LEGACY STATUS AS A SIGNAL IN COLLEGE ADMISSIONS

by

LEONARD D. CABRERA

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2006

ACKNOWLEDGMENTS

I would like to thank my advisor, David Figlio, and committee members Larry Kenny, Rich Romano, and Bob Emerson for their guidance and support. I would also like to thank Dennis Epple, David Denslow, Doug Waldo, Chunrong Ai, Sarah Hamersma, Josh Kneifel, and Jennifer Shelamer for their assistance, comments, and suggestions. Additionally, I appreciate the assistance of the Air Force Academy, specifically, Rich Fullerton and Mike Lucchesi from the Department of Economics and Geography, William "Trapper" Carpenter and Rollie Stoneman from the Admissions Office, and Jeff Thompson, Kathy O'Donnell, Dave Skowron, and Jau Tsau from the Plans and Analysis Division, for their generous support in providing background material and data.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

LEGACY STATUS AS A SIGNAL IN COLLEGE ADMISSIONS

By

Leonard D. Cabrera

August 2006

Chair: David Figlio
Major Department: Economics

Opponents of legacy admit policies claim such policies are inherently discriminatory and contrary to a merit-based system, yet many universities award admissions points to legacy applicants. The term "legacy" is used to describe a college student whose parent is an alumnus of the same university. This dissertation looks at measurable performance benefits to investigate the idea that legacy status provides some information to admissions offices. Empirical data from the Air Force Academy graduating classes of 1994 to 2005 are used. The variables of interest include traditional academic measures as well as student choices of academic major and career field and several post-educational measures.

Logit or multinomial logistic regressions are run for each performance measure while controlling for high school performance, standardized test scores, and demographic data. Legacy status has no significant impact on grades, order

of merit, college major or Air Force rank. However, legacy status is associated with a 0.10 increase in the probability of graduation and 0.04 point higher military performance average. The graduation figure results from legacy admits being less likely to voluntarily quit, and the results are even more dramatic for less qualified students. For graduates, legacy status leads to a 0.09 increase in the probability of being a rated officer and 0.11 increase in the probability of serving at least 8 years in the Air Force. These results are robust to model specification.

A theoretical model of the admissions process is developed that formalizes the influence of legacy status: a direct effect on graduation probability, a selection impact through enrollment, and a signaling effect for unobserved student characteristics. These effects cannot be estimated separately, so empirical results measure the overall impact of legacy status. The model suggests a technique for testing the optimality of the admissions process, but requires data on all applicants. The additional data are also required to examine other potential sources of bias in the empirical work.

CHAPTER 1
INTRODUCTION

In a 2004 speech on affirmative action, President Bush was asked whether colleges should eliminate legacy policies because, in the reporter's view, they are not based on merit, but on where an applicant's parent went to college.[1] Despite this view, many colleges defend the practice and insist that legacy admits are equally (or better) qualified than their peers, they perform better, and they bring in more donations as alumni.[2] This dissertation studies the effects of legacy status on educational outcomes, student choices, and post-educational outcomes.

Some schools have admissions policies that favor legacy admits. The policies can be as innocuous as awarding a few extra points to the application or as blatant as accepting the student regardless of qualification. Arguments for and against these legacy policies center around economic equity and efficiency arguments, but the question of whether to use legacy status is not resolved. A formal theory is proposed in this dissertation that shows legacy status could be used as a signal of unobserved student characteristics which do lead to increased student performance.

---

[1] A student is considered a legacy admit if either parent is an alumnus of the school. For this paper, the terms school, college, and university are used interchangeably.

[2] Schmidt (2004), Sanoff (2004), Lassila (2004)

Empirical data from the United States Air Force Academy graduating classes of 1994 to 2005 are used to verify the assertion. By focusing only on data available during the admissions process, it is possible to determine whether legacy status is a valid signal of future performance, especially when compared to other signals used for college entry. Traditional academic measures such as graduation rates, grades and graduation order of merit are considered. Using data from the Academy eliminates possible confounding effects of monetary contributions and gives clear post-educational outcomes.[3] Graduate performance is measured by student choice of college major as well as Air Force career field, time in service, and Air Force rank.

A probit model is used to predict the probability of graduation as a function of admissions data and legacy status. Control variables for high school state, gender, and race are also included. Splines are used to allow for nonlinear relationships between the admissions data and graduation rates. Subsets of the data are used to determine if legacy status affects students differently. A multinomial logistic regression is used to identify the effect of legacy status on students who fail and those who quit for non-academic reasons. Ordinary least squares (OLS) models are run using the same control variables to predict student grade point average (GPA), military performance average (MPA), and graduation order of merit, a measure that combines academic, military, and athletic performance.

[3] Theoretically, if a school is receiving monetary compensation for a legacy admit, there is a tradeoff between student performance and alumni donations that could result in legacy admits having lower performance than non-legacy students.

Multinomial logistic regressions are used to predict the probability of graduates attaining engineering or scientific majors and the probability of going on to flying or technical careers. To predict time in service and Air Force rank, binary variables are created for cutoff values. These variables are then predicted using logit models. These latter models are severely limited by the available data, so an extension is made by using Academy performance measures as control variables.

The average impact of legacy status is a 0.10 increase in the probability of graduation. When the sample is restricted to the least academically qualified students, legacy status has a stronger impact on student success. Therefore, in the cases in which the legacy policy is more likely to help an applicant get admitted, the signal of legacy status is more important. The 10 point difference in graduation probability stems mostly from non-legacy students who choose to not graduate (i.e., quit for issues other than grades). Legacy status does not have a significant effect on a student's GPA or graduation order of merit, but does result in graduates whose average MPA score is 0.04 points higher than non-legacy graduates.

Legacy status has no statistically significant relationship with academic major or Air Force rank, but is positively correlated with career field and time in service. Legacy graduates are roughly 9 percentage points more likely to be rated officers and nearly 11 percentage points more likely to serve beyond 8 years. Extending the data set back to 1982 shows that military performance at the Academy is at least ten times as important as grades in predicting time in service and rank.

Several robustness tests are performed. The impact of legacy status is independent of the other control variables and not very sensitive to model specification. The results may not generalize to all universities because of the unique characteristics of the Air Force Academy, but they are likely to be evident in high skill programs such as medical school.

Unfortunately, these results may be biased because of selection issues. A theoretical model of Academy admissions is developed that allows legacy status to have a direct impact on graduation probability, a selection impact through enrollment, and a signaling effect for unobserved student characteristics. These effects cannot be estimated separately, so empirical results measure the overall impact of legacy status, which is the correct measure to evaluate the admissions policy. The model suggests a technique for testing the optimality of the admissions process, but requires data on all applicants. The additional data are also required to examine other potential sources of bias in the empirical work.

CHAPTER 2
BACKGROUND AND LITERATURE REVIEW

Legacy Policy Debate

Recent discussions about affirmative action have contained criticisms of legacy admit policies. In 2004, President Bush gave a speech before a journalism convention, and questions about affirmative action quickly shifted to legacy admits. The President quipped about his own family ties to Yale, but ultimately said universities should stop giving preference to legacy admits (Goldstein 2004). Prominent Democrats share the Republican president's position. Senator Edward Kennedy (D-MA) submitted wording into the College Quality, Affordability, and Diversity Improvement Act (S1793) to require colleges to disclose information about legacy admits, and John Edwards vowed to eliminate the use of legacy policies when he made his bid for President (Schmidt 2004).

Despite the strong political support against legacy admit policies, there is little economic reasoning and almost no empirical support for any claims about legacy admits in the literature. The main assertion in favor of legacy admits is financial.[1] William R. Fitzsimmons, dean of admissions and financial aid at Harvard, defends the school's legacy policy because it helps raise funds that

[1] Although alumni contributions do not go directly to USAFA, the Academy's Association of Graduates (AOG) does use alumni contributions to fund some cadet activities at the Academy superintendent's discretion. In order to verify the predominant claims in the literature about legacy donations, the AOG was approached but refused to make data available for this study.

"make it possible for Harvard to admit many students from moderate or low-income backgrounds" (Schmidt 2004, p.A1). His argument is echoed by Yale University President Rick Levin (Lassila 2004). Opponents say legacy policies go against a merit-based system and can freeze out qualified applicants (Goldstein 2004). Several schools reviewed by Schmidt (2004) claim legacy policies are not sufficient for admission and legacy admits perform at least as well as their peers.

From an economic perspective, the proponents of legacy policies use an efficiency argument: allowing legacy admits increases the total resources of the school, which allows more students overall to attend the university. Critics tend to focus on the equity of legacy policies. Neither argument is addressed directly in the economics literature, and very little data are publicly available to support the claims of either side. More importantly for this study, there are no articles that discuss the potential information content of legacy status.

Air Force Academy Experience

There are unique aspects of the Air Force Academy that make it different from other universities. On the academic side, students must complete all graduation requirements within a four-year period (eight semesters), and the core curriculum is sufficiently technical that all graduates receive a bachelor of science degree regardless of major. In addition to military training throughout the year, all students are required to participate in intercollegiate or intramural athletics and take two physical fitness exams each semester.

Perhaps the most striking differences observed by outsiders are the structured environment and social life at the Academy. Cadets have a very regimented schedule during the week, and weekends can involve inspections, parades, military training, or home football games (which all cadets are required to attend). Cadets must have a pass in order to leave the Academy, but enjoying a pass may be difficult because the South Gate (leading to Colorado Springs) is almost eight miles from the cadet area, and cadets are not allowed to own or maintain a vehicle in their first two years (and sometimes not in the third year).

Given the myriad of requirements and restrictions, students at the Air Force Academy face a combination of intellectual, physical, and emotional challenges that are not present at most other universities. Any additional information a student possesses about these challenges prior to attending the Academy could help deal with the added hardships. Motivation or understanding provided by alumni parents could also help. Therefore, the impact of legacy status on student success could be more significant at the Academy than it is at other universities.

Air Force Academy Admissions

As with any university, the exact admissions process for the Air Force Academy is a guarded procedure. The description here is a purposely vague summary based on information provided by the Associate Director of Admissions. Note that in addition to satisfying the Academy's admissions guidelines, applicants must be nominated by their U.S. senator or representative.[2]

Each applicant is awarded an overall admissions score that uses a weighted compilation of SAT/ACT score, PAR score, extracurricular activities,

[2] There are several other nominating sources, but they only apply to a small fraction of applicants. Data were not available to determine the impact of legacy status on the nomination process. Arguably, legacy applicants are more informed and better prepared to deal with the process because of their parent's experience. Although this could have implications for the pool of applicants and acceptance rates, these issues are not the focus of this study.

leadership qualities (e.g., team captain vs. team member), and a subjective assessment. The PAR score is an Academy-generated measure based on high school GPA, class rank and size, percentage of graduates going on to higher education, rigor of curriculum, and average number of academic courses taken per semester. Not all the data are available for all applicants, so PAR score is somewhat subjective, but it is a powerful tool that consolidates all high school academic performance into a single measure that also captures high school and neighborhood specific effects.

The subjective assessment includes an evaluation from the liaison officer who helps the applicant through the process, comments from teachers, letters of recommendation, and a writing sample from the applicant. In addition, some credit is awarded for legacy status.[3] Despite these extra points, the Associate Director of Admissions was emphatic that all applicants who are accepted to the Academy, whether legacy or not, meet all admissions guidelines. Summary statistics similar to Maloney and McCormick (1993) are displayed in Table 2-1. Unlike their results, which revealed significant differences between athletes and non-athletes at Clemson, there is little practical difference (and no statistical difference) between legacy and non-legacy admits at the Air Force Academy. Figures 2-1 and 2-2 emphasize the similarity between legacy and non-legacy admits.

[3] The exact number of points is not important for the purposes of this study. Schmidt (2004) and Pruden (2004) review the legacy policies of several public and private universities. A typical public university's legacy policy awards 4 points on a scale of 100.

The use of legacy consideration at the Air Force Academy is different from most other schools, which makes it ideal for this study. As noted earlier, many schools use legacy admits to loosen alumni wallets. Alumni funding issues are not a concern at the Academy, which allows this study to look at non-monetary effects. Also, overall performance is a great concern for the service academies, since the graduates will go on to serve in the armed forces. The institutions want to use all the information available during the admissions process to ensure the best crop of new officers. Each applicant who is admitted and fails to graduate is one less officer the Air Force will have for that year group. This implies that a good measure of success for the admissions board at the Air Force Academy is the graduation rate of each class.

