bias (3)	MSE (4)	Z-Value (5)
	α				
Adjusted Method	8.9610	40.3906	-1.039	41.470121	-1.6348405
Bootstrapping	8.9510	40.8262	-1.049	41.926601	-1.6417462
Buckley and James	8.9555	40.9967	-1.0445	42.08768	-1.1313006
Miller	8.9731	41.0542	-1.0269	42.108724	-1.6026894

Method \ Estimator	$\hat{\beta}$ (6)	Var($\hat{\beta}$) (7)	Bias (8)	MSE (9)	Z-Value (10)
	β				
Adjusted Method	0.2114	0.0112	0.0114	0.0113299	1.0771987
Bootstrapping	0.2103	0.0116	0.0103	0.0117060	0.9563309
Buckley and James	0.2227	0.0127	0.0227	0.0132152	2.0142993
Miller	0.2150	0.0129	0.0150	0.013125	1.3206764

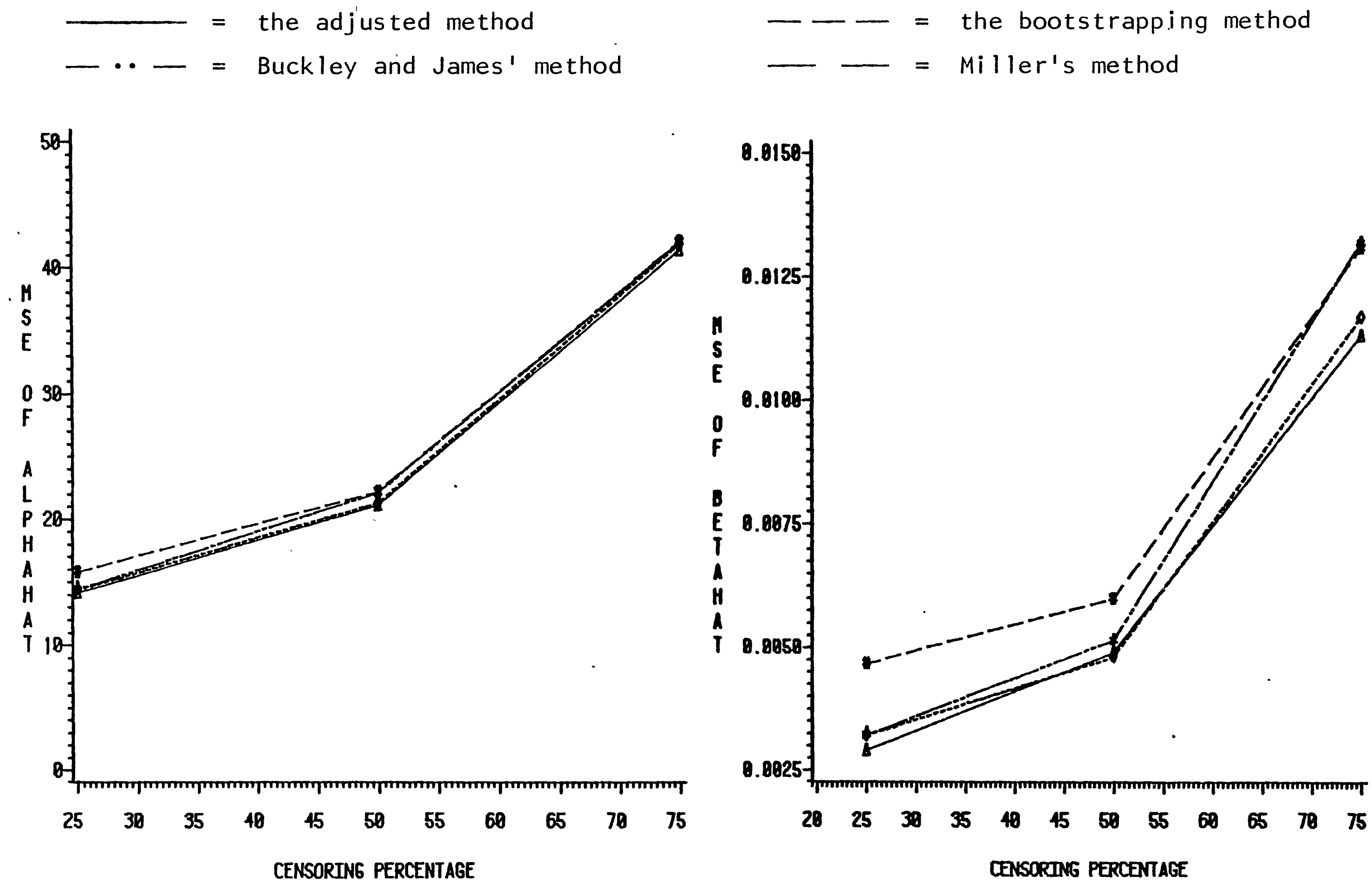


Figure 12. MSE of the Estimates of α and β Based on 100 Replications (α = 10, β = 0.2, $X_i \sim U(0,100)$, $\varepsilon_i \sim N(0,100)$ and Sample Size = 50)

TABLE XXXVII

SIMULATIONS CALCULATING THE ESTIMATES OF α AND β BASED ON 100 REPLICATIONS (α = 10, β = 0.2, $X_i \sim U(0,100)$ $\varepsilon_i \sim N(0,100)$, SAMPLE SIZE = 75 AND 25% CENSORING)

Method \ Estimator	$\hat{\alpha}$ (1)	Var($\hat{\alpha}$) (2)	Bias (3)	MSE (4)	Z-Value (5)
Adjusted Method	10.2116	9.3562	0.2116	9.4009746	0.6917766
Bootstrapping	10.1753	10.2012	0.1753	10.231930	0.5488533
Buckley and James	11.0994	9.9962	1.0994	11.20488	3.4772688
Miller	9.7524	10.3611	-0.2476	10.622406	-0.7618966

Method \ Estimator	$\hat{\beta}$ (6)	Var($\hat{\beta}$) (7)	Bias (8)	MSE (9)	Z-Value (10)
Adjusted Method	0.1973	0.0027	-0.0027	0.00270729	-0.5196152
Bootstrapping	0.1986	0.0028	-0.0014	0.00280196	-0.2645751
Buckley and James	0.1951	0.0028	-0.0049	0.00282401	-0.9260129
Miller	0.1902	0.0041	-0.0098	0.00419604	-1.5305029

TABLE XXXVIII

SIMULATIONS CALCULATING THE ESTIMATES OF α AND β BASED ON 100 REPLICATIONS ($\alpha = 10$, $\beta = 0.2$, $X_i \sim U(0,100)$ $\varepsilon_i \sim N(0,100)$, SAMPLE SIZE = 75 AND 50% CENSORING)

Method \ Estimator	α				
	$\hat{\alpha}$ (1)	Var($\hat{\alpha}$) (2)	Bias (3)	MSE (4)	Z-Values (5)
Adjusted Method	10.5980	12.8917	0.5980	13.249304	1.6655056
Bootstrapping	10.3961	13.1572	0.3961	13.314095	1.0920012
Buckley and James	10.4114	13.2627	0.4114	13.431950	1.1296615
Miller	10.2919	15.0021	0.2919	15.087306	0.7536298

Method \ Estimator	β				
	$\hat{\beta}$ (6)	Var($\hat{\beta}$) (7)	Bias (8)	MSE (9)	Z-Values (10)
Adjusted Method	0.1947	0.0040	-0.0053	0.00402809	-0.8380035
Bootstrapping	0.1926	0.0041	-0.0074	0.00415476	-1.1556858
Buckley and James	0.1943	0.0049	-0.0057	0.00493244	-0.8142857
Miller	0.1928	0.0057	-0.0072	0.00575184	-0.9536133

