NAVAL POSTGRADUATE SCHOOL
Monterey, California

THESIS

AN ANALYSIS OF THE COAST GUARD ENLISTED ATTRITION

by

Laureano Enrique Oñate Rubiano

September 1993

Thesis Advisor: So Young Sohn

REPORT DOCUMENTATION PAGE

Form Approved OMB Np. 0704

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
September 1993

3. REPORT TYPE AND DATES COVERED
Master's Thesis

4. TITLE AND SUBTITLE An Analysis of the Coast Guard Enlisted Attrition

5. FUNDING NUMBERS

6. AUTHOR(S) LCDR Laureano Enrique Oñate Rubiano

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Postgraduate School
Monterey CA 93943-5000

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES 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.

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(maximum 200 words)*

In this this thesis, survival analysis is used to study US Coast Guard enlisted attrition behavior in terms of individual personnel characteristics such as sex, marital status, race, paygrade and rating. Results obtained based on 8 years of historical data from FY83 to FY90 are as follows: males and married individuals have higher survival probabilities than their counter parts, respectively; paygrades E-1 to E-5 have higher attrition than paygrades E-6 to E-9; American Indians have the highest attrition and Asian members have the highest survival probabilities; rating 170 (Gunner's Mate) has the highest attrition over all ratings followed by rating 180 (Fire Control Technician); the rating with the highest survival probability is 570 (Aviation Machinist's Mate); a decreasing trend in attrition was found during the last 4 years of the observation period; it was also observed that there was significantly high attrition at the end of the four years service contract and when the enlisted member reachs twenty years of service. Aditionally, this thesis provides the a regression model in order to predict monthly enlisted attrition figures. Significant predictors selected are the prior month's attrition, the number of enlistments four years prior and the current unemployment rate. The selected regression model explains almost 97% of the total variation of monthly attrition. It turns out to perform better than the current method used by the CG.

14. SUBJECT TERMS Separated Enlisted; Monthly Attrition; Predictors; Survival Functions; Individual Personnel Characteristics

15. NUMBER OF PAGES 51

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

Approved for public release; distribution is unlimited.

AN ANALYSIS OF THE COAST GUARD ENLISTED ATTRITION

by

Laureano Enrique Oñate Rubiano
Lieutenant Commander, Colombian Navy
B.S., Escuela Naval de Cadetes Cartagena Colombia, 1987

Submitted in partial fulfillment
of the requirements for the degree of

MASTER OF SCIENCE IN OPERATION ANALYSIS

from the

NAVAL POSTGRADUATE SCHOOL
September 1993

Author: ______________________
Laureano Enrique Oñate Rubiano

Approved by: ______________________
So Young Sohn, Thesis Advisor

____
Robert R. Read, Second Reader

____
Peter Purdue, Chairman
Department of Operation Research

ABSTRACT

In this this thesis, survival analysis is used to study US Coast Guard enlisted attrition behavior in terms of individual personnel characteristics such as sex, marital status, race, paygrade and rating. Results obtained based on 8 years of historical data from FY83 to FY90 are as follows: males and married individuals have higher survival probabilities than their counter parts, respectively; paygrades E-1 to E-5 have higher attrition than paygrades E-6 to E-9; American Indians have the highest attrition and Asian members have the highest survival probabilities; rating 170 (Gunner's Mate) has the highest attrition over all ratings followed by rating 180 (Fire Control Technician); the rating with the highest survival probability is 570 (Aviation Machinist's Mate); a decreasing trend in attrition was found during the last 4 years of the observation period; it was also observed that there was significantly high attrition at the end of the four years service contract and when the enlisted member reachs twenty years of service.

Additionally, this thesis provides a regression model in order to predict monthly enlisted attrition figures. Significant predictors selected are the prior month's attrition, the number of enlistments four years prior and the current unemployment rate. The selected regression model explains almost 97% of the total variation of monthly attrition. It turns out to perform better than the current method used by the CG.

TABLE OF CONTENTS

LIST OF FIGURES

Executive Summary

One of the responsibilities of the US Coast Guard Personnel Workforce Planning Office is to forecast enlisted monthly attrition. In order to forecast enlisted monthly attrition, it is necessary to analyze attrition behavior. The current method to forecast the number of enlisted attrition is based on the mean of the past eight years attrition figures of the corresponding month. This method facilitates simple implementation but it does not utilize other useful information such as economy, attrition behavior and individual characteristics associated with a specific kind of enlisted member.

In this thesis, first, survival analysis was applied to investigate the attrition behavior of the US Coast Guard enlisted in terms of their individual characteristics such as sex, marital status, race, paygrade and rating. Results obtained based on the past eight years data from October of 1982 to September of 1990 are as follows: males and married individuals have higher survival probabilities than their counterparts, respectively; paygrades E-1 to E-5 have higher attrition than paygrades E-6 to E-9; American Indians have the highest attrition and the Asian members the highest survival functions; rating 170 (Gunner's Mate) has the highest attrition over all ratings followed by rating 180

(Fire Control Technician); the rating with the highest survival probability is 570 (Aviation Machinist's Mate); there was a decreasing trend in attrition in the last 4 years of the observation period. The common feature of the survival functions was the significant fall at the end of the four years service contract and when the enlisted member reach twenty years of service.

Secondly, based on the observations made in the survival analysis a candidate set of predictors was selected to fit a regression model for enlisted monthly attrition. The model finally selected contains the prior month's attrition, the number of enlistments four years prior and the current unemployment rate as significant predictors. The selected regression model explains almost 97% of the total variation of monthly attrition. Performance of the selected model is better than the current method used by US Coast Guard in terms of the mean squared error and the mean relative error.

I. INTRODUCTION

A. PROBLEM DESCRIPTION

One of the responsibilities of the United States Coast Guard (CG) Personnel Workforce Planning (PWP) Office, located in Washington D.C., is to forecast personnel stocks, promotion requirements, monthly attrition and recruitment needs. The current method to forecast the number of enlisted attrition by paygrade per month is based on the mean of the past previous eight years attrition figures. For the corresponding month this method facilitates simple implementation. However, it does not utilize other useful information such as attrition behavior and individual characteristics associated with a specific kind of enlisted member. The CG Office of Personnel and Training maintains enlisted personnel records in addition to information concerning the main reasons for enlisted attrition. Recently a study was written [Ref. 1], examining enlisted attrition behavior and developing a model that projects these attrition figures. The study lacked of temporal stability, as it used personnel data for only one year (FY91). The conclusions and modelling results were necessarily not very reliable.

B. OBJECTIVES

This thesis uses eight years of historical data from FY83 through FY90 to obtain the survival functions of the individual personnel characteristics for survival analysis. Additionally, it applies regression models to predict monthly attrition figures and compares the performance of the resulting prediction model to those previously in use.

C. PERSONNEL BACKGROUND

A summary from USGC LT Douglas Allen Blakemore's thesis is used to describe the personnel background of the Coast Guard enlisted personnel structure.

The Coast Guard enlisted personnel structure is composed of nine paygrades from E-1 to E-9. The paygrade E-1 contains recruits attending Coast Guard basic training school; paygrade E-2 is composed of enlisted personnel who have completed basic training and have been assigned to active duty commands in preparation for attending a CG specialty school. Paygrades E-3 through E-9 are divided into 24 active duty military occupational skills (MOS) or subspecialities. Of these, only 22 are of concern since one (Musician), has an extremely small attrition rate and the other (Sonar Technician), no longer exists. The following list contains MOS; their respective CG abbreviation for subspecialty name and the rating code. It will be used throughout this thesis to calculate survival probabilities.

MOS Subspecialty	Rating	Code
- Aviation Machinist's Mate	AM	570
- Aviation Electrician's Mate	AE	560
- Aviation Damage Controlman	AD	520
- Aviation Survivalman	ASM	530
- Aviation Electronic Technician	AT	550
- Boatwain's Mate	BM	100
- Damage Controlman	DC	210
- Electrician's Mate	EM	270
- Electronics Technician	ET	240
- Fire Control Technician	FT	180
- Gunner's Mate	GM	170
- Health Service Technician	HS	870
- Machinery Technician	MK	200
- Marine Science Technician	MST	790
- Public Affairs Specialist	PA	340
- Quartermaster	QM	110
- Radarman	RD	130
- Radioman	RM	350
- Storekeeper	SK	420
- Subsistence Specialist	SS	500
- Telephone Technician	TT	280
- Yeoman	YN	360

An enlisted man (EM) receives an MOS after completing a specialty school or completing an intensive on-the-job training program. Both methods are usually administered while the EM is in the E-3 paygrade. An E-3 who has not obtained an MOS is called non-rated; E-3's through E-9's hold only one MOS at a time and usually maintain that MOS throughout their careers.

Promotion to the next highest paygrade is determined by MOS and dictated by the needs of the CG. Promotions to paygrade E-5 through E-9 are vacancy driven while those to paygrades E-3 and E-4 occur on qualification for advancement. Promotion to paygrade E-2 occurs upon completion of basic training.