Legacy Admit Literature

There is very little analysis of the impact of legacy policies in either the economics or education literature. The only explicit references to legacy policies are found in education articles, but these give descriptions of the practice rather than any analysis.[4] Perhaps the closest area of study is the theoretical literature on the transfer of human capital.[5] There are also many empirical papers dealing with parental impacts on their children's outcomes and papers that address student achievement directly.[6] While somewhat dated, Havemen and Wolfe (1995) provides a review of many earlier studies that look at educational choices

[4] See, for example, Pruden (2004), Sanoff (2004), Schmidt (2004)

[5] Becker and Tomes (1986), Coleman (1988), Benabou (1996), Shea (2000), Black, Devereux and Salvanes (2003), Oreopoulos, Page and Stevens (2003)

[6] Coelli (2004) references Shavit and Blossfeld (1993), Haveman and Wolfe (1995), Duncan and Brooks-Gunn (1997), Mayer (1997), Levy and Duncan (2000), and Shea (2000)

and attainments. The "return to schooling" measures in their review and most of the literature since then cover a wide array of topics including high school completion,[7] grades or test scores,[8] college acceptance or completion,[9] post-graduate earnings,[10] and criminal behavior.[11] Statistical discrimination is another area that is applicable to the study of legacy admissions policies. There are several papers that address how firms use easily observable characteristics, such as educational attainment, to forecast performance and then rely less on these signals as they observe actual performance.[12] Other names for statistical discrimination in the case of educational attainment include "screening theory" and "sheepskin effects."

Lentz and Laband (1989) and Laband and Lentz (1992) come closest to investigating legacy issues. They argue for intergenerational transfers of career-specific human capital that motivate children to pursue the same careers as their parents. The 1989 paper uses a logit model to estimate the probability of acceptance into medical school and concludes acceptance is more likely for

---

[7] Eckstein and Wolpin (1999), Sander and Krautmann (1995), Evans and Schwab (1995), and Coelli (2004)

[8] Maloney and McCormick (1993), Betts and Morell (2000), Cascio and Lewis (2005)

[9] Blanchfield (1972), Corazzini, Dugan and Grabowsky (1972), Bishop (1977), Datcher (1982), Fuller, Manski and Wise (1982), Dolan, Jung and Schmidt (1985), Lentz and Laband (1989), Laband and Lentz (1992), Sander and Krautmann (1995), Evans and Schwab (1995), Light and Strayer (2000), Coelli (2004)

[10] Datcher (1982), Daymont and Andrisani (1984), Bound, Griliches and Hall (1986), Hungerford and Solon (1987), Jones and Jackson (1990), Card and Krueger (1992), Laband and Lentz (1992), Kane and Rouse (1993), Loury and Garman (1995), Behrman, Rosenzweig and Taubman (1996), Brewer, Eide and Ehrenberg (1999), Shea (2000)

[11] Thornberry, Moore and Christenson (1985)

[12] Lazear (1977), Hungerford and Solon (1987) Altonji and Pierret (2001), Epple, Romano and Seig (2003), Autor and Scarborough (2004)

children of doctors. The latter paper uses a similar model and gets the same result using data for lawyers. This paper also concludes that sons of lawyers are more likely to graduate law school and make more money as lawyers than other lawyers do. More importantly, the second paper specifically looks at whether or not lawyer parents talk about their careers with their sons. Having a parent talk about being a lawyer is more important than merely having a parent that is a lawyer.

There are several theoretical papers that examine university policies.[13] The general model in this dissertation is closest to the one developed by Epple, Romano and Seig (2006), which shows how schools use color-blind signals of race to achieve diversity goals. Fryer, Loury and Yuret (2003) also develop a similar model that focuses on optimal admissions policies from the perspective of the university. There are other sequential admissions models in the literature, but none of them explicitly model differences between students.[14] Most other theoretical models of college admissions focus on supply and demand constraints, and are not as closely related.[15]

While this dissertation builds on previous work, it is unique for several reasons. First, the focus of this paper is purely on the signals observed by the admissions board. This is to resolve the question of whether legacy status is a valid signal of potential success. Another unique aspect is the focus on various

[13] Rothschild and White (1995), Winston (1999), Ehrenberg (1999), Epple, Romano and Seig (2003)

[14] Olmstead and Sheffrin (1981), Fuller, Manski and Wise (1982), Eckstein and Wolpin (1999)

[15] Radner and Miller (1970), Tuckman (1971), Corazzini, Dugan and Grabowsky (1972), Willis and Rosen (1979), Brewer, Eide and Ehrenberg (1999)

post-educational performance measures: major selection, career field selection, time in service, and Air Force rank. These are admittedly unique to service academies, but they are potentially better than the common use of wage, which Daymont and Andrisani (1984) show is very dependent on major selection. Finally, this dissertation addresses the potential bias of trying to use empirical results based on enrollment data to evaluate admissions policies. This is done formally with a theoretical model and with numerical examples.

Table 2-1. Legacy Admit Summary Statistics

Legacy Admits

	Obs	Mean	Std Dev	Min	Max
SAT Score	449	1309.53	95.32	1040	1580
PAR Score	449	648.25	96.04	425	804
High School GPA	405	3.78	0.39	2.42	4.91

Non-legacy Admits

	Obs	Mean	Std Dev	Min	Max
SAT Score	13891	1297.54	98.68	860	1600
PAR Score	13891	653.52	92.28	354	809
High School GPA	11791	3.80	0.37	2	5

Notes:

- Table is based on the classes of 1994 to 2005 from the Air Force Academy
- Zero values are not included, nor are the 730 records identified as bad data (see "Data" section of Chapter 3). Including the bad data does not change the result that there is no statistically significant difference between legacy and non-legacy admits.
- SAT Score is either (i) the sum of a student's math and verbal scores, using recentered scores for high school classes prior to 1996 or (ii) the converted composite ACT score based on formulas from *The College Board* (see Appendix A).
- High School GPA only includes values from 2 to 5.
- Simple means tests show no statistical difference between the mean value for legacy and non-legacy admits in each category. Two-sample Wilcoxon rank-sum tests suggest no difference between legacy and non-legacy admits for PAR scores and high school GPAs, but a statistically significant difference for SAT scores.
- See "Data" section of Chapter 3 and Appendix A for clarification on data issues.

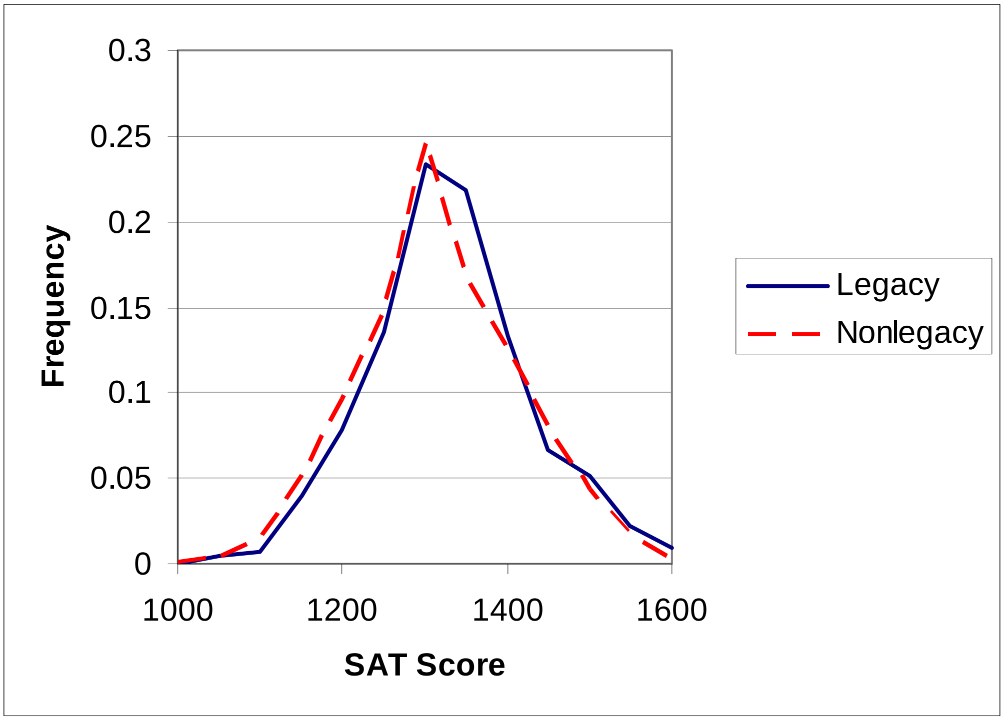


Figure 2-1. SAT Scores for Legacy and Non-legacy Admits

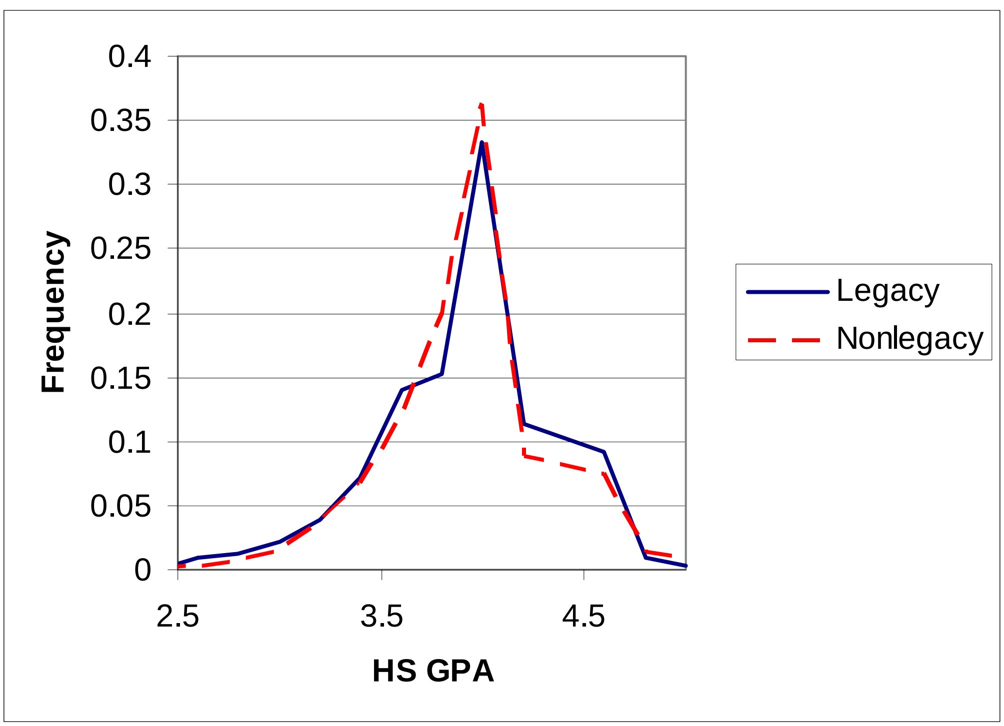


Figure 2-2. High School GPA for Legacy and Non-legacy Admits

CHAPTER 3
TRADITIONAL EDUCATIONAL MEASURES

This chapter studies the effects of legacy status on educational outcomes at the U.S. Air Force Academy. Colleges may use legacy status as a signal for potential student success and/or potential monetary contributions (from the parent). A theory is developed which claims legacy status is a signal of student success when monetary contributions are not a factor. Empirical data from the graduating classes of 1994 to 2005 are used to verify the assertion. While legacy status has no significant impact on grades or order of merit, it is associated with a 0.10 increase in the probability of graduation and a military performance average that is 0.04 points higher. This result is robust to model specification, and the increased graduation rate stems from legacy admits being less likely to voluntarily quit. While the results may not generalize to all universities, they are likely to be similar for other demanding, high-skill professions such as medical school or PhD programs.

Theoretical Framework

There are two aspects to understanding a legacy policy: the university and the student. Presumably, the university has specific objectives in mind when designing its policies. In order to incorporate legacy status into these policies, there must be knowledge of how legacy status makes a student different from his or her peers. This section provides a conceptual theory for legacy status. Chapter 5 develops a formal theory.

University Objective

Several economic models explain university behavior, and almost all use some type of utility maximizing framework. There can be many components of a school's objective function (e.g., diversity in race, gender, geography, income, proposed major, etc.), but for the most part a college is looking to select students with strong academic backgrounds who have a reasonable chance of success at the university. Epple, Romano and Seig (2003) develop a theoretical model of college admissions, with and without affirmative action, in which schools want to maximize a quality index that increases with academic qualification of the student body. The authors limit diversity to race and income and conclude that a school with a preference for racial diversity will employ alternative signals of race (i.e., income) to satisfy its goals if it is prohibited from using affirmative action (i.e., using race blind admissions). This result suggests that schools will use any signals legally available to them in order to achieve their objectives.