TABLE XXXIX

SIMULATIONS CALCULATING THE ESTIMATES OF α AND β BASED ON 100 REPLICATIONS (α = 10, β = 0.2, $X_i \sim U(0,100)$ $\varepsilon_i \sim N(0,100)$, SAMPLE SIZE = 75 AND 75% CENSORING)

Method \ Estimator	α				
	$\hat{\alpha}$ (1)	Var($\hat{\alpha}$) (2)	Bias (3)	MSE (4)	Z-Value (5)
Adjusted Method	10.6324	28.2364	0.6324	1.1901103	1.1901103
Bootstrapping	10.4519	29.0556	0.4519	0.8383539	0.8383539
Buckley and James	10.2977	32.4192	0.2977	0.5228506	0.5228506
Miller	10.2913	32.9909	0.2913	0.5071581	0.5071581
Method \ Estimator	**β**				
	$\hat{\beta}$ (6)	Var($\hat{\beta}$) (7)	Bias (8)	MSE (9)	Z-Value (10)
Adjusted Method	0.1888	0.0077	-0.0112	0.00782544	-1.2713585
Bootstrapping	0.1863	0.0081	-0.0137	0.00828769	-1.5222222
Buckley and James	0.1891	0.0083	-0.0109	0.00841881	-1.1964304
Miller	0.1936	0.0102	-0.0064	0.0102409	-0.6336944

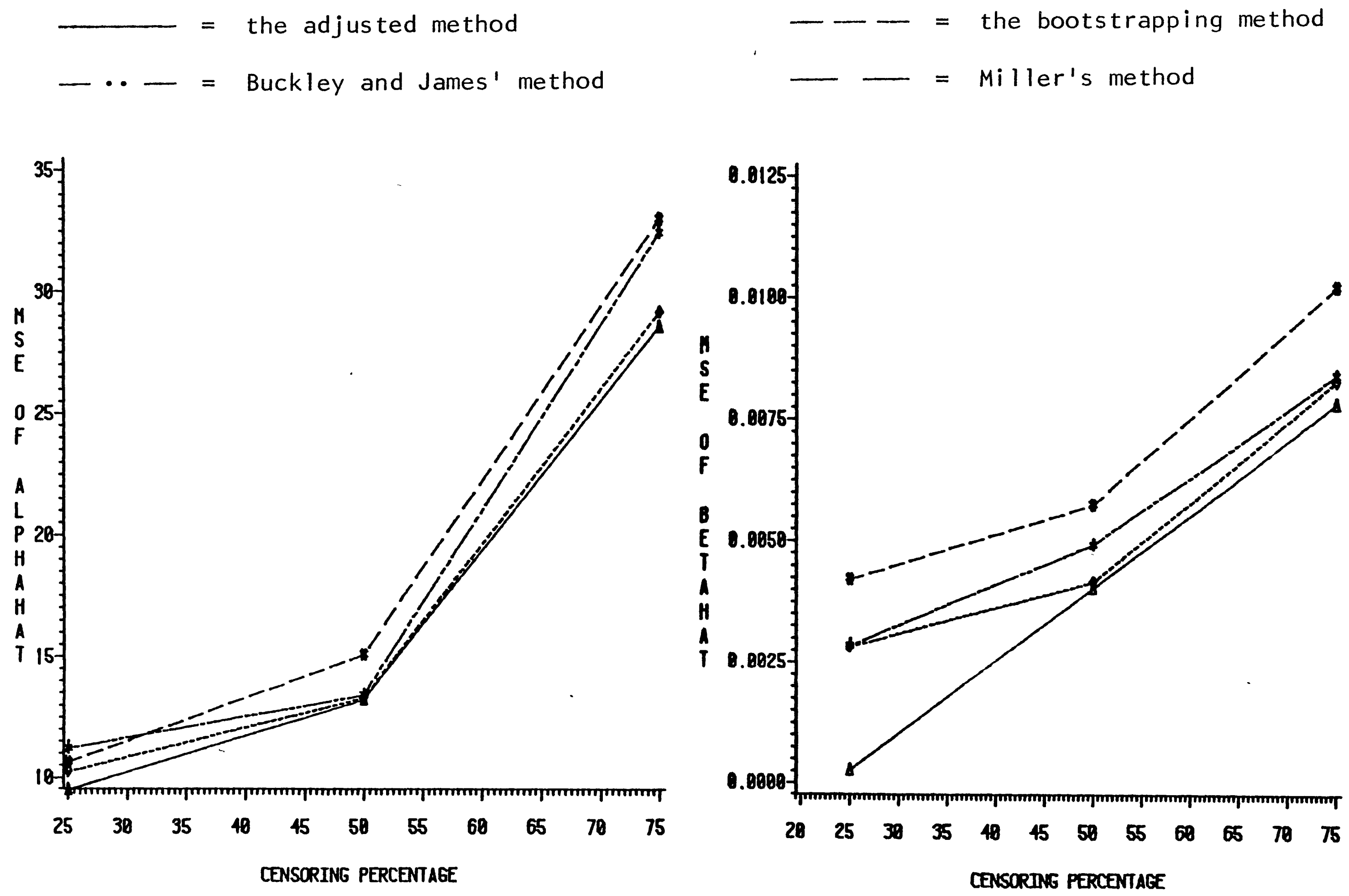


Figure 13. MSE of the Estimates of α and β Based on 100 Replications (α = 10, β = 0.2, $X_i \sim U(0,100)$, $\varepsilon_i \sim N(0,100)$ and Sample Size = 75)

as the sample size is increased regardless of the fraction of censoring in each trial.

Table XXXI - Table XXXIX and Figure 11 - Figure 13 show that if we fix the sample size and change the amount of censoring level 25, 50, 75%, the MSE for all methods would increase as the amount of censoring level increases. At the same time, if the sample size is increased, the MSE are decreased.

The results from these tables show that the adjusted method and the bootstrapping method are good choices to estimate regression coefficients even though these are some violations of independence between Y_i and C_i.

4.3 Heart Transplant Data

The Stanford Heart Transplantation program was begun in October 1967. By February 1980, 184 patients had received heart transplants. A few of these had multiple transplants. Their survival times (uncensored or censored at 2/1980) are displayed in Appendix B along with their ages at the time of the first transplant. Also included are their T5 mismatch scores which measure the degree of tissue incompatibility between the donor and recipient hearts with respect to HLA antigens.

Other variables such as waiting time to transplant, time since program inception, and previous open-heart surgery which were analyzed in some of the previous studies have not been included in this study. Also, those patients who entered the program but never received a transplant are excluded.

In analyzing the T5 mismatch scores, Miller (1976) and Crowley and Hu (1977) made a distinction between deaths primarily due to rejection of the donors' hearts by the recipients' immune system and non-rejection related deaths. The latter were treated as censored observations. This distinction is maintained in this study.

Table XXXXX - Table XXXXI gives the regression coefficient' estimators for age along and T5 mismatch scores alone and their estimated standard deviations. Figures 14 and 15 show how the estimated regression lines fit the data in both age and T5 mismatch scores for all methods.

