NAVAL POSTGRADUATE SCHOOL
Monterey, California

THESIS

AN ANALYSIS AND DEVELOPMENT OF A PRODUCTION
PREDICTOR MODEL FOR THE   ARMY RESERVE OFFICER
TRAINING CORPS PROGRAM

by

William C. Hopkinson

September 1988

Thesis Advisor: Harold J. Larson

REPORT DOCUMENTATION PAGE

1a Report Security Classification Unclassified			1b Restrictive Markings			
2a Security Classification Authority			3 Distribution/Availability of Report			
2b Declassification Downgrading Schedule			Approved for public release; distribution is unlimited.			
4 Performing Organization Report Number(s)			5 Monitoring Organization Report Number(s)			
6a Name of Performing Organization Naval Postgraduate School		6b Office Symbol *(if applicable)* 55	7a Name of Monitoring Organization Naval Postgraduate School			
6c Address *(city, state, and ZIP code)* Monterey, CA 93943-5000			7b Address *(city, state, and ZIP code)* Monterey, CA 93943-5000			
8a Name of Funding Sponsoring Organization		8b Office Symbol *(if applicable)*	9 Procurement Instrument Identification Number			
8c Address *(city, state, and ZIP code)*			10 Source of Funding Numbers			
			Program Element No	Project No	Task No	Work Unit Accession No
11 Title *(include security classification)* AN ANALYSIS AND DEVELOPMENT OF A PRODUCTION PREDICTOR MODEL FOR THE ARMY RESERVE OFFICER TRAINING CORPS PROGRAM.						
12 Personal Author(s) WILLIAM C. HOPKINSON						
13a Type of Report Master's Thesis	13b Time Covered From To		14 Date of Report *(year, month, day)* September 1988		15 Page Count 66	
16 Supplementary Notation The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.						

17 Cosati Codes			18 Subject Terms *(continue on reverse if necessary and identify by block number)*
Field	Group	Subgroup	Army ROTC Program, Regression Analysis, *Army personnel, aurad, (ITB)*

19 Abstract *(continue on reverse if necessary and identify by block number)* The purpose of this thesis is to examine and model the Army Reserve Officer Training Corps (ROTC) commissioning process in terms of several possible explanatory variables. Each of the variables, which included unemployment rate, average yearly college tuition, ROTC enrollment by class, advertising budget, scholarship program and propensity towards military service, were analyzed for trends in the data with respect to the dependent variable, the number of second lieutenants commissioned in a year. Four regression models were fitted to numerous combinations of the explanatory variables and the dependent variable. Three variables were found to be significant and possess the potential to predict the number of second lieutenants commissioned each year within the range of the data used for modeling. *Keywords:*

20 Distribution Availability of Abstract ☒ unclassified unlimited ☐ same as report ☐ DTIC users	21 Abstract Security Classification Unclassified		
22a Name of Responsible Individual Harold J. Larson	22b Telephone *(include Area code)* (408) 646-2473		22c Office Symbol 55La

DD FORM 1473,84 MAR

83 APR edition may be used until exhausted
All other editions are obsolete

security classification of this page

An Analysis and Development of a Production Predictor Model
for the Army Reserve Officer Training Corps Program.

by

WILLIAM C. HOPKINSON
CPT, United States Army
B.S., United States Military Academy, West Point, NY, 1980

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
September 1988

Author: _____

WILLIAM C. HOPKINSON

Approved by: _____
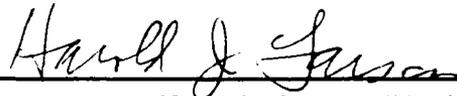
Harold J. Larson, Thesis Advisor

_____

Glenn F. Lindsay, Second Reader

_____
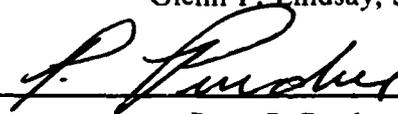
Peter P. Purdue, Chairman,
Department of Operations Research

_____
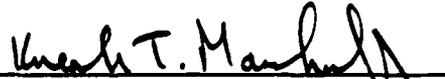
Kneale T. Marshall,
Dean of Information and Policy Sciences

ii

ABSTRACT

The purpose of this thesis is to examine and model the Army Reserve Officer Training Corps (ROTC) commissioning process in terms of several possible explanatory variables. Each of the variables, which included unemployment rate, average yearly college tuition, ROTC enrollment by class, advertising budget, scholarship program and propensity towards military service, were analyzed for trends in the data with respect to the dependent variable, the number of second lieutenants commissoned in a year. Four regression models were fitted to numerous combinations of the explanatory variables and the dependent variable. Three variables were found to be significant and possess the potential to predict the number of second lieutenants commissioned each year within the range of the data used for modeling.

iii

Accession For

NTIS GRA&I	☑
DTIC TAB	☐
Unannounced	☐
Justification	

By

Distribution/

Availability Codes

| Dist | Avail and/or Special |
| A-1 | |

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

I. INTRODUCTION

The Army Reserve Officer Training Corps (ROTC) plays a vital role in the development of future Army officers and serves as the primary source of commissioning of second lieutenants for the Active and Reserve Army and for the National Guard. During the early 1970's, the Army ROTC program suffered a decrease in enrollment partially due to the anti-war sentiment prevalent at many colleges. Many schools discontinued their association with the Army ROTC program. Curtailment of the draft and the introduction of the all volunteer army also contributed to the drop in Army ROTC enrollment from a high of 177,000 in 1966 to a low of 33,000 in 1973 [Ref. 1: p. 3-1]. As a result of the decrease in enrollment, the Army ROTC program of the late 1970's and early 1980's was unable to meet the demands of the reserve components. In an effort to improve the Army ROTC program, in 1980 Congress increased the number of scholarships from 6500 to 12,000. This attracted more quality students into the program. Currently, the Army ROTC program provides for approximately 75% of the commissioned officers in any particular year group, with Officer Candidate School (OCS) and the United States Military Academy (USMA) providing the remaining 25% of the officers. This makes the Army ROTC program the principle commissioning source of second lieutenants for the Army. At the present time, the Army ROTC program is commissioning approximately 8000 second lieutenants per year. The Army ROTC mission, as directed by the Department of the Army, is expected to increase by 35% to 10,800 lieutenants per year between now and 1995 in order to meet projected demands. During this same time frame, a 15% decrease in the number of males and females between the ages of 18 and 24 is expected leading to a decrease in college enrollment and a possible decrease in Army ROTC enrollment. With this reduction in the eligible population, it will become increasingly difficult for the Army ROTC program to meet its officer requirements in the future.

A. OBJECTIVE

The Army ROTC commissioning process consists of recruiting individuals into the Army ROTC program, training them to be officers in the United States Army, retaining them in the Army ROTC program, and finally upon successful completion of the program commissioning these individuals as second lieutenants. There are numerous factors which influence the Army ROTC commissioning process such as economics,

attitude of the country towards military service, cost of college, and college enrollment. The United States Army ROTC Cadet Command identified eight variables which they felt may have some influence on the commissioning process and proposed the following question. Can the number of second lieutenants commissioned in a particular year be explained by one or more of these variables and if so what is this relationship? The objective of this thesis is to explore the Army ROTC commissioning process and attempt to develop a model which can predict the number of second lieutenants commissioned in a year by the Army ROTC program as a function of these proposed explanatory variables. This thesis will use regression analysis to investigate if any single or multiple variable relationships exist between the dependent variable, the number of second lieutenants commissioned in a year, and the proposed explanatory variables.

B. BACKGROUND

The Army ROTC program consists of a scholarship and a non-scholarship program, both of which lead to commissioning as a second lieutenant if successfully completed. There are several ways an individual can enter either of these programs. For the non-scholarship program, an individual may enter the Army ROTC program during his or her freshmen year as a Military Science I (MSI) cadet or enter the Army ROTC program during his or her sophomore year as a Military Science II (MSII) cadet and could then be commissioned as a second lieutenant four or three years later respectively. If an individual enters the Army ROTC non-scholarship program following their sophomore year, they must attend a summer training program known as Camp Challenge. Following successful completion of this training, the individual is designated as a Military Science III (MSIII) cadet at the beginning of his or her junior year and may be commissioned as a second lieutenant two years later.

There are three types of scholarships available to students entering the Army ROTC scholarship program. The four year scholarship program provides the student with tuition and fees during all four years of college. An individual enters the four year scholarship program as an MSI cadet during his or her freshmen year and may be commissioned as a second lieutenant four years later. The three year scholarship program provides for tuition and fees for the last three years an individual is in college. An individual entering the three year scholarship program may be a cadet who entered the non-scholarship program as a MSI cadet and then received a three year scholarship at the beginning of his or her sophomore year, or the individual may have entered the Army ROTC scholarship program directly as an MSII cadet without having any prior

2

Army ROTC experience. The third scholarship program available is the two year scholarship program which provides tuition and fees for the final two years of college. Typically, the student receiving a two year scholarship has entered the Army ROTC program as a non-scholarship cadet during his or her freshmen or sophomore year and is then selected for a two year scholarship. A student may receive a two year scholarship without prior Army ROTC experience provided he or she successfully completes the summer training program Camp Challenge prior to the start of his or her junior year. Regardless of whether a cadet is in the scholarship or non-scholarship program, all cadets must attend advance camp prior to commissioning either following their junior or senior year to receive training in precommissioning skills.

Both the scholarship and non-scholarship programs have various obligations associated with them, depending on the amount of time the individual has been in the Army ROTC program and whether or not he or she has received a scholarship. Prior to the start of their junior year, cadets in the non-scholarship program are required to sign a contract obligating them to some military service. Following the signing of this contract, the individual receives a subsistence allowance of $100 per month during his or her junior and senior year. The obligations associated with the scholarship program vary with the type of scholarship the individual has received. For the four year scholarship program, the cadet does not incur a military obligation until the begining of his or her sophomore year. For the two and three year scholarship programs, an individual incurs a military obligation immediately upon accepting the scholarship.

C. DATA COLLECTION

The following sets of data were provided by the United States Army ROTC Cadet Command and identified as possible explanatory variables which may influence the dependent variable, the number of second lieutenants commissioned each year by the Army ROTC program.

1. Opening and closing Army ROTC enrollment reports from 1970 through 1987 provided the number of college students enrolled in the Army ROTC program for each year, by year group.

2. Scholarship reports from 1970 through 1987 provided the number and type of scholarships for each year, by year group.

3. Advertising budget reports from 1974 through 1987 provided the amount in current dollars spent on print advertising (newspapers, magazines) each year. The Army ROTC program currently does not advertise on television or radio and has not in the past.

4. Leads reports from 1973 through 1986 provided the number of individuals per year who responded to an advertisement by either filling out a postcard or making a telephone inquiry. A lead is defined as a response by an individual to some sort of print advertising such as a card in a magazine. This data does not provide any information as to whether or not these individuals ever enrolled in the Army ROTC program.

5. The national unemployment rate for each year from 1970 through 1987 was gathered from a Statistical Abstract [Ref. 2: p. 129].

6. Annual Freshmen College Enrollment reports from 1970 through 1987 provided the number of full time freshmen students who enrolled each year [Ref. 3: p. 130].

7. The average yearly college tuition from 1970 through 1987 provided historical data on the cost for tuition and fees for public and private schools [Ref. 3: p. 222].

8. Youth Attitude Tracking Survey (YATS) results from 1973 though 1985 provided a measurement of the propensity toward military service based on responses to a questionaire by high school students [Ref. 4: p. 20]. The primary use of this report is for enlisted recruiting but it is assumed to provide information on general sentiment toward military service.

Data was provided on the number of second lieutenants commissioned in each year by the Army ROTC program from 1970 through 1986. This was identified as the dependent variable whose behavior is to be modeled. Chapter II will examine each of the variables for trends over time and for a relationship with the dependent variable.

4

II. EXPLORATORY DATA ANALYSIS

A. INTRODUCTION

Exploratory data analysis consisted of investigating how each of the variables affected the Army ROTC commissioning process. The proposed explanatory variables were plotted against time and against the dependent variable in order to take an initial look at the problem and at each of the data sets. Recall from Chapter I that the dependent variable is the number of second lieutenants commissioned by the Army ROTC program in a year. These graphical representations provided a quick look at the data for trends over time and provided some insight into the relationship between the possible explanatory variables and the dependent variable. The dependent variable was plotted against time to examine the fluctuations in the number of second lieutenants commissioned each year over the past 17 years.