Assume a university wants to maximize the academic quality of its students. The exact measure is not important, but it could be the graduation rate, the average GPA, the percentage of graduates who go on to graduate school, or the average starting salary of graduates. To attain this objective, the admissions board is limited to observable student characteristics. Typical measures include high school performance and college entry exam scores, but these are noisy indicators of a student's potential performance, especially at selective colleges, because high school is not necessarily a challenging experience for top students. Standardized tests mitigate some problems with high school data, but these exams only measure intellect; they do not reflect work ethic, maturity, or other factors that are important in determining college success. Unfortunately, these other factors are rarely observable. Many schools attempt to capture these unobservable, non-academic factors with extracurricular activities or letters of recommendation. These measures have limited value because students join clubs for "square filling," and only request letters of recommendation from people who will write favorable ones. One factor a student cannot manipulate is legacy status. An investigation into the nature of legacy status can determine if it is a valid signal for the student's future college performance.

Student's Legacy Status

Most of the economics literature identifies parental effects through their educational attainment or household income.[1] Although they do not consider legacy status, these studies do provide a framework for analyzing the impact of legacy status. There are two ways legacy status can affect a student's performance: genetic and cultural.

The genetic argument for parental effects says a child's performance is a function of breeding or innate ability inherited from the parents' genetic code. This is an argument about the student's overall quality, which is found to be more important than cultural aspects by Black, Devereux and Salvanes (2003). Unfortunately, testing this result is difficult because students can choose to not graduate for non-performance related reasons.

The second avenue for parental impact comes from the interaction between the parent and child. The parent may impart school-specific information or a level

---

[1] Datcher (1982), Lentz and Laband (1989), Black, Devereux and Salvanes (2003), Oreopoulos, Page and Stevens (2003)

of motivation or maturity that helps the student succeed more than peers who do not have such a benefit. The information shared by the parent could ensure a better fit between the student and the college. Light and Strayer (2000) find students have higher chances of graduating if the quality level of their college matches their observed skill level. For legacy admits, one could argue that information passed by the parents ensures a better fit. The information could also better prepare or motivate the students so they are more likely to succeed than their non-legacy peers.

These theories can be tested empirically. Although the causal mechanism of legacy status (genetic vs. cultural) cannot be determined with the available data for the Air Force Academy, the impact on student performance can be observed through graduation rates, GPA, MPA, and order of merit. To consider all aspects, non-graduates can be divided into those who leave because of grades and those who leave for other (non-academic) reasons. Based on the cultural arguments of motivation passed from alumni parents and better fit between student and school, legacy admits should be less likely to drop out for non-academic reasons. The quality (genetic) and preparation (cultural) arguments predict legacy admits will be less likely to drop out for academic reasons and they should have higher grades than non-legacy admits. Therefore, the overall theory that legacy status provides valuable information to admissions boards can be confirmed if legacy admits are more likely to graduate and have better grades than their peers.

Empirical Strategy

Several different models are needed to confirm the predictions of the theoretical framework, but all are built on the basic model which uses each student's admissions data to predict some performance characteristic:

$$\text{Performance} = \mathbf{x}'\boldsymbol{\beta} + \gamma\,\text{Legacy} + \varepsilon \qquad (3\text{-}1)$$

where $\mathbf{x}$ is a vector containing:

SAT_Score
Math_Ratio
PAR_Score
Intercollegiate
Prior
Other_Academy
Military_Background
Dummies for gender, race, AFA class year, and high school state
Constant term

Four different performance measures are considered: probability of graduation, GPA, MPA, and order of merit. Graduation is considered first. It is a binary variable so a probit model is used.[2] Ideally, the vector $\mathbf{x}$ would contain all the measures used by the admissions office. See "Threats to Identification" later in this chapter.

The SAT_Score measures overall ability, so higher scores are expected to result in higher performance.[3] The total score combines two different types of

---

[2] For graduation probability, the model (3-1) is modified to be a probit as follows:

$$\Pr[\text{AFA\_Grad} = 1 \mid \mathbf{x}, \text{Legacy}] = \int_{-\infty}^{\mathbf{x}'\boldsymbol{\beta}+\gamma\text{Legacy}} \phi(t)\,dt = \Phi(\mathbf{x}'\boldsymbol{\beta} + \gamma\text{Legacy})$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are the density and cumulative distribution of a standard normal distribution. The difference between probit and logit are inconsequential for this data set. Probit is used for computational simplicity because Stata automatically computes marginal effects. An OLS linear probability model for graduation probability also gives similar results.

[3] The Air Force Academy only records an applicant's best standardized test score. All ACT scores are converted to their recentered SAT equivalents. See Appendix B.

scores, each measuring a different skill set. This is handled by using a process similar to Maloney and McCormick (1993), which computes the math to verbal ratio (or simply Math_Ratio). Since the academy is a technical school, the Math_Ratio is also expected to have a positive effect on performance. For example, two students who are equal in all other measures and have a total SAT_Score of 1300 are not identical if one scores 760 Math and 540 Verbal while the other scores the reverse, 540 Math and 760 Verbal. The student with the higher math score is expected to perform better.[4]

A student's PAR_Score is a single number calculated by Academy admissions that combines various high school academic measures (high school GPA, class rank and size, percentage of graduates going on to higher education, rigor of curriculum, and average number of academic courses taken per semester). The higher the score, the better the student is expected to perform at the Academy; therefore, a positive coefficient is expected.

Since a school is expected to make tradeoffs between student performance and a student's other contributions to the school (athletics, funding, publicity, etc.), the coefficient for the binary variable Intercollegiate is expected to be negative. Maloney and McCormick (1993) provide evidence that intercollegiate athletes, on average, do not perform as well academically as non-athletes, even after controlling for high school grades and SAT scores.

---

[4] An interaction term between SAT score and math ratio could be added to allow the impact of the ratio to vary for different SAT scores. The result is negative, meaning the ratio is not as important for higher scoring students. Using the interaction does not affect the coefficient of legacy status, but it adds unnecessary complexity to the interpretation of the results.

Similar studies are not available for prior enlisted military members. Arguably, these students are more mature and thus should perform better. However, they have more time between graduating high school and entering college and could forget some of the academic knowledge and skills required to succeed. Therefore, the coefficient for Prior is ambiguous.

Given the hypothesis that legacy status provides positive information, the coefficient for Legacy should be positive. An interesting comparison is the coefficient for Other_Academy, a dummy variable for all other service academies. Although the parents of these students did not experience the exact same environment as parents who attended the Air Force Academy, the other service academies are similar, so the Other_Academy students may have similar advantages. Theoretically, then, the coefficient should be positive and similar to Legacy. Another interesting test of the theory is the dummy variable Military_Background, which equals one if either of the student's parents has military experience, not including graduates from service academies. This is an approximation of the military component of the effect of legacy status (other portions being specific to the Academy culture). The coefficient is expected to be positive but smaller than Legacy.

Variation 1: Nonlinear Relationships (Splines)

According to a source at the Air Force Academy, internal studies show nonlinear relationships between student performance and the student's SAT and PAR scores. As the scores increase, student performance improves, but only to a certain point, above which higher scores do not affect performance. A piecewise linear, continuous function (spline) is used for SAT_Score, Math_Ratio, and PAR_Score, using a technique similar to Lott and Kenny (1999).[5] That is, for each variable, the slope is allowed to change discretely at a specific value, creating a kink. For example, the SAT_Score variable is replaced with two new variables:

$$\text{Low\_SAT} = \begin{cases} \text{SAT\_Score} & \text{if SAT\_Score} \leq \text{S} \\ \text{S} & \text{if SAT\_Score} > \text{S} \end{cases} \tag{3-2}$$

$$\text{High\_SAT} = \begin{cases} 0 & \text{if SAT\_Score} \leq \text{S} \\ \text{SAT\_Score} - \text{S} & \text{if SAT\_Score} > \text{S} \end{cases} \tag{3-3}$$

where S is the kink. An automated search is performed for the cutoff value for all three variables simultaneously, in order to get the best fit for the model based on the log likelihood value. The optimal kinks occur at 1280 SAT_Score, 0.97 Math_Ratio, and 600 PAR_Score.[6]

Variation 2: Student Quality (Quartiles)

The probit model using the three splines gives an estimate of the contribution of legacy status overall, which answers the question of whether legacy status provides useful information about graduation probability to an admissions board. Although not specifically addressed by the theoretical framework, legacy status may affect different types of students differently. To resolve this question, the data is broken into distinct subgroups by using the

[5] Several techniques can model the nonlinear effect of these variables. A quadratic model has significant squared terms which verifies the nonlinearity, but the model is fairly restrictive and does not fit the data as well as the spline model does. Dummy variables also work, but they do not ensure a continuous relationship. (There is no reason to believe performance jumps or falls dramatically for a specific value of any of these variables.) These alternative specifications do not have a substantial impact on the effect of legacy status.

[6] The search includes over 10,000 regressions that systematically vary the pivot point for all three variables. The ranges investigated are: 1200-1400 for SAT_Score, 0.90-1.20 for Math_Ratio, and 550-700 for PAR_Score.

intersection of the bottom quartiles of both SAT and PAR scores and the intersection of the upper quartiles.[7] The intersection of the bottom quartiles (which turns out to be about 10 percent of the data) attempts to isolate students for whom legacy status plays a larger role in the acceptance decision. The result could support or counter the equity argument against legacy policies.

Variation 3: Quitting vs. Failing (Mlogit)

In order to verify the individual predictions of the cultural view of legacy status, it is necessary to break down students who do not graduate into two groups: those who fail and those who quit. This information is not directly available in the data, but it can be estimated by using AFA_GPA. Anything less than 2.0 is a failing GPA at the Academy, so any non-graduate with AFA_GPA between zero and two is labeled as someone who failed (or quit because of academics). Non-graduates with AFA_GPA equal to zero drop out before grades are issued in the first semester, so they are assumed to leave the Academy for non-academic reasons. Similarly, non-graduates with AFA_GPA of 2.0 or better are assumed to quit for non-academic reasons. An unordered multinomial logit model is estimated to explain how legacy status impacts the decision to graduate, quit, or fail.[8] Greene (2003) describes a formal test of the mlogit's Independence from Irrelevant Alternatives (IIA) assumption as specified by

---

[7] Quartiles are used to keep the sample size sufficiently large for statistical significance. Intersecting the top and bottom deciles is more dramatic, but the sample size drops below 400 observations, so the estimates are insignificant unless the state fixed effects are removed.

[8] The switch from probit to logit is used for convenience because Stata has an mlogit function, but no equivalent procedure for probit.

Hausman and McFadden (1984). This test is performed using the *suest* command in Stata to verify that IIA is satisfied.

Variation 4: Other Performance Measures: GPA, MPA, and OM (OLS)

Grades, military performance, and order of merit are other measures of student performance which can be estimated by the model in (3-1). The dependent variable is replaced with AFA_GPA, AFA_MPA, or AFA_OMp, and the data are restricted to graduates only. The latter measure is order of merit as a fraction of class size, which means lower numbers are better, so the expected signs of the coefficients are reversed. Since the new dependent variables are continuous, simple OLS estimation can be used.[9]

Data

Data for every cadet from the classes of 1994 through 2005 come from the Academy's Plans and Analysis Division, with considerable collaboration with the Admissions office.[10] Some of the fields in the data set are supplied to the Academy by the Air Force Personnel Center. There are a total of 15,070 records, each containing information on Academy performance, high school performance, and legacy status. The data also contain each graduate's Air Force status as of July 2005. Summary statistics for variables used in the empirical model are included in Table 3-1, and a complete description of the variables is in Appendix A.

---

[9] Technically, the predictions must be constrained to the [0,4] and (0,1] intervals in order for OLS to be valid. For AFA_GPA and AFA_MPA all predictions are within the correct interval; for AFA_OMp, all but one are.

[10] USAFA/XPX and USAFA/RRS. Based on the agreement for the release of data, the author is not permitted to share the data.

Given the long period of time, the complexities of data passed between multiple organizations, and inevitable coding errors, the data set is not perfect. The Academy is aware of the errors but does not have the resources to investigate data issues. Individuals can only be identified by class year and order of merit, so outside data verification is not possible.[11] In order to ensure more accurate results, general rules are used to reduce the possibility of corrupt data in the analysis. If there are obvious errors for a particular field, the entire record is suspect and not included in the analysis. Missing information also makes a record questionable, so records missing a variable are also removed as long as the number removed for each variable is less than one percent of the data.[12]

High school data are considered first. There are 18 records missing high school state and 36 with either missing or invalid high school year.[13] There are also 15 records with possible errors in high school size because they list over 1,500 students in the graduating class.[14] There are many records missing either SAT or ACT score because the Academy only records an applicant's best score. After combining SAT and ACT scores, there are only six records missing a standardized test score (see Appendix B).

