NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

IDENTIFYING U.S. MARINE CORPS RECRUIT CHARACTERISTICS THAT CORRESPOND TO SUCCESS IN SPECIFIC OCCUPATIONAL FIELDS

by

Ben E. McCaleb III

June 2016

Thesis Advisor:	Robert A. Koyak
Second Reader:	Chad W. Seagren

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			*Form Approved OMB No. 0704–0188*
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington, DC 20503.			
1. AGENCY USE ONLY ***(Leave blank)***	**2. REPORT DATE** June 2016	**3. REPORT TYPE AND DATES COVERED** Master's Thesis	
4. TITLE AND SUBTITLE IDENTIFYING U.S. MARINE CORPS RECRUIT CHARACTERISTICS THAT CORRESPOND TO SUCCESS IN SPECIFIC OCCUPATIONAL FIELDS			**5. FUNDING NUMBERS**
6. AUTHOR(S) Ben E. McCaleb III			
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) N/A			**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. IRB Protocol number ____NPS.2016.0008-IR-EP5-A____.			
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited			**12b. DISTRIBUTION CODE**

13. ABSTRACT (maximum 200 words)

This thesis investigates how Marine recruit information available at entry can be used to predict which occupational field (OCCFLD) is best suited to an individual and if a Marine successfully completes the first term of enlistment. Multinomial regression models are developed to calculate estimated probabilities that a given recruit will attain United States Marine Corps (USMC) Computed Reenlistment Tiers I, II, III, or IV in a particular OCCFLD. Optimization of OCCFLD assignment based on the developed models illustrates the potential value of insight gained from recruit information available prior to enlistment.

The relationship of recruit characteristics available prior to enlistment and the USMC Computed Tier Score assigned in the last year of a Marine's first enlistment is dependent upon the OCCFLD assigned. We recommend identifying OCCFLDs with the highest estimated probabilities of Tier I or Tier II attainment at the recruitment phase. Providing recruits and recruiters a tool that provides estimated probabilities of attaining Tier I or Tier II in descending order for each OCCFLD during initial assignment has the potential to increase the caliber of Marines across all OCCFLDs and to aid in assessing the current OCCFLD assignment practices.

14. SUBJECT TERMS assignment, ASVAB, first term enlistment, MOS, Multinomial Elastinet Regression, OCCFLD			**15. NUMBER OF PAGES** 77
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified	**18. SECURITY CLASSIFICATION OF THIS PAGE** Unclassified	**19. SECURITY CLASSIFICATION OF ABSTRACT** Unclassified	**20. LIMITATION OF ABSTRACT** UU

NSN 7540–01-280-5500

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

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release; distribution is unlimited

IDENTIFYING U.S. MARINE CORPS RECRUIT CHARACTERISTICS THAT CORRESPOND TO SUCCESS IN SPECIFIC OCCUPATIONAL FIELDS

Ben E. McCaleb III
Captain, United States Marine Corps
B.S., Texas A&M University, 2010

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
June 2016

Approved by: Robert A. Koyak
Thesis Advisor

Chad W. Seagren
Second Reader

Patricia A. Jacobs
Chair, Department of Operations Research

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

This thesis investigates how Marine recruit information available at entry can be used to predict which occupational field (OCCFLD) is best suited to an individual and if a Marine successfully completes the first term of enlistment. Multinomial regression models are developed to calculate estimated probabilities that a given recruit will attain United States Marine Corps (USMC) Computed Reenlistment Tiers I, II, III, or IV in a particular OCCFLD. Optimization of OCCFLD assignment based on the developed models illustrates the potential value of insight gained from recruit information available prior to enlistment.

The relationship of recruit characteristics available prior to enlistment and the USMC Computed Tier Score assigned in the last year of a Marine's first enlistment is dependent upon the OCCFLD assigned. We recommend identifying OCCFLDs with the highest estimated probabilities of Tier I or Tier II attainment at the recruitment phase. Providing recruits and recruiters a tool that provides estimated probabilities of attaining Tier I or Tier II in descending order for each OCCFLD during initial assignment has the potential to increase the caliber of Marines across all OCCFLDs and to aid in assessing the current OCCFLD assignment practices.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

LIST OF FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF ACRONYMS AND ABBREVIATIONS

AFQT	Armed Forces Qualification Test
AI	Auto Information
AO	Assembling Objects
AR	Arithmetic Reasoning
AS	Auto and Shop Information
ASVAB	Armed Services Vocational Aptitude Battery
BMI	Body Mass Index
CDF	Cumulative Density Function
CFT	Combat Fitness Test
CL	Clerical
CDC	Centers for Disease Control
CMC	Commandant of the Marine Corps
CNA	Center for Naval Analyses
DOD	Department of Defense
EI	Electronic Information
EL	Electronics
FITREP	Fitness Report
FY	Fiscal Year
GLM	Generalized Linear Model
GS	General Science
GT	General Technical
HQMC	Headquarters Marine Corps
IST	Initial Strength Test
LAV	Light Armored Vehicle
M&RA	Manpower and Reserve Affairs
MAGTF	Marine Air Ground Task Force
MC	Mechanical Comprehension
MCMAP	Marine Corps Martial Arts Program
MCO	Marine Corps Order
MCRC	Marine Corps Recruiting Command

MCRISS	Marine Corps Recruiting Information Support System
MEPS	Military Entrance Processing Station
MK	Mechanical Knowledge
MM	Mechanical Maintenance
MNGLM	Multinomial Generalized Linear Regression Model
MOS	Military Occupational Specialty
NAVMC	Navy Marine Corps
NJP	Non-Judicial Punishment
OCCFLD	Occupational Field
OLS	Ordinary Least Squares
PC	Paragraph Comprehension
PEF	Program Enlisted For
PFT	Physical Fitness Test
PII	Personally Identifiable Information
SI	Shop Information
TFDW	Total Forces Data Warehouse
USMC	United States Marine Corps
WK	Word Knowledge

EXECUTIVE SUMMARY

The 35th Commandant of the Marine Corps (CMC), General James Amos, stated in his 2010 Commandant's Planning Guidance that, "The goal of retention is to retain the most qualified instead of the 'first to volunteer,' while meeting manpower requirements goals" (Commandant of the Marine Corps, 2010, p. 14). While the Marine Corps acknowledges the need to retain the "most qualified" Marines, the reenlistment screening process can only retain the highest caliber Marines within a specific reenlistment year. Once a Marine reaches the end of his or her first term of enlistment, the scores used during the reenlistment screening are already a matter of record. Identifying the occupation for which a recruit is best suited prior to enlistment has the potential to increase the caliber of Marines across all occupational fields (OCCFLDs), thus enabling the retention of the most qualified.

Solutions addressing the CMC's directive should address the initial placement of recruits in appropriate occupational fields (OCCFLDs) to maximize recruits' potential for success. An analytical approach that identifies occupations for which a Marine is most likely to succeed would enable the Marine Corps to proactively identify the "most qualified" Marines for each occupation upon enlistment. Assigning recruits to OCCFLDs in which they are more likely to succeed improves not only the individual Marine's enlisted experience, but also the quality of the specific OCCFLD assigned and the United States Marine Corps (USMC) as a whole.

This study provides evidence that the relationship of recruit characteristics available prior to enlistment and the USMC Computed Tier Score is dependent upon the OCCFLD assigned. Individual statistical models are constructed for each OCCFLD with the recruit characteristics used as predictors. Multinomial regression provides estimated probabilities that a given recruit attains Tiers I, II, III, or IV in each OCCFLD with a Marine attaining either a Tier I or Tier II categorization of the USMC Computed Tier Score defined as a successful first term enlistment. These models are applied to all Marines in the study for each of the 38 OCCFLDs. Optimization of OCCFLD assignment based on the developed models illustrates the potential insight provided by recruit

information available prior to enlistment and results in reassigning approximately 91 percent of the Marines in our study to an OCCFLD other than the Marine's actual OCCFLD.

The contribution of predictor variables based on recruit information available prior to enlistment is dependent upon the OCCFLD. Identifying OCCFLDs with the greatest estimated probabilities of a successful first term enlistment takes a proactive approach to increasing the caliber of Marines assigned to each OCCFLD. The number of Marines in each Tier level will not change due to the normalized percentile definition within MOSs, or OCCFLDs as in this study, but the potential for increasing Marine performance levels required for obtaining each Tier exists. Consequently, Headquarters Marine Corps (HQMC) screening of the most qualified individual should take place prior to the reenlistment screening process, ideally during initial assignment.

This study is meant to assist USMC Manpower Studies & Analysis Branch and recruiters in better assigning recruits to OCCFLDs in which they will succeed. The information provided by the models is intended to become a new tool recruiters use in addition to those currently used, ultimately benefiting the Marine Corps and future Marines by identifying the OCCFLDs that maximize the estimated probability of a successful first term of enlistment for each new recruit.

References

Commandant of the Marine Corps. (2010). *35th Commandant of the Marine Corps commandant's planning guidance*. Washington, DC: Headquarters United States Marine Corps.

ACKNOWLEDGMENTS

I would like to thank Dr. Robert Koyak for his support and patience throughout my thesis process. Dr. Koyak's knowledge, mentorship, and unrelenting commitment to my success were instrumental in all aspects of my work; without his assistance, this thesis would not have been possible. I would also like to thank Dr. Chad Seagren for his assistance and time as a second reader and advisor.

Additionally, the Marines and civilians at HQMC Manpower & Reserve Affairs (M&RA) were essential to my success. Captain William Wathen's and Mr. Timothy Johnson's assistance with my data requirements from the Total Force Data Warehouse (TFDW) and the Marine Corps Recruiting Information Support System (MCRISS) ensured I received all requested information in a timely manner. Throughout the thesis process, M&RA continued their support and encouragement.

Finally, I want to thank my family for their undying support and love throughout my thesis process, Naval Postgraduate School attendance, and entire military career. Despite the trials I experience in my life, I can always be sure that I have the love and support of my wife, Caroline, daughter, Kailey, and son, Ben. Their support is truly the reason for my success.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

The 35th Commandant of the Marine Corps (CMC), General James Amos, stated in his 2010 Commandant's Planning Guidance that, "the goal of retention is to retain the most qualified instead of the 'first to volunteer,' while meeting manpower requirements goals" (Commandant of the Marine Corps, 2010, p. 14). While the Marine Corps acknowledges the need to retain the "most qualified" Marines, the reenlistment screening process can only retain the highest caliber Marines within a specific reenlistment year. Once a Marine reaches the end of his or her first term of enlistment, the scores used during the reenlistment screening are already a matter of record. Identifying the occupation for which a recruit is best suited prior to enlistment has the potential to increase the caliber of Marines across all occupational fields (OCCFLDs), thus enabling the retention of the most qualified.

Solutions addressing the CMC's directive should address the initial placement of recruits in appropriate OCCFLDs to maximize recruits' potential for success. An analytical approach that identifies occupations for which a Marine is most likely to succeed would enable the Marine Corps to proactively identify the "most qualified" Marines for each occupation upon enlistment. The Marine Corps Manpower and Reserve Affairs (M&RA) command is investigating the use of recruit characteristics available prior to enlistment to use as predictors of first term enlistment success within specific occupational fields. An OCCFLD is a grouping of similar military occupational specialties (MOS). The 03 Infantry OCCFLD encompasses the MOSs 0311 Rifleman, 0312 Light Armored Vehicle (LAV) Crewman, 0321 Reconnaissance Man, 0331 Machine Gunner, 0341 Mortarman, 0351 Infantry Assaultman, and 0352 Antitank Missileman. Assigning recruits to OCCFLDs in which they are more likely to succeed improves not only the individual Marine's enlisted experience, but also the quality of the specific OCCFLD assigned and the United States Marine Corps (USMC) as a whole.

Under the current assignment process, recruits are assigned to OCCFLDs based on minimum entrance criteria specific to each MOS, the recruit's preference, and the needs of the Marine Corps. This approach enables the Marine Corps to identify recruits that meet the minimum requirements, but it does not provide insight into which MOS or OCCFLD the recruit would have the highest probability of success. Our objective is to provide the Marine Corps with tools to better understand the correlation between a recruit's quantifiable characteristics available prior to enlistment and metrics of a successful first term enlistment in specific OCCFLDs.

The Marine Corps can incorporate insights gained from our approach to better utilize human capital in the current MOS and OCCFLD assignment process. Our research is not espousing alteration of current entrance criteria for MOSs. Instead, our goal is to provide the Marine Corps an approach to assess the effectiveness of its assignment process.

A. SCOPE AND LIMITATIONS

The purpose of this study is to demonstrate how information obtained from a Marine recruit at entry can be used to predict which OCCFLD is best suited to that person. As additional information about recruits becomes available in the future, (e.g., psychometrics), this approach may serve as a template for evaluating improvement in prediction. The predictive analysis discussed in this thesis is intended to assist recruiting and manpower personnel in the evaluation of the current OCCFLD assignment process to benefit both the Marine Corps and the individual Marine.

The USMC's Total Force Data Warehouse (TFDW) and Marine Corps Recruiting Information Support System (MCRISS) provide data on active-duty Marines. Our focus is on those who entered service in fiscal years (FY) 2008 through 2011. These four years are the most recent period when relationships between recruit characteristics, OCCFLD assignment, and effectiveness metrics are expected to be stable. Recruits who entered the Marine Corps after FY 2011 do not have sufficient time in their assigned OCCFLD for effectiveness metrics to be calculated. Recruits enlisting prior to FY 2008 did so during a time of significant engagement by the USMC in theaters of conflict, potentially creating incongruities relative to later enlistees.

Our investigation begins with an exploratory data analysis phase in which incomplete or invalid data are removed and relationships between potential predictor variables are identified. Then a generalized linear model (GLM) is fit to the remaining data to predict the outcome variable, which in our study is the USMC Computed Tier Score. Every first-term Marine has a Computed Tier Score, which is based on assessments accrued throughout a Marine's enlistment regardless of intent to reenlist. Marines are categorized into four groups based on their Computed Tier Scores: Tier I (top ten percent), Tier II (between the 60th and 90th percentiles), Tier III (between the 10th and 60th percentiles), and Tier IV (below the 10th percentile). The Tier assignment facilitates normalization within an OCCFLD and comparison across OCCFLDs, making it suitable as an outcome variable in a statistical analysis.

This study addresses the following questions:

1. How can information obtained from a Marine recruit at entry be used to predict which OCCFLD is best suited to that individual?

2. How can information obtained from a Marine recruit at entry be used to predict if a Marine successfully completes the first term of enlistment?

Determining if a recruit is placed in the appropriate MOS and corresponding OCCFLD is difficult for several reasons. First, there are no systematically collected metrics that attempt to directly measure the quality of a Marine's placement. Available metrics that are often used—such as the time required for promotion to a certain paygrade, proficiency and conduct marks, and fitness report (FITREP) scores—do not directly measure how well a Marine is matched to the requirements of a particular MOS or OCCFLD. Additionally, the rate at which a Marine is promoted and the marks he or she receives depend on the MOS and the Marine supervisor's subjective assessment, thus preventing the direct comparison of Marines from different MOSs.