B. EXPLANATORY VARIABLES

The following observations were made from the plots of the explanatory variables against time and against the dependent variable in Figures 1 through 11. Note there is a difference in the number of points between the time plots and the scatter plots due to the lag associated with each explanatory variable. A lag is defined as the amount of time between when the explanatory variable occurred and when its effect was felt by the dependent variable. For example, the number of MSI cadets enrolled in a particular year affects the number of second lieutenants commissioned four years later. A lag of four years is applied since commissioning occurs four years following enrollment as an MSI cadet. The plot of MSI cadets versus time from 1966 to 1986 consists of 21 data points. This lag of four years reduces the number of data points for the scatter plot by four from 21 to 17 points. Similarly, lags were applied to each of the proposed explanatory variables. The lags chosen for each variable and how they were obtained will be discussed further in Chapter III.

1. Army ROTC Enrollment

Figure 1 on page 6 depicts the enrollment of freshmen as Military Science I (MSI) cadets from 1966 through 1986. The first plot shows the trend in enrollment of MSI cadets versus time while the scatter plot shows the number of cadets who enrolled in MSI versus the number of second lieutenants commissioned four years later. Note that the scatter plot has four fewer points than the time plot due to the four year lag.

5
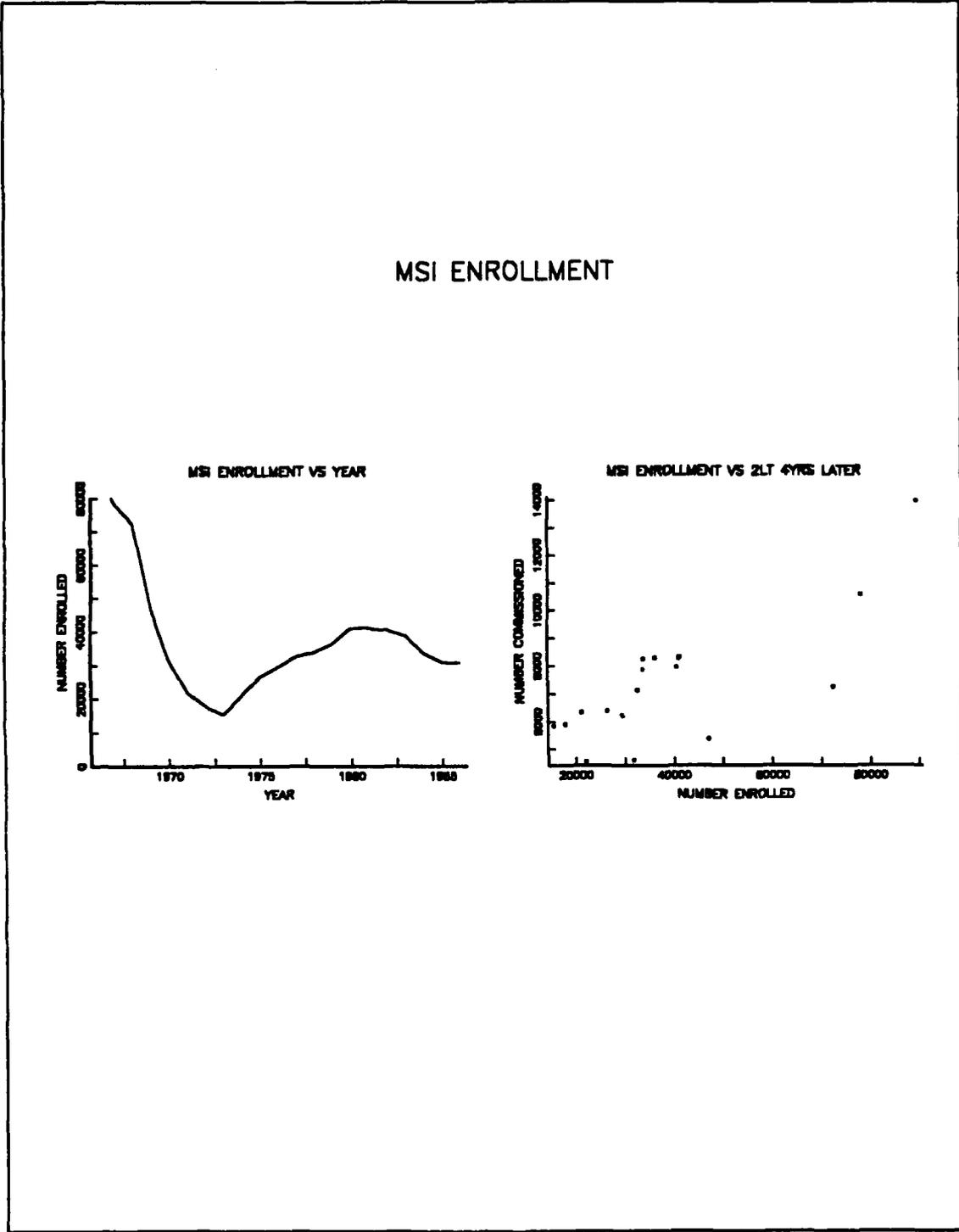
MSI ENROLLMENT

Figure 1.   MSI Enrollment

MSII ENROLLMENT
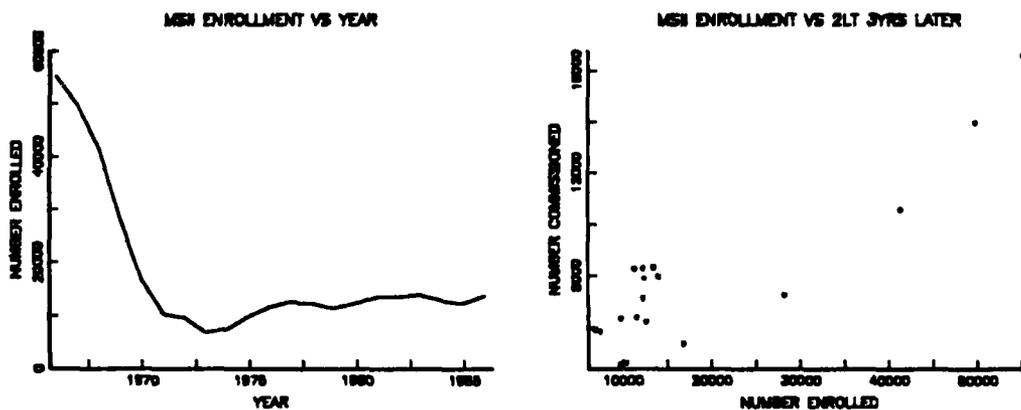
Figure 2.   MSII Enrollment

MSIII ENROLLMENT

MSIII ENROLLMENT VS YEAR

MSIII ENROLLMENT VS 2LT 2YRS LATER
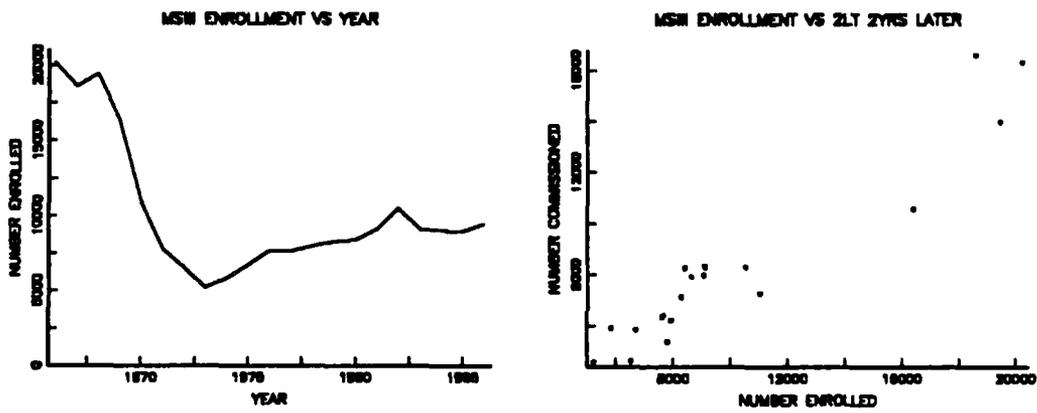
Figure 3.   MSIII Enrollment

8

MSIV ENROLLMENT

MSIV ENROLLMENT VS YEAR

**MSIV ENROLLMENT VS 2LT 1YR LATER**
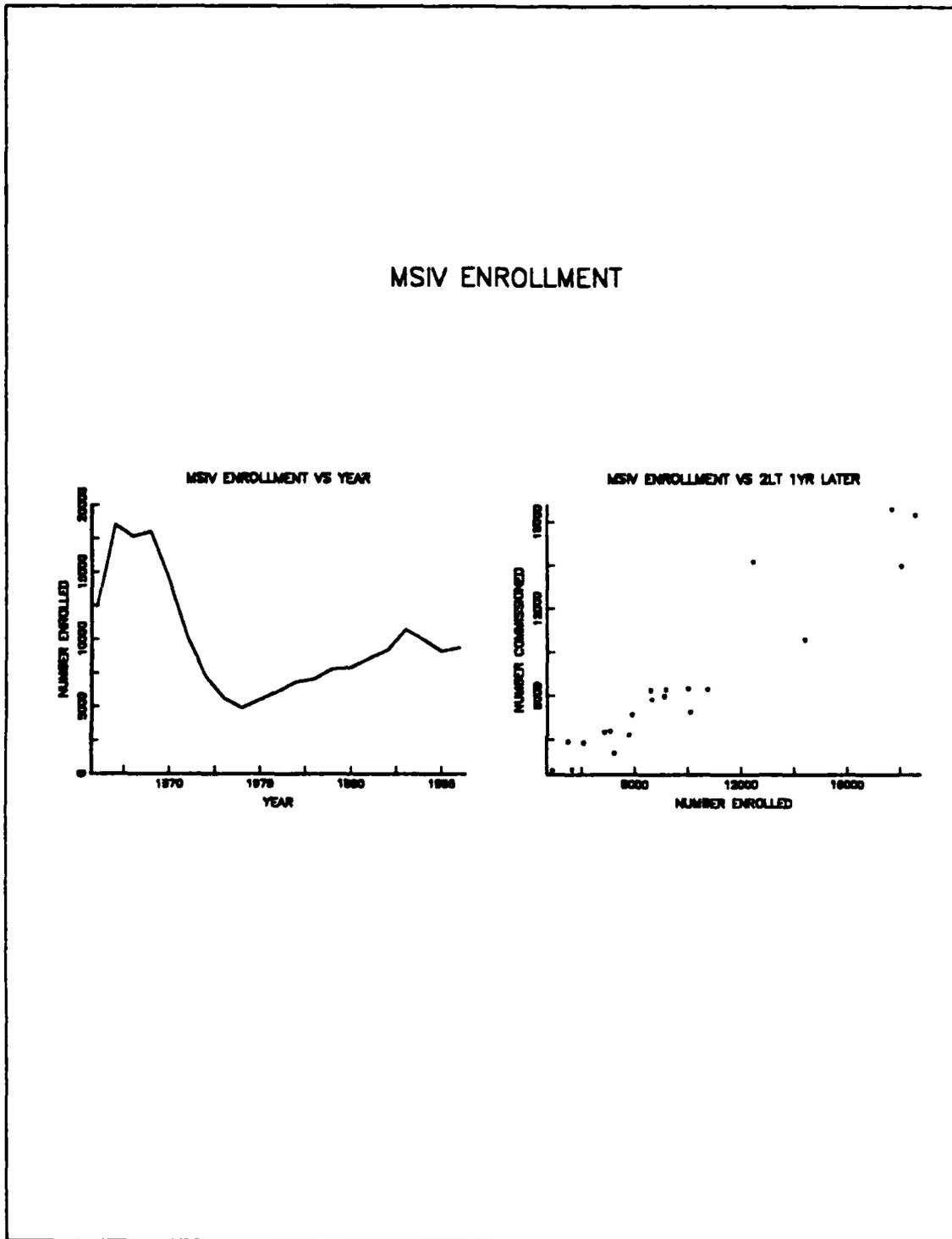
Figure 4.   MSIV Enrollment

9

The peak year in enrollment for the Army ROTC program was 1966 and is the starting point for looking at the trend in enrollment over time for MSI cadets. As the plot of MSI enrollment versus time shows, enrollment was at a high in 1966 and steadily decreased through 1973. This highpoint in 1966 represents the build up for the Vietnam conflict and the subsequent lowpoint in 1973 corresponds to the withdrawal of the United States forces from Vietnam when the Army no longer had a need for a large officer corps. The three rightmost points on the scatter plot correspond to enrolling in the Army ROTC program as an MSI cadet in 1966, 1967 and 1968 and being commissioned in 1970, 1971, and 1972. These outliers are a result of the build up for the Vietnam conflict and are not representative of the current Army ROTC program. Following the Vietnam conflict in 1973, enrollment slowly began increasing taking on its current form with an average enrollment of 33,585 cadets in MSI during the last four years. As Figure 1 shows, historical data prior to 1974 reflects the growth of the officer corps for a wartime situation and does not accurately represent the current peacetime Army ROTC program. Use of this data would bias any attempts at model development. Therefore it was decided not to use data prior to 1974 in model development. This decision reduced the number of data points available for regression analysis thus limiting the variety of possible models to investigate this problem. While this loss of data restricted modeling efforts, it allowed for developing models which more accurately represent the current Army ROTC program. Figures 2, 3, and 4 provide similar results for MSII, MSIII, and MSIV respectively. All three figures show enrollment at a high point during the Vietnam era and decreasing as the conflict ended. Disregarding these early points, all four scatter plots indicate there may be a relationship between the enrollment data and the number of second lieutenants commissioned.