---

[11] Although not a scientific sample, personal contact with five Academy graduates revealed no major discrepancies in their records.

[12] An alternative method, used by Attiyeh and Attiyeh (1997), is to substitute the average value for the variable and create a dummy variable equal to one if the value is missing. This technique is more appropriate when there are many records missing the same field. It is used in some of the alternative specifications to test for robustness.

[13] Examples of invalid high school year include 618 and 1900.

[14] These schools were contacted to verify the class sizes, but only one school replied, which updated the class size from 8181 to 80. An attempt to download school sizes from the U.S. Department of Education's National Center for Education Statistics was also unsuccessful.

High school rank as a percentage of class size can only be calculated if both rank and class size are available. There are 2,973 records missing one or both of these measures. There are also many problems with high school GPA, since the values range from 0.04 to 9.98. There are 83 records between 0 and 2 and 370 records above 5. In addition, there are 1,832 records with missing GPA. The number of records with these errors is too large to simply eliminate the data, so PAR score is used in lieu of high school rank and GPA. This substitution eliminates the data problems because there are only nine records missing PAR score. In addition, the use of PAR score is more appropriate because it is the measure used by the Academy admissions office to capture high school performance.[15]

There are several filters that are applied to Academy and Air Force data to identify problems. First, graduates from the Academy must maintain at least a 2.0 GPA. There is one record for which this is not the case. Similarly, graduates must maintain a 2.0 MPA. There are 3 records that do not and 18 records with MPA values greater than 4.0. All graduates incur a service commitment of at least five years. There are legitimate reasons for someone to leave the Air Force before the commitment expires, but there is no way to identify these cases with this data set. Therefore, all records for graduates prior to 2002 with less than 3 years in service are labeled as bad data (197). Another problem is graduates whose time in service does not correspond to rank. Promotions for junior officers are based primarily on time in service, so the time should coincide with the appropriate

[15] One of the robustness checks uses high school rank and GPA instead of PAR score. The change in the marginal effect of legacy status is inconsequential.

rank. Two filters are used: Second Lieutenants with more than 4 years service (127) and First Lieutenants with more than 6 years service (26). These records are labeled as bad data.

Bad data for non-graduates are identified by looking at any records for non-graduates that have positive years of service or valid Air Force Specialty Codes (AFSCs). Although there is the possibility that non-graduates have to serve in the military to repay their commitment, they typically serve as enlisted troops, and all the ranks listed are for officers. There are 310 bad records based on these criteria.

The final filter applied is to drop data with missing demographic data. Only two records fall under this category.

The filters applied on the data are summarized in Table 3-2. They are not mutually exclusive, so the total number of records removed is 730, which accounts for less than 5 percent of the 15,070 observations. Not all of the filters apply directly to the empirical model (i.e., they do not directly affect variables in the model). The purpose of these filters is to ensure higher quality results by eliminating data that are known to have errors.[16]

Empirical Results

Graduation Rate

Results for the probit model with the three splines are presented in Table 3-3. The marginal effect of legacy status on graduation probability is very

[16] The filters do not drive the results. There is no substantial difference between the means and standard deviations of each variable using "good" and "bad" data. In addition, the models described in the previous section are run with and without these filters and with additional filters. The marginal effect of legacy status remains nearly identical in all cases.

significant both statistically (better than 1%) and practically (a little more than 10 percentage points added to the probability of graduating). To put this in perspective, note that legacy status has a more substantial impact than gender or any of the race controls. Compared to SAT scores, being a legacy admit is equivalent (in terms of impact on graduation probability) to just over 230 points, which is greater than two standard deviations for SAT scores.[17] Similarly, legacy status corresponds to 88 points in the student's PAR score.[18] This is almost as much as a standard deviation for PAR score.

The other variables of interest have the expected signs. SAT scores increase the probability of graduation by almost half a percentage point for each ten points on the SAT up to 1280 (i.e., Low_SAT). Above 1280 (High_SAT), SAT scores are no longer statistically significant at the five percent level, but even so, the point estimate is negative and nearly a quarter of the impact of the lower SAT scores. A one standard deviation improvement in SAT score increases the probability of graduation by 4.4 percentage points. This is the maximum improvement assuming the SAT score remains below 1280.

Similarly, increased Math_Ratio greatly improves the probability of graduating up to the pivot point of 0.97. A one standard deviation improvement in Math_Ratio below the pivot point (i.e., Low_Math_Ratio) increases the likelihood of graduation by 4.2 percentage points. For example, given two identical students with total SAT scores of 1260, a student with 660 verbal and 600 math is roughly

---

[17] The point equivalence is found by dividing the Legacy marginal effect by the Low_SAT marginal effect: 0.0136903/0.0004474 = 238.08.

[18] The PAR equivalence is found by dividing the Legacy marginal effect by the Low_PAR marginal effect: 0.0136903/0.0011817 = 87.81.

4 percentage points more likely to graduate than a student with 700 verbal and 560 math (Math_Ratio 0.9 versus 0.8). For ratios above 0.97 (High_Math_Ratio), however, improved math scores relative to verbal scores no longer matter. This suggests students with math skills at least as good as their verbal skills are most likely to succeed at the Air Force Academy.

PAR score is the Academy's best internal predictor of academic success at the Academy. Based on the marginal effects in this model, a one standard deviation increase in PAR score (92 points) increases the probability of graduation by almost 11 points. This relationship holds up to a PAR score of 600 (i.e., Low_PAR), above which the impact of increased PAR score is not as strong. For High_PAR, an increase of one standard deviation only increases graduation probability by 4.5 percentage points. Note that the effects of PAR score are much greater than SAT or math ratio.

The non-academic variables for intercollegiate athletics and prior enlisted status do not have a statistically significant effect on graduation rates. Other specifications such as a basic linear model, a probit without splines, or including high school GPA and rank instead of PAR score occasionally result in significant intercollegiate and prior status. Regardless of significance, the marginal effects are always negative for Intercollegiate and positive for Prior. Both variables are sensitive to model specification so their impact is uncertain, but in all models the effect of each is smaller than those of the academic characteristics.

Perhaps more interesting than the traditional predictors of performance are the two variables most closely associated with legacy status: Other_Academy and Military_Background. Students whose parents attended another service academy have nearly the same advantage as the legacy admits: roughly 11 percentage points more likely to graduate. A military background has a marginal effect of almost two percentage points,[19] which suggests the academy culture imparted by the parents is more significant than the military background instilled in the students.

The other variables in the model are dummy controls for class year, high school state, race, and gender. They are included to absorb variation in the data, and their interpretation is not the primary focus of this study.

The Air Force Academy is about more than just academics (see Chapter 2). All specifications result in a statistically significant regression, but they do not have a lot of predictive power. For the probit in the first column of Table 3-3, for example, the pseudo R^2 is only 0.0455. Attiyeh and Attiyeh (1997) look at predictive accuracy by comparing their estimated model to a naïve model. In this case, a naïve model is one that predicts everyone graduates because the median for AFA_Grad is greater than 0.5. The probit model only improves predictive accuracy by 0.43 percentage points. This results from the fact that many highly qualified students at the Academy choose to not graduate. In fact, there are two people in the data set with 1600 SAT scores who did not graduate. After adding converted ACT scores, the graduation rate for students with perfect test scores is only 80 percent, which is not much higher than the overall average

[19] In all the models discussed in this paper, the point estimate for the marginal effect of Military_Background ranges from 1.7 to 2.1 percentage points. The result could be different if the variable were divided between enlisted and officer parents, or by career (20 years of service) versus non-career parents, but data are not available at that level of detail.

of 74.6 percent. To emphasize the point that academic success does not necessarily translate to graduating, note that seven students in the data set have a perfect 4.0 GPA at the Academy, and none of them graduated.

Marginal Students

The main concern for opponents of legacy policies is that awarding the extra points may eliminate qualified candidates from consideration. In order to test this assertion, one would need to clearly identify marginal students who are accepted by the margin of the points awarded by legacy status. Such data are not available, so an alternative is to look at students in the bottom of the academic qualifications. The second and third columns of Table 3-3 show the probit output for the intersections of the lower and upper quartiles based on SAT and PAR scores.[20] The lower quartile intersection only includes students whose SAT scores are 1230 or lower and PAR scores are 578 or lower. "Quartiles" seems misleading here because the actual amount of data in the intersection is roughly 10 percent (1490 of 14340). The cutoffs for the upper quartiles are SAT scores above 1370 and PAR scores above 726. In both cases, the kinks in the splines fall outside the cutoffs, so only one side of the spline is used in each probit model. (The computer automatically drops the other variable.)

The results for these models at first do not appear as strong as the model with the full data. The variables for SAT scores and math ratios, which are significant with all the data, are not significant for the smaller subsets, primarily

---

[20] The bottom 10 percent may be a better cut off, but the sample size is too small (353), and only Low_PAR is significant. If all fixed effects are removed, the lower cutoff results in a substantial marginal effect of legacy status of 26.0 percentage points. If the fixed effects are removed from the full data set, the marginal effect of legacy is basically unchanged.

because of the smaller sample sizes. Other_Academy is also strongly significant with the full data, but loses its significance in the lower quartiles model and is dropped completely in the upper quartiles model. In the latter case, the variable perfectly predicts graduation, so the variable is automatically dropped because there is no variation in graduation success. For the lower quartile, Other_Academy does not appear important because there are so few students with parents from other service academies in the intersection of the lower quartiles.

Despite the loss of significance for many control variables, legacy status is the primary focus, and the new probit results show a dramatic impact. For the intersection of the upper quartiles of students, legacy status does not have a significant effect on graduation. For students in the lower quartiles, however, being a legacy admit makes graduation 18.2 percentage points more likely. As with the full data set, that figure is equivalent to one standard deviation (92 points) in a student's PAR score. A comparison to SAT score is not valid because Low_SAT is not significant.

There is also a substantial improvement in the predictive accuracy of the lower quartiles model relative to a naïve model. With the full data, the spline probit model only improves predictions over a naïve model by 0.43 percentage points. This figure jumps to 3.41 for the lower quartiles and drops to 0.31 for the upper quartiles. Since most of the other variables lose their significance in the smaller models, the change in predictive accuracy may be caused by the change in the impact of legacy status.

To drive home the point, consider the overall graduation rates for legacy and non-legacy admits for the full data set: 84.4 versus 74.3 percent for legacy and non-legacy admits, respectively. When looking at the upper quartiles model, this gap narrows: 86.9 versus 81.1 percent. At the lower end, however, it widens considerably: 79.1 versus 60.8 percent. It seems the motivation or preparation of alumni parents has a greater impact for more academically-challenged students. Since legacy status contributes so much more to the probability of graduation for marginal students, there is little evidence to support the claim that the legacy policy prevents otherwise qualified students from being admitted.

Quitting vs. Failing

Several possible explanations for why legacies outperform non-legacies are presented in the "Theoretical Framework." A multinomial logit model is used to distinguish how legacy status influences the probability of not graduating for academic or non-academic reasons. For simplicity, these events are referred to as failing and quitting, respectively. Normally, mlogit coefficients are not easily interpreted because the marginal effect of any one variable is dependent on the coefficient of all the variables.[21] Table 3-4 shows the results of the marginal effect command (newly available in Stata 9) for the mlogit procedure. The table shows the marginal effect of legacy status on graduation probability is nearly identical to the result of the probit model: 0.1037.

The advantage of this method is that it shows how the increased probability breaks down between the likelihood of failing and quitting. The third column

[21] See Greene (2003).

shows that nearly all of the improvement comes from legacy students being less likely to quit. From the 10 percentage points improvement for graduation, 9 points come being less likely to quit, and 1 point comes from being less likely to fail. Similar results could be listed for the other explanatory variables, but that would detract from the purpose of this section.

Another way to look at the breakdown is to follow the procedure identified by Greene (2003) and the Stata 7 reference manual. This method was used prior to software advances and has its weaknesses because it does not provide a standard error, but it does provide an informal test for the orthogonality of legacy status. "Adjusted" probabilities for graduating, failing, and quitting are computed for both legacy and non-legacy admits. The probabilities come from the mlogit predictions, first assuming all students are legacy admits (i.e., Legacy = 1) and then assuming they are non-legacy admits. These probabilities are "adjusted" because they account for the other control variables.

