AN EMPIRICAL ANALYSIS OF THE DETERMINANTS OF INITIAL OCCUPATIONAL CHOICE BY MALE HIGH SCHOOL GRADUATES

by

Donald Francis Cox

Dissertation submitted to the Faculty of the Graduate School
of the University of Maryland in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

1986

Copy 1

APPROVAL SHEET

Title of Thesis: AN EMPIRICAL ANALYSIS OF THE DETERMINANTS OF INITIAL OCCUPATIONAL CHOICE BY MALE HIGH SCHOOL GRADUATES

Name of Candidate: Donald Francis Cox

Doctor of Philosophy, 1986

Thesis and Abstract Approved: Frank Brechling

Dr. Frank Brechling

Professor

Department of Economics

Date Approved: April 30 1986

ABSTRACT

Title of Dissertation: AN EMPIRICAL ANALYSIS OF THE DETERMINANTS OF INITIAL OCCUPATIONAL CHOICE BY MALE HIGH SCHOOL GRADUATES

Donald Francis Cox, Doctor of Philosophy, 1986

Dissertation directed by: Dr. Frank Brechling, Professor, Department of Economics

This dissertation consisted of an empirical analysis of the determinants of initial occupational choice by male high school graduates. The approach used was based on the theory of random utility. According to this approach, the individual selects a particular outcome from a set of possible outcomes based on both observed and unobserved characteristics of the individual and the particular possible outcome. In this analysis, the occupational choice set contained three possible outcomes. These possibilities were civilian sector employment, military service and college enrollment.

For the empirical analysis, a sample of 1,748 male high school graduates was drawn from the National Longitudinal Survey of Youths (1979-1981). The empirical model consisted of a mixed discrete/continuous simultaneous 4 equation system. Three estimation strategies were used. The first was a simple two stage

logit/ordinary least squares procedure. The second was a modified two stage logit/ordinary least squares procedure that corrected for self-selectivity bias. The third strategy consisted of a modified two stage logit/ordinary least squares procedure that corrected for both self-selectivity and choice-based sampling bias.

The estimation results indicate that the decision to enlist is most sensitive to the net income of the individual's family and the predicted civilian sector wage. The military experience of the individual's father and the desire to acquire additional training are also important in this decision. In addition, the differences in the estimates across the three estimation procedures illustrate the importance of correcting for sample biases.

This dissertation is dedicated to the memory of

Joseph E. Cox

CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction and Background Information

1. Introduction and Statement of Research

Since the inception of the All Volunteer Force in 1973, a considerable amount of research has been devoted to the study of military enlistment supply. For the most part, however, this research has been limited to the analysis of aggregate level data of a predominately economic nature. The present research consists of an empirical analysis of the determinants of enlistment that differs in both the type of data and analytical approach from most of the previous work in this area.

This chapter has several purposes. The following section provides a description of the present analysis. Sections 2 and 3, respectively, provide background information on the current Armed Force and previous research on enlistment supply. The last section is reserved for an outline of the remaining chapters.

1.1 Statement of Research

While the analysis of aggregate level data has proved useful in forecasting enlistment supply,[1] it has not been able to provide the needed insight into the actual determinants of the enlistment decision. The present analysis utilizes a micro level longitudinal data base drawn from the National Longitudinal Survey of Youths (1979-1981) that is rich in economic, sociological, and demographic information about the potential enlistee. By the use of such data, specific hypotheses with respect to the determinants of enlistments, as well as other initial occupational choices, can be tested. A potential limitation of the present research is that, unless the sample size is sufficiently large to permit disaggregation by service, the analysis will not be service specific.

1.1.1 Hypotheses

The hypotheses that will be tested in the empirical portion of the analysis center on the effects of individual specific attributes in the determination of initial occupational choice. These hypotheses are grouped under the following three headings.

i.) The role of socio-psychological and family-demographic characteristics in the enlistment decision. These characteristics include motivation, parental education levels, the number of family members currently serving in the military, attitudes, marital status, father's occupation and other characteristics that could be important to the enlistment decision as well as other initial occupational choices.

ii.) The role of economic attributes in the enlistment decision: Previous studies have utilized average aggregate level economic attributes. These attributes are rather imprecise in explaining and predicting individual behavior. The present analysis will include economic measures that are more individual-specific. This will allow testing the effects on the enlistment decision of acquired human capital, civilian wages (actual and expected), individual-relevant unemployment rates and the ability to finance a college education.

iii.) The impact of recruitment activities on the enlistment decision: Prior analysis on this aspect of the enlistment decision has met with limited success. This failure could be attributed to imprecise instruments, insufficient variation in the explanatory variables, or model misspecification. In the current analysis, recruitment activity will be measured by individual contact with a recruiter, advertising (location specific), and recruiter density.

1.2 Analytical Approach

The analytical approach consists of the estimation of a mixed discrete/continuous simultaneous equations model of occupational status for male high school graduates up to one year after receiving a high school diploma or General Education Diploma (G.E.D.). The occupational status choice set consists of three possibilities: civilian employment, college enrollment or military service. The sample selection is based upon several considerations. First, during the period of analysis (1979-1981), high school graduates can be viewed as unconstrained by military manpower demand considerations. Secondly, by allowing approximately one year to elapse before observing the individual's occupational status, a more accurate assessment of the initial career decision can be made. Lastly, this sample specification will simplify the analysis by excluding occupational switching, which would complicate the analysis if the individual were tracked for more than one year.

A major difficulty with qualitative choice models is sample self-selectivity bias. Various techniques have been developed to correct for this bias in binary choice models. However, very limited work has been done to extend these correction procedures to the trichotomous case. The present analysis will include a trichotomous correction procedure that is based upon the method suggested by Lee (1982).

1.3 The Data

The data base will consist of a sub-sample of approximately 2,000 observations from the National Longitudinal Survey of Youth, 1979-1981. Approximately 650 of these observations are enrolled in college, 230 are enlisted in the military and 1,180 are employed in civilian sector jobs. This sub-sample will be augmented with locational specific information on economic conditions and recruitment related activities. The military data will be drawn from the Defense Manpower Data Center's (DMDC) Enlistment Master Files.[2]

2. Background Information on the All Volunteer Force

The current commitment to an All Volunteer Armed Force is not a unique experience in United States history. Rather, historically, U.S. Armed Services have satisfied their manpower requirements using volunteers. Only during periods of national emergency has this country turned to a policy of mandatory conscription to satisfy these unexpected manpower requirements. What is unique about the current force is its size. In particular, the current force is approximately five times the size of the pre-World War II force (the last pre-draft era force).[3]

The first serious movement towards the reinstitution of an all volunteer force came in 1969 with President Nixon's establishment of the Gates Commission. The resulting commission report recommended a return to an all volunteer force accompanied by an increase in military wages, inproved recruiting activities and the establishment of a standby draft system.[4] Almost three years were to elapse before the then Secretary of Defense, Mr. Melvin Laird, was to announce the end of the draft in January 1973.

Since the return to the policy of voluntary enlistment, the force has been periodically subjected to critical commentary about its continuing viability.[5] Typically, these comments focus on the quantity and quality of the incoming recruits. Actual recruiting performance has lent mixed support to these comments. Table 1-1 provides a breakdown of actual recruiting trends for the Army. Army data is reported because, out of the four services, the Army has had the most difficulty in achieving its recruitment objectives.[6]

The first column of Table 1-1 illustrates that, except for Fiscal Years (FY) 1978-1979, the Army has been able to satisfy its recruiting objectives. Two measures of recruit quality are reported in columns 2 and 4. With the exception of FY1979 and FY1980, the Army has been increasingly able to attract a higher proportion of high school graduates. To control for general population trends in educational attainment, the ratio of high school graduate recruits to 18 year old high school graduates is reported in column 3. The data in this column indicates that the Army had a below average proportion of high school recruits for FY1974-FY1980. However, ever though data for the percentage of population high school graduates in the most recent years is unavailable, an overall trend towards a higher proportion of high school graduates can be seen. Information on recent recruiting performance has tended to support this trend.[7]

Table 1-1: Army Recruiting Trends (non-prior service enlistments)

Fiscal Year	% of Manpower Obj. Filled (1)	% High Sch. Graduates (2)	% High Sch. Grads/Pop % H.S. Grad (3)	% TCAT I-IIIA (4)	% Black (5)	% Blk/ % Blk Pop (6)
1974	98.7	50.1	0.66	52.5	27.2	2.39
1975	100.4	57.8	0.78	57.6	23.0	2.00
1976	100.1	58.6	0.79	54.8	24.4	2.12
1977	100.3	59.2	0.80	34.2	29.4	2.53
1978	97.7	73.7	0.99	37.9	34.3	2.93
1979	86.7	64.1	0.87	30.6	36.8	3.12
1980	100.2	54.3	0.75	26.0	29.8	2.53
1981	101.0	80.3	1.11	40.0	27.4	2.32
1982	104.1	86.0	*	53.0	24.6	2.07
1983	100.3	87.6	*	61.4	22.0	*

* Data not available.

Sources by column number:

(1), (2), (4), (5) : Data provided by S. Castledine, Department of the Army Personnel Office of the Deputy Chief of Staff for Military Personnel Accessions (DAPE-MPA).

(3) : Pop % H.S. Grad refers to the percentage of 18 year olds that are high school graduates. The Statistical Abstracts of the United States: 1984, Table No. 255, pp. 160, Dec. 1983.

(6) : The Statistical Abstracts of the United States: 1984, Table No. 32, pp. 32, Dec. 1983 (for years 1979-1982). Data for 1974-1978 came from The Statistical Abstracts of the United States: 1979, Table No. 27, pp. 28, Sept. 1979.

A second indicator of recruit quality is the relative performance of a group of recruits on the Armed Forces Qualification Test (AFQT).[8] The AFQT is a measure of the general "trainability" of a military applicant. It is computed by combining various subtests on the Armed Services Vocational Aptitude Battery (ASVAB) and used by all services to select recruits. The AFQT is a percentile score that reflects an applicant's standing as compared to a particular reference population. The AFQT percentile scores are divided into five test categories with a further division of two of the test categories into subcategories:

TCAT	AFQT Percentile Scores
I	93-99
II	65-92
IIIA	50-64
IIIB	31-49
IVA	21-30
IVB	16-20
IVC	10-15
V	1-9 .

By law, TCAT V personnel have always been ineligible for military service. Applicants who score in the upper 50 percent on the AFQT are classified into Test Categories (TCAT) I-IIIA and are considered to be above average in quality.

The decline in the percentage of TCAT I-IIIA Army Recruits during FY1977-FY1980 is primarily due to a calibration error in the operational ASVAB used during these years. When the error was detected and corrected norms were applied to the recruits accessed during those years, the number of TCAT IV accessions rose from 10 percent to 45 percent. In addition, prior to October 1984, all versions of the AFQT had been normed using the 1944 mobilization population as the reference population.[9] Renorming the test scores to the 1980 U.S. youth population (18-23 years old) resulted in an additional, albeit small, decline in the percentage of TCAT I-IIIA accessions. Even so, the years following the recalibration (FY1981 to present) indicate a steady increase in the percentage of TCAT I-IIIA recruits.

A third criticism periodically directed to the current force is that it is disproportionately manned by minorities.[10] The information in columns 5 and 6 tend to support this, particularily for the late 1970's. Since FY1979, however, the trend has been towards a more proportionate representation of blacks in the military.

In general, the information in Table 1-1 indicates that the Army has been able to attract a sufficient quantity of recruits, while upgrading the quality of these recruits (particularily in the most recent years). The ability of the Army, and the Armed Forces in general, to maintain this trend in the coming years may become increasingly difficult. The projected demographic trends illustrated by Figure 1-1 indicates that over the next decade the eligible potential recruit pool will decline considerably.

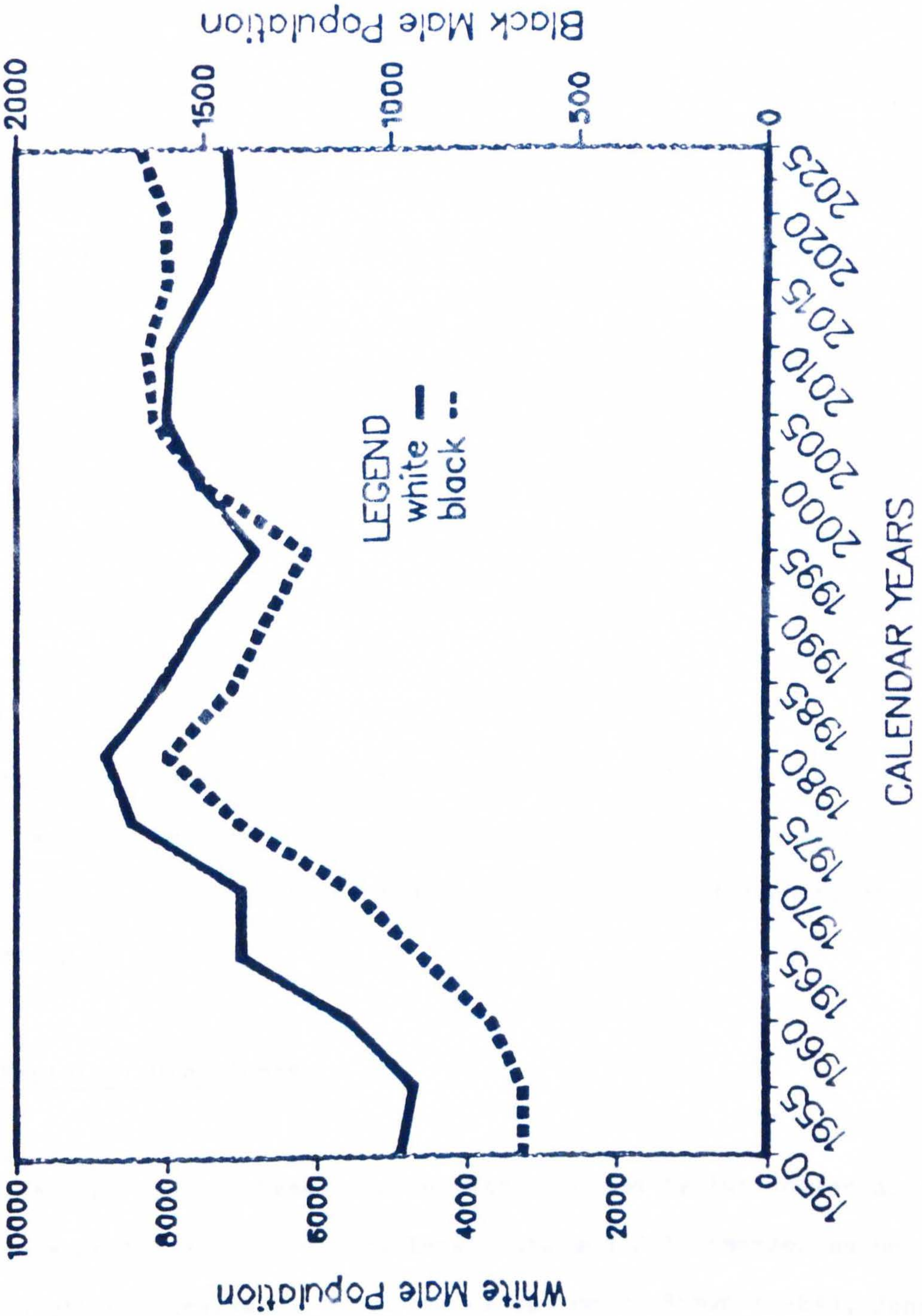
3. A Brief Discussion on Prior Research

The bulk of reseach on enlistment supply has been of an empirical nature. What follows is a brief discussion of the theoretical underpinnings and major empirical findings in this area.[11]

3.1 Theoretical Background

The economic theory of military enlistments is dominated by supply side utility maximization principles.[12] The lack of demand side analysis is largely attributed to the peculiar characteristics of military enlistment demand. As stated by Brown (1984), "... the Armed Forces are neither pure price takers nor pure quantity takers. Rather, they attempt to fill a predetermined number of positions at a predetermined wage, with recruit quality varying to equate supply and demand."[13] This implies that recruits which are deemed high quality (i.e.- Test Categories I-IIIA) are not demand constrained and, given the number of high quality recruits, the residual number of lower quality recruits is determined via a more standard market clearing process.

Figure 1-1: MALE POPULATION
BY RACE
Ages 17–21



SOURCE: Bureau of the Census, U.S. Dept. of Commerce
(Unpublished Tabulations)

The theory of enlistment supply was first explicitly discussed by Oi (1967) and Fisher (1969). Simply stated, the theory predicts that an individual will enlist if military monetary compensation exceeds the sum of civilian monetary compensation and taste (distaste) for military service. Or, using Fisher's notation, the individual will enlist if

$$(1.1) \qquad W_m > W_c(1 + D),$$

where W_m and W_c are the respective military and civilian income streams (discounted to present value) and D is a measure of taste (distaste) for military service.

This basic framework has been the foundation for most of the empirical work in this area. Subsequent researchers have focussed on the more technical aspects of model specification and econometric estimation.[14]

3.2 Empirical Highlights

Previous empirical analyses can be characterized by two criteria: 1.) Aggregate or individual (micro) level data and 2.) time-series or cross-sectional analysis. With the exception of Brown (1984), Jehn and Shughart (1977), Ellwood and Wise (1984) and Daula and Smith (1984), all of the aggregate level studies have also been time-series analyses. A summary of the major characteristics and findings of these studies is presented in Table 1-2. The studies covered in this summary should not be considered exhaustive of the work done in this area, but merely representative.[15] These highlights are classified as either economic, recruiter or demographic related.

3.2.1 Economic Findings

The two major economic issues relevent to enlistment supply are the effects of wages (directly) and unemployment (both directly and indirectly). Most of the studies have found significant and positive (negative) military (civilian) wage effects. The range of estimated wage elasticities, however, is quite large. This diversity of findings is partially attributed to the branches studied, specification of the wage variable, period of analysis, and stratification of the data by race and/or recruit quality.

While these estimates range from very inelastic (Ash, Udis and McNown) to rather elastic (Dale and Gilroy, Baldwin, et. al. and Daula, et. al.), several underlying patterns have been fairly consistently reported. First, almost all of the studies that included race effects found higher wage elasticities for whites vs blacks.[16] In addition, those individuals deemed high quality had larger estimated wage elasticies than those of lower quality individuals.[17]

The estimated effects of civilian labor market unemployment have been somewhat less consistent. One would expect significant unemployment

Table 1-2: Summary of Previous Empirical Findings

Author(s)	Period of Analysis	Branches[1] Analyzed	Elasticities[2]		Recruitment Effects[3]		Demographics	
			Wage	Unemployment	Recruiters	Advertising	Race	Others
(time-series)								
Ash, Udis and McNown (1983)	1967:2-1976:2	A, N, AF, MC	-.088 to -.99[4]	.015 to .135[5]	NA	NA	Y[6]	NA
Cooper (1977)	FY71-FY75	D	.95 to 1.23[7]	.11 to .27	NA	NA	Y[6]	NA
Dale and Gilroy (1983a)	10/75-3/82	A, N, AF, MC	2.30[7, 8]	.94	YS[9]	NA	Y[8]	NA
Dale and Gilroy (1983b)	10/75-3/82	A	.6 to 2.12[7]	.69 to 1.3	YS[9]	NA	Y[6]	NA
DeVany and Saving (1982)	6/69-6/76	AF	1.41[7]	-1.60[10]	YS[11]	NA	NA	NA
Fechter (1978)	1958:2-1974:4	D	.64 to 1.5[7]	-1.4 to -2.6[5, 10]	NA	NA	NA	NA

Table 1-2 (cont) : Summary of Previous Empirical Findings

Author(s)	Period of Analysis	Branches[1] Analyzed	Elasticities[2]		Recruitment Effects[3]		Demographics	
			Wage	Unemployment	Recruiters	Advertising	Race	Others
(time-series)								
Fernandez (1979)	7/70-9/78	A, N, AF, MC	-1.36 to[7] .48	.65 to 1.4	YNS	NA	NA	NA
Fisher (1969)	1957:3-1965:4	D	-.46[4]	YNS	NA	NA	NA	NA
Goldberg (1980)	1971:3-1977:4	N	YNS[7]	YS[12]	YS	YNS	NA	NA
Grissmer (1977)	6/70-6/75	D, A, N, AF, MC	.50 to[7] 1.70	-.84 to 1.25	NA	NA	Y[6]	NA
Horne (1984)	1977:2-1984:2	A	2.70[7]	.73	YS	YS	NA	NA

Table 1-2 (cont) : Summary of Previous Empirical Findings

Author(s)	Period of Analysis	Branches[1] Analyzed	Elasticities[2]		Recruitment Effects[3]		Demographics	
			Wage	Unemployment	Recruiters	Advertising	Race	Others
(cross-sectional)								
Baldwin, Daula and Fagan (1982)	1978	D	2.30[7]	.94[13]	YS	NA	Y	Y[14]
Daula, Fagan and Smith (1982)	1978	D	2.30[7]	3.36	YS[11]	NA	Y	Y[15]
Detrouzos[16] (1984)	1980-1981	A	-.81 to -.16[17]	.31 to .58	YS	NA	NA	NA
Jehn and Shughart (1977)	1973	N	-1.69 to -1.96[4]	.14 to .38	YS	NA	Y[18]	Y[19]

Table 1-2 (cont) : Summary of Previous Empirical Findings

Author(s)	Period of Analysis	Branches Analyzed[1]	Elasticities[2]		Recruitment Effects[3]		Demographics	
			Wage	Unemployment	Recruiters	Advertising	Race	Others
(pooled ts-cs)								
Brown (1984)	1975:4-1982:3	A	.5 to 1.5[20]	.4 to .8	YS	YNS	NA	NA
Daula and Smith (1984)	10/80-6/83	A	.49 to 1.89[7]	.28 to 1.36	YS	YS	Y[18]	Y[21]
Ellwood and Wise (1984)	1973-1982	D	.36 to .62[20]	.01 to .05	NA	NA	Y	NA

1. D = All branches combined, A = Army only, N = Navy only, AF = Air Force only, MC = Marine Corps only.
2. Only statistically significant estimates reported.
3. Y = Included, YS = Included and statistically significant, YNS = Included but not statistically significant, NA = Not included.
4. Estimated from the ratio of civilian to military wages.
5. Most estimates statistically insignificant and/or wrong signed.
6. Estimated seperate equations by race.
7. Estimated from the ratio of military to civilian wages.
8. Reported estimates are for the Army only.
9. Used a dummy variable to approximate increased recruiter effort between 11/79 to 8/81.
10. Estimated from the employment rate.
11. Statistically significant, but wrong signed.
12. Elasticity not reported.

13. Estimate based on cyclical change in the unemployment rate. The unemployment rate by itself was statistically insignificant.
14. Variables included were southern (<0 and insignificant) and north-central (>0 and significant) location of residence.
15. Variables included were marital status (>0 and insignificant), black/hispanic (<0 and insignificant), south (>0 and insignificant) and north-central (>0 and significant) location of residence.
16. Estimated seperate equations for 1980 and 1981.
17. Estimate based on civilian wage only.
18. Measured as the percentage black residents in the district.
19. Consisted of the percentage of residents in an urban area (>0 and insignificant), median education level (>0 and significant) and per capita income (<0 and significant) for the district.
20. Estimate based on military wage only.
21. Consisted of the percentage of Republican voters in the district (>0 and significant) in 1980.

rate effects.[18] For most of these studies this is not the case. Rather, the estimated effect has been insignificant and/or incorrectly signed and sensitive to changes in the functional form. In the studies that found statistically significant effects, the elasticities have been generally quite low. Comparisons across race and recruit quality have resulted in fairly consistent differences in effects. Blacks were less sensitive to unemployment effects than whites. High quality recruits were more sensitive to the unemployment rate than lower quality recruits.

A third economically relevant issue is the effectiveness of various military college assistance programs. These programs were the GI Bill (up to 1977) and the Veterans Educational Assistance Program (from 1977 to the present).[19] Only Dale and Gilroy (1983a,1983b), Brown (1984) and Daula and Smith (1984) have empirically tested the effectiveness of these programs.[20] With the exception of Daula and Smith (1984),[21] the findings of these studies generally indicate that these programs have been successful in attracting recruits. Brown's results are particularily surprising because the estimated effect of the VEAP was larger than that of the military wage.[22]

3.2.2 Recruitment Effects

Recruitment activities consist of advertising and recruiter effort. The analysis of the effectiveness of these activities, while important to military manpower planners, has yielded generally inconclusive results. Most studies found negligible and/or statistically insignificant effects. The extreme case is Daula, Fagan and Smith (1982) who found statistically significant, but negative recruiter impacts.[23] The most consistently significant (over model specification) estimates are those of Brown (1984), Dertouzos (1984), Goldberg (1980), Jehn and Shughart (1977) and Horne (1984). Dale and Gilroy (1983b) found insignificant results when measuring recruiter effectiveness by the number of production recruiters. However, by using a dummy variable approximation for recruiter effort, a significant positive relationship was discovered.

Brown (1984), Daula and Smith (1984), Goldberg (1980) and Horne (1984) included advertising measures. Brown (1984) and Goldberg (1980) had generally nonsignificant findings. Brown used both local and national advertising expenditures. His results indicated that, at best, only national advertising had a positive and significant estimated effect. However, this result was not consistent across equation specifications. Local advertising had no discernable effect. Goldberg (1980) also found consistently insignificant effects across samples and functional forms.

Daula and Smith (1984) and Horne (1984) had somewhat better success. Both of these studies included national and local advertising measures. Daula and Smith found both of these measures to be positive and statistically significant. In addition, national advertising (measured by potential exposure time) was found to have a consistently greater effect than local advertising expenditures. Horne estimated similar effects for national advertising, but local advertising was found not to be statistically significant.

3.2.3 Demographic Effects

The only demographic effects used in the time-series studies consisted of the estimation of different equations for blacks and whites. The results generally indicate a lower sensitivity to changes in wages and unempolyment rates for blacks in comparison to whites.

The cross-sectional and pooled studies were able to introduce additional demographic measures due to the variation in demographic characteristics between individuals and/or local regions. Daula, Fagan and Smith (1982) and Baldwin, Daula and Fagan (1982) were the only studies to utilize individual level data. The variables that these studies used, however, were restricted to locational and marital status dummy variables. Their significant finding was that individuals from the north-central part of the country had a higher probability of enlisting than individuals from other regions.

Jehn and Shughart (1977) used average demographic characteristics of the recruiting district. Their results indicate that districts with a higher median education and a lower per capita income have a higher rate of enlistment.

4. Outline of Chapters

The remainder of this research is organized into four chapters. The following chapter will present a model of occupational choice, as applied to military enlistment. The relevant econometric issues and empirical model hypotheses will also be discussed in this chapter.

The third chapter is used to describe how the data set was constructed. In order to generate a "feel" for the data, various descriptive statistics are presented.

The empirical model estimation results will be presented in the fourth chapter. These results will then be compared with the empirical findings of prior studies. A summary of the empirical results and a discussion of potential future research is given in the last chapter.

End Notes

1.) For example see Dale and Gilroy (1983a,1983b), Fernandez (1979) and Horne (1984).

2.) See Chapter 3 for a description of this data.