Enlisted personnel sign service obligation contracts (enlistments) that require the individual to serve in the CG for a pre-determined number of years, usually four years. (In the past, there have been two year enlistments but these have been terminated.) Upon completion of an enlistment and upon approval of the CG, a person may sign a new contract (re-enlist), or separate from the CG. Enlisted attrition generally occurs due to :

- Retirement - After 20 years of active duty service.
- Non-re-enlistments - An EM chooses not to reenlist or the CG chooses not to reenlist the individual.
- Administrative reasons - An EM may depart the CG prior to the end of his/her contract enlistment -" for the convenience of the government ".
- Selection to an Officer program.
- Death or disability.

This study will concentrate on the first four reasons listed above and will use the word "separation" referring to the EM leaving active duty service.

D. SCOPE OF THE THESIS

The first goal of this study is to develop survival functions for USCG EM personnel, in which survival analysis techniques can be used to analyze data on the length of time an EM remained in CG. This technique takes on different names, depending on the particular application at hand.

Recently the term, "event history analysis", has been used by social scientists to describe applications for the analysis of the length of time it takes an employee to retire or resign from a given job [Ref. 2].

Survival analysis is a method for describing the distribution of the length of time for a given event, such as the termination of the service in the CG. One way to perform survival analysis is to construct a histogram for the length of time that individuals spend in service. Alternatively, one can use the length of service time as a dependent variable and determine if it can be predicted by variables such as marital status, race, gender or military occupational skill.

The main tool used in the survival analysis is the survival function calculated from the data using the frequency histogram in which for any given time t, the area under the curve to the left of t is the proportion of individuals in the population who separate from the CG to time t. A common feature of survival data is the presence of right censored observations due either to withdrawal of experimental units or termination of the experiment. For such observations it is only known that the lifetime exceeded the given value. The exact life time remains unknown. Such data cannot be analyzed by ignoring the censored observations because, among other considerations, the longer-lived observations are generally more likely to be censored. The survival distribution function evaluated at t is the probability that an observation from the population will have a lifetime exceeding t, that is $S(t) = Prob(T > t)$. Their estimators are called product limit estimators. [Ref 2].

An analysis with the entire data set will be elaborated, in order to know some specific trends, behaviors and predominance of the individual characteristics of the data under study.

A regression model will be developed to forecast monthly attrition and to establish relationships between explanatory variables and the attrition.

II. DATA OVERVIEW

A. POPULATION

The USCG Office of Personnel and Training located in Washington D.C., provided the data for this study. The data set contained 27,2160 individual personal records for all CG enlisted personnel for the observation period from FY83 to FY90. The fields contained in each record are : Personnel Identification Number (PID), a generic number which identifies one individual in the CG; rating, according to the military occupational skill; paygrade, from E-1 to E-9; sex, males and females; minority designator, race of the individual; marital status, married and single; date of entrance into the CG; date of separation from the CG; GAOCD, is a separation designator code for CG enlisted personnel; censor, which indicates whether the person remained in the CG or not. The data contains a record for each enlisted member per year of service in the CG, in other words, if a person remains in service for 8 years there exists 8 records for that individual. These multiple records were collapsed as one record per person, without loss of information. After this process, there were 50,036 records from which 29,405 belonged to people who left active duty service during the observation period and 20631 which belonged to censored active duty CG members, at the end of FY90 (30 Sep/90). This data includes cases where the CG allowed a member to separate from the CG and then rejoin in the future at the member former paygrade.

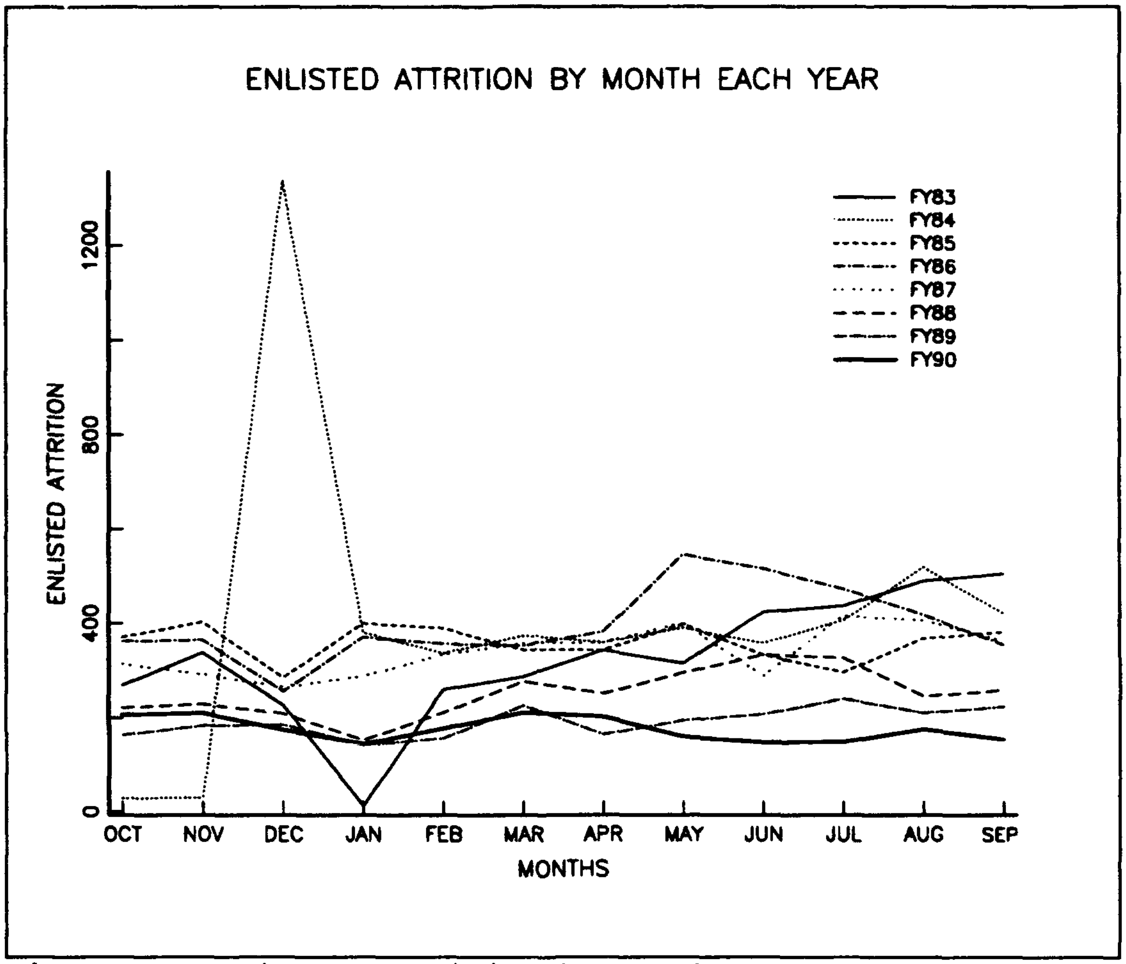

Figure 1. Enlisted attrition by month each year.

Based on this data set the number of months a person spent in the CG can be calculated, to analyze total time in service. In addition, censored information can be obtained. Figure 1 shows the monthly attrition for each year from FY83 to FY90. The attrition for Dec/83 appears to be an outlier since the attrition values for the other months that remain are no are no greater than 650. However, this value is preceded by the two smallest attrition values in the sample (25 and 30) and they both are under the next minimum value of attrition (182).