Second, there is no way to observe how well a Marine would have performed in an alternate assignment. A controlled experiment that assigns recruits of similar qualities to random OCCFLDs to compare the individual outcomes is conceptually impossible. A feasible design, however, is to compare Marines with similar entry qualities that are assigned different OCCFLDs under the current assignment methodology. Although it remains an observational study, control is exercised over important factors that otherwise would be confounding, either through statistical modeling or through the use of case-control methodology.

Finally, this study is limited in scope to the four years spanning FY 2008 to 2011. During those years, the Marine Corps began to drawdown from a manpower level of 202,000 during Operation Enduring Freedom and Operation Iraqi Freedom to a post-conflict level of 184,000 (Gibson, 2016). The number of available reenlistment slots was affected, though not uniformly across years or OCCFLDs. This variability was realized prior to the study and we proceeded with the understanding that such variability across OCCFLDs and years is present during any period of military activity.

B. ORGANIZATION OF THE THESIS

The organization of the thesis is as follows: Chapter I gives an introduction to the thesis topic, motivation for conducting the study, the study's scope, and limitations. This thesis builds upon and compares several other published efforts on the assignment of military personnel to occupational specialties. Chapter II discusses these previous papers in the literature review and details the current enlisted Marine MOS assignment's background. Chapter III describes the TFDW and MCRISS data along with the methodology used to conduct the study. This chapter also includes a description of the data collection and formatting process. Chapter IV describes the study's results and analysis. Chapter V provides a summary of the study, conclusions, and recommendations for future work.

II. BACKGROUND

The USMC assessed approximately 30,000 to 35,000 new recruits each year in FY 2008 to FY 2011 to maintain congressionally mandated manpower levels based on data drawn from TFDW. Recruits are assigned a Program Enlisted for (PEF) based on the recruit's preferences and aptitude, as well as the needs of the Marine Corps. A PEF is a grouping of similar MOSs from which an OCCFLD and MOS are later selected during the Marine's basic training. A Marine's aptitude is measured by the Armed Services Vocational Aptitude Battery (ASVAB). ASVAB scores are determining factors in the Marine's PEF assignment and significant in the Marine's potential for success during the first term of enlistment and future USMC career. The following section describes the ASVAB in greater detail.

A. ARMED SERVICES VOCATIONAL APTITUDE BATTERY (ASVAB)

ASVAB tests were introduced in 1968 as part of Armed Services' Student Testing Program and adopted by the Department of Defense (DOD) in 1974 for enlistee screening and OCCFLD assignment. To date, over 40 million tests have been administered ("Official Site of the ASVAB," n.d.). Tests are taken at a Military Entrance Processing Station (MEPS) and subdivided into ten subtests measuring four aptitude domains shown in Table 1.

Table 1. ASVAB Subtest Names, Descriptions, and Aptitude Domains

Test	Description	Domain
General Science (GS)	Knowledge of physical and biological sciences	Science/ Technical
Arithmetic Reasoning (AR)	Ability to solve arithmetic word problems	Math
Word Knowledge (WK)	Ability to select the correct meaning of a word presented in context and to identify the best synonym for a given word	Verbal
Paragraph Comprehension (PC)	Ability to obtain information from written passages	Verbal
Mathematics Knowledge (MK)	Knowledge of high school mathematics principles	Math
Electronics Information (EI)	Knowledge of electricity and electronics	Science/ Technical
†Auto Information (AI)	Knowledge of automobilie technology	Science/ Technical
†Shop Information (SI)	Knowledge of tools and shop terminology and practices	Science/ Technical
Mechanical Comprehension (MC)	Knowledge of mechanical and physical principles	Science/ Technical
Assembling Objects (AO)	Ability to determine how an object will look when its pars are put together	Science/ Technical

†AI and SI are administered as separate tests, but combined into one single score (labeled AS).

Source: "Official Site of the ASVAB" (n.d.)

The ASVAB subtest scores are combined to calculate the total ASVAB raw score used to place the recruit into a PEF. The tests are tailored to an individual's ability as subsequent questions increase or decrease in difficulty dependent upon the individual's answer to the previous question. The ASVAB's final score is based on percentiles ranging from 1–99.

Currently, the ASVAB subtests are administered through a computer interface. Answers provided during the test determine the subsequent question; a correct answer results in an increase in difficulty, while an incorrect answer results in a decrease in difficulty. The total ASVAB raw score is compared to all ASVAB tests taken in 1997 by potential enlistees between the ages of 18 to 23. The corresponding percentile is the individual's final ASVAB score ("Official Site of the ASVAB," n.d.).

A partial score computed using the world knowledge (WK), paragraph comprehension (PC), AR, and mechanical knowledge (MK) ASVAB subtests form the Armed Forces Qualification Test (AFQT) that determines an individual's enlistment eligibility. Similar to the final ASVAB score, the AFQT is scored on a 1–99 scale based on percentiles which are then grouped into categories for the purpose of enlistment as shown in Table 2.

Table 2. AFQT Categories Based on Final AFQT Score Percentile

Category	Percentile
I	93-99
II	65-92
IIIA	50-64
IIIB	31-49
IVA	21-30
IVB	16-20
IVC	10-15
V	0-9

Source: "Official Site of the ASVAB" (n.d.)

In the early 1990s, Congress issued two quality controls for first-term, non-prior service enlistees that mandated 90 percent of new recruits be high school diploma graduates and 60 percent be classified AFQT Category I, II, or III (Kapp, 2013, p. 2). Once an enlistee is determined eligible to enlist, a PEF comprising several MOSs is assigned based on requirements for each MOS.

B. MARINE CORPS MILITARY OCCUPATIONAL SPECIALTY (MOS) ASSIGNMENT

The Marine Corps has identified the effective assignment of a USMC recruit to his or her MOS as an important objective. Related MOSs within the active and reserve components of the USMC are grouped into OCCFLDs and allow generalization of over 100 enlisted Marine MOSs. A MOS is identified by a four digit numerical code that corresponds to a Marine's military specialty. The first two MOS digits represent the

Marine's OCCFLD and the second two the military specialty within the broader specialty area. The 38 OCCFLDs considered in this study are shown in Table 3.

Table 3. OCCFLD Codes and Descriptions

OCCFLD	Description	OCCFLD	Description
01	Personnel & Admin	43	Public Affairs
02	Intel	44	Legal
03	Infantry	46	Combat Camera
04	Logistics	48	Recruiting and Retention Specialist
05	MAGTF Plans	55	Music
06	Comm	57	CBRN
08	Arty	58	MP
11	Utilities	59	Electronics Maint
13	Engineer	60	Aircraft Maint
18	Tank and AAV	61	Aircraft Maint
21	Ground Ordnance Maint	62	Aircraft Maint
23	Ammo & Explosive Ord Disposal	63	Avionics
26	SigInt	64	Avionics
28	Data/Comm Maint	65	Aviation Ordnance
30	Supply	66	Avionics Logistics
31	Dist Management	68	METOC
33	Food Service	70	Airfield Services
34	Financial Management	72	Air control/Air Supt/Anti-Air Warfare/ATC
35	Motor Transport	73	Flight Crew

Source: (HQMC, 2015)

The Marine Corps uses four composite scores calculated from an individual's ASVAB subtest scores to partially determine a recruit's ability to meet perquisites prior to MOS assignment. These composite scores include the General Technical (GT), Mechanical Maintenance (MM), Clerical (CL), and Electronics (EL), which are calculated by summing ASVAB subtest scores as shown in Table 4 ("Official Site of the ASVAB," n.d.).

Table 4. Composite Scores Calculated from ASVAB Scores

Composite	Computational Formula†
General Technician (GT)	VE + AR + MC
Mechanical Maintenance (MM)	AR + MC + AS + EI
Clerical (CL)	VE + MK
Electronics (EL)	AR + MK + EI + GS

†VE is a verbal composite formed from an optimally weighted

Source: "Official Site of the ASVAB" (n.d.)

Prerequisites for each MOS are updated annually as directed by Marine Corps Order (MCO) 1200.18 (HQMC, 2014) and delineated in NAVMC 1200.1A (HQMC, 2015). These prerequisites include both minimum composite scores and other required characteristics as shown in Table 5 for a single MOS, 0313, LAV Crewman in the 03 Infantry OCCFLD.

Table 5. USMC MOS 0313 Light Armored Vehicle Crewman Enlistment Prerequisites

Prerequisites
ASVAB GT Score $\geq$ 90
Qualified Basic Infantryman
Normal Color Vision
Vision Correctable to 20/20
Valid Driver's License
65 $\leq$ Height $\leq$ 75

Source: (HQMC, 2015, pp. 361-362)

It is important to understand that MOS prerequisites are minimum requirements, which include both ASVAB and non-ASVAB components; and requirements for each MOS.

C. LITERATURE REVIEW

Several previous studies examine the MOS assignment process and its effect on the manpower community. Of interest and detailed in this section are studies of USMC Field Radio Operator time for promotion and reenlistment eligibility, USMC communications OCCFLD entrance standards, USMC reenlistment incentives and process, and Navy reenlistment incentives and process.

Wathen (2014) examines the factors that predict promotion time to E-4, Corporal, and reenlistment Computed Tier Score for USMC Field Radio Operators. The focus of his thesis is the 0621 MOS, which is under the 06 OCCFLD. The objective of Wathen's study is to identify a statistical relationship between Marine Corps recruits' entry-level attributes and two measures of success: time to promotion and reenlistment. A statistical model is developed using 1,100 Marines entering in FY 2007 through 2014 to identify the most influential entry-level attributes for success.

Wathen (2014) identifies the 1.5 mile run time during the Initial Strength Test (IST), IST crunches, rifle qualification score, ASVAB GT composite score, Marine's weight, and the requirement for a weight waiver during the recruit's enlistment process as the most influential predictors for the Computed Tier Score. He also identifies the 1.5 mile run time during the IST, IST crunches, rifle qualification score, ASVAB general science (GS) score, ASVAB MK score, ASVAB PC score, ASVAB CL composite score, and the requirement for a weight waiver during the recruit's enlistment process as the most influential predictors of the time for promotion to E-4. Wathen concludes that new job performance metrics should be developed and the ASVAB scores identified be included in the Field Radio Operator entrance criteria. This conclusion provides insight into a recruit's potential success in a specific MOS, although the model and final conclusions are based on information not available prior to enlistment, such as the Marine's rifle qualification score.

Rautio (2011) examines the relationship between ASVAB composite scores and successful completion of communications OCCFLD, 06, schools. Similar to Wathen (2014), this study uses a statistical model to predict success, but using only ASVAB subtest scores as predictors and school completion as a success metric. Rautio analyses data from 9,921 Marines entering in FYs 2006 through 2009 to identify ASVAB EL scores as the most significant factor for successful school completion. Rautio also identifies ASVAB CL scores, year of enlistment, whether the recruit attended the 0612 or 0651 MOS schools, and being married, Hispanic, or American Indian as additional positive factors. Although these conclusions reinforce the idea that ASVAB scores have a significant effect on successful completion of 06 OCCFLD schools, ethnic and marital status information is not formally used by the USMC for decision making at the enlistment stage.

Cole (2014) investigates retention of first-term Marines using the Computed Tier Score as a metric of success. Her study uses the Tier Score as an outcome variable similar to Wathen (2014), but is focused on reenlistment and retention rather than assignment. Both studies identify the Composite Tier Score as an objective assessment of a Marine's success at the end of a first term enlistment, while Wathen (2014) and Rautio (2011) identify ASVAB and other pre-enlistment factors as effective predictors of success metrics. Although ASVAB scores are effective in predicting success in early performance, Koopman (2007) finds that the AFQT, which is comprised of ASVAB subtest scores, should be combined with information gathered during enlistment to improve the accuracy of predictions.

Each study discussed above focuses on different aspects of a Marine's or Sailor's enlistment, using different measures of success. Again our research examines variables that each of these studies identify as useful for predicting an enlisted Marine's success at the end of a first term enlistment using only information available prior to enlistment as predictors and Composite Tier Scores as a metric of success.

D. CHAPTER SUMMARY

Marine Corps recruiters working in coordination with M&RA attempt to assign enlisting Marines to OCCFLDs in which they will succeed and to maintain the Marine Corps' warfighting capability at the highest level. The current process uses minimum entrance criteria to assign these enlistees to OCCFLDs and MOSs that are available when the recruit enlists. Although the timing of specific OCCFLD or MOS availability is beyond the scope of this study, the process can be improved through identification of the OCCFLDs in which a recruit is most likely to succeed.

Based on the literature review and initial data analysis, the approach of using ASVAB scores and information available prior to enlistment as predictors with a normalized Computed Tier Score as a success metric is appropriate to identify OCCFLDs with the highest probability of success for an enlisting Marine. This study broadens the scope of earlier studies to include all Marine OCCFLDs, which allows for an informed decision during initial assignment. The intent is to provide all interested parties with the insight that benefits the individual, selected OCCFLD, and Marine Corps as a whole.

III. DATA AND METHODOLOGY

A. DATA AND DATA FORMATING

This section provides an explanation of the process to obtain the data and detailed description of the original personnel records. Chapter III also explains data processing steps undertaken to support our subsequent analysis.

1. Data Summary

The data in our research was obtained from the MCRISS and TFDW databases which are managed by M&RA. The scope of our study is limited to use of personnel records for all Marines entering active duty in FY2007 through FY2011; this captures 137,333 Marines in 237 data fields. Every Marine Corps MOS and OCCFLD is represented in the data. Data for a recruit prior to enlistment was provided by MCRISS and data for the first-enlistment period was provided by TFDW.

Identities of subjects are protected through the use of a non-Personally Identifiable Information (PII) code, which enables the combining of 19 individual files provided by M&RA into one comprehensive database. Duplicates and omissions were removed.

2. Data Formatting

a. Data Consolidation

The multiple MCRISS and TFDW files were merged into a database of 137,333 records, one per Marine recruit. Of these 137,333 records, 27,520 contained erroneous data, and 6,244 had omissions. These incomplete and erroneous records were removed, resulting in a final data set that contains information on 103,569 Marines. We describe the process of merging and filtering the data in the paragraphs below.

Prior to creating the merged database, the files contain variable numbers of observations due to the presence of multiple records for each individual and omissions. Multiple records occur in data bases that are updated to provide a history of events such as promotions during a Marine’s initial enlistment period. Files that contain fewer observations are a result of Marines not recording information in a key data field such as the score on an advanced test that a Marine may not have taken.