TABLE XXXX

REGRESSION ESTIMATES AND STANDARD DEVIATIONS FOR LOG_{10} OF TIME TO DEATH VERSUS AGE AT TRANSPLANT WITH n = 157 STANFORD HEART TRANSPLANT PATIENTS

	Intercept		Age	
Estimator	$\hat{\alpha}$	$\hat{S}D(\hat{\alpha})$	$\hat{\beta}$	$\hat{S}D(\hat{\beta})$
Adjusted Method	3.9761	0.6256	-0.0454	0.0140
Bootstrapping	3.7993	0.6175	-0.0412	0.0138
Buckley and James	4.2421	0.6314	-0.0513	0.0141
Miller	3.6486	0.6315	-0.0389	0.0141

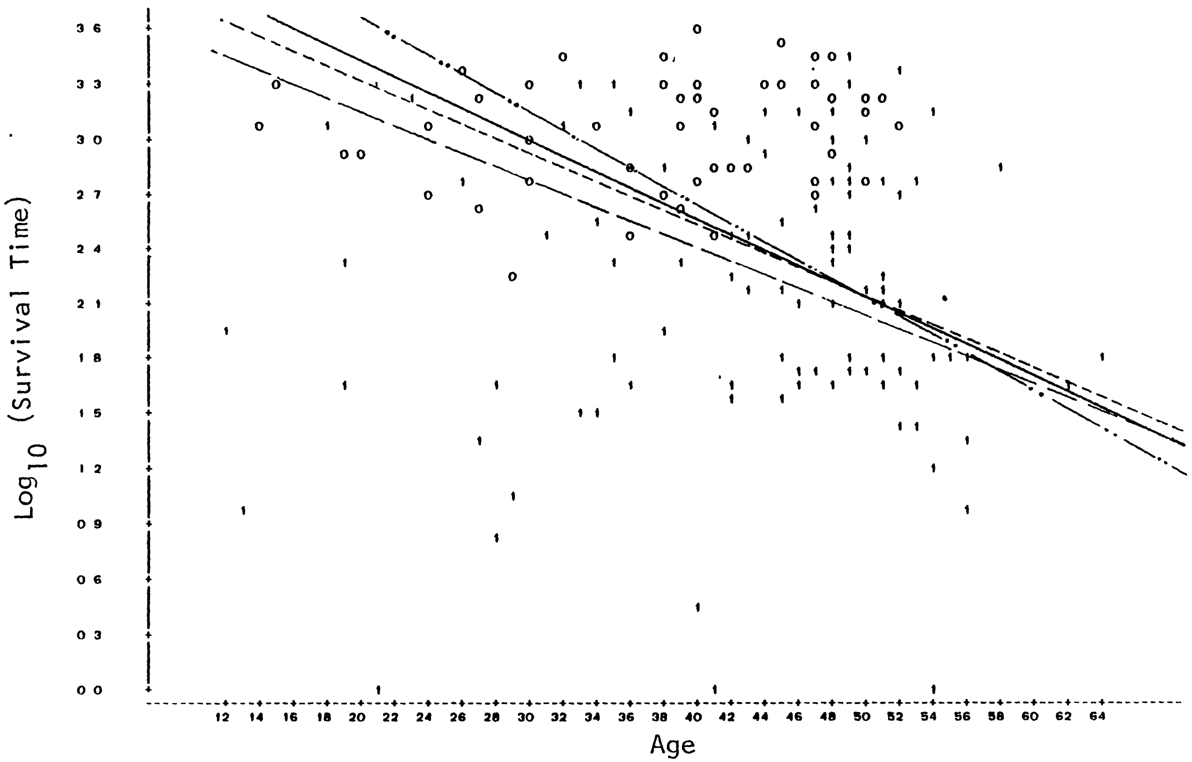
NOTE: 30 iterations are repeated for both Buckley and James' method and Miller's method. 100 bootstrap samples are calculated for the bootstrapping method.

TABLE XXXXI

REGRESSION ESTIMATES AND STANDARD DEVIATIONS FOR LOG_{10} OF TIME TO DEATH VERSUS T5 MISMATCH SCORES WITH m = 157 STANFORD HEART TRANSPLANT PATIENTS

Estimator	Intercept $\hat{\alpha}$	Intercept $\hat{S}D(\hat{\alpha})$	T5 $\hat{\beta}$	T5 $\hat{S}D(\hat{\beta})$
Adjusted Method	3.2186	0.2810	-0.0124	0.0120
Bootstrapping	3.2144	0.2800	-0.0136	0.0118
Buckley and James	3.2289	0.2826	-0.0130	0.0124
Miller	3.2401	0.2863	-0.0041	0.0133

NOTE: 30 iterations are repeated for both Buckley and James' method and Miller's method. 100 bootstrap samples are calculated for the bootstrapping method.



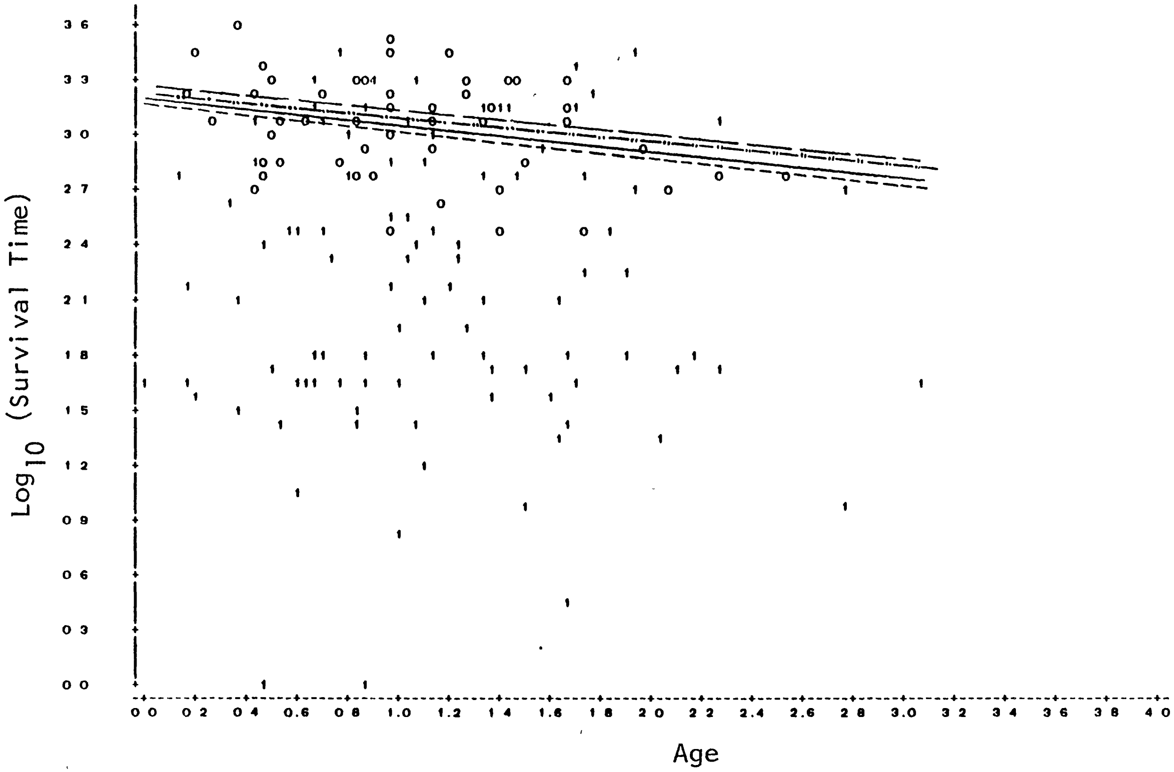
———— = the adjusted method

— — — — = the bootstrapping method

— •• — = Buckley and James' method

—— —— = Miller's method

Figure 14. Scatterplot of Log_{10} Survival Time (in Days) Versus Age at Transplant (in Years) for 157 Stanford Heart Transplant Patients. Patients Denoted by "1" are Deceased and Those by "0" Were Still Alive as of February 1980



= the adjusted method

= the bootstrapping method

= Buckley and James' method

= Miller's method

Figure 15. Scatterplot of Log$_{10}$ Survival Time (in Days) Versus T5 Mismatch Score for 157 Stanford Heart Transplant Patients. Patients Denoted by "1" are Deceased and Those by "0" Were Still Alive as of February 1980

CHAPTER V

SUMMARY AND CONCLUSIONS

This chapter outlines an application for the randomly censored linear regression model, summarizes some implications of the results of the regression parameter estimates, lists areas for future work.