3.) For more detailed background information on the current volunteer force see Janowitz and Moskos (1979) and Bachman, Blair and Segal (1977).

4.) See The Report of the President's Commission on an All-Volunteer Force (1970), pp. 10.

5.) For example see Binkin (1984) or Janowitz and Moskos (1979).

6.) See Holden (1980).

7.) See the Washington Post, Dec. 9, 1984.

8.) The following discussion is based on information provided by Ms. F. Grafton, Data Base Management Project Leader, Manpower and Personnel Policy Research Group, U.S. Army Research Institute for the Behavioral and Social Sciences, Alexandria, VA.

9.) See Office of the Assistant Secretary of Defence/Manpower, Reserve Affairs, and Logistics (1980) or Shields and Grafton (1983) for a discussion of the misnorming problem.

10.) See Binkin (1982).

11.) For a more extensive review of the literature, see Morey and McCann (1983).

12.) Notable exceptions are DeVany and Saving (1982), Detrouzos (1984), Ellwood and Wise (1984) and Daula and Smith (1984).

13.) Pp. 4.

14.) These issues include the specification of the dependent variable (accessions vs contracts signed), the appropriate independent variable lag structure(s) and the appropriate model estimation technique. See Brown (1984), pps. 3-11 for a discussion of the first two types of issues.

15.) For additional examples of empirical work in this area see Altman and Fechter (1967), Amey, Fechter, Grissmer and Sica (1976), Burton (1970), Dale and Gilroy (1984), Goldberg and Greenston (1983), Hosek and Peterson (1984), Kim (1982), Kim, Farrell and Clague (1971), McNown, Ash and Udis (1980), and Withers (1979).

16.) See Dale and Gilroy (1983a,1983b) and Ash, Udis and McNown (1983).

17.) See Ash, Udis and McNown (1983), Dale and Gilroy (1983a,1983b) and Grissmer (1977).

18.) This is supported in a survey of Army enlistees by Elig, Gade and Shields (1982). The results of this survey indicate that over 40 percent of enlistees mentioned that unemployment was a factor in the decision to enlist.

19.) See Fernandez (1980) or Huck, Kusmin and Shepard (1982) for a discussion of these programs.

20.) Ellwood and Wise (1984) used Brown's formulation of education benefits but dropped it due to anamalous results (pp. 14).

21.) Daula and Smith (1984) found, in all but one equation, insignificant VEAP effects. The only statistically significant estimate indicated that the VEAP had a negative effect on high quality recruits. They did find, however, that the Army College Fund (which augmented the basic VEAP as of 1982) had a positive and statistically significant effect across all types of recruits.

22.) See Brown (1984), Tables 2 and 3.

23.) DeVany and Saving (1982) found similar results. However, they did find that recruiters had a positive effect on the quality of the recruit.

Chapter 2

Occupational Choice of High School Graduates: Theory and Model Specification

1. Introduction

The primary purpose of this chapter is to present a microeconomic model of occupational choice and specify the empirical model estimation equations. This model will be based upon the theory of random utility. A general model of occupational choice is presented in the following section. Section 3 will apply this model to a three choice setting, which is the basis of the empirical analysis.

There are several econometric issues that must be addressed prior to estimation of the empirical model. These issues are addressed in Section 4. The actual specification of the variables in the estimation equations will be presented in Section 5. The last section is reserved for a chapter summary.

2. A General Model of Occupational Choice

The decision to choose a particular occupation from a set of possible occupations will be analyzed in a standard utility maximization framework. Two initial assumptions will be made. First, the individual is assumed to face a set of occupational possibilities for any particular point in time. Secondly, it is assumed that each of these occupational possibilities has an associated set of monetary and non-pecuniary attributes. The individual's problem therefore consists of selecting the occupation that maximizes the returns to these attributes over some given length of time. For an individual, the utility function associated with these possible occupations is given as

(2.1) $U_k = U(Y_k, X_k)$, for $k = 1,2,3,\ldots,K$.

X_k is defined as a vector of non-pecuniary attributes associated with occupation k and Y_k is a vector of monetary returns to this occupation. Initially, it is assumed that both Y_k and X_k are observable and nonstochastic. This assumption will be partially relaxed in the following section.

This simple utility function is now expanded to include the effects of individual tastes. Individual tastes are assumed to play a large role in the determination of the occupational choice. Given identical observed attributes between two or more possibilities, the individual will choose the alternative with the highest taste valuation. Introducing tastes explicitly into the model transforms the above utility function into a stochastic function. This transformation is because, unlike pecuniary and non-pecuniary occupational attributes (which are usually observable), tastes cannot be observed. Therefore, explicitly including tastes into the model results in the possibility of an outcome that the observable attributes would not predict. This randomness in outcome is explicitly represented by rewriting equation (2.1) as

$$(2.2) \qquad U_k = V(Y_k, X_k) + \xi_k^*, \text{ for } k = 1,2,3,\ldots,K.$$

The first term in the right hand side of this equation, $V(.)$, is the non-stochastic measurable portion of the utility function. The taste effect on the valuation of occupation k by the individual is represented by ξ_k^*. Tastes are assumed to be distributed randomly over individuals.

This particular formulation of the utility function is an example of what is referred to in the literature as a random utility model. First introduced by Thurstone (1927),[1] this formulation of utility has proven useful for the analysis of a variety of qualitative choice problems.[2] It is used in this analysis because it presents a convenient and realistic framework for the analysis of occupational choice.

The assumed distribution of the stochastic elements in equation (2.2) will become important for the statistical specification of the empirical model. This will be covered in the following section. For the present discussion, no assumptions will be made as to the explicit distribution of these elements. Rather, this exposition will proceed on a more general level.

Now consider the choice of a particular occupation j out of K possible occupations. Occupation j will be selected if and only if $U_j > U_k$ for all $k \neq j$. Or, in the framework of equation (2.2), j will be selected if and only if

$$(2.3) \qquad V(Y_j, X_j) + \xi_j^* > V(Y_k, X_k) + \xi_k^*,$$

for all $k \neq j$.[3] By rearranging terms, this can be equivalently expressed as

$$(2.4) \qquad (V_j - V_k) > (\xi_k^* - \xi_j^*),$$

where $V_m = V(Y_m, X_m)$, for $m = 1,2,\ldots,K$. This equation states that only if the observable difference of V_j and V_k (measurable utility) exceeds the unobservable difference in the elements ξ_k^* and ξ_j^* (tastes) will occupation j be selected.

Given the specification of the utility function in equation (2.2), the probability that the individual chooses a particular occupation from the set of occupations K can now be expressed. Let P_1 represent the probability of observing the selection of occupation 1 from the occupational set K. By rearranging the terms in equation (2.4), this probability can be expressed as

$$(2.5) \qquad P_1 = \text{prob}(V_1 - V_2 + \xi_1^* > \xi_2^* \text{ and } V_1 - V_3 + \xi_1^* > \xi_3^* \ldots$$

$$\text{and } V_1 - V_K + \xi_1^* > \xi_K^*)$$

$$= \int_{-\infty}^{+\infty} \int_{-\infty}^{V_1-V_2+\xi_1^*} \int_{-\infty}^{V_1-V_3+\xi_1^*} \cdots \int_{-\infty}^{V_1-V_K+\xi_1^*} f(\xi_1^*, \xi_2^*, \ldots, \xi_K^*) d\xi_K^* \ldots d\xi_2^* d\xi_1^*,$$

where $f(\xi_1^*, \xi_2^*, \ldots, \xi_K^*)$ is the joint density function of the stochastic elements ξ_k^* (k=1,2,...,K). Similar expressions can be found for the other occupational possibilities. Once the exact form of the density function f(.) is specified, this probability is completely defined.

3. Application of the Model to Military Enlistments

The purpose of this section is to apply the general model presented in the previous section to the decision to enlist in the Armed Forces. In order to make the present analysis more tractable, several additional simplifications will be introduced. First, the occupational choice set will consist of three possible states:

civilian employment, full-time education and military service. This simplification is required due to the econometric complexities and data constraints that would result with a more disaggregate choice set.[4] Secondly, the possibility of occupational switching will not be considered. Rather, this analysis will focus on the occupational status of the individual one year after graduation from high school.[5]

In light of these simplifications, the utility associated with the three occupational possibilities is represented as

$$(2.6) \qquad U_k = V(Y_k, X_k) + \xi_k^*$$

where k = 0 for civilian employment,

= 1 for full-time education,

= 2 for military service.

The variables Y_k, X_k and ξ_k^* are the same as defined in the previous section. Assume that the observable elements of the utility function are linear in a vector of unknown parameters. Introducing the subscript i to denote the individual, equation (2.6) is rewriten as

$$(2.7) \qquad U_{ik} = Y_{ik}\alpha_k + X_{ik}\beta_k + \xi_{ik}^*, \text{ for } k = 0,1,2,$$

where α_k and the vector β_k are unknown parameters.

In the previous section it was assumed that both Y_{ik} and X_{ik} (k=0,1,2) were nonstochastic. This assumption is now relaxed with respect to the monetary returns to civilian employment (Y_{i0}) only. It is assumed that these returns are determined by a set of observable nonstochastic personal attributes and local labor market conditions that are linear in a vector of parameters. In addition, it is further assumed that a certain amount of unaccountable randomness exists in the determination of the actual observed value of these returns. Therefore, these returns are represented as

$$(2.8) \qquad Y_{i0} = Z_{i0}\gamma_0 + e_{i0},$$

where Z_{i0} is a vector of observable nonstochastic attributes, γ_0 is a vector of unknown parameters and e_{i0} is a random error term, assumed distributed independently $N(0,\sigma_0^2)$.

Note that the monetary returns to the other possible occupations are still assumed to be nonstochastic. In an ideal situation, it could be argued that a certain amount of randomness also exists in the determination of the monetary returns to college (Y_{i1}). However, in the present analysis these returns will never be observed. Instead, a nonstochastic proxy variable will be used to represent these returns. As this proxy will be both exogenous to the individual and known with certainty, there is no reason to relax this assumption. A discussion of this proxy variable is deferred to Section 4.2. For the present discussion, C_{i1} will be used to denote this proxy.

The initial monetary returns to military service (Y_{i2}) are also both exogenous to the individual (at the date of decision) and known with certainty. Initial military monetary compensation is determined primarily via the Federal budgetary process, with the only real source of variation due to differences in the individual's number of dependents. A discussion of the compensation components is given in Section 5.2.

Given these considerations, the basic empirical model can be obtained by substituting equation (2.8) into equation (2.7). Or,

$$(2.9) \qquad U_{i0} = \bar{Y}_{i0}\alpha_0 + X_{i0}\beta_0 + \xi_{i0},$$

$$U_{i1} = C_{i1}\alpha_1 + X_{i1}\beta_1 + \xi_{i1}, \text{ and}$$

$$U_{i2} = Y_{i2}\alpha_2 + X_{i2}\beta_2 + \xi_{i2},$$

where $\bar{Y}_{i0} = Z_{i0}\gamma_0$, C_{i1} is the monetary returns to college proxy variable and ξ_{ik}, $k=0,1,2$ is the reduced form error term. All of the other variables are as previously defined. To simplify the following discussion define:

$$W_{i0} = [\bar{Y}_{i0}, X_{i0}],$$

$$W_{i1} = [C_{i1}, X_{i1}],$$

$$W_{i2} = [Y_{i2}, X_{i2}], \text{ and}$$

$$\delta_k' = [\alpha_k', \beta_k'].$$

The final step necessary to complete the statistical specification of the basic general empirical model is to specify the distribution of ξ_{ik}. For probabilistic models of this type, there are two commonly assumed distributions of the error terms in equation (2.9) : the normal and the type I extreme value distributions. The former distribution leads to the specification of a probit model and the latter to a logit model. For the present analysis, it will be assumed that these error terms are distributed i.i.d. type I extreme value.

Recall from equation (2.5) the general model specification of the probability of selecting a particular occupation. The probability of an enlistment can be expressed in a similar fashion. Or,

(2.10) $$\text{Prob(Enlistment)} = \text{prob}(U_{i2} > U_{i0} \text{ and } U_{i2} > U_{i1})$$

$$= \text{prob}(W_{i2}\delta_2 - W_{i0}\delta_0 > [\xi_{i0} - \xi_{i2}] \text{ and}$$

$$W_{i2}\delta_2 - W_{i1}\delta_1 > [\xi_{i1} - \xi_{i2}]).$$

This expression can be simplified by introducing the following notation. Define: $\eta_{02} = (\xi_{i0} - \xi_{i2})$ and $\eta_{12} = (\xi_{i1} - \xi_{i2})$. Equation (2.10) can now be restated as

$$(2.11) \qquad \text{Prob(Enlistment)} = P_{i2} = \text{prob}(W_{i2}\delta_2 - W_{i0}\delta_0 > \eta_{02}$$

$$\text{and } W_{i2}\delta_2 - W_{i1}\delta_1 > \eta_{12}).$$

As noted above, the logit model is based on an assumed type I extreme value distribution of the error terms ξ_{ik}, k=0,1,2. This implies that

$$(2.12) \qquad Pr\ (\xi_{ik} < \xi) = \exp(-\exp(-[\xi - \alpha]/\beta)),$$

where α and β are parameters bounded by $-\infty < \alpha < +\infty$ and $\beta > 0$.[6] It can be shown that the convolution of two random variables that are independently and identically distributed type I extreme value is distributed logistically.[7] Therefore, the error terms η_{02} and η_{12} in equation (2.11) are distributed logistically.

As a final consideration, note that the utility associated with each of the various possible occupations (the dependent variable in equation (2.9)) is never observed. What is observed is the actual outcome of the decision process. A common convention is to use a discrete variable, that is assigned a value contingent on the observed outcome, as a proxy for the underlying utility valuation of these possible outcomes. Let I_i be such a proxy variable. In particular, define

I_i = 0 if individual i chooses civilian employment,

= 1 for college enrollment, and

= 2 for military service.

The probability of observing an enlistment, as given in equation (2.11), can now be expressed within the logit model specification as

$$(2.13) \quad P_{i2} = pr(I_i = 2) = \frac{\exp(W_{i2}\delta_2)}{\sum_k \exp(W_{ik}\delta_k)},$$

where the right hand side term is a cumulative logistic distribution. The multinomial logit model is a product of a series of independently and identically distributed logistic functions. The covariance matrix of this model is a (K x K) diagonal matrix, where the diagonal elements are equal to $\pi^2/3$, which is the variance of the logistic distribution.

This a priori restriction on the structure of the covariance matrix is a potential disadvantage of the logit model specification. In the literature, this restriction leads to a property that is usually referred to as the Independence of Irrelevent Alternatives (IIA). If the model specification imposes IIA, then the relative odds of any two possible outcomes are unaffected by the presence or introduction of additional alternatives. A simple demonstration of this property can be seen within the context of the logit specification of equation (2.13). Initially assume that the choice set consists of two possible outcomes (k=0,1). The probabilities of these outcomes are given as

$$(2.14) \qquad P_{ik} = \frac{\exp(W_{ik}\delta_k)}{\exp(W_{i0}\delta_0) + \exp(W_{i1}\delta_1)}, \text{ for } k = 0,1.$$

The relative odds of these probabilities are

$$(2.15) \qquad \frac{P_{i0}}{P_{i1}} = \frac{\left[\dfrac{\exp(W_{i0}\delta_0)}{\exp(W_{i0}\delta_0) + \exp(W_{i1}\delta_1)}\right]}{\left[\dfrac{\exp(W_{i1}\delta_1)}{\exp(W_{i0}\delta_0) + \exp(W_{i1}\delta_1)}\right]} = \frac{\exp(W_{i0}\delta_0)}{\exp(W_{i1}\delta_1)}.$$

Now consider the effect of the introduction of a third alternative. The denominator of the right hand side of equation (2.14) will be expanded to include the effect of this additional alternative. Or, equation (2.14) becomes

$$(2.16) \qquad P_{ik} = \frac{\exp(W_{ik}\delta_k)}{\exp(W_{i0}\delta_0) + \exp(W_{i1}\delta_1) + \exp(W_{i2}\delta_2)}, \text{ for } k= 0,1,2.$$

However, note that the odds in equation (2.15) are unaffected by the introduction of this additional alternative. If the alternatives are sufficiently different in their characteristics, then this property does not impose unrealistic restrictions on the model. In fact, the existence of IIA reduces the computational requirements for the evaluation of alternative impacts.[8]

Counterintuitive outcomes could result if the newly introduced alternative is characteristically closer to one of the alternatives than the other(s). Domencich and McFadden (1975) illustrated this problem in the context of transportation mode choice.[9] They considered the situation of using an auto or taking a bus for a particular trip. Intuitively, one would expect that the introduction of an additional bus option would affect the probability of selecting the original bus (i.e. - the closer substitute) more than that of the auto. However, this intuitive outcome is not what results under the assumption of IIA. Instead, the IIA property results in both probabilities being altered by the same percentage. Therefore, dependent on how characteristically similar the possible outcomes are, the imposition of IIA could lead to some rather unreasonable predicted outcomes.

4. Additional Econometric Considerations

If there were complete information on the occupational selection explanatory variables, then the estimation of the empirical model would be relatively straightforward. However, for the current analysis, an ideal model cannot be estimated because of three complications. First, the monetary returns to education are not observed for any of the individuals in the current sample. Secondly, monetary returns to civilian employment are only observed for individuals who opted for civilian employment. This partial observance is an example of a self-selectivity bias. Thirdly, the sample is choice-based: individuals in the military were overly represented in the sample, relative to the true population percentage.

The purpose of this section is to discuss how the estimation method must be modified to adjust for these additional complications. The following subsection describes how a simple polychotomous logit model is estimated. The subsequent subsections describe how this basic estimation method will be modified to correct for the effects of these further complications.

4.1 Maximum Likelihood Estimation of a Simple Logit Model

By generalizing the specification of the polychotomous logit model in equation (2.13), it can be seen that this model is nonlinear in the parameters. An appropriate technique to estimate models of this type is the maximum likelihood estimation (MLE) procedure. For a trichotomous logit model with N observations, the likelihood function to be maximized is given as

$$(2.17) \qquad L = \prod_{i=1}^{N} P_{i0}^{D_{i0}} \cdot P_{i1}^{D_{i1}} \cdot P_{i2}^{D_{i2}} ,$$

where D_{ik} $(k=0,1,2)$ is a dummy variable equal to 1 if individual i chooses occuaption k and equal to 0 otherwise.[10] P_{ik} $(k=0,1,2)$ is the generalized form of the logit specification in equation (2.13). Or,

$$(2.18) \qquad P_{ik} = \exp(W_{ik}\delta_k)/\sum_k \exp(W_{ik}\delta_k).$$

In order to obtain the maximum likelihood estimates of the model parameters, a set of equations involving the partial derivatives of the likelihood function in equation (2.17), with respect to these parameters, must be solved. As these equations are nonlinear in δ_k, they must be solved by using an iterative procedure. The procedure that will be used for the current analysis is the Newton-Raphson method.

4.2 The Non-Observance of Returns to Education

As stated above, the monetary returns to education (Y_{i1}) is not observed for any of the individuals in the current data sample. There are several methods of creating a proxy variable for these unobserved returns.[11]

One possible method to create an educational returns proxy variable consists of estimating the education monetary returns equation with data for individuals who have already completed college, discounting the estimated $\hat{Y}_{i1,t+n}$ to the relevant time period, and using the resulting value as a proxy for these expected returns. The problem with this approach is that these estimates will most likely be biased and imprecise. One potential bias of this approach is due to these estimates being based on a sample of individuals that have completed college, whereas those who are still involved in the education process face a non-trivial probability of not completing. Hence, the estimated income would be upwardly biased. In addition, determination of the proper rate of discount and differences in the vector of explanatory variables would further reduce the accuracy of this method.

A second method of approximating Y_{i1} consists of estimating what the individual would have earned if employed in the civilian sector. Assuming that the returns to human capital investment are an increasing function (over the relevant range) of unobservable "ability", this estimate would create a lower bound on Y_{i1}. There are two difficulties with this approach. First, there is no reason to expect "ability" to have a homogeneous impact across options. Rather, it is more reasonable to expect the opposite. This implies that the observed earnings of those who elected not to enroll in college could very well exceed the estimated earnings of those who did, at the date of enrollment. Secondly, the proper context in which to view the monetary returns to education is over the individual's lifetime. Therefore, even if "ability" were homogeneous, a cross-sectional analysis at the time of enrollment, would have difficulty capturing these returns.

This discussion illustrates the difficulty in estimating these returns to education. However, it should be clear that some form of a proxy for these returns must be included in the model. If no attempt is made to approximate these returns, then these returns are in effect assumed to have no influence on the decision to enter college, an assumption which is clearly incorrect. Rather than attempting to generate a value for the monetary returns to education, it has been decided that a cost of college approximation will be utilized instead. As stated above, the monetary returns to college are spread out over the individual's work career. The costs of college, however, are incurred only in the actual education process. In addition to foregone civilian income, these costs predominately consist of tuition charges. It is argued that, in conjunction with estimated (foregone) civilian income, tuition costs reflect, albeit imprecisely, the lower bound on the monetary returns to education. Therefore, these monetary returns will be proxied by both civilian income and in-state tuition rates.

4.3 Self-Selectivity Bias

Self-selectivity bias is a common econometric problem in models with mixed discrete/continuous dependent variables. This type of bias is similar to a standard omitted variable bias. The only difference is that the direction of the self-selectivity bias can be more easily inferred. An intuitive explanation of this bias can be seen by the following simple example. Consider two individuals (A and B). Individual A has a comparative advantage in hunting and individual B has a similar advantage in farming. However, only these individuals are totally aware of these advantages. In all other respects these individuals are identical. Left to their own, these individuals select the occupation where their comparative advantage lies. Now, let a third individual (who is unaware of these unobserved comparative advantages) attempt to determine what these individuals would have earned in the occupation not chosen (i.e. - farming for A and hunting for B). Based on the observed characteristics of the individuals (which are identical), this third person would reason that A would earn the same as B in farming and B would earn the same as A for hunting. However, by not taking into account the comparative advantages, this reasoning leads to an over prediction of what the individuals would have earned in the occupations not selected.

In a more rigorous context, possible bias arises when one or more of the dependent variables is observed, conditional on the satisfaction of a selection criteria. Unless the error terms of the selection equation and the other equations are uncorrelated, estimation of these other equations, based on the sample of selected observations, will yield inconsistent estimates of the true model parameters. When the dependent variable in the selection equation is binary, there are several computationally simple two stage bias correction procedures available.[12] However, with a polychotomous dependent variable, bias correction becomes more difficult. The bias correction procedure that will be used in the present analysis is based upon the work of Lee (1983).[13,14]

The nature of the self-selectivity bias can be observed by considering the 4 equation model given by equations (2.8) and (2.9) in Section 3. Or,

(2.19) $U_{ik} = W_{ik}\delta_k + \xi_{ik}$, for k=0,1,2 and

$$Y_{i0} = Z_{i0}\gamma_0 + e_{i0}.$$

In this system, the dependent variable in the civilian earnings equation is observed only if the individual chooses civilian employment. Recall from equation (2.10) the condition for the choice of military service. A similar condition for civilian employment can be found. Or, the individual will choose civilian employment iff

(2.20) $G_{02}\pi_{02} > \eta_{20}$ and $G_{01}\pi_{01} > \eta_{10}$, where

$$G_{02} = [\ W_{i0} \quad W_{i2}\ ],$$

$$\pi_{02}' = [\ \delta_0' \quad -\delta_2'\ ],$$

$$G_{01} = [\ W_{i0} \quad W_{i1}\ ],$$

$$\pi_{01}' = [\ \delta_0' \quad -\delta_1'\ ],$$

$$\eta_{20} = (\xi_{i2} - \xi_{i0}) \text{ and } \eta_{10} = (\xi_{i1} - \xi_{i0}).$$

Now, keeping this observance criteria in mind, the expected value of civilian income $E(Y_{i0})$ can be expressed as

(2.21) $E(Y_{i0} \mid Z_{i0} \text{ and } I_i = 0 \text{ are observed}) =$

$$Z_{i0}\gamma_0 + E(e_{i0} \mid G_{01}\pi_{01} > \eta_{10} \text{ and } G_{02}\pi_{02} > \eta_{20}).$$

Unless e_{i0} is uncorrelated with η_{10} and η_{20}, the expected value of the disturbance term in equation (2.21) will not equal 0 and $E(Y_{i0}) \neq Z_{i0}\gamma_0$. This implies that simple ordinary least squares estimation of this equation will result in biased estimates of Y_{i0}. To obtain unbiased estimates of the parameters in this equation, the effects of the correlation between e_{i0} and η_{10}, η_{20} must be controlled for.

The starting point for obtaining a proper bias correction procedure is to find a more explicit expression for this bias. Once such an expression is obtained, an appropriate bias correction instrument can be created.

The strategy that is used to find an exact expression for the conditional expection of e_{i0} in (2.21) is the one employed by Lee (1983). Briefly, this strategy consists of re-expressing the trivariate distribution of e_{i0}, η_{10} and η_{20} in terms of a distribution with well known properties (the bivariate normal).

To allow the expression of the self-selectivity bias in terms of a bivariate normal distribution, it is convenient to look at the choice of occupation within an order statistic framework. Within an order statistic framework, the individual is viewed as ranking the returns to the various alternatives from lowest to highest. This implies that a particular occupation will be selected only if the returns to that occupation exceed the maximum returns over the set of alternative occupational choices. There is little conceptual difference between this order statistic approach and the previously used multiple binary comparision approach. Both approaches imply the choice of the alternative with the highest associated returns.

Within an order statistic decision process, the probability that the individual will choose civilian employment (P_{i0}) can be expressed as

$$(2.22) \qquad P_{i0} = \text{prob}(I_i = 0) = \text{prob}(U_{i0} > \max_k U_{ik},\ k = 1,2),$$

where I_i is the discrete choice indicator variable as defined in equation (2.16) above. Recall that W_{ik}, k=0,1,2 represents the observable nonstochastic components of U_{ik} and δ_k represents the vector of unknown parameters for the components in W_{ik}. Therefore equation (2.22) can be rewritten as

$$(2.23) \qquad P_{i0} = \text{prob}(W_{i0}\delta_0 > \max_k(W_{ik}\delta_k + \xi_{ik}) - \xi_{i0}),\ k=1,2$$

$$= \text{prob}(W_{i0}\delta_0 > \psi_{i0}),$$

where $\psi_{i0} = \max_k(W_{ik}\delta_k + \xi_{ik}) - \xi_{i0}$, $k = 1,2$, is assumed i.i.d. logistic. A bivariate function in e_{i0} and Ψ_{i0} can now be specified. Let $B_0(e_{i0},\psi_{i0};\rho_0)$, where ρ_0 is the correlation coefficient of e_{i0} and ψ_{i0}, represent this bivariate function. Note that as ψ_{i0} is assumed distributed logistically, this bivariate function is not bivariate normal. A bivariate normal distribution function can be obtained by transforming ψ_{i0} to a standard normal random variable. Or, following Lee (1983), define

$$(2.24) \qquad \psi^*_{i0} = J_0(\psi_{i0}) = \Phi^{-1}(F_0(\psi_{i0})), \text{ and}$$

$$J_0(W_{i0}\delta_0) = \Phi^{-1}(F_0(W_{i0}\delta_0)),^{15}$$

where $F_0(.)$ denotes the cumulative logistic distribution functions and Φ^{-1} is the inverse of a standard normal distribution function. As long as the distributions of ψ_{i0} and $W_{i0}\delta_0$ are well defined, this transformation is possible. Further, as both the standard normal and the logistic distributions are symmetric around 0, this transformation will be quite accurate for all but extreme value probabilities.