The "adjusted" probabilities are shown on the right side of Table 3-5. The difference between these probabilities determines the marginal effect of legacy status. As with the original probit model, the marginal effect on graduation probability is roughly a 10 percentage point increase. The marginal effects of legacy status on the probability of failing and quitting show how those 10 points break down. Legacy status has a much larger impact on quitting than on failing. Legacies are 8.9 percentage points less likely to quit than non-legacies and only 1.5 percentage points less likely to fail. In percentage terms, the effect of legacy status seems even more substantial: legacy admits are 43.5 percent less likely to quit and 28.8 percent less likely to fail.[22]

Table 3-5 also presents "unadjusted" probabilities for graduating, failing, and quitting. These probabilities are found by simply dividing the data into graduates, non-graduates who fail, and non-graduates who quit for both legacy and non-legacy admits. Comparing the unadjusted and adjusted probabilities shows little change in the difference between legacy and non-legacy admits. That is, after adjusting for gender, race, class year, high school state, SAT score, math ratio, PAR score, intercollegiate status, and prior enlisted status, the difference in graduation rates for legacy versus non-legacy admits is practically unchanged (i.e., legacies are still roughly 10 percentage points more likely to graduate). Therefore, the impact of legacy status is orthogonal to those associated with the other control variables. This evidence supports the assertion that legacy admits possess some non-academic motivational factor not captured by other admissions data that makes them more likely to succeed at the Air Force Academy.

Other Performance Measures: GPA, MPA, and OM

Table 3-6 presents the OLS results for Academy GPA, MPA, and graduation order of merit as a fraction of class size.[23] Recall these models only

---

[22] Running individual probit models to compare graduating versus failing and graduating versus quitting yields similar results: legacies are 9.4 percentage points less likely to quit and 1.5 percentage points less likely to fail. Running the mlogit procedure for the intersection of the lower quartiles of SAT and PAR scores results in the same 8 to 1 quit/fail ratio even though the probabilities themselves nearly double.

[23] The OLS results are computed using robust standard errors so heteroscedasticity is not a problem. Alternative specifications optimize the spline kinks for each dependent variable, but the

look at graduates and AFA_OMp has opposite signs because smaller numbers are better. Only the MPA model reveals any significant effect of legacy status.[24] The lack of significance for GPA is not surprising since most of the impact of legacy status on graduation probability comes from the reduced probability of quitting (rather than failing).

The marginal effect of legacy status on MPA is only 0.04 points, but it is highly statistically significant and is rather large when compared to the other variables. In terms of SAT scores, being a legacy admit is equivalent to over 200 points, more than two standard deviations. The equivalence in terms of PAR score is not as strong as the graduation model, but still large at 80 points, about 85% of a standard deviation. Despite this seemingly large impact, the legacy advantage in MPA is washed out in the order of merit model.[25]

The academic control variables have the expected signs. Higher SAT scores contribute to higher GPA, MPA, and order of merit (a lower fraction of class size). Below a score of 1280 (i.e., Low_SAT), a one standard deviation increase in SAT score (roughly 100 points) results in an increase of 0.09 grade points, 0.02 military points, and a drop of 6 percentage points in order of merit. Above 1280, the impact of SAT on MPA is cut in half and only marginally significant statistically. In contrast, high SAT scores have a bigger effect on

---

results do not vary enough to justify the potential confusion of using different kinks for each model.

[24] An alternative specification forces a logit for continuous data by running OLS on $\ln[p/(1-p)]$, where p = GPA/4, MPA/4, or AFA_OMp. The statistical significance of each variable is virtually identical, as are the signs, but the magnitudes of some marginal effects are noticeably different between the OLS and makeshift logit models. The main result remains unchanged: legacy status does not have a significant effect on GPA or order of merit.

[25] Order of merit is a weighted average of GPA, MPA, and APA (athletic performance average).

grades and order of merit. A standard deviation increase in SAT score increases GPA by 0.14 points and improves graduation order of merit by 7.5 percentage points. Given a class size of 1000 students, this 100 point increase in SAT score translates into 60 places in the order of merit for lower scores, and 75 places for higher scores.

Math ratios below 0.97 do not contribute significantly to GPA, MPA, or order of merit. Higher math ratios have statistically significant but practically inconsequential impacts. A one standard deviation increase in the math ratio above 0.97 increases GPA by 0.014 points, decreases MPA by 0.01 points, and decreases order of merit percentage by less than 0.7 points. These effects are roughly a tenth of the SAT score effects, so the math ratio does not have the same practical significance as the total SAT score.

PAR scores are just as important as SAT scores in predicting student success and similarly more important for GPA than MPA. For lower scores (below 650), a one standard deviation increase in PAR score results in increases of 0.15 points on GPA, 0.05 points on MPA, and a 10 percentage point decrease (100 places) in order of merit. These results increase to 0.17 points and 11 percentage points (110 places) for PAR scores above 650. There is no change in the marginal effect of PAR score on MPA.

One unexpected result in Table 3-6 is the coefficient for Intercollegiate. According to the results of the model, intercollegiate athletes on average have 0.03 higher GPA than comparable non-athletes. This result is different from what Maloney and McCormick (1993) find for athletes at Clemson. One possible explanation is that their study involved students while they were still in school, and the results in this study focus on students who finished school (so potentially lower performing athletes are not included). Intercollegiates have MPAs that are almost 0.07 points lower on average, which suggests the added input from the coaches does not make up for the time the cadets spend away from their squadrons during games and practices. The impact on MPA outweighs the GPA advantage for athletes because intercollegiate status is not significant in predicting order of merit.

The impact of prior enlisted status produces potentially disturbing results. These students, on average, have GPAs that are 0.14 points lower and MPAs that are 0.02 points lower than their peers. The prior enlisted cadets also graduate with order of merit 9.3 percentage points higher (93 places lower). Part of this result could be because prior enlisted students are further removed from high school, and they struggle to regain their academic skills. Another potential explanation is that students who attend the Air Force Academy Prep School are considered prior enlisted based on one year of active duty service before entering the Academy. These students attend the prep school because of lower academic preparation. A more controversial explanation could be that prior enlisted students do not think top academic performance is necessary for their careers in the "real" Air Force.

The other non-academic background characteristics, Other_Academy and Military_Background, do not have significant effects on GPA, MPA, or order of merit.

Robustness

It almost seems implausible that legacy status can have such a large impact on the likelihood of graduation. Throughout the study, many alternative specifications are tried in order to derive the correct relationship. These models include a basic linear model, a probit without splines, and a probit using high school GPA and class rank instead of PAR score. In all cases, legacy status is statistically significant and increases the probability of graduation with marginal effects ranging from 10.3 to 10.7 percentage points.

For the spline probit model presented in Table 3-3, the search for optimal kinks in the splines could be considered a robustness check. After 10,416 iterations, the marginal effect of Legacy fluctuated between 10.3 and 10.5 percentage points. This may not be a sufficient robustness check because it is the same basic model, but it does show the results are consistent over a large range of kinks in the splines.

It could be that the legacy impact is sensitive to the data used in the study. To verify such a claim, the general spline model is re-run using the entire data set. The marginal effect of legacy status on graduation probability in this case is an increase of 10.9 percentage points, not much different than omitting the bad data. Another alternative is to more aggressively eliminate potentially bad data. If the model is re-run without any records that are incomplete, the marginal effect for Legacy is still 10.4. A more dramatic test of the model's sensitivity to data is to randomly use subsets of the data. This can be done by using the PID code, a unique identifier from the Academy's database which should be unrelated to any other variables. Running the probit model for even and odd PID yields marginal effects for Legacy of 7.7 and 13.2, respectively. Both are within the 95% confidence interval for Legacy using the result from Table 3-3.

A final robustness check is a falsification test to determine the likelihood that the impact of legacy resulted from some random event. An automated procedure is established where legacy status is randomly assigned to students whose parents are not from other service academies or do not have military background (i.e., Other_Academy = 0 and Military_Background = 0). The assignment is made by generating uniform(0,1) random variables and using the overall proportion of legacy admits (0.031311). If the random value is equal to or less than this proportion, the student is labeled as a legacy admit. Others are non-legacies. The model is then re-run and the marginal effect of legacy is recorded. After 1,000 iterations, only 63 of the regressions result in a statistically significant marginal effect for legacy status. Of these, the values range from 3.70 to 7.98. This lends support to the conclusion that the strong result of 10.38 percentage points is not a random event.

Limitations and Further Research

Threats to Identification

There are several problems with the identification strategy of this empirical study. The most obvious is the use of mostly academic variables in conjunction with legacy status. As the summary of Air Force Academy admissions indicates, part of the process includes extracurricular activities, leadership qualities, and other subjective areas. These characteristics are observed by the admissions office, but are not available in the data set. There is the possibility that legacy status is capturing the impact of these unobserved variables. If the missing variables are correlated with one of the regressors (SAT score, PAR score, legacy status, etc.), there is the potential for that regressor to be correlated with the random error term. The normal solution would be to use a proxy variable in place of the omitted variable. In this case, there are no other data available.

Fortunately, these subjective measures are arguably limited in predicting student performance because of the potential lack of variability and other reasons listed in the "Theoretical Framework" section. The data only include students who were accepted to the Academy. Given the selective nature of the process and the vetting in the Congressional nomination stage, there is probably little variation in the subjective measures. Even if the subjective measures do help predict performance, they are more likely to be correlated with the other academic variables rather than legacy status. The previous section shows these academic measures are orthogonal to legacy status, so it is likely that subjective measures are also unrelated to legacy status. Unfortunately, the claim that omitted variables are not a problem cannot be verified without access to all the data used by the admissions office.

There could also be omitted variables that are not observed by the admissions office. One obvious variable that is definitely correlated with legacy status is parents' education. It could be that legacy status is simply capturing the fact that the student's parent is a college graduate. This is unlikely since the percentage of legacy admits is small. If the only contribution of legacy status is a college graduate parent, the relationship would not be as significant because many non-legacy admits would also have parents who are college graduates.

Still, it would be nice to add a control for parent's education, similar to the Other_Academy variable, to compare the effect of an alumni parent (legacy) to a parent who is a regular college graduate. If the Academy's only concern is using legacy status as a signal for student performance, the fact that legacy status could be correlated to omitted variables that are not used is unimportant. Such correlation is the whole point behind using a signal: the correlation is more important than the causality.

Selection issues are another potential problem with this study. There is a sequence of choices a student must make before entering the Academy. First, the student must choose to apply. Then, if accepted, the student must choose whether to attend the Academy. Legacy and non-legacy students may make these decisions differently. In fact, research by Lentz and Laband (1989) suggests intergenerational transfers of career-specific human capital make it more likely for children to pursue the same careers as their parents. In that case, one would expect a disproportionate number of legacy students to apply to (and choose to accept an appointment from) the Academy. This should mean the results of this study understate the true effect of legacy status, but this claim cannot be verified without data on all applicants. Chapter 5 addresses more selection issues.

Applicability

The results are based on data from the United States Air Force Academy. As Chapter 2 demonstrates, the Academy is not representative of most universities. The structure and rigor (both academic and non-academic) of the Academy may exaggerate the impact of legacy status. The information or motivation provided by alumni parents may be more significant at the Academy, relative to other schools. Also, since alumni contributions do not directly benefit the Academy, the tradeoff between student performance and alumni donations is not an issue as it is in most private universities. At these schools, it is possible that legacy admits have lower performance than non-legacy students. Still, legacy status may be an equally important signal for other intense programs, such as medical school.

Future Research

This study is limited to looking at the impact of legacy status on students who attend the Academy. Since the available data only include students who enrolled at the Academy, there is no way to determine what impact legacy status has on all applicants. Opponents of legacy admits are mostly concerned with the fairness of the application process. Admissions offices may be more concerned with yield: are legacy applicants more likely to matriculate once accepted? Without data on all applicants, it is impossible to fully address those concerns.

Another intriguing question that cannot be resolved because of data limitations is following up on non-graduates, both legacies and non-legacies. If it were possible to track these students, one could determine if legacy status at the Academy is a significant influence on graduation from another college. An additional extension could build on Winston and Zimmerman (2003) and study the peer effects of legacy status. This would require very detailed data on cadets and their roommates. Due to the complication of potentially different roommates each semester, such a study would probably have to be limited to first year performance.

Legacy siblings could also be an interesting area of research, although slightly more complicated than alumni parents.[26] If detailed data were available, one could determine if having a sibling who is currently attending or has already graduated from the Academy has a similar legacy effect. Another angle would be to consider siblings who attend the Academy, but do not graduate.

There are also avenues of further research that may be of greater concern to the Air Force. These include the impact of legacy status on a student's academic major or a graduate's career choice, time in service, or rank in the Air Force. These are the focus of Chapter 4.