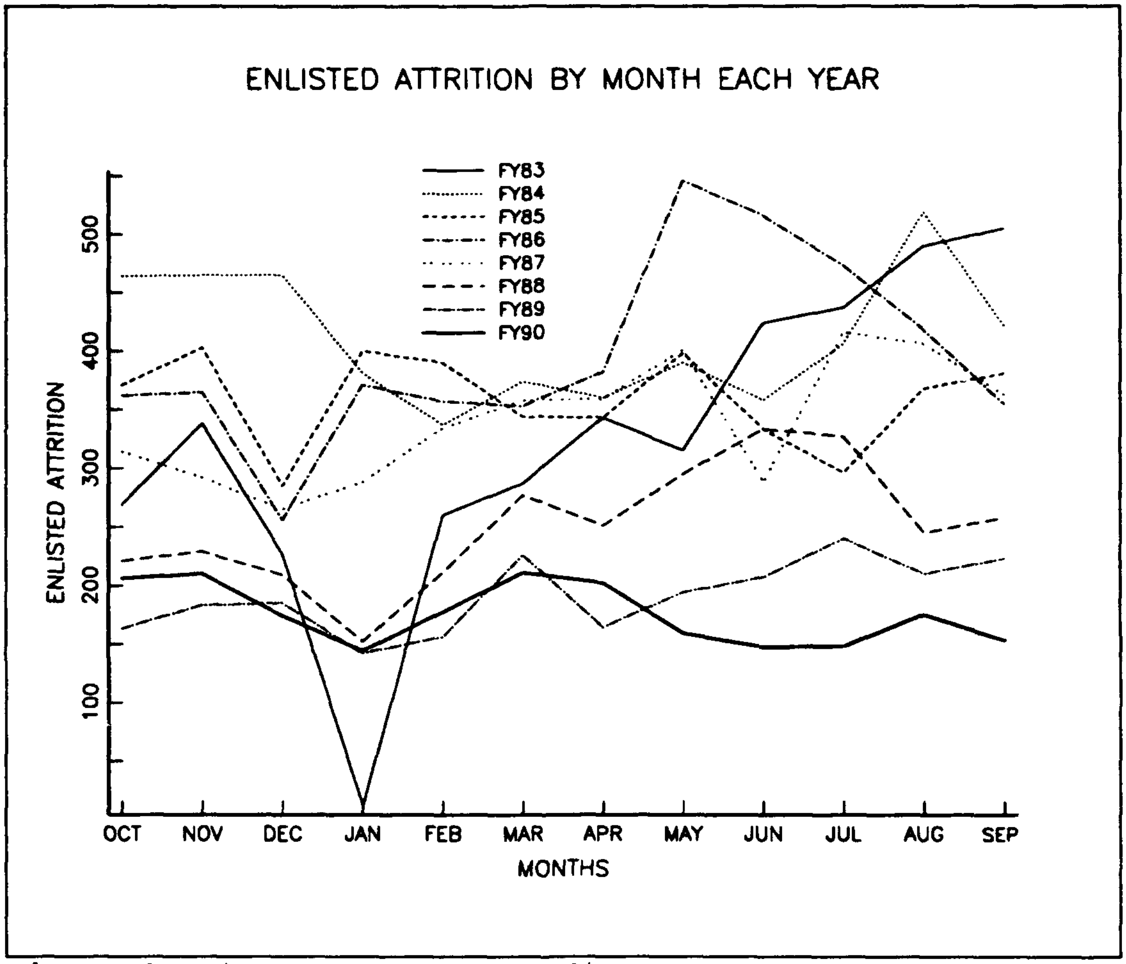

Figure 2. First quarter FY84 fixed.

In the future, analysis will conducted on the first quarter of FY84 as an average of the three attrition values corresponding tc Oct/83 (25), Nov/83 (30) and Dec/83 (1,339). As can be seen in Figure 2, the new values for FY84 are in the ranges of all other fiscal years values.

B. SUMMARY OF PERSONNEL

Table 2 provides a comprehensive view of the stocks of personnel by paygrade of enlisted active duty (AC) at 30 sep/90, and the total of enlisted members who separed (RE) during the observation period from starting FY83 to the end of FY90. Of note is the fact that 22 % of the total data is comprised of those members who are in the E-2 paygrade and the highest amount of attrition in paygrade E-4.

Table 2. Active Duty Enlisted at 30 Sep/90 and Separated Enlisted Personnel from FY83 to FY90

PAYGRADE	ACTIVE DUTY	SEPARATED	TOTAL
E-1	1781	4778	6559
E-2	6894	4024	10918
E-3	4116	4918	9034
E-4	1893	7619	9512
E-5	2003	4252	6255
E-6	2548	1880	4428
E-7	1199	1084	2283
E-8	137	441	578
E-9	60	409	469
TOTAL	20631	29405	50036

Table 3. Personnel by sex and Marital Status

PAYGRADE		SEX		MARITAL STATUS		TOTALS
		MALE	FEMALE	SINGLE	MARRIED	
E-1	AC	1592	189	1666	115	1781
	RE	4009	769	4272	506	4778
E-2	AC	6200	694	5900	994	6894
	RE	3359	665	3226	798	4024
E-3	AC	3808	308	2941	1175	4116
	RE	4328	590	3615	1303	4918
E-4	AC	1765	128	1424	469	1893
	RE	6947	672	5413	2206	7619
E-5	AC	1895	108	1292	711	2003
	RE	3888	364	2533	1719	4252
E-6	AC	2497	51	1220	1328	2548
	RE	1802	78	734	1146	1880
E-7	AC	1198	1	325	874	1199
	RE	1083	1	237	847	1084
E-8	AC	136	1	16	121	137
	RE	441	0	85	356	441
E-9	AC	60	0	4	56	60
	RE	409	0	90	319	409
TOT	AC	19151	1480	14788	5943	20631
	RE	26266	3139	20205	9200	29405

Table 3, present the stocks of personnel by sex and marital status. As can be observed, males, and single members are the dominant groups in the CG; females have reached paygrade E-8; married members (4,813) compared to single (16,526), left from the CG in low Paygrades (E-1 to E-4), in lower proportion compared with paygrades E-5 to E-9 (married (4,387), single (3,679)); there are small amount of single members in paygrades E-8 and E-9 in active duty.

Table 4 contains the enlisted stocks classified by race. The findings observed here are: Caucasian has the largest stocks followed by Blacks, Hispanic, American Indians and Asian.

Table 4. Active Duty and Retired personnel by race

PAYGRADE		RACE					TOTAL
		BLACK	HISPA	A IND	ASIAN	CAUCA	
E-1	AC	82	62	34	18	1585	1781
	RE	563	238	118	40	3819	4778
E-2	AC	412	323	121	57	5981	6894
	RE	534	258	89	51	3092	4024
E-3	AC	306	233	50	35	3492	4116
	RE	656	301	89	38	3834	4918
E-4	AC	184	84	18	18	1589	1893
	RE	691	325	103	66	6434	7619
E-5	AC	244	85	7	23	1644	2003
	RE	309	150	33	49	3711	4252
E-6	AC	165	61	9	31	2282	2548
	RE	107	49	8	90	1626	1880
E-7	AC	44	25	6	14	1110	1199
	RE	36	17	0	140	891	1084
E-8	AC	2	1	0	1	133	137
	RE	10	7	0	51	373	441
E-9	AC	3	0	0	1	56	60
	RE	8	1	0	27	373	409
TOT	AC	1442	874	245	198	17872	20631
	RE	2914	1346	440	552	24153	29405

During the eight years of observation American Indians have reached only paygrade E-7. Blacks, Hispanics, American Indians and Asian together are only the 13% of the population.

Table 5, classifies each paygrade by rating. The non-rated enlisted (MOS 150 and 320) comprise 49 % of the total population, the rating with the greatest amount of enlisted is 200 (Machinery technician) with 9%, followed by rating 100 (Boatwain's Mate) with 6.6%. The rating least populated is 180 (Fire Control Technician) with 114 enlisted (0.2%).

Table 5. Enlisted personnel by MOS

RATE	E-1	E-2	E-3	E-4	E-5	E-6	E-7	E-8	E-9
100	0	25	131	1122	681	696	450	80	119
110	0	13	62	407	266	233	134	25	15
130	0	9	29	187	86	84	41	6	8
150	6559	8083	6002	0	0	0	0	0	0
170	0	8	24	240	85	83	38	9	4
180	0	4	4	43	33	12	12	3	3
200	0	72	300	1647	1137	771	426	83	59
210	0	9	66	418	287	188	84	19	11
240	0	16	65	790	517	327	158	46	25
270	0	23	86	633	193	144	88	37	31
280	0	2	10	115	85	46	29	8	6
320	0	2472	1650	0	0	0	0	0	0
340	0	0	2	43	31	23	13	3	2
350	0	28	112	590	419	284	118	40	36
360	0	6	65	698	545	358	174	44	42
420	0	11	75	510	460	255	87	32	21
500	0	121	230	822	361	262	423	35	26
520	0	4	24	303	201	131	59	39	22
530	0	0	6	63	44	29	12	5	2
550	0	3	24	184	223	136	56	17	9
560	0	2	17	164	117	78	39	10	6
570	0	1	18	198	124	88	40	13	9
790	0	1	14	65	59	35	20	4	5
870	0	5	18	270	301	165	82	18	10
TOT	6559	10918	9034	9512	6255	4428	2283	578	469

III. SURVIVAL ANALYSIS

This chapter presents methods for further quantifying the probability that an enlisted member will separate from the CG. The time reference is established as the number of months an enlisted member served in the CG. This number is computed from the date at which the individual joined the CG and the date when he or she retired. Only integer months were calculated and the number of days less than 30 were neglected. For example, an individual who served between 87/09/05 (year, month, day) and 90/02/20, has a calculated value, named MY, of 29 months in active duty. With this information survival functions were developed by paygrade and individual characteristics such as sex, race, marital status and rating. These calculations were performed using the SAS LIFETEST procedure. A program listing is included as Appendix A.

A. SURVIVAL FUNCTIONS

A first step in the analysis of the survival data is the estimation of the distribution of the failure times. The survival distribution function (SDF), also known as the survival function, is used to describe the lifetimes of the enlisted personnel in the CG. The SDF evaluated at t, is the probability that an enlisted member will have a lifetime exceeding t, that is $S(t) = Prob(T>t)$ where $S(t)$ denotes the survival function and T is the lifetime of a randomly selected experimental unit. There are three types of plots commonly used with survival functions: a plot of the estimated survival function against time, a plot of the negative natural log of the estimated survival function against time, and a plot of the natural log of the negative natural log of the estimated survival function against the natural log of time. The last two plots provide empirical checks of the appropriateness of the exponential model and the Weibull model respectively, for the survival data [Ref. 3]. The most important task in the analysis of CG attrition data is the comparison of survival curves. It is of interest to determine whether two or more strata share the same survival functions. The life test procedure can be used with data that may be right censored, (enlisted who actually served in the CG at 30 Sep/90) to compute nonparametric estimates of the survival distributions for each stratum and to perform rank tests for the association of CG attrition with other individual characteristics. The survival estimates are computed within the defined strata levels. Rank statistics and likelihood ratio tests are used to test homogeneity of the survival function over different strata.