Examining the attrition experienced by the Army ROTC program within each year group provided some interesting insight into the retention phase of the commissioning process. A year group is defined as the year in which an individual is commissioned. It is composed from the four years of Army ROTC enrollment in MSI through MSIV prior to commissioning For example, year group 1980 was composed from MSI cadets in 1976, MSII cadets in 1977, MSIII cadets in 1978, and MSIV cadets in 1979 leading to commissioning in 1980. Figure 5 on page 11 is a plot of all four Military Science years and the number of second lieutenants commissioned by year group. The differences in height between any two lines for any year group represents the number of cadets who left the program between those particular years. For example, the

ROTC ATTRITION BY YEAR GROUP
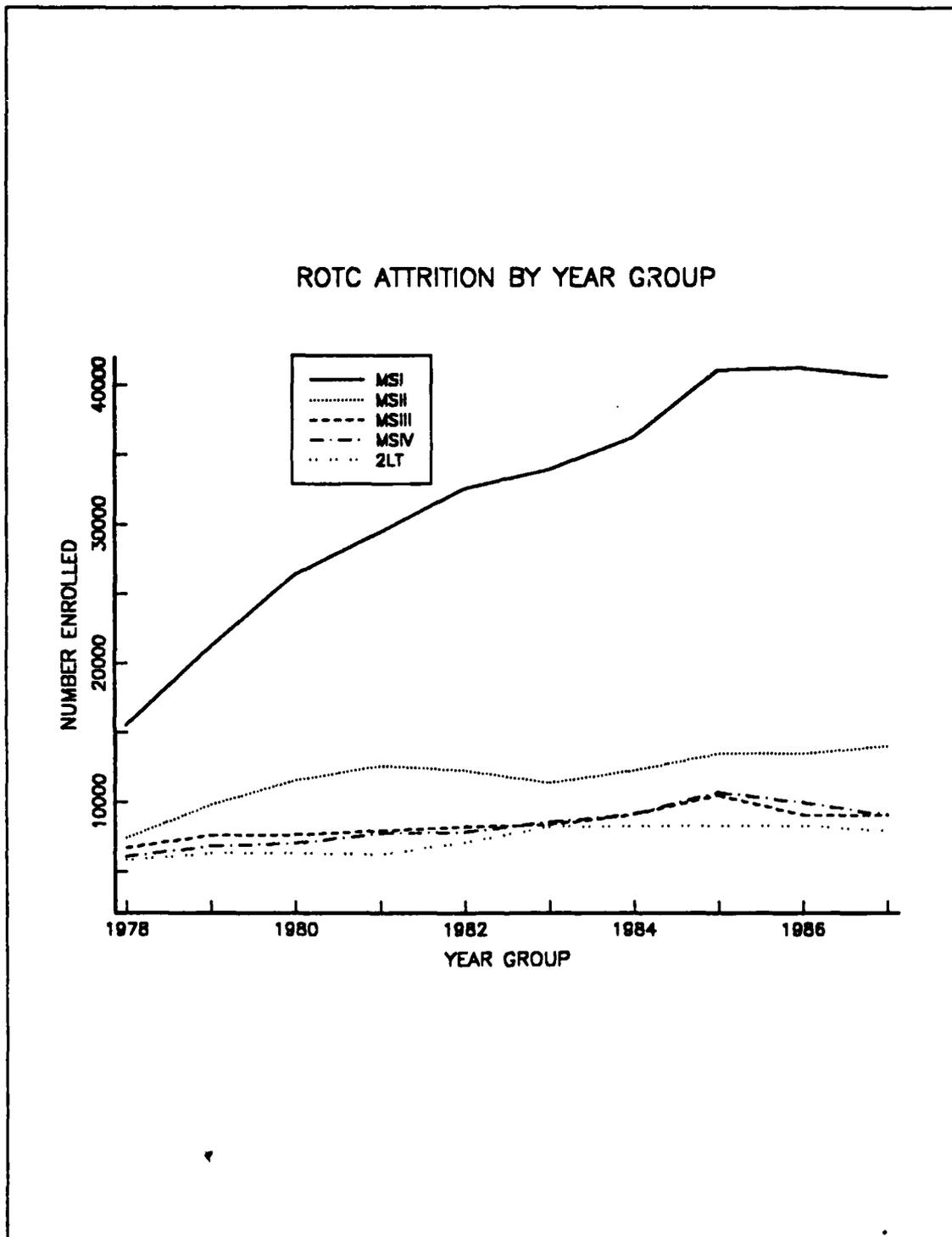
Figure 5.   Army ROTC Attrition

difference between the MSI line and the MSII line for year group 1980 is 16,587. This is the number of MSI cadets who were in the Army ROTC program in 1976 but did not enroll in the program as MSII cadets in 1977. While attrition has remained fairly constant between the MSII, MSIII, MSIV, and commissioning years, it has risen steadily for MSI cadets between their freshmen and sophomore years. From year group 1978 through year group 1986, enrollment of MSI cadets has increased while the ability to retain these cadets and have them continue in the Army ROTC program as MSII cadets has decreased. The large dropout of cadets from MSI to MSII may reflect the normal attrition of college students between their freshmen and sophomore years. However, it would seem this should be a fairly constant attrition rate and not increasing as it is for the Army ROTC program. This inability to retain cadets following their MSI year may indicate the Army ROTC program is attractive to students initially but for some reason is not appealing as a long term program.

2. Scholarship Program

The Army ROTC scholarship program plays an important role in the recruiting and retaining phases of the commissioning process. The scholarship program was initiated in 1964 with a total of 5,500 scholarships per year being allocated to pay for tuition and fees. The number of scholarships was later increased to 6,500 per year in 1970 and again increased in 1983 to its current level of 12,000 per year. A comparison between the Army ROTC scholarship program and the Air Force and Navy ROTC programs reveals a disparity in the number of scholarships allocated to each service. In 1984, the Army ROTC program had 67,727 cadets enrolled with 12,000 cadets receiving scholarships. By comparison, the Air Force ROTC program in 1984 had 24,883 cadets enrolled with 7,500 receiving scholarships and the Navy ROTC program had 10,920 cadets enrolled with 9,500 cadets receiving scholarships. While 18% of the cadets enrolled in Army ROTC in 1984 attended college on some sort of Army ROTC scholarship, 30% of Air Force ROTC cadets and 87% of Navy ROTC cadets received some sort of scholarship [Ref. 1: p. 4-9]. These percentages are significantly different and indicate the Army ROTC program may not be receiving a fair share of the number of ROTC scholarships allotted by Congress. The Army ROTC program produces the largest number of second lieutenants per year of any service while having the lowest percentage of cadets on scholarship.

The scholarship data provided little use from a modeling standpoint due to the fact that the maximum number of scholarships allowed were used each year. This made

the data of little use for predictive purposes. An analysis of the scholarship data did provide some interesting insights into the retention problem experienced by the Army ROTC program. Table 1 below depicts the historical retention rates from 1974 through 1986 for scholarship and nonscholarship cadets.

Table 1. **PERCENTAGE OF ARMY ROTC RETENTION:** Scholarship versus Nonscholarship Cadets

	Scholarship Program			Nonscholarship
	4 year	3 year	2 year	
MSI to MSII	91.2%			33.0%
MSII to MSIII	73.3%	89.2%		32.0%
MSIII to MSIV	91.6%	92.7%	94.2%	88.0%
MSIV to 2LT	94.4%	94.0%	94.0%	92.0%
Enrollment to 2LT	57.0%	78.0%	88.5%	8.0%

This table indicates that scholarships greatly increase retention in the Army ROTC program. A combination of the financial assistance provided by the scholarship and the military obligation incurred for failing to complete the program, lead to a greater likelihood of staying in the Army ROTC program. The low retention rate of the nonscholarship cadets may be the result of many reasons such as medical disqualification, withdrawal from school, academic or Army ROTC failure, lack of interest in the Army ROTC program, or any number of other reasons. While scholarship cadets face many of those same problems, it appears the screening and qualifications required to win a scholarship generally results in the selection of a quality cadet who stays in the program.

The retention rates from the MSIII or junior year on through commissioning as a second lieutenant for the scholarship programs are 86.5%, 87.1%, and 88.5% for the four, three and two year scholarship program respectively The retention rate for the nonscholarship cadets from their MSIII year through commissioning is 81.0%. These rates were obtained by multiplying the retention rates in Table 1 between the MSIII to MSIV year and the MSIV to commissioning year. These retention rates indicate that from the junior year on, retention among the the scholarship programs is almost equal across the board and retention among the non-scholarship cadets is only slightly lower than retention among the scholarship cadets. From this observation, it appears that

13

retaining cadets from their sophomore year to their junior year, which corresponds to the time when all cadets become obligated, is a key point in the commissioning process.

Among the scholarship programs, the two year program is the least expensive and has the highest over all retention rate followed by the three year program. The high overall retention rates associated with the two and three year scholarship programs appear to be the result of the military obligation incurred immediately upon receiving a scholarship. These programs have no attrition accounted for between the MSI and MSII years which is when the largest amount of attrition has historically occurred. Prior to 1984, four year scholarship cadets did not incur any obligation until the start of their junior year. Since 1984, four year scholarship recipients have incurred a military obligation upon entering their sophomore year. This may explain why the average historical retention rate for four year scholarship cadets was low between their sophomore and junior years (MSII to MSIII). Since 1984, the retention rate for four year scholarship students between their MSII and MSIII year has increased to 94% while retention between their MSI and MSII year has decreased to 85%. During this same time frame, retention in the two year scholarship program between the MSIII and MSIV years has remained at 94%. This indicates that once a cadet incurs a military obligation, the retention rate is almost the same regardless of which scholarship program the individual is in. While the two year program is the most cost effective scholarship program in terms of dollars spent per cadet, the four year scholarship program is recognized as a valuable recruiting tool for the Army ROTC program which attracts quality high school students into the program. The value of the four year scholarship program can not be measured in terms of cost alone.