Conclusions

This chapter studies the effects of legacy status on educational outcomes at the Air Force Academy. Data from the classes of 1994 to 2005 are used to verify the assertion that legacy status provides some information about a student's future performance in college, above and beyond the information contained in traditional measures such as high school academic performance. A probit model is used to predict the probability of graduation as a function of admissions data and legacy status. Control variables for high school state, gender, and race are also included. A multinomial logistic regression is used to identify the effect of legacy status on failing and quitting. In addition, OLS models are run using the

---

[26] The Air Force Academy actually gives legacy bonus points for either parents or siblings (not additive). USAFA/XPX could not confirm whether the Legacy field included both parents and sibling legacies. As a precaution, an attempt to separate parent and siblings uses the Parent_Service field: if Legacy = 1 and Parent_Service = 0, the student is assumed to be a sibling legacy. The marginal effects are nearly identical: 0.1042 for parents and 0.0979 for siblings.

same control variables to predict student GPA, MPA, and graduation order of merit.

Legacy status has no significant effect on GPA or order of merit, but legacy admits are 10 percentage points more likely to graduate, and those legacy graduates have 0.04 points higher MPA. The increase in graduation probability comes mainly from a reduction in the likelihood that a legacy admit will voluntarily quit the Academy. The effect on probability of graduation increases as the academic qualifications of the students decrease. That means legacy status is more important for those students for whom the additional points awarded by a legacy policy are most beneficial.

The results may not generalize to other universities because of the unique aspects of the Air Force Academy, but a similar result could hold for intense programs such as medical school. It is possible that legacy status is picking up the effects of other student characteristics that increase the probability of graduation. If these other variables are not observed or used in the admissions process, then the use of legacy status to capture these other variables is good policy.

Table 3-1. Summary Statistics for Relevant Variables

Variable	**Obs**	**Mean**	**Std. Dev.**	**Min**	**Max**
AFA_Grad	14340	0.7465	0.4350	0	1
AFA_GPA	10705	2.925	0.4314	2	3.99
AFA_MPA	10705	2.905	0.2687	2.075	4
AFA_OMp	10682	0.5032	0.2878	.0010	1
Female	14340	0.1535	0.3605		Binary
Asian	14340	0.0401	0.1962		Binary
Black	14340	0.0566	0.2311		Binary
Hispanic	14340	0.0669	0.2498		Binary
Indian	14340	0.0120	0.1089		Binary
Unknown	14340	0.0042	0.0646		Binary
SAT_Score	14340	1297.92	98.59	860	1600
Low_SAT	14340	1249.00	50.66	860	1280
High_SAT	14340	48.92	64.20	0	320
Math_Ratio	14340	1.0363	0.1136	.6471	1.9714
Low_Math_Ratio	14340	0.9523	0.0379	.6471	.9700
High_Math_Ratio	14340	0.0840	0.0922	0	1.0014
PAR_Score	14340	653.35	92.40	354	809
Low_PAR	14340	583.45	32.89	354	600
High_PAR	14340	69.90	71.71	0	209
Intercollegiate	14340	0.2538	0.4352		Binary
Prior	14340	0.1351	0.3419		Binary
Legacy	14340	0.0313	0.1742		Binary
Other_Academy	14340	0.0139	0.1173		Binary
Military_Background	14340	0.1706	0.3761		Binary

Notes:

- Table is based on the classes of 1994 to 2005 from the Air Force Academy.
- The 730 records identified as "bad data" are not included.
- AFA_GPA, AFA_MPA, and AFA_OMp only include students who graduated from the Academy. There are 23 students who graduated, but were not assigned an order of merit.
- SAT_Score is either (i) the sum of a student's math and verbal scores, using recentered scores for high school classes prior to 1996 or (ii) the converted composite ACT score based on formulas from *The College Board* (see Appendix A).
- High_* and Low_* variables are the upper and lower components of respective splines using kinks optimized for the graduation model (1280 for SAT, 0.97 for Math Ratio, 600 for PAR).
- See "Data" section and Appendix A for clarification on data issues.

Table 3-2. Filters Applied to Identify Bad Data

Type of Error	**Number of Records**
HS State	18
HS Year	36
HS Size	15
No SAT/ACT	6
No PAR Score	9
AFA GPA	1
AFA MPA (too low)	3
AFA MPA (too high)	18
Service Commitment	197
2Lt Service	127
1Lt Service	26
Non-grads	310
No Race	2
Total	730

Notes:

- See "Data" section for a thorough description of each type of error.

Table 3-3. Marginal Effects for Graduation Probit with Splines

	Full Model	**Lower Quartiles**	**Upper Quartiles**
Female	-0.0289 (0.0108)***	-0.0732 (0.0404)*	-0.0832 (0.0305)***
Black	0.0374 (0.0156)**	0.0011 (0.0404)	-0.2703 (0.1777)*
Hispanic	-0.0200 (0.0159)	-0.0145 (0.0527)	0.0931 (0.0465)
Indian	-0.0843 (0.0369)**	-0.0845 (0.0985)	-0.0501 (0.1099)
Asian	-0.0120 (0.0199)	0.0806 (0.0847)	-0.0346 (0.0547)
Unknown	-0.0342 (0.0592)	0.0392 (0.1577)	
Low_SAT	0.00045 (0.000090)***	0.00043 (0.00027)	
High_SAT	-0.00013 (0.000067)*		0.00013 (0.00020)
Low_Math_Ratio	0.3677 (0.1050)***	-0.0842 (0.4529)	0.5306 (0.2728)*
High_Math_Ratio	0.0255 (0.0446)	0.0401 (0.1427)	0.0950 (0.1472)
Low_PAR	0.0012 (0.00012)***	0.0020 (0.00037)***	
High_PAR	0.00048 (0.000063)***		0.00017 (0.00037)
Intercollegiate	-0.0125 (0.0097)	0.0431 (0.0330)	0.0015 (0.0348)
Prior	0.0137 (0.0114)	0.0200 (0.0322)	-0.0051 (0.0629)
Legacy	0.1038 (0.0172)***	0.1824 (0.0628)**	0.0624 (0.0448)
Other_Academy	0.1115 (0.0249)***	-0.0266 (0.1352)	
Military_Background	0.0197 (0.0098)**	0.0592 (0.0369)	0.0449 (0.0256)*
Observations	14340	1490	1567
Pseudo R^2	0.0455	0.0747	0.0635
Accuracy			
Naïve Model	74.65%	61.36%	81.38%
Estimated Model	75.08%	64.77%	81.69%

Notes:
- Standard errors are given in parentheses.
- All models include dummies for high school state and Academy class year.
- For dummy variables, marginal effect is for discrete change from 0 to 1.
- Lower and upper quartiles refer to the intersection of the respective quartiles for both SAT and PAR scores.

* Significant at the 10% level; ** Significant at the 5% level; *** Significant at the 1% level

Table 3-4. Marginal Effects for Graduation Mlogit Model

	0 (Grad)	**1 (Fail)**	**2 (Quit)**
Female	-0.0236 (0.0104)**	0.0012 (0.0035)	0.0224 (0.0100)**
Black	0.0433 (0.0147)***	0.0019 (0.0045)	-0.0452 (0.0141)***
Hispanic	-0.0214 (0.0154)	0.0059 (0.0049)	0.0155 (0.0148)
Indian	-0.0796 (0.0361)**	0.0213 (0.0130)	0.0583 (0.0347)*
Asian	-0.0077 (0.0186)	0.0223 (0.0086)**	-0.0146 (0.0172)
Unknown	-0.0285 (0.0555)	-0.0040 (0.0145)	0.0325 (0.0537)
Low_SAT	0.00035 (0.000090)***	-0.00014 (0.000030)***	-0.00022 (0.000080)**
High_SAT	-0.000094 (0.000070)	-0.00012 (0.000030)***	0.00021 (0.000060)***
Low_Math_Ratio	0.3650 (0.1003)***	-0.0893 (0.0309)***	-0.2757 (0.0961)***
High_Math_Ratio	0.0241 (0.0429)	-0.0277 (0.0145)*	0.0037 (0.0410)
Low_PAR	0.00091 (0.00012)***	-0.00030 (0.000030)***	-0.00062 (0.00011)***
High_PAR	0.00048 (0.000060)***	-0.00029 (0.000030)***	-0.00019 (0.000060)***
Intercollegiate	-0.0180 (0.0094)*	-0.0123 (0.0025)***	0.0303 (0.0092)***
Prior	0.0190 (0.0109)*	0.0062 (0.0036)*	-0.0252 (0.0104)**
Legacy	0.1037 (0.0157)***	-0.0105 (0.0055)*	-0.0932 (0.0149)***
Other_Academy	0.1091 (0.0227)***	-0.0086 (0.0087)	-0.1005 (0.0211)***
Military_Background	0.0220 (0.0092)**	0.0036 (0.0032)	-0.0257 (0.0088)***

Notes:
- Standard errors are given in parentheses.
- Model includes dummies for gender, race, and Academy class year.

* Significant at the 10% level; ** Significant at the 5% level; *** Significant at the 1% level

Table 3-5. Orthogonality of Legacy Status

	Unadjusted		Adjusted		
	Non-legacy	**Legacy**	**Non-legacy**	**Legacy**	**Difference**
Graduate	74.34 %	84.41 %	74.32 %	84.73 %	10.41 %
Fail	5.15	3.56	5.14	3.66	–1.48
Quit	20.52	12.03	20.54	11.61	–8.93

Notes:

- Unadjusted probabilities simply tabulate cadets who graduate, who don't graduate with GPA between zero and two ("Fail"), and who don't graduate with GPA equal to zero or greater than two ("Quit").
- Adjusted probabilities use predictions of the mlogit model to estimate the same probabilities after accounting for the other control variables.
- The marginal effect of legacy status is the difference between the legacy and non-legacy adjusted probabilities.
- The complete procedure is described on page 668 of Greene (2003).

Table 3-6. Effects of Legacy Status on GPA, MPA, and OM Using OLS

	GPA	MPA	OM
Female	-0.0109 (0.0091)	0.0166 (0.0066)**	-0.00029 (0.0063)
Black	-0.0582 (0.0139)***	0.0209 (0.0115)*	0.0284 (0.0098)***
Hispanic	-0.0627 (0.0144)***	-0.0265 (0.0104)**	0.0439 (0.0099)***
Indian	-0.0548 (0.0317)*	-0.0548 (0.0243)**	0.0547 (0.0225)**
Asian	-0.0653 (0.0189)***	-0.0192 (0.0137)	0.0471 (0.0128)***
Unknown	0.0339 (0.0699)	0.00021 (0.0417)	-0.0194 (0.0443)
Low_SAT	0.00094 (0.00008)***	0.00020 (0.000060)***	-0.00065 (0.000056)***
High_SAT	0.0014 (0.000063)***	0.000084 (0.000047)*	-0.00077 (0.000041)***
Low_Math_Ratio	0.1052 (0.1022)	0.0953 (0.0732)	-0.0511 (0.0702)
High_Math_Ratio	0.1250 (0.0411)***	-0.0908 (0.0299)***	-0.0605 (0.0280)**
Low_PAR	0.0016 (0.00012)***	0.00052 (0.000090)***	-0.0011 (0.000083)***
High_PAR	0.0019 (0.000057)***	0.00051 (0.000042)***	-0.0012 (0.000039)***
Intercollegiate	0.0339 (0.0085)***	-0.0670 (0.0063)***	-0.00058 (0.0059)
Prior	-0.1402 (0.0101)***	-0.0209 (0.0080)***	0.0933 (0.0072)***
Legacy	0.0220 (0.0179)	0.0419 (0.0139)***	-0.0195 (0.0123)
Other_Academy	0.0106 (0.0249)	-0.0072 (0.0208)	-0.0071 (0.0171)
Military_Background	-0.0162 (0.0092)*	0.0070 (0.0067)	0.0081 (0.0063)
Constant	0.4109 (0.1582)***	2.2865 (0.1164)***	2.1815 (0.1095)***
Observations	10705	10705	10682
R^2	0.3677	0.1113	0.3337

Notes:
- Robust standard errors are given in parentheses.
- All models include dummies for high school state and Academy class year.
- Logit models give different marginal effects, but the statistical significance of each variable is unchanged.

* Significant at the 10% level; ** Significant at the 5% level; *** Significant at the 1% level

CHAPTER 4
POST-EDUCATIONAL MEASURES

This chapter looks at measurable performance benefits to investigate the idea that legacy status provides some information to admissions offices. Empirical data from the Air Force Academy graduating classes of 1994 to 2005 are used to predict student choices in terms of college major and Air Force career field, as well as time in service and rank achieved by graduates. While legacy status has no significant impact on college major or Air Force rank, it is associated with a 0.09 increase in the probability of being a rated officer and 0.11 increase in the probability of serving at least 8 years in the Air Force. These results are robust to model specification. Extending the data back to 1982 (where admissions data are not available) shows that military performance at the Academy is at least ten times as important as grades in predicting time in service and rank. Since previous work shows that legacy status leads to higher military performance, it appears that using legacy status as a signal of future merit may be a good policy.