In the following section global survival functions are computed for individual characteristics such as sex, marital status, race, paygrade and rating.

1. Enlisted attrition by sex.

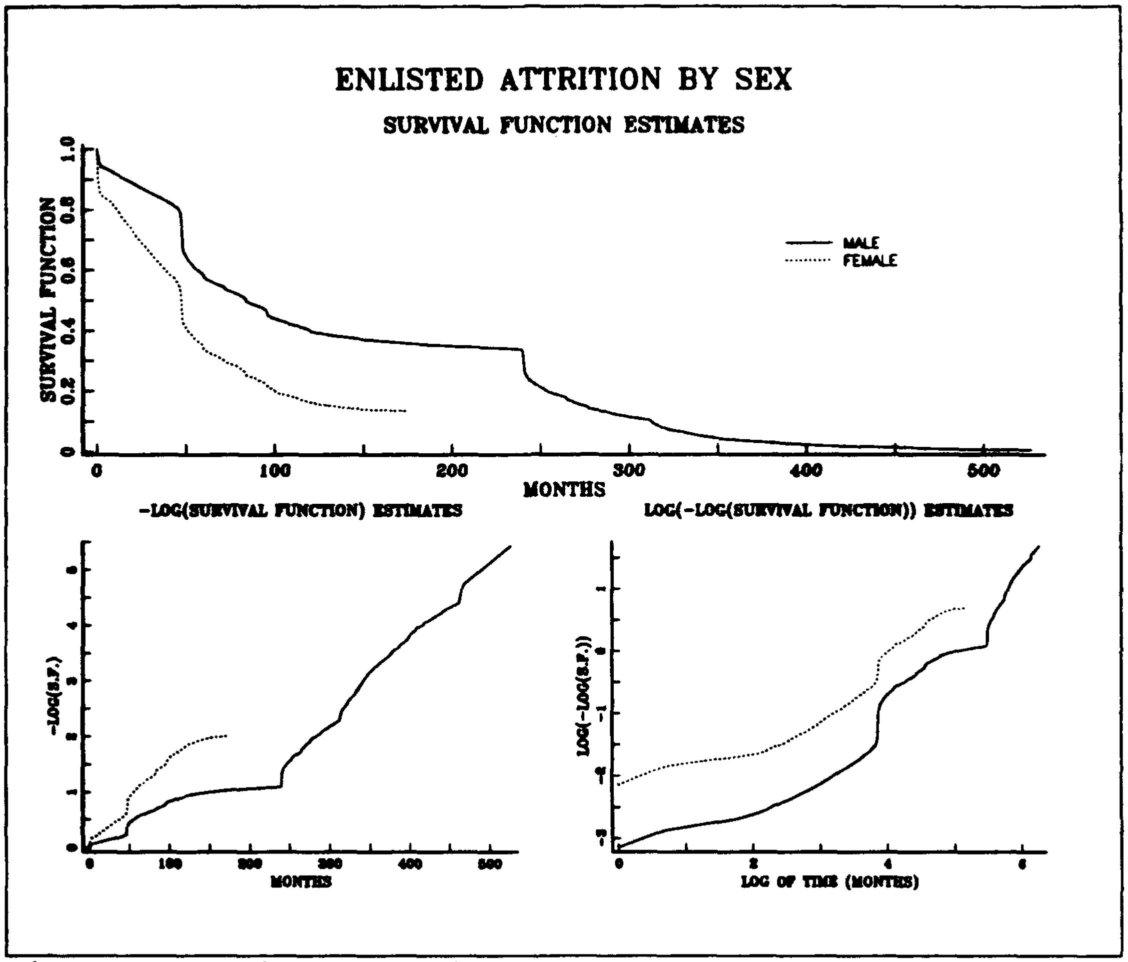

Figure 3. Enlisted attrition by sex.

As observed from the plot of the survival functions in Figure 3, records for females are limited to a maximum of 180 months since they initially started serving in CG in 1975. Females have demonstrated a higher attrition rate than their counterparts. Comparatively higher attrition rates are found close to 48 months of service for both groups, which corresponds to the end of the first contract. The three tests used for homogeneity indicate that the survival functions for the two sexes are significantly different. The log survival plot in figure 3, does not present a linear pattern through the origin which is required to support an exponential model, and the plot of the log of the negative log of the survival does not show a linear behavior in support of a Weibull model.

2. Enlisted attrition by marital status.

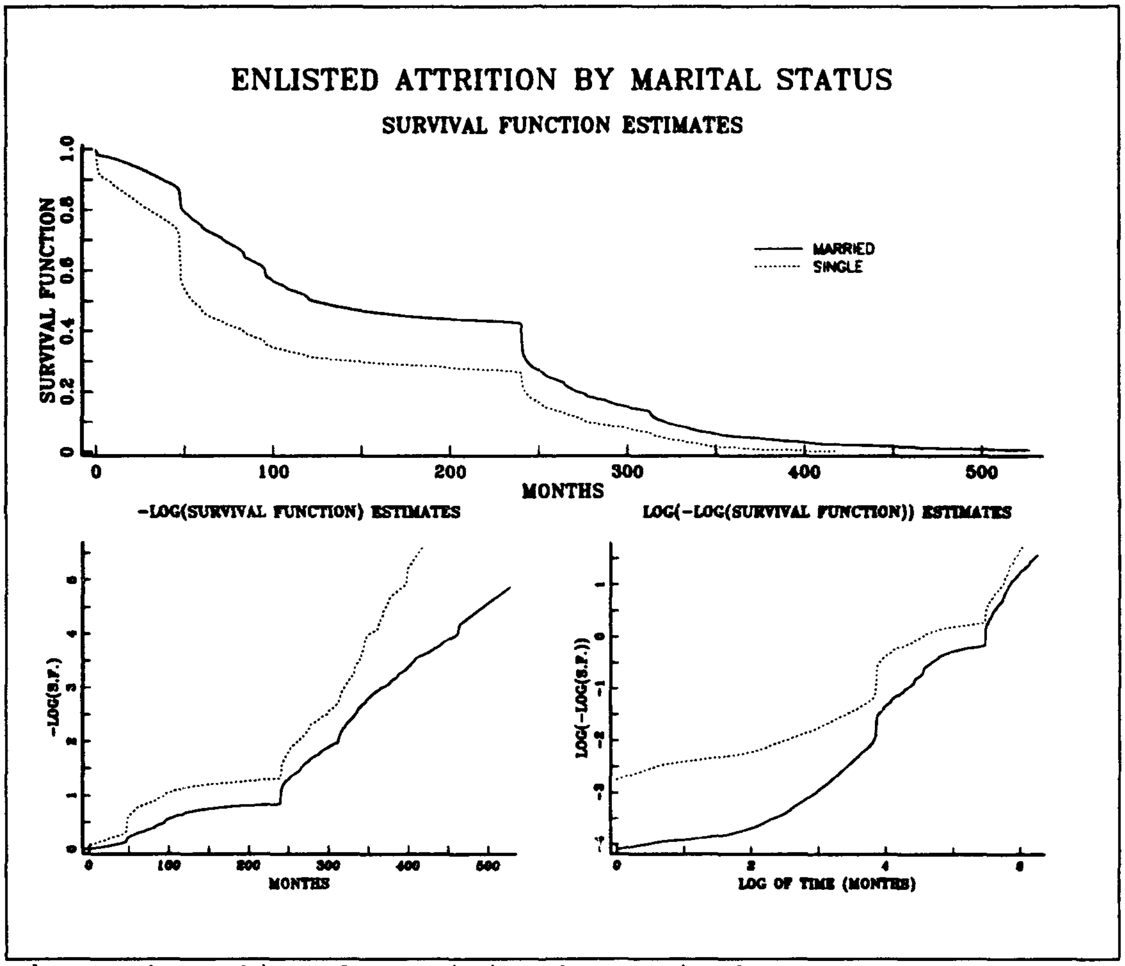

Figure 4. Enlisted attrition by marital status.

Figure 4 shows the plot of the survival function estimates where the probability to survive in the CG is higher for a married person than for a single member. The attrition is high in the starting period of service and decreases rapidly for single members once they finish their first four year contract as compared to that of married members. This behavior is inverted when a member reaches 20 years of service. Tests for homogeneity again, indicate that the survival functions for singles and married enlisted are significantly different at alpha = 0.05. The log plots support neither exponential nor Wiebull models for the data.