Based on this transformation of ψ_0, define $B_0^*(e_{i0}, \psi_{i0}^*; \rho_0^*)$ to be the the bivariate normal distribution of e_{i0} and ψ_{i0}^*, with ρ_0^* the correlation coefficient of (e_{i0}, ψ_{i0}^*) and σ_0^2, σ_*^2 the variances of e_{i0} and ψ_{i0}^*. The conditional expectation of e_{i0} in (2.21) can now be explicitly expressed. Or, by substitution from equations (2.24) and (2.21),

$$
\begin{aligned}
(2.25)\quad E(e_{i0} \mid J_0(W_{i0}\delta_0) > \psi_{i0}^*) &= -\rho_0^*\sigma_0/\sigma_* E(\psi_{i0}^* \mid J_0(W_{i0}\delta_0) > \psi_{i0}^*) \\
&= -\rho_0^*\sigma_0\phi(J_0(W_{i0}\delta_0))/\Phi(J_0(W_{i0}\delta_0)) \\
&= -\rho_0^*\sigma_0\phi(J_0(W_{i0}\delta_0))/F_0(W_{i0}\delta_0) \\
&= -\lambda_0\phi(J_0(W_{i0}\delta_0))/F_0(W_{i0}\delta_0),
\end{aligned}
$$

as $\sigma_* = 1$ and $\Phi(J_0(W_{i0}\delta_0)) = \Phi(\Phi^{-1}(F_0(W_{i0}\delta_0))) = F_0(W_{i0}\delta_0)$. $\phi(.)$ represents the standard normal density function and $\lambda_0 = \rho_0\sigma_0$. It should be noted that this expression follows only if the conditional expectation of e_{i0} given

ψ^*_{i0} is assumed to be linear.[16] In the most restrictive case, where e_{i0} is also assumed distributed standard normal, λ_0 reduces to σ_{0*} (i.e. - the covariance of e_{i0} and ψ^*_{i0}).

The ratio $\phi(J_0(.))/F_0(.)$ is more commonly referred to as the "Mills ratio". The derivation of this ratio from the conditional expectation of a bivariate normal distribution function can be found in Johnson and Kotz (1972).[17] Intuitively, this ratio can be described as representing the effect of the probability of choosing civilian employment on the civilian income equation error term (e_{i0}). Those individuals with the lowest probability of choosing this occupation would have the largest (uncorrected) overestimates of Y_{i0}. Conversely, those individuals with the highest probability would have the smallest bias. This is basically what this ratio indicates. As the probability of choosing this occupation goes to 1, this ratio goes to 0 and as the probability approaches 0, the ratio approaches 1.[18]

Now, given the expression for the self-selectivity bias in equation (2.25), the correctly specified civilian wage equation can be written. Or, substituting from equation (2.25) into equation (2.20),

$$(2.26) \qquad Y_{i0} = Z_{i0}Y_0 - \lambda_0[\phi(J_0(W_{i0}\delta_0))/F_0(W_{i0}\delta_0)] + \Delta_0$$

$$= Z_{i0}Y_0 - \lambda_0 M_0 + \Delta_0,$$

where Δ_0 is an error term that is uncorrelated with ψ^*_{i0}, with $E(\Delta_0) = 0$. If λ_0 and M_0 were known, the value for the self-selectivity bias could be easily calculated and the estimation of this income equation would be straightforward. However, as M_0 is not known, an appropriate estimate (say M_0) must be found.

There are two possible methods for obtaining estimates of M_0. The first consists of maximizing the limited information likelihood function (LIML) based on the bivariate normal density function $b_0(e^*_{i0}, \psi^*_{i0}; \rho^*_0)$. The second method is to use a two stage logit/ordinary least squares estimation procedure.[19] This second method is the one that is utilized in the present analysis.

The first stage of this two stage procedure consists of estimating a reduced form logit model. This will be done by maximizing the likelihood function given in equation (2.17). Or,

$$(2.27) \qquad L = \prod_{i=1}^{N} P_{i0}^{D_{i0}} \cdot P_{i1}^{D_{i1}} \cdot P_{i2}^{D_{i2}},$$

where $P_{ik} = \exp(W_{ik}\delta_k)/\sum_k \exp(W_{ik}\delta_k)$. Using the reduced form parameter estimates obtained from this first stage, an estimate of the Mills ratio $(\hat{M}_0)$ can be calculated. This estimate is given as

$$(2.28) \qquad \hat{M}_0 = \phi(J_0(W_{ik}\hat{\delta}_k))/F_0(W_{ik}\hat{\delta}_k).$$

The second stage consists of estimating the civilian wage equation given in (2.26), after substituting $\hat{M}_0$ for M_0. Or,

$$(2.29) \qquad Y_{i0} = Z_{i0}\gamma_0 - \lambda_0\hat{M}_0 + \Delta_0 \ .$$

Once estimates of Y_{i0} ($\hat{Y}_{i0}$) are obtained, the structural probability model can be estimated. Substituting $\hat{Y}_{i0}$ into equation (2.18), the structural probability model is now given as

$$(2.30) \qquad P_{ik} = \exp(\hat{W}_{ik}\delta_k)/\sum_k \exp(\hat{W}_{ik}\delta_k) \ ,$$

where $\hat{W}_{ik} = [\hat{Y}_{ik} \ X_{ik}]$, for $k = 0$,

$= [C_{ik} \ X_{ik}]$, for $k = 1$, and

$= [Y_{ik} \ X_{ik}]$, for $k = 2$.

To obtain estimates of the structural probability model parameters, the likelihood function given in equation (2.27) is reestimated, after substituting $\hat{W}_{ik}$ for W_{ik} (k=0,1,2).

It should be noted that the procedure outlined in this section differs from the procedure found in Lee (1983) in several respects. First, Lee's procedure assumes that the reduced form parameters are equal across outcomes (i.e. $\delta_0 = \delta_1 = \delta_2 = \delta$). Here, this assumption has been implicitly relaxed. It is assumed that this relaxation will not affect the consistency of the estimators obtained from the estimation of the second stage earnings equation.

Secondly, and more importantly, the procedure outlined by Lee is limited to the estimation of the earnings equation. The properties of the estimators for the parameters of the structural probability model are not discussed in Lee (1983). A formal proof of the consistency of these structural model estimators would be desirable. At the present time, however, such a proof does not exist in the literature, and will not be attempted here. Therefore, the consistency of these estimators and the regular formula for the asymptotic VC matrix is assumed.

4.4 Choice Based Sampling Bias

A data set can be defined as choice based if the probability of an individual being included in the sample is contingent on a decision made by the individual. For example, if a researcher was interested in the probability of an individual choosing a particular mode of transportation, a random sample of individuals could be collected. However, to guarantee a sufficient number of observations for a particular mode choice, this sample may have to be rather large. A less costly approach would be to sample individuals who have opted for the particular transport mode. This less costly sampling procedure is frequently used in qualitative response data collection.

An estimation bias is introduced by a choice based sample as the sample does not represent the true population distribution of individuals in the various possible outcome groups. Therefore, estimation of a qualitative choice model, using this non-random sample, will yield inconsistent estimates of the true population parameters.[20]

The data used for the present analysis is choice based as the 1979 interview year NLS sample was merged with a military sample.[21] The individuals in the military sample were drawn from the military with a different probability than the individuals in the civilian sample. Unless the data set is adjusted to reflect the true population percentage of military enlistees, the model parameter estimates would predict a higher probability of military enlistment than exists in the population.

A choice based bias correction procedure has been suggested by Manski and Lerman (1977). Basically, this procedure consists of weighting each observation's contribution to the likelihood function. The parameter estimates from this weighted likelihood function will be consistent estimators of the true population parameters.

Now, following Manski and Lerman, define

$$(2.31) \qquad w_{ij} = Q(j)/H(j)$$

to be the choice based sampling weight for the ith individual from subsample j. Q(j) and H(j) are respectively defined as the fraction of the population and the fraction of the total sample in subsample j. By inspection, it can be seen that this weight correctly compensates for the over(under) representation of individuals in subsample j, relative to the population. For highly choice based samples (where H(j) is large relative to Q(j)), this weight will be small. In the situation where a choice based sample does not exist (i.e. - H(j) = Q(j)), this weight will equal 1.

The modification to the likelihood function necessary to correct for this choice-based sampling bias is rather straightforward. All that is required is to multiply the observed outcome dummy variable (D_{ik}) by the appropriate weight (w_{ij}). For example, consider the likelihood function in equation (2.27). The choice-based weighted version of this likelihood function is written as

$$(2.32) \qquad L_{weighted} = \prod_{i=1}^{N} P_{i0}^{w_{ij}D_{i0}} \cdot P_{i1}^{w_{ij}D_{i1}} \cdot P_{i2}^{w_{ij}D_{i2}}$$

Manski and Lerman refer to this specification of the likelihood function as the Weighted Exogenous Sampling Likelihood (WESML) estimator. This likelihood function can then be maximized in the standard fashion.[22,23]

It should be noted that using the WESML estimator to obtain the first stage estimates of the two stage self-selectivity bias correction procedure will most likely alter the second stage civilian income estimates. As the military is overrepresented in the sample, the effect of the WESML estimator is to reduce the estimated military enlistment probability (P_{i2}). Therefore, as $P_{i0} + P_{i1} + P_{i2} = 1$, $\Delta(P_{i0} + P_{i1}) > 0$. The effect on the probability of civilian employment (P_{i0}), however, is ambiguous. If $\Delta P_{i0} > 0$, the predicted value of the Mills ratio ($\hat{M}_0$) in equation (2.28) will be reduced and the resulting income estimates ($\hat{Y}_{i0}$) will be increased. This possible result should not be considered to be a further complication of the analysis. In fact, such a possible effect is thoroughly consistent with what the theory of military enlistments would predict.

4.4 Estimation Strategy Summary

Before moving on to the actual equation specifications, the various estimation strategies that will be employed should be summarized. This summary is to make it perfectly clear how the models will be estimated. Three strategies will be used.

(i) The Simple Two-Stage: This strategy consists of a.) estimating an uncorrected civilian wage equation, inserting the estimated wage into the structural occupational selection equations and b.) then estimating the structural occupational selection equations.

(ii) The Self-Selectivity Corrected Two-Stage: This two-stage procedure consists of a.) estimating the reduced form selection equations, as given in equation (2.27). The resulting probability estimates are then used to compute the Mills ratio (M_0) in equation (2.28).

b.) Based on the estimates of the Mills ratio, the civilian wage equation is then estimated. The resulting wage estimates are then inserted into the structural occupation selection equations. These selection equations are then estimated.

(iii) The Self-Selectivity/Choice Based Corrected Two-Stage: This procedure is identical to (ii) above except that the reduced form and structural selection equations are estimated with a choice-based weighted likelihood function.

5. <u>Model Equations Variable Specification</u>

The purpose of this section is to specify the variables that will be used in the estimation of the empirical model. The data set that will be employed consists of a subsample of the National Longitudinal Survey of Youths (1979-1981), supplemented by additional data on economic and recruitment related activities. The variables are briefly described in Table 2-1. A more thorough description of the data set and the variables is presented in the following chapter. A summary of the hypothesized variable impacts can be found in Table 2-2. The Discussion in the following subsections will be devoted to various aspects of the equation specifications and variable hypotheses that may not be readily apparent upon examination of Tables 2-1 and 2-2.

5.1 Occupational Selection Equations.

The hypothesized variable impacts for the selection equations in Table 2-2 are all logical. There is one point, though, that must be clarified.

In Table 2-2 there are no variable hypotheses for civilian employment. This is because, to insure that the model parameters are identified, a normalization constraint must be imposed. There are two types of normalization constraints. The first constrains a particular parameter (say δ_{jk}) to sum to 0 across the possible outcomes.[24] This constraint implies that, for the present trichotomous model, $\delta_{j0} = -(\delta_{j1} + \delta_{j2})$. The second type of normalization constraint sets all of the δ_{jk}'s equal to 0 for a particular k.[25] This constraint is based on the notion that one of the possible outcomes is the normal state of behavior. The other possible outcomes are considered to be deviations from this normal state of behavior. As shown by Avery (1980), either constraint will yield the identical statistical outcome.[26] The second method will be used in the present analysis as it is considered to yield more easily interpretive results.

Given this choice of normalization constraint, the coefficients of the civilian employment specific attributes are constrained to 0 and will not be included in the model estimation. Civilian wage effects, though, will be retained in the model via the other selection equations.

5.2 The Wage Equations

The specification of the civilian wage equation is predominately based on the theory of Human Capital.[27] In general, human capital investment is usually represented by labor market experience, job specific training and educational attainment.[28] For the present analysis, the education level is relatively constant across the sample (i.e. - all of the individuals are high school graduates). Therefore, human capital is approximated by labor market experience and non-high school training programs.

Military wages do not exhibit as much variation as those in the civilian sector, at entry level positions. For most intents and purposes, the entry level military wage for non-prior service (NPS) enlistees is exogenous to the individual's personal characteristics. The wage for this group is relatively constant cross-sectionally and determined by the Federal budgetary process. However, it is possible to introduce some variation into this wage. The sample covers three years. This allows the introduction of some variation due to periodic cost of living increases (COLA) and real wage increases. In

Table 2-1: Definition of Variables

Variable	Definition
ADDT	: Desires additional training outside of college (= 1 if yes, = 0 otherwise).
BLACK	: Black dummy variable (= 1 if black, = 0 otherwise).
COLPROG	: Participation in college preparatory program (= 1 if participated, = 0 otherwise).
COSTO	: In-state tuition rates for the year of graduation ($'s).
CWAGE	: Civilian wage rate (annual, in $'s).
DADMIL	: Father in military during year of graduation (= 1 if yes, = 0 otherwise).
EDAD	: Highest year of education completed by father.
EMOM	: Highest year of education completed by mother.
ESIB	: Highest year of education completed by oldest sibling.
HEXP	: Highest year of education respondent expects to complete.
HISP	: Hispanic dummy variable (= 1 if Hispanic, = 0 otherwise).
LADV	: Local military advertising expenditures ($'s/18-22 state pop.).
MSTAT	: Marital status (= 1 if married, = 0 otherwise).
MWAGE	: Military wage rate (annual, in $'s).
NADV	: National military advertising expenditures ($'s/18-22 state pop.).
PSIBS	: Percentage of siblings in school.

Table 2-1 (cont.)

Variable	Definition
REC	: Number of recruiters per 18-22 state pop.
TDEP	: Number of dependents.
TFMAFS	: Number of family members in the military.
TGPROG	: Number of government training courses completed.
TNFINC	: Total net family income for the year of graduation ($'s).
TOTHRS	: Number of weeks of labor market experience.
TSCORE	: Armed Forces Qualification Test score.
TVTPROG	: Number of vocational/training programs completed.
UNRATE	: Local unemployment rate.
VEAP	: Veterans Educational Assistance Program maximum benefits ($'s).
VOCPROG	: In vocational program while in high school (= 1 if yes, = 0 otherwise).

Table 2-2: Summary of Coefficient Hypotheses

Variable	Equation: College Enrollment	Military Enlistment	Civilian Wage (CWAGE)	Military Wage (MWAGE)
ADDT	-	+		
BLACK	-	+	-	
COLPROG	+	-		
COSTO	-	+/?		
CWAGE	-	-		
DADMIL	-/?	+		
EDAD	+	-		
EMOM	+	-		
ESIB	+	-		
HEXP	+	+/?		
HISP	-	+	-	
LADV	-/?	+		
MSTAT	-	?	+	+
MWAGE	-/?	+		
NADV	-/?	+		
PSIBS	-	?		
REC	-/?	+		
TDEP				+

Table 2-2 (cont.)

Variable	Equation			
	College Enrollment	Military Enlistment	Civilian Wage (CWAGE)	Military Wage (MWAGE)
TFMAFS	-/?	+		
TGPROG			+	
TNFINC	+	-		
TOTHRS			+	
TSCORE	+	+/?	+	
TVTPROG			+	
UNRATE	?	+	-	
VEAP	-/?	+		
VOCPROG			+	

addition, total military compensation comprises Basic Pay (BP) and additional allowances and incentive pays.[29] The primary allowances consist of Basic Allowances for Quarters (BAQ) and Basic Allowance for Subsistance (BAS). The amount of these allowances varies with the number of dependents. There is also a Federal Tax Advantage (TA) for members of the military. This tax advantage also varies with the number of dependents. Inclusion of these allowances and the tax advantage will introduce some additional variation to total compensation. Special incentive pays will not, however, be included as these pays are conditional upon occupational speciality and location of duty assignment.

Given this exogenous structure of military compensation, the equation for total military compensation (Y_{i2}) can be expressed as

$$\text{(2.33)} \quad MWAGE = BP + BAQ(TDEP) + BAS(TDEP) + TA(TDEP),$$

where TDEP is the total number of dependents.

6. Chapter Summary

In brief, the purpose of this chapter was to:

A.) Present a theoretical framework for the analysis of the decision to enlist in the Armed Forces. The model presented was based upon the theory of random utility.

B.) Discuss the econometric issues relevant to the estimation of the empirical model. The results of this discussion were the choice of a logit model specification, the use of the "Lee Approach" for the correction of selectivity bias, and the application of sample weights (Manski and Lerman (1977)) for the correction of choice based sampling bias.

C.) Specify the empirical model equations. The variables specified consisted of various economic and non-economic attitudinal attributes that were hypothesized to influence the decision of initial occupational choice. A more detailed description of these variables (and the data base in general) is presented in the following chapter.

End Notes

1.) In a technical sense, Thurstone did not develop a theory of "random utility". Rather, he developed a "Law of Comparative Judgement" for the analysis of responses to various stimuli in a psychological experimentation setting. An excellent discussion of Thurstone's model and its' applications in presented in Bock and Jones (1968).

2.) For examples of the application of this model see Domencich and McFadden (1975), Daula (1981) or Hausman and Wise (1978). Additional example references can be found in the survey article of Amemiya (1981).

3.) This inequality relationship is assumed to be strict as the probability of $U_k = U_j$, for all $k \neq j$ is zero, by definition.

4.) For example, a more disaggregate choice set would break down civilian employment by job type, military service by branch and education by major field of study and/or type of school (i.e. -private vs public).

5.) The decision to exclude non-high school graduates from the sample is based on the demand constrant complexities that would otherwise result. This problem is discussed in Ash, Udis and McKnown (1983), pps. 147 and 154 and in Brown (1984), pp. 4.

6.) In practice, the parameter β is usually constrained to equal 1. If this constraint is not imposed, the variance of the logit specification ($\beta^2\pi^2/3$) is no longer known and therefore must be estimated.

7.) See Domencich and McFadden (1975), pps. 63-65.

8.) See Domencich and McFadden (1975), pps. 70-71.

9.) Pps. 78-79.

10.) A more thorough discussion of this likelihood function can be found in Maddala (1983), pps. 35-37.

11.) Santos (1981) and Kalton (1982) present a discussion of various methods of imputing missing values. These methods, however, are based on the assumption that the missing values are the result of a random process. This is not the situation for the present analysis.

12.) For examples of the application of a binary choice model self-selectivity correction procedure see Heckman (1976,1979), Lee (1978), or Maddala (1983).

13.) A presentation of this approach can also be found in Maddala (1983), pps. 275-276.

14.) Discussions of slightly different correction procedures can be found in Dubin and McFadden (1984) or Hay (1984).

15.) A computationally simple and accurate approximation of this transformation is found in Bock and Jones (1968), Appendix C.

16.) See Olsen (1980).

17.) Pps. 112-113.

18.) As the probability of civilian employment → 1, the numerator (a density function, symmetric around 0) → 0, the denominator (a cumulative distribution) → 1 and the entire ratio → 0.

19.) See Maddala (1983), pps. 273-274.

20.) See Manski and Lerman (1978), pps. 1985-1986.

21.) See The National Longitudinal Surveys Handbook (1983), pps. 11-13, for a description of how the military and civilian samples were obtained.

22.) Pp. 1981.

23.) For examples of the application of the WESML estimator see Daula, Fagan and Smith (1982) and Daula and Smith (1984).

24.) See McFadden (1976).

25.) See Avery (1980). The identification problem that would result if some type of normalization constraint was not imposed can be clearly seen by considering the logit equation specification given in equation (2.18). From this equation, the probability of the ith individual choosing a particular occupation (say occupation 0) is expressed as

$$P_{i0} = \frac{e^{W_{i0}\delta_0}}{e^{W_{i0}\delta_0} + e^{W_{i1}\delta_1} + e^{W_{i2}\delta_2}}$$

$$= \frac{1}{e^{-W_{i0}\delta_0}\,(e^{W_{i0}\delta_0} + e^{W_{i1}\delta_1} + e^{W_{i2}\delta_2})},$$

and if the elements in W_{ik} are not unique to that outcome, then

$$P_{i0} = \frac{1}{e^{W_i(\delta_0-\delta_0)} + e^{W_i(\delta_1-\delta_0)} + e^{W_i(\delta_2-\delta_0)}}.$$

Therefore, unless one of the coefficient vectors is constrained (normalized), the estimated coefficients for this particular outcome cannot be uniquely identified.

26.) Pp. 17.

27.) See Becker (1975), Mincer (1974) or Rosen (1977) for a discussion of the theory of human capital.

28.) See Mincer (1974), Ashenfelter (1979), Wise (1975) and Medoff and Abraham (1980) for examples of empirical estimations of the returns to human capital investment.

29.) See U.S. General Accounting Office report GAO/NSID-84-41 (1984) for background information on military compensation.

Chapter 3

The Data Base

1. Introduction

A major strength of the current analysis consists of the data that are utilized for the empirical model estimation. The purpose of this chapter is to present a thorough description of these data.

The primary data source is the National Longitudinal Survey of Youths (NLS), 14-22 Years Old. A sub-sample of male high school graduates was drawn from this survey. This sub-sample was then augmented by additional data on military compensation and recruitment activities, local labor market conditions, and education costs. These additional data were drawn from: A.) the Defense Manpower Data Center's Enlistment Master Files, B.) The Current Population Survey for 1981 and C.) statistical publications provided by the National Center for Education Statistics.

The remainder of this chapter is organized into three sections. The following section contains a description of the primary data source (NLS). The third section provides a similar description of the additional data sources. Section 4 is reserved for concluding comments on the data base.

2. The National Longitudinal Survey of Youths (NLS)

The NLS Youth Survey is one of five surveys conducted by the Center for Human Resource Research, The Ohio State University, under a contract with the Office of Manpower Policy, Evaluations, and Research, U.S. Department of Labor.[1] The purpose of the survey was twofold: 1.) to assist in the evaluation of expanded employment and training programs, as legislated by 1977 amendments to the Comprehensive Employment and Training Act (CETA), and 2.) to replicate most of the material of the earlier youth cohort surveys. The total sample consisted of 12,700 individuals who were between the ages of 14-22 years old in 1979. Of these individuals, 5,700 were civilian males, 5,700 were civilian females, and 1,300 were military males and females. The individuals were interviewed annually between the years 1979-1983.

The 1982 and 1983 surveys were unavailable for use in the present analysis. Individuals were selected for inclusion in the analysis sub-sample only if they were males that graduated for high school between 1978 and 1980. The occupational status of these individuals

Table 3-1: Sample Breakdown by Year of Graduation and Occupation

Occupational Status in (t + 1)	Year of Graduation (t)			Total
	1978	1979	1980	
Military Service	123	57	40	220
Civilian Employment	291	346	277	914
College Enrollment	205	181	228	614
Total	619	584	545	1748

was then observed in the following interview year. A breakdown of the sub-sample by occupational status and year of graduation is presented in Table 3-1. In this table, an individual is classified as in the military if he is in the active force as of the survey interview date. The large number of military observations in the 1979 interview sample, relative to 1980 and 1981, is due to the overrepresentation of the military in the 1979 survey year. An individual falls into the civilian employment category if he was currently working, employed - but not working, or unemployed - but seeking work as of the interview date. College enrollees consist of those attending college on a full-time basis.

2.1 Description of the NLS Variables

A total of 470 variables was extracted from the NLS files. A complete list of these variables originally extracted is available from the author upon request. After aggregations, deletions and transformations, the resulting set consisted of 23 variables plus various identification codes. The following discussion will focus on this final group of variables. It should be noted that, unless otherwise stated, these variables are binary. These variables are based upon the responses to qualitative questions. An affirmative answer is represented by a value of 1, a negative response is coded as a 0.

ADDT: Indicator of the individual's desire for additional training. This variable is based on the response to the question "Not counting regular schooling or college, would you like to get any other occupational or job training?". This question was asked in the 1979 interview only.

BLACK: Indicator that the individual is Black.

COLPROG: Indicator of participation in a college preparatory program while attending grades 9-12. This question was asked in all of the 3 interview years.

CWAGE: The total wages and salaries (civilian) earned by the individual in the last calendar year. This was reported in all survey years.

DADMIL: Indicator of respondent's father/stepfather being a member of the military in 1978. Based on the response to the question "What kind of work did he (father/stepfather) do (in 1978)?". This question was asked in the 1979 interview only.

EDAD: The highest grade of formal education completed by the respondent's father. Reported in the number of years of education. This question was asked in the 1979 interview only.

EMOM: The highest grade of formal education completed by the respondent's mother. Reported in the number of years of education. This question was asked in the 1979 interview only.

ESIB: The highest grade of formal education completed by the respondent's oldest sibling. Reported in the number of years of education. This question was asked in the 1979 interview only.

HEXP: Indicator of the individual's educational expectations. Based on the response to the question "... what is the highest grade or year you think you will actually complete?". This question was asked in the 1979 and 1981 interviews. The 1981 response was used only if there was no recorded response for the 1979 interview.

HISP: Indicator that the individual is Hispanic.

MSTAT: The marital status of the individual at the time of the interview. This question was asked in all interview years.

PSIBS: The percentage of siblings in school. This was calculated from the ratio of the number of siblings in school to the total number of siblings in 1979.

TDEP: The number of family dependents as of the date of the interview. Reported for all interview years.

TFMAFS: The total number of family members (excluding the father) that were in the Armed Forces. Based on the reported occupational status of the respondent's family members during the 1979 interview.

TGPROG: The total number of government training programs completed by the respondent as of the date of interview. This question was asked in all interview years.

TNFINC: The total net family income for the previous calendar year. Reported in actual dollars and cents. This question was asked in all interview years.

TOTHRS: The total number of full-time equivalent weeks of work experience as of the date of interview. This variable was constructed from the individual's work history in three stages. The first stage consisted of determining the duration of employment at each job (there is a maximum of 5 possible jobs per interview date). This was done by either subtracting the starting date from the ending date (if the individual was no longer employed at that job) or subtracting the starting date from the interview date (if the individual was still at that job at the date of interview). The resulting job duration was calculated in terms of weeks. The second stage consisted of converting job duration into full-time equivalent weeks. This was accomplished by multiplying job duration by a full-time equivalent weight. This weight was the ratio of average hours worked per week at the job to hours worked per week, if employed full-time. A full-time work week was assumed to be 35 hours in length. The last stage consisted of aggregating the number of full-time equivalent weeks per job into a total measure of job experience.