Theoretical Framework

Three theoretical areas apply to this chapter: university objectives, student legacy status, and statistical discrimination. It builds directly on the previous chapter, so the theoretical framework is essentially the same.

The utility-maximizing framework used by a university is best described by Epple, Romano and Seig (2003). They show that a school prevented from using race will use alternative signals of race in order to satisfy its diversity goals. This result suggests that schools will use any signals legally available to them in order to achieve their objectives.

The economics literature identifies two possible avenues for parental influence on children: genetic and cultural. The genetic argument says a child's performance is a function of breeding or innate ability inherited from the parents' genetic code. This view is supported by Black, Devereux and Salvanes (2003). The cultural argument says parental impact comes from the interaction between the parent and child. The parent may impart school-specific information or a level of motivation or maturity that helps the student succeed more than peers who do not have such a benefit. This view is supported by Laband and Lentz (1992).

The previous chapter shows evidence of improved performance associated with legacy status in the form of increased probability of graduation and higher MPAs, but the exact causal relationship of legacy status is not important. The admissions office looks at many signals they associate with future Academy performance: SAT scores, PAR scores, legacy status, etc. This is a form of statistical discrimination in which the admissions office uses past performance of previous cadets as indicators of the potential performance of prospective cadets. If there is a positive correlation between legacy status and student performance, then legacy is a valid signal to the Academy.

Empirical Strategy

This chapter extends the previous chapter to build a linear progression of models to analyze the impact of legacy status. The earlier chapter focuses on performance measures specific to the Air Force Academy: graduation probability, grades, MPA, and order of merit. This chapter focuses on student choices and post-college performance, all conditional on graduation. There are four different performance measures: college major, career field, time in service, and Air Force rank.

College Major

There are many ways to evaluate student choices for major field of study. Former Secretary of the Air Force James Roche stated an objective of increasing the number of scientists and engineers. Therefore, to evaluate student selection of major, the following variable is used:

$$\text{AFA\_Major}_i = \begin{cases} 2 & \text{If graduate } i \text{ is a science major} \\ 1 & \text{If graduate } i \text{ is an engineering major} \\ 0 & \text{Otherwise} \end{cases} \tag{4-1}$$

A science major includes all degrees in biology, chemistry, physics, meteorology, computer science, mathematics, and operations research. Engineering fields include aeronautical, astronautical, civil, environmental, electrical, and mechanics. Space operations, engineering science, and general engineering are also included as engineering degrees.

The probability that a graduate receives a degree in either science or engineering is predicted using a multinomial logit model:

$$\Pr[\text{AFA\_Major}_i = j \mid \mathbf{x}_i, \text{Legacy}_i] = \frac{e^{\boldsymbol{\beta}_j'\mathbf{x}_i + \gamma_j \text{Legacy}_i}}{\sum_{k=0}^{2} e^{\boldsymbol{\beta}_k'\mathbf{x}_i + \gamma_k \text{Legacy}_i}} \tag{4-2}$$

where $\mathbf{x}$ is a vector containing:

SAT_Score
Math_Ratio
PAR_Score
Intercollegiate
Prior

Other_Academy
Military_Background
Dummies for gender, race, and Academy class year[1]
Constant term

There is a substantial difference between graduates and non-graduates in major field of study. Over 45 percent of graduates have a technical major (science or engineering), while only 12 percent of non-graduates do. This study focuses on graduates in order to get an idea of actual returns for the Air Force. Chapter 3 addresses the effect the variables in $\mathbf{x}$ have on the probability of graduation. Rather than compound the effect of graduation with major selection, it is better to look at major conditional on graduation.

The technical majors are divided into science and engineering because there is a large disparity in the effects for gender, math ratio, and other variables. The biggest difference is in gender, where females are more likely to be scientists, but less likely to be engineers.

Ideally, the vector $\mathbf{x}$ would contain all the measures used by the admissions office (see "Threats to Identification" below). The expected marginal effects are discussed after the presentation of the four models because many effects are similar for each performance measure.

Air Force Career

As with academic major, there are many ways to break down Air Force career fields. There are only two areas that are large enough each year to derive

[1] Including state fixed effects causes problems because some states do not have graduates with each major. This results in large standard errors for the respective coefficient estimates. Also, only a handful of state fixed effects are statistically significant, even at the 10 percent level. These large errors are compounded when marginal effects are computed, resulting in insignificant results (Greene 2003). Dropping state fixed effects does not have much impact on the marginal effects.

statistical significance. Fortunately, these are also the most important career fields for the Air Force. The largest field follows directly from the Air Force's primary flying mission: rated officers (pilots and navigators). Given the Air Force's recent emphasis on new missions in space and cyberspace, there is also high demand for officers in technical careers. Therefore, the following variable is used:

$$\text{AF\_Job}_i = \begin{cases} 2 & \text{If graduate } i \text{ goes into a technical field} \\ 1 & \text{If graduate } i \text{ goes into a rated job} \\ 0 & \text{Otherwise} \end{cases} \qquad (4\text{-}3)$$

A career field is identified by an Air Force Specialty Code (AFSC), a sequence of five characters. The first is a number that indicates a broad career area: operations (1), logistics (2), support (3), medical (4), professional (5), acquisition (6), etc. Subsequent numbers or letters further break down the career into increasingly specific specialties. For example, the second digit separates a pilot (1) from a navigator (2); the third, a bomber pilot (B) from a fighter pilot (F); and the remaining characters specify the exact platform. For the most part, only the first two characters are used in this paper.

Technical fields include astronaut (13A), space and missiles (13S), weather (15), civil engineer (32), scientist (61), and developmental engineer (62). Rated fields include pilots (11) and navigators (12), including those in training (92T). There are not enough graduates from each class in other types of careers to use them in this model.

The probability that a graduate is in a technical or rated career field is predicted using a multinomial logit model:

$$\Pr[\text{AF\_Job}_i = j \mid \mathbf{x}_i, \text{Legacy}_i] = \frac{e^{\boldsymbol{\beta}_j'\mathbf{x}_i + \gamma_j \text{Legacy}_i}}{\sum_{k=0}^{2} e^{\boldsymbol{\beta}_k'\mathbf{x}_i + \gamma_k \text{Legacy}_i}} \tag{4-4}$$

where $\mathbf{x}$ is the same as in (4-2).

Time in Service

Perhaps the best measure of return on investment for the Air Force is the time an Academy graduate stays in the service. According to Air Force Instruction 36-2107 (22 Apr 2005), officers who graduate from service academies incur a five-year active-duty service commitment (ADSC).[2] Officers can add to their ADSC by undergoing voluntary training programs such as flight school or advanced academic degrees. These commitments can be as long as 10 years. Unfortunately, the admissions and legacy data are not sufficient to consider 10 years of service. Instead, a logit model is used to predict the probability that graduates stay in the service for at least eight years:

$$\Pr[\text{8\_Years}_i = 1 \mid \mathbf{x}_i, \text{Legacy}_i] = \frac{e^{\mathbf{x}_i'\boldsymbol{\beta} + \gamma_i \text{Legacy}_i}}{1 + e^{\mathbf{x}_i'\boldsymbol{\beta} + \gamma_i \text{Legacy}_i}} = \Lambda(\mathbf{x}_i'\boldsymbol{\beta} + \gamma_i \text{Legacy}_i) \tag{4-5}$$

where $\mathbf{x}$ has all the variables in (4-2) and $\Lambda(\cdot)$ is the logistic cumulative distribution function.

To make the most of the available data, Academy GPA and MPA are added to the model to see if the marginal effects of the admissions data or legacy status

[2] There is some confusion because the National Defense Authorization Act for Fiscal Year 1990 changed the commitment to six years beginning with the class of 1996. This change was repealed by the National Defense Authorization Act for FY 1996.

change. Then all available data back to the class of 1982 are used to predict time in service. This technique shows which available measures are most closely associated with graduates staying in the Air Force. While it is not as good as the model in (4-5), it is the best way to link student performance to time in service with the data available.

Air Force Rank

Another valuable indicator for how well graduates perform in the Air Force is the rank they attain. Unfortunately, junior officer rank is primarily correlated with time in service. All officers are considered for promotion to first lieutenant at 2 years, captain at 4 years, and major between 10 and 12 years (based on Air Force needs, but the entire year group is considered at the same time). In addition, promotions to first lieutenant and captain are nearly automatic, with promotion rates well above 90 percent. Ideally, a logit model could be used to predict whether a graduate attains the rank of major:

$$\Pr[\text{Major}_i = 1 \mid \mathbf{x}_i, \text{Legacy}_i] = \frac{e^{\mathbf{x}_i'\boldsymbol{\beta} + \gamma_i \text{Legacy}_i}}{1 + e^{\mathbf{x}_i'\boldsymbol{\beta} + \gamma_i \text{Legacy}_i}} = \Lambda(\mathbf{x}_i'\boldsymbol{\beta} + \gamma_i \text{Legacy}_i) \qquad (4\text{-}6)$$

where $\mathbf{x}$ has all the variables in (4-2) and $\Lambda(\cdot)$ is the logistic cumulative distribution function.

This model is severely limited by the data available. Only the oldest two classes have any graduates with the rank of major. The class of 1995 has 941 graduates but only 9 with the rank of major, which is not sufficient for any statistical inferences. Using only the data for the class of 1994 limits the sample size to 974, only 25 of which are legacy admits.

One way to make use of the data available is to use the same technique as the time in service model. That is, add Academy GPA and MPA to (4-6) and then run a reduced form of the model for classes prior to 1994.

Predictions

The SAT_Score measures overall ability, so higher scores are expected to result in higher performance.[3] Although higher ability implies a lower marginal cost for more difficult majors (i.e., science or engineering), this does not necessarily translate into increased likelihood of being a pilot or spending more time in service. It could be that graduates with higher ability face higher opportunity costs by virtue of being qualified for more lucrative careers outside the Air Force. Therefore, the impact of SAT_Score on pilot careers, time in service, and rank is indeterminate.

The total SAT score combines two different types of scores, each measuring a different skill set. Ideally, the model should include both scores, but then the method used to convert ACT composite scores to SAT scores would not be possible. Instead, the different scores are handled by computing the math to verbal ratio (or simply Math_Ratio), a process similar to Maloney and McCormick (1993). Science and engineering are technical college majors, so the Math_Ratio is expected to have a positive effect. There should also be a positive effect for

[3] The Air Force Academy only records an applicant's best standardized test score. All ACT scores are converted to their recentered SAT equivalents. See Appendix B.

technical career fields, but there is no clear theory to predict the impact on other performance measures.[4]

A student's PAR_Score is a single number calculated by Academy admissions that combines various high school academic measures (high school GPA, class rank and size, percentage of graduates going on to higher education, rigor of curriculum, and average number of academic courses taken per semester). The higher the score, the better the student is expected to perform at the Academy. Students with higher PAR scores should be more likely to declare technical majors and choose technical career fields. The impact on other performance measures is uncertain for the same reason as SAT_Score. Higher scores imply greater ability, but they also increase the opportunity cost of staying in the Air Force.

Given the increased time pressures on intercollegiate athletes, they are expected to be less likely to declare more difficult majors. This may also make them less likely to have technical careers, but the impact on rated status cannot be predicted. Also, there is no clear theory on how intercollegiate athletics would affect time in service or Air Force rank.

There are no known studies or theories about the performance of prior enlisted military members who become officers. A surprising result from the previous chapter is that prior enlisted cadets have slightly lower GPAs, but this does not suggest anything about what major they declare. One could speculate

[4] An interaction term between SAT score and math ratio could be added to allow the impact of the ratio to vary for different SAT scores. In all four models, the interaction is statistically insignificant and there is no change in the marginal effect of legacy status.

that graduates who are prior enlisted are more likely to stay in and achieve higher ranks because of their military background.

Given the hypothesis that legacy status provides positive information, the coefficient for Legacy should be positive. The prediction best supported by theory is the likelihood for legacy graduates to be rated officers. Laband and Lentz (1992) and Lentz and Laband (1989) both conclude that children are more likely to select the same careers as their parents. In the case of legacy admits, it is much more likely that their parents were rated officers.

Other_Academy is a binary variable indicating whether one of a student's parents graduated from a different service academy. Given that this chapter deals with student choices and life outside the Air Force Academy, a close relationship between Legacy and Other_Academy is not expected. The Other_Academy students could be significantly different from legacy students when it comes to their major and career choices. Although these students will likely serve in a different branch than their parents, they still come from families with a military background, so there may be a positive correlation with time in service and rank.