3. Enlisted attrition by race

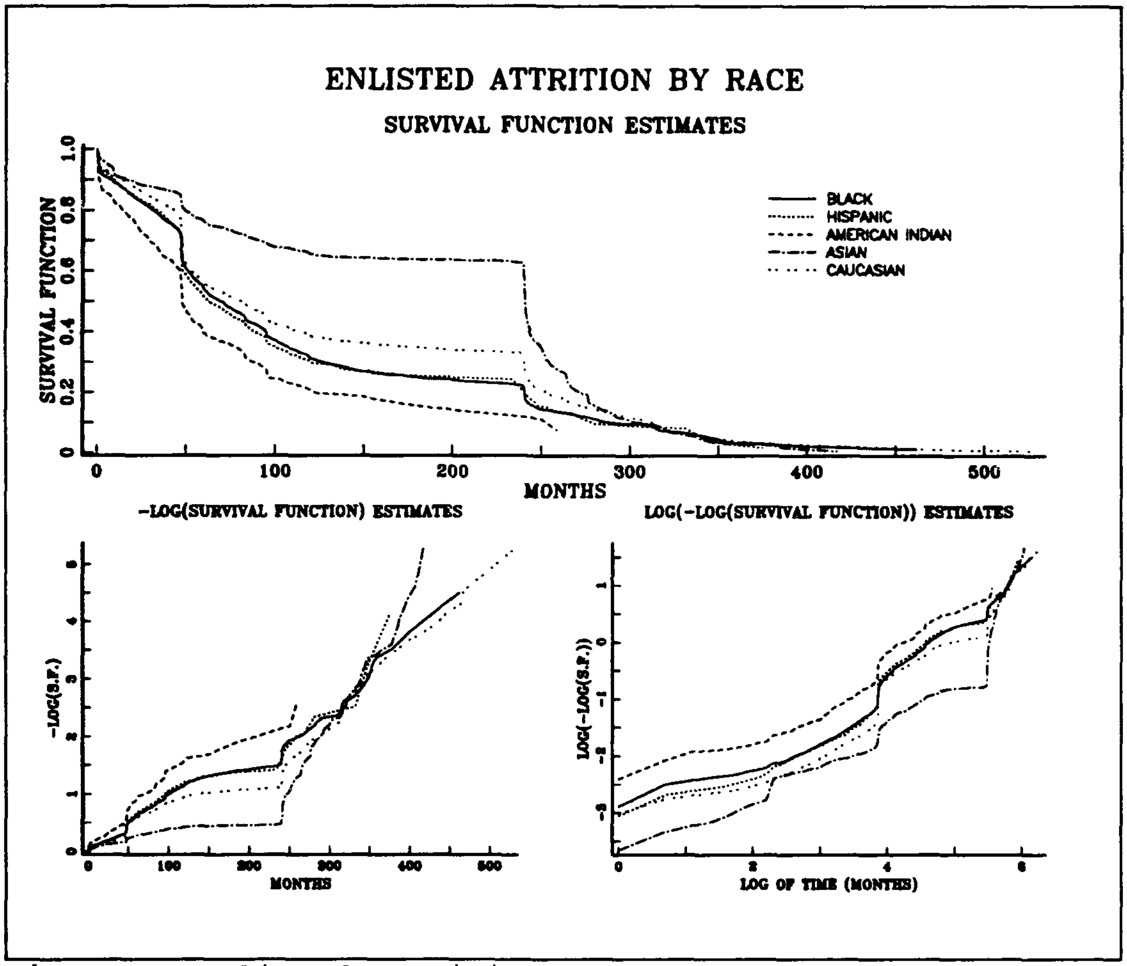

Figure 5. Enlisted attrition by race.

The survival functions here show a similar pattern for Blacks and Hispanics. American Indians have the highest attrition. Asian groups show the highest survival probability followed by Caucasians. The plots of the logs conform neither to exponential nor Weibull models.

4. Enlisted attrition by paygrade.

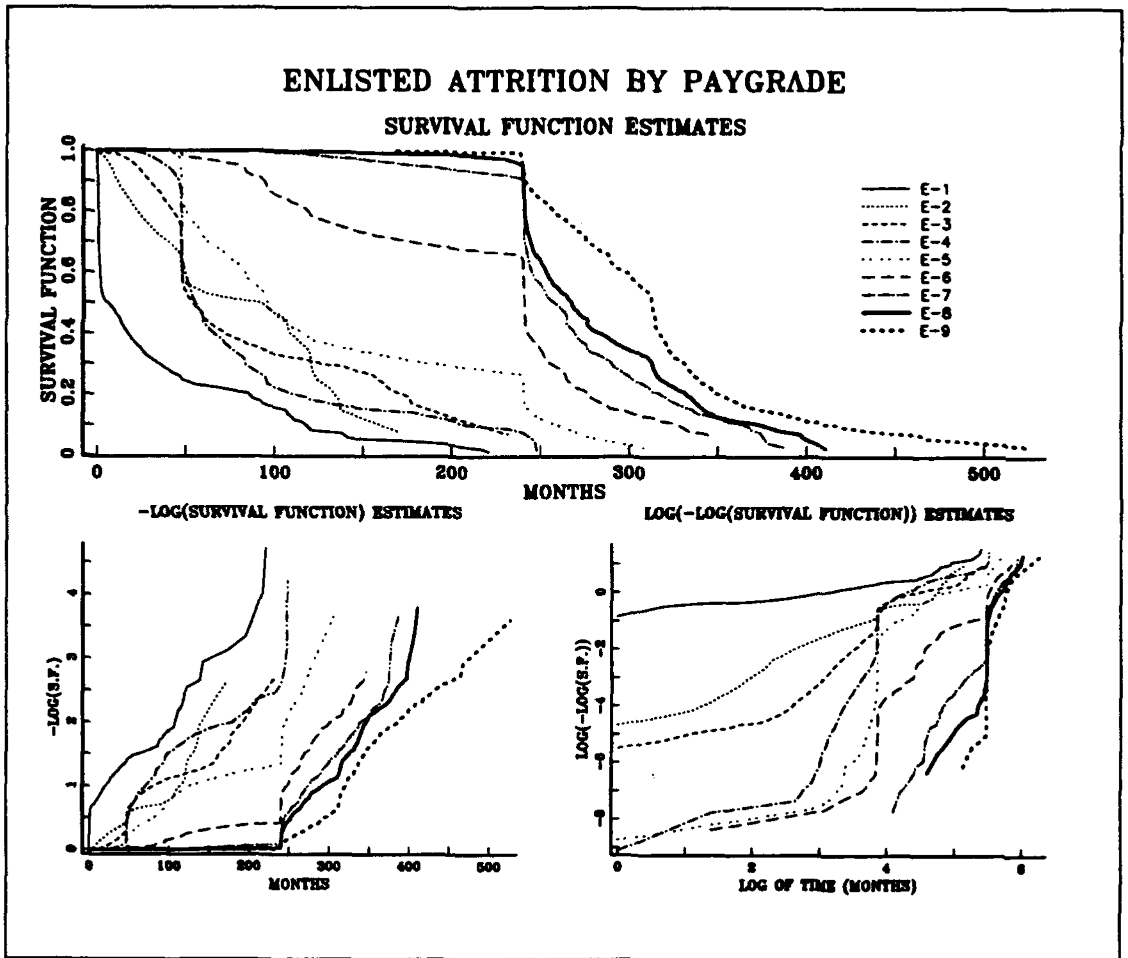

Figure 6. Enlisted attrition by paygrade.

The graph of the survival function estimates in Figure 6, represents the highest attrition on low paygrades until the end of the 4 year contract. The attrition for paygrades E-6 to E-9, increases abruptly after reaching 20 years of service.

The tests for homogeneity indicate that at least one of the survival functions are different from the others. However, in the negative log of the survival function against time plot, Figure 6, the paygrades E-1 to E-3 appear to be candidates for exponential models, while paygrades E-7 to E-9 appear to follow that of a Wiebull model. Note that the general pattern is piecewise linear for each paygrade until the two deep drops around 48 and 240 months.

5. Enlisted attrition by rating

Figure 7 represent the attrition by each rating. There are three of the survival functions that do not follow the pattern of the others. They correspond to the rating 170 (Gunner's Mate), followed in decreasing order by rating 280 (Telephone Technician) and rating 870 (Health Service Technician). The rating 570 (Aviation Machinist's Mate) is the group with the highest survival probability followed by rating 100 (Boatwain's Mate). All survival functions drop sharply around months 48 and 240.