5.1 Application

The statistical use is for making individual inference which includes statement about the estimation. For a person with a given covariate, the regression parameters in a linear model when the data is randomly censored are estimated. Often in medical studies when patients are entering a study randomly for a fixed time period, the observation on the survival time of a patient is incomplete in the sense that it is right censored. This censoring can be due to a number of causes; the patient was alive at the termination of the study, the patient withdrew alive during the study or the patient died of causes other than those under study. The problem arising is how to estimate parameters for such model, $T_i = \alpha + \beta x_i + \varepsilon_i$ where the variable T_i has been observed and subjected to a censoring variable. The objective of this thesis is to provide other reasonable choices of selecting the methods of analyzing such data since a few methods have been invented in the past years. Most of those methods require iterative routines which require much computer time. This has been intuitive disadvantage for those methods. In this thesis, we develop two methods: the adjusted method and the bootstrapping method, which do not need iterative schemes. However, the computer is still the main tool for these methods. We show that these methods provide the better choices in case one does not prepare using the other methods. For numerical comparisons, we present simulation results under various experiments.

5.2 Result Conclusions

The objective of this section is to summarize the numerical results of the proposed estimation methods. The more the amount of censoring level changes, the more the biases from all methods increase. Nevertheless, the adjusted method and the bootstrapping method are reasonable choices in terms of MSE of the estimates (in almost all the simulations). The adjusted method and the bootstrapping method can be good alternatives for one another in some simulations. However, the bootstrapping method needs a lot more computer memory than the adjusted method does. The biases of the estimates from both methods are very significant in some simulation experiments. This has been affected by increasing the sample size. Therefore, the performances of the estimates from both proposed methods are shown so that one is not reluctant to use both methods as the better candidates than Miller's method and as the reasonable methods comparing to Buckley and James' method. An estimate of the variance (σ^2) proposed in both methods has not been evaluated in the simulation study. However, it is estimated in Heart Transplant Data for both methods. One last conclusion from the simulation is that the bootstrapping method and the adjusted method cannot beat one another in terms of MSE basis. It sometimes provides higher MSE than the other does.

5.3 Further Work

Further works suggested are as follows:

1. The effect of various weighted matrices Σ instead of I in the model could be studied.

2. Simulation studies with general covariates with greater dimension (more than 1) should be evaluated.

3. In theoretical point of view, the estimates of α, β and σ^2 from both proposed methods have not been considered. This matter should be studied and more simulation should be done.

4. Numerous applications are possible in health administration as indicated by the examples mentioned throughout this thesis. This is an area that has been much explored.

5. Finally, the sample size needed for each problem should be evaluated.

A SELECTED BIBLIOGRAPHY

Amemiya, Takeshi, "Regression Analysis When the Dependent Variable is Truncated Normal." Econometrica, Vol. 41 (1973), 997-1016.

Blight, B. J. N., "Estimation From a Censored Sample for the Exponential Family." Biometrika, Vol. 37 (1970), 389-395.

Breslow, N., "Covariance Analysis of Censored Survival Data." Biometrics, Vol. 30 (1974), 89-99.

Breslow, N., Crowley, J., "A Large Sample Study of the Life Table and Product Limit Estimates Under Random Censorship." Annals of Statistics, Vol. 2 (1974), 437-453.

Buckley, J., James, I., "Linear Regression With Censored Data." Biometrika, Vol. 66, 3 (1979), 429-436.

Chatterjee, S., MeLeish, D. L., "Fitting Linear Regression Models to Censored Data by Least Squares and the Method of Maximum Likelihood." Technical Report No. 51, (1981), Stanford University, Department of Statistics.

Cox, D. R., "Regression Models and Life-Tables." Royal Statistical Society of London, Series B, Vol. 34 (1972), 187-202.

Crowley, J., Hu, M., "Covariance Analysis of Heart Transplant Survival Data." Journal of the American Statistical Association, Vol. 72 (1977), 27-36.

Csörgö, S., Horváth, L., "On the Koziol-Green Model for Random Censorship." Biometrika, Vol. 68 (1981), 391-401.

Dempster, A. P., Laird, N. M., Rubin, D. B., "Maximum Likelihood From Incomplete Data Via the EM Algorithm." Royal Statistical Society of London, Series B, Vol. 39 (1977), 1-21.

Durongwatana, S., "A Comparison of Least Squares Methods of Estimating Linear Regression Parameters With Censored Data." Unpublished M.S. Report (1983), Oklahoma State University.

Dyer, A. R., "An Analysis of the Relationship of Systalic Blood Pressure, Serum Cholesterol, and Smoking to 14-Year Mortality in the Chicago Peoples Gas Company Study-II. Coronary and Cardiovascular-Renal Mortality in the Two Competing Risk Models". Journal of Chronical Diseases, Vol. 38 (1973), 571-578.

Efron, B., "The Two Sample Problem With Censored Data." Proceedings of the Fifth Berkeley Symposium, Vol. 4 (1967), 831-853.

Efron, B., "Bootstrap Methods: Another Look at the Jackknife." Annals of Statistics, Vol. 7 (1979), 1-26.

Efron, B., "Censored Data and the Bootstrap." Journal of the American Statistical Association, Vol. 76 (1981), 312-320.

Efron, B., "Nonparametric Estimates of Standard Error: The Jackknife, the Bootstrap and Other Methods." Biometrika, Vol. 68 (1981), 589-599.

Efron, B., The Jackknife, the Bootstrap and Other Resampling Plans, Society for Industrial and Applied Mathematics (1982). Philadelphia, Pennsylvania 19103.

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

Elandt-Johnson, R. C. and Johnson, N. L., Survival Models and Data Analysis. John Wiley & Sons (1980), New York.

Eubank, R. L., LaRiccia, V. N., "Location and Scale Parameter Estimation From Randomly Censored Data." Communication Statistics-Theory and Methods, Vol. 11 (1982), 2869-2888.

Fetter, R. B., Shim, Y., Freeman, J. L., Averill, R. F., and Thompson, J. D., "Case Mix Definition by Diagnosis-Related Groups." Medical Care, Vol. 18 (1980), 1-53.

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

Friedman, J. H., Stuetzle, W., "The In-Out Method for Linear Regression With Censored Data." Technical Report No. 65, (1981), Stanford University, Division of Biostatistics.

Gehan, E. A., "A Generalized Uilcoxon Test for Comparing Arbitrarily Singly-Censored Samples." Biometrika, Vol. 62 (1965), 203-223.

Gilbert, J. D., "Random Censorship." Unpublished Ph.D. Dissertation (1962), University of Chicago.

Glasser, M., "Regression Analysis With Dependent Variable Censored." Biometrics, Vol. 21 (1965), 300-307.

Grundy, D. M., "The Fitting of Grouped Truncated and Grouped Censored Normal Distributions." Biometrika, Vol. 39 (1952), 252-259.