<u>TPROG</u>: The total number of training courses completed as of the date of interview. This includes both government sponsored (TGPROG) and other vocational/training programs (TVTPROG). This variable was reported for all of the interview years.

<u>TSCORE</u>: The raw score results of the Armed Forces Qualification Test (AFQT). This score was constructed by summing the first four sub-sections of the Armed Services Vocational Aptitude Battery (ASVAB), of which the AFQT is a part. The ASVAB sub-sections used

were Arithmetic Reasoning, Word Knowledge, Paragraph Comprehension, and 1/2 of Numerical Operations. This examination was administered to the total sample in the 1981 interview. Raw score results were used as they provide more variation than the test category classifications would have.

TVTPROG: The total number of vocational/training programs (not government sponsored) completed by the individual as of the date of interview. This was asked in all survey years.

VOCPROG: Indicator of participation in a vocational program while in high school. This was reported in all survey years.

2.2 Descriptive Statistics

In order to generate a "feel" for the data, several descriptive statistical breakdowns are given. Table 3-2 contains sample means and standard deviations for the NLS variables, stratified by the observed choice of occupation.

Briefly, the statistics in this table illustrate several discernable differences between individuals in the various occupational categories. Measures of family educational attainment (EDAD,EMOM,ESIB,PSIBS) indicate that, on average, college enrollees come from the highest educated families. Military enlistees come from the next highest group and civilian employees from the lowest. Measures of individual expectations and attainment (COLPROG, HEXP, TSCORE) illustrate a similar pattern.

Individuals in the military and civilian sectors appear to come from larger families and have a greater number of dependents than those who enroll in college. Also, civilian sector individuals have a greater amount of labor market experience (TOTHRS) and training program participation (TGPROG, TVTPROG) than those in the other occupational categories.

The average values for total net family income (TNFINC) are, in general, consistent with expectations. Average family income for college enrollees is greater than that of individuals in the other occupational groups. The low average value of income for military enlistees appears to be somewhat unreasonable.[2] This is particularly noticeable upon controlling for parental education levels. This average value indicates that it may be possible that the reported income (family) income values are downward biased. However, these income values are based upon the individual's response. Therefore, if these values are measures of perceived family income, this variable may be a more accurate predictor of occupational choice than the actual family income would be.

Table 3-2: Descriptive statistics of NLS Variables by Occupation

Variable (unit of measure)	Occupation Status			Total Sample mean (Std. Dev)
	Civilian Emp. mean (Std. Dev.)	Military Serv. mean (Std. Dev.)	College mean (Std. Dev.)	
ADDT (0/1)	0.79 (0.41)	0.87 (0.34)	0.59 (0.49)	0.73 (0.45)
BLACK (0/1)	0.25 (0.44)	0.29 (0.45)	0.24 (0.43)	0.25 (0.44)
COLPROG (0/1)	0.15 (0.36)	0.20 (0.40)	0.55 (0.50)	0.30 (0.33)
CWAGE ($'s)	3,691.48 (3,341.80)			
DADMIL (0/1)	0.02 (0.13)	0.07 (0.26)	0.01 (0.10)	0.02 (0.14)
EDAD (Years)	8.33 (5.47)	10.97 (2.89)	12.69 (3.90)	11.65 (2.99)
EMOM (Years)	10.73 (2.79)	11.23 (2.38)	12.30 (3.04)	11.34 (2.92)
ESIB (Years)	9.36 (5.51)	12.61 (2.20)	13.26 (1.66)	11.15 (4.58)
HEXP (Years)	13.33 (1.86)	14.01 (2.12)	15.88 (1.57)	14.32 (2.15)
HISP (0/1)	0.14 (0.35)	0.09 (0.28)	0.14 (0.35)	0.14 (0.34)
MSTAT (0/1)	0.03 (0.16)	0.08 (0.27)	0.002 (0.04)	0.02 (0.15)
PSIBS (%)	49.52 (35.58)	50.58 (35.39)	64.75 (36.86)	54.78 (37.69)
TDEP (#)	0.09 (0.40)	0.16 (0.51)	0.01 (0.11)	0.07 (0.35)
TFMAFS (#)	0.03 (0.19)	0.02 (0.13)	0.02 (0.11)	0.02 (0.19)

Table 3-2: Descriptive Statistics of NLS Variables (Cont.)

Variable (unit of measure)	Occupational Status: Civilian Emp. mean (Std. Dev.)	Military Serv. mean (Std. Dev)	College mean (Std. Dev.)	Total Sample mean (Std. Dev.)
TGPROG (#)	0.02 (0.14)	0.02 (0.13)	0.02 (0.13)	0.02 (0.14)
TNFINC ($'s)	19,267.09 (12,972.60)	7,690.44 (5,539.45)	25,767.71 (16,700.56)	20,940.35 (16,257.61)
TOTHRS (Weeks)	56.80 (85.50)	33.67 (55.65)	39.089 (62.74)	47.64 (75.37)
TPROG (#)	0.16 (0.43)	0.14 (0.40)	0.11 (0.35)	0.14 (0.40)
TSCORE (#)	64.16 (19.79)	71.35 (16.54)	79.99 (16.53)	70.75 (20.05)
TVTPROG (#)	0.14 (0.40)	0.13 (0.38)	0.09 (0.32)	0.12 (0.37)
VOCPROG (0/1)	0.18 (0.38)	0.11 (0.31)	0.18 (0.24)	0.13 (0.33)

3. Supplemental Data

As stated above, the NLS sub-sample was augmented with state level data on various military and economic attributes. The military data were supplied by the Defense Manpower Data Center (DMDC). The state level economic data came from the Bureau of Labor Statistics (BLS). The college cost data came from various publications of the National Center of Education Statistics (NCES). What follows is a brief description of these data; descriptive statistics are found in Table 3-3. Several of the local economic and military specific variables were lagged one year. These variables are denoted by (t-1) in Table 3-3.

3.1 Military Specific Data

The military specific data consist of military income (MWAGE), the number of production recruiters (REC(t), REC(t-1)), maximum potential educational benefits from the Veterans Educational Assistance Program (VEAP), and national and local advertising expenditures (LADV(t), LADV(t-1), NADV(t), NADV(t-1)). With the exception of MWAGE, these variables are specific to the individual only in terms of the individual's state of residence at the time of high school graduation.

VEAP is institutionally determined with no cross-sectional or individual specific variation. However, there is variation in this attribute across sample years. The values used are in present value dollars, discounted over the average enlistment term of 3 years.[3]

MWAGE was determined by the functional relationship in equation (2.33). Cross-sectional variation in this measure is due to differences in the various supplemental allowances between enlistees with and without dependents. The data for these allowances and basic pay came from the DOD Military Compensation Background Papers (1982). Tax advantage estimates were provided by the U.S. Army Finance Center, Ft. Harrison, IN. The number of production recruiters per state was determined by a mapping of recruiting districts to the state.[4] This mapping was necessary because the recruiting districts did not always fall entirely within a given state, but overlapped state boundaries. Local and national advertising expenditures are expressed in terms of dollars per 18-20 year old males for the state of residence. Ideally, data on actual advertising exposure time should be used.[5] However, this type of data is unavailable for DOD-wide advertising.

In general, the descriptive statistics of these variables (see Table 3-3) indicate little discernable difference between the occupational

Table 3-3: Descriptive Statistics of Supplemental Variables

Variable (unit of measure)	Occupational Status			Total Sample mean (Std. Dev.)
	Civilian Emp. mean (Std. Dev.)	Military Serv. mean (Std. Dev.)	College mean (Std. Dev.)	
COST ($'s)	876.25 (319.08)	836.31 (300.14)	871.08 (284.74)	869.40 (305.14)
LADV(t) ($'s/Pop)	0.115 (0.029)	0.156 (0.030)	0.153 (0.029)	0.155 (0.029)
LADV(t-1) ($'s/Pop)	0.158 (0.030)	0.159 (0.030)	0.157 (0.027)	0.158 (0.029)
MWAGE ($'s)	8,385.65 (506.43)	8,219.81 (511.74)	8,412.82 (529.15)	8,8374.32 (518.37)
NADV(t) ($'s/Pop)	0.821 (0.119)	0.788 (0.111)	0.802 (0.112)	0.810 (0.116)
NADV(t-1) ($'s/Pop)	0.775 (0.133)	0.730 (0.119)	0.792 (0.142)	0.776 (0.136)
REC(t) (#)	434.66 (295.89)	407.27 (292.42)	449.46 (292.14)	436.41 (294.26)
REC(t-1) (#)	420.87 (290.10)	413.56 (307.46)	438.49 (292.28)	426.14 (293.08)
UNRATE(t) (%)	6.32 (1.45)	6.09 (1.12)	6.43 (1.57)	6.33 (1.46)
UNRATE (t-1) (%)	6.27 (1.35)	6.49 (1.33)	6.32 (1.43)	6.31 (1.37)
VEAP ($'s)	4,914.44 (589.84)	4,653.65 (609.34)	4,847.19 (617.41)	4,858.00 (596.30)

categories. It should be noted that the appparent lower than average values of COST, VEAP, NADV, REC and MWAGE for the military category is partially attributed to the overrepresentation of this category in the 1979 sample.

3.2 Local Economic Data

The only local economic attribute used was the state unemployment rate (UNRATE). The data used was supplied by C. Brown, University of Maryland.[6] Various other state level economic indicators (i.e. - average civilian wages) could have been included. But, given the available individual level information on earnings, this type of data would most likely not have added much insight on the choice of occupation.

3.3 Cost of College Data

The cost of a year of college education (COST) is represented by the basic in-state tuition charges of 4-year state universities.[7] These tuition charges were drawn from a series of NCES publications on college costs for selected public and private universities.[8] For states with more than one 4-year public institution, tuition costs were constructed as an arithmetic average across these universities.

For 1979, Washington, D.C. posed a particular problem. Prior to 1980, there were no reported public university tuition rates. To correct for this, an arithmetic average of the out-of-state tuition charges for Maryland and Virginia was used.

4. Concluding Comments

Several concluding comments about the data set should be made. First, in order to maintain a reasonable sample size, several potentially interesting attitudinal (NLS) variables had to be excluded due to missing value problems. Second, for a small number of cases, values had to be imputed for net family income (TNFINC). The method of imputation is presented in Appendix A.

The final data base was constructed by merging the state level data with the NLS sub-sample. The NLS Public Use Tapes provide very limited information of the individual's place of residence. This limitation necessitated a primary merge of the NLS sub-sample with the NLS Geocode Files. The Geocode Files provided data on the individual's state of residence at the age of 14 and between the years of 1979-1981. Because the 1979-1981 state residential codes proved to be unreliable, the age 14 residence code was used instead. It is recognized that, due to interstate migration, this code could introduce some inaccuracies. But, given the present situation, this was the most accurate measure available. A secondary merge with the state level data was then undertaken to produce the final data base.

End Notes

1.) For more detailed background information on the NLS see the National Longitudinal Surveys Handbook (1983) or Parnes (1975).

2.) For example, a 1982 survey of Army personnel found an average total annual family income that ranged between $15,000-26,000, in 1982 dollars. See The 1982 DA Survey of Personnel Entering the Army, Vol. 2, pps. 372-373 (1984).

3.) This data was provided by C. Brown, University of Maryland. See Brown (1984), Data Appendix for a more complete description of how this variable was constructed.

4.) The mapping procedure used was provided by C. Brown, University of Maryland.

5.) Daula and Smith (1984) used an exposure time variable (for the Army only) and found it to have better explanatory power than expenditure measures.

6.) See Brown (1984), Data Appendix.

7.) A truer measure of the cost of college would take into account room and board costs. However, inclusion of these additional costs would require additional information (or assumptions) as to who would become an on campus resident. Willis and Rosen (1979) dealt with this problem by assuming that all individuals that lived beyond a certain distance from the nearest university would become on campus residents. Due to the strong nature of this assumption, this approach will not be followed in the present analysis.

8.) The NCES publications used were: 1981 Digest of Educational Statistics (for academic year 1979/1980), College Costs 1980-1981 (for academic year 1980/1981) and College Costs 1981-1982 (for academic year 1981/1982).

Chapter 4

Empirical Estimates

1. Introduction

The empirical model specifications were based upon the hypotheses presented in Chapter 2. Prior to the discussion of the empirical results, several preliminary comments are in order.

A total of 8 model specifications were estimated. Two general specifications were used. These specifications were: A.) A "Full Model" that contained the complete set of exogeneous variables, as defined in Table 2-2, and B.) A "Limited Model" that excluded the advertising and recruiter variables. In order to keep the discussion as focussed as possible, only the "Limited Model" estimates are presented in this chapter. The other model specification estimates are found in Appendix B. For the same reason, the reduced form model estimates are not presented. However, these estimates can be obtained directly from the author upon request.

Three estimation methodologies were used. The first method consisted of a simple two-stage procedure. The second method was a two-stage procedure that corrected for self-selectivity bias. The third method was a two-stage procedure that corrected for both self-selectivity bias and choice-based sampling bias.

The actual process of estimating these models turned out to be more complex than was anticipated. There were no readily available econometric software packages that contained the procedures necessary to calculate the self-selectivity and choice-based sampling bias correction factors.[1] The software package that came the closest to meeting these estimation requirements was LIMDEP.[2] This package was modified to compute the polychotomous inverse Mills ratio and the correct variance-covariance matrix for the choice-based sample estimations.[3,4]

The Logit equations were estimated via a maximum likelihood technique. Most of the functions converged after only 6 or 7 interations.[5] Even so, these estimations proved to be rather costly in both time and estimated computer cost.[6] The civilian wage equations were estimated by a standard ordinary least squares technique.

As a final point, it should be noted that all of the monetary variables were deflated to 1978 dollars. This was done to control for the effects of inflation between the years 1979-1981.

The remainder of this chapter is organized into four sections. The following section presents the civilian wage equation estimates. Section 3 reports the structural selection equation estimates. These estimates are interpreted in Section 4. The last section is reserved for a summary of the major empirical findings.

2. Civilian Wage Equation Estimates

Several specifications of the civilian wage equation were tested. These specifications differed only in the forms of the unemployment rate, labor force work experience and training course variables. Two forms of the unemployment rate were used. These were the unemployment rate during the year of graduation (UNRATEO) and the difference in the unemployment rate between the year of graduation and the year prior to graduation (DUNRATE). This second formulation was used primarily because of its expected impact in the selection equations. Estimates with both forms are reported.

Labor force work experience was tested under three forms. These were a simple linear form (TOTHRS), a quadratic form (TOTHRS and TOTHRSQ) and a natural logarithm form (LNHRS). The quadratic and natural logarithm forms performed better than the simple linear form. The logarithm form performed marginally better than the quadratic form and is the only one presented in this chapter.

The training course variables (TVTPROG and TGPROG) were entered both separately and as an aggregate variable (TPROG). There was no significant difference between either specification and only the aggregate form estimates are reported.

Before proceeding, two points should be kept in mind. First, the civilian wage estimates are not the primary focus of this analysis. Therefore, the following discussion of these estimates is relatively brief. To facilitate this discussion, a description of the variables is given in Table 4-1. The civilian wage equation estimates associated with the "Limited Model" specifications are given in Tables 4-2 to 4-5. The wage equations reported in these tables differ only in the specification of the unemployment rate variable and the estimated values of the Mills ratio (MILLS). In Models 1 and 3, the straight unemployment rate (UNRATEO) was used. Models 2 and 4 used the first difference in the unemployment rate (DUNRATE). Recall from Chapter 2 that the Mills ratio is calculated from the reduced form civilian employment selection equation estimates. The reduced form selection equation used to estimate the variable MILLS in Models 1 and 2 included the net family income (NETY). NETY is the total net family income (TNFINC), deflated to 1978 dollars. The MILLS for Models 3 and 4, however, are based on a reduced form model that used a standardized value of NETY (STDY). The differences in these two specifications of family income are discussed in more detail in the following section. The wage equation estimates for the other model specifications appear in Appendix B.[7]

Second, the difference in the reported wage equation estimates within each table may appear small. This is particularly noticeable for the differences between the self-selectivity corrected estimates and the self-selectivity/choice-based sample corrected estimates. The reason for this is also due to minor differences in the calculated values for the variable MILLS. Even though these estimates are very similar in some instances, they are reported to illustrate the effects of these correction procedures.

Table 4-1: Civilian Wage Equation Variable Definition

Variable Name	Definition
BLACK	: Black dummy variable (= 1 if Black).
CWAGE	: Dependent Variable - Reported annual civilian wages and salaries (in $'s).
DUNRATE	: Change in unemployment rate.
HISP	: Hispanic dummy variable (= 1 if Hispanic).
LNHRS	: Natural logarithm of number of weeks of work experience.
MILLS	: Inverse Mills ratio self-selectivity correction factor.
MSTAT	: Marital status (= 1 if married).
TPROG	: Total number of training courses completed.
TSCORE	: Armed Forces Qualification Test score.
UNRATEO	: Unemployment rate in the year of graduation.
VOCPROG	: In vocational program while in high school (= 1 if yes).

Table 4-2: Model 1 Civilian Wage Equation Estimates

Variable	Uncorrected	Corrected(1)	Corrected(2)
Constant	37.527	-73.820	-74.413
	(0.077)	(0.152)	(0.153)
BLACK	-812.827 **	-580.642 **	-609.031 **
	(2.979)	(2.043)	(2.109)
DUNRATE	-16.334	-26.974	-15.156
	(0.213)	(0.352)	(0.198)
HISP	-14.590	86.435	83.960
	(0.048)	(0.285)	(0.276)
LNHRS	887.782 **	831.548 **	847.038 **
	(12.434)	(11.247)	(11.477)
MILLS		-771.222 **	-622.953 **
		(2.790)	(2.124)
MSTAT	1589.160 **	1600.800 **	1549.450 **
	(2.485)	(2.512)	(2.426)
TPROG	260.703	221.301	228.749
	(1.100)	(0.935)	(0.965)
TSCORE	9.684 *	19.626 **	17.465 **
	(1.666)	(2.887)	(2.546)
UNRATE0			
VOCPROG	398.743	321.601	337.294
	(1.503)	(1.210)	(1.267)
R^2 (Adjusted)	0.198	0.204	0.201
F-Statistic	29.095	26.921	26.464

Note: 1. Corrected(1) is self-selectivity bias corrected and Corrected(2) is both self-selectivity and choice base sampling bias corrected.
2. Absolute value T-Statistics are in parens below the coefficient estimates.
* - Statistically significant at the 10% level.
** - Statistically significant at the 5% level.

Table 4-3: Model 2 Civilian Wage Equation Estimates

Variable	Uncorrected	Corrected(1)	Corrected(2)
Constant	807.731	653.608	648.710
	(1.201)	(0.972)	(0.960)
BLACK	-821.440 **	-592.687 **	-621.700 **
	(3.017)	(2.088)	(2.156)
DUNRATE			
HISP	-35.646	68.152	61.309
	(0.119)	(0.226)	(0.203)
LNHRS	881.531 **	827.489 **	842.271 **
	(12.351)	(11.210)	(11.427)
MILLS		-753.451 **	-609.309 **
		(2.735)	(2.077)
MSTAT	1537.510 **	1553.710 **	1501.330 **
	(2.406)	(2.440)	(2.353)
TPROG	274.636	232.530	241.386
	(1.161)	(0.984)	(1.020)
TSCORE	9.299	19.066 **	16.942 **
	(1.602)	(2.804)	(2.468)
UNRATEO	-114.670 *	-108.684	-107.359
	(1.662)	(1.580)	(1.557)
VOCPROG	415.373	340.799	353.778
	(1.569)	(1.285)	(1.331)
R^2 (Adjusted)	0.200	0.206	0.203
F-Statistic	29.522	27.261	26.817

Note: 1. Corrected(1) is self-selectivity bias corrected and Corrected(2) is both self-selectivity and choice base sampling bias corrected.
2. Absolute value T-Statistics are in parens below the coefficient estimates.
* - Statistically significant at the 10% level.
** - Statistically significant at the 5% level.

Table 4-4: Model 3 Civilian Wage Equation Estimates

Variable	Uncorrected	Corrected(1)	Corrected(2)
Constant	37.527	-74.038	-74.933
	(0.077)	(0.153)	(0.154)
BLACK	-812.827 **	-582.783 **	-609.563 **
	(2.979)	(2.055)	(2.116)
DUNRATE	-16.334	-26.751	-15.073
	(0.213)	(0.349)	(0.197)
HISP	-14.590	86.497	84.680
	(0.048)	(0.286)	(0.278)
LNHRS	887.782 **	830.830 **	846.386 **
	(12.434)	(11.242)	(11.472)
MILLS		-782.453 **	-633.104 **
		(2.839)	(2.165)
MSTAT	1589.160 **	1594.630 **	1545.070 **
	(2.485)	(2.503)	(2.419)
TPROG	260.703	221.781	228.843
	(1.100)	(0.937)	(0.965)
TSCORE	9.684 *	19.741 **	17.573 **
	(1.666)	(2.909)	(2.566)
UNRATEO			
VOCPROG	398.743	321.553	336.770
	(1.503)	(1.211)	(1.265)
R^2 (Adjusted)	0.198	0.204	0.201
F-Statistic	29.095	26.959	26.488

Note: 1. Corrected(1) is self-selectivity bias corrected and Corrected(2) is both self-selectivity and choice base sampling bias corrected.
2. Absolute value T-Statistics are in parens below the coefficient estimates.
* - Statistically significant at the 10% level.
** - Statistically significant at the 5% level.

Table 4-5: Model 4 Civilian Wage Equation Estimates

Variable	Uncorrected	Corrected(1)	Corrected(2)
Constant	807.731	650.907	646.584
	(1.201)	(0.968)	(0.957)
BLACK	-821.440 **	-594.880 **	-622.090 **
	(3.017)	(2.100)	(2.163)
DUNRATE			
HISP	-35.646	68.003	61.955
	(0.119)	(0.226)	(0.205)
LNHRS	881.531 **	826.847 **	841.630 **
	(12.351)	(11.207)	(11.422)
MILLS		-763.991 **	-619.446 **
		(2.780)	(2.118)
MSTAT	1537.510 **	1547.650 **	1497.040 **
	(2.406)	(2.431)	(2.346)
TPROG	274.636	233.095	241.534
	(1.161)	(0.987)	(1.021)
TSCORE	9.299	19.174 **	17.051 **
	(1.602)	(2.825)	(2.488)
UNRATEO	-114.670 *	-108.298	-107.122
	(1.662)	(1.575)	(1.554)
VOCPROG	415.373	340.698	353.226
	(1.569)	(1.285)	(1.329)
R^2 (Adjusted)	0.200	0.206	0.203
F-Statistic	29.522	27.296	26.842

Note: 1. Corrected(1) is self-selectivity bias corrected and Corrected(2) is both self-selectivity and choice base sampling bias corrected.
2. Absolute value T-Statistics are in parens below the coefficient estimates.

* - Statistically significant at the 10% level.

** - Statistically significant at the 5% level.

In general, the estimates reported in these tables indicate no unexpected results. The summary statistics were low, but not unreasonable, given the type of data used.[8] The selectivity bias corrected estimates (columns 2-3) had marginally better overall fits than the uncorrected equation estimates. This result is directly attributed to the bias correction factor (MILLS). This variable was significant and correctly signed in all of the various functional forms. Evaluated at the sample means of the explanatory variables, the average effect of the selectivity bias correction factor (in monetary terms) was to reduce the average predicted wage approximately 190 (current value) dollars.

The human capital measures (LNHRS, TPROG, and VOCPROG) were all positive, although only the labor force experience measure (LNHRS) was statistically significant. The "ability" proxy (TSCORE) was also positive (as expected) and statistically significant for all of the bias corrected estimates. It is interesting to note that the size of the estimated coefficient for this variable is approximately twice as large for the bias corrected estimates, compared to the uncorrected ones. A possible explanation for this result is that the majority of higher "ability" individuals chose not to enter the civilian labor market immediately. If so, the estimated (corrected) returns to those higher "ability" individuals who chose civilian employment would be larger than uncorrected estimates would indicate.

The estimated effects of the black dummy variable and marital status were slightly surprising. It was hypothesized that marital status reflected the individual's labor market attachment and would therefore positively affect the civilian wage.[9] The size of this estimated effect was larger than expected. For example, evaluating equation 4 at the sample means of the exogeneous variables indicates that a married individual would earn on average approximately 55% more than a single individual. However, without additional information, it is difficult to arrive at an alternative explanation (with any level of confidence) of this estimated effect.

The estimated black effect was also stronger than expected. Given that all the individuals in the sample are high school graduates, it was not expected that such a large difference in the estimated wage would exist between blacks and the other ethnic/racial groups. This result is supported by the estimated (insignificant) Hispanic effect. It can only be postulated that this effect reflects either a lower labor force attachment and/or the existence of a possible direct/indirect discrimination effect.[10]

The unemployment rate failed to have a significant effect under either functional form. This is not all that surprising, for several reasons. First, part of the unemployment rate effect is to discourage individuals from seeking jobs (i.e. - the "discouraged worker hypothesis"). This discouragement effect is partially reflected in the self-selectivity bias correction variable. This effect can be clearly seen by comparing the uncorrected and corrected coefficient estimates for Models 2 and 4. Secondly, it may be possible that the unemployment rate used was not the most appropriate one. If the data were available, a state level teenage rate, by race/ethnic group, would perhaps have yielded more statistically solid estimates.

3. Structural Selection Equation Estimates

As stated in the introduction, only the 4 "Limited Model" specifications are presented in this section. The only differences between these specifications is in the formulation of the unemployment rate and net family income variables. Two forms of each of these variables were used. The unemployment rate was entered as either a straight rate for the year of graduation (UNRATE0) or the difference in the rates between the year prior to graduation and the year of graduation (DUNRATE). The rationale for this second formulation was to test if the individual is more sensitive to changes in labor market conditions compared to the static characteristics of the market.[11] The net family income variable was reformulated as a deviation from an ethnic/racial group standardized value.[12] This formulation was used to test if relative family income was more important than absolute income in the occupational choice decision.

The major difference between the specifications in this section and those in Appendix B is the absence of most of the "military specific" variables. In brief, the results reported in Appendix B indicated that these "military specific" variables were in all cases statistically insignificant.[13] Also, it was found that the education level of the individual's oldest sibling was highly correlated with that of the individual's father.[14]

As in the previous section, a brief definition of the variables is given to facilitate the following discussion (see Table 4-6). The empirical estimates are reported in Tables 4-7 to 4-18. The large number of tables is due to the three estimation techniques. It should be noted that the interpretation of logit estimates differs considerably from that of a more standard ordinary least squares model. A primary difference is that the coefficient estimates should not be interpreted as having a direct effect on the dependent variable. Rather, the coefficient estimates should be interpreted as representing the nonlinear effects of the independent variables on the occupational probabilities. The signs of the coefficient estimates, however, can be directly interpreted. Therefore, the present discussion will focus on the direction and level of statistical significance of the estimates. A more standard interpretation of the estimates is given in Section 4, below.