B. SUMMARY

In this chapter, the survival functions were estimated in terms of individual characteristics of the enlisted member. Results obtained are as follow: males and married individuals have higher survival probabilities than their counterparts, respectively; paygrades E-1 to E-5 have higher attrition than

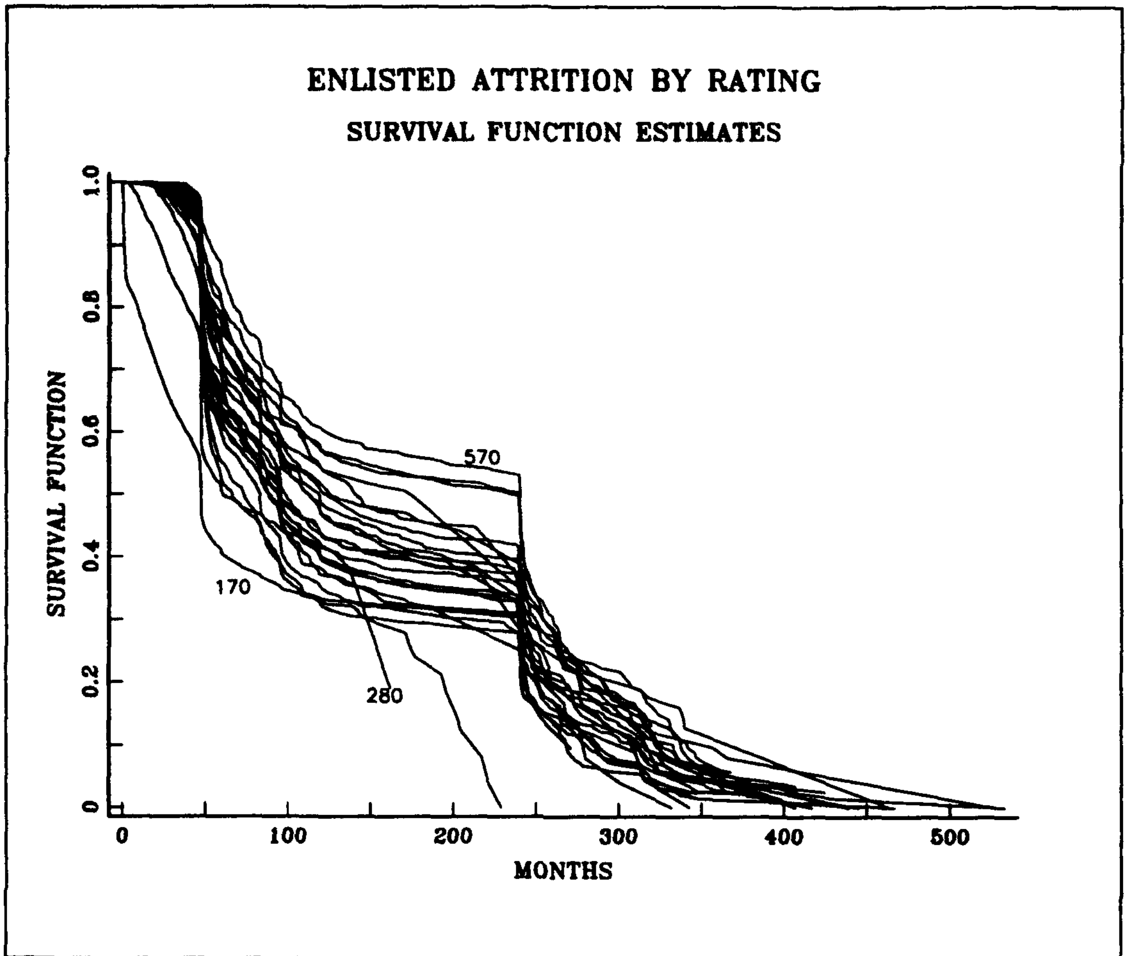

Figure 7. Enlisted attrition by rating.

paygrades E-6 to E-9; American Indians have the highest attrition and the Asian members the highest survival functions; rating 170 (Gunner's Mate) has the highest attrition over all ratings followed by rating 180 (Fire Control Technician); the rating with the highest survival probability is 570 (Aviation Machinist's Mate); there was a decreasing trend in attrition in the last 4 years of the observation period. The common feature of the survival functions was the significant fall at the end of the four years service contract and when the enlisted member reach twenty years of service.

In general the survival functions developed in this thesis have similar patterns for the paygrades, marital status and sex to those analyzed in the previous work by LT Blakemore. For rating and race, this thesis employed more strata than his and the results could not be compared.

It appear that neither Weibull nor the exponential models can be fixed to the survival data.

IV PREDICTION MODEL FOR MONTHLY ATTRITION

This chapter deals with the study of multiple regression models to predict monthly attrition. Monthly attrition is the dependent variable. The potential independent variables for predicting monthly attrition are: monthly attrition in the three previous months, the number of enlisted personnel who joined the CG four years ago and twenty years ago, respectively, seasonality, monthly unemployment rate and time.

A. MULTIPLE REGRESSION MODEL

The reason for using regression models is the desire to forecast. A forecast is a quantitative estimate (or set of estimates) about the likelihood of future events based on past and current information. Two types of forecasting can be applied: the point forecast, that predicts a single number in each forecast period; and the interval forecast, that indicates an interval in which the realized value will lie.

The variables in the following list were used as a set of candidate predictors to forecast monthly attrition (which will be denoted CONTEO).

EXPLANATORY VARIABLES:

YT_1 number of attrition in the previous month

YT_2 number of attrition two months ago

YT_3 number of attrition three months ago

XT_48 number of enlisted who entered four year ago

XT_240 number of enlisted who entered twenty years ago

MM time period starting from Oct/82

TS square term of MM

D1 Dummy variable indicating first semester of each FY (Oct - Mar)

D2 Dummy variable indicating second semester of each FY (Apr - Sep)

RATA monthly unemployment rate of U.S.

XT_48 and XT_240 were included because of the prominences in the survival analysis. These two time periods (48 and 240) showed a significant change in the survival probabilities. The variables MM and TS were used to consider the observational period time trend in the attrition behavior. Two dummy variables, D1 and D2, were included to distinguish the potential difference in attrition behavior in two different semesters. Finally the monthly unemployment rate of U.S. was added to relate the attrition behavior to the condition of U.S. economy. These rates were taken from the Monthly Labor Review, a publication of the US Department of Labor and Bureau of Labor Statistics. Based on these, the initial model considered was as follow:

$$CONTEO= \beta_1 YT\_1+\beta_2 YT\_2+\beta_3 YT-3+\beta_4 XT\_48+\beta_5 XT\_240+\beta_6 MM+\beta_7 TS+\beta_8 D1+\epsilon$$

Note that this model does not contain an intercept term. Using a backward elimination option of the PROC REG procedure of the statistical package SAS the following three models were selected to predict the monthly attrition.

MODEL 1 CONTEO = YT_1 XT_48 MM TS RATA

MODEL 2 CONTEO = YT_1 XT_48 RATA

MODEL 3 CONTEO = YT_1 YT_2

Details are in Appendix B.

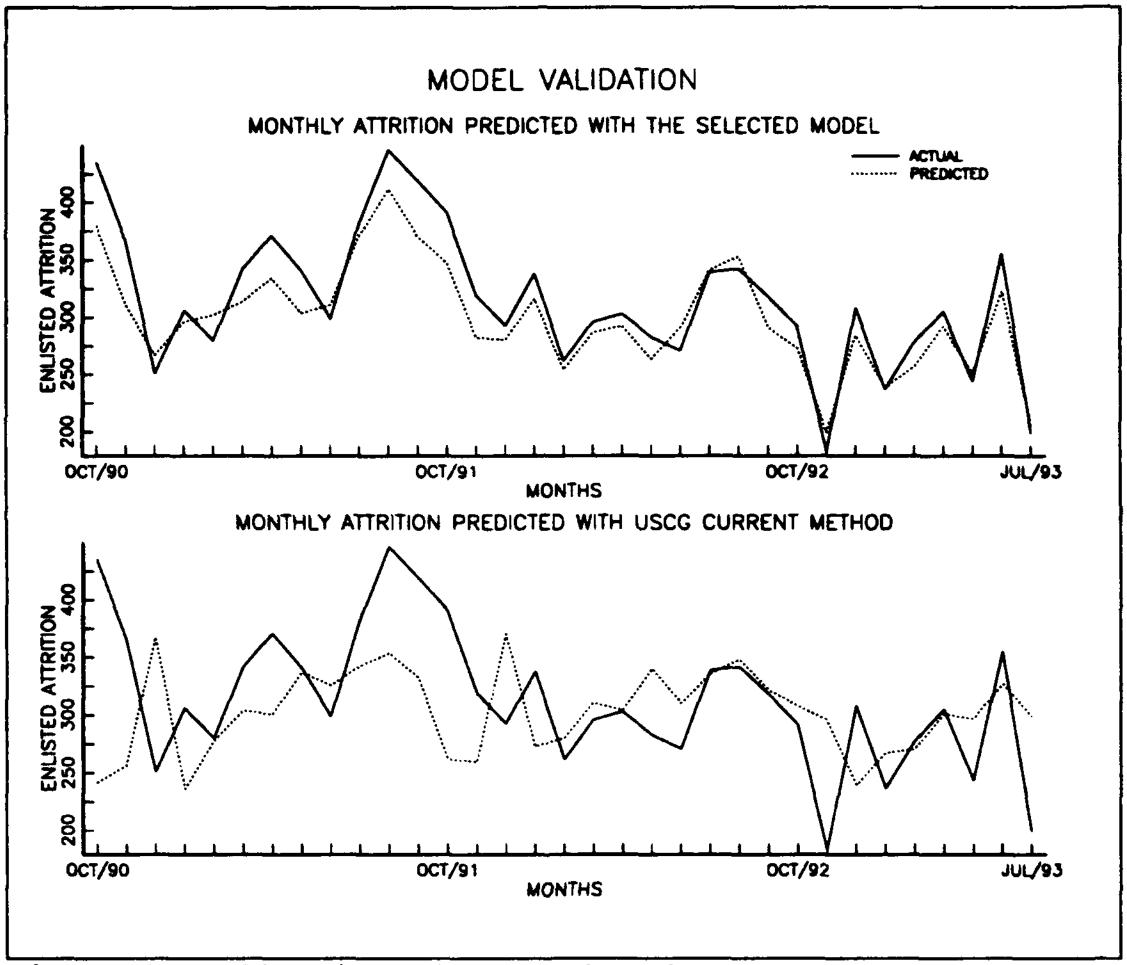

Figure 8. Validation of model 2 and current CG method.