Gupta, A. K., "Estimation of the Mean and Standard Deviation of a Normal Population From a Censored Sample." Biometrika, Vol. 39 (1952), 260-273.

Kalbfleisch, J. D., Prentice, R. L., The Statistical Analysis of Failure Time Data. John Wiley & Sons (1980), New York.

Kaplan, E. L., Meier, P., "Nonparametric Estimation From Incomplete Observations." Journal of the American Statistical Association, Vol. 53 (1958), 457-481.

Koul, H., Susarla, V., Van Ryzin, J., "Regression Analysis With Randomly Right-Censored Data." The Annals of Statistics, Vol. 9 (1981), 1276-1288.

Lagakos, S. W., "A Stochastic Model for Censored-Survival Data in the Presence of an Auxiliary Variable." Biometrics, Vol. 32 (1976), 551-559.

Lagakos, S. W., "General Right Censoring and It's Impact on the Analysis of Survival Data." Biometrics, Vol. 35 (1979), 139-156.

Lawless, J. F., Statistical Models and Methods for Lifetime Data. John Wiley and Sons, Inc., New York (1982).

Meier, P., "Estimation of a Distribution Function From Incomplete Observations." Perspectives in Probability and Statistics, Academic Press, London (1975), 67-87.

Miller, R. G., "Least Squares Regression With Censored Data." Biometrika, Vol. 63 (1976), 449-464.

Miller, R. G., Survival Analysis. John Wiley & Sons, Inc., New York (1980).

Miller, R. G., Halpern, J., "Regression With Censored Data." Biometrika, Vol. 69 (1982), 521-531.

Nelson, W., Hahn, G. J., "Linear Estimation of a Regression Relationship From Censored Data, Part I - Simple Methods and Their Applications." Technometrics, Vol. 14 (1972), 247-269.

Nelson, W., Hahn, G. J., "Linear Estimation of a Regression Relationship From Censored Data, Part II - Best Linear Unbiased Estimation and Theory." Technometrics, Vol. 15 (1973), 133-150.

Nelson, W., Hahn, G. J., "A Comparison of Methods for Analyzing Censored Life Data to Estimate Relationships Between Stress and Product Life." IEEE Transactions on Reliability, Vol. R-23 (1974), 2-20.

Oakes, D., "The Asymptotic Information in Censored Survival Data." Biometrika, Vol. 64 (1977), 441-448.

Parr, W. C., "A Note on the Jackknife, the Bootstrap and the Delta Method Estimators of Bias and Variance." Biometrika, Vol. 70 (1983), 719-722.

Prentice, R. L., "Exponential Survival With Censoring and Explanatory Variables." Biometrika, Vol. 60 (1973), 279-288.

Pierce, D. A., Stewart, W. H., Kopecky, K. J., "Distribution-Free Regression Analysis of Grouped Survival Data." Biometrics, Vol. 35, (1979), 785-793.

Quenouille, M., "Approximate Tests of Correlation in Time Series." Royal Statistical Society of London, Series B, Vol. 11 (1949), 18-84.

Schmee, J., Hahn, G. J., "A Simple Method for Regression Analysis With Censored Data." Technometrics, Vol. 21 (1979), 417-434.

Schmee, J., Hahn, G. J., "A Computer Program for Simple Linear Regression With Censored Data." Journal of Quality Technology, Vol. 13 (1981), 264-269.

Susarla, V., Vau Ryzin, J., "Large Sample Theory for an Estimation of the Mean Survival Time From Censored Samples." The Annals of Statistics, Vol. 8 (1980), 1002-1016.

Tiku, M. L., "A Note on Estimating the Location and Scale Parameters of the Exponential Distribution From a Censored Sample." The Australian Journal of Statistics, Vol. 9 (1967), 49-54.

Tiku, M. L., "Estimating the Parameters of Normal and Logistic Distributions From Censored Samples." The Australian Journal of Statistics, Vol. 10 (1968), 64-74.

Tiku, M. L., "Estimating the Parameters of Log-Normal Distributions From Censored Samples." Journal of the American Statistical Association, Vol. 63 (1968), 134-140.

Tiku, M. L., "Testing Group Effects From Type II Censored Normal Samples in Experimental Design." Biometrics, Vol. 29 (1973), 25-33.

Tiku, M. L., "Estimating and Testing Group Effects From Type I Censored Normal Samples in Experimental Design." Communication Statistics-Theory and Methods, Vol. A6 (1977), 1485-1501.

Tiku, M. L., "Linear Regression Model With Censored Observations." Communication Statistics-Theory and Methods, Vol. A7 (1978), 1219-1232.

Van Ryzin, J., Koul, H., Susarla, V., "Multi-Step Estimation of Regression Coefficients in Linear Model With Censored Survival Data." Technical Report B-17, Division of Biostatistics, Columbia University (1981).

Van Ryzin, J., Koul, H., Susarla, V., "Least Squares Regression Aanlysis With Survival Data." Technical Report B-16, Division of Biostatistics, Columbia University (1981).

APPENDIX A

A PROGRAM FOR THE ADJUSTED METHOD

```
DATA SIMULATE;
     INPUT SEED;
      LIST;
      CARDS;
1671983
;
PROC MATRIX;
     SEED1=1671983;
     SEED2=2354076;
     N=20;
     BETA=1/0.2;
     NU=100;
     CENRATE=0;
  DO NTRLS=1 TO NU;
     SEED1=SEED1+10;
     SEED2=SEED2+20;
       E=J(N,1,0);
       C=J(N,1,0);
       DO K=1 TO N;
          E(K,1)=RANNOR(SEED1);
          C(K,1)=1+1.5*2*K-0.015*(2*K)**2;
       END;
     I1=J(N,1,1);
     I2=1:N;
        II=I2#2;
        X=I1||II';
        Y=X*BETA;
        Y=Y+E;
     Z=Y><C;
        DELTA=J(N,1,1);
        P=J(N,1,0);
        DIST=J(N,1,0);
        YNEW=J(N,1,0);
        ID=I(N);
       NUN=0;
   DO I=1 TO N;
      IF Y(I,1)>C(I,1) THEN DELTA(I,1)=0;
      END;
         A=DIAG(DELTA);
          NUN=TRACE(A);
         CENRATE=CENRATE+(N-NUN);
         BETAK_1=INV(X'*A*X)*X'*A*Z;
                R=Z-X*BETAK_1;
                YHAT=X*BETAK_1;
           DO KK=1 TO N;
              YNEW=J(N,1,YHAT(KK,1));
              YNEW=YNEW+R;
              YDEL=YNEW||R||DELTA;
              ERROR=YDEL;
              YDEL(RANK(YDEL(,1)),)=ERROR;
                   ANEW=DIAG(YDEL(,3));
                   RESD=YDEL(,2);
                DO I=1 TO N;
                   P(I,1)=SQRT((N+1-I)#/(N+2-I));
                END;
```

```
                DO I=1 TO N;
                   IND=ID;
                   KP=1;
                 DO K=1 TO I;
                    IND(K,K)=0;
                 END;
                 IND(I,I)=1;
                 CHECK=VECDIAG(IND*(ID-ANEW));
                 C1=CHECK;
            DO M=1 TO N;
               IF C1(M,1)=1 THEN KP=KP*P(M,1)*P(M,1);
            END;
              DIST(I,1)=KP;
           END;
          BIAS=(RESD'*ANEW*DIST)#/(TRACE(ANEW)-SUM(ANEW*DIST));
          Z(KK,1)=Z(KK,1)+BIAS;
      END;
    BETAHAT=INV(X'*A*X)*X'*A*Z;
    BETAH=BETAHAT';
    ALLTRLS=ALLTRLS//BETAH;
  END;
    CENRATE=CENRATE#/NU;
    RESULT=ALLTRLS;
  OUTPUT RESULT OUT=TEMP1;
  OUTPUT CENRATE OUT=TEMP2;
 DATA TEMP3;SET TEMP1;
      DROP ROW:
      RENAME COL1=ALPHAHAT
             COL2=BETAHAT:
 DATA TEMP4;SET TEMP2;
      DROP ROW;
      RENAME COL1=CENSOR;
PROC UNIVARIATE DATA=TEMP3;
    VAR ALPHAHAT BETAHAT;
```