We use the following rule for merging datasets: records that do not contain information are recorded as blanks, and only the most recent records are kept when multiple records occur. This approach ensures that the most recent information is used to conduct our analysis. The number of observations remaining corresponds to the number of individual Marines in the study and is verified by unique non-PII codes to be 137,333 Marines. Although this ensures each observation of a Marine only includes a single data point for each data field, errors and omissions remain.

b. Observation Removal and Substitution

Records with missing data fields are either set to the lowest qualification level or the entire observation is removed. The only blank data field set to the lowest qualification record is the Marine Corps Martial Arts Program (MCMAP) belt level. Setting the MCMCAP belt level to "NOT TRAINED" is required to maintain the MCMAP records for future analysis, and is appropriate as this belt level is administrative in nature and does not imply the acquisition of skill. Removing records with blank or erroneous data fields reduces the total number of Marines in the study by approximately 25 percent to 103,569 complete records.

Marine Corps body weight standards vary dependent upon the Marine's height and gender. To prevent misclassifying a Marine as overweight or outside of established Marine Corps height and weight standards (HQMC, 2008, p. 26), height and weight data fields are replaced with a body mass index (BMI). The Centers for Disease Control's (CDC) formula for adult BMI is utilized and is shown in Equation 1 ("About Adult BMI," 2015).

$$BMI = \frac{weight(lb)}{height(in)^2} \times 703 \tag{1}$$

c. *Consolidating Sparse Groups*

Categorical variables considered for inclusion in the analysis are screened to ensure each level contains a sufficient number of observations for use as a predictor variable. The data fields for civilian education level, recruiting station enlistment waiver, recruiting district enlistment waiver, recruiting region enlistment waiver, and Marine Corps Recruiting Command (MCRC) enlistment waiver are identified as containing sparse groups.

The civilian education variable includes entries for grade levels seven through doctoral degrees. These groups are consolidated into five categories: less than high school diploma, high school diploma, some college, bachelor degree, and above bachelor degree. The four categorical variables for enlistment waivers are individually grouped into three levels: zero waivers, one waiver, and two or more waivers. This consolidation results in four categorical variables with three levels each.

d. *Assumptions and Limitations of Data*

The purpose of this study is to derive analytical models that can be used to identify OCCFLDs for which an enlistee is most likely to succeed during the Marine's first enlistment. Several assumptions were required to use the available TFDW data and construct models that provide insight into the effective assignment of Marine recruits to OCCFLDs. Our study makes the following assumptions:

1. A Marine assigned an MOS within an OCCFLD has met all applicable prerequisites. Each OCCFLD is composed of several MOSs; therefore, an individual assigned an OCCFLD may be disqualified for a requirement that is outside the scope of this study. An example of such a disqualifier is color blindness.

2. The Marine Corps' Computed Tier Score used for reenlistment decisions is an appropriate metric of success of first-term Marines. This study does not attempt to redefine or reweight the computation of the Computed Tier Score as a quantitative metric.

3. Decisions by Marines to reenlist or accept a discharge at the end of the first enlistment is not incorporated in this study. This study also does not investigate the Marine's success past the first term of enlistment.

B. VARIABLES USED IN THE STUDY

This section provides descriptions of variables provided by MCRISS and TFDW that are considered for inclusion in the analysis.

1. Dependent Variable

The dependent (response) variable considered for analysis is the Computed Tier Score used by the Marine Corps for reenlistment purposes. A Computed Tier Score is calculated using assessments accrued throughout a Marine's enlistment for all first-term enlisted Marines regardless of their intent to reenlist. The Computed Tier Score introduced in May 2011 provides commanders a quantitative tool to compare Marines against other Marines in the same reenlistment year and MOS (HQMC, 2011). A Computed Tier Score is calculated using the Marine's physical fitness test (PFT), combat fitness test (CFT), proficiency marks, conduct marks, rifle qualification score, MCMAP belt level, and history of meritorious promotion. The individual assessment scores are combined as shown in Table 6 by adding the PFT, CFT, proficiency marks multiplied by 100, conduct marks multiplied by 100, rifle qualification score, MCMAP belt score, and 100 points if the Marine was meritoriously promoted to the current rank with no non-judicial punishments (NJP) in the last year (M&RA, 2015).

Table 6. Computed Tier Score Calculations, Values, and Weights

Event	Score
PFT	PFT Score
CFT	CFT Score
Proficiency	Proficiency Marks x 100
Conduct	Conduct Marks x 100
Rifle	Rifle Score
MCMAP	(See MCMCAP Chart)
Meritorious Promotion	(+100) Meritoriously Promoted to Current Rank No NJP within last year

Source: (M&RA, 2015)

MCMAP has belt levels from Not Trained to 6th Degree Black Belt and instructor levels from Green Belt Instructor to Chief Instructor. Each belt level has a corresponding point value as shown in Table 7 used to calculate the Composite Tier Score.

Table 7. MCMAP Belt Weights for Computed Tier Score Calculation

MCMAP	DESCRIPTION	Point Value
MMA	NOT TRAINED	0
MMB	TAN BELT	5
MMC	GRAY BELT	10
MMD	GREEN BELT	15
MMF	BROWN BELT	20
MMH	BLACK BELT, 1ST DEGREE	25
MMM	BLACK BELT, 2ND DEGREE	30
MMN	BLACK BELT, 3RD DEGREE	35
MMP	BLACK BELT, 4TH DEGREE	40
MMQ	BLACK BELT, 5TH DEGREE	45
MMR	BLACK BELT, 6TH DEGREE	50
MME	GREEN BELT MARTIAL ARTS INSTRUCTOR	60
MMG	BROWN BELT MARTIAL ARTS INSTRUCTOR	70
MMJ	BLACK BELT, 1ST DEGREE MARTIAL ARTS INSTRUCTOR	80
MMK	BLACK BELT, 1ST DEGREE MARTIAL ARTS INSTRUCTOR TRAINER	90
MML	BLACK BELT, 1ST DEGREE MARTIAL ARTS INSTRUCTOR SECOND DEGREE	95
MMS	CHIEF INSTRUCTOR	100

Source: (M&RA, 2015)

The raw Computed Tier Scores are calculated for all Marines eligible to reenlist and normalized by calculating each Marine's percentile relative to other Marines in the same MOS. Computed Tier Scores are then broken into four classes, I to IV as shown in Table 8, and separated by percentile values of 90, 60, and 10, respectively. For the purposed of this study, raw Computer Tier Scores are compared within OCCFLDs rather than MOS to determine percentiles. This approach to Tier assignment allows for normalization within an OCCFLD and comparison across OCCFLDs, making it suitable as an outcome variable in a statistical analysis.

Table 8. Computed Tier Score Classifications

Tier	Percentile[1]	Description[2]	Explanation[2]
I	90th	EMINENTLY QUALIFIED MARINE	Does superior work in all duties, even extremely difficult or unusual assignments can be given with full confidence that they will be handled in a thoroughly competent manner. Demonstrates positive effect on others by example and persuasion.
II	60th	HIGHLY COMPETITIVE MARINE	Does excellent work in all regular duties, but needs assistance in dealing with extremely difficult or unusual assignments. Demonstrates reliability, good influence, sobriety, obedience, and industry.
III	10th	COMPETITIVE MARINE	Can be depended upon to discharge regular duties thoroughly and competently but usually needs assistance in dealing with problems not of a routine nature.
IV	0th	BELOW AVERAGE MARINE	May or may not meet minimum standards.

Source: (M&RA, 2015)[1], (HQMC, 2011)[2]

A Marine's legal history affects the maximum possible Tier classification assigned depending on what legal action was taken against the Marine. A NJP or court martial conviction results in an automatic highest Tier classification as shown in Table 9.

Table 9. Highest Computed Tier Classification Based on Legal History

Legal History	Highest Tier
NJP x 1	Tier II
NJP x 2	Tier III
Court Martial	Tier IV

Source: (M&RA, 2015)

An example of the Marine Corps Tier Worksheet is shown in Figure 1 using fictional inputs. It shows the Marine's raw Composite Tier Score is on the right, MOS raw Composite Tier Score on the left, and percentile breakdown at the bottom.

Figure 1. Quality Comparison Worksheet: Composite Tier Score Example

CPL I. M. MARINE MOS XXXX		
Event	**MOS Avg**	**SNM's Scores**
PFT	246	274
CFT	282	284
Proficiency Marks	430	430
Conduct Marks	430	430
Rifle Qualification	293	303
MCMAP	MMB - Tan Belt	MMD - Green Belt
Meritorious Promotion	N/A	0
	1691	1751
Legal History	**Type**	**Date**
NJP(s) / Court Martial	N/A	N/A
	Tier Chart	
Tier I (10%)		
Tier II (30%)		
Tier III (50%)	X	
Tier IV (10%)		
SNM ranks 53rd of 100 Marines reenlisting in FY'XX in MOS XXXX SNM falls in the 10 - 59 Percentile SNM is a Tier III Marine		

Source: (M&RA, 2015)

2. Predictor Variables

The consolidated database contains 237 data fields as potential independent (predictor) variables in a statistical analysis. These variables provide an overview of the Marine's career from the point of enlistment to the end of his or her first term of enlistment. The database also includes information not applicable to OCCFLD assignment or Composite Tier Score calculation such as a Marine's present command and medical examination dates. A list of the 25 independent variables considered in this study is given in Table 10.

Table 10. Independent Variables Selected for Inclusion in Analysis

Variable	Description	Type
AGE_ENLIST	Recruit's Age at Enlistment	Numeric
CIV_ED	Civilian Education	Categorical
RS_Waiver	Waiver: Recruiting Station	Categorical
District_Waiver	Waiver: Distric	Categorical
REGION_Waiver	Waiver: Regional	Categorical
MCRC_Waiver	Waiver: MCRC	Categorical
GT_SCORE	ASVAB: Composite GT Score	Numeric
MM_SCORE	ASVAB: Composite MM Score	Numeric
CL_SCORE	ASVAB: Composite CL Score	Numeric
EL_SCORE	ASVAB: Composite EL Score	Numeric
AR_SCORE	ASVAB: Subtest AR Score	Numeric
AS_SCORE	ASVAB: Subtest AS Score	Numeric
EI_SCORE	ASVAB: Subtest EI Score	Numeric
GS_SCORE	ASVAB: Subtest GS Score	Numeric
MC_SCORE	ASVAB: Subtest MC Score	Numeric
MK_SCORE	ASVAB: Subtest MK Score	Numeric
PC_SCORE	ASVAB: Subtest PC Score	Numeric
VE_SCORE	ASVAB: Subtest VE Score	Numeric
WK_SCORE	ASVAB: Subtest Score	Numeric
BMI	Calculated BMI	Numeric
GENDER	Gender	Binary
IST_PASS	Initial Strength Test - Pass/Fail	Binary
MARITAL_STAT	Marital Status	Categorical
MENTAL_GRP	AFQT Group Assigned	Categorical

C. METHODOLOGY

This section explains the approach used to conduct our statistical analysis, which includes variable selection, OCCFLD assignment optimization, and model validation. Linear and logistic regressions are described to lay a foundation to explain the model used in this study, which is multinomial generalized linear regression. The statistical techniques presented are basic for fitting multinomial generalized linear models and the reader is referred to Friedman, Hastie, Simon, & Tibshirani (2016) for a treatment of generalized linear regression modeling under constraints that enforce sparseness.

Statistical models are used to express a relationship between a dependent variable and a set of independent variables. The Composite Tier Score is a categorical dependent variable for which the probabilities of attaining each of the four Tier classification levels are related to the independent variables. Prior to conducting an analysis, we set aside ten percent of the observations as a "test" set with the remaining ninety percent used as a "training" set. The 93,213 observations in the training data set are used to construct and estimate models to predict the dependent variable, which are probabilities that a Marine with a given set of values of the predictor variables is classified as Tier I, II, III, or IV respectively. The remaining 10,356 observation are reserved to validate the model once constructed.

1. Linear and Logistic Regression

Linear regression models are used to fit a single dependent variable, Y, using k independent variables $X = \{x_1, x_2, \cdots, x_k\}$, and $k+1$ unknown coefficients $\beta = \{\beta_0, \beta_1, \beta_2, \cdots, \beta_k\}$ corresponding to the independent variables. The β_0 coefficient represents the model's intercept term. A final prediction error term, ε, is included to represent prediction error that is random and unrelated to the independent variables X. A generalized form for a linear model is shown in Equation 2 (Faraway, 2015, p. 13).

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon \tag{2}$$

Standard linear regression is not suitable for a model with categorical outcomes for the dependent variable. For dichotomous (binary) outcomes logistic regression expands upon the linear regression model in which the independent variable Y_i takes on the value 0 for "failure" and 1 for "success." The probability of success, p_i, is related to the predictor variables as shown in Equation 3.

$$\eta_i = \log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik} \tag{3}$$

The regression parameters $\beta_0,\ldots,\beta_k$ are estimated by maximizing the log likelihood function with Equation 4.

$$\ell(\beta_0,\ldots,\beta_k)=\sum_{i=1}^{n}\left[y_i\eta_i-\log(1+e^{\eta_i})\right] \tag{4}$$

Variable selection is an important aspect of constructing both linear and logistic regression models, which entails determining the best subset of independent variables to use in the model. Including too many independent variables in the model results in “overfitting” which reduces the prediction accuracy of the model due to increased estimation variability. On the other hand, including too few independent variables results in “underfitting” which introduces excessive bias. In order to balance the two types of error, criterion-based procedures that attempt to estimate prediction error often are used (Faraway, 2015, pp. 23, 155). With larger data sets cross-validation is widely regarded as the preferred method for assessing how well an estimated statistical model can fit new data.

2. Multinomial Regression with Elastinet Variable Selection

Our study uses a multinomial generalized linear regression model (MNGLM), which generalizes logistic regression to handle categorical dependent variables that assume more than two possible values. Let p_{ir} denote the probability that Y_i assumes categorical level $r, r=1,\ldots,R.$ The MNGLM assumes that $\eta_{ir}=\log(p_{ir})$ can be expressed using Equations 5, 6, and 7 (Friedman, Hastie, Simon, & Tibshirani, 2010, p. 11).

$$\eta_{ir}=\beta_{0r}+\beta_{1r}X_{i1}+\cdots+\beta_{kr}X_{ik} \tag{5}$$

$$e^{\eta_{i1}}+\cdots+e^{\eta_{iR}}=1 \tag{6}$$

$$i=1,\ldots,n;\ \ r=1,\ldots,R \tag{7}$$

The constraints in (5, 6, and 7) ensure that the MNGLM reduces to ordinary logistic regression when $R=2.$ For outcome variables with multiple categories the number of parameters in a MNGLM is substantially greater than in a logistic regression model due to the need for a separate model component for each categorical probability. Because of this, model selection is more effectively conducted using a penalized maximum likelihood approach that forces many of the coefficients β_{jr} toward zero. Techniques to control model complexity using penalization include ridge and lasso regression which have been extended to generalized linear models (Faraway, 2015, pp. 174, 177). The elastinet technique incorporates features of both ridge and lasso penalization (Zou & Hastie, 2005, pp. 302-303).