3. Advertising Budget and Leads

The Army ROTC advertisement program has a significant influence on the recruiting phase of the commissioning process. It is the primary method of attracting individuals to the Army ROTC progam. All advertising is currently conducted by mail, public service announcements, or print advertising. There are plans for a television advertising campaign in the future. Figure 6 shows how significantly the advertising budget has increased from 1981 through 1986. Figure 7 shows how the number of leads resulting from the advertising has decreased. Recall from Chapter 1 that a lead is defined as a response to an advertisement such as mailing in a postcard. A measure of the cost effectiveness of the advertising program is the amount in dollars spent on advertising each year divided by the number of leads produced. Table 2 shows how the cost

14

ADVERTISING BUDGET

ADVERTISING BUDGET VS YEAR

ADVERTISING BUDGET VS 2LT THREE YRS LATER
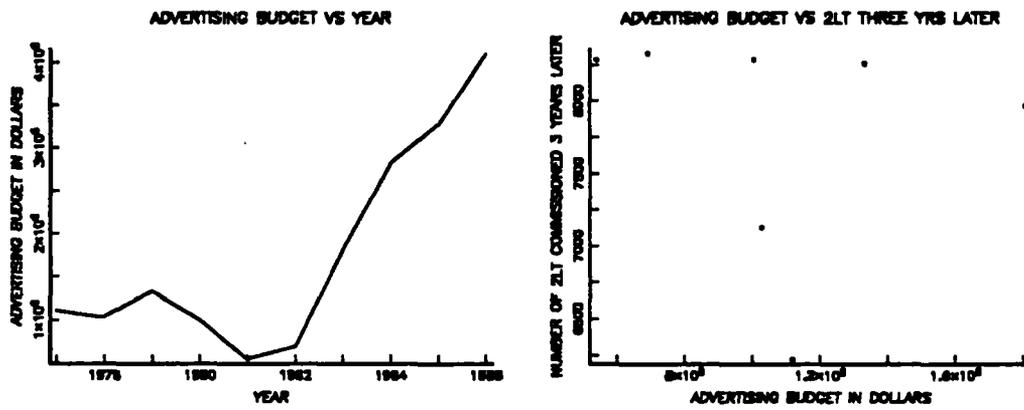
Figure 6.   Annual Advertising Budget

LEADS FROM ADVERTISING

NUMBER OF ADVERTISING LEADS VERSUS YEAR

NUMBER OF LEADS

YEAR

NUMBER LEADS VS NUMBER OF 2LT THREE YEARS LATER

NUMBER OF 2LT COMMISSIONED 3 YEARS LATER

NUMBER OF LEADS

Figure 7.   Leads

16
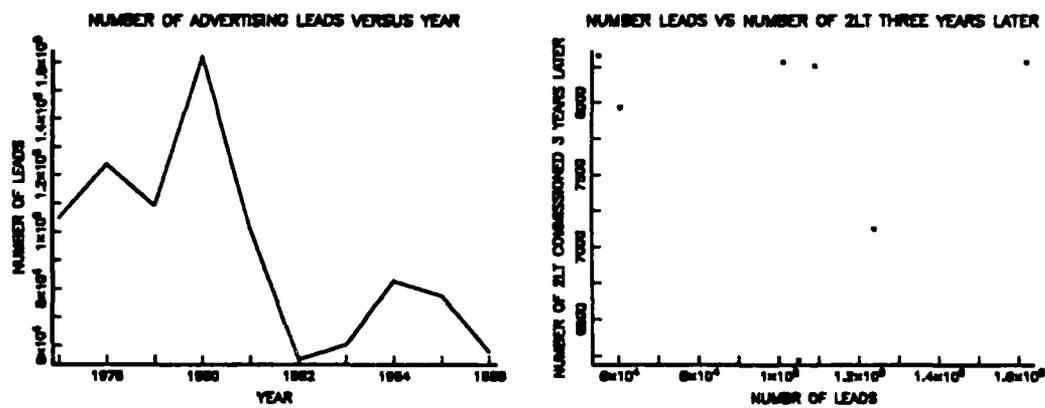
effectiveness of the advertising program has deteriorated over the past decade. This may be the result of cost increases in the advertising industry and inflation. However, to go from a cost effectiveness ratio of $5.33 per lead in 1981 to $70.90 per lead in 1986 indicates the advertisements are not reaching their target audience or are of poor quality.

The scatter plots for advertising budget and leads do not appear to indicate any relationships exist between these variables and the dependent variable. The limited number of data points makes it difficult to draw any conclusions from the scatter plots. This data would be more useful for studying the effectiveness of the advertising program if each of the leads were correlated with whether or not the individual enrolled in the Army ROTC program.

Table 2. ADVERTISING AND LEADS

Year	Advertising Budget in then year dollars	Number of Leads	Cost/Lead
1977	$1,118,460	104,875	$10.66
1978	$1,028,470	123,979	$8.30
1979	$1,332,478	109,094	$12.21
1980	$1,005,788	162,251	$6.20
1981	$539,843	101,278	$5.33
1982	$692,169	54,993	$12.59
1983	$1,808,384	60,293	$29.99
1984	$2,823,427	82,823	$34.09
1985	$3,270,304	77,805	$42.03
1986	$4,100,994	57,838	$70.90

4. Annual Unemployment Rate

The use of the unemployment rate represented an attempt to tie some economic indicator to the commissioning process. The United States Army ROTC Cadet Command recommended the use of the annual average unemployment rate, believing there should be some correlation between the unemployment rate and enrollment in the Army ROTC program. As the economy deteriorated, it was felt more students would

be interested in the military as a possible career and the Army ROTC scholarship program to assist in paying their college tuition. As seen in Figure 8, from 1970 through 1986, the unemployment rate has fluctuated between five and ten percent. The scatter plot revealed there may be some relationship between unemployment and the number of second lieutenants commissioned three years later.

5. Freshmen Enrollment

Figure 9 shows the trend in enrollment of college freshmen. Since 1980, the number of freshmen enrolling in college has begun to decline. This decline is expected to continue into the 1990's due to the projected reduction in the number of college age students. This reduction in the number of students may have a significant effect on the ability of the Army ROTC program to meet its projected demand. Colleges will be competing to attract quality students into their programs and industry will be competing for college graduates. It will become increasingly difficult to attract and retain high caliber students into the Army ROTC program. The scatter plot of freshmen enrollment reveals there may be some relationship with the number of second lieutenants commissioned four years later. It appears from the scatter plot, as freshmen enrollment increases so does the number of second lieutenants commissioned four years later.

6. Average Yearly College Tuition

Figure 10 depicts the rising cost of going to college over the last 16 years. During the 1980's, the cost of college tuition has experienced a significant growth. Between 1978 and 1984, the average yearly cost for tuition and fees has doubled. While there is some inflation built into these numbers, this still represents a significant growth. The scatter plot indicates that a relationship between these rising costs and the number of second lieutenants commissioned may exist. It appears that as college tuition goes up so does the number of second lieutenants commissioned three years later. This indicates that an interest may exist in the Army ROTC program as college costs rise.

7. Youth Attitude Tracking Survey

The use of the youth attitude tracking survey was an attempt to develop a relationship between propensity towards service among high school age students and the number of second lieutenants commissioned. The limited number of data points made it difficult to draw any valid conclusions about this relationship. Figure 11 shows that the youth attitude tracking survey in the mid to late 1970's was on the decline. Since 1979, the results of the survey have been rising indicating propensity toward military service among high school students has begun to increase. The scatter plot, however

18

ANNUAL UNEMPLOYMENT RATE

UNEMPLOYMENT RATE VS YEAR

UNEMPLOYMENT RATE VS NUMBER OF 2LT THREE YRS LATER
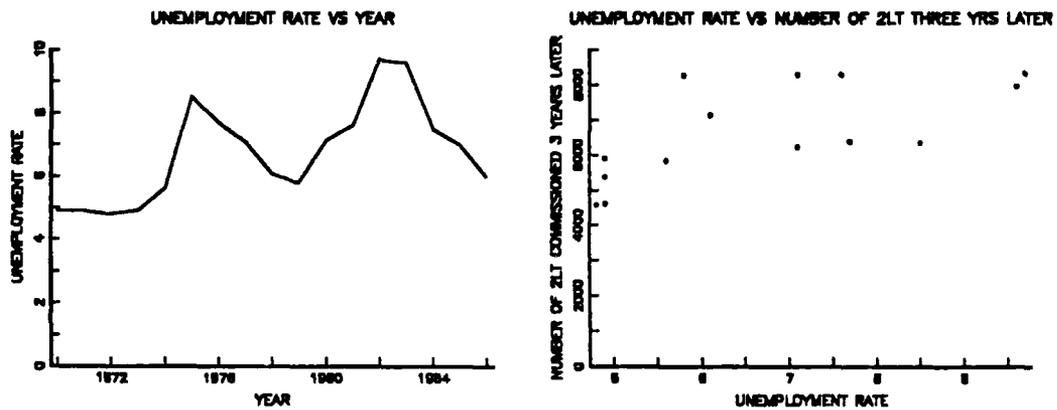
Figure 8.   Annual Unemployment Rate

19

FRESHMEN ENROLLMENT

NUMBER OF FRESHMEN ENROLLED VERSUS YEAR

FRESHMEN ENROLLED VS 2LT FOUR YEARS LATER
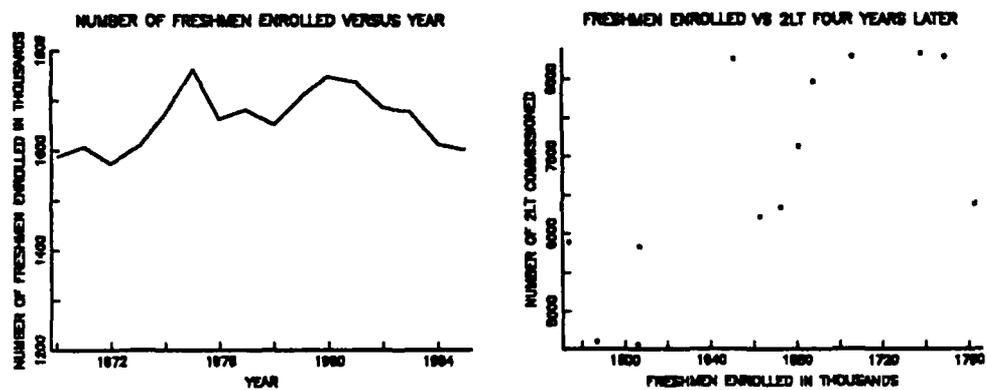
Figure 9.   Freshmen Enrollment

20

AVERAGE YEARLY COLLEGE TUITION
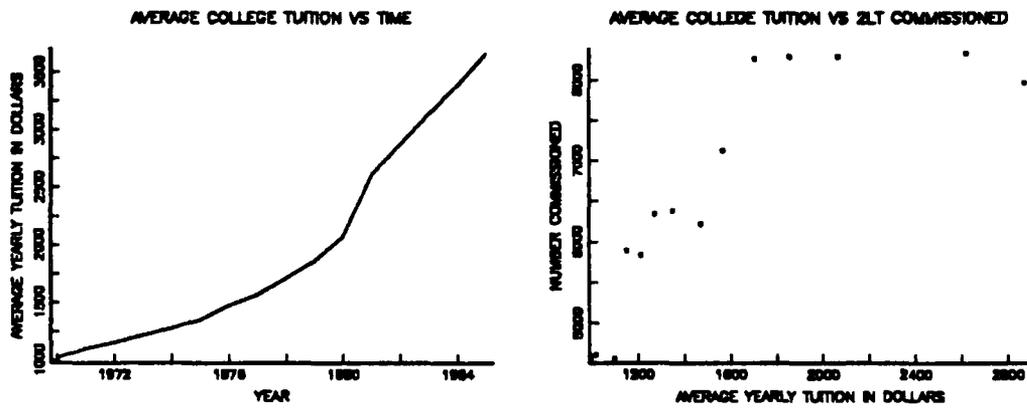
Figure 10.   Average Yearly College Cost

21

YOUTH ATTITUDE TRACKING SURVEY
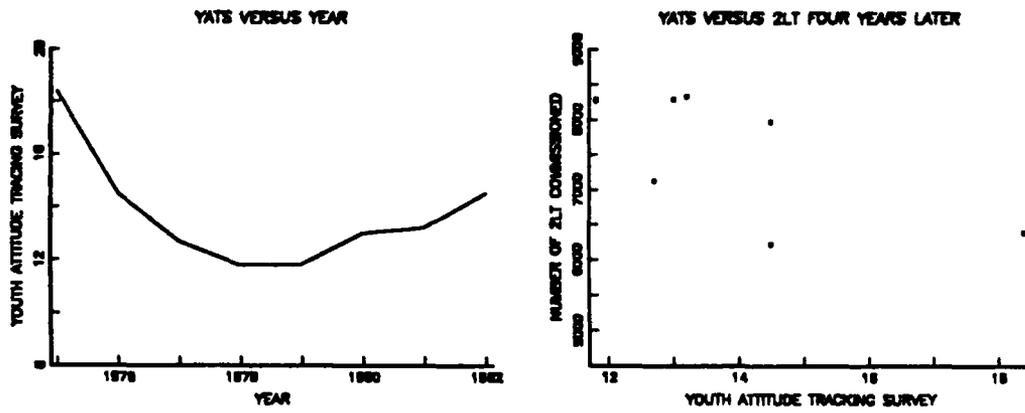
Figure 11.   Youth Attitude Tracking Survey

22

does not reveal any strong relationship between the survey data and the number of second lieutenants commissioned. One would expect that as the attitude toward military service improves, the number of second lieutenants commissioned each year would increase but this is not revealed in the scatter plot.