A second difference in estimate interpretation is with "Goodness of Fit" statistics. With ordinary least squares analysis, an R^2 Statistic yields a fairly precise measure of how well the model does at explaining the variation of the dependent variable around its' mean value. There is no precisely equivalent statistic for multinomial logit models. In fact, there appears to not even be a consensus in the literature on an appropriate "Goodness of Fit" statistic in general.[15] Two fit measures are used in the present analysis. The first consists of the percentage improvement in the percentage of outcomes correctly predicted by the model, relative to a "simple" percentage predicted.[16] A "simple" percentage prediction is calculated by dividing the number of observations in the "dominant" outcome category by the total number of observations. This "simple" prediction is then subtracted from the model percentage correctly predicted and the result is then divided by the "simple" prediction to calculate the percentage improvement in the number of correctly predicted outcomes. Or,

$$(4.1) \qquad \%\ \text{Improvement} = (P_m - P_s)/P_s,$$

where, P_m is the percentage correctly predicted by the model and P_s is the "simple" percentage predicted.[17] The rationale for this statistic can be seen in the following example. Consider a binary choice model where 75% of the observations choose outcome "A" and 25% do not. If a model is estimated that correctly predicts 80% of the total outcomes, on face value, one would consider this a fairly good fit. However, this model only predicted 5% more correct outcomes than if all of the observations were assigned outcome "A". Therefore, this statistic illustrates how well the model performs relative to simply assigning all of the observations to the dominant outcome category. For the current analysis, the "simple" percentage predicted (letting civilian employment be the dominant category) was calculated as 52.29%. The percentage correctly predicted (P_m) was determined by assigning the observations to the outcome with the highest predicted probability.

The second "Goodness of Fit" statistic used is McFadden's ρ^2 statistic.[18] This statistic is defined as

$$(4.2) \qquad \rho^2 = 1 - [L(\beta)/L(0)],$$

where $L(\beta)$ is the value of the likelihood function evaluated at the mean values of the independent variables and $L(0)$ is the value of the likelihood function constraining all of the model paremeters to 0. McFadden argues that this statistic can be used to find an approximate R^2 equivalent value. Using McFadden's equivalency approximation,[19] the ρ^2 ranges (.368 - .394) are approximately equal to an R^2 range of .72 - .75.

The following discussion will first center on Models 1 and 2 (see Tables 4-7 to 4-12). These models were estimated first and are used to logically motivate the formulation of the standardized family income variable. At first glance, it appears that the straight unemployment rate specification (Model 2) yielded a slightly better overall fit, compared to the difference specification (Model 1). Also, the coefficient estimates (again, in general) were relatively stable between the two bias corrected estimations.

The individual and family demographic variables (BLACK, HISP, EDAD, EMOM, PSIBS, MSTAT, DADMIL) yielded several unexpected results. The most surprising of these was for the Black and Hispanic dummy variables. It was hypothesized that Blacks and Hispanics would have a higher propensity towards the military and a lower one towards college. The uncorrected estimates, however, indicated an exactly opposite tendency. This result could partially be attributed to the fact that the sample is not a random sample of these two groups. Rather, all of the individuals in the sample are high school graduates. With lower graduation rates for these groups, relative to the national average, it is likely that the Blacks and Hispanics in this sample are in the upper end of the distribution (of college bound) for their respective groups.

The bias corrected estimates yielded considerably different Black and Hispanic effects for military service. The black effect goes from a statistically significant negative estimate to a weakly positive, statistically insignificant estimate. This result is most likely due to the characteristics of the civilian and military samples. The uncorrected estimates indicated that, ceteris paribus, the proportion of blacks in the military, relative to the occupational groups, is small. Weighting the military sample reduced the importance of each observation in this sample by approximately two thirds. The weights did not significantly alter the importance of the observations in the civilian sample. Therefore, any Blacks who joined the military and were in the civilian sample are given a greater importance than those who joined and are in the military sample. This could have significantly altered the proportion of weighted observations in the military. The bias corrected Hispanic estimates indicate a similar, though weaker, effect.

The estimated effects of the parental education variables were, for the most part, consistent with the hypothesized effects. The father's education level was estimated as having a positive and statistically significant influence on the college enrollment decision. A similar positive effect was estimated for enlisting. This result was not anticipated. A plausible explanation for this result is not readily apparent. The lack of a statistically significant effect of the mother's education is not of major importance. It simply indicates that the individual (recall, they are all males) is more influenced by the father than the mother.

It appears that the estimated effect of the percentage of siblings in school (PSIBS) does not reflect a "cost of college". Rather, the estimates seem to indicate that this variable is more of a proxy for propensity to enter college. The lack of effect on the enlistment equation was expected. This lack of effect confirmed the view of this variable being "college specific".

The estimates of marital status on the decision to join the military was also counter to the hypothesized effect. It was argued that the frequent extended absences from home that military service requires would have a disincentive effect. The estimates tend to indicate that this disincentive effect is not an important determinant of the enlistment decision. Instead, it appears that the monetary benefits enjoyed by service family members are a more important factor in the choice.[20] The effects on college enrollment were more consistent with expectations.

The father in the military dummy variable (DADMIL) was marginally significant (i.e. - at the 10% level) and correctly signed for the enlistment equation in the uncorrected and self-selectivity corrected specifications. It was significant in the self-selectivity/choice based corrected specifications. It had no statistically significant effect in the college enrollment equation. This indicates that there is an influence on the individual's decision to enlist. It can only be speculated upon whether this variable represents a true influence, or a source of information on military life. The statistically insignificant effect of this variable on the enrollment equation was consistent with the hypothesized indeterminate impact.

The "ability" proxy variable (TSCORE) yielded consistently stable and statistically significant parameter estimates over all specifications. In the college enrollment equations, the estimates were as expected. Higher "ability" does increase the probability of going to college, ceteris paribus. The similar estimated relationship for the military cannot be interpreted as clearly.

The question on the individual's desire to seek additional training (ADDT) came out to be a better predictor of enlistment than was expected. It was statistically significant and correctly signed over all specifications. Two tentative interpretations of this result are made. First, if the individual seeks more training, and is indifferent to the source of this training, then there should be no discernible effect on the enlistment probability. However, if the military is considered a least cost method of obtaining such training, this estimated effect would result. Secondly, these estimates may also reflect some indirect effects of military recruitment policy. The military has periodically orchestrated various advertising campaigns to represent the military as a method of obtaining jobs skills, particularly in more technical fields. Although this appears to be a tenuous explanation, it cannot be ruled out. This explanation is further supported by the estimated effects of the ability variable TSCORE. The estimated effect on college enrollment is also consistent with expectations. In fact, by the structure of the question any other outcome would have been rather disconcerting.

Two additional variables were used to represent the individual's intentions towards college. These were the highest grade expected to complete (HEXP) and a dummy variable for participation in a college preparatory program (COLPROG). The estimates indicate that both of these variables have strong positive effects on this decision. Also, consistent with the hypotheses that these attributes are "college specific" neither one had a statistically significant effect in the military enlistment equations.

The last group of variables consists of purely economic type measures (UNRATEO/DUNRATE, NETY, COSTO, VEAP, MWAGE, PWAGE). The unemployment rate variable estimates were consistently statistically insignificant across functional form and possible outcome. It is difficult to believe that these individuals are unaffected by local labor market conditions, as these estimates indicate. There are two additional explanations for these estimates. First, the unemployment rates used were state-level, average rates. These rates may not truly reflect local labor market conditions. Secondly, there may not have been sufficient variation in these unemployment rate measures. This implies that perhaps a longer time-series analysis is required to truely capture the unemployment rate effect.

Net family income (NETY) appears to be a major determinant of occupational choice. The estimates were consistently statistically significant and stable across functional forms and estimation techniques. The tuition cost variable (COSTO) and the military educational benefits variable (VEAP) did not perform as well. The apparent insignificance of tuition costs most likely indicates that this lower bound approximation does not reflect the true costs of college very precisely.

The uncorrected effect of VEAP was the opposite of the hypothesized effect on military service. The choice-based bias corrected estimates were of the correct sign, although statistically insignificant. This result is interpreted as implying that those who enter the military do not consider post-service educational benefits as an important determinant of this decision.

The last two economic variables are the civilian and military wages. The estimated effects of the civilian wage (PWAGE) were as expected for both military service and college. For military service, the military wage (MWAGE) estimates were correctly signed across functional forms and estimation techniques. They were statistically significant in only the uncorrected and self-selectivity bias corrected equations. There were no measurable effects of this wage on the college decision.

The discussion will now focus on Models 3 and 4. These models differ from Models 1 and 2 only in the formulation of the net family income variable. The motivation for the formulation of family income in these models (STDY) rests with the apparent anomalous racial/ethnic dummy variable estimates in Models 1 and 2. This respecification of family income was used to test if relative income was a more important determinant of initial occupational choice than absolute income. If so, it was hypothesized that the racial/ethnic dummy variable estimates would become more consistent with the original set of hypotheses.

The estimates indicate that a relative income effect does exist. The Black and Hispanic dummy variable estimates changed dramatically, particularilly across the uncorrected models. This result implies that the individual's choice of occupation is more sensitive to the level of family income relative to the ethnic/racial group average than to that of the total population.

The only other major difference between the the estimates of Models 3 - 4 and 1 - 2 was for marital status. The estimated coefficients for this variable were approximately twice as large in Models 3 - 4 for the military enlistment outcome. This implies that by controlling for the individual's relative net family income, marital status plays a larger role in the determination of initial occupational choice. This effect is particularily true for the military possibility.

Table 4-6: Structural Selection Equation Variable Definitions

Variable Name	Definition
ADDT	: Desires additional training outside of college (=1 if yes).
BLACK	: Black dummy variable (=1 if Black).
COLPROG	: College prepartory program participation (=1 if participated).
COSTO	: In-state tuition costs for the year of graduation.
DADMIL	: Father in military during year of graduation (=1 if yes).
DUNRATE	: Change in unemployment rate between graduation year and prior.
EDAD	: Highest grade completed by respondent's father.
EMOM	: Highest grade completed by respondent's mother.
HEXP	: Highest grade respondent expects to complete.
HISP	: Hispanic dummy variable (=1 if Hispanic).
MSTAT	: Marital status (= 1 if married).
MWAGE	: Military wage ($'s).
NETY	: Total net family income in year of graduation ($1000's).
PSIBS	: Percentage of siblings in school.
PWAGE	: Predicted civilian wage ($'s).
STDY	: Total net family income standardized to ethnic/racial cohort.

Table 4-6: Structural Equation Variable Definitions (Cont.)

Variable Name	Definition
TSCORE	: Armed Forces Qualification Test score.
UNRATEO	: Unemployment rate in the year of graduation.
VEAP	: Education benefits under the VEAP ($'s).

Table 4-7: Model 1 Structural Selection Equation Estimates

- Uncorrected -

Variable	Military Service	College
Constant	-20.4290 **	-10.6115 **
	(3.420)	(2.588)
ADDT	0.724732 **	-0.300998 *
	(2.807)	(1.929)
BLACK	-0.609139 **	0.477392 **
	(2.155)	(2.195)
COLPROG	-0.135143	0.763704 **
	(0.549)	(4.838)
COSTO	-0.872348-3	-0.855401-3
	(1.234)	(1.538)
DADMIL	1.05767 *	-0.851483
	(1.816)	(1.248)
DUNRATE	-0.113575	0.691228-1
	(0.865)	(0.807)
EDAD	0.124649 **	0.137333 **
	(5.253)	(7.374)
EMOM	0.169380-1	-0.831646-2
	(0.407)	(0.270)
HEXP	0.559670-1	0.460593 **
	(1.091)	(11.453)
HISP	-0.797852 **	0.738216 **
	(2.185)	(3.168)
MSTAT	-0.168921-1	-1.47786
	(0.022)	(1.274)
MWAGE	0.554651-2 **	0.150389-3
	(3.226)	(0.125)
NETY	-0.404752 **	0.305530-1 **
	(11.524)	(2.410)
PSIBS	0.305425	0.694926 **
	(1.157)	(3.557)
PWAGE	-0.101780-2 **	-0.586036-3 **
	(6.915)	(4.585)
TSCORE	0.326546-1 **	0.322234-1 **
	(4.769)	(6.141)
VEAP	-0.108142-2 **	-0.251665-3
	(2.596)	(0.778)

Percentage Prediction Improvement = 41.57 $\rho^2 = 0.389$

Note: 1. Coefficient estimates for the civilian sector outcome are normalized to zero.
2. Absolute value T-Statistics are in parens below the coefficient estimates.

* - Statistically significant at the 10% level.

** - Statistically significant at the 5% level.

Table 4-8: Model 1 Structural Selection Equation Estimates

- Self-Selectivity Bias Corrected -

Variable	Military Service	College
Constant	-18.9946 **	-9.87675 **
	(3.187)	(2.420)
ADDT	0.710196 **	-0.291678 *
	(2.757)	(1.867)
BLACK	-0.500354 *	0.524995 **
	(1.793)	(2.437)
COLPROG	-0.175039	0.726376 **
	(0.711)	(4.578)
COSTO	-0.781804-3	-0.804711-3
	(1.106)	(1.447)
DADMIL	1.00821 *	-0.830970
	(1.714)	(1.225)
DUNRATE	-0.116170	0.626207-1
	(0.886)	(0.731)
EDAD	0.111903 **	0.129053 **
	(4.719)	(6.967)
EMOM	0.148641-1	-0.839768-2
	(0.357)	(0.273)
HEXP	0.296146-1	0.438596 **
	(0.572)	(10.773)
HISP	-0.746866 **	0.741400 **
	(2.052)	(3.180)
MSTAT	0.943369-2	-1.46288
	(0.012)	(1.262)
MWAGE	0.515870-2 **	0.318508-5
	(3.010)	(0.003)
NETY	-0.387260 **	0.313814-1 **
	(11.172)	(2.468)
PSIBS	0.247655	0.650784 **
	(0.941)	(3.346)
PWAGE	-0.949137-3 **	-0.547410-3 **
	(6.672)	(4.394)
TSCORE	0.332942-1 **	0.325921-1 **
	(4.859)	(6.198)
VEAP	-0.985761-3 **	-0.213146-3
	(2.370)	(0.660)

Percentage Prediction Improvement = 41.57 ρ^2 = 0.388

Note: 1. Coefficient estimates for the civilian sector outcome are normalized to zero.

2. Absolute value T-Statistics are in parens below the coefficient estimates.

* - Statistically significant at the 10% level.

** - Statistically significant at the 5% level.

Table 4-9: Model 1 Structural Selection Equation Estimates

- Self-Selectivity and Choice-Base Bias Corrected -

Variable	Military Service	College
Constant	-5.92906 (1.113)	-10.655 ** (2.543)
ADDT	0.582569 ** (2.541)	-0.273558 (1.701)
BLACK	0.111708-2 (0.004)	0.450747 ** (2.044)
COLPROG	-0.170697 (0.798)	0.704818 ** (4.332)
COSTO	-0.610698-3 (0.968)	-0.824691-3 (1.455)
DADMIL	1.43689 ** (2.597)	-0.804629 (1.142)
DUNRATE	-0.179689 (1.593)	0.747758-1 (0.846)
EDAD	0.101780 ** (4.744)	0.127029 ** (6.577)
EMOM	0.426104-2 (0.115)	-0.129727-1 (0.407)
HEXP	-0.470021-2 (0.102)	0.440612 ** (10.580)
HISP	-0.588882 * (1.896)	0.669198 ** (2.794)
MSTAT	1.31668 ** (2.077)	-1.48382 (1.256)
MWAGE	0.799076-3 (0.528)	0.283366-3 (0.229)
NETY	-0.389682 ** (12.780)	0.319443-1 ** (2.434)
PSIBS	0.312427 (1.359)	0.625861 ** (3.144)
PWAGE	-0.820000-3 ** (6.753)	-0.575095-3 ** (4.567)
TSCORE	0.336329-1 ** (5.548)	0.315419-1 ** (5.827)
VEAP	0.281139-3 (0.734)	-0.217295-3 (0.653)

Percentage Prediction Improvement = 38.62 $\rho^2 = 0.369$

Note: 1. Coefficient estimates for the civilian sector outcome are normalized to zero.

2. Absolute value T-Statistics are in parens below the coefficient estimates.

* - Statistically significant at the 10% level.

** - Statistically significant at the 5% level.

Table 4-10: Model 2 Structural Selection Equation Estimates

\- Uncorrected -

Variable	Military Service	College
Constant	-22.1435 **	-8.34291 **
	(5.313)	(2.807)
ADDT	0.726856 **	-0.298373 *
	(2.815)	(1.909)
BLACK	-0.618976 **	0.472092 **
	(2.187)	(2.166)
COLPROG	-0.134858	0.757706 **
	(0.547)	(4.803)
COSTO	-0.599014-3	-0.927156-3
	(0.780)	(1.529)
DADMIL	1.08635 *	-0.863048
	(1.877)	(1.271)
EDAD	0.126641 **	0.136026 **
	(5.345)	(7.344)
EMOM	0.206168-1	-0.787499-2
	(0.494)	(0.257)
HEXP	0.586438-1	0.460146 **
	(1.141)	(11.447)
HISP	-0.739603 **	0.702798 **
	(2.050)	(3.046)
MSTAT	-0.287598	-1.26477
	(0.474)	(1.128)
MWAGE	0.629014-2 **	-0.503995-3
	(5.497)	(0.596)
NETY	-0.403455 **	0.301023-1 **
	(11.522)	(2.373)
PSIBS	0.302493	0.690790 **
	(1.145)	(3.541)
PWAGE	-0.102891-2 **	-0.577511-3 **
	(6.923)	(4.480)
TSCORE	0.317565-1 **	0.324018-1 **
	(4.629)	(6.166)
UNRATEO	-0.126893	-0.325911-2
	(1.499)	(0.056)
VEAP	-0.123020-2 **	-0.178300-3
	(2.980)	(0.554)
Percentage Prediction Improvement	= 41.46	ρ^2 = 0.389

Note: 1. Coefficents estimates for the civilian sector outcome are normalized to zero.

2. Absolute value T-Statistics are in parens below the coefficient estimates.

* - Statistically significant at the 10% level.

** - Statistically significant at the 5% level.

Table 4-11: Model 2 Structural Selection Equation Estimates

\- Self-Selectivity Bias Corrected -

Variable	Military Service	College
Constant	-20.8023 **	-7.81013 **
	(5.012)	(2.637)
ADDT	0.712582 **	-0.289412 *
	(2.766)	(1.849)
BLACK	-0.508642 *	0.519647 **
	(1.820)	(2.408)
COLPROG	-0.174319	0.722271 **
	(0.707)	(4.555)
COSTO	-0.523323-3	-0.880970-3
	(0.680)	(1.452)
DADMIL	1.03736 *	-0.842455
	(1.775)	(1.248)
EDAD	0.114048 **	0.128157 **
	(4.821)	(6.953)
EMOM	0.183812-1	-0.797328-2
	(0.441)	(0.260)
HEXP	0.325491-1	0.439058 **
	(0.628)	(10.792)
HISP	-0.687272 *	0.709004 **
	(1.911)	(3.071)
MSTAT	-0.259200	-1.26588
	(0.425)	(1.130)
MWAGE	0.590805-2 **	-0.603038-3
	(5.206)	(0.716)
NETY	-0.386233 **	0.309105-1 **
	(11.170)	(2.429)
PSIBS	0.245829	0.648417 **
	(0.933)	(3.338)
PWAGE	-0.958256-3 **	-0.539077-3 **
	(6.675)	(4.286)
TSCORE	0.324360-1 **	0.327572-1 **
	(4.724)	(6.221)
UNRATEO	-0.117555	-0.438002-4
	(1.394)	(0.001)
VEAP	-0.112983-2 **	-0.143597-3
	(2.746)	(0.447)

Percentage Prediction Improvement = 41.46 $\rho^2 = 0.388$

Note: 1. Coefficient estimates for the civilian sector outcome are normalized to zero.

2. Absolute value T-Statistics are in parens below the coefficient estimates.

* - Statistically significant at the 10% level.

** - Statistically significant at the 5% level.

Table 4-12: Model 2 Structural Selection Equation Estimates

- Self-Selectivity and Choice-Base Bias Corrected -

Variable	Military Service	College
Constant	-10.5184 **	-7.92580 **
	(2.780)	(2.600)
ADDT	0.570723 **	-0.273677
	(2.484)	(1.699)
BLACK	0.150456-1	0.442397 **
	(0.058)	(2.002)
COLPROG	-0.160269	0.698897 **
	(0.747)	(4.296)
COSTO	-0.423000-3	-0.840420-3
	(0.622)	(1.356)
DADMIL	1.47527 **	-0.828383
	(2.691)	(1.184)
EDAD	0.105206 **	0.125864 **
	(4.905)	(6.551)
EMOM	0.760456-2	-0.126145-1
	(0.206)	(0.397)
HEXP	-0.360713-2	0.440847 **
	(0.078)	(10.590)
HISP	-0.487396	0.634183 **
	(1.562)	(2.675)
MSTAT	0.787564	-1.23394
	(1.583)	(1.084)
MWAGE	0.230910-2 **	-0.484779-3
	(2.260)	(0.559)
NETY	-0.387795 **	0.316203-1 **
	(12.730)	(2.408)
PSIBS	0.317538	0.620372 **
	(1.381)	(3.118)
PWAGE	-0.820118-3 **	-0.571561-3 **
	(6.671)	(4.500)
TSCORE	0.328057-1 **	0.317076-1 **
	(5.352)	(5.849)
UNRATEO	-0.965150-1	-0.122473-1
	(1.280)	(0.205)
VEAP	0.972177-4	-0.149452-3
	(0.254)	(0.451)

Percentage Predict Improvement = 38.40 ρ^2 = 0.368

Note: 1. Coefficient estimates for the civilian sector outcome are normalized to zero.

2. Absolute value T-Statistics are in parens below the coefficient estimates.

* - Statistically significant at the 10% level.

** - Statistically significant at the 5% level.

Table 4-13: Model 3 Structural Selection Equation Estimates

- Uncorrected -

Variable	Military Service	College
Constant	-13.2328 **	-11.1683 **
	(2.191)	(2.718)
ADDT	0.724423 **	-0.302844 *
	(2.792)	(1.940)
BLACK	0.556529 *	0.345484
	(1.957)	(1.607)
COLPROC	-0.143217	0.763671 **
	(0.576)	(4.835)
COSTO	-0.826824-3	-0.853054-3
	(1.157)	(1.534)
DADMIL	1.07404 *	-0.895842
	(1.823)	(1.306)
DUNRATE	-0.105537	0.709493-1
	(0.799)	(0.828)
EDAD	0.122989 **	0.138198 **
	(5.150)	(7.424)
EMOM	0.181531-1	-0.854999-2
	(0.431)	(0.277)
HEXP	0.581371-1	0.461520 **
	(1.126)	(11.463)
HISP	-0.281555	0.662040 **
	(0.759)	(2.833)
MSTAT	1.12882	-1.55980
	(1.471)	(1.345)
MWAGE	0.229281-2	0.402248-3
	(1.317)	(0.332)
PSIBS	0.310618	0.697976 **
	(1.171)	(3.570)
PWAGE	-0.100639-2 **	-0.586358-3 **
	(6.800)	(4.585)
STDY	-5.99300 **	0.417956 **
	(11.724)	(2.313)
TSCORE	0.329313-1 **	0.319933-1 **
	(4.753)	(6.095)
VEAP	-0.826315-3 **	-0.274473-3
	(1.962)	(0.828)

Percentage Prediction Improvement = 41.24 ρ^2 = 0.394

Note: 1. Coefficient estimates for the civilian sector outcome are normalized to zero.

2. Absolute value T-Statistics are in parens below the coefficient estimates.

* - Statistically significant at the 10% level.

** - Statistically significant at the 5% level.

Table 4-14: Model 3 Structural Selection Equation Estimates

- Self-Selectivity Bias Corrected -

Variable	Military Service	College
Constant	-12.0815 **	-10.4475 **
	(2.004)	(2.554)
ADDT	0.709623 **	-0.293220 *
	(2.740)	(1.877)
BLACK	0.607595 **	0.389628 *
	(2.146)	(1.829)
COLPROG	-0.183708	0.725471 **
	(0.739)	(4.570)
COSTO	-0.739406-3	-0.802235-3
	(1.035)	(1.443)
DADMIL	1.02065 *	-0.870421
	(1.716)	(1.278)
DUNRATE	-0.108556	0.644545-1
	(0.823)	(0.752)
EDAD	0.110203 **	0.129757 **
	(4.614)	(7.010)
EMOM	0.159390-1	-0.866002-2
	(0.379)	(0.281)
HEXP	0.316265-1	0.439119 **
	(0.607)	(10.772)
HISP	-0.253622	0.662343 **
	(0.685)	(2.832)
MSTAT	1.10629	-1.55047
	(1.441)	(1.339)
MWAGE	0.203955-2	0.263716-3
	(1.173)	(0.218)
PSIBS	0.252815	0.653175 **
	(0.956)	(3.356)
PWAGE	-0.941660-3 **	-0.547163-3 **
	(6.579)	(4.391)
STDY	-5.73335 **	0.431405 **
	(11.357)	(2.380)
TSCORE	0.335860-1 **	0.323584-1 **
	(4.848)	(6.152)
VEAP	-0.741411-3 *	-0.236876-3
	(1.763)	(0.732)

Percentage Prediction Improvement = 41.35 $\rho^2 = 0.392$

Note: 1. Coefficient estimates for the civilian sector outcome are normalized to zero.

2. Absolute value T-Statistics are in parens below the coefficient estimates.

* - Statistically significant at the 10% level.

** - Statistically significant at the 5% level.

Table 4-15: Model 3 Structural Selection Equation Estimates

- Self-Selectivity and Choice-Base Bias Corrected -

Variable	Military Service	College
Constant	0.742900	-11.2229 **
	(0.139)	(2.672)
ADDT	0.581110 **	-0.275268
	(2.522)	(1.712)
BLACK	1.09632 **	0.312627
	(4.257)	(1.438)
COLPROG	-0.167890	0.703831 **
	(0.776)	(4.323)
COSTO	-0.553631-3	-0.825122-3
	(0.870)	(1.456)
DADMIL	1.43900 **	-0.846173
	(2.557)	(1.194)
DUNRATE	-0.172026	0.758783-1
	(1.518)	(0.859)
EDAD	0.100439 **	0.127750 **
	(4.650)	(6.621)
EMOM	0.635180-2	-0.134537-1
	(0.170)	(0.422)
HEXP	-0.274713-2	0.441327 **
	(0.059)	(10.590)
HISP	-0.101911	0.586808 **
	(0.317)	(2.443)
MSTAT	2.39116 **	-1.57231
	(3.737)	(1.331)
MWAGE	-0.225117-2	0.544877-3
	(1.481)	(0.439)
PSIBS	0.314276	0.628629 **
	(1.360)	(3.155)
PWAGE	-0.807598-3 **	-0.575455-3 **
	(6.607)	(4.564)
STDY	-5.70542 **	0.444971 **
	(12.690)	(2.381)
TSCORE	0.338038-1 **	0.313222-1 **
	(5.509)	(5.782)
VEAP	0.518737-3	-0.241049-3
	(1.341)	(0.723)

Percentage Prediction Improvement = 39.38 $\rho^2 = 0.373$

Note: 1. Coefficient estimates for the civilian sector outcome are normalized to zero.
2. Absolute value T-Statistics are in parens below the coefficient estimates.