Military_Background is a binary variable indicating that a cadet's parent has military experience but is not a service academy graduate. There is no known theory to predict how these students will make major and career choices, but it could be argued that the military background will increase their time in service and rank. Table 4-1 shows a summary of all the expected effects.

Data

Data for every cadet from the classes of 1982 through 2005 come from the Academy's Plans and Analysis Division, with considerable collaboration with the Admissions office.[5] Some of the fields in the data set are supplied to the Academy by the Air Force Personnel Center. There are a total of 11,103 records for graduates from the classes of 1994 through 2005, each containing information on Academy performance, high school performance, and legacy status. The data also contain each graduate's Air Force status as of July 2005. This includes rank, AFSC, and time in service, but these fields are not available for the class of 2005 since they had just graduated. The data for the classes of 1982 to 1993 (11,821 records) do not contain the admissions and legacy status data. Summary statistics for variables used in the empirical models are included in Tables 4-2 and 4-3. A complete description of the variables is listed in Appendix A.

Given the long period of time, the complexities of data passed between multiple organizations, and inevitable coding errors, the data set is not perfect. The same filters described in Chapter 3 are applied to the expanded data set, and the results are summarized in Table 4-4. They are not mutually exclusive, so the total number of records removed is 641 for 1982-1993 and 398 for 1994-2005, which accounts for less than 5 percent of the observations. Not all of the filters apply directly to the empirical model (i.e., they do not directly affect

---

[5] USAFA/XPX and USAFA/RRS. Based on the agreement for the release of data, the author is not permitted to share the data.

variables in the model). The purpose of these filters is to ensure higher quality results by eliminating data that are known to have errors.[6]

Empirical Results

College Major

Table 4-5 shows the distribution of Academy majors broken down between legacy and non-legacy graduates. The table clearly shows that there is little practical difference between legacy and non-legacy graduates in terms of academic major. If anything, legacy graduates are slightly less likely to have technical majors, but this result does not account for the other admissions data. Table 4-6 shows the marginal effects estimated from the multinomial logit model. As the raw data suggest, legacy status has no impact on academic major, neither practical or statistical.

Other admissions data result in the expected marginal effects. A one point gain in total SAT score increases the probability of declaring an engineering or scientific major by 0.077 and 0.057 percentage points, respectively. Considering a one standard deviation increase in SAT score (97 points), these effects translate into 7.4 and 5.5 point increases. These are rather large results relative to the overall likelihood of declaring engineering or science, 27.8 and 18.8 percent, respectively.

The distribution of points on the SAT is also very significant. The marginal effect of math ratio is 0.8405 for engineering and 0.3252 for science. In terms of

[6] The filters do not drive the results. There is no substantial difference between the means and standard deviations of each variable using "good" and "bad" data. In addition, the models are run with and without these filters and with additional filters. The marginal effect of legacy status remains nearly identical in all cases.

a standard deviation increase (0.1124), the impacts are 9.4 and 3.7 percentage point increases, respectively. For example, given two identical graduates with total SAT scores of 1260, a graduate with 660 verbal and 600 math is roughly 9.4 percentage points more likely to be an engineer (and 3.7 points more likely to be a scientist) than a graduate with 700 verbal and 560 math (Math_Ratio 0.9 versus 0.8). This result shows the importance of quantitative skills in completing technical majors, especially for engineering.

High school performance is not as important as standardized test scores, but it still has a large impact on the probability of a graduate having a technical major. The PAR score marginal effects for the likelihood of engineering and science majors are 0.00057 and 0.00046. These translate into increases of 5.2 and 4.1 percentage points for a one standard deviation increase in PAR score (90 points).

The remaining variables of interest are not statistically significant, with a couple of exceptions. Intercollegiate athletes who graduate are 5.8 percentage points less likely to be engineering majors. Prior enlisted graduates are 7.8 percentage points less likely to be science majors. A military background is the least important statistically significant factor. These graduates are 2.3 percentage points less likely to be engineers and 2.1 points more likely to be scientists.

The predictive ability of the college major model is not very strong (0.093 pseudo R^2), but it is fairly consistent over various specifications. Dropping class year fixed effects, removing the data filter, and adding a more aggressive data filter do not change the results. Introducing piecewise linear, continuous functions (splines) for SAT score, math ratio and PAR score also has little effect.[7]

Air Force Career

Table 4-7 shows the distribution of Academy career fields, broken down between legacy and non-legacy graduates. Unlike the similar table for academic majors, there appears to be a clear difference between legacy and non-legacy graduates, especially for the rated career field. The estimated marginal effects from the multinomial logit model in Table 4-8 confirm this difference. Legacy graduates are 9.3 percentage points more likely to be rated officers. However, legacy status does not have a statistically significant relationship on the probability of being in a technical career field.

The relationship between the other admissions data and Air Force career is not as strong as it is with academic major. SAT scores do not help predict the probability of a graduate being a rated officer. The marginal effect of SAT score on the likelihood of a technical career is 0.00029. That means a one standard deviation increase in SAT score makes a graduate 2.8 percentage points more likely to have a technical career. This is less than half of the effect on technical majors, which makes sense because a technical major is required for a technical career. (Not all graduates with technical majors go on to technical career fields.)

As expected, more mathematically oriented graduates are more likely to be in rated or technical career fields. Math ratio has a statistically significant effect on the probability of being in a rated or technical career: 0.1306 and 0.3018,

[7] The only major impact of adding the splines is that math ratio nearly doubles its effect below the 0.97 kink. Above this region the effect of math ratio drops by about 25 percent.

respectively. A one standard deviation increase leads to increased likelihood of 1.5 and 3.4 percentage points. It is reasonable that better math skills are more important for technical careers than rated careers.

High school performance has a small but surprising effect on career choice. The marginal effects of PAR score are –0.00017 and 0.00019 for rated and technical careers, respectively. These translate into 1.6 points less likely to be rated and 1.7 points more likely to be technical for a one standard deviation change in PAR score. While this is a statistically significant result, it is not particularly strong.

Other variables of interest also have surprising results. Intercollegiate status has no impact on a technical career, but these graduates are 10.8 percentage points less likely to be rated. A similar result exists for prior enlisted graduates: there is no impact on technical careers, but these graduates are 9.3 percentage points less likely to be rated. Graduates with parents from another service academy are less likely to be in technical careers (by 4.1 points). This is statistically significant at the 0.1 level, but it is consistent throughout all variations of the model. Military background has no significant effect on career choice.

As with the academic major model, the predictive ability of this model is not very strong (0.064 pseudo R^2). It is still fairly consistent over various specifications. Removing other fixed effects, using splines, or adding Academy GPA and MPA does not change the basic relationship between legacy status and career choice. The marginal effect of legacy status on a rated career is between

8.4 and 11.8 percentage points. Using splines does change the effect of math ratio and PAR score, but there is no substantial change in the other variables.

Time in Service

A simple examination of the distribution of time in service reveals a difference between legacy and non-legacy graduates (see Table 4-9). This relationship is also reflected in the marginal effects of the logit model. Table 4-10 shows that legacy graduates are nearly 11 percentage points more likely to stay in the Air Force for at least eight years.

None of the other variables in the model (except gender) is as strongly related to time in service. Math ratio is not significant. SAT and PAR scores are statistically significant with nearly identical inconsequential marginal effects. A one standard deviation increase in these scores results in only 1.7 and 1.6 percentage point increases in the probability of serving at least eight years. Graduates who were intercollegiate athletes are 6.4 percentage points less likely to stay beyond eight years, while graduates from families with military backgrounds are 3.3 points more likely to stay.

These results are fairly robust to model specification. Removing fixed effects, using splines, or adding Academy GPA and MPA does not change the basic relationship between legacy status and time in service.

The sample size for this model is considerably smaller than the previous two models. Admissions data are not available for classes prior to 1994, but it is possible to look at the relationship between Academy performance measures (rather than admissions data) and time in service. This can be combined with the results from Chapter 3 to link legacy status to time in service via the Academy performance measures. Table 4-11 shows three separate logit model results looking at graduates who stay for 10, 15, and 20 years. Each successive model has fewer data points because fewer classes can be included in the model.

Academy GPA is not significant for the 10 year model, but it is for 15 and 20 with marginal effects of 0.0302 and 0.0421, respectively. A one standard deviation increase in GPA results in an increased probability of staying beyond 15 years by 1.4 percentage points. The same change in GPA increases the probability of staying beyond 20 years by 1.9 points. These results are dwarfed by the effects of MPA: 0.1937, 0.2271, and 0.2735 for 10, 15, and 20 years, respectively, all significant at the 0.01 level. These translate into increases of 5.6, 6.6, and 7.8 percentage points for a one standard deviation increase in MPA. Note that a standard deviation in MPA scores is only 60 percent of that for GPA.

Chapter 3 shows that legacy graduates have slightly higher MPAs than non-legacies. Combined with the results above, this confirms the result that legacy graduates are likely to serve longer than their non-legacy peers.

Air Force Rank

Table 4-12 shows the distribution of graduates of the class of 1994 who have attained the rank of major. As with all the previous models, there appears to be initial evidence that legacy status has a large impact, nearly 10 percentage points in this case. Unfortunately, with only 25 legacy admits in the class, it is difficult to ascertain any level of statistical significance. In fact, Table 4-13 shows the results of the logit model, which confirms there is no statistically significant effect of legacy status. None of the admissions variables has a significant relationship to the probability of a graduate pinning on major.

As with the time in service model, the main problem here is a lack of sample size. With only 25 legacy graduates in the class of 1994, all spread among 15 different AFSCs (16 rated, 2 technical, and 7 others), it is difficult to have any statistical confidence in any results. The same technique from the time in service model is used to increase the sample size by adding classes prior to 1994. Table 4-14 shows two separate logit model results that look at the probability of graduates achieving at least the rank of lieutenant colonel and colonel.

For the LtCol model, Academy GPA is statistically significant with a marginal effect of 0.0660. A one standard deviation increase in GPA makes it 3.0 percentage points more likely for a graduate to make the rank of LtCol. Grades, however, are not significant for attaining the rank of Col. Academy MPA is a better predictor for rank. It has marginal effects of 0.2423 and 0.0692 for LtCol and Col, respectively. This is over twice as important as GPA for LtCol since a one standard deviation increase in MPA results in 7.0 percentage points more likely for a graduate to attain the rank of LtCol. The effect drops for the rank of Col (2.0 points).

The LtCol model includes the classes of 1982 through 1989; the Col model includes 1982-1985. The cutoff for LtCol is not important because that rank is awarded following a time-based promotion board similar to earlier ranks. The cutoff for Col is very sensitive because there are large variations in the number of colonels per class. The fewer classes that are included (i.e., move the cutoff closer to 1982), the greater the marginal effect of MPA and GPA. Regardless of the cutoff, however, the marginal effect of MPA is always ten times the size of that for GPA.

Limitations and Further Research

Threats to Identification

There are several problems with the identification strategy of this empirical study. First, it shares all the same problems as Chapter 3 since it uses the same data set. These problems include the lack of non-academic admissions data, the potential for omitted variables not observed by the admissions office, and selection issues related to a student's decision to apply to and accept an appointment from the Academy. These issues could jeopardize any identification of causal relationships, but that problem is minor since the study is considering whether legacy status is a valid signal of performance (not necessarily a cause of performance). The selection issue is a bigger problem because different application and acceptance decisions between legacy and non-legacy students could result in a disproportionate number of legacy students. While this could mean the results of this study understate the true effect of legacy status, that claim cannot be verified without data on all applicants.

There are other problems specific to this study which mainly stem from the linear relationship (in time) of the dependent variables. It makes sense that the admissions data used as regressors in this model lose predictive power as the dependent variables move further away from college admission (as evidenced by decreasing pseudo R^2 as the models progress). Using the same variables in each model could cause problems because there is a link between the dependent variables. For example, graduates can only be in a technical career field if they have a technical major. Graduates can only attain a certain rank if they have been in the service for the required amount of time.

The career model is also limited because the allocation of each type of field for each year group is constrained by Air Force requirements. Although this is handled somewhat by the class year fixed effects, the fact remains that the career field a graduate gets is a function of the student's request, their Academy performance, their academic major, Air Force needs, and training availability. Since career fields are not simply chosen by the cadets, the model looks at the relationship between legacy status (and other variables) to actual career fields, not necessarily the desired career fields.

The time in service model is critically linked to the career field model because of service commitments incurred for training programs, specifically the ten year commitment from pilot training. If