APPENDIX B

A PROGRAM FOR THE BOOTSTRAPPING METHOD

```
DATA SIMULATE;
     INPUT NUM_TRLS ALPHA BETA N;
     SEED1=1671983;
     SEED2=2354076;
       SEED3=632704;
     LIST;
     CARDS;
100 1 0.2 10
;
DATA TEMP1;SET SIMULATE;
     DO I=1 TO NUM_TRLS;
        SEED1=SEED1+10;
       SEED2=SEED2+20;
        DO BS=1 TO N;
           ERROR=RANNOR(SEED1)*10;
           X=2*BS;
           C=20*RANUNI(SEED2)+(ALPHA+BETA*X);
           Y=ALPHA+BETA*X+ERROR;
           T=MIN(Y,C);
        OUTPUT;
        END;
     END;
PROC SYSREG DATA=TEMP1 NOPRINT OUT=B OUTEST=B1;BY I;
     MODEL T=X;
     OUTPUT P=THAT
            R=TRESID;
PROC DELETE DATA=TEMP1;
DATA TEMP2;SET SIMULATE;
     DO I=1 TO NUM_TRLS;
        DO TRIAL=1 TO 100;
           DO SAMPLE=1 TO N;
              BS=INT(RANUNI(SEED3)*N)+1;
           OUTPUT;
           END;
        END;
     END;
PROC SORT DATA=TEMP2;BY I BS;
PROC SORT DATA=B;BY I BS;
     DATA BNEW;SET B;
         DROP ERROR X C Y T THAT;
         OUTPUT;
     DATA SIMUL;MERGE TEMP2 BNEW;BY I BS;
          RENAME TRESID=RESD;
          IF SAMPLE=. THEN DELETE;
     DATA SIMUL1;MERGE SIMUL B1;BY I;
          DROP _TYPE_ _MODEL_ _DEPVAR_ T;
          RENAME INTERCEP=ALPHAHAT X=BETAHAT _SIGMA_=SIGMA1;
PROC DELETE DATA=TEMP2 SIMUL B B1 BNEW;
PROC SORT DATA=SIMUL1;BY I TRIAL BS;
     DATA TEMP3;SET SIMULATE;
          DO I=1 TO NUM_TRLS;
             DO TRIAL=1 TO 100;
                DO KK=1 TO N;
```

```
                    INDV=2*KK;
                    OUTPUT;
                    END;
              END;
           END;
      DATA TEMP4;MERGE SIMUL1 TEMP3;
           YBOOT=ALPHAHAT+BETAHAT*INDV+RESD;
PROC DELETE DATA=SIMUL1 TEMP3;
PROC SYSREG DATA=TEMP4 NOPRINT OUTEST=EST1
             OUT=A;BY I TRIAL;
      MODEL YBOOT=INDV;
            OUTPUT P=YBHAT
                   R=YBRESID;
      DATA TEMP5;SET EST1;
         DROP _TYPE_ _MODEL_ _DEPVAR_;
         RENAME INTERCEP=ALPBOOT INDV=BETABOOT _SIGMA_=SIGMAB;
         OUTPUT;
      DATA TEMP6;MERGE TEMP4 TEMP5;BY I TRIAL;
PROC DELETE DATA=TEMP4 TEMP5;
PROC MEANS DATA=TEMP6 NOPRINT;BY I;
      VAR ALPBOOT ALPHAHAT BETABOOT BETAHAT;
      OUTPUT OUT=MNBOOT
             N=NBOOT
             MEAN=MBOOT1 M1 MBOOT2 M2;
      DATA FINAL;SET MNBOOT;
           ABOOT=M1-(MBOOT1-M1);
           BBOOT=M2-(MBOOT2-M2);
           OUTPUT;
PROC UNIVARIATE DATA=FINAL;
      VAR ABOOT BBOOT;
```

APPENDIX C

A PROGRAM FOR BUCKLEY AND JAMES' METHOD

```
DATA SIMULATE;
     INPUT SEED;
      LIST;
       CARDS;
1672983
;
PROC MATRIX;
     SEED1=1671983;
     SEED2=2354076;
     N=10;
    BETA=1/0.2;
     NU=100;
     CENRATE=0;
  DO NTRLS=1 TO NU;
     SEED1=SEED1+10;
     SEED2=SEED2+20;
       E=J(N,1,0);
       C=J(n,1,0);
       DO K=1 TO N;
          E(K,1)=RANNOR(SEED1)*10;
          C(K,1)=20*RANUNI(SEED2)+1+0.2*2*K;
       END;
     I1=J(N,1,1);
     I2=1:N;
     II=I2#2;
        X=I1||II';
        Y=X*BETA;
        Y=Y+E;
     Z=Y><C;
        DELTA=J(N,1,1);
        P=J(N,1,0);
        DIST=J(N,1,0);
        JUMP=J(N,1,0);
        Q=J(N,1,0);
        ID=I(N);
       NUN=0;
    DO I=1 TO N;
       IF Y(I,1)>C(I,1) THEN DELTA(I,1)=0;
       END;
          A=DIAG(DELTA);
           NUN=TRACE(A);
          CENRATE=CENRATE+(N-NUN);
          BETAK_1=INV(X'*A*X)*X'*A*Z;
          ITER=0;
          DIFF=J(2,1,1);
             DO WHILE(MAX(ABS(DIFF))>0.0001 AND ITER<20);
          ITER=ITER+1;
          R=Z-X*BETAK_1;
          RDEL=R||DELTA||X||Z;
          ERROR=RDEL;
          RDEL(RANK(RDEL(,1)),)=ERROR;
              ANEW+DIAG(RDEL(,2));
          X0=RDEL(,3);
          X1=RDEL(,4);
              XNEW=X0||X1;
```

```
              ZNEW=RDEL(,5);
     DO I=1 TO N;
          P(I,1)=SQRT((N+1-I)#/(N+2-I));
          END;
     DO I=1 TO N;
          IND=ID;
          KP=1;
       DO K=1 TO I;
          IND(K,K)=0;
          END;
          CHECK=VECDIAG(IND*ANEW);
          C1=CHECK;
         DO M=1 TO N;
             IF C1(M,1)=1 THEN KP=KP*P(M,1)*P(M,1);
             END;
          DIST(I,1)=KP;
      END;
   JUMP(1,1)=DIST(1,1);
       DO I=2 TO N;
           JUMP(I,1)=DIST(I,1)-DIST(I-1,1);
            END;
   HD=SQRT(JUMP);
   IND=ID;
        DO I=1 TO N;
           DO K=1 TO I;
               IND(K,K)=0;
           END;
 Q(I,1)=XNEW(I,)*BETAK_1+((JUMP'*IND*RDEL(,1))#/(HD'*IND*HD));
 END;
     YSTAR=ANEW*ZNEW+(ID-ANEW)*Q;
     BETAK=INV(XNEW'*XNEW)*XNEW'*YSTAR;
     DIFF=BETAK-BETAK_1;
    BETAK_1=BETAK;
     END;
 SIGMA2=((YSTAR-XNEW*BETAK)'*(YSTAR-XNEW*BETAK))#/(N-2);
 COVMTR=SIGMA2*INV(XNEW'*XNEW);
 ESTVAR=VECDIAG(COVMTR);
    BETAHAT=BETAK'||SIGMA2||ESTVAR';
    ALLTRLS=ALLTRLS//BETAHAT;
  END;
    CENRATE=CENRATE#/NU;
    RESULT=ALLTRLS;
 OUTPUT RESULT OUT=TEMP1;
 OUTPUT CENRATE OUT=TEMP2;
      DATA TEMP3;SET TEMP1;
           DROP ROW;
           RENAME COL1=ALPHAHAT
                  COL2=BETAHAT;
      DATA TEMP4;SET TEMP2;
           DROP ROW;
           RENAME COL1=CENSOR;
 PROC PRINT DATA=TEMP3;
 PROC PRINT DATA=TEMP4;
 PROC CHART DATA=TEMP3;
      VBAR ALPHAHAT BETAHAT;
```

```
PROC UNIVARIATE DATA=TEMP3;
     VAR ALPHAHAT BETAHAT;
```