The penalty function forces the coefficients toward zero rather than selecting the number of variables to include in the model. Construction of the penalty function is shown in Equation 8.

$$Penalty = \sum_{j=1}^{k}\left[\frac{(1-\alpha)}{2}\beta_j^{\,2} + \alpha\left|\beta_j\right|\right] \tag{8}$$

The mixing parameter α is restricted to the unit interval ($0 \le \alpha \le 1$). For $\alpha = 0$ the penalty uses only the squared regression coefficients and is the same as the penalty used in ridge regression. Regression coefficients estimated under a ridge constraint tend to be shrunk towards zero, which makes ridge regression a type of shrinkage estimator. For $\alpha = 1$ the penalty uses only the absolute values of the coefficients and is the same as the penalty used in lasso regression. An attractive feature of the lasso is that it produces estimated regression models with a number of the coefficients set equal to zero, which in effect makes the lasso a variable-selection technique. Intermediate values of α produce estimates that have features of both ridge and lasso regression.

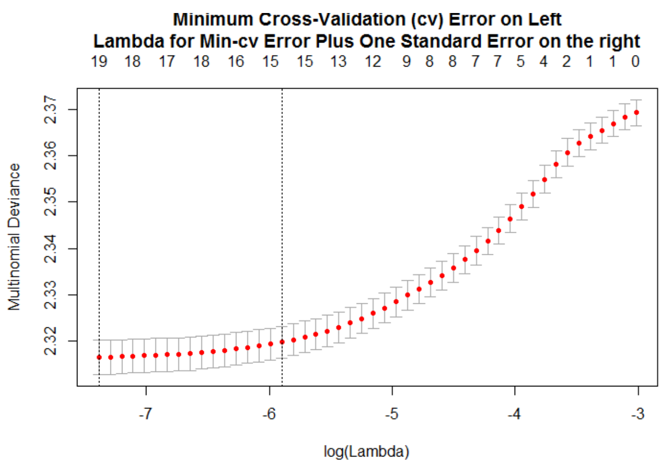
The estimated model is obtained by maximizing a profile penalized maximum likelihood function, shown in (9), with respect to α and λ.

$$g(\alpha,\lambda) = \max_{\beta_1,\ldots,\beta_k}\left(\frac{\text{log}-\text{likelihood}}{n} - \lambda * Penalty\right) \tag{9}$$

To conduct the maximization in R (R Core Team, 2015) we use the cv.glmnet function in the glmnet package (Friedman et al., 2010) which uses ten-fold cross-validation for α varying from 0 to 1 by increments of 0.1, and λ varying over 100 distinct values

(Friedman et al., 2016, p. 11) internally chosen by the software. Each fold consists of a randomly selected subset of approximately ten percent of the data. Ten models are estimated by withholding each fold in turn and fitting the model to the remaining ninety percent of the data. The log-likelihood is evaluated by considering how well the model fits the data in the held-out fold. The λ associated with one standard error above the minimum cross-validation value and the corresponding α are selected as the best mixing and complexity parameters. An illustration of selecting λ from the minimum cross-validation error is illustrated in Figure 2 where the λ associated with the minimum cross-validation error is indicated as the vertical dotted line to the left and the λ associated with one standard error above the minimum cross-validation error indicated by the vertical dotted line to the right.

Figure 2. Cross-Validation Error vs log(λ) for Multinomial GLM



For a given OCCFLD the final estimated MNGLM is used to produce estimated probabilities that a Marine attains each of the four Composite Tier Score categories, based on characteristics available prior to enlistment. Although the model is fit only to data that correspond to Marines who were assigned to that OCCFLD, the predicted model can be applied to any Marine regardless of the OCCFLD to which he or she was assigned. By fitting a separate MNGLM to each of the 38 OCCFLDs we obtain an $N \times m$ matrix of estimated probabilities where $N = 93,213$ and $m = 38 \times 4 = 152$. These probabilities can be used to assess whether a recruit could have been assigned to a different OCCFLD in which his or her estimated probability of achieving a better Tier score would have been greater. Taken in the aggregate, this information can be useful to the USMC as a tool for evaluating the effectiveness of its OCCFLD assignment process. It is, however, necessary to evaluate this information in light of the needs of the USMC to achieve specific manning levels in each of the OCCFLDs. In the following section we demonstrate how to formulate an optimization problem that incorporates manning levels as constraints for assigning recruits to OCCFLDs.

3. Occupational Field Selection Optimization

Optimization of OCCFLD assignment identifies Marines from the set $I = \{i_1, ..., i_n\}$ where n is the total number of Marines in the study for assignment to one of the 38 OCCFLDs in the set $J = \{j_1, ..., j_{38}\}$. Our objective is to maximize the expected number of either Tier I, or Tiers I and II, attainments across the set of recruits. This is equivalent to maximizing the sum of the probabilities of achieving these Tier groups based on how recruits are assigned to OCCFLDs. The summation of the Marine's probability of success, p_{ij}, and a binary variable x_{ij} that is equal to 1 if and only if Marine i is assigned to OCCFLD j, is maximized as shown in (10). That each Marine is assigned to exactly one OCCFLD is enforced in (11), and achieving the quota D_j for each OCCFLD j is enforced in (12). Each Marine is either assigned a specific OCCFLD or not, there are no partial assignments. This latter constraint is a consequence of a linear assignment problem and requires no direct enforcement other than (11).

$$\max \sum_{i \in I, j \in J} p_{ij} \, x_{ij} \tag{10}$$

$$\begin{aligned} s.t. \sum_{j \in J} x_{ij} &= 1, \quad \forall i \in I \\ 0 \le x_{ij} &\le 1, \quad \forall i \in I, j \in J \end{aligned} \tag{11}$$

$$\sum_{i \in I} x_{ij} = D_j \qquad \forall j \in J \tag{12}$$

To ensure OCCFLDs with small populations are not disproportionately assigned low probabilities of success to increase the overall probability of success, a second optimization is conducted. The second optimization maximizes the average probability of success in each individual OCCFLD, which is accomplished by replacing the objective function, Equation 10, in the previous optimization with the new objective function shown in (13).

$$\max \sum_{i \in I} \frac{p_{ij}\, \mathrm{x}_{ij}}{D_J} \qquad \forall j \in J \tag{13}$$

4. Model Validation

It is useful to assess the classification accuracy of the estimated MNGLM models developed in this study. For this purpose we use the ten-percent validation sample that was initially extracted from the larger data set. Tier-attainment probabilities are obtained for all observations in the validation samples pertaining to each of the OCCFLDs and a classification rule is devised for attaining either Tier I or Tiers I and II. Because the percentages of validation-sample attainments in these groups are known to be approximately 10 percent for Tier I and 40 percent for Tier I or II, we enforce these quotas in the classification rule. For Tier I, this reduces to classifying recruits with the largest K probabilities as belonging to that Tier group, where K is the quota for Tier I applied to the validation sample for a particular OCCFLD.

Let X denote the number of recruits who are classified as Tier I according to this rule and who achieve Tier I. A hypergeometric distribution is used to describe X due to each Tier's fixed quotas for both the number of actual successes and the number of classified successes. This distribution is used to calculate the probability that there would be as many or more successes identified given that the selections were random. Equation 14 shows the hypergeometric cumulative density function (CDF) for the probability of the random variable X taking values greater than or equal to x, the number of observed successes. The population, N, contains K success states and the number of draws from the population also is represented by K.

$$P(X \geq x) = \sum_{k=x}^{K} \frac{\binom{K}{k}\binom{N-K}{K-k}}{\binom{N}{K}} \tag{14}$$

Calculating the hypergeometric CDF for each of the 38 models' results provides a statistical method to verify the model.

5. Statistical and Optimization Software

The statistical programming language R with the glm.net package (Friedman et al., 2010) loaded is used for data combination, formatting, and statistical analysis. Resulting statistical models and probabilities of success are exported for analysis in the optimization program GAMS using CPLEX (GAMS Development Corporation, 2014).

D. CHAPTER SUMMARY

This chapter details the process of data acquisition and formatting, variable selection, and the methodology used to construct analytical models. Data formatting that encompasses data consolidation, observation removal and substitution, and sparse group consolidation ensures the data set is complete prior to statistical analysis. Predictor and response variables from the formatted data are considered for use in the multinomial elastinet regression with the goal of optimizing Marines' OCCFLD assignments upon entry into the Marine Corps. Finally, the method used to validate the model is described to ensure the model provides meaningful statistical results.

THIS PAGE INTENTIONALLY LEFT BLANK

IV. ANALYSIS

A. MODEL EVALUATION

As mentioned in Chapter III, the data set is partitioned into a "training" set containing ninety percent of the observations and "test" set containing the remaining ten percent. Partitioning is effected using a random number generator to select the observations to be included in the training set without respect to OCCFLD. Although this approach ensures the two sets are random samples of the population, randomly sampling the entire population does not enforce stratification in the same proportion at which each OCCFLD occurs in the total sample. The random sampling of the entire population into the training set containing 93,213 observations and test set containing 10,356 is broken down by OCCFLD in Table 11.

Table 11. OCCFLD Manpower Demands for Training and Test Sets

		OCCFLD Demand				OCCFLD Demand	
OCCFLD	Description	Train Set	Test Set	OCCFLD	Description	Train Set	Test Set
01	Personnel & Admin	3,779	400	43	Public Affairs	216	33
02	Intel	1,365	119	44	Legal	209	29
03	Infantry	20,309	2,208	46	Combat Camera	256	31
04	Logistics	2,097	257	48	Recruiting and Retention Specialist	138	11
05	MAGTF Plans	174	15	55	Music	353	46
06	Comm	8,577	932	57	CBRN	495	60
08	Arty	3,019	347	58	MP	2,462	278
11	Utilities	1,878	206	59	Electronics Maint	856	102
13	Engineer	5,242	571	60	Aircraft Maint	2,443	290
18	Tank and AAV	1,759	215	61	Aircraft Maint	3,767	418
21	Ground Ordnance Maint	2,706	291	62	Aircraft Maint	2,018	220
23	Explosive Ord Disposal	1,019	114	63	Avionics	2,130	238
26	SigInt	1,686	165	64	Avionics	1,516	153
28	Data/Comm Maint	2,539	292	65	Aviation Ordnance	1,496	186
30	Supply	4,057	479	66	Avionics Logistics	1,155	128
31	Dist Management	305	34	68	METOC	190	25
33	Food Service	1,283	151	70	Airfield Services	1,221	131
34	Financial Management	697	69	72	Air Control/Air Supt/Anti-Air Warfare/ATC	990	125
35	Motor Transport	8,712	982	73	Flight Crew	99	5

The two data sets are subdivided based on OCCFLD enabling construction of 38 individual elastinet GLMs for estimation of probabilities that a given recruit will attain Tiers I, II, III, or IV in a particular OCCFLD. The training data set is used to construct each of the GLMs. Each model is constructed using only observations for Marines assigned to that specific OCCFLD. For example, the 20,309 Marines in the 03 Infantry OCCFLD are used to construct an elastinet GLM for the 03 OCCFLD. Regressions for each of the 38 OCCFLDs are constructed to predict the probability of a Marine attaining each of the four Tier levels resulting in four models for each OCCFLD, totaling 152 regression models. Estimated regression coefficients for the Tier I and Tier II models are included as Appendix A and Appendix B, respectively.

As explained in Chapter III, the inclusion of both the L_1-norm and L_2-norm in the elastinet GLM penalty function forces regression coefficients either toward zero or sets the coefficient to zero depending on how the penalty function is constructed. This characteristic of the elastinet GLM automatically selects which of the 25 potential predictor variables are included in each model. Regressions models for each of the 38 OCCFLDs and four Tier levels are different from each other, as are the number and magnitude of the non-zero regression coefficients. The number of non-zero coefficients included in the 38 models for probability estimates that a Marine attains Tier I range from 9 to 16 with an average of approximately 13 non-zero coefficients. Similarly, the number of non-zero coefficients included in the 38 models for probability estimates that a given Marine attains Tier II ranges from 7 to 18 with an average of approximately 13 non-zero coefficients.

Referring to Tables 12 and 13, the 38 models predicting Tier I attainment have several common features. BMI contributed negatively to probability estimates that a Marine attains Tier I for 34 of the 38 models. The ASVAB scores EL, VE, MM, CL, and AFTQ categorization Mental Group did not contributed to the majority of the models while Civilian Education and Age at Enlistment contributed positively.

Table 12. Tier I Predictor Variables Listed Most Frequently in Models with Negative Contribution, No Contribution, or Positive Contribution

Negative Contribution		No Contribution		Positive Contribution	
Predictor	Number of Models	Predictor	Number of Models	Predictor	Number of Models
BMI	34	EL_SCORE	33	CIV_ED	34
GS_SCORE	29	VE_SCORE	31	AGE_ENLIST	26
District_Waiver	25	MM_SCORE	29	MK_SCORE	20
WK_SCORE	22	MENTAL_GRP	28	GT_SCORE	19
EI_SCORE	18	CL_SCORE	26	AR_SCORE	18

A "positive contribution" increases the probability estimates that a Marine attains Tier I due to a positive coefficient, a "negative contribution" decreases the probability estimate due to a negative coefficient, and "no contribution" has no impact on the model due to a coefficient of zero.

Table 13. Comparison of Regression Coefficients for OCCFLDs 03 and 04

	OCCFLD Regression Coefficients	
	03	04
AGE_ENLIST	0.0482	0.0770
CIV_ED	0.2543	0.4696
RS_Waiver	-0.0134	0
District_Waiver	-0.0630	-0.0989
REGION_Waiver	0.1020	0.0369
MCRC_Waiver	0	0
GT_SCORE	0.0176	0.0055
MM_SCORE	0	0.0107
CL_SCORE	0	0
EL_SCORE	0	0
AR_SCORE	0.0069	0.0003
AS_SCORE	0.0099	-0.0015
EI_SCORE	0	-0.0163
GS_SCORE	-0.0110	-0.0143
MC_SCORE	0	-0.0079
MK_SCORE	0.0112	0.0372
PC_SCORE	0.0037	0
VE_SCORE	0	0
WK_SCORE	-0.0074	-0.0092
BMI	-0.0112	-0.0414
GENDER	0	0.2176
IST_PASS	0	0
MARITAL_STAT	-0.0952	-0.2928
MENTAL_GRP	0	0

Example of a Side-by-Side OCCFLD comparison illustrating the similarities and differences between OCCFLD regressions and individual regression coefficients.

Referring to Table 14, the 38 models predicting Tier II attainment have several common features. BMI contributed negatively to probability estimates that a Marine attains Tier II for 36 of the 38 models. The ASVAB scores VE, EL, MM, CL, and GT did not contributed to the majority of the models while Civilian Education, Gender, and Age at Enlistment contributed positively as shown in Table 14.