* - Statistically significant at the 10% level.

** - Statistically significant at the 5% level.

Table 4-16: Model 4 Structural Selection Equation Estimates

- Uncorrected -

Variable	Military Service	College
Constant	-14.7740 **	-8.86009 **
	(3.523)	(2.977)
ADDT	0.727187 **	-0.300063 *
	(2.801)	(1.919)
BLACK	0.544691 *	0.342206
	(1.911)	(1.588)
COLPROG	-0.144382	0.757266 **
	(0.580)	(4.799)
COSTO	-0.565969-3	-0.929478-3
	(0.729)	(1.533)
DADMIL	1.10086 *	-0.907610
	(1.880)	(1.329)
EDAD	0.124867 **	0.136855 **
	(5.236)	(7.393)
EMOM	0.218158-1	-0.810137-2
	(0.517)	(0.264)
HEXP	0.610070-1	0.461024 **
	(1.180)	(11.457)
HISP	-0.228123	0.626786 **
	(0.621)	(2.709)
MSTAT	0.878203	-1.34113
	(1.422)	(1.197)
MWAGE	0.298227-2 **	-0.268185-3
	(2.586)	(0.316)
PSIBS	0.307089	0.693813 **
	(1.157)	(3.554)
PWAGE	-0.101759-2 **	-0.577449-3 **
	(6.807)	(4.477)
STDY	-5.97386 **	0.411577 **
	(11.724)	(2.277)
TSCORE	0.320639-1 **	0.321864-1 **
	(4.624)	(6.122)
UNRATEO	-0.123264	-0.180087-2
	(1.447)	(0.031)
VEAP	-0.968249-3 **	-0.198252-3
	(2.320)	(0.615)

Percentage Prediction Improvement = 41.24 ρ^2 = 0.393

Note: 1. Coefficient estimates for the civilian sector outcome are normalized to zero.

2. Absolute value T-Statistics are in parens below the coefficient estimates.

* - Statistically significant at the 10% level.

** - Statistically significant at the 5% level.

Table 4-17: Model 4 Structural Selection Equation Estimates

- Self-Selectivity Bias Corrected -

Variable	Military Service	College
Constant	-13.7266 **	-8.34443 **
	(3.283)	(2.813)
ADDT	0.712609 **	-0.290821 *
	(2.750)	(1.859)
BLACK	0.598052 **	0.386398 *
	(2.108)	(1.810)
COLPROG	-0.184364	0.721044 **
	(0.740)	(4.454)
COSTO	-0.492853-3	-0.883374-3
	(0.635)	(1.456)
DADMIL	1.04804 *	-0.882262
	(1.773)	(1.300)
EDAD	0.112256 **	0.128840 **
	(4.711)	(6.996)
EMOM	0.194256-1	-0.822769-2
	(0.461)	(0.268)
HEXP	0.347476-1	0.439581 **
	(0.666)	(10.791)
HISP	-0.198488	0.630173 **
	(0.542)	(2.721)
MSTAT	0.857486	-1.34789
	(1.386)	(1.204)
MWAGE	0.273730-2 **	-0.358063-3
	(2.384)	(0.423)
PSIBS	0.250479	0.650842 **
	(0.946)	(3.348)
PWAGE	-0.950737-3 **	-0.538478-3 **
	(6.582)	(4.280)
STDY	-5.71802 **	0.424705 **
	(11.357)	(2.342)
TSCORE	0.327556-1 **	0.325379-1 **
	(4.719)	(6.177)
UNRATEO	-0.114047	0.153245-2
	(1.344)	(0.026)
VEAP	-0.878870-3 **	-0.164558-3
	(2.111)	(0.512)

Percentage Prediction Improvement = 41.13 ρ^2 = 0.392

Note: 1. Coefficient estimates for the civilian sector outcome are normalized to zero.

2. Absolute value T-Statistics are in parens below the coefficient estimates.

* - Statistically significant at the 10% level.

** - Statistically significant at the 5% level.

Table 4-18: Model 4 Structural Selection Equation Estimates

- Self-Selectivity and Choice-Base Bias Corrected -

Variable	Military Service	College
Constant	-3.67492	-8.48389 **
	(0.972)	(2.778)
ADDT	0.570568 **	-0.275197 *
	(2.471)	(1.708)
BLACK	1.10519 **	0.306159
	(4.261)	(1.405)
COLPROG	-0.159364	0.697590 **
	(0.735)	(4.286)
COSTO	-0.372954-3	-0.846264-3
	(0.544)	(1.366)
DADMIL	1.47538 **	-0.869041
	(2.645)	(1.234)
EDAD	0.103740 **	0.126579 **
	(4.805)	(6.596)
EMOM	0.985259-2	-0.130944-1
	(0.263)	(0.412)
HEXP	-0.135780-2	0.441579 **
	(0.029)	(10.600)
HISP	-0.546610-2	0.552179 **
	(0.017)	(2.322)
MSTAT	1.88341 **	-1.32012
	(3.739)	(1.160)
MWAGE	-0.791204-3	-0.230564-3
	(0.775)	(0.264)
PSIBS	0.318088	0.623326 **
	(1.376)	(3.130)
PWAGE	-0.807658-3 **	-0.571515-3 **
	(6.520)	(4.494)
STDY	-5.68010 **	0.439906 **
	(12.650)	(2.352)
TSCORE	0.329660-1 **	0.315031-1 **
	(5.313)	(5.808)
UNRATEO	-0.942853-1	-0.107758-1
	(1.245)	(0.181)
VEAP	0.340800-3	-0.171258-3
	(0.882)	(0.516)

Percentage Prediction Improvement = 38.73 $\rho^2 = 0.372$

Note: 1. Coefficient estimates for the civilian sector outcome are normalized to zero.
2. Absolute value T-Statistics are in parens below the coefficient estimates.
* - Statistically significant at the 10% level.
** - Statistically significant at the 5% level.

3.1 A Test for the Independence of Irrelevant Alternatives

As stated in Chapter 2, one of the potential problems with logit estimations is the Independence of Irrelevant Alternatives (IIA) property. This property could lead to unrealistic predictions when additional alternatives are added (or subtracted) from the choice set. A simple test of this property has been devised by Hausmann and McFadden (1984).[21] This test consists of estimating the model with the full set of possible outcomes (K), reestimating with (K-1) outcomes and constructing a test statistic based on the difference in the structural model coefficients and variance-covariance matrices. This test statistic is represented as

$$(4.3) \qquad Q = N^*[\beta_r - \beta_u]'[V_r - V_u]^{-1}[\beta_r - \beta_u],$$

where β_r and V_r are the estimated parameter vector and variance-covariance matrix for K-1 possible outcomes (restricted) model and β_u and V_u are the estimated parameter vector and variance covariance matrix for the K possible outcomes (unrestricted) model. The test statistic Q is distributed χ^2 (N).

For the IIA test, the college enrollment option was omitted. The calculation of the test statistic did not yield actual values for Q. This was because the differences in the elements of V_r and V_u were very small and the attempt to invert the matrix of these differences was not successful.[22] However, it can be inferred that if the inversion was possible, the resulting Q statistic would be close to zero.[23] Therefore, the null hypothesis (the existence of IIA) cannot be rejected.

The implication of failing to reject the presence of IIA is not necessarily detrimental. This failure simply indicates that the probability of entering the military, relative to staying in the civilian labor market, is independent of the decision to enroll in college. The above structural model estimates for the "occupation specific" variables tend to support this test failure.

4. Interpretation of Structural Selection Equation Estimates

This section presents two more standard methods of interpreting the selection equation estimates. The first method involves calculation of elasticities. The second consists of constructing individual "profiles". These profiles will allow the evaluation of the effects of the binary (dummy) variables on the choice of occupation.

4.1 Elasticity Estimates

Unlike elasticities calculated from linear coefficient models, multinomial qualitative response model elasticities are a direct function of where the individual is on the cumulative distribution function. It is intuitively apparent that the effect of a given X_i

should be larger on the individual who is on the margin between two (or more) outcomes compared to the individual who is not. To correctly account for this effect, the elasticity is estimated as

$$(4.4) \qquad E_{mj} = \delta_{jm}(1-\overline{P}_m)\overline{X}_j - \sum_{\substack{k=0,\\ k\neq m}}^{K} \delta_{jk}\overline{P}_k\overline{X}_j,$$

where E_{mj} = the elasticity of probability m with respect to independent variable X_j, δ_{jk} = the estimated coefficient of X_j in equation k, and P_k = the estimated probability for outcome k. The second term in this equation is necessary to take into account the cross effects of X_j on the other alternatives. Note that if the coefficients of one outcome are normalized to zero, as in the present analysis, this cross effect term is summed over [K-2] outcomes. A derivation of this expression is found in Appendix C.

The following four tables present elasticities calculated for the self-selectivity and choice-based bias-corrected versions of Models 1 through 4 above. These elasticities were calculated at the mean values of the X's and the estimated probabilities. Elasticity estimates are reported only for the continuous variables, since elasticities estimates are not relevant for binary variables.

As seen in these tables, only the elasticities for EDAD, NETY, PWAGE and TSCORE are based on coefficient estimates that are statistically significant in both the military and college equations. Therefore, the following discussion will focus on these elasticity estimates.

The largest estimated difference in the effects is for family income (NETY). Not only is there a very elastic negative effect on the military (Models 1 and 2), but there is also an asymmetric college effect. For illustrative purposes, the military family income elasticity shows that, at the sample average, an increase in real net family income of $1,000 (11.5%) will decrease the probability of enlisting from approximately 9% (the estimated sample average) to 5.6% (a decrease of 38%). For college, the same increase in income will increase the probability of going to college from 36.7% to 38.8% (an increase of 5.7%). A possible explanation for this asymmetrical effect is that the model does not control for total family wealth. The total family wealth of individuals who go to college most likely exceeds that of individuals in the other occupational groups. If so, the college elasticities underestimate the true family wealth effect. Further, if individuals who enter the military come from families with little or no additional assets, the estimated family income effects for the military more accurately reflect the true effects.

The estimated predicted civilian wage elasticities are correctly signed, but inelastic for both occupations. The estimates indicate that the decision to enlist in the military is more sensitive to changes in the civilian wage than is the decision to enroll in college. This result is consistent with the hypothesized (unobserved) family wealth effect.

The estimated elasticities for the "ability" measure TSCORE indicate that the decision to join the military or enter college is responsive to the individual's "ability". The slightly larger elasticity estimates for military enlistments is not sufficiently different from those for college enrollment to warrant further explanation.

Two final points should be noted. First, the decision to enter college is highly elastic with respect to the highest expected grade completed (HEXP). This finding indicates that this variable is a very significant measure of the individual's propensity toward pursuing post-secondary education. Secondly, the estimated military wage elasticities are highly unstable. The most probable reason for this result is that there is not sufficient variation in this measure to correctly estimate its effect.

A standard procedure would be to now compare the above elasticity estimates to those in earlier studies. In the present situation it is difficult to make this kind of comparison, for several reasons. First, the variables used in this analysis differ considerably from those in most of the studies reviewed in Chapter 1 (see Table 1-2). Second, the key continuous variables used in these earlier studies (the unemployment rate and military wages) were found to be statistically insignificant in the present analysis. Further, most of the previous analyses reported wage elasticities based on the estimated civilian to military or military to civilian wage ratio. Therefore, a comparison of the estimated civilian wage elasticities (which were statistically significant) to these wage ratio elasticities would be difficult.

The only study where a direct comparison can be made is that of Dertouzos (1984). Using cross-sectional Army data for 1980 and 1981 (in separate equations) Dertouzos estimated civilian wage elasticities that ranged from -.81 to -.165. These estimates are entirely consistent with the above estimated elasticity range of -.76 to -.74.

Table 4-19: Model 1 Estimated Elasticities

Variable	Military	College
COSTO	-0.095	-0.175
DUNRATE	-0.012	0.004
EDAD	0.467 **	0.724 **
EMOM	0.098	-0.098
HEXP	-2.261	3.999 **
MWAGE	2.261	0.384
NETY	-3.292 **	0.494 **
PSIBS	0.030	0.205 **
PWAGE	-0.758 **	-0.411 **
TSCORE	1.343 **	1.195 **
VEAP	0.708	-0.344

* - Indicates that the coefficient estimate was statistically significant at the 10% level.
** - Indicates that the coefficient estimate was statistically significant at the 5% level.

Table 4-20: Model 2 Estimated Elasticities

Variable	Military	College
COSTO	-0.029	-0.186
EDAD	0.503 **	0.713 **
EMOM	0.131	-0.098
HEXP	-2.364	4.001 **
MWAGE	8.267 *	-1.867
NETY	-3.273 **	0.493 **
PSIBS	0.034	0.203 **
PWAGE	-0.759 **	-0.408 **
TSCORE	1.284 **	1.206 **
UNRATEO	-0.558	-0.006
VEAP	0.303	-0.182

* - Indicates that the coefficient estimate was statistically significant at the 10% level.
** - Indicates that the coefficient estimate was statistically significant at the 5% level.

Table 4-21: Model 3 Estimated Elasticities

Variable	Military	College
COSTO	-0.076	-0.178
DUNRATE	-0.008	0.004
EDAD	0.452 **	0.739 **
EMOM	0.123	-0.103
HEXP	-2.355	4.004 **
MWAGE	-8.158	0.516
PSIBS	0.031	0.206 **
PWAGE	-0.745 **	-0.412 **
STDY	-0.104 **	0.016 **
TSCORE	1.357 **	1.183 **
VEAP	1.186	-0.421

* - Indicates that the coefficient estimate was statistically significant at the 10% level.
** - Indicates that the coefficient estimate was statistically significant at the 5% level

Table 4-22: Model 4 Estimated Elasticities

Variable	Military	College
COSTO	-0.011	-0.189
EDAD	0.487 **	0.719 **
EMOM	0.647	-0.104
HEXP	-2.338	4.005 **
MWAGE	-2.305	-0.271
PSIBS	0.034	0.204 **
PWAGE	-0.743 **	-0.409 **
STDY	-0.097 **	0.015 **
TSCORE	1.299 **	1.196 **
UNRATEO	-0.508	0.097
VEAP	0.789	-0.294

* - Indicates that the coefficient estimate was statistically significant at the 10% level.
** - Indicates that the coefficient estimate was statistically significant at the 5% level.

4.2 Profiles

A second method of interpreting the empirical estimates is to develop "composite" individuals that have a higher than average propensity to choose a particular outcome. The primary purpose for these profiles is to examine the effects of the binary type variables on the choice of occupation. The predicted probabilities in Tables 4-23 to 4-26 were evaluated at the sample mean values of the continuous variables, for the self-selectivity/choice-based bias-corrected version of Model 1 only. The values in these tables should not be interpreted as the predicted probabilities of individuals with these binary characteristics. Instead, they should be interpreted as the instantaneous effects of these variables on the choice of occupation.

These profiles could be used to develop a policy instrument. For example, if the military is trying to focus recruiting resources on a higher success rate group, such a tool could serve as a screening device. Construction of this tool requires the estimation of the total effects of these binary variables. To estimate these total effects, the probabilites must be evaluated at the mean values of the continuous variables, for those individuals with the observed binary variable attribute(s). Again, these total effects are not presented, as the purpose of this section is to isolate the effects of these binary variables.

The values in these tables indicate that the highest probability of enlistment is caused by the interaction of the person being white, married, and having a father in the military. The lowest enlistment probability is for Hispanic singles in a college preparatory program.

The largest dummy variable effect is for marital status. On average, married individuals have approximately a 5 times larger probability of joining the military, in comparison to single individuals. This effect is consistent across all ethnic/racial groups and other dummy variable effects. The effect on the college enrollment decision varies between ethnic/racial groups. The largest effect is for Whites (3.2 - 5.3 times lower) and the lowest is for Hispanics (2.9 - 4.2 times lower).

After marital status, the most important estimated impact is for the individual's father being in the military. This increases the probability of a military enlistment 4.6 (for Whites) - 5.6 (for Hispanics) times for singles and 3.2 (for Whites) - 3.8 (for Hispanics) times for married individuals. The effect on college enrollment is slightly lower. The largest effect is for married Blacks. Married Blacks with fathers in the military have approximately a 4.3 times lower probability of going to college. Married Whites and Hispanics have 2.9 and 2.5 times lower probabilities, respectively. For single individuals, the estimated probabilities are 1.7 (for Hispanics) - 2.0 (for Whites) times lower.

The effects of desiring additional training and college program participation on the predicted probabilities are not as substantial. As the structural equation estimates indicated, desiring additional training increases the military enlistment probability and decreases the college enrollment probability. For singles, the reduction in the college enrollment probability is split between increases in the military and civilian probabilities. For married individuals, this reduction appears to be almost entirely transferred to increases in the military probability. Further, increases in the military probability also came from reductions in the civilian sector probability. This result is interpreted as indicating that married individuals consider the military as a preferable means of obtaining additional job training.

In terms of absolute changes in predicted probabilities, the effects of college program participation are greatest for single individuals. For individuals within this group, most of the increase in predicted college probability (approximately 95%) comes from reductions in the civilian sector employment probability. However, for married individuals, approximately 71% of the increased college probability comes from decreases in the civilian sector probability. This indicates that the effects of program participation on the military enlistment probability are greatest for married individuals. Or, college program participation has little relative effect on the probability of single individuals entering the military.

Table 4-23: Occupational Choice Probabilities by Selected Characteristics: Total Sample Averages

Marital Status	Occupation	Ethnic/Racial Group			
		Black	Hisp	White	Total Sample
Single	Civilian	0.577	0.523	0.676	0.631
	Military	0.021	0.010	0.024	0.021
	College	0.402	0.467	0.300	0.348
Married	Civilian	0.774	0.783	0.810	0.801
	Military	0.104	0.058	0.108	0.099
	College	0.122	0.159	0.082	0.100

Table 4-24: Occupational Choice Probabilities by Selected Characteristics: Participated in College Program

Marital Status	Occupation	Ethnic/Racial Group			
		Black	Hisp	White	Total Sample
Single	Civilian	0.462	0.405	0.570	0.520
	Military	0.015	0.007	0.018	0.015
	College	0.523	0.588	0.412	0.465
Married	Civilian	0.727	0.717	0.780	0.762
	Military	0.086	0.047	0.093	0.083
	College	0.187	0.236	0.128	0.155

Table 4-25: Occupational Choice Probabilities by Selected Characteristics: Desires Additional Training

Marital Status	Occupation	Ethnic/Racial Group			
		Black	Hisp	White	Total Sample
Single	Civilian	0.593	0.541	0.689	0.793
	Military	0.025	0.013	0.029	0.025
	College	0.382	0.446	0.282	0.329
Married	Civilian	0.767	0.784	0.799	0.793
	Military	0.121	0.069	0.126	0.115
	College	0.112	0.147	0.075	0.091

Table 4-26: Occupational Choice Probabilities by Selected Characteristics: Father in the Military

Marital Status	Occupation	Ethnic/Racial Group			
		Black	Hisp	White	Total Sample
Single	Civilian	0.683	0.673	0.742	0.722
	Military	0.102	0.056	0.110	0.099
	College	0.215	0.271	0.148	0.179
Married	Civilian	0.615	0.715	0.626	0.638
	Military	0.341	0.220	0.346	0.326
	College	0.044	0.065	0.028	0.036

5. Summary

In sum, all of the model estimates presented in this chapter appear to tell a similar story. Those who enter the military are significantly different from those who enter college. The estimates indicate that military enlistees are not affected by educational benefits and longer term formal educational expectations. The ability of the military to provide short-run technical training is important. Net family income (and unobserved family wealth) and predicted civilian wage have a strong negative effect on the decision to enter the military. In addition, the military experience of the individual's father and marital status have a considerable positive effect on the decision to enlist.

Those who enter college appear to be dominated by longer term considerations. Short term economic attrubutes, while somewhat important, do not seem to be as dominant as for those who enter the military. Rather, long run educational expectations appear to be far more important for individuals entering college.

End Notes

1.) The choice based sampling weights were calculated by taking the ratio of the population percentage of the particular group relative to the sample percentage. The weights were equal to .343 and 1.072 for the military and civilian groups, respectively.

2.) Copywright 1985 by W. Greene, New York University.

3.) Manski and Lerman (1977) have shown that with a choice based sampling estimator the standard estimator of the variance-covariance matrix is not appropriate. The appropriate variance-covariance estimator is given as

$$V = A^{-1}HA^{-1} ,$$

where V is the correct variance-covariance estimator, H is the standard (weighted) variance-covariance matrix estimator and A is the unweighted variance-covariance estimator. The intuitive explanation for this correction is slightly subtle. The weighted matrix (H) will deflate the standard errors for the overrepresented group and inflate these errors for the underrepresented group. The product of $A^{-1}H$ will create a new weight that is the inverse of the sampling weights. This new weight will correctly inflate the estimated standard errors for the overrepresented group and deflate those for the underrepresented group. This is also discussed in Manski and Mcfadden (1981) and Greene (1985).

4.) Mr. C. Capps and Dr. W. Greene were instrumental in the software modifications.

5.) The Newton - Raphson algorithm was used for the maximum likelihood estimates.

6.) Using a VAX-11 computer, some of the model estimations used as much as 2 hours of CPU time. Needless to say, if a dollar cost was associated to this, the amount would be quite large.

7.) See Tables B-1 to B-4.

8.) Using the same basic data source (NLS), Daula, Fagan and Smith (1982) found even lower wage equation fits.

9.) See Table 2-2.

10.) See Welch (1967, 1973) for a discussion on these types of discrimination.

11.) Preliminary analysis indicated that entering the unemployment rate for the year of graduation and the prior year as separate variables, yielded opposite signed (but mainly statistically insignificant) coefficient estimates. This was interpreted as showing that it was the difference in the unemployment rate between years that mattered. Copies of these preliminary estimates are available from the author.

12.) This standardized variable was calculated as

$$STDY_{ij} = \frac{NETY_{ij} - \mu_{NETY_{ij}}}{\sigma_{NETY(j)}},$$

where $STDY_{ij}$ is the standardized net family income for individual i in ethnic/racial group j, $NETY_{ij}$ is the reported net family income, $\mu_{NETY(ij)}$ is the mean net family income in ethnic/racial group j, and $\sigma_{NETY(j)}$ is the standard deviation of net family income in ethnic/racial group j.

13.) See tables B-5 to B-16 in Appendix B.

14.) See the differences in the estimates of EDAD in Tables B-5 to B-10 (that include the sibling education variable) and Tables B-11 to B-16 (that exclude it).

15.) A good discussion on goodness of fit measures for qualitative response can be found in Amemiya (1981).

16.) The idea for this measure came originally from a discussion with Dr. Frank Levy, University of Maryland, College Park. The final formulation was developed with the assistance of Mr. Roy Nord, U.S. Army Research Institute, Alexandria, VA.

17.) This statistic would be less meaningful for models where the magnitude of the "miss" matters (ie - an ordered probit or logit model).

18.) See McFadden (1974).

19.) See Domencich and McFadden (1975), Figure 5.5, pp. 124.

20.) A discussion of additional benefits to military dependents that cannot be easily quantified can be found in the U.S. General Accounting Report GAO/NSIAD-84-41 (1984).

21.) See equation 1.16, pp. 1225.

22.) The differences in V_r and V_u averaged approximately 0.1E-6.

23.) It is not difficult to show that as $[V_r - V_u] \rightarrow 0$, $[V_r - V_u]^{-1} \rightarrow \infty$ and $Q \rightarrow 0$.

Chapter 5

Conclusions and Extensions

1. Introduction

After reading through the wealth of estimates in the previous chapter, the reader may have a somewhat fuzzy picture of what we have learned. The primary purpose of this chapter is to bring this picture into focus by summarizing the major empirical findings. The following contains a discussion of these findings. The last section will suggest avenues of future research in this area.

2. Major Findings

Two major findings emerged from this analysis. The first concerns the actual empirical evidence. The second is in regard to the econometric technique.

2.1 Empirical Findings

The empirical estimates tell a consistent story on the key determinants of initial occupational choice. There appear to be two, not necessarily independent, determinants of whether the individual decides to enlist in the military or go to college, relative to entering the civilian labor market. These are:

A.) The financial status of the individual's family: The decision to enter the military is far more sensitive to net family income and predicted civilian wages than is the decision to enter college. The strength of this result, though, needs to be qualified. The accuracy of the family income variable is somewhat suspect. As mentioned in Chapter 3, there is a very large difference in reported family incomes across the occupational groups. It is possible that reported family income for those in the military is not that of the parents. Rather, it could be that of the individual alone. Given the nature of the data used, however, it was not possible to determine if this was so. In addition, a value for family income had to be imputed for approximately 25% of the sample (see Appendix A). Taken together, these two qualifications reduce the strength of this finding.

Financial responsibility, measured by marital status, does support the existence of this relationship. The statistically insignificant estimates for military wages and unemployment rates are largely attributed to insufficient variation in these variables.

B.) The long term expectations of the individual: Those who enter the military do not appear to be influenced by long term educational expectations. There was no statistically significant effect of VEAP, highest expected grade of completion or participation in a college preparatory program. The statistically significant effect of desiring additional training is interpreted as indicating that the individual is more interested in acquiring short term technical training instead of broader, longer term training that more typically characterizes a college education.

Individuals who enter college appear to be more affected by longer term expectations. The estimates of highest expected grade of completion and college preparatory program participation support this. This finding, however, also needs to be qualified.

Instead of reflecting the individual's planning horizon, this variable could also be interpreted as representing a preference and/or a propensity toward continued education. Further, it should be noted that these expectations are most likely determined by the family's financial status. A reasonable argument can be made that as family wealth increases, the ability to finance longer term activities also increases. However, to correctly account for this, data on total family wealth, not just reported net income, is required.

As mentioned above, the data used consisted of only high school graduates. Therefore, interpretation of the estimated racial/ethnic effects is restricted to individuals within this group. The estimates indicated that the largest difference in estimated racial/ethnic effects was for Hispanics, relative to the total sample. Hispanic high school graduates, on average, had the lowest predicted military enlistment probability and the highest predicted college enrollment probability. After controlling for choice based sampling bias and relative family income, the estimated instantenous difference in occupational choice for blacks and whites was not large.

For immediate policy formulation, this analysis does not provide many significant results. A major concern of military manpower analysts is how well the various monetary incentives offered by the military perform in attracting "high quality" recruits (i.e. - those who have scored in the upper 50% of the AFQT and are high school graduates). A typical procedure would be to use the model estimates to A.) make projections of aggregate population enlistments, and B.) perturbe the policy variables to determine the effects on these aggregate enlistments. As the estimated effects of these monetary incentive variables were not statistically significant, projections based on these estimated coefficients would be very tenuous. Therefore, this type of analysis is not reported. It should be noted that the lack of significance of these monetary variables should not be interpreted as indicating that the decision to enter the military is exogenous to these incentives. Rather, this lack of significance is attributed to insufficient variation in these attributes.