APPENDIX D

A PROGRAM FOR MILLER'S METHOD

```
DATA SIMULATE;
     INPUT NNN;
      LIST;
       CARDS;
3
;
PROC MATRIX;
     SEED1=1671983;
      SEED2=2354076;
     N=10;
     BETA=1/-0.4;
     NU=100;
     CENRATE=0;
  DO NTRLS=1 TO NU;
     SEED1=SEED1+10;
     SEED2=SEED2+20;
     E=J(N,1,0);
     C=J(N,1,0);
     DO K=1 TO N;
        E(K,1)=RANNOR(SEED1)*10;
        C(K,1)=14*RANNOR(SEED2)+41;
     END;
     I1=J(N,1,1);
     I2=1:N;
     II=I2#2;
        X=I1||II';
        Y=X*BETA;
        Y=Y+E;
     Z=Y><C;
        DELTA=J(N,1,1);
        P=J(N,1,0);
        DIST=J(N,1,0);
        JUMP=J(N,1,0);
        Q=J(N,1,0);
        ID=I(N);
       NUN=0;
  DO I=1 TO N;
     IF Y(I,1)>C(I,1) THEN DELTA(I,1)=0;
     END;
        A=DIAG(DELTA);
         NUN=TRACE(A);
        CENRATE=CENRATE+(N-NUN);
        BETAK_1=INV(X'*A*X)*X'*A*Z;
        ITER=0;
        DIFF=J(2,1,1);
            DO WHILE(MAX(ABS(DIFF))>.0001 and ITER<20);
                ITER=ITER+1;
                R=Z-X*BETAK_1;
                RDEL=R||DELTA||X||Z;
                ERROR=RDEL;
                RDEL(RANK(RDEL(,1)),)=ERROR;
                    ANEW=DIAG(RDEL(,2));
                X0=RDEL(,3);
                X1=RDEL(,4);
                    XNEW=X0||X1;
```

```
                    ZNEW=RDEL(,5);
          DO I=1 TO N;
               P(I,1)=SQRT((N+1-I)#/(N+2-I));
               END;
          DO I=1 TO N;
               IND=ID;
               KP=1;
            DO K=1 TO I;
               IND(K,K)=0;
               END;
               CHECK=VECDIAG(IND*ANEW);
               C1=CHECK;
              DO M=1 TO N;
                  IF C1(M,1)=1 THEN KP=KP*P(M,1)*P(M,1);
                  END;
               DIST(I,1)=KP;
           END;
        JUMP(1,1)=DIST(1,1);
            DO I=2 TO N;
                JUMP(I,1)=DIST(I,1)-DIST(I-1,1);
                 END;
       WSTAR=GINV(DIAG(JUMP));
       BETAK=GINV(XNEW'*WSTAR*XNEW)*XNEW'*WSTAR*ZNEW;
          DIFF=BETAK-BETAK_1;
          BETAK_1=BETAK;
      END;
    SIGMA2=((ZNEW-XNEW*BETAK)'*WSTAR*(ZNEW-XNEW*BETAK))#/(N-2);
    COVMTR=SIGMA2*INV(XNEW'*WSTAR*XNEW);
    ESTVAR=VECDIAG(COVMTR);
       BETAHAT=BETAK'||SIGMA2||ESTVAR';
       ALLTRLS=ALLTRLS//BETAHAT;
     END;
       CENRATE=CENRATE#/NU;
       RESULT=ALLTRLS;
     OUTPUT RESULT OUT=TEMP1;
     OUTPUT CENRATE OUT=TEMP2;
  DATA TEMP3;SET TEMP1;
       DROP ROW;
       RENAME COL1=ALPHAHAT
              COL2=BETAHAT;
  DATA TEMP4;SET TEMP2;
       DROP ROW;
       RENAME COL1=CENSOR;
PROC PRINT DATA=TEMP3;
PROC PRINT DATA=TEMP4;
PROC CHART DATA=TEMP3;
     VBAR ALPHAHAT BETAHAT;
PROC UNIVARIATE DATA=TEMP3;
     VAR ALPHAHAT BETAHAT;
```

APPENDIX E

STANFORD HEART TRANSPLANT DATA

Observation No.	Patient No.	Survival Time	Dead=1 Alive=0	Age	T5 Mismatch Score	Log_{10} (Survival Time)
1	1	15	1	54	1.11	1.17609
2	2	3	1	40	1.66	0.47712
3	3	46	1	42	0.61	1.66276
4	4	623	1	51	1.32	2.79449
5	5	126	1	48	0.36	2.10037
6	6	64	1	54	1.89	1.80618
7	7	1350	1	54	0.87	3.13033
8	8	23	1	56	2.05	1.36173
9	9	279	1	49	1.12	2.44560
10	10	1024	1	43	1.13	3.01030
11	11	10	1	56	2.76	1.00000
12	12	39	1	42	1.38	1.59106
13	13	730	1	58	0.96	2.86332
14	14	1961	1	33	1.06	3.29248
15	15	136	1	52	1.62	2.13354
16	16	1	1	54	0.47	0.00000
17	17	836	1	44	1.58	2.92221
18	18	60	1	64	0.69	1.77815
19	19	3695	0	40	0.38	3.56761
20	20	1996	1	49	0.91	3.30016
21	21	1	1	41	0.87	0.00000
22	22	47	1	62	0.87	1.67210
23	23	54	1	49	2.09	1.73239
24	25	2878	1	49	0.75	3.45909
25	26	3410	0	45	0.98	3.53275
26	27	44	1	36	0.00	1.64345
27	28	994	1	48	0.81	2.99739
28	29	51	1	47	1.38	1.70757
29	30	1478	1	36	1.35	3.16967
30	31	254	1	48	1.08	2.40483