C. DEPENDENT VARIABLE

As seen in Figure 12, the number of second lieutenants commissioned per year declined as the end of the Vietnam conflict approached in 1973. Between 1975 and 1982 the number of second lieutenants being commissioned steadily increased. From 1982 until the present, the number being commissioned has leveled off and appears to be declining. This leveling off and subsequent decline appears to correlate with the decline in freshmen college enrollment as seen in Figure 9 on page 20. This decrease in the eligible population may be a significant factor in whether or not the United States Army ROTC Cadet Command will be able to meet its future commissioning goals.

NUMBER OF SECOND LIEUTENANTS COMMISSIONED VS YEAR
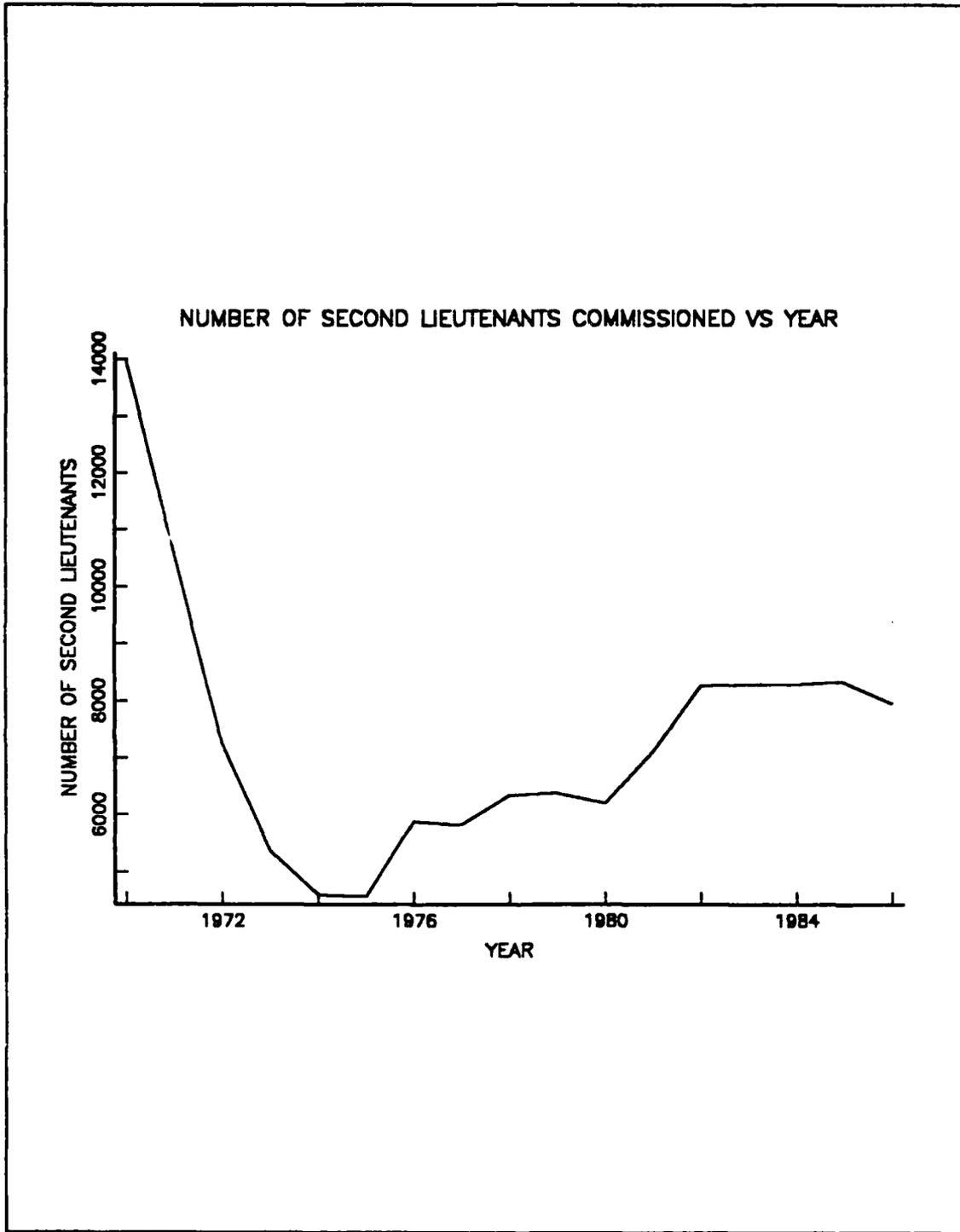
Figure 12.   Second lieutenants Commissioned each Year

24

III. MODEL DEVELOPMENT

A. METHODOLOGY

Least squares regression analysis was used to explore the relationships between the proposed explanatory variables and the dependent variable, the number of second lieutenants commissioned by the Army ROTC program in a year. The method of least squares estimates the coefficient for each variable, β , in the regression equation by minimizing the sums of squares of the residuals. Two single variable regression models were used to examine if a trend existed between a single explanatory variable and the dependent variable. Two multiple regression models were used to analyze the relationships between several different combinations of the explanatory variables and the dependent variable. All regression analysis was conducted using the computer software package GRAFSTAT.

Hypothesis testing was conducted to see how well each of the proposed models fit the data. A t test was conducted to obtain the level of significance for each coefficient estimated. Values for the level of significance can range between zero and one with a small value leading to the conclusion that the coefficient is significant and a large value (greater than .20) indicating that the coefficient is not significant. Similarly, an F test was conducted for each model to determine the overall model level of significance. Again, a small value for the level of significance would indicate the coefficients are significant and the model may have some value for predictive purposes and a large value would indicate the coefficients are not significant and the model has little value for predictive purposes. For a discussion of the underlying assumptions, derivations of the test statistics, and null and alternative hypotheses of the t test and the F test see DeGroot [Ref. 5: p. 617-623]. In addition to the t statistic and the F statistic, a value for the coefficient of determination, R^2, was calculated for each model. Values for R^2 can range from zero to one with higher values indicating a greater amount of variability is explained by the model. Typically, values of R^2 greater than .75 indicate a significant amount of the variabilty in the regression model is explained and that the model provides a good fit to the data. Depending on the data and the regression model, lower values of R^2 may also be acceptable. For each of the models, an example of how it may be used and a 95% prediction interval based on normal theory is provided [Ref. 6: p. 153]. This prediction interval represents the range within which the dependent variable should fall

25

95% of the time for the proposed values of the explanatory variable(s). For a discussion and an example of how these prediction intervals were developed see Appendix A.

All four of the proposed regression models have the standard normality assumptions associated with least squares regression. The observations of the dependent variable are assumed to be independent and normally distributed with a constant variance. These assumptions are checked in the analysis of the residuals using the Kolmogorov- Smirnov (K-S) test with a level of significance of .05. For a discussion of the Kolmogorov-Smirnov test see Degroot [Ref. 5: p. 554]. Regression analysis also assumes the error terms, ε , are independent of each other. To check this assumption, the Durbin-Watson test for serial correlation of the residuals with a level of significance of .05 was conducted. The null hypothesis associated with this test is the residuals are serially correlated. Rejecting this null hypothesis leads to the conclusion the residuals are not serially correlated and are therefore independent. There is also a region associated with this test which is inconclusive and the hypothesis can neither be accepted or rejected. For a complete discussion of the Durbin-Watson test see Johnston [Ref. 6: p. 251].

B. REGRESSION ANALYSIS

1. Model one: the simple linear regression model

Simple linear regression provides a method to examine the relationship between a dependent variable, Y , and a single predictor or explanatory variable, X . Recall from Chapter I, the dependent variable is the number of second lieutenants commissioned in a year by the Army ROTC program and the predictor variables are the proposed explanatory variables. A straight line relating these two variables can be described by the equation

$$Y = \beta_0 + \beta_1 X \tag{1}$$

In equation 1, β_0 is the intercept term corresponding to the value of Y when X equals zero. β_1 is the slope of the line which is defined as the rate of change in Y for a unit change in X . The natural predictor variables for this thesis lag behind the dependent variable by a number of years which will be called i . For example as discussed in Chapter II, the lag associated with the variable MSI is four meaning the effect from the predictor variable is experienced by the response variable four years later, i.e., the number of cadets enrolled in MSI in 1976 will affect the number of second lieutenants commissioned by the Army ROTC program four years later.

When fitting the data to equation 1, all of the points will not fall directly on the line. Hence, an error term, ε , must be introduced to account for this deviation from the line. Combining the residual term and the lag with equation 1 leads to the simple linear regression equation

$$Y_t = \beta_0 + \beta_1 X_{(t-l)} + \varepsilon_t \tag{2}$$

where t represents time or the year. This model allows for examining whether or not a linear relationship exists between the dependent variable and each of the individual explanatory variables. For example, this model can explore for trends between the unemployment rate and the number of second lieutenants commissioned. Each of the possible explanatory variables were fitted to this model with an appropriate lag depending on the particular variable. For example, for the MSII, MSIII, and MSIV enrollment data the lags used were three, two and one year(s) respectively. For the other variables which did not have a set lag as to when they affected the dependent variable, lags of two, three, and four years were applied and the model with the highest value of R^2 was accepted as being the best fit for that particular variable.

a. Model one results

Table 3 provides a summary of the best fit between each of the proposed explanatory variables and the dependent variable for the simple linear regression model. For the enrollment data MSI through MSIV, the enrollment of MSIV cadets had the best fit indicating it was the best predictor for this model. The value of R^2 , .86, indicates the variation explained by the regression is high and the significance level of .005 is extremely strong. As one might expect, the number of students enrolled one year prior to commissioning is the most accurate single variable linear predictor of the number of second lieutenants commissioned in a year. MSI and MSIII had fairly good values for R^2 (.54 and .63) and their coefficients were significant indicating they may be reasonable predictors also. The variable MSII enrollment had a very low value for R^2 , .14, indicating the model was a poor fit. No reason could be determined as to why MSII had such a poor fit other than it may be associated with the large amount of attrition experienced by the Army ROTC program between the MSI and MSII year. Of the remaining variables, tuition and unemployment rate had the highest values for R^2 , (.65 and .44) and were fairly significant indicating there may be a trend between the response variable and these predictor variables. The explanatory variables budget, leads and youth attitude tracking survey had extremely low values for R^2 and were not significant

27

indicating they had little use as predictors for this model. For each of these variables, the Durbin-Watson test at the .05 level of significance proved to be inconclusive. Scatter plots of the residuals and a fit to the normal cumulative distribution function for the variables MSIV, tuition, and unemployment rate with the Kolmogorov-Smirnov (K-S) bounds can be found in Appendix B. The limited number of residuals makes it difficult to draw any conclusions about the constant variance assumption. The points are within the K-S bounds for all of the variables indicating the residuals are consistent with the normal distribution.

Table 3. MODEL ONE RESULTS

Variable (Lag)	Coefficient	Estimate of the coefficient	t statistic	level of significance	R^2
MSI (4 years)	β_0	3362.9	3.39	.005	.54
	β_1	.11384	3.58	.005	
MSII (3 years)	β_0	4424.3	2.69	.019	.14
	β_1	.19418	1.40	.186	
MSIII (2 years)	β_0	1379.4	1.61	.151	.63
	β_1	.66278	4.73	.000	
MSIV (1 year)	β_0	1892.1	3.38	.005	.86
	β_1	.61898	9.40	.000	
Tution (3 years)	β_0	3642.4	5.07	.004	.65
	β_1	1.9154	4.60	.008	
Unemployment (3 years)	β_0	3096.1	2.60	.023	.44
	β_1	531.35	3.09	.009	
Budget (3 year)	β_0	808.21	8.24	.000	.02
	β_1	-.0003	-.33	.758	
Leads (3 years)	β_0	8018.6	7.59	.000	.01
	β_1	-.0023	-.24	.821	
YATS (3 years)	β_0	9968.5	4.20	.005	.17
	β_1	-186.8	-1.09	.320	

b. Model one use

The most significant results obtained using model one were with the variables MSIV and tuition. To use MSIV as a predictor, substitute the estimates in

Table 3 for β_0 and β_1 into equation 2 and replace the predictor variable X with MSIV with a lag of one year. This leads to the prediction equation

$$\hat{Y}_t = 1892.1 + .61898(MSIV_{t-1}). \tag{3}$$

Equation 3 is a predictor model for the number of second lieutenants commissioned each year as a function of the number of cadets enrolled in MSIV one year prior. To use this linear model as a predictor, a hypothesized value for MSIV would be substituted into equation 3. For example, if the number enrolled in MSIV were 10,000, the predicted value for the number of second lieutenants commissioned one year later would be 8,081 with a 95% prediction interval ranging from 6,700 to 9,462. Similarly, for the predictor tuition, substituting the results in Table 3 into equation 2 yields

$$\hat{Y}_t = 3642.4 + 1.9154(Tuition_{t-3}). \tag{4}$$

For a yearly tuition cost of \$4000, the predicted number of second lieutenants commissioned three years later would be 11,304 with a 95% prediction interval of 9,663 to 12,944. Similar equations could be developed from equation 2 and the results in Table 3 for other predictor variables.