The predictors variables in the models are significant at level of 5 % and the R^2 for all models is greater than 0.9569.

In order to fit the model, the 96 observations (one for each month in eight fiscal years from Oct/82 to Sep/90) were used while 33 observations (Oct/90 to Jun/93) were employed to validate the model. As performance criteria, the mean squared error (MSE) and mean relative error (MRE) were used.

The predicted attrition values of each true monthly attrition for FY91, FY92 and the first nine months of FY93 were computed and are in Appendix C. Figure 8 contains the corresponding plots. The mean squared error and the mean relative error of each model are given as follows:

Mean Squared Error and Mean Relative Error for FY91

	Model 1	Model 2	Model 3	CG Method
MSE	1410.166	515.931	817.225	4331.973
MRE	0.30325	0.13904	0.18486	1.65557

Mean Squared Error and Mean Relative Error for FY92

	Model 1	Model 2	Model 3	CG Method
MSE	1286.171	94.482	134.652	1346.880
MRE	0.38381	0.08839	0.10979	1.01372

Mean Squared Error and Mean Relative Error for FY93 (9 months)

	Model 1	Model 2	Model 3	CG Method
MSE	2304.833	505.083	533.362	3586.590
MRE	0.52702	0.27350	0.28616	1.86012

B. RESULTS

The results shown in the previous section indicate that model 2 predicts best the twelve month attrition for FY91, FY92 and the first nine months of FY93 in terms of not only the MSE but also the MRE. In model 2, 96.66% of the total variation in the attrition is explained by the predictor variables such as XT_48, YT_1 and RATA. The correlation among these three predictors are low and the Durbin-Watson statistic is close to two indicating there is no significant first order autocorrelation in the residuals.

In summary, in order to forecast a future monthly attrition one can use:

$$\widehat{CONTEO} = 0.674663 * YT\_1 + 0.170029 * XT\_48 + 6.996215 * \widehat{RATA}$$

When YT_1 and RATA are unknown at the time of forecast, predicted values of $\widehat{YT}\_1$ and RATA can replace actual YT_1 and RATA.

V. CONCLUSION AND RECOMMENDATIONS

A. CONCLUSIONS

The objective of this thesis was to analyze the attrition behavior and to predict monthly enlisted attrition of the US Coast Guard.

First, survival analysis was used to investigate the attrition behavior of a USCG enlisted in terms of individual characteristics such as sex, marital status, rating, race and paygrade. Results obtained based on the past eight years (Oct/82 to Sep/90) data are as follows: males and married individuals have higher survival probabilities than their counterparts, respectively; paygrades E-1 to E-5 have higher attrition than paygrades E-6 to E-9; American Indians have the highest attrition and the Asian members the highest survival functions; rating 170 (Gunner's Mate) has the highest attrition over all ratings followed by rating 180 (Fire Control Technician); the rating with the highest survival probability is 570 (Aviation Machinist's Mate); there was a decreasing trend in attrition in the last 4 years of observation period. The common feature of the survival functions was the significant fall at the end of the four years service contract and when the enlisted member reach twenty years of service.

Secondly, a regression model was estimated to forecast the monthly attrition: significant predictors are the prior month's attrition, the number of enlistments four years prior and the current unemployment rate as significant predictors. This model explains almost 97% variation in the monthly attrition. Performance of the regression model turns out to be better than that of the current method used in CG.

B. RECOMMENDATIONS

The goals of this thesis were met. Recommendations for further studies in CG enlisted attrition are listed bellow:

1. Recommendation 1

For future research it will be necessary to have a more manageable enlisted data base in order to facilitate necessary analysis.

2. Recommendation 2

It is recommended to formulate the CG enlisted attrition as a time series model such as moving average models, autoregressive models or a combination of both.

3. Recommendation 3

It is recommended that the forecast model be formulated for each rating.

APPENDIX A

This source file was used to calculate survival functions for each individual characteristics such as sex, marital status, race, paygrade and rating.

```
OPTION LINESIZE=80;
DATA ONE; SET HELEN.HELEN;
   CENSORED = (SEPAR > 900930 )      ;
   IF PAY = 1   THEN GROUP = 'P1' ;
   ELSE IF PAY = 2   THEN GROUP = 'P2' ;
   ELSE IF PAY = 3   THEN GROUP = 'P3' ;
   ELSE IF PAY = 4   THEN GROUP = 'P4' ;
   ELSE IF PAY = 5   THEN GROUP = 'P5' ;
   ELSE IF PAY = 6   THEN GROUP = 'P6' ;
   ELSE IF PAY = 7   THEN GROUP = 'P7' ;
   ELSE IF PAY = 8   THEN GROUP = 'P8' ;
   ELSE GROUP = 'P9' ;
PROC LIFETEST OUTSURV=OUT1 PLOT=(S) NOTABLE;
   TIME MY*CENSORED(1);
   STRATA GROUP ;
DATA DOS ; SET OUT1 ;
DROP GROUP SDF_LCL SDF_UCL          ;
IF _CENSOR_ = 1 THEN DELETE ;
DATA TRES ; SET DOS(DROP= _CENSOR_) ;
PROC PRINT ;
```

APPENDIX B

This is the source file to estimate and validate the 3 regression models.

```
OPTION LINESIZE=80;
DATA ONE;
INPUT MM CONTEO XT_48 RATA D1 D2 ;
CARDS;
;
DATA DOS; SET ONE;
YT_1 = LAG1(CONTEO) ;
YT_2 = LAG2(CONTEO) ;
TS = MM * MM ;
IF MM < 97 ;
   PROC CORR;
   PROC REG ;
   MODEL CONTEO =  YT_1 YT_2 XT_48 MM TS RATA / NOINT DW ;
   PROC REG ;
   MODEL CONTEO =  YT_1 XT_48 MM TS RATA / NOINT DW ;
   PROC REG ;
   MODEL CONTEO =  YT_1 XT_48 MM RATA / NOINT DW ;
   PROC REG ;
   MODEL CONTEO =  YT_1 XT_48 RATA / NOINT DW ;
   PROC REG ;
   MODEL CONTEO =  YT_1 YT_2 / NOINT DW ;
```

```
DATA NEW ; SET DOS ;
DROP YT_2 ;
IF MM > 108 AND MM < 121 ;
NEWM1 = 0.558604*YT_1 + 0.162241*XT_48 + 9.085655*RATA +
2.094146*MM - 0.025143*TS ;
SQ1 = (CONTEO - NEWM1)**2 ;
RE1 = ABS(CONTEO - NEWM1) / CONTEO ;
NEWM2 = 0.674663*YT_1 + 0.170029*XT_48 + 6.996215*RATA  ;
SQ2 = (CONTEO - NEWM2)**2 ;
RE2 = ABS(CONTEO - NEWM2) / CONTEO ;
NEWM4 = 0.785271*YT_1 + 0.193608*YT_2 ;
SQ4 = (CONTEO - NEWM4)**2 ;
RE4 = ABS(CONTEO - NEWM4) / CONTEO ;
DATA NEW1 ; SET NEW ;
DROP MM YT_1 XT_48 TS D1 D2 RATA  ;
PROC SUMMARY ; VAR SQ1 SQ2 SQ4 RE1 RE2 RE4 ;
OUTPUT  OUT = OUT1 MEAN = MSQ1 MSQ2 MSQ4 MRE1 MRE2 MRE4 ;
PROC PRINT DATA = OUT1 ;
PROC PRINT DATA = NEW1;
```

Correlation Analysis

9 'VAR' Variables: MM CONTEO XT_48 RATA D1 D2 YT_1 YT_2 TS

Simple Statistics

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
MM	96	48.50000	27.85678	4656	1.00000	96.00000
CONTEO	96	306.29167	106.72453	29404	12.00000	546.00000
XT_48	96	303.02083	131.30767	29090	38.00000	649.00000
RATA	96	6.87917	1.56937	660.40000	5.00000	10.90000
D1	96	0.51042	0.50252	49.00000	0	1.00000
D2	96	0	0	0	0	0
YT_1	95	307.90526	106.10694	29251	12.00000	546.00000
YT_2	94	309.31915	105.77235	29076	12.00000	546.00000
TS	96	3120	2789	299536	1.00000	9216

Correlation Analysis

Pearson Correlation Coefficients / Prob > !R! under Ho: Rho=0 / Number of Observations