A potentially important policy tool that did come out of this analysis is the use of individual "composites" for the allocation of recruiting resources. The profile tables presented in the previous chapter illustrated the instantaneous effects of various binary demographic variables on the occupational choice probabilities. A recruiter policy tool could be constructed by extending these instantaneous profiles to evaluate the total effects of these binary measures. These total effects can be found by evaluating the occupational probabilities at the average values of the continuous variables for individuals with the given binary variable(s) characteristic(s). The resulting "total" profile could then be used by recruiters to determine whether the individual with these binary variable characteristics has a high probability of enlistment.

For illustrative purposes only, an example of these total profiles is given in Table 5-1. Referring to this table, it can be seen that married blacks have the highest total effect probability of enlisting in the military and single whites have the lowest. Recruiters could use this information to more effectively allocate their time (and effort). As married blacks have the highest probability of enlisting, a potential gain in recruiting efficiency could be realized by shifting recruiting resources to the lower probability groups. The amount of resources shifted would be dependent upon a measure of the marginal productivity of recruiters across the various probability groups, which this analysis failed to successfully estimate.

This crude illustration should not be interpreted as a policy recommendation. Rather, it just demonstrates a possible use for probabilistic model estimates. Given better quality data, this type of screening device could be a useful tool for recruiters faced with serious resource constraints.

As a final point it should be noted that the estimates obtained in this analysis are relevant only to the initial occupational choice of the individuals in this sample. Therefore, these estimates should not be interpreted as indicating the individual's occupational status 2 or more years beyond graduating high school. A great amount of occupational switching most likely will occur. Further, predicting that an individual with a given set of attributes has a high probability of enrolling in college does not exclude this individual

Table 5-1: Occupational Choice Probabilities by Selected Characteristics - Total Effect Estimates

Marital Status	Occupation	Ethnic/Racial Group			
		Black	Hispanic	White	Total Sample
Single	Civilian	0.532	0.557	0.540	0.540
	Military	0.121	0.050	0.077	0.085
	College	0.347	0.393	0.383	0.375
Married	Civilian	0.552	0.788	0.683	0.673
	Military	0.430	0.158	0.276	0.289
	College	0.018	0.054	0.041	0.038

from eventually entering the military. Rather, the individual may participate in a college ROTC program (which entails a military service commitment). However, controlling for this possibility was both beyond the scope of this analysis and the data presently available.

2.2 Econometric Issues

One aspect of this analysis was to determine the effects of self-selectivity and choice based sampling bias on the estimated occupational probability models. As stated in Chapter 2, the consistency of the structural model parameter estimates for the self-selectivity corrected models were not proven, but logically assumed. With this qualification in mind, the empirical estimates indicated that controlling for these biases did matter. The estimated effects of controlling for self-selectivity alone were not that substantial. The only structural model variable that was affected by this procedure was the predicted civilian wage.

The estimates generated after controlling for both self-selectivity and choice based sampling, however, were substantially different. This was most noticeable in the differences between the estimated black effects. The implication of this difference is rather obvious. When the data is not representative of the population, the possibility of large biases in the parameter estimates exists.

Therefore, when using data of this type, a great deal of caution must be exercised in the model estimation.

3. Extensions

As with most empirical analyses, the questions raised exceed the questions answered. This analysis has demonstrated that the application of self-selectivity and choice-based sampling bias correction procedure in a polychotomous choice logit model is not only possible, but important in order to estimate occupational choice correctly. Extending the analysis is only constrained by data limitations. If the data were available, the following extensions could be pursued.

A.) Three Choice Model

i.) Analysis of the effects of family wealth on occupational choice. It is argued above that the asymmetrical family income estimates are due to unobserved family wealth. To correctly determine the effects of family wealth several additional pieces of information are required. First, a clean estimate of parental income is desired. As mentioned above, the precision of the net family income variable used in this study is questionable. In addition, it would be desirable to know the financial assets of the parents. This could be determined by data on reported interest income and large assets (i.e. - home ownership). Finally, it would be desirable to know if the individual considers himself/herself financially dependent on his/her parents. By incorporating this type of data, a more precise estimate of the family wealth effect could be found.

ii.) Increase the number of years covered by the analysis. The estimated statistically insignificant effects of the military specific variables were most likely due to insufficient variation in these measures. Increasing the number of years will allow the effects of these variables to be more conclusively estimated.

iii.) Respecifiy the cost of college. The lower bound college cost proxy variable appears not to be a very good instrument for the true cost of college. A better instrument, for example, would control for the availability of financial aid and include additional educational costs (i.e. - commuting costs and/or room and board expenses).

B.) <u>Augmented Choice Model</u>

Expand the present structure to include more potential outcomes. In particular, the decision to enter the military should be disaggregated to the branch of service level. Further, by estimating a branch level model, the effects of interservice competition could be determined.

C.) Probit Analysis

The logit model constrained the ability of the model to capture the intercorrelations of the effects of the explanatory variables across the occupational choice set. It is difficult to determine the total effects of this constraint. Therefore, to validate the estimated relationships, a probit analysis could be utilized.

Appendix A

Net Family Income Imputations

Reported total net family income (TNFINC) contained a nontrivial number of missing values (25.92%). In order to maintain a reasonable sample size, impuited values were created for those observations with missing values. The imputation method employed consisted of regressing TNFINC on a set of exogeneous characteristics, for those observation with actual values for TNFINC. The estimated coefficents were then used to impute values for the observations with missing values. The follwing equation gives the list of regressors and their estimated coefficients. T-Statistics are reported below the coefficient estimates.

(A.1)

$$\begin{aligned} TNFINC = {} & \underset{(1.177)}{2292.54} - \underset{(0.717)}{939.12}(HISP) - \underset{(7.577)}{7462.64}(BLACK) \\ & - \underset{(5.981)}{1232.90}(D1) + \underset{(2.394)}{2208.18}(D2) + \underset{(1.651)}{1640.60}(D3) \\ & + \underset{(2.814)}{2784.13}(D4) + \underset{(5.411)}{499.30}(EDAD) + \underset{(7.624)}{1252.56}(EMOM), \end{aligned}$$

with $R^2 = 0.21$ and $F = 42.30$.

The right hand side variables in this equation are defined as:

HISP = 0/1 dummy variable for Hispanic (=1 if Hispanic),

BLACK = 0/1 dummy variable for Black (=1 if Black),

D1 = 0/1 dummy variable for military service (=1 if in),

D2 = 0/1 dummy variable for college enrollment (=1 if enrolled),

D3 = 0/1 dummy variable for 1979 high school graduate,

D4 = 0/1 dummy variable for 1980 high school graduate,

EDAD = Education of father (in years),

EMOM = Education of mother (in years).

Appendix B

This appendix contains the estimation results for the "Full" model specifications. The following tables are laid out in the identical format as the tables in Chapter 4. The T-Statistics are in parens below the estimated coefficients. The variables are the same as defined in Table 2-1.

Table B-1: Model 5 Civilian Wage Equation Estimates

Variable	Uncorrected	Corrected(1)	Corrected(2)
Constant	37.527	-32.816	-32.037
	(0.077)	(0.068)	(0.066)
BLACK	-812.827 **	-662.643 **	-686.935 **
	(2.979)	(2.353)	(2.412)
DUNRATE	-16.334	-23.765	-15.827
	(0.213)	(0.310)	(0.206)
HISP	-14.590	42.125	41.020
	(0.048)	(0.139)	(0.135)
LNHRS	887.782 **	850.838 **	862.588 **
	(12.434)	(11.589)	(11.780)
MILLS		-535.419 **	-413.642
		(2.094)	(1.529)
MSTAT	1589.160 **	1572.350 **	1555.110 **
	(2.485)	(2.463)	(2.432)
TPROG	260.703	224.396	233.337
	(1.100)	(0.946)	(0.982)
TSCORE	9.684 *	16.230 **	14.585 **
	(1.666)	(2.463)	(2.199)
VOCPROG	398.743	349.847	360.800
	(1.503)	(1.316)	(1.355)
R^2 (Adjusted)	0.198	0.201	0.199
F-Statistic	29.095	26.446	26.161

Table B-2: Model 6 Civilian Wage Equation Estimates

Variable	Uncorrected	Corrected(1)	Corrected(2)
Constant	807.731	707.724	706.814
	(1.201)	(1.052)	(1.047)
BLACK	-821.440 **	-672.311 **	-696.584 **
	(3.017)	(2.391)	(2.450)
HISP	-35.646	23.516	19.527
	(0.119)	(0.078)	(0.065)
LNHRS	881.531 **	845.905 **	857.050 **
	(12.351)	(11.540)	(11.719)
MILLS		-526.671 **	-409.429
		(2.066)	(1.515)
MSTAT	1537.510 **	1524.450 **	1506.020 **
	(2.406)	(2.390)	(2.357)
TPROG	274.636	236.266	245.950
	(1.161)	(0.998)	(1.037)
TSCORE	9.299	15.769 **	14.164 **
	(1.602)	(2.394)	(2.136)
UNRATEO	-114.670 *	-110.609	-109.899
	(1.662)	(1.606)	(1.593)
VOCPROG	415.373	367.894	376.910
	(1.560)	(1.387)	(1.418)
R^2 (Adjusted)	0.200	0.203	0.201
F-Statistic	29.522	26.811	26.535

Table B-3: Model 7 Civilian Wage Equation Estimates

Variable	Uncorrected	Corrected(1)	Corrected(2)
Constant	37.527	-65.707	-67.708
	(0.077)	(0.136)	(0.139)
BLACK	-812.827 **	-588.330 **	-615.982 **
	(2.979)	(2.071)	(2.135)
DUNRATE	-16.334	-27.301	-15.525
	(0.213)	(0.356)	(0.202)
HISP	-14.590	87.162	84.444
	(0.048)	(0.288)	(0.277)
LNHRS	887.782 **	831.833 **	847.294 **
	(12.434)	(11.234)	(11.463)
MILLS		-757.071 **	-609.106 **
		(2.728)	(2.067)
MSTAT	1589.160 **	1590.510 **	1544.590 **
	(2.485)	(2.496)	(2.418)
TPROG	260.703	222.251	230.313
	(1.100)	(0.939)	(0.971)
TSCORE	9.684 *	19.388 **	17.253 **
	(1.666)	(2.853)	(2.515)
VOCPROG	398.743	322.286	338.470
	(1.503)	(1.213)	(1.271)
R^2 (Adjusted)	0.198	0.203	0.200
F-Statistic	29.095	26.873	26.430

Table B-4: Model 8 Civilian Wage Equation Estimates

Variable	Uncorrected	Corrected(1)	Corrected(2)
Constant	807.731 (1.201)	664.945 (0.989)	657.430 (0.974)
BLACK	-821.440 ** (3.017)	-598.097 ** (2.108)	-625.603 ** (2.172)
HISP	-35.646 (0.119)	69.656 (0.231)	63.009 (0.208)
LNHRS	881.531 ** (12.351)	827.377 ** (11.194)	841.989 ** (11.408)
MILLS		-745.868 ** (2.696)	-604.553 ** (2.052)
MSTAT	1537.510 ** (2.406)	1544.060 ** (2.425)	1496.190 ** (2.344)
TPROG	274.636 (1.161)	233.433 (0.988)	242.561 (1.025)
TSCORE	9.299 (1.602)	18.908 ** (2.783)	16.835 ** (2.454)
UNRATEO	-114.670 * (1.662)	-109.370 (1.590)	-107.893 (1.565)
VOCPROG	415.373 (1.560)	341.187 (1.286)	354.309 (1.332)
R^2 (Adjusted)	0.200	0.206	0.203
F-Statistic	29.500	27.232	26.803

Table B-5: Model 5 Structural Selection Equation Estimates
- Uncorrected -

Variable	Military Service	College
Constant	367.716	-739.442
	(0.961)	(0.327)
ADDT	0.798408 **	-0.315515 **
	(3.026)	(1.976)
BLACK	-0.850003 **	0.266292
	(2.829)	(1.185)
COLPROG	-0.102368	0.760927 **
	(0.410)	(4.657)
COSTO	-0.977141-3	-0.956416-3
	(1.364)	(1.657)
DADMIL	1.25052 **	-0.520531
	(2.147)	(0.765)
DUNRATE	-0.132024	0.670749-1
	(0.976)	(0.749)
EDAD	-0.147965-1	0.138969-1
	(0.439)	(0.545)
EMOM	0.543627-1	0.185596-1
	(1.240)	(0.570)
ESIB	0.223114 **	0.257400 **
	(5.622)	(7.418)
HEXP	0.666834-1	0.463273 **
	(1.271)	(11.133)
HISP	-1.10498 **	0.392718
	(2.861)	(1.561)
LADO	-2.71204	-1.61378
	(0.358)	(0.257)
MSTAT	34.6604	-66.3888
	(1.020)	(0.328)
MWAGE	-0.922168-1	0.182324
	(0.961)	(0.321)
NADO	604.958	-1134.55
	(1.012)	(0.322)
NETY	-0.408352 **	0.272683-1 **
	(11.367)	(2.156)
PSIBS	0.319371	0.826503 **
	(1.178)	(4.069)
PWAGE	-0.956195-3 **	-0.517803-3 **
	(6.321)	(3.902)
RECO	-0.301993-3	-0.960619-4
	(0.888)	(0.375)
TFMAFS	-0.164400	-0.331456
	(0.327)	(1.006)
TSCORE	0.298783-1 **	0.298167-1
	(4.253)	(0.749)
VEAPO	-0.119085	0.221553
	(1.020)	(0.321)

Percentage Prediction Improvement = 43.65 $\rho^2 = 0.417$

Table B-6: Model 5 Structural Selection Equation Estimates
- Self-Selectivity Bias Corrected -

Variable	Military Service	College
Constant	360.292	-734.981
	(0.935)	(0.330)
ADDT	0.786030 **	-0.309989 *
	(2.983)	(1.940)
BLACK	-0.769019 **	0.302349
	(2.630)	(1.357)
COLPROG	-0.131374	0.737301 **
	(0.526)	(4.512)
COSTO	-0.902803-3	-0.918248-3
	(1.260)	(1.592)
DADMIL	1.20520 **	-0.519132
	(2.054)	(0.765)
DUNRATE	-0.131995	0.635291-1 **
	(0.977)	(7.164)
EDAD	-0.143837-1	0.134153-1
	(0.427)	(0.526)
EMOM	0.508150-1	0.172906-1
	(1.159)	(0.531)
ESIB	0.209525 **	0.249424 **
	(5.288)	(7.164)
HEXP	0.481130-1	0.449097 **
	(0.911)	(10.701)
HISP	-1.05504 **	0.403515
	(2.737)	(1.603)
LADO	-2.57588	-1.59529
	(0.339)	(0.254)
MSTAT	33.9150	-66.0281
	(0.991)	(0.331)
MWAGE	-0.903456-1	0.181245
	(0.436)	(0.324)
NADO	591.882	-1128.36
	(0.984)	(0.325)
NETY	-0.396588 **	0.794579 **
	(11.130)	(2.196)
PSIBS	0.277076	0.794579 **
	(1.025)	(3.924)
PWAGE	-0.915980-3 **	-0.493062-3 **
	(6.186)	(3.775)
RECO	-0.283989-3	-0.891418-4
	(0.834)	(0.348)
TFMAFS	-0.146248	-0.315366
	(0.291)	(0.958)
TSCORE	0.302671-1 **	0.300091-1 **
	(4.305)	(5.609)
VEAPO	-0.116477	0.220365
	(0.991)	(0.324)

Percentage Prediction Improvement = 40.15 ρ^2 = 0.416

Table B-7: Model 5 Structural Selection Equation Estimates - Self-Selectivity and Choice-Based Bias Corrected -

Variable	Military Service	College
Constant	48.942 (0.998)	-570.303 (0.251)
ADDT	0.659533 ** (2.220)	-0.299402 ** (1.927)
BLACK	-0.268224 (0.829)	0.240872 (1.100)
COLPROG	-0.124672 (0.432)	0.715868 ** (4.497)
COSTO	-0.770646-3 (0.952)	-0.920930-3 (1.626)
DADMIL	1.58710 ** (2.523)	-0.526561 (0.800)
DUNRATE	-0.201728 (1.299)	0.803009-1 (0.924)
EDAD	-0.220200-1 (0.587)	0.127030-1 (0.515)
EMOM	0.458747-1 (0.925)	0.133838-1 (0.425)
ESIB	0.200065 (0.251)	0.245626 ** (7.344)
HEXP	0.114403-1 (0.192)	0.452011 ** (11.053)
HISP	-0.840827 * (1.881)	0.347188 (1.413)
LADO	-4.97165 (0.550)	-0.181110 (0.029)
MSTAT	41.9987 (1.052)	-51.4053 (0.253)
MWAGE	-0.113613 (1.008)	0.140022 (0.246)
NADO	709.257 (1.010)	-870.764 (0.246)
NETY	-0.396049 ** (9.634)	0.284649-1 ** (2.320)
PSIBS	0.370122 (1.191)	0.785743 ** (3.958)
PWAGE	-0.767851-3 ** (4.240)	-0.508744-3 ** (3.863)
RECO	-0.370311-3 (0.966)	-0.127960-3 (0.512)
TFMAFS	-0.461533-1 (0.092)	0.333455 (1.055)
TSCORE	0.306384-1 ** (3.860)	0.290219-1 ** (5.579)
VEAPO	-0.138127 (1.007)	0.170011 (0.246)

$\rho^2 = 0.398$

Percentage Prediction Improvement = 40.15

Table B-8: Model 6 Structural Selection Equation Estimates
- Uncorrected -

Variable	Military Service	College
Constant	[illegible].364	-734.524
	([illegible].933)	(0.319)
ADDT	[illegible].795802 **	-0.310402 *
	([illegible].021)	(1.941)
BLACK	0.[illegible]51986 **	0.2648[illegible]5
	(2.8[illegible]9)	(1.177)
COLPROG	-0.[illegible]0[illegible]402	0.757273 **
	(0.[illegible]06[illegible])	(4.651)
COSTO	-0.8[illegible]5[illegible]3-3	-0.111146-2 *
	(1.0[illegible])	(1.775)
DADMIL	1.27[illegible]34 **	-0.518478
	(2.199)	(0.767)
EDAD	-0.1231[illegible]-1	0.128104-1
	(0.367)	(0.504)
EMOM	0.5747[illegible]6-[illegible]	0.190532-1
	(1.308)	(0.586)
ESIB	0.257415 **	0.257415 **
	(7.420)	(7.420)
HEXP	0.0689119-1	0.462294 **
	(1.312)	(11.116)
HISP	-1.06793 **	0.371162
	(2.785)	(1.483)
LADO	-2.27892	-1.87502
	(0.300)	(0.301)
MSTAT	33.4436	-65.9480
	(0.989)	(0.321)
MWAGE	-0.888394-1	0.180990
	(0.931)	(0.314)
NADO	590.498	-1129.53
	(0.903)	(0.315)
NETY	-0.407682 **	0.267040-1 **
	(11.392)	(2.111)
PSIBS	0.318407	0.821181 **
	(1.175)	(4.049)
PWAGE	-0.961642-3 **	-0.504917-3 **
	(6.306)	(3.777)
RECO	-0.228912-3	-0.157388-3
	(0.658)	(0.607)
TFMAFS	-0.148839	-0.326045
	(0.298)	(0.989)
TSCORE	0.291714-1 **	0.300860-1 **
	(4.149)	(5.625)
UNRATEO	-0.949114-1	0.187060-1
	(1.059)	(0.301)
VEAPO	-0.116394	0.220662
	(1.002)	(0.315)

$\rho^2 = 0.416$

Percentage Prediction Improvement = 43.65

Table B-9: Model 6 Structural Selection Equation Estimates - Self-Selectivity Bias Corrected -

Variable	Military Service	College
Constant	347.974	-729.892
	(0.909)	(0.322)
ADDT	0.783565 **	-0.305289 *
	(2.978)	(1.908)
BLACK	-0.770258 **	0.300446
	(2.637)	(1.347)
COLPROG	-0.130199	0.734890 **
	(0.521)	(4.498)
COSTO	-0.743998-3	-0.107249-2 *
	(0.952)	(1.713)
DADMIL	1.23293 **	-0.517137
	(2.108)	(0.767)
EDAD	-0.119851-1	0.123894-1
	(0.357)	(0.487)
EMOM	0.538551-1	0.178231-1
	(1.227)	(0.548)
ESIB	0.208688 **	0.249802 **
	(5.273)	(7.174)
HEXP	0.505351-1	0.178231-1
	(0.957)	(10.700)
HISP	-1.01750 **	0.382978
	(2.659)	(1.529)
LADO	-2.15524	-1.83994
	(0.284)	(0.295)
MSTAT	32.7068	-65.5710
	(0.962)	(0.324)
MWAGE	-0.869878-1	0.179865
	(0.906)	(0.317)
NADO	577.397	-1122.95
	(0.965)	(0.318)
NETY	-0.396057 **	0.272715-1 **
	(11.146)	(2.151)
PSIBS	0.276782	0.790692 **
	(1.023)	(3.910)
PWAGE	-0.919711-3 **	-0.480327-3 **
	(6.165)	(3.648)
RECO	-0.211919-3	-0.149283-3
	(0.608)	(0.575)
TFMAFS	-0.131961	-0.310716
	(0.265)	(0.943)
TSCORE	0.295742-1 **	0.302672-1 **
	(4.203)	(5.651)
UNRATEO	-0.909054-1	0.199047-1
	(1.016)	(0.320)
VEAPO	-0.113778	0.219393
	(0.974)	(0.318)

Percentage Prediction Improvement = 43.65 $\rho^2 = 0.416$

Table B-10: Model 6 Structural Selection Equation Estimates - Self-Selectivity and Choice-Base Bias Corrected -

Variable	Military Service	College
Constant	439.518	-558.629
	(0.978)	(0.241)
ADDT	0.639891 **	-0.295514 *
	(2.164)	(1.899)
BLACK	-0.251378	0.235974
	(0.780)	(1.076)
COLPROG	-0.115112	0.711266 **
	(0.398)	(4.469)
COSTO	-0.695438-3	-0.102958-2 *
	(0.784)	(1.678)
DADMIL	1.62797 **	-0.533135
	(2.587)	(0.814)
EDAD	-0.182724-1	0.115667-1
	(0.489)	(0.470)
EMOM	0.478409-1	0.138262-1
	(0.963)	(0.440)
ESIB	0.198812 **	0.245741 **
	(4.546)	(7.342)
HEXP	0.136282-1	0.451509 **
	(0.229)	(11.053)
HISP	-0.771610 *	0.323313
	(1.737)	(1.322)
LADO	-4.27904	-0.775689
	(0.468)	(0.127)
MSTAT	41.0373	-50.3548
	(1.029)	(0.243)
MWAGE	-0.110915	0.136991
	(0.984)	(0.236)
NADO	703.941	-856.209
	(1.003)	(0.238)
NETY	-0.395045 **	0.279556-1 **
	(9.654)	(2.277)
PSIBS	0.383059	0.777628 **
	(1.234)	(3.927)
PWAGE	-0.763848-3 **	-0.499523-3 **
	(4.197)	(3.761)
RECO	-0.293577-3	-0.185248-3
	(0.748)	(0.732)
TFMAFS	-0.247912-1	-0.328039
	(0.050)	(1.038)
TSCORE	0.298651-1 **	0.292882-1 **
	(3.789)	(5.624)
UNRATEO	-0.680436-1	0.111701-1
	(0.682)	(0.185)
VEAPO	-0.137268	0.167261
	(1.001)	(0.237)

Percentage Prediction = 40.81 $\rho^2 = 0.397$

Table B-11: Model 7 Structural Selection Equation Estimates
- Uncorrected -

Variable	Military Service	College
Constant	348.497	-682.168
	(0.984)	(0.372)
ADDT	0.722469 **	-0.303422 *
	(2.792)	(1.942)
BLACK	-0.60511 **	0.484200 **
	(2.127)	(2.221)
COLPROG	-0.128967	0.772078 **
	(0.522)	(4.884)
COSTO	-0.806432-3	-0.860969-3
	(1.136)	(1.518)
DADMIL	1.05712 *	-0.871479
	(1.824)	(1.273)
DUNRATE	-0.119990	0.670203-1
	(0.892)	(0.765)
EDAD	0.125776 **	0.138206 **
	(5.275)	(7.401)
EMOM	0.165525-1	-0.963409-2
	(0.397)	(0.313)
HEXP	0.592401-1	0.461666 **
	(1.152)	(11.457)
HISP	-0.721423 *	0.783706 **
	(1.921)	(3.246)
LADO	0.107646	0.613954
	(0.015)	(0.102)
MSTAT	32.7382	-61.4325
	(1.042)	(0.375)
MWAGE	-0.869275-1	0.168508
	(0.980)	(0.367)
NADO	575.409	-1046.850
	(1.040)	(0.366)
NETY	-0.404931 **	0.302866-1 **
	(11.494)	(2.384)
PSIBS	0.285678	0.676661 **
	(1.077)	(3.453)
PWAGE	-0.100871-2 **	-0.590011-3 **
	(6.826)	(4.608)
RECO	-0.277653-3	-0.153488-3
	(0.833)	(0.615)
TFMAFS	-0.271340	-0.380486
	(0.532)	(1.154)
TSCORE	0.323495-1 **	0.323534-1 **
	(4.719)	(6.165)
VEAPO	-0.113504	0.204306
	(0.892)	(0.366)

Percentage Prediction Improvement = 41.46 $\rho^2 = 0.391$

Table B-12: Model 7 Structural Selection Equation Estimates
- Self-Selectivity Bias Corrected -

Variable	Military Service	College
Constant	340.957	-687.247
	(0.956)	(0.378)
ADDT	0.707147 **	-0.294184 *
	(2.739)	(1.880)
BLACK	-0.501729 *	0.529648 **
	(1.784)	(2.451)
COLPROG	-0.169666	0.734616 **
	(0.686)	(4.624)
COSTO	-0.719154-3	-0.811361-3
	(1.013)	(1.431)
DADMIL	1.00917 *	-0.849355
	(1.724)	(1.248)
DUNRATE	-0.121722	0.607118-1
	(0.907)	(0.693)
EDAD	0.113105 **	0.129916 **
	(4.747)	(6.996)
EMOM	0.147598-1	-0.961329-2
	(0.354)	(0.313)
HEXP	0.332638-1	0.439774 **
	(0.641)	(10.786)
HISP	-0.677529 *	0.784260 **
	(1.810)	(3.246)
LADO	0.107303	0.556587
	(0.015)	(0.092)
MSTAT	31.9693	-61.9307
	(1.010)	(0.380)
MWAGE	-0.850635-1	0.169819
	(0.952)	(0.372)
NADO	561.570	-1055.87
	(1.008)	(0.372)
NETY	-0.387944 **	0.311257-1 **
	(11.147)	(2.443)
PSIBS	0.231429	0.634055 **
	(0.875)	(3.249)
PWAGE	-0.943677-3 **	-0.552092-3 **
	(6.596)	(4.418)
RECO	-0.249729-3	-0.141304-3
	(0.749)	(0.566)
TFMAFS	-0.244284	-0.353953
	(0.480)	(1.073)
TSCORE	0.329650-1 **	0.326948-1 **
	(4.805)	(6.218)
VEAPO	-0.110709	0.206105
	(1.017)	(0.372)