2. Model two: the single variable distributed lag model

The simple linear regression model implies that Y depends on X at only one preceding point in time. This holds for variables such as enrollment data where the effect felt by the response variable is associated with a change in the predictor variable at a specified time, i.e., the number of second lieutenants commissioned in 1986 is a function of the number of cadets enrolled in MSIII in 1984 and not the number of cadets enrolled in MSIII in 1985. For some predictor variables, their influence over the response variable may be spread over a period of time. This leads to a more general approach which says Y depends on several previous X values. One way of representing this type of relationship is using a distributed lag model [Ref. 7: p. 160]. For example, the unemployment rate over a period of three years may affect the commissioning of second lieutenants in a specific year. To study this effect, the single variable distributed lag model was proposed. This model is represented by the nonlinear equation

$$Y_t = \beta_0 + \beta_1(X_{(t-2)} + \lambda X_{(t-3)} + \lambda^2 X_{(t-4)}) + \varepsilon_t \tag{5}$$

29

where Y is the response variable, X is the predictor variable, β_0 is the intercept term, β_1 is the coefficient associated with the predictor variable, ε is the residual term, t represents the time or year, and λ is a constant between zero and one.

This model assumes the predictor variable has a decreasing weighted effect on the response variable over a specified time period. This weighted effect is captured in the λ term which distributes the effect of the explanatory variable in a decreasing manner over a three year time period. Different values of λ yield different time profiles. For example, if λ were .5, this would mean the weights of the explanatory variables effect distributed over three years would be 1.0 two years prior, .5 three years prior, and $.5^2$ (.25) four years prior. Carrying this example further to one of the proposed explanatory variables, the number of second lieutenants commissioned in 1980 could be a weighted function of the unemployment rate in 1976 through 1978 with the unemployment rate for each year having weights of .25, .5, and 1.0 respectively rather than a function of the unemployment rate in just one of these years as proposed in model one. The selection of the appropriate λ is the constant which yields the greatest value of R^2, the coefficient of determination [Ref. 7: p. 164]. The estimation of the λ term lead to a loss of one degree of freedom which had a minimal effect on the level of significance obtained for this model during hypothesis testing. To transform the data for this model, an APL function was written that created a single vector which could then be used in a single variable regression in GRAFSTAT. An increment of .1 starting at .1 through .9 was used for the values of λ in searching for the highest value for R^2. This transformation of the proposed explanatory variables resulted in the loss of three degrees of freedom due to the reduction of the size of the data. The predictor variables examined for this distributed effect were unemployment rate and tuition. The limited number of data points prohibited the use of the variables advertising budget, leads, and youth attitude tracking survey for this model.

a. Model two results

Table 4 represents a summary of the variables unemployment and tuition using model two. The distributed lag model for the variable unemployment showed an increase in the value of R^2 over model one, the simple linear regression model, from .44 to .53. The level of significance for the variable unemployment decreased slightly from .009 to .005 indicating the distributed lag model was slightly more significant than model one. This indicates the effects of unemployment may occur over a time period rather than at one particular discrete time. The variable average yearly college tuition did not

change in the value of R^2 (.65) when compared with the results from model one. The level of significance using model two decreased slightly from .008 to .002 indicating the estimated coefficient for the distributed lag model was slightly more significant for model two. These results indicate the simple linear model and the distributed lag model were of equal value for predicting the number of second lieutenants commissioned in a year for the independent variable college tuition. The Durbin-Watson statistics for the unemployment model (.94) and for the tuition model (.88) were inconclusive. Scatter plots of the residuals and K-S bounds for the normal cumulative distribution function are in Appendix B. Again, the limited number of data points makes it difficult to draw any conclusions about the variance of the residuals. The points are within the K-S bounds indicating the residuals are consistent with the normal distribution.

Table 4. MODEL TWO RESULTS

Variable (λ)	Coefficient	Estimate of the coefficient	t statistic	level of significance	R^2
Unem- ployment (.9)	β_0	1902.3	1.36	.205	.53
	β_1	264.33	3.51	.005	
Tuition (.9)	β_0	3924.1	5.30	.000	.65
	β_1	.66554	4.30	.002	

b. Model two use

For model two, the most significant results were obtained using the variable average yearly college tuition with a λ of .9 . Substituting the values obtained in Table 4 into equation 5 yields

$$\hat{Y}_t = 3924.1 + .66554(Tuition_{t-2} + .9\ Tuition_{t-3} + .81\ Tuition_{t-4}) \qquad (6)$$

For tuition costs of $3600, $3800, and $4000 for successive years, the predicted value for the number of second lieutenants commissioned in a year is 10,803 with a prediction interval from 9,591 to 12,014.

3. Model three: the general linear model

The general linear model is a multiple regression model in which several predictor or explanatory variables, $X_1, X_2, ..., X_p$, are used to model a single dependent variable, Y . For our problem, the number of predictor variables, p , was limited to four

explanatory variables due to the limited number of data points and the loss of one degrees of freedom for each coefficient estimated. A multiple linear regression equation that expresses the dependent variable as a linear funtion of several predictor variables is

$$Y_t = \beta_0 + \beta_1 X_{1(t-i)} + \beta_2 X_{2(t-i)} + \beta_3 X_{3(t-i)} + \beta_4 X_{4(t-i)} + \varepsilon_t \qquad (7)$$

where, as in the previous models, the β's are the unknown parameters to be estimated, t represents the time or year, i is the lag associated with each predictor variable, and ε is the error term. This general linear model allows for examining how linear combinations of the lagged explanatory variables affect the number of second lieutenants commissioned in a year. For example, equation 7 can be used to investigate what, if any, effect the explanatory variables MSI enrollment, unemployment rate, freshmen college enrollment, and average college tuition have on the commissioning of second lieutenants. The lag chosen for each model was based on the enrollment data used for each regression, i.e., a lag of two years was used if MSIII enrollment data was used in the model. Enrollment data was included in each combination to establish the lag. All possible combinations of military science enrollment data (MSI through MSIV) with three explanatory variables were analyzed using this model.

The stepwise algorithm used to obtain the best model was to first fit all four variables to the model. If the levels of significance for each variable were less than .20 and the model level of significance was below .10, the model was accepted as being significant. If the model was not accepted, the least significant variable, i.e., the variable with the highest value for the level of signicicance, was removed and the regression on Y was recalculated using the remaining variables. This was repeated until all the estimated coefficients had a level of significance of less than .20 and the model had a level of significance of less than .10 . The selection of .20 and .10 for stopping points for the levels of significance was a subjective decision based on experience gained during model development. The models selected as being the best fit were those which, after completing the algorithm, had the highest value for R^2.

a. Model three results

For model three with a four year lag, Table 5 represents the first step of the algorithm and Table 6 is the the final result for the best model fit. Table 5 has an extremely high value for the level of significance of the model (.008) and a high value for R^2 (.76) indicating a good fit of the data to the model. However, the level of significance for the coefficient MSI (.8419) indicated this variable had very little use in this model.

32

The regression model was run again this time removing the variable MSI and the results are shown in Table 6. The increased R^2 (.78) and levels of significance for the model indicated a better fit. This was the best model obtained for the general linear model with a four year lag. The Durbin-Watson statistic of 1.78 indicated the residuals were not serially correlated. A scatter plot and fit of the residuals to the normal cumulative distribution function can be found in Appendix B. The residuals are within the K-S bounds and they are randomly scattered indicating the residuals are consistent with the normal distribution.

Table 5. MODEL THREE RESULTS WITH FOUR YEAR LAG

Variable	Estimate (level of significance)	F Statistic for Model	Level of Significance for Model	R^2
Intercept β_0	-12373 (.1733)	7.97	.008	.76
MSI β_1	.0090 (.8419)			
Unemployment Rate β_2	-357.0 (.1844)			
Avg College Tuition β_3	1.797 (.0462)			
Freshmen enrolled β_4	10.94 (.0800)			

33

Table 6. MODEL THREE RESULTS, FOUR YEAR LAG MSI REMOVED

Variable	Estimate (level of significance)	F Statistic for Model	Level of Significance for Model	R^2
Intercept β_0	-12803 (.1250)	10.87	.002	.78
Unemployment Rate β_1	-382.7 (.1503)			
Avg College Tuition β_2	1.902 (.0065)			
Freshmen enrolled β_3	11.29 (.0522)			

Table 7 represents the best results of the stepwise regression using a two year lag. The levels of significance for the coefficients for both variables, MSIII enrollment and average college tuition, are less than .20 and the model level of significance of .000 is outstanding. The R^2 value of .84 indicates a large amount of the variablity in the model is explained. The Durbin-Watson statisic of 1.89 indicated the residuals were not serially correlated. A scatter plot and fit of the residuals to the normal cumulative distribution function can be found in Appendix B. The residuals are within the K-S bounds and they are randomly scattered indicating the residuals are consistent with the normal distribution.

Table 7. MODEL THREE RESULTS, TWO YEAR LAG

Variable	Estimate (level of significance)	F Statistic for Model	Level of Significance for Model	R^2
Intercept β_0	1559 (.0699)	32.02	.000	.84
MSIII β_1	.4328 (.0021)	.		
Avg College Cost β_2	.9008 (.0018)			

34

A final regression using model three was run for the enrollment data MSI through MSIV. This was done to investigate how military science enrollment alone across all grades affected commissioning. The lag for this model was varied depending on the variable. Table 8 represents the first step of the regression of the MS enrollment data on the dependent variable. While the value of R^2 (.91) and the model level of significance of .000 are extremely strong, the level of significance associated with the MSI and MSIII coefficients (.85 and .63 respectively) are high indicating they are not significant. Table 9 represents the final step of the regression with the variables MSI and MSIII enrollment removed. The value of R^2 (.90) and all the levels of significance are high indicating enrollment in the sophomore and senior year have some value in predicting commissioning of second lieutenants. The Durbin-Watson statistic for Table 9 was 1.34 which was inconclusive. A scatter plot and fit of the residuals to the normal distribution can be found in Appendix B. These plots lead to the conclusion the residuals are consistent with the normal distribution.

Table 8. MODEL THREE RESULTS, ROTC ENROLLMENT DATA

Variable (lag)	Estimate (level of significance)	F Statistic for Model	Level of Significance for Model	R^2
Intercept β_0	2392 (.0691)	19.48	.000	.91
MSI Enrollment β_1 (four year lag)	-.0098 (.8464)			
MSII Enrollment β_2 (three year lag)	-.1583 (.3933)			
MSIII Enrollment β_3 (two year lag)	-.2463 (.6306)			
MSIV Enrollment β_4 (one year lag)	1.0984 (.0332)			

Table 9. MODEL THREE RESULTS, ROTC ENROLLMENT DATA MSII AND MSIV

Variable (lag)	Estimate (level of significance)	F Statistic for Model	Level of Significance for Model	R^2
Intercept β_0	2030 (.0121)	47.05	.000	.90
MSII Enrollment β_1 (three year lag)	-.1903 (.1026)			
MSIV Enrollment β_2 (one year lag)	.8926 (.0000)			

The negative coefficient associated with the MSII variable in Table 9 indicates a negative partial correlation exists with the dependent variable. This implies that if the number of cadets enrolled in MSII increased, the number of second lieutenants commissioned three years later would decrease. This does not make sense intuitively. An additional regression was conducted replacing the variable MSII with the variable MSIII enrollment in an attempt to develop a model which did not have a negative coefficient. For the regression of the variables MSIII and MSIV on the dependent variable, the coefficient for MSIII was negative (-.7376) and the value of R^2 (.84) was lower than the previous model. This was not an improvement and lead to the acceptance of the regression of MSII and MSIV enrollment on the number of second lieutenants commissioned as being the best fit for enrollment data. Caution should be used with this model due to the negative coefficient.