Observation No.	Patient No.	Survival Time	Dead=1 Alive=0	Age	T5 Mismatch Score	Log_{10} (Survival Time)
31	34	51	1	52	1.51	1.70757
32	35	323	1	48	1.82	2.50920
33	36	3021	0	38	0.98	3.48015
34	37	66	1	49	0.66	1.81954
35	38	2984	0	32	0.19	3.47480
36	39	2723	1	32	1.93	3.43505
37	40	550	1	48	0.12	2.74036
38	41	66	1	51	1.12	1.81954
39	42	65	1	45	1.68	1.81291
40	43	227	1	19	1.02	2.35603
41	44	2805	0	48	1.20	3.44793
42	45	25	1	53	1.68	1.39794
43	46	631	1	26	1.46	2.80003
44	47	2734	0	47	0.97	3.43680
45	48	12	1	29	0.61	1.07918
46	49	63	1	56	2.16	1.79934
47	50	2474	1	52	1.70	3.39340
48	51	1384	1	46	1.41	3.14114
49	52	544	1	52	1.94	2.73560
50	53	29	1	53	1.08	1.46240
51	54	48	1	53	3.05	1.68124
52	55	297	1	42	0.60	2.47276
53	56	1318	1	48	1.44	3.11992
54	57	1352	1	54	0.68	3.13098
55	58	50	1	46	2.25	1.69897
56	59	547	1	49	0.81	2.73799
57	60	431	1	47	0.33	2.63448
58	61	68	1	51	1.33	1.83251
59	62	26	1	52	0.82	1.41497
60	63	161	1	43	1.20	2.20683

Observation No.	Patient No.	Survival Time	Dead=1 Alive=0	Age	T5 Mismatch Score	Log_{10} (Survival Time)
61	65	2313	0	26	0.46	3.36418
62	66	1634	1	23	1.78	3.21325
63	67	146	1	45	0.16	2.16435
64	68	48	1	28	0.77	1.68124
65	69	2127	1	35	0.67	3.32777
66	70	263	1	49	0.48	2.41996
67	71	2106	0	40	0.86	3.32346
68	72	293	1	43	0.70	2.46687
69	73	2025	0	30	1.44	3.30643
70	74	2006	0	15	1.26	3.30233
71	75	2000	0	45	1.46	3.30103
72	76	1995	0	47	1.65	3.29994
73	77	1945	0	38	1.28	3.28892
74	78	65	1	55	0.69	1.81291
75	79	731	1	38	0.42	2.86392
76	80	1866	0	49	0.51	3.27091
77	81	538	1	49	2.76	2.73078
78	82	1846	0	44	0.83	3.26623
79	83	68	1	35	0.85	1.83251
80	84	1773	0	27	0.70	3.24871
81	85	1722	0	40	0.95	3.23603
82	86	928	1	50	1.12	2.96755
83	87	1718	0	39	1.77	3.23502
84	88	22	1	27	1.64	1.34242
85	89	40	1	42	1.59	1.60206
86	90	7	1	28	1.00	0.84510
87	91	1638	0	48	0.43	3.21431
88	92	1612	0	51	1.25	3.20737
89	93	25	1	52	0.53	1.39794
90	94	1534	1	44	1.71	3.18583
91	95	1547	0	50	0.18	3.18949
92	96	1271	1	32	1.05	3.10415

Observation No.	Patient No.	Survival Time	Dead=1 Alive=0	Age	T5 Mismatch Score	Log_{10} (Survival Time)
93	97	44	1	46	1.71	1.64345
94	98	1247	1	41	0.43	3.09587
95	99	1232	1	18	0.70	3.09061
96	100	191	1	42	1.74	2.28103
97	101	1393	0	46	0.95	3.14395
98	103	1378	0	41	1.65	3.13925
99	104	1373	0	41	1.38	3.13767
100	105	274	1	31	0.58	2.43775
101	106	31	1	33	0.36	1.49136
102	107	1341	0	50	1.13	3.12743
103	108	42	1	19	0.63	1.62325
104	109	381	1	45	0.98	2.58092
105	110	1264	0	52	0.64	3.10175
106	111	1262	0	34	1.68	3.10106
107	112	1261	0	47	0.82	3.10072
108	113	47	1	36	0.16	1.67210
109	114	1193	0	24	1.15	3.07664
110	115	626	1	53	1.74	2.79657
111	116	48	1	51	0.99	1.68124
112	117	1150	1	32	2.25	3.06070
113	118	45	1	48	0.65	1.65321
114	119	1116	0	14	0.54	3.04766
115	120	1107	0	18	0.25	3.04415
116	121	1102	0	39	1.35	3.04218
117	122	195	1	39	0.73	2.29003
118	123	30	1	34	0.84	1.47712
119	124	1040	0	43	0.50	3.01703
120	125	993	0	30	0.95	2.99695
121	127	729	1	49	1.10	2.86273
122	129	202	1	48	1.24	2.30535
123	130	841	0	48	0.86	2.92480

Observation No.	Patient No.	Survival Time	Dead=1 Alive=0	Age	T5 Mismatch Score	Log_{10} (Survival Time)
124	132	265	1	49	1.22	2.42325
125	133	1	1	21	0.47	0.00000
126	134	793	0	19	1.98	2.89927
127	135	328	1	34	1.02	2.51587
128	136	781	0	20	1.12	2.89265
129	137	752	0	43	1.50	2.87622
130	138	738	0	41	0.53	2.86806
131	139	86	1	12	1.26	1.93450
132	140	132	1	46	1.09	2.12057
133	141	663	0	36	0.47	2.82151
134	142	660	0	42	0.75	2.81954
135	143	221	1	35	1.04	2.34439
136	144	90	1	38	1.00	1.95424
137	145	619	0	47	0.90	2.79169
138	146	618	0	50	0.82	2.79099
139	147	576	0	53	2.25	2.76042
140	149	36	1	45	0.20	1.55630
141	150	549	0	40	2.53	2.73957
142	151	548	0	30	0.47	2.73878
143	152	541	0	47	0.43	2.73320
144	154	169	1	51	1.89	2.22789
145	155	122	1	51	1.33	2.08636
146	157	468	0	24	1.39	2.67025
147	158	464	0	38	2.07	2.66652
148	159	10	1	13	1.49	1.00000
149	162	406	0	39	1.18	2.60853
150	163	391	0	27	1.17	2.59218
151	165	50	1	50	0.50	1.69897
152	166	139	1	51	0.96	2.14301
153	167	322	0	36	1.73	2.50786
154	168	292	0	43	1.40	2.46538

Observation No.	Patient No.	Survival Time	Dead=1 Alive=0	Age	T5 Mismatch Score	Log_{10} (Survival Time)
155	169	278	0	41	0.98	2.44404
156	172	145	1	50	0.96	2.16137
157	174	176	0	29	1.72	2.24551

VITA

Supol Durongwatana

Candidate for the Degree of

Doctor of Philosophy

Thesis: REGRESSION MODEL WITH CENSORED OBSERVATIONS

Major Field: Statistics

Biographical:

Personal Data: Born in Bangkok, Thailand, January 22, 1957, the youngest son of Mr. Chalerm and Mrs. Supanee Durongwatana.

Education: Attended elementary school in Bangkok, Thailand; graduated from Triam Udom Suksa School, Bangkok, Thailand, in 1975; received the Bachelor of Arts degree in Statistical and Computing Science from Chulalongkorn University, Bangkok, Thailand, in 1979; received Master of Science degree in Statistics from Oklahoma State University, Stillwater, Oklahoma, in May, 1983; completed requirements for the Doctor of Philosophy degree at Oklahoma State University in December, 1985.

Professional Experience: Programmer at NCR Company, Bangkok, Thailand, 1979-1980; Graduate Teaching Assistant, Department of Statistics, Oklahoma State University, 1980-1982; Graduate Teaching Associate, Department of Statistics, Oklahoma State University, 1983-1985.

Professional Organizations: American Statistical Association, Mu Sigma Rho.