Percentage Prediction Improvement = 41.35 $\rho^2 = 0.389$

Table B-13: Model 7 Structural Selection Equation Estimates
- Self-Selectivity and Choice-Based Bias Corrected -

Variable	Military Service	College
Constant	388.252 (0.928)	-577.101 (0.337)
ADDT	0.588399 ** (2.010)	-0.276564 * (1.819)
BLACK	0.211307-2 (0.007)	0.457897 ** (2.157)
COLPROG	-0.158521 (0.552)	0.712492 ** (4.604)
COSTO	-0.561509-3 (0.701)	-0.806723-3 (1.449)
DADMIL	1.45180 ** (2.316)	-0.842965 (1.275)
DUNRATE	-0.193770 (1.245)	0.745944-1 (0.878)
EDAD	0.104370 ** (3.958)	0.128307 ** (7.192)
EMOM	0.555941-2 (0.117)	-0.141247-1 (0.475)
HEXP	-0.120119-2 (0.020)	0.441989 ** (11.142)
HISP	-0.465451 (1.058)	0.719254 ** (3.045)
LADO	-2.07532 (0.243)	1.90924 (0.340)
MSTAT	36.3261 (0.979)	-52.1149 (0.340)
MWAGE	-0.979765-1 (0.935)	0.142264 (0.331)
NADO	614.597 (0.941)	-883.382 (0.331)
NETY	-0.389487 ** (9.662)	0.316907-1 ** (2.574)
PSIBS	0.284884 (0.933)	0.607153 ** (3.176)
PWAGE	-0.808927-3 ** (4.575)	-0.581681-3 ** (4.611)
RECO	-0.365277-3 (0.962)	-0.180675-3 (0.742)
TFMAFS	-0.183506 (0.359)	-0.374311 (1.180)
TSCORE	0.331470-1 ** (4.253)	0.317109-1 ** (6.208)
VEAPO	-0.119781 (0.939)	0.172391 (0.330)

$\rho^2 = 0.371$

Percentage Prediction Improvement = 37.96

Table B-14: Model 8 Structural Selection Equation Estimates
- Uncorrected -

Variable	Military Service	College
Constant	337.589	-678.027
	(0.959)	(0.366)
ADDT	0.721717 **	-0.299598 *
	(2.791)	(1.914)
BLACK	-0.609641 **	0.489454 **
	(2.141)	(2.200)
COLPROG	-0.129316	0.766950 **
	(0.523)	(4.856)
DADMIL	1.08076 *	-0.870847
	(1.872)	(1.279)
EDAD	0.127382 **	0.137185 **
	(5.346)	(7.383)
EMOM	0.198055-1	-0.920857-2
	(0.474)	(0.300)
HEXP	0.612546-1	0.461148 **
	(1.190)	(11.448)
HISP	-0.683592 *	0.761781 **
	(1.1834)	(3.171)
LADO	0.303757	0.279313
	(0.042)	(0.047)
MSTAT	31.6511	-61.0546
	(1.1031)	(0.369)
MWAGE	-0.839010-1	0.167383
	(0.951)	(0.361)
NADO	562.067	-1043.37
	(1.022)	(0.362)
NETY	-0.404252 **	0.297888-1 **
	(11.502)	(2.344)
PSIBS	0.28508	0.673313 **
	(1.074)	(3.440)
PWAGE	-0.101610-2 **	-0.579983-3 **
	(6.815)	(4.492)
RECO	-0.197902-3	-0.208814-3
	(0.581)	(0.828)
TFMAFS	-0.2544202	-0.377350
	(0.503)	(1.143)
TSCORE	0.316067-1 **	0.325582-1 **
	(4.601)	(6.197)
UNRATEO	-0.110814	0.752854-2
	(1.255)	(0.125)
VEAPO	-0.111036	0.203712
	(1.034)	(0.361)

Percentage Prediction Improvement = 41.35 ρ^2 = 0.390

Table B-15: Model 8 Structural Selection Equation Estimates - Self-Selectivity Bias Corrected -

Variable	Military Service	College
Constant	330.425	-682.983
	(0.933)	(0.372)
ADDT	0.706569 **	-0.290816 *
	(2.738)	(1.856)
BLACK	-0.504242 *	0.526057 **
	(1.793)	(2.431)
COLPROG	-0.169616	0.731130 **
	(0.686)	(4.605)
COSTO	-0.520745-3	-0.930747-3
	(0.677)	(1.522)
DADMIL	1.03300 *	-0.849332
	(1.773)	(1.254)
EDAD	0.114791 **	0.129248 **
	(4.825)	(6.989)
EMOM	0.178267-1	-0.920654-2
	(0.427)	(0.300)
HEXP	0.354729-1	0.440034 **
	(0.683)	(10.796)
HISP	-0.638132 *	0.764791 **
	(1.718)	(3.181)
LADO	0.331859	0.264470
	(0.046)	(0.044)
MSTAT	30.9247	-61.5401
	(0.984)	(0.375)
MWAGE	-0.821482-1	0.168661
	(0.925)	(0.366)
NADO	548.827	-1051.940
	(0.992)	(0.368)
NETY	-0.387432 **	0.306229-1 **
	(11.155)	(2.402)
PSIBS	0.231463	0.632219 **
	(0.875)	(3.243)
PWAGE	-0.948660-3 **	-0.542283-3 **
	(6.578)	(4.300)
RECO	-0.172120-3	-0.194939-3
	(0.505)	(0.772)
TFMAFS	-0.229213	-0.351245
	(0.454)	(1.064))
TSCORE	0.322556-1 *	0.328867-1 **
	(4.692)	(6.248)
UNRATEO	-0.103404	0.981650-2
	(1.175)	(0.163)
VEAPO	-0.108353	0.205419
	(1.002)	(0.367)

Percentage Prediction Improvement = 41.46 ρ^2 = 0.389

Table B-16: Model 8 Structural Selection Equation Estimates - Self-Selectivity and Choice-Base Bias Corrected -

Variable	Military Service	College
Constant	382.212	-568.381
	(0.917)	(0.329)
ADDT	0.572854 **	-0.274246 *
	(1.963)	(1.801)
BLACK	0.195136-1	0.451641 **
	(0.064)	(2.124)
COLPROG	-0.149768	0.706774 **
	(0.521)	(4.572)
COSTO	-0.44001-3	-0.884797-3
	(0.505)	(1.478)
DADMIL	1.48737 **	-0.850153
	(2.377)	(1.291)
EDAD	0.107244 **	0.127200 **
	(4.078)	(7.167)
EMOM	0.806535-2	-0.13733-1
	(0.170)	(0.463)
HEXP	-0.197736-3	0.442097 **
	(0.003)	(11.151)
HISP	-0.390527	0.696819 **
	(0.896)	(2.964)
LADO	-1.50684	1.30258
	(0.174)	(0.222)
MSTAT	35.6644	-51.3246
	(0.964)	(0.332)
MWAGE	-0.961221-1	0.139991
	(0.920)	(0.323)
NADO	613.723	-873.413
	(0.943)	(0.324)
NETY	-0.38822 **	0.312828-1 **
	(9.671)	(2.539)
PSIBS	0.294752	0.602008 **
	(0.965)	(3.155)
PWAGE	-0.804532-3 **	-0.575370-3 **
	(4.521)	(4.519)
RECO	-0.283305-3	-0.231385-3
	(0.731)	(0.940)
TFMAFS	-0.159796	-0.370632
	(0.315)	(1.168)
TSCORE	0.323998-1 **	0.319129-1 **
	(4.184)	(6.242)
UNRATEO	-0.792186-1	0.103277-2
	(0.801)	(0.018)
VEAPO	-0.119792	0.170528
	(0.942)	(0.324)

$\rho^2 = 0.370$

Percentage Prediction Improvement = 38.09

Appendix C

Derivation of Elasticity Formulas

In general, for a given X_j, the elasticity of the dependent variable (Y) with respect to X_j is given as

$$(C.1) \qquad E_{Yj} = \partial Y/\partial X_j * \overline{X}_j/\overline{Y}.$$

However, in a logistic model specification, the formula for calculating the elasticity is slightly different. Within a multinomial logit specification, the probability of a given outcome (ie - the dependent variable) is expressed as

$$(C.2) \qquad P_k = \frac{\exp(\sum_{i=1}^{I} \beta_{ik} X_i)}{\sum_{k=1}^{K}(\exp(\sum_{i=1}^{I} \beta_{ik} X_i))},$$

where P_k = the probability of the kth outcome,

β_{ik} = the coefficient of the ith explanatory in the kth equation, and

X_i = the ith explanatory variables. For simplicity it is assumed that $X_{ik} = X_i$, for all k.

For the present case, the outcome set has three elements, or k = 0,1,2. As a simplification let $V_k = sum(i=1,I)\beta_{ik}X_i$. Therefore, the outcome probabilities are expressed as

(C.3) $$P_0 = exp(V_0)/(exp(V_0) + exp(V_1) + exp(V_2)),$$

(C.4) $$P_1 = exp(V_1)/(exp(V_0) + exp(V_1) + exp(V_2)), \text{ and}$$

(C.5) $$P_2 = exp(V_2)/(exp(V_0) + exp(V_1) + exp(V_2)).$$

This can be further simplified by normalizing the coefficients of the first outcome to equal 0. Or,

(C.3') $$P_0 = 1/(1 + exp(V_1) + exp(V_2)),$$

(C.4') $$P_1 = exp(V_1)/(1 + exp(V_1) + exp(V_2)), \text{ and}$$

(C.5') $$P_2 = exp(V_2)/(1 + exp(V_1) + exp(V_2)).$$

The first step in the derivation is to find the partial derivative of a particular P_k (say P_1) with respect to an X_j. This is given as

(C.6) $$\partial P_1/\partial X_j =$$

$$\frac{\beta_{j1}[\exp(V_1)][1+\exp(V_1)+\exp(V_2)]-[\exp(V_1)]\{\beta_{j1}[\exp(V_1)]+\beta_{j2}[\exp(V_2)]\}}{[1 + \exp(V_1) + \exp(V_2)]^2}$$

$$= \frac{\beta_{j1}[\exp(V_1)+\exp(2V_1)+\exp(V_1+V_2)]-\beta_{j1}[\exp(2V_1)]-\beta_{j2}[\exp(V_1+V_2)]}{[1 + \exp(V_1) + \exp(V_2)]^2}$$

$$= \frac{\beta_{j1}[\exp(V_1)] + [\beta_{j1} - \beta_{j2}][\exp(V_1+V_2)]}{[1 + \exp(V_1) + \exp(V_2)]^2} .$$

Substituting from equations (C.3') - (C.5') into equation (C.6) yields

$$\text{(C.7)} \qquad \partial P_1/\partial X_j = \beta_{j1}P_0P_1 + [\beta_{j1} - \beta_{j2}]P_1P_2$$

$$= \beta_{j1}(P_0 + P_2)P_1 - \beta_{j2}P_1P_2 .$$

Now, define $P_0 = 1 - P_1 - P_2$. Substituting this expression into equation (C.7) and further simplifying, gives

$$\text{(C.8)} \qquad \partial P_1/\partial X_j = \beta_{j1}(1 - P_1 - P_2 + P_2)P_1 - \beta_{j2}P_1P_2$$

$$= \beta_{j1}(1 - P_1)P_1 - \beta_{j2}P_1P_2$$

The final step needed to find the elasticity formula is to multiply equation (C.8) by $\overline{X}_j/\overline{P}_1$ and set $P_1 = \overline{P}_1$ and $P_2 = \overline{P}_2$. This yields an expression for the elasticity similar to that in equation (C.1). Or,

(C.9) $$E_{1j} = \partial P_1/\partial X_j * \overline{X}_j/\overline{P}_1 = [\beta_{j1}(1 - \overline{P}_1)\overline{P}_1 - \beta_{j2}\overline{P}_1\overline{P}_2] * [\overline{X}_j/\overline{P}_1]$$

$$= [\beta_{j1}(1 - \overline{P}_1) - \beta_{j2}\overline{P}_2] * \overline{X}_j \ .$$

It is not difficult to derive similar expressions for the other model outcomes.

References

Altman, S.H. and A.E. Fechter (1967). "The Supply of Military Manpower in the Absence of a Draft", American Economic Review, Vol. 77, No. 2, pps. 19-31.

Amemiya, T. (1981). "Qualitiative Response Models: A Survey", Journal of Economic Literature, Vol. 19, No. 4, pps. 483-536.

Amey, D.M., A.E. Fechter, D.W. Grissmer, and G.P. Sica (1976). "Supply Estimation of Enlistees to the Military", McLean, VA: General Research Corporation.

Ash, C., B. Udis, and R.F. McNown (1983). "Enlistments in the All-Volunteer Force: A Military Personnel Supply Model and Its Forecasts", American Economic Review, Vol. 73, No. 1, pps. 145-155.

Ashenfelter, O. (1979). "Estimating the Effect of Training Programs on Earnings with Longitudinal Data", in Research in Labor Economics, Supplement 1, Greenwich, CT: JAI Press, Inc, pps. 97-117.

Avery, R.B. (1979). Qualitative Dependent Variable Program, Pittsburg, PA: Carnegie-Mellon University.

Bachman, J.G., J.D. Blair, and D.R. Segal (1977). The All-Volunteer Force, Ann Arbor: University of Michigan Press.

Baldwin, R.H., T.V. Daula and T.W. Fagan (1982). "As Our Policy Maker's Questions Become More Sophisticated, Our Research Methods Must Match that Sophistication. Some Micro Data Evidence of Enlistment, Attrition and Reenlistment", West Point, New York: Department of Social Sciences, Army Center for Economic and Manpower Research.

Becker, G.S. (1975). Human Capital, second edition, New York: National Bureau of Economic Research.

Binkin, M. (1982). Blacks and the Military, Washington, D.C.: The Brookings Institution.

_____ , (1984). America's Volunteer Military: Progress and Prospects, Washington, D.C.: The Brookings Institution.

Bock, R.D. and L.V. Jones (1968). The Measurement and Prediction of Judgement and Choice, San Francisco: Holden-Day.

Brown, C. (1984). "Military Enlistments: What Can We Learn from Geographic Variation?", working paper no. 1261, Cambridge, MA: National Bureau of Economic Research, Inc.

Burton, G.C. (1970). "Supply of First Term Military Enlistees", in Studies Prepared for the President's Commission on an All-Volunteer Armed Forces, Washington, D.C.: U.S. Government Printing Office.

Cooper, R.V.L. (1977). "Military Manpower and the All-Volunteer Force", report no. R-1450-ARPA, Santa Monica: The Rand Corporation.

_____ , (1979). Defence Manpower Policy, Santa Monica: The Rand Corporation.

Daganzo, C. (1980). Multinomial Probit, New York: Academic Press.

Dale, C. and C. Gilroy (1983a). "The Effects of the Business Cycle on the Size and Composition of the U.S. Army", Atlantic Economic Journal, Vol. 11, No. 1, pps. 42-53.

_____ , (1983b). "Forecasting Army Enlistments", working paper no. PPRG-83-1, Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

_____, (1984). "Determinants of Enlistments: A Macroeconomic Time-Series View", Armed Forces and Society, Vol. 10, No. 2, pps. 192-210.

Daula, T.V. (1981). "The Retention of First-Term Soldiers: Theory and Empirical Evidence", Ph.D. dissertation, Cambridge, MA: Massachusetts Institute of Technology.

Daula, T.V., T.W. Fagan and D.A. Smith (1982). "A Microdata Model of Enlistment in the Armed Forces", paper presented at the summer meetings of the Econometric Society.

Daula, T.V. and D.A. Smith (1984). "Recruiting Goals, Enlistment Supply and Enlistments in the U.S. Army", paper presented at the U.S. Army Manpower Economics Conference, Williamsburg, VA.

Dertouzos, J.N. (1984). "Enlistment Supply, Recruiting Objectives, and the All-Volunteer Army", report no. P-7022, Santa Monica: The Rand Corporation.

DeVany, A.S. and T.R. Saving (1982). "Life Cycle Job Choice and the Demand and Supply of Entry-Level Jobs--Some Evidence from the Air Force", Review of Economics and Statistics, Vol. 64, No. 3, pps. 457-465.

Domencich, T.A. and D. McFadden (1975). Urban Travel Demand: A Behavioral Analysis, Amsterdam: North-Holland Publishing Company.

Dubin, J.A. and D. McFadden (1984). "An Econometric Analysis of Residential Electric Appliance Holdings and Consumption", Econometrica, Vol. 52, No. 2, pps. 345-362.

Elig, T., P.A. Gade and J.L. Shields (1982). "The 1982 Department of the Army Survey of Personnel Entering the Army", paper presented at the meeting of Military Testing Association, San Antonio, TX.

Ellwood, D.T. and D.T. Wise (1984). "Uncle Sam Wants You-Sometimes", unpublished draft.

Evans, T. (undated). "Army Advertising", unpublished mimeo, U.S. Army Recruiting Command, Fort Sherridan, IL.

Faris, J.H. (1984). "Economic and Non-Economic Factors of Personnel Recruitment and Retention in the AVF", Armed Forces and Society, Vol. 10, No.2, pps. 251-275.

Fechter, A.E. (1979). "The Supply of Enlisted Volunteers in the Post-Draft Environment: An Evaluation Based on Pre-1972 Experience", in R.V.L. Cooper, ed., Defence Manpower Policy, Santa Monica: The Rand Corporation, pps. 87-99.

Fernandez, R.L. (1979). "Forecasting Enlisted Supply: Projections for 1970-90", report no. N-1297-MRAL, Santa Monica: The Rand Corporation.

______, (1980). "Issues in the Use of Post-Service Educational Benefits as Enlistment Incentives", report no. N-1510-MRAL, Santa Monica: The Rand Corporation.

Fisher, A. (1969). "The Cost of the Draft and the Cost of Ending the Draft", American Economic Review, Vol. 59, No. 3, pps. 239-254.

Goldberg, L. (1980). "Recruiters, Advertising, and Navy Enlistments", professional paper no. 275, Alexandria, VA: The Center for Naval Analysis.

Goldberg, L. and P. Greenston (1983). "A Time-Series, Cross-Sectional Study of Enlistment Supply: Army, Navy, Air Force, Marine Corps 1976-1982", paper presented at the Fifty-Third Annual Conference of the Southern Economic Association, Washington, D.C.

Greene, W.H. (1985). LIMDEP, New York: Department of Economics, Graduate School of Business Administration, New York University.

Grissmer, D.W. (1979). "The Supply of Enlisted Volunteers in the Post-Draft Environment: An Analysis Based on Monthly Data, 1970-1975", in R.V.L. Cooper, ed., Defence Manpower Policy, Santa Monica: The Rand Corporation, pps. 100-115.

Gwartney, J. and J. Long (1978). "The Relative Earnings of Blacks and Other Minorities", International Labor Relations Review, Vol. 31, pps. 336-346.

Harworth, J., J. Gwartney and C. Harworth (1975). "Earnings, Productivity, and Changes in Employment Discrimination During the 1960's", American Economic Review, Vol. 65, No. 1, pps. 158-168.

Hausman, J.A. and D. McFadden (1984). "Specification Tests for the Multinomial Logit Model", Econometrica, Vol. 52, No. 5, pps. 1219-1240.

Hausman, J.A. and D.A. Wise (1978). "A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Hetrogeneous Preferences", Econometrica, Vol. 46, No. 2, pps. 403-426.

Hay, J.W. (1984). "Physician's Specialty Choice and Specialty Income", Center for Health Affairs, Millwood, VA: Project HOPE.

Heckman, J. (1976). "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models", Annals of Economic and Social Measurement, Vol. 5, pps. 475-492.

______ , (1979). "Sample Bias as a Specification Effor", Econometrica, Vol. 47, No. 1, pps. 153-162.

Holden, C. (1980). "Doubts Mounting About All-Volunteer Force", Science, Vol. 209, pps. 1095-1099.

Horne, D.K. (1984). "An Economic Analysis of Army Enlistment Supply", working paper no. MPPRG-84-5, Alexandria, VA: U.S Army Research Institute for the Behavioral and Social Sciences.

Hosek, J.R. and C.E. Peterson (1984). "Enlistment Decisions of Young Men", report no. WD-2450-MIL, Santa Monica: The Rand Corporation.

Huck, D., L. Kusmin and E. Shepard (1982). "Improving Educational Benefits: Effects on Costs, Recruiting and Retention", U.S. Congressional Budget Office, Washington, D.C.

Janowitz, M. and C. Moskos (1979). "Five Years of the All-Volunteer Force: 1973-1978", Armed Forces and Society, Vol. 5, No. 2, pps. 171-218.

Jehn, C. and W.F. Shughart (1979). "Modeling Recruiting District Performance", in R.V.L. Cooper, ed., Defence Manpower Policy, Santa Monica: The Rand Corporation, pps. 137-148.

Johnson, N.L. and S. Kotz (1972). Distributions in Statistics: Continuous Multivariate Distributions, New York: John Wiley and Sons.

Judge, G.G., W.E. Griffiths, R.C. Hill and T.C. Lee (1980). The Theory and Practice of Econometrics, New York: John Wiley and Sons.

Kalton, G. (1982). "Imputing for Missing Survey Responses", Proceedings of the Section on Survey Research Methods, American Statistical Association.

Kenny, L.W., L.F. Lee, G.S. Maddala and R.P. Trost (1979). "Returns to College Education: An Investigation of Self-Selection Bias Based on the Project Talent Data", International Economic Review, Vol. 20, No. 3, pps. 775-789.

Kim, C. (1982). "The All-Volunteer Force: 1979 NLS Studies of Enlistment, Intentions to Serve, and Intentions to Re-Enlist", Columbus, Ohio: Center for Human Resource Institute, The Ohio State University.

Kim, K.H., S. Farrell and E. Clague (1971). The All-Volunteer Army: An Analysis of Demand and Supply, New York: Praeger Press.

King, A. and C. Knapp (1978). "Race and the Determinants of Lifetime Earnings", International Labor Relations Review, Vol. 31, No. 3, pps. 347-355.

Lee, L.F. (1978). "Unionism and Wage Rates: A Simultaneous Equation Model with Qualitiative and Limited Dependent Variables", International Economic Review, Vol. 19, No. 2, pps. 415-433.

_____ , (1982). "Some Approaches to the Correction of Selectivity Bias", Review of Economic Studies, Vol. 49, pps. 355-72.

_____ , (1983). "Generalized Econometric Models with Selectivity", Econometrica, Vol. 51, No. 2, pps. 507-512.

Levy, F. (1979). "The Labor Supply of Female Household Heads, or AFDC Incentives Don't Work Too Well", The Journal of Human Resources, Vol. 14, No. 1, pps. 76-97.

_____ , (1980). "Changes in Employment Prospects for Black Males", in Brookings Papers on Economic Activity, Vol. 2, pps. 513-538.

Maddala, G.S. (1977). Econometrics, New York: McGraw-Hill Book Company.

_____ , (1983). Limited-Dependent and Qualititive Variables in Econometrics, Cambridge, England: Cambridge University Press.

Maier, M.H. and F.C. Grafton (1980). "Renorming ASVAB 6/7 at Armed Forces Examining and Entrance Stations", a managerial report of Technical Memorandum 80-1, Directorate for Accession Policy, Office of the Secretary of Defence, Washington, D.C..

Manski, C.F. and S. Lerman (1977). "The Estimation of Choice Probabilities for Choice Based Samples", Econometrica, Vol. 45, No. 8, pps. 1977-1988.

Manski, C.F. and D. McFadden (1981). "Alternative Estimators and Sample Designs for Discrete Choice Analysis", in C.F. Manski and D. McFadden, eds., Structural Analysis of Discrete Data with Econometric Applications, Cambridge, MA: The MIT Press, pps. 2-50.

McFadden, D. (1976). "Quantal Choice Analysis: A Survey", Annals of Economic and Social Measurement, Vol. 5, pps. 511-523.

McNown, R.F., B. Udis and C. Ash (1980). "Economic Analysis of the All-Volunteer Force", Armed Forces and Society, Vol. 7, No. 1, pps. 113-132.

Medoff, J.L. and K.G. Abraham (1980). "Experience, Performance and Earnings", The Quarterly Journal of Economics, Vol. 95, No. 4, pps. 703-736.

Mincer, J. (1974). Schooling, Experience and Earnings, New York: National Bureau of Economic Research.

Mood, A.M., F.A. Graybill and D.C. Boes (1974). Introduction to the Theory of Statistics, Third Edition, New York: McGraw-Hill Book Company.

Morey, R.C. and J.M. McCann (1983). "Armed Services Recruiting Research: Issues, Findings, and Needs", unpublished paper.

Office of the Assistant Secretary of Defence/Manpower, Reserve Affairs, and Logistics (1980). "Aptitude Testing of Recruits", A Report to the House Committee on Armed Services.

Office of the Secretary of Defence (1982). Military Compensation Background Papers, second edition, Washington, D.C.: U.S. Government Printing Office.

Olsen, R. J. (1980). "A Least Squares Correction for Selectivity Bias", Econometrica, Vol. 48, No. 7, pps. 1815-1820.

Oi, W. (1967). "The Economic Cost of the Draft", American Economic Review, Vol. 77, No. 2, pps. 39-62.

Parnes, H.S. (1975). "The National Longitudinal Surveys: New Vistas for Labor Market Research", American Economic Review, Vol. 65, No. 2, pps. 244-249.

Pindyke, R.S. and D.L. Rubinfeld (1981). Econometric Models and Economic Forecasts, second edition, New York: McGraw-Hill Book Company.

Rosen, S. (1977). "Human Capital: A Survey of Empirical Research", in Research in Labor Economics, Greenwich, Conn: JAI Press, Inc.

Santos, R. (1981). "Effects of Imputation on Regression Coefficients", Proceedings of the Section on Survey Research Methods, American Statistical Association.

Shields, J.L. and F.C. Grafton (1983). "A Natural Experiment: Analysis of an Almost Unselected Army Population", Alexandria, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

The National Longitudinal Surveys Handbook 1983-1984 (1983), Columbus, Ohio: Center for Human Resource Research, The Ohio State University.

The Report of the President's Commission on an All-Volunteer Force (1970), New York: Macmillian.

The Washington Post, 9 December, 1984.

U.S. General Accounting Office (1976). "Advertising for Military Recruiting: How Effective is it?", report no. FPCD-76-168, Washington, D.C..

_____ , (1984). "Military and Federal Civilian Disposable Income Comparisons and Extra Pays Received by Military Personnel", report no. NSIAD-84-41, Washington, D.C..

Welch, F. (1967). "Labor-Market Discrimination: An Interpretation of Income Differences in the Rural South", _Journal of Political Economy_, Vol. 75, pps. 225-240.

_____ , (1973). "Black-White Differences in Returns to Schooling", _American Economic Review_, Vol. 63, No. 5, pps. 893-907.

Willis, R.J. and S. Rosen (1979). "Education and Self-Selection", _Journal of Political Economy_, Vol. 87, No. 5, pps. s7-s36.

Wise, D.A. (1975). "Acedemic Achievement and Jobe Performance", American Economic Review, Vol 65, No. 3, pps. 350-366.

Withers, G.A. (1979). "International Comparisons in Manpower Supply", in R.V.L. Cooper, ed., Defence Manpower Policy, Santa Monica: The Rand Corporation, pps. 116-136.