Table 14. Tier II Predictor Variables Listed Most Frequently in Models with Negative Contribution, No Contribution, or Positive Contribution

Negative Contribution		No Contribution		Positive Contribution	
Predictor	Number of Models	Predictor	Number of Models	Predictor	Number of Models
BMI	36	VE_SCORE	34	CIV_ED	36
District_Waiver	27	EL_SCORE	33	GENDER	25
GS_SCORE	22	MM_SCORE	32	AGE_ENLIST	25
WK_SCORE	20	CL_SCORE	28	MK_SCORE	20
EI_SCORE	18	GT_SCORE	25	PC_SCORE	20

A "positive contribution" increases the probability estimates that a Marine attains Tier II due to a positive coefficient, a "negative contribution" decreases the probability estimate due to a negative coefficient, and "no contribution" has no impact on the model due to a coefficient of zero. Gender indicates male.

B. MODEL VALIDATION

The regression models provide estimates of probabilities that a given recruit will attain Tiers I, II, III, or IV in a particular OCCFLD without specifying which Tier the Marine actually obtained. The Tier obtained is known as a result of Tier Score calculations and normalization described in Chapter III. To compare the regression models' results to the actual Tier obtained, a process of classifying a Marine as a specific Tier is developed. By definition, the top 10^{th} percentile in each OCCFLD is considered Tier I. For the purposes of this study, the top 10 percent of Marines ordered by probability of Tier I attainment in each OCCFLD of the test set are identified as Tier I. The remaining Marines in each OCCFLD are now ordered by the sum of their probability of attaining either Tier I or Tier II. For example, Marine A has a probability of attaining Tier I, P(Tier I), equal to 0.12 and a probability of attaining Tier II, P(Tier II), equal to 0.32. Marine A's probability of attaining either Tier I or Tier II is the sum of the two probabilities, P(Tier I) + P(Tier II), equal to 0.44. The Marines with the highest probability of attaining Tier I or Tier II are identified as Tier II. The number of Marines in Tier II is equal to 30 percent of the number of Marines in the specific OCCFLD in the test set. Tier III and Tier IV are defined in the same manner as Tier II and the number of Marines in each of these last two Tiers are equal to 50 and 10 percent of Marines in the specific OCCFLD in the test set, respectively.

The estimated MNGLM regression models should outperform random Tier level assignment for the models to be regarded as exhibiting skill. The number of Marines correctly identified as Tier I is compared to a discrete hypergeometric distribution. The percentages represent the probability of randomly assigning ten percent of an OCCFLD to Tier I and the number of correctly assigned Marines to Tier I is the same or greater than the number of Marines correctly assigned to Tier I using the regression models and assignment process from the previous paragraph.

Nine OCCFLDs for assignment as Tier I have probabilities that are greater than .01% (05 MAGTF Plans, 33 Food Service, 34 Financial Management, 43 Public Affairs, 48 Recruiting and Retention Specialist, 57 CBRN, 63 Avionics, 68 METOC, and 72 Air Control/Air support/Anti-Air Warfare/ATC. Of these, the only OCCFLDs with a probability greater than one percent are 33 Food Service with 3.49 percent and 34 Financial Management with 2.41 percent. Repeating the process for OCCFLD assignments as Tier I or II resulted in only one probability greater than zero, 73 Flight Crew with 0.01 percent.

The probability calculations for Tier assignments are repeated for the probabilities of random assignment outperforming the results of the models for assignment to either Tier I or Tier II and are shown in Table 16. Only the "73" OCCFLD is calculated to have a probability greater than 0.01%.

C. OCCFLD ASSIGNMENT OPTIMIZATION

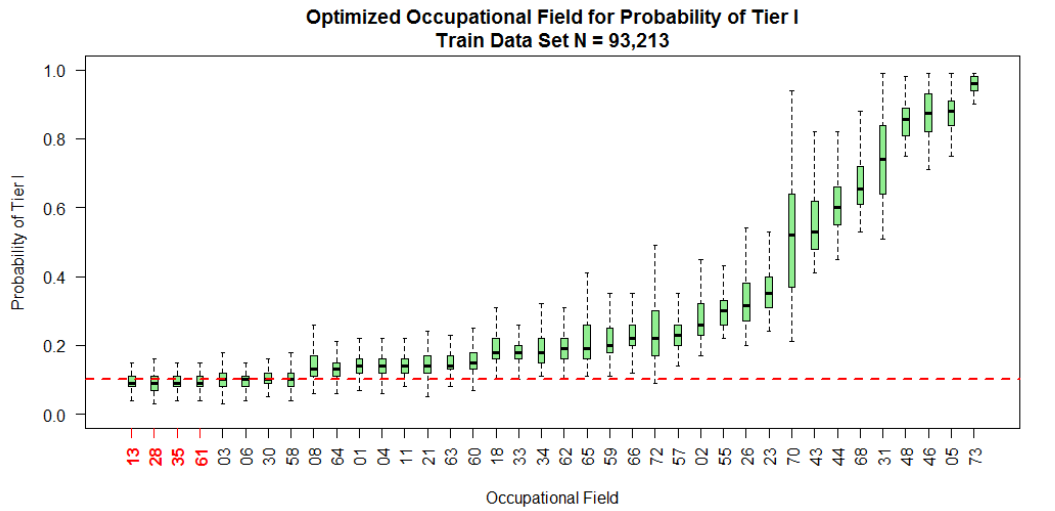
Results presented to this point are regression models and hypergeometric distributions to estimate the probability of a Marine attaining each of the four Tier levels in the OCCFLD the Marine was actually assigned upon enlistment. Investigating the effect of using recruit information available prior to enlistment requires calculating probability estimates of a Marine attaining each of the four Tier levels in all 38

OCCFLD. These estimates provide insight into which OCCFLDs have the highest probability of success for a specific recruit.

Probability estimates for each of the four Tier levels and all 38 OCCFLDs are calculated for each of the 93,213 Marines in the test set resulting in 152 probabilities for each Marine and 14,168,376 probabilities in total. The process of OCCFLD assignment is optimized to maximize the sum of all probabilities of success in the “new” OCCFLD a Marine is assigned. The two metrics of success investigated are a Marine attaining Tier I and a Marine attaining either Tier I or Tier II.

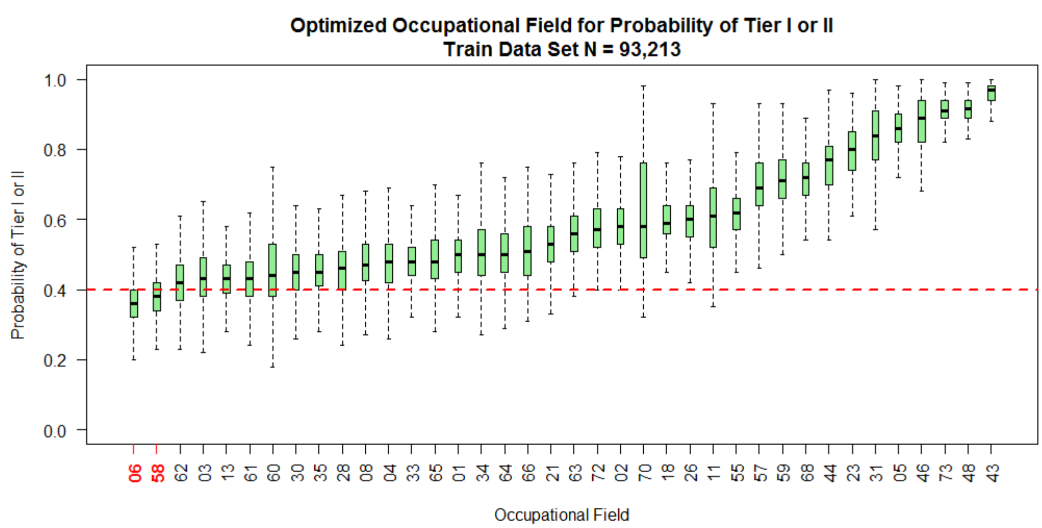
The baseline for the probability of a Marine attaining Tier I is 0.10 by definition of Tier I as the top tenth percentile. Probability of Marines attaining Tier I in the “new” optimized OCCFLD represents an increase above the baseline. Figure 3 shows the optimization output for Tier I in the training set of 93,213 Marines as a boxplot. The boxplot shows the median estimated probability of attaining Tier I as a horizontal black bar within the colored vertical box. The colored vertical box range from the 75th percentile down to the 25th percentile and the dashed “whiskers” extend from the edge of the colored bar to the data points with the maximum and minimum value. The average estimated probability of a Marine attaining Tier I is calculated as 0.151 compared to the 0.10 baseline shown in Figure 3 as the horizontal red dashed line. Similar to Figure 3, Figure 4 shows the optimization output for Tier I or II. The average estimated probability of a Marine attaining Tier I or II is calculated as 0.477 compared to the 0.40 baseline shown in Figure 4 as the horizontal red dashed line.

Figure 3. Training Data Set: Optimized OCCFLD for Probability of Tier I



The average estimated probability of a Marine attaining Tier I is calculated as 0.151 compared against the 0.10 baseline shown as a horizontal red dashed line.

Figure 4. Training Data Set: Optimized OCCFLD for Probability of either Tier I or Tier II



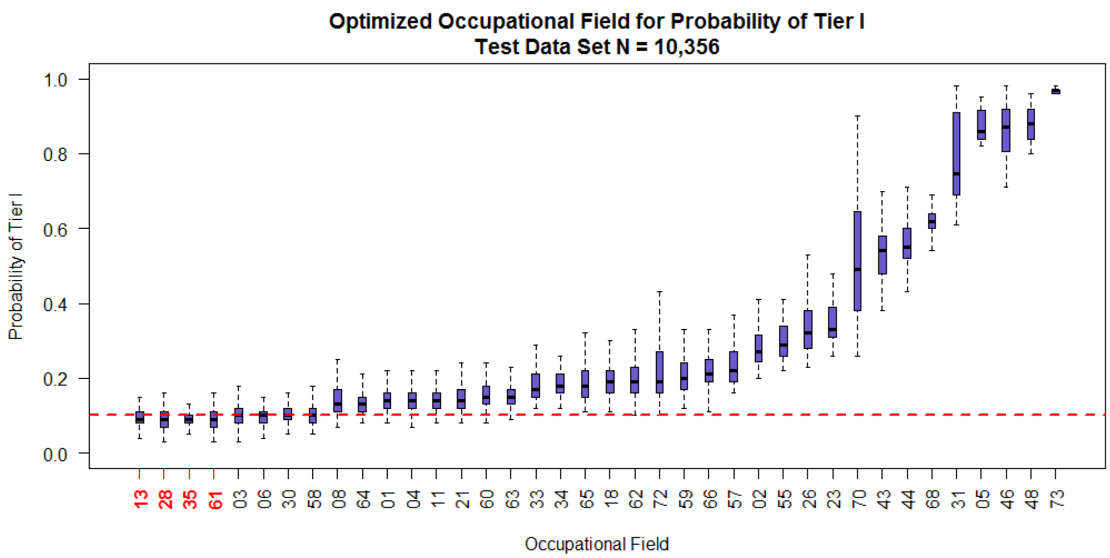
The average estimated probability of a Marine attaining Tier I or Tier II is calculated as 0.477 compared against the 0.40 baseline shown as a horizontal red dashed line.

The median estimated probability of a Marine attaining Tier I is below the ten percent baseline for OCCFLDs 13 Engineer, 28 Data/Communications Maintenance, 35 Motor Transport, and 61Aircraft Maintenance. The median estimated probability of a Marine attaining Tier I or II is below the forty percent baseline for OCCFLDs 06

Communications and 58 Military Police. Optimizing OCCFLD assignments in the training set results in reassigning approximately 91 percent of the Marines to an OCCFLD other than the Marine's actual OCCFLD for success as Tier I and reassignment of approximately 92 percent for optimizing for success as either Tier I or II.

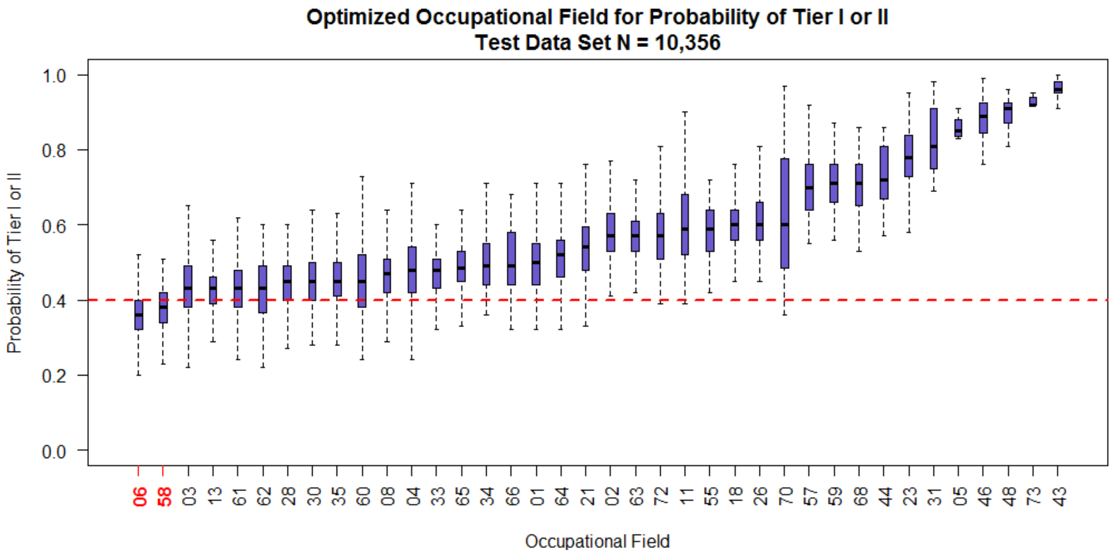
Calculation of the estimated probability estimates are now conducted for the test set consisting of the 10,356 Marine records not used to build the regression models. Similar to the training set, probability of Marines attaining Tier I in the "new" optimized OCCFLD represents an increase above the baseline. Figure 5 shows the optimization output for Tier I in the test set. The average estimated probability of a Marine attaining Tier I is calculated as 0.150 and the 0.10 baseline. Figure 6 shows the optimization output for Tier I or II with an average estimated probability of a Marine attaining Tier I or II calculated as 0.476 compared to the 0.40 baseline.

Figure 5. Test Data Set: Optimized OCCFLD for Probability of Tier I



The average estimated probability of a Marine attaining Tier I is calculated as 0.150 compared against the 0.10 baseline shown as a horizontal red dashed line.

Figure 6. Test Data Set: Optimized OCCFLD for Probability of either Tier I or Tier II



The average estimated probability of a Marine attaining Tier I or Tier II is calculated as 0.476 compared against the 0.40 baseline shown as a horizontal red dashed line.