The results from model three indicate unemployment rate, college tuition and enrollment in the Army ROTC program have some value for predicting the number of second lieutenants commissioned each year. The two year lag model had higher values for R^2 and for the levels of significance than the four year lag model. This indicates the closer to commissioning, the better the model will fit. The four year lag model may be useful in explaining the retention of MSI cadets in the Army ROTC program based on the cost of college and the unemployment rate. This indicates that economics play a key role in retaining individuals.

b. Model three use

The best results obtained using model three with a combination of variables are those in Table 7. This model has a two year lag with the variables MSIII enrollment, and average yearly college tuition. For model three, the equation obtained was

$$\hat{Y}_t = 1559 + .4328\,(MSIII_{t-2}) + .9008\,(Tuition_{t-2}). \tag{8}$$

For MSIII enrollment of 9,500, and tuition of $4000, the predicted number of second lieutenants commissioned two years later is 9,274 with a 95% prediction interval from 8,109 to 10,439. For the Military Science enrollment data, the best results using model three were a combination of MSII and MSIV data. For this regression model the equation obtained was

$$\hat{Y}_t = 2023 - .1903\,(MSII_{t-3}) + .8926\,(MSIV_{t-1}). \tag{9}$$

For values of 13,000 and 9,400 for MSII and MSIV respectively, the predicted value for the number of second lieutenants commissioned each year is 7940 with a 95% prediction interval of 7,029 to 8,852.

4. Model four: the general distributed lag model

The fourth regression model used was a general distributed lag model. This multiple regression model was a combination of model two, the single variable distributed lag model, and model three, the general linear model. It assumes the dependent variable is a function of three explanatory variables, $X_1, X_2,$ and X_3, where the first variable has a fixed lag and the second and third variables have a distributed lag. The fixed lag variable, X_1, for our problem represents military science enrollment while the distributed lag variables X_2 and X_3 represent unemployment rate and average college tuition. This model can be specified by the equation

$$Y_t = \beta_0 + \beta_1 X_{1(t-i)} + \beta_2(X_{2(t-2)} + \lambda_2 X_{2(t-3)} + \lambda_2^2 X_{2(t-4)}) \tag{10}$$

$$+ \beta_3(X_{3(t-2)} + \lambda_3 X_{3(t-3)} + \lambda_3^2 X_{3(t-4)}) + \varepsilon_t$$

where the β and λ parameters are estimated coefficients as discussed in the previous models. To determine the best fit for this model a stepwise algorithm similar to the one used in model three was developed with the additional step of incrementing the λ terms by .10 from .10 to .90. The first step was to fit all three variables to the model with λ_2

and λ_3 both assigned a value of .10. The least significant variable, i.e., the variable having a level of significance greater than .20, was removed and the regression on the dependent variable was recalculated. This was repeated until all the estimated coefficients had a significance level of less than .20 and the model level of significance was less than .10. The resulting R^2 was recorded and then the value of λ_2 was incremented by .1 while holding λ_3 constant. The stepwise algorithm was repeated until λ_2 equaled .90. Then the value for λ_3 was incremented by .10, the value for λ_2 was reset to .10 and the algorithm was repeated. The results were then compared to see which combinations of λ 's provided the model with acceptable levels of significance and the highest value of R^2.

a. Model four results

Tables 10 and 11 represent a summary of the best stepwise regression obtained using model four. For Table 10, the R^2 value of .90 and the model level of significance of .001 were extremely good. However, the level of significance for MSI enrollment, β_1, of .8341 was high. Removing MSI enrollment from the model produced the results in Table 11. The R^2 value of .90 and the level of significance of .000 indicate a good model fit. The level of significance for the individual coefficients improved to an acceptable level. The values of λ_2 and λ_3 were both .90 as in model two which may or may not have some significance. As in model three, this model indicates average college tuition and unemployment rate over a period of time may have some value in predicting the number of second lieutenants commissioned in a year. While the numerical results obtained by model four are slightly better then those obtained using model three, the introduction of the λ term adds an additional variable which must be estimated. The Durbin-Watson statistic for Table 11 was 1.78. This value lead to the conclusion the residuals were not serially correlated. Scatter plots and a fit of the residuals to the normal distribution with K-S bounds can be found in Appendix B. These plots lead to the conclusion the residuals were normally distributed with a constant variance thus validating the assumptions of the regression model.

Table 10. MODEL FOUR RESULTS

Variable	Estimate (level of significance)	F Statistic for Model	Level of Significance for Model	R^2
Intercept β_0	2016 (.3769)	24.09	.001	.90
MSI β_1	-.0123 (.8341)			
Unemployment Rate β_2 ($\lambda = .9$)	100.9 (.4533)			
Avg College Tuition β_3 ($\lambda = .9$)	.5511 (.1541)			

Table 11. MODEL FOUR RESULTS WITH MSI REMOVED

Variable	Estimate (level of significance)	F Statistic for Model	Level of Significance for Model	R^2
Intercept β_0	1613 (.1501)	40.41	.000	.90
Avg College Tuition β_1 ($\lambda = .9$)	.4798 (.0021)			
Unemployment Rate β_2 ($\lambda = .9$)	123.2 (.1231)			

b. Model four use

For model four, the best results were obtained using average yearly college tuition and the unemployment rate. Substituting the results in Table 11 into equation 10 yields

$$\hat{Y}_t = 1613 + .4798(Tuition_{t-2} + .9\ Tuition_{t-3} + .81\ Tuition_{t-4}) \tag{11}$$

$$+ 123.2\ (Unemployment_{t-2} + .9\ Unemployment_{t-3} + .81\ Unemployment_{t-4}).$$

For successive tuition costs of $3600, $3800, and $4000 and successive unemployment rates of 6%, 7%, and 8%, the predicted value for the number of second lieutenants commissioned each year is 8,929 with a prediction interval of 7,637 to 10,221.

39

C. MODEL DEVELOPMENT CONCLUSIONS

From model development, it can be concluded that of the variables examined, unemployment rate, average yearly college tuition and ROTC enrollment may have some value in predicting the dependent variable , the number of second lieutenants commissioned each year by the ROTC program. Based on the modeling done, the other proposed variables had little or no value in predicting the dependent variable. The single variable regression models indicated a trend exists between the dependent variable and the explanatory variables unemployment rate, average college tuition, and MSI, MSIII, and MSIV enrollment. The best single variable predictor was MSIV enrollment using the estimates obtained by the simple linear regression model. This single variable prediction model is represented by equation 3. The best multiple variable predictors were the combination of average college tuition and unemployment rate using the general distributed lag model. This prediction model is represented by equation 11.

IV. CONCLUSIONS, OBSERVATIONS AND RECOMMENDATIONS

A. CONCLUSIONS

The exploratory data analysis and the regression analysis indicated that of the variables examined, unemployment rate, average yearly college tuition, and Army ROTC enrollment have some value in predicting the number of second lieutenants commissioned each year. The other variables examined, (advertising budget, leads resulting from advertising, scholarships, freshman enrollment, and youth attitude tracking survey), were not significant and appeared to have little value in explaining the dependent variable. Model one, the simple linear regression model, indicated a positive correlation existed between the number of second lieutenants commissioned in a year and the variables unemployment rate, average yearly college tuition, and Army ROTC enrollment. This indicates as the value of these variables increases or decreases the dependent variable will act in a similar manner. In terms of college costs and the unemployment rate, this may be interpreted as saying high unemployment and increasing costs may make the Army ROTC program more attractive while a strong economy with low unemployment and inflation may make the program less attractive to college students. Of the multiple regression models used, model four using the variables average college tuition and unemployment rate appears to be the most useful in predicting the commissioning of second lieutenants in terms of multiple variables. This model had a high value for R^2, .90, and the model level of significance, .000, was also high. As a predictor model, caution should be taken in using data outside the range of the data used in this analysis. It would be inappropriate to use extremely high unemployment rates (above 11%) and average yearly college costs above $4000. Any results obtained using figures outside this range should be viewed with caution.

B. OBSERVATIONS

The following observations were made concerning the variables during this analysis.

1. Enrollment in MSI has increased at a significantly higher rate than in MSII, MSIII, and MSIV indicating a problem may exist in retaining cadets in the program.

2. The Army ROTC scholarship program appears to have a significant effect on retention.

3. The number of leads produced as a result of the advertising budget indicates the advertisements may not be attracting the right audience or the collection of the leads data is inaccurate.

4. The unemployment rate appears to be positively correlated with the number of second lieutenants commissioned each year indicating economics is an important factor in attracting and retaining students in the Army ROTC program.

5. Average yearly college tuition is also positively correlated to the dependent variable indicating economics is an important factor.

6. The limited number of data points for the youth attitude tracking survey makes it difficult to draw any conclusions about the relationship between this variable and the number of second lieutenants commissioned in a year.

C. RECOMMENDATIONS

The following recommendations are made as a result of this study.

1. Additional data should be collected to validate and improve the three variable regression model obtained.

2. The Army ROTC scholarship program should be further studied and reviewed for cost effectiveness and increased to meet retention and commissioning goals.

3. The advertising budget effectiveness should be reviewed and the data collection method for leads be validated and designed to correlate names with enrollment.

4. The youth attitude tracking survey should be redesigned to include specific questions pertaining to high school student's interest in joining Army ROTC.

5. Further research should be made into investigating why attrition between the MSI and MSII years is so high.

APPENDIX A. PREDICTION INTERVAL DEVELOPMENT

The following is an example of how a 95% prediction interval can be developed. There is an underlying assumption with these prediction intervals that the observed values for the dependent variable are normally distributed with a constant variance. The matrix notation below is introduced for ease of discussion. Let Y be a n by one vector, n being the number of data points, whose elements are given by y_i, for example,

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ ... \\ y_n \end{bmatrix}$$

where the y_i's are the observed values of the dependent variable. Let $\hat{\beta}$ be a p by one vector, p being the number of predictors, whose elements are given by $\hat{\beta}_i$, for example,

$$\hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ ... \\ \hat{\beta}_p \end{bmatrix}$$

where the $\hat{\beta}_i$'s are the estimated values for β. Let C be a p by one vector whose elements are given by c_i, for example,

$$C = \begin{bmatrix} 1 \\ c_1 \\ c_2 \\ ... \\ c_p \end{bmatrix}$$

where the individual c's are the hypothesized values for the predictors. The first element of the vector C, one, corresponds to the intercept term. Let X be defined as a n by (p+1) matrix given by

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

where the individual x's are the observed data points of the independent variables. The ones in the first column represent the intercept term. $\hat{Y}$, the point estimate for the number of second lieutenants commissioned in a year, may be estimated by the equation

$$\hat{Y} = C^T \hat{\beta}. \tag{12}$$

For model one using an estimate of 10,000 for MSIV enrollment, the values for C and $\hat{\beta}$ are

$$C = \begin{bmatrix} 1 \\ 10000 \end{bmatrix}$$

and

$$\hat{\beta} = \begin{bmatrix} 1892.1 \\ .61898 \end{bmatrix}$$

Multiplying the two matrices together leads to a $\hat{Y}$ of 8081. The 95% prediction interval for $\hat{Y}$ is given by

$$\hat{Y} \pm t_{.05/2}\hat{\sigma}(1 + C^T(X^TX)^{-1}C)^{1/2}$$

where $t_{.05/2}$ is the 97.5th quantile of the student's t distribution with n-2 degrees of freedom. For this example, the value of $(X^TX)^{-1}$, which is the variance-covariance matrix divided by the standard error squared, is

$$(X^TX)^{-1} = \begin{bmatrix} 8.2965E-1 & -9.3785E-5 \\ -9.3785E-5 & 1.1465E-8 \end{bmatrix}$$

and the value for $\hat{\sigma}$ is 615.01. The 97.5th quantile of the student's t distribution for a 95% prediction interval (α equal to .05) with n-2 degrees of freedom (n= 16) is 2.14. Performing the matrix multiplication leads to a value of .10048 for $C^T(X^TX)^{-1}C$.

44

Substituting these values into the prediction interval formula leads to an interval of 8,081 ± 1,381. Similarly, prediction intervals for each of the other models can be developed.