	MM	CONTEO	XT_48	RATA	D1
MM	1.00000 0.0 96	-0.61797 0.0001 96	-0.41096 0.0001 96	-0.92129 0.0001 96	-0.10791 0.2953 96
CONTEO	-0.61797 0.0001 96	1.00000 0.0 96	0.53645 0.0001 96	0.44194 0.0001 96	-0.24383 0.0167 96
XT_48	-0.41096 0.0001 96	0.53645 0.0001 96	1.00000 0.0 96	0.32342 0.0013 96	-0.14932 0.1465 96
RATA	-0.92129 0.0001 96	0.44194 0.0001 96	0.32342 0.0013 96	1.00000 0.0 96	0.12041 0.2426 96
D1	-0.10791 0.2953 96	-0.24383 0.0167 96	-0.14932 0.1465 96	0.12041 0.2426 96	1.00000 0.0 96
YT_1	-0.60800 0.0001 95	0.81003 0.0001 95	0.43099 0.0001 95	0.42469 0.0001 95	-0.26260 0.0101 95
YT_2	-0.59948 0.0001 94	0.69724 0.0001 94	0.42722 0.0001 94	0.39914 0.0001 94	-0.20203 0.0509 94
TS	0.96888 0.0001 96	-0.69331 0.0001 96	-0.43166 0.0001 96	-0.82029 0.0001 96	-0.09517 0.3563 96

	D2	YT_1	YT_2	TS
MM	.	-0.60800	-0.59948	0.96888
		0.0001	0.0001	0.0001
	96	95	94	96
CONTEO	.	0.81003	0.69724	-0.69331
		0.0001	0.0001	0.0001
	96	95	94	96
XT_48	.	0.43099	0.42722	-0.43166
		0.0001	0.0001	0.0001
	96	95	94	96
RATA	.	0.42469	0.39914	-0.82029
		0.0001	0.0001	0.0001
	96	95	94	96
D1	.	-0.26260	-0.20203	-0.09517
		0.0101	0.0509	0.3563
	96	95	94	96
YT_1	.	1.00000	0.80644	-0.68380
		0.0	0.0001	0.0001
	95	95	94	95
YT_2	.	0.80644	1.00000	-0.67642
		0.0001	0.0	0.0001
	94	94	94	94
TS	.	-0.68380	-0.67642	1.00000
		0.0001	0.0001	0.0
	96	95	94	96

Model: **MODEL CONTEO = f(XT_48, RATA, YT_1, YT_2, TS, MM)**
NOTE: No intercept in model. R-square is redefined. Dependent Variable: CONTEO

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	6	9611402.0244	1601900.3374	485.667	0.0001
Error	88	290254.97562	3298.35200		
U Total	94	9901657			

Root MSE	57.43128	R-square	0.9707
Dep Mean	306.35106	Adj R-sq	0.9687
C.V.	18.74688		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > !T!
YT_1	1	0.569087	0.10135072	5.615	0.0001
YT_2	1	-0.009225	0.10036742	-0.092	0.9270
XT_48	1	0.164531	0.05108831	3.221	0.0018
MM	1	2.209609	0.97305540	2.271	0.0256
TS	1	-0.026221	0.01081079	-2.425	0.0173
RATA	1	8.529324	3.38234431	2.522	0.0135

Durbin-Watson D	1.961
(For Number of Obs.)	94
1st Order Autocorrelation	0.009

Model: **MODEL CONTEO = f(XT_48, RATA, YT_1, TS, MM)**
NOTE: No intercept in model. R-square is redefined. Dependent Variable: CONTEO

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	5	9723932.3348	1944786.467	599.485	0.0001
Error	90	291968.66517	3244.09628		
U Total	95	10015901			

Root MSE	56.95697	R-square	0.9708
Dep Mean	306.68421	Adj R-sq	0.9692
C.V.	18.57186		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > !T!
YT_1	1	0.558604	0.07950773	7.026	0.0001
XT_48	1	0.162241	0.05018357	3.233	0.0017
MM	1	2.094146	0.90905311	2.304	0.0235
TS	1	-0.025143	0.01015032	-2.477	0.0151
RATA	1	9.085655	3.13547291	2.898	0.0047

Durbin-Watson D	1.979
(For Number of Obs.)	95
1st Order Autocorrelation	0.007

Model: **MODEL CONTEO = f(XT_48, RATA, YT_1, MM)**
NOTE: No intercept in model. R-square is redefined. Dependent Variable: CONTEO

Analysis of Variance

Source	**DF**	**Sum of Squares**	**Mean Square**	**F Value**	**Prob>F**
Model	4	9704026.5996	2426006.6499	707.870	0.0001
Error	91	311874.40041	3427.19121		
U Total	95	10015901			

Root MSE	58.54222	R-square	0.9689
Dep Mean	306.68421	Adj R-sq	0.9675
C.V.	19.08876		

Parameter Estimates

Variable	**DF**	**Parameter Estimate**	**Standard Error**	**T for H0: Parameter=0**	**Prob > !T!**
YT_1	1	0.675515	0.06576463	10.272	0.0001
XT_48	1	0.173063	0.05138446	3.368	0.0011
MM	1	-0.122828	0.16371164	-0.750	0.4550
RATA	1	7.569406	3.16073313	2.395	0.0187

Durbin-Watson D	2.087
(For Number of Obs.)	95
1st Order Autocorrelation	-0.050

Model: **MODEL CONTEO = f(XT_48, RATA, YT_1)**
NOTE: No intercept in model. R-square is redefined. Dependent Variable: CONTEO

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	3	9702097.409	3234032.4697	948.144	0.0001
Error	92	313803.59101	3410.90860		
U Total	95	10015901			

Root MSE	58.40298	R-square	0.9687
Dep Mean	306.68421	Adj R-sq	0.9676
C.V.	19.04336		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > !T!
YT_1	1	0.674663	0.06559844	10.285	0.0001
XT_48	1	0.170029	0.05110324	3.327	0.0013
RATA	1	6.996215	3.05971911	2.287	0.0245

Durbin-Watson D	2.072
(For Number of Obs.)	95
1st Order Autocorrelation	-0.044

Model: **MODEL** **CONTEO = f(YT_1, YT_2)**
NOTE: No intercept in model. R-square is redefined. Dependent Variable: CONTEO

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	2	9521013.4894	4760506.7447	1150.595	0.0001
Error	92	380643.51056	4137.42946		
U Total	94	9901657			

Root MSE	64.32285	R-square	0.9616
Dep Mean	306.35106	Adj R-sq	0.9607
C.V.	20.99645		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > !T!
YT_1	1	0.785271	0.10156464	7.732	0.0001
YT_2	1	0.193608	0.10135320	1.910	0.0592

Durbin-Watson D	2.003
(For Number of Obs.)	94
1st Order Autocorrelation	-0.012

APPENDIX C

True and fitted attrition calculated with the selected model and the current method used by USCG.

	True Attrition	Predicted Attrition Model 2	Predicted Attrition Current Method
FY91			
Oct/90	435	379.454	241.750
Nov/90	366	310.328	256.250
Dec/90	251	266.937	367.500
Jan/91	306	296.300	236.250
Feb/91	280	302.150	277.750
Mar/91	342	313.636	303.750
Apr/91	371	334.110	300.500
May/91	341	303.537	337.250
Jun/91	299	310.849	325.750
Jul/91	383	371.072	342.875
Aug/91	446	411.705	353.875
Sep/91	419	370.354	332.375
FY92			
Oct/91	391	347.414	262.500
Nov/91	319	282.724	259.750
Dec/91	293	280.335	370.500
Jan/92	338	316.155	273.000
Feb/92	262	254.189	280.250
Mar/92	296	286.479	310.625
Apr/92	303	292.431	304.000
May/92	283	263.674	340.500
Jun/92	271	290.604	310.125
Jul/92	339	340.863	336.125
Aug/92	342	352.768	348.375
Sep/92	318	291.120	321.625
FY93 (nine months)			
Oct/92	292	273.408	307.875
Nov/92	183	199.190	295.875
Dec/92	308	284.484	239.750
Jan/93	237	238.624	267.625
Feb/93	278	256.763	270.875
Mar/93	305	291.982	300.875
Apr/93	244	250.998	296.875
May/93	355	322.655	327.125
Jun/93	199	204.822	299.250

LIST OF REFERENCES

1. Douglas Allen Blakemore, Predicted Coast Guard Enlisted Attrition, Master's Thesis, Naval Postgraduate School, Monterey, California, March 1992.

2.A.A. Afifi and Virginia Clark, Computed-Aided Multivariate Analysis, Second Edition, Van Nostrand Reinhold.

3.SAS Institute Inc., SAS/STAT User's Guide, Version 6, Vol 2, NC:SAS Institute Inc., 1988.

INITIAL DISTRIBUTION LIST

		No. Copies
1.	Defense Technical Information Center Cameron Station Alexandria VA 22304-6145	2
2.	Library, Code 052 Naval Postgraduate School Monterey CA 93943-5002	2
3.	Professor So Young Shon, Code OR/Sh Naval Posgraduate School Monterey CA 93940	2
4.	Professor Robert R. Read, Code OR/Rd Naval Postgraduate School Monterey CA 93940	2
5.	USCG Personnel Workforce Planning Office Washington D.C.	2
6.	LCDR Enrique Oñate Rubiano A.A. 5204 Cartagena Colombia	2