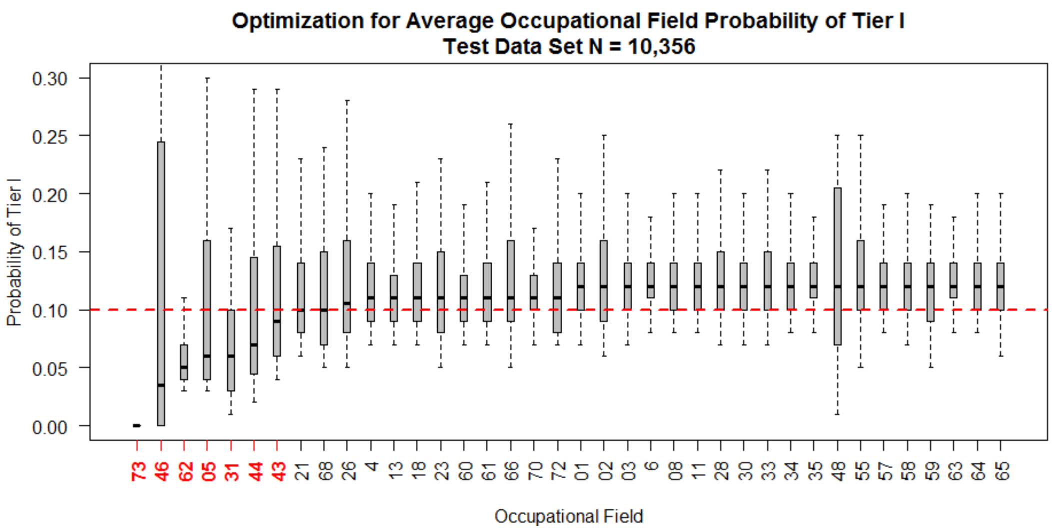
The test set optimization results in the same OCCLDs identified as having a median estimated probability of a Marine attaining Tier I and Tier I or II below the respective baselines. Optimizing OCCFLD assignments in the test set results in reassigning approximately 91 percent of the Marines to an OCCFLD other than the Marine’s actual OCCFLD for success as Tier I and reassignment of approximately 92 percent for optimizing for success as either Tier I or II.

Optimizations conducted thus far have maximized the overall estimated probability of a Marine attaining Tier I or attaining Tier I or II. To investigate the potential for a small OCCFLD being disproportionately penalized to increase the overall estimated probability across all OCCFLDs, a second set of optimizations are calculated. As explained in Chapter III, the second set of optimization maximizes the average estimated probability in each OCCFLD.

Probability of Marines attaining Tier I in the “new” optimized OCCFLD represents an increase above the baseline. Figure 7 shows the optimization output for Tier I in the test set. The average estimated probability of a Marine attaining Tier I across each

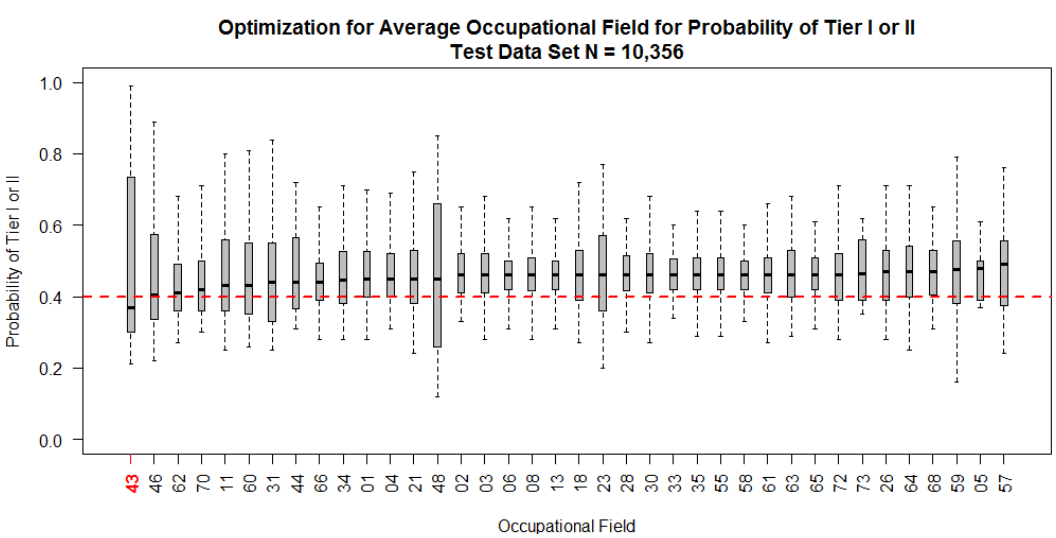
OCCFLD is calculated as 0.130 compared to the 0.10 baseline. Figure 8 shows the optimization output for Tier I or II with an average estimated probability of a Marine attaining Tier I or II calculated as 0.460 compared to the 0.40 baseline.

Figure 7. Test Data Set: Optimization for Average OCCFLD Probability of Tier I



The average estimated probability of a Marine attaining Tier I across each OCCFLD is 0.130. The 0.10 baseline is shown as a horizontal red dashed line.

Figure 8. Test Data Set: Optimization for Average OCCFLD Probability of Tier I or II

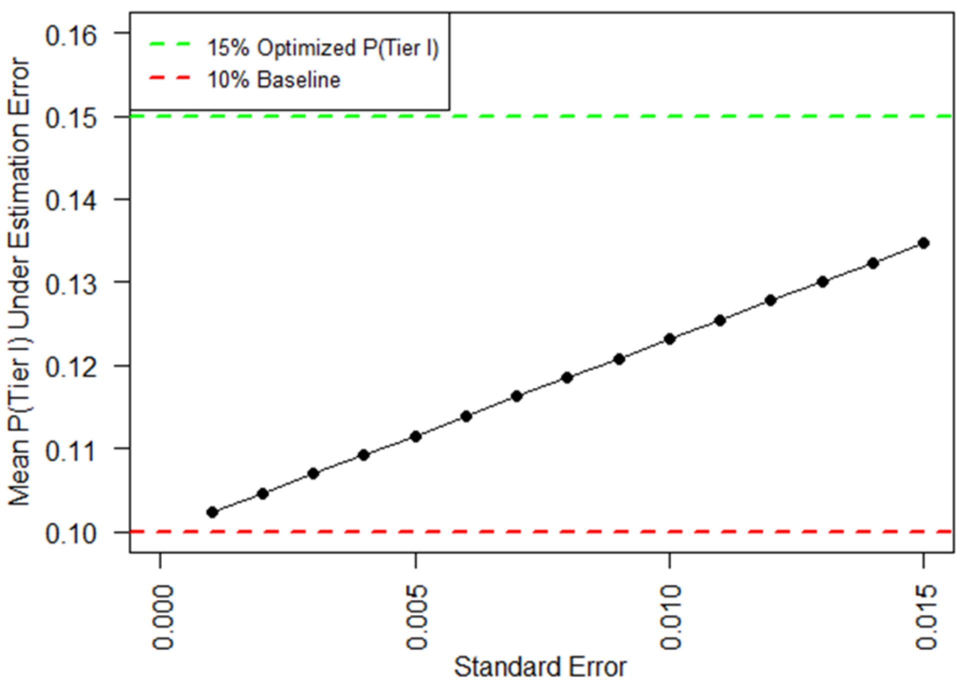


The average estimated probability of a Marine attaining Tier I or II across each OCCFLD is 0.460. The 0.40 baseline is shown as a horizontal red dashed line.

The median estimated probability of a Marine attaining Tier I across each OCCFLD is below the ten percent baseline for OCCFLDs 73 Flight Crew, 46 Combat Camera, 62 Aircraft Maintenance, 05 MAGTF Plans, 31 Distribution Management, 44 Legal, and 43 Public Affairs. The median estimated probability of a Marine attaining Tier I or II is below the forty percent baseline for OCCFLD 43 Public Affairs. The OCCFLDs identified as having medians below the respective baseline differ for the two objective function, though each result in an overall average increase above the baseline.

To investigate the effects of bias due to optimization with estimated probabilities, a third optimization is conducted. This final optimization assumes a normal distribution with a mean of 10 percent (0.10) and standard error ranging from 0.001 to 0.015 as shown in Figure 9.

Figure 9. Mean Probability of Tier I under Estimation Error for Standard Errors Ranging from 0.001 to 0.015



The upper green dotted line indicates the optimized 15 percent probability of attaining Tier I and the lower red dotted line indicates the baseline 10 percent probability of attaining Tier I due to normalization within OCCFLDs.

Figure 9 illustrates that, despite the inherent bias associated with optimizing using estimated probabilities, there remains the potential for an increase in probability of attaining Tier I under the regression models within these standard errors.

D. CHAPTER SUMMARY

This chapter details the process of model evaluation, verification, and exploitation through optimization. Model evaluation searches for trends common across multiple models for Tier I or Tier II. Although trends for Tier I and Tier II are presented separately in Tables 12 and 14, respectively; several predictor variable contributions are common to both models. The predictor variables BMI, District Waiver, and the ASVAB Subtest scores for GS, WK, and EI all contribute negatively to the majority of models for both Tier I and Tier II, thus lowering the estimated probability of attaining the specific tier. The ASVAB Subtest scores for EL, VE, MM, and CL did not contribute to the majority of models for both Tier I and Tier II due to regression coefficients of zero. Predictor variables common to both Tier I and Tier II that positively contributed to the majority of models, increasing the estimated probability of attaining that tier, included Civilian Education Level, Age at Enlistment, and the ASVAB Subtest score for MK.

Model verification considers the probability that random tier-level assignment performs as well or better than assignments from the regression models. All OCCFLDs have p-values that are statistically significant at the 5 percent significance level with several having very small p-values. We therefore find strong and consistent evidence across all OCCFLDs that the MNGLM models have skill in predicting either Tier I or Tiers I/II attainment by USMC recruits at the end of their first enlistment.

Finally, the OCCFLD assignments are optimized using a success metric of attaining Tier I and attaining Tier I or II. Assignments are optimized to maximize the overall probability of attaining Tier I across all OCCFLDs, resulting in an increase in average probability from the baseline of 10 percent to an optimized 15.1 percent for the training set and to an optimized 15.0 percent for the test set. The maximization for overall probability is repeated for attaining either Tier I or Tier II resulting in an increase from the baseline of 40 percent to an optimized 47.7 percent for the training set and to an optimized 47.6 for the test set. To ensure the increase in probability of success is not disproportionately penalizing OCCFLDs due to size, a second optimization is conducted to maximize average probability of attaining Tier I and of attaining either Tier I or II within each OCCFLD. The second optimization results in an increase for the probability of attaining Tier I from the baseline of 10 percent to an optimized 13 percent for the test set and an increase in probability of attaining either Tier I or II from the baseline of 40 percent to an optimized 46 percent. These optimizations include random estimation error associated with estimated probability of Tier I or Tier I/II that artificially biases the outcome, increasing the probability of attaining Tier I or Tier I/II. This inherent bias for optimizing Tier assignments remains a topic for future research.

THIS PAGE INTENTIONALLY LEFT BLANK

V. CONCLUSIONS AND RECOMMENDATIONS

A. SUMMARY

This thesis investigates how information on a Marine recruit available at entry can be used to predict the Tier level for assignment to a specific OCCFLD. Multinomial elastinet regression models are developed to calculate estimated probabilities that use entry-level information as predictor variables. Our models are applied to all Marines in the study for each of the 38 OCCFLDs to generalize the analysis. Optimization of OCCFLD assignment based on the developed models illustrates the potential insight provided by recruit information available prior to enlistment. Two questions are considered in our analysis and are presented in this section with our findings.

1. How can information obtained from a Marine recruit at entry be used to predict which OCCFLD is best suited to that individual?

This study provides evidence that the relationship of recruit characteristics available prior to enlistment and the USMC Computed Tier Score are dependent upon the OCCFLD assigned. Individual models are constructed for each OCCFLD with the recruit characteristics used as predictors. The multinomial elastinet regression provides estimated probabilities that a given recruit will attain Tiers I, II, III, or IV in each OCCFLD. Ranking the OCCFLDs in order of estimated probability of success for attainment of Tier I or attainment of Tier I or Tier II provides recruits and recruiters with insight concerning which OCCFLDs are best suited for that recruit.

2. How can information obtained from a Marine recruit at entry be used to predict if a Marine successfully completes the first term of enlistment?

This study defines a successful first term of enlistment as a Marine attaining either a Tier I or Tier II categorization of the USMC Computed Tier Score. As illustrated for two OCCFLDs in Table 14 of Chapter IV, the contribution of predictor variables based on recruit information available prior to enlistment is dependent upon the OCCFLD. Trends for estimated probability of attaining Tier I and attaining Tier II from the models are provided in Tables 13 and 15 of Chapter IV. Providing recruits and recruiters a tool that provides estimated probabilities of attaining Tier I or Tier II in descending order for each OCCFLD during initial assignment, aids in the selection of an OCCFLD that maximizes the estimated probability of a successful first term enlistment.

B. CONCLUSIONS

Identifying OCCFLDs with the greatest estimated probabilities of a successful first term enlistment takes a proactive approach to increasing the caliber of Marines assigned to each OCCFLD. The number of Marines in each Tier level will not change due to the normalized percentile definition within MOSs, or OCCFLDs as in this study, but the potential for increasing Marine performance levels required for obtaining each Tier exists. Optimization of OCCFLD assignment based on the developed models illustrates the potential insight provided by recruit information available prior to enlistment and results in reassigning approximately 91 percent of the Marines in our study to an OCCFLD other than the Marine’s actual OCCFLD. The large percentage of Marines reassigned to a different OCCFLD does not imply that recruiters or manpower HQMC are performing their duties incorrectly; rather the large percentage of reassignments illustrates that there exists information that can aid in these duties. Consequently, Headquarters Marine Corps (HQMC) screening of the most qualified individual should take place prior to the reenlistment screening process, ideally during initial assignment.

This study is meant to provide a tool that enables USMC Manpower Studies & Analysis Branch and recruiters to identify possible recruit-OCCFLD pairings in which the recruit has the greatest probability of first term success. The models developed for each of the OCCFLDs can be packaged in such a way as to enable recruiters with no statistical knowledge to implement the model and interpret the results provided for each of the OCCFLDs. We recommend identifying OCCFLDs with the highest estimated probabilities of Tier I or Tier II attainment at the recruitment phase. Providing recruits and recruiters a tool that provides estimated probabilities of attaining Tier I or Tier II in descending order for each OCCFLD during initial assignment has the potential to increase the caliber of Marines across all OCCFLDs and aid in assessing the current OCCFLD assignment practices.

C. RECOMMENDATIONS FOR FUTURE WORK

Based on insights gained from our study, the following two topics for future work are suggested to improve the impact of our findings.

First, optimization used in Chapter IV illustrates an overall average increase in Marines' probabilities of attaining Tier I and Tier I or II across all OCCFLDs. Due to the impossibility of reassigning Marines to the new optimized MOSs, further simulations could determine the level of bias introduced during regression and optimization. Such simulations would provide a process to determine the level of true improvement in comparison to optimizing to the upper extent of the variance.

Second, the models developed are based solely on information available prior to a recruit's enlistment with the Computed Tier Score categories as the single metric of success. The recruit information contains demographics, education level, a single pass-fail initial strength test, and limited ASVAB test scoring. The incorporation of non-cognitive tests prior to enlistment has the potential to provide greater predictive power of a Marine's probability of success during a first term enlistment. Additionally, the use of OCCFLD specific testing for use as a performance metric would provide insight into the characteristics required of Marines in specific OCCFLDs rather than a single Tier Score common across all OCCFLDs.