APPENDIX B.   RESIDUAL ANALYSIS

The following are residual plots for each model. Each set of residuals has been fitted to a cumulative distribution function of the normal distribution with the Kolmogorov-Smirnov 95% bounds. The randomness of the scatter plots of the residuals versus the predicted values indicates the residuals are randomly distributed with a constant variance.

MODEL ONE MSIV RESIDUALS FIT TO NORMAL CDF
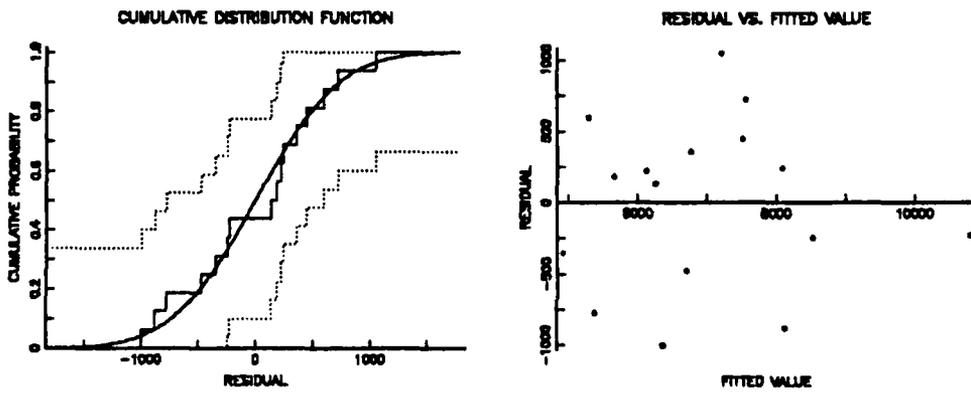
Figure 13.   Model one residuals for MSIV

MODEL ONE UNEMPLOYMENT RESIDUALS FIT TO NORMAL CDF
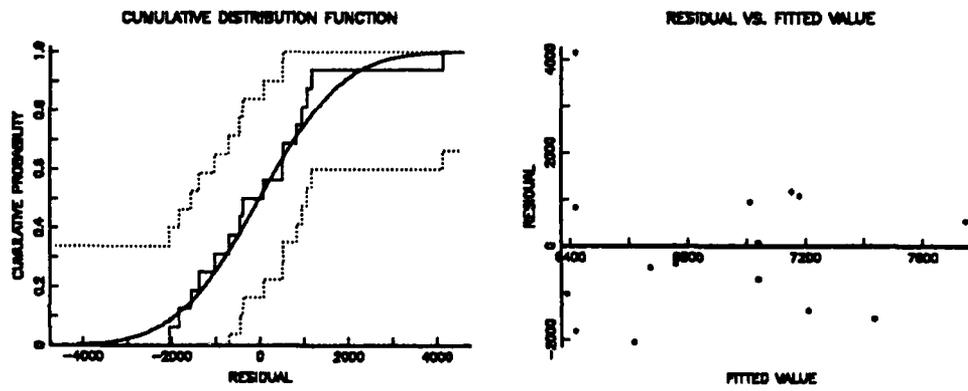
Figure 14.   Model one residuals for unemployment

MODEL ONE COLLEGE COST RESIDUAL FIT TO NORMAL CDF
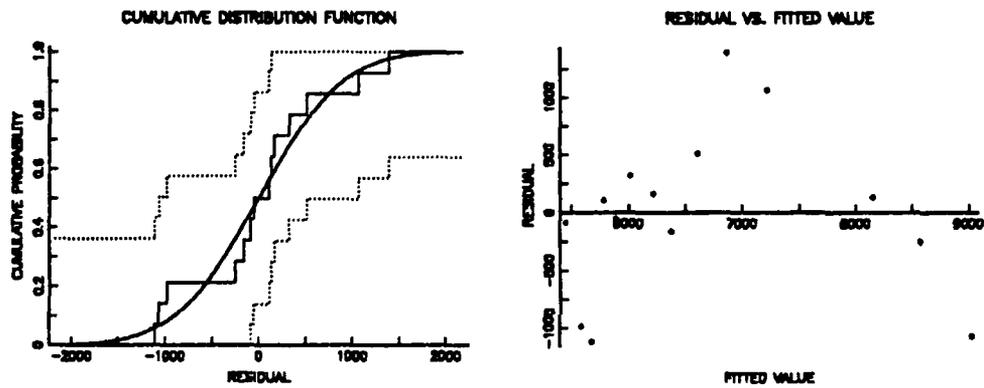
Figure 15.   Model one residuals for college costs

MODEL TWO UNEMPLOYMENT RESIDUALS FIT TO NORMAL CDF
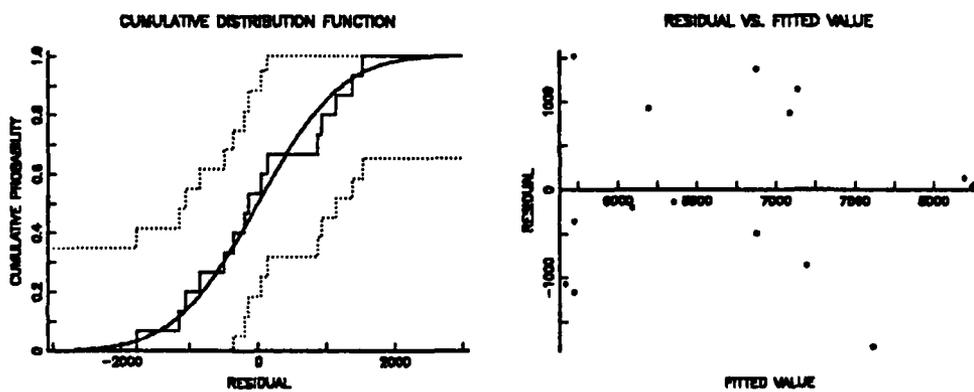
Figure 16. Model two residuals for unemployment

50

MODEL TWO COST RESIDUALS FIT TO NORMAL CDF
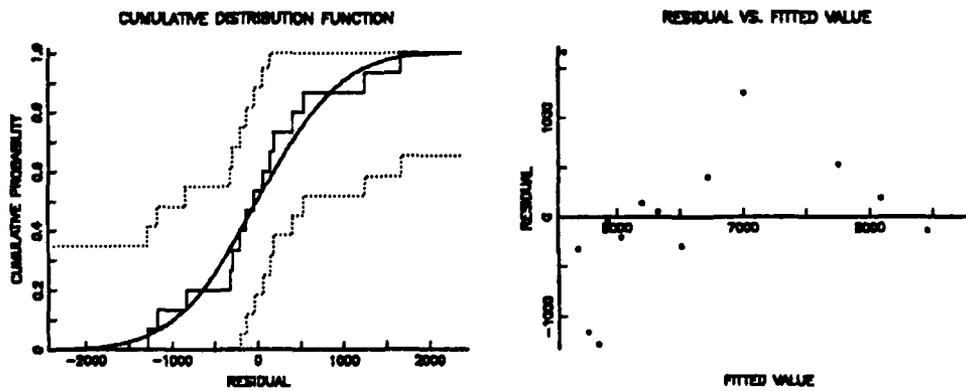
Figure 17. Model two residuals for college costs

51

MODEL THREE RESIDUALS WITH FOUR LAG FIT TO NORMAL
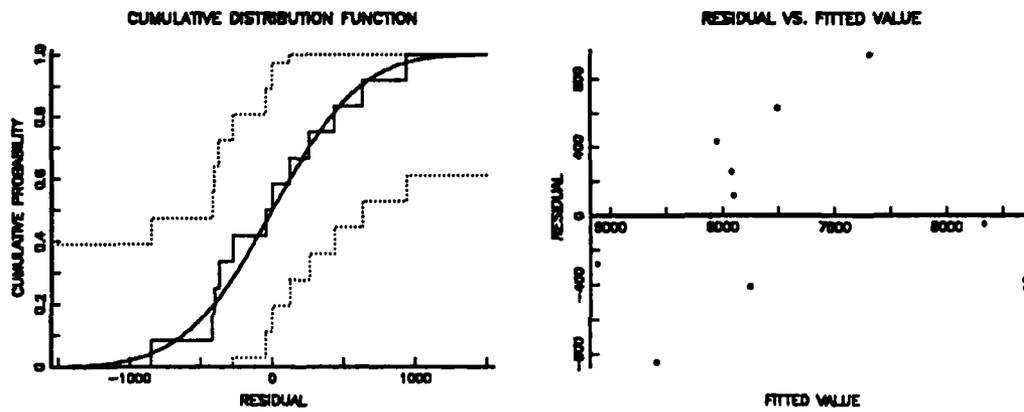
Figure 18.   Model three residuals with a four year lag

MODEL THREE RESIDUALS TWO YEAR LAG FIT TO NORMAL CDF
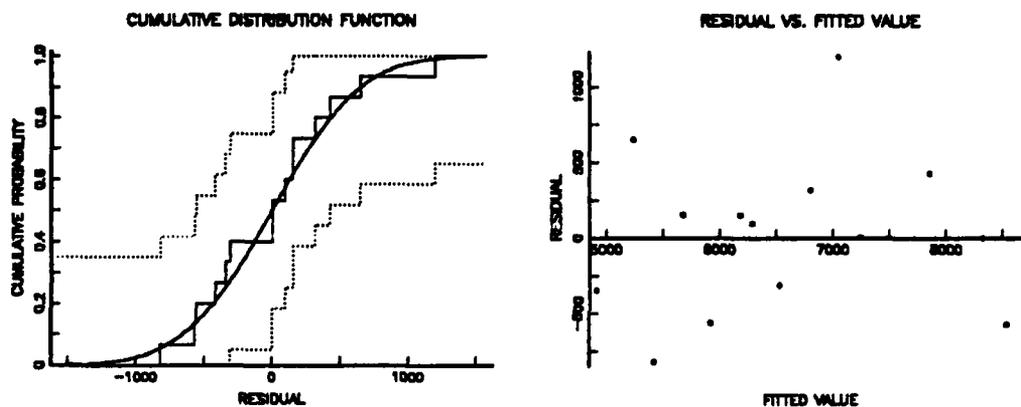
Figure 19.   Model three residuals with a two year lag

MSII AND MSIV RESIDUALS FIT TO NORMAL CDF
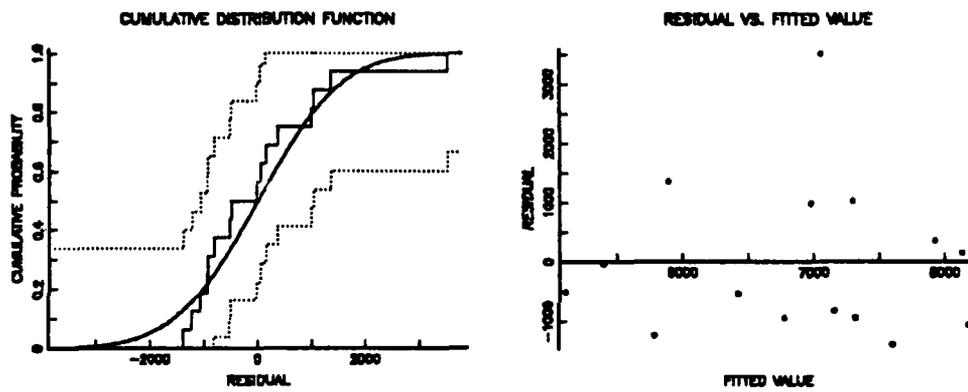
Figure 20.   MSII and MSIV residuals with varying lags

Figure 21. Model four residuals

LIST OF REFERENCES

1. ROTC Study Group, *ROTC Study Final Report*, 1986.

2. U.S. Department of Commerce, *National Data Book and Guide to Resources*, 106th ed., U.S. Government Printing Office, 1987.

3. U.S. Department of Education, *Digest of Education Statistics*, U.S. Government Printing Office, 1987.

4. Defense Manpower Data Center, *Youth Attitude Tracking Study: Historical Evolution and Characteristics*, by M.T. Laurence and S.T. Bridges, 1986.

5. DeGroot, M.H., *Probability and Statistics*, 2d ed., Addison Wesley Co., 1986.

6. Johnston, J., *Econometric Methods*, 2d ed., McGraw-Hill Book Co., 1972.

7. Rao, P. and Miller, R.L., *Applied Econometrics*, Wadsworth Publishing Co., 1971.

INITIAL DISTRIBUTION LIST