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX A. REGRESSION COEFFICIENTS: TIER I

	Admin	Intel	Infantry	Logistics	MAGTF Plans	Comm	Arty	Utilities
	01	**02**	**03**	**04**	**05**	**06**	**08**	**11**
AGE_ENLIST	0	0	0.048159943	0.076983033	0.119575192	0.055603311	0.111202151	0.074585454
CIV_ED	0.592219514	0.731274261	0.254311141	0.469609652	0.190185479	0.294327051	0.136932538	0.233860625
RS_Waiver	-0.009902115	0.416255069	-0.013444639	0	-0.181962852	0	-0.025605394	0
District_Waiver	-0.074601217	0	-0.063000518	-0.098892689	0.21859981	-0.155380353	-0.070311542	-0.088323628
REGION_Waiver	0	0	0.102046254	0.036908315	0.648876538	0	-0.072383854	0
MCRC_Waiver	0	-0.006911625	0	0	1.236974258	0	0	0
GT_SCORE	0.004128784	0.000222607	0.017628397	0.005512061	0.020530391	0.014593179	0	0
MM_SCORE	0	0.030454264	0	0.010675608	0	0	0	0
CL_SCORE	0	0	0	0	0	0.008462227	0.041132626	0.024634335
EL_SCORE	0	0	0	0	0	0	0	0
AR_SCORE	0.014486712	0.007577537	0.006854559	0.000339456	0	0	0	0.023235434
AS_SCORE	0.011317901	0.015502919	0.009940452	-0.001545291	-0.043766753	-0.001865391	0.003183243	0
EI_SCORE	0	-0.028476612	0	-0.016289042	0	-0.010360467	0	-0.017962736
GS_SCORE	-0.025684556	-0.024224543	-0.010982454	-0.014296234	-0.065917506	-0.015117308	-0.013067775	-0.002905137
MC_SCORE	0.002404709	0	0	-0.007933806	0	0	0	0
MK_SCORE	0.026160153	0.032278399	0.01120079	0.037177036	-0.001085519	0	0	0.007914585
PC_SCORE	-0.007507398	-3.31E-05	0.003744205	0	0	0.013998834	0	0
VE_SCORE	-0.001518784	0	0	0	0	0	0	0
WK_SCORE	0	0	-0.007448165	-0.009193927	-0.08320968	-0.027610891	-0.030335931	-0.034321348
BMI	-0.055806805	-0.005119405	-0.011236221	-0.041412931	0	-0.027333128	-0.014233832	-0.003019255
GENDER	0.010029903	0.139002959	0	0.217633239	0	0.069174938	0	0.096401931
IST_PASS	0	0	0	0	0	0.039452842	-0.452712586	-1.255750551
MARITAL_STAT	-0.383182167	0.194688774	-0.095227652	-0.292772832	0.121767628	-0.106713146	-0.068552903	0
MENTAL_GRP	0	0	0	0	0	0	0	0

	Engineer	Tank/AAV	Ground Ord	EOD	SigInt	Data/Comm	Supply	Dist Mgmt
	13	**18**	**21**	**23**	**26**	**28**	**30**	**31**
AGE_ENLIST	0.036023084	0.021105298	0.069595416	0	0.036885711	0.023613383	0.022626105	-0.008978241
CIV_ED	0.563038145	0	0.450148587	0.28408327	0.542932593	0.27613052	0.223084763	1.82904343
RS_Waiver	0.044869877	0	0	0	0.338337325	0.010509276	0.051832348	-0.488044513
District_Waiver	-0.269083747	-0.107106807	0	-0.04955907	-0.029928722	-0.10391097	0	-0.126986778
REGION_Waiver	0.079965175	0.619772836	0	0	0.588508661	0.473039022	0.13674491	-0.674054322
MCRC_Waiver	0.032087858	-0.045339624	0	-0.142656696	-0.121933512	0	0	0.312295019
GT_SCORE	0.006800989	0	0.010703056	0.000140526	0	0	0.022749588	0.022579946
MM_SCORE	0	0.000607145	0	0.011864948	0	0.005970899	0	0.110393537
CL_SCORE	0.000679721	0	0	0	0	0	0.03172131	0
EL_SCORE	0	0	0	0	0	0.008665556	0	0
AR_SCORE	0	0	0.008058335	0	0	0	0	0
AS_SCORE	0.001274756	0.008903369	0	0	0	0	0	0.002708519
EI_SCORE	0	0	0	0	0	0	-0.012520068	-0.110760088
GS_SCORE	-0.014409579	-0.000796789	-0.00500191	-0.008649036	0	0	0	-0.075827108
MC_SCORE	0.00055068	0	0	0.039603236	0.009972861	0	0	-0.076728086
MK_SCORE	0.007009227	0.03733664	0.02594853	0	0.018133622	0.031176353	0	-0.008674537
PC_SCORE	3.88E-05	0	0	0.004517053	0.006476634	0	-0.004284078	0
VE_SCORE	0	-0.025914215	0	0	0	0	-0.000157888	-0.057944391
WK_SCORE	0	0	-0.023109565	0	-0.054512358	-0.009820214	-0.027142314	0
BMI	-0.036021027	-0.02075658	-0.055509351	0	-0.048894363	-0.035229942	-0.048310651	-0.139346422
GENDER	0.081766072	0	1.452810702	0.243579809	0	0	0.122582861	-0.122063542
IST_PASS	0	0.071096182	-0.041253668	0.157520987	-1.244875027	0.493839329	0	0
MARITAL_STAT	0	-0.105844927	-0.275650354	0	0.027511339	-0.134940504	-0.106183708	0
MENTAL_GRP	0	0	0	-0.397520305	-0.193906529	0.110063166	0.203332316	0

	Food Service	Financial Mgmt	Motor Trans	Public Affairs	Legal	Combat Cam	Recruit & Retention	Music
	33	**34**	**35**	**43**	**44**	**46**	**48**	**55**
AGE_ENLIST	0.105279083	0	0.054757441	0.071893614	0.117805458	0	0.040896558	0
CIV_ED	0	0.612779588	0.304579211	0.56238452	0.176059516	1.13261779	0.27518001	0.085994761
RS_Waiver	-0.083806277	-0.065322948	0	0	0.003402699	0	0	-0.184926761
District_Waiver	-0.087044423	0	0	0.055179514	-0.118774345	0	-0.489484373	-0.088771116
REGION_Waiver	-0.133795992	0	0	-0.141785473	-1.042593281	1.301433209	0.496651518	-1.260439204
MCRC_Waiver	-0.051215773	-0.748343804	0	0.058630751	0	0	-1.452802447	0.133770469
GT_SCORE	0	0.012836012	0.003769386	0	0	0	-0.007507269	0
MM_SCORE	0	0.009495191	0	0	0	0	0	0
CL_SCORE	0.017542673	0	0.005243927	0	0.034493953	0	0	0
EL_SCORE	0.004415187	0	0	0	0	-0.05420244	0	0
AR_SCORE	0.009688264	0	0.015532293	0.023591796	0.004465299	0	0	0
AS_SCORE	0	0	0	0	-0.016939811	0.095490161	-0.118666915	0.005919091
EI_SCORE	0	0	0	0	-0.03984251	0	0.053025337	-0.009357765
GS_SCORE	0	-0.019796837	-0.011030874	-0.003654614	-0.093809593	-0.12290565	-0.008156686	-0.018293682
MC_SCORE	0	0	0	0.009749958	0	0	-0.034839954	0
MK_SCORE	0	0	0.003871598	0	0.01887459	0.077009912	0	0.046048918
PC_SCORE	-0.012507975	0	0	0	0.096190336	0.005925383	0.075490909	0.017127873
VE_SCORE	-0.023397345	0	0	0	0	0	0	0
WK_SCORE	0	-0.016623356	-0.017378609	-0.011866145	0	0	0	0
BMI	-0.104692176	-0.038536251	-0.028533936	-0.064533033	-0.048282224	-0.151654405	0.161018926	-0.096121847
GENDER	0	0	0	1.408260784	0	0.121049861	0	0.096648162
IST_PASS	-0.571069686	0	0	0	0	0	0	0.388264087
MARITAL_STAT	0.297904519	0.422242048	-0.019796554	0	0	-1.27355885	0.825039435	0.092623074
MENTAL_GRP	0.19693339	0	0	0	0	-1.181718245	0	0

	CBRN	MP	Elect Maint	Aircraft Maint	Aircraft Maint	Aircraft Maint	Avionics	Avionics
	57	**58**	**59**	**60**	**61**	**62**	**63**	**64**
AGE_ENLIST	0.020327615	0.085668909	0.061947679	0	0.019149418	0.049495894	0	0
CIV_ED	0.450250699	0.245315339	0.245756016	0.518321129	0.518191263	0.549972245	0	0.525076867
RS_Waiver	-0.006509558	-0.050764783	0	-0.165721078	0.006361316	0	0	0.015727144
District_Waiver	-0.04936747	0	-0.215850325	-0.135413412	0	0	-0.018208437	-0.288993071
REGION_Waiver	-0.315732087	0.076004937	0	0.008092733	0	-0.112047032	0	0.007152045
MCRC_Waiver	0	0	-0.020289946	0.031775567	0.040773788	0.18081342	0	-0.047744824
GT_SCORE	0	0.007703536	0	0.017733868	0.012209163	0.031722221	0	0
MM_SCORE	0	0	0	0	0.003122037	0	0	0.017732826
CL_SCORE	0	0.021055459	0	0	0	0	0	0.010866807
EL_SCORE	0	0	0	0	0	0	0	0.014048708
AR_SCORE	0.023094084	0	0	0.038987188	4.72E-05	0.000731677	0.014969322	0
AS_SCORE	0	0	0.002360816	0.010120076	0	-0.003367956	0.006048574	0.011771176
EI_SCORE	0.016482054	0	0	-0.002530572	-0.011618267	-0.02678092	-0.02215775	-0.047062662
GS_SCORE	-0.016703721	-0.014679123	0	-0.019343214	-0.013317787	-2.86E-05	-0.006819661	0
MC_SCORE	0	0	0.008914549	0	0.003061895	0	0	-0.00181948
MK_SCORE	0	0	0.058921206	0	0.037156458	0.008369232	0	0
PC_SCORE	0	0	0	0	0	-0.00328393	0	0
VE_SCORE	0	0	-0.048794831	0	0	0	0	0
WK_SCORE	0	-0.028341154	0	-0.016051386	-0.002029313	-0.025856678	0	-0.044758107
BMI	-0.023841682	-0.015470085	-0.091046624	-0.039366191	-0.025053014	-0.080159296	-0.013767233	-0.016392759
GENDER	1.830897428	0.247185255	0.02916615	-0.014383539	0	-0.031196738	0	0
IST_PASS	0	0.427585235	0.212111859	-1.26551283	0	-4.55478812	0	0.159372053
MARITAL_STAT	-0.23840335	-0.135717935	0	-0.492752113	-0.025513555	0	0	-0.177655932
MENTAL_GRP	0	0	-0.238184463	0	-0.002759905	0	0	0.008839417

	Aviation Ord	Avionics Log	METOC	Airfield Serv	Air Control/Air Supt/AAW/ATC	Flight Crew
	65	**66**	**68**	**70**	**72**	**73**
AGE_ENLIST	0.015296784	0.091399154	-0.020334117	0.011865893	0.070682975	0
CIV_ED	0.862509545	0	0.804877423	1.293524558	0.659506082	2.579103472
RS_Waiver	-0.161239386	0	0.090084232	-0.111292247	-0.069987129	-0.22996626
District_Waiver	-0.023924944	-0.063603707	-0.690431609	0	-0.098945225	0
REGION_Waiver	0	0	0	-0.252275514	0	0
MCRC_Waiver	-0.519893278	0.040523384	0	0	-0.066148487	-0.51922818
GT_SCORE	0.010557403	0	0	0	0.003990427	0
MM_SCORE	0	0	0	0	0	0
CL_SCORE	0	0.044091468	0	0	0.019509694	0
EL_SCORE	0	0	0	0	0	0.081499747
AR_SCORE	0.01195178	0.01048751	-0.023344843	0	0.025047244	0
AS_SCORE	-0.0207882	-1.98E-05	0.003172634	0	-0.003585512	0
EI_SCORE	-0.001963264	-0.016569017	0	-0.039659262	-0.010541764	-0.028953291
GS_SCORE	0	-0.017476576	-0.037045717	0	-0.002196816	0
MC_SCORE	0	0.011653962	0	0	-0.018361452	0
MK_SCORE	0.022372499	0	0.091362483	0.026093405	0	0
PC_SCORE	-0.008616254	0.001197504	0	0	0.004003527	0
VE_SCORE	0	0	0	0	0	-0.014300957
WK_SCORE	-0.010133082	-0.018606873	0	-0.008354646	0	-0.085563412
BMI	-0.047536367	0	-0.131861249	-0.0145115	-0.01221974	-0.407043596
GENDER	-0.028545455	-0.003057268	2.169548342	0.086998456	-0.003882631	0.472630476
IST_PASS	-1.697224704	0	0	0	0	0
MARITAL_STAT	0	0	-0.103379449	0	0	0
MENTAL_GRP	0	0	0	0	0.109517218	0

THIS PAGE INTENTIONALLY LEFT BLANK

APPENDIX B. REGRESSION COEFFICIENTS: TIER I & II

	Admin	Intel	Infantry	Logistics	MAGTF Plans	Comm	Arty	Utilities
	01	**02**	**03**	**04**	**05**	**06**	**08**	**11**
AGE_ENLIST	0.018996627	0.037103803	0.014228784	0.013300049	0.026150584	0.008648068	0.004730373	0.013903197
CIV_ED	0.197513269	0.012203655	0.166085853	0.160465913	0.159216612	0.141739314	0.11162368	0.394668587
RS_Waiver	0	0.054013197	-0.033620701	0	0.459820795	0	0	0
District_Waiver	-0.161186012	0	-0.128252497	0	0	-0.023713573	-0.286453444	-0.221534082
REGION_Waiver	-0.143446204	0.078403008	-0.158942205	0.057214862	0	0	-0.044160183	-0.347754678
MCRC_Waiver	0	0	-0.068875016	0	-0.296443372	0	0	-0.009859738
GT_SCORE	0.015764576	0.04655018	0.007310523	0.006332117	0	0	0.022936904	0.013986827
MM_SCORE	0.003771671	0	0	0	-0.040983752	0.00543813	0	0.000509334
CL_SCORE	0.001720591	0.004695425	0	0.002955612	0	0	0	0
EL_SCORE	0	0	0	0	0	0	0	0
AR_SCORE	0	-0.015961319	0.00072346	0	0	0	0	0
AS_SCORE	0	0	0	0.002742931	-0.005684425	0.006657504	0	0.004037132
EI_SCORE	-0.003238193	-0.005442685	0	-0.006989392	-0.000909399	-0.001108261	-0.005922596	0
GS_SCORE	-0.000618856	-0.000506743	0	-0.014457197	-0.001489452	-0.001560417	-0.017196172	-0.013900138
MC_SCORE	-0.014987685	-0.019149669	0	0.003224849	0	0	0	0
MK_SCORE	0	0	0.008605647	0	0	0	0.013240592	0.013208805
PC_SCORE	0	3.67E-03	0.000285959	0	0.000508655	4.77866E-05	0	0.000709914
VE_SCORE	0	0	0	0	0	0	0	0
WK_SCORE	-0.033430584	-0.005948698	-0.00183371	0	0	-0.011306368	0	-0.009544813
BMI	-0.029893273	-0.043308022	-0.015032978	-0.025683304	-0.01871981	-0.025497846	-0.044258202	-0.027857126
GENDER	0.02260213	0.352019704	0	0.167931649	1.532164676	0	0	0.506395201
IST_PASS	0	-0.547592031	0	0	0	0	0	-0.022210598
MARITAL_STAT	-0.054145119	0.217744118	0	0.142796287	0.46251438	0.016117464	-0.177802645	-0.147342623
MENTAL_GRP	0	0.180816806	-0.007815329	0	0.33662258	-0.090170535	0	0

	Engineer	Tank/AAV	Ground Ord	EOD	SigInt	Data/Comm	Supply	Dist Mgmt
	13	**18**	**21**	**23**	**26**	**28**	**30**	**31**
AGE_ENLIST	0.014946994	0.007179659	0.010285512	0	0	0.040526347	0.01816785	0.067336432
CIV_ED	0.152386528	0.19281935	0.115270089	0.572631401	0.074992222	0.095864357	0.086587094	0.299667856
RS_Waiver	-0.044869877	0	0	0	0	-0.08099587	-0.012257679	0.134700438
District_Waiver	-0.076198499	-0.033933628	-0.146600538	0.003488231	-0.317627877	0	-0.146672565	-0.294159966
REGION_Waiver	0	0	-0.030817049	0.110956536	-0.261818244	-0.067177465	-0.370781601	-0.33838924
MCRC_Waiver	-0.054020748	0.045339624	0	0	0.162321664	-0.093264141	0	-0.130411382
GT_SCORE	0	0	0.008190429	0.024797865	0	0	0.009316673	0
MM_SCORE	0	0	0	0	0	0	0.000254084	0
CL_SCORE	0	0.002558427	0	0	0	0	0.008663424	0.018799958
EL_SCORE	0	0	0	0	0	0	0	-0.031586797
AR_SCORE	0.003113956	0	0	0.010012573	0.024935721	0.001062139	0	0
AS_SCORE	-7.90056E-05	0	0	0	0	0	0	-0.002801069
EI_SCORE	-0.002132968	-0.009360733	0	0	-0.017161762	-0.009241043	0	0
GS_SCORE	-0.006492788	0	-0.006287943	0	-0.002940949	0	0	-0.007033799
MC_SCORE	0	0.020375647	0.003186251	-0.015285765	0.002581321	0	0	0.047945432
MK_SCORE	0.010153272	0.001713903	0.008550036	0	0	0.003340226	0	0.05307177
PC_SCORE	1.76E-03	-0.000221145	0.005216296	0	0.038295858	0.022074588	0	0
VE_SCORE	0	0	0	0	0	0	-0.00250466	0
WK_SCORE	-0.007924137	0	0	-0.000228923	0	-0.009383683	-0.003574222	0
BMI	-0.033756715	-0.053936513	-0.038121568	-0.064732549	-0.028765096	-0.034002604	-0.029136142	-0.022220657
GENDER	0	0	0	0	0.451781963	0	0.080882375	0.156809593
IST_PASS	0	-0.071096182	0	1.164476035	1.401124838	-0.044904652	1.817643234	0
MARITAL_STAT	-0.097470011	0	-0.057018523	0	0	0	0	0
MENTAL_GRP	0	0	0	0.041306177	0	0	0	0.015725969

	Food Service	Financial Mgmt	Motor Trans	Public Affairs	Legal	Combat Cam	Recruit & Retention	Music
	33	**34**	**35**	**43**	**44**	**46**	**48**	**55**
AGE_ENLIST	-0.023931533	0	0.005847061	-0.030506073	0	0	-0.096792104	-0.03039864
CIV_ED	0.005719575	0.26964255	0.072575418	0.159154162	0.627147418	0.260799874	0.668747095	0.131501186
RS_Waiver	-0.076129286	0.212995139	0	0	0	-0.057675407	0.734996436	0.195027247
District_Waiver	-0.337537741	-0.086724033	-0.096406056	-0.850454951	-0.449284836	0.026165711	0	-0.28526198
REGION_Waiver	0.035183914	0.281217482	0	1.752593028	0.026199707	-0.150998409	-0.987815029	0.119278476
MCRC_Waiver	0.038964473	-0.05109234	0	-0.519703059	0.049554366	0	0.171100057	0.150891265
GT_SCORE	0	0	0	0	0	0	0.092847354	0
MM_SCORE	0	0	0	0	0	0	0	0
CL_SCORE	0	0.013738715	0	0.000376989	0	0	0	0
EL_SCORE	0	0	0	0.00889759	0	0.010254086	-0.043528542	0.013891693
AR_SCORE	0	-0.001209027	0.006224994	0	-0.004465299	0	-0.062557023	0
AS_SCORE	-0.013322974	0	-7.52945E-05	-0.026083345	-0.014452525	0	0.050142131	-0.013946269
EI_SCORE	0	0	-0.001849277	-0.041871131	0	0	-0.048258431	0.003759535
GS_SCORE	-0.007939279	0	-0.003583698	0.003654614	0	0	0	0.01591266
MC_SCORE	0	-0.003463655	0.000402593	0.010136828	0	0	-0.008458736	-0.015320553
MK_SCORE	0	0	0.004223162	0.016835629	-0.005525351	0	0	0
PC_SCORE	0	-0.007024412	0.00940326	0.012959216	0.001082959	0.003537689	0	0
VE_SCORE	0	0	0	0	0	0	0	0
WK_SCORE	0	0	-0.004634969	-0.001513003	0	-0.079320772	0.002869516	-0.01200898
BMI	-0.022803687	-0.02965192	-0.026137064	-0.015484158	0.003053255	-0.00642596	-0.247901082	-0.009537068
GENDER	0.143604282	0.17621146	0.161057307	0.579885562	0	0.150420004	0.770861039	-0.030850433
IST_PASS	-0.060719337	0	1.645532053	-2.486548798	0	0	0	-0.768820666
MARITAL_STAT	-0.278070792	-0.183672425	0.00666362	0	0	-0.630275112	-0.293065679	0.746148632
MENTAL_GRP	0	0	-0.003824904	0	0	-0.197384543	0	0

	CBRN	MP	Elect Maint	Aircraft Maint	Aircraft Maint	Aircraft Maint	Avionics	Avionics
	57	**58**	**59**	**60**	**61**	**62**	**63**	**64**
AGE_ENLIST	0.037702885	0.009941838	-0.029737721	0.034321359	0.025846848	0.019835127	0.019214775	0.013752118
CIV_ED	0.196698829	0.002206743	0	0.193900555	0.091830038	0.147991834	0.26482044	0.147587734
RS_Waiver	0.006509558	-0.021313395	0	0.021408671	-0.006361316	0.005902852	-0.095997067	0.042093481
District_Waiver	0.040018906	0	-0.706026108	-0.072140222	-0.045523694	-0.048885986	-0.016599057	-0.178959901
REGION_Waiver	0	-0.076004937	0	-0.047868671	0	-0.048556115	0	0.237667905
MCRC_Waiver	0	0.016106039	0.020289946	-0.031775567	-0.098998875	-0.015608265	0.027306248	0.000493767
GT_SCORE	0	0	0	0.003376688	0.00371353	0	0	0
MM_SCORE	0	0	0	0.007249864	0	0	0	0
CL_SCORE	0	1.84822E-05	0.029115433	0	0	0	0	0
EL_SCORE	0	0	0	0	0	0	0	0
AR_SCORE	0	0	0.007749813	0	9.30E-03	0	-0.002831642	0.018303891
AS_SCORE	0	0	0.022811072	0	0	0.000629569	-0.003700543	0
EI_SCORE	-0.006966692	0	-0.013222742	-0.008452054	0	0	-0.008318003	0
GS_SCORE	-4.3481E-05	-0.001817097	-0.002594436	0	-0.008521358	2.86E-05	0	0
MC_SCORE	0	0.002129179	-0.002502285	0.003612003	0	0	0.027337635	0.00181948
MK_SCORE	0.021847146	0.013000849	0	0.01621885	0.001126295	0	0	0.011564567
PC_SCORE	0.012641712	0	0.023216424	0.003879358	0.009738282	0.003020657	0.005302972	0
VE_SCORE	0	0	0	0	0	0	0	-0.031202346
WK_SCORE	0	-0.003400584	-0.080793104	-0.011238319	-0.005624876	-0.013204513	-0.002579435	0
BMI	-0.048819704	-0.04380551	-0.030645935	-0.020062867	-0.038784005	-0.023628598	-0.049749995	-0.023061794
GENDER	0.089050095	0.150435779	0.460636314	0.043798363	-0.057095192	0.721060851	0	0.078289472
IST_PASS	0	0.217647468	1.38273051	1.621893679	-0.348199063	0	0	-0.789435174
MARITAL_STAT	-0.157180905	0.019410497	-0.765401092	-0.065821097	0.149534689	0.096810218	0	0.01982736
MENTAL_GRP	0	0.029766622	-0.010558694	-0.028526032	0	0	0	-0.020599616

	Aviation Ord	Avionics Log	METOC	Airfield Serv	Air Control/Air Supt/AAW/ATC	Flight Crew
	65	**66**	**68**	**70**	**72**	**73**
AGE_ENLIST	0	0.026446817	0.089924796	0.029610748	0	0
CIV_ED	0.09924808	0.349696961	0.141864172	0.034455643	0.213351234	0
RS_Waiver	0	-0.064240783	-0.341826445	-0.004375761	0.001633875	-0.054833511
District_Waiver	-0.392542657	-0.224895255	0	0	-0.473722079	-2.658724414
REGION_Waiver	-0.212241294	0	-2.77689278	0	0	0
MCRC_Waiver	0.098280508	-0.044492692	0.327389227	0	0	0.310374853
GT_SCORE	0	0	0	0	0	-0.075086767
MM_SCORE	0	0	0	0	0	0
CL_SCORE	0	0	0	0	0	0
EL_SCORE	0	0	0	0	0	0
AR_SCORE	0.020624294	0	-0.006914372	0.004729443	-0.009808062	0
AS_SCORE	0	0.00E+00	-0.001503571	0	0.003585512	0
EI_SCORE	0.001963264	-0.004091607	0	0	0.000565595	0.0177781
GS_SCORE	-0.003385824	-0.004381922	-0.001410346	-0.021246808	-0.019062374	0
MC_SCORE	0	0	0	0	0	0.056816039
MK_SCORE	0.003848024	0.000266042	0	0.007082396	0.008059784	0.000814989
PC_SCORE	0.005173663	-0.005189784	0	0	-0.004003527	-0.005099138
VE_SCORE	0	0	0	0	-0.020862407	-0.105313412
WK_SCORE	-0.015463753	0	0	0	0	0
BMI	-0.01329797	-0.048981874	-0.016167739	-0.05740172	-0.022598647	0
GENDER	0.773084217	0.394180237	0.026380232	0.107872035	0.019871012	-1.408831318
IST_PASS	-0.273032974	-3.578538098	0	-3.089421489	0	0
MARITAL_STAT	-0.354050552	0.339439186	0.41577754	0	-0.450748715	-1.359027025
MENTAL_GRP	-0.029393601	0	-0.604886808	0	-0.285103061	-0.200839489

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF REFERENCES

About adult BMI. (2015, May 15). Retrieved from: http://www.cdc.gov/healthyweight/assessing/bmi/adult_bmi/index.html

Cole, L. A. (2014). U.S. Marine Corps enlisted retention: An analysis of stakeholder incentives for the retention of tier I first term Marines. Master’s thesis, Naval Postgraduate School, Monterey, CA.

Commandant of the Marine Corps. (2010). *35th Commandant of the Marine Corps commandant’s planning guidance*. Washington, DC: Headquarters United States Marine Corps.

Official site of the ASVAB. (n.d.). Retrieved February 23, 2016, from http://www.official-asvab.com/applicants.htm

Faraway, J. J. (2015). *Linear models with R* (2nd ed.). Boca Raton, FL: CRC Press, Taylor & Francis Group.

Friedman, J., Hastie, T., Simon, N., & Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software, 33*(1), 1–22.

Friedman, J., Hastie, T., Simon, N., & Tibshirani, R. (2016, February 23). Lasso and elastic-net regularized generalized linear models. *Package ‘glmnet.’* CRAN.

GAMS Development Corporation. (2014). *General algebraic modeling system (GAMS)* (GAMS Distribution 24.7.1). Washington, DC: Author.

Gibson, C. (2016, February 11). *Protecting our security through utilizing right-sized end-strength act of 2016* (H.R. 4534). Washington, DC: 114th Congress, 2D Session.

HQMC. (2008, Aug 8). *Marine Corps body composition and military appearance program* (MCO 6110.3 W/CH 1). Washington, DC: Author.

HQMC. (2011, May). *Commandant approved updated reenlistment procedures* (MARADMIN 273/11). Washington, DC: Author.

HQMC. (2014, Dec 8). *Military occupational specialties (MOS) program manual (Short title: MOS manual order)* (MCO 1200.18). Washington, DC: Author.

HQMC. (2015). *Military Occupational specialties manual* (NAVMC 1200.1A W/ CH 1). Washington, DC: Author.

Kapp, L. (2013). *Recruiting and retention: An overview of FY2011 and FY2012 results for active and reserve component enlisted personnel* (CRS Report No.RL32965l). Washington, DC: Congressional Research Service.

Koopman, M. E. (2007). *Improving reenlistment incentives and processes.* Alexandria, VA: Center for Naval Analysis.

M&RA. (2015, September 15). Quality comparison worksheet. PowerPoint Presentation. Quantico, VA: Author.

R Core Team. (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing (Version 3.1.1)[Computer software]. Retrieved from URL https://www.R-project.org/.

Rautio, L. J. (2011). *Study of standards used to screen recruits for assignment to the communications field in the U.S. Marine Corps*. Master's thesis, Naval Postgraduate School, Monterey, CA.

Wathen, W. G. (2014). *Identifying factors that predict promotion time to E-4 and re-enlistment eligibility for U.S. Marine Corps field radio operators*. Master's thesis, Naval Postgraduate School, Monterey, CA.

Zou, H., & Hastie, T. (2005, Sept). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B*, *67*(2), 302–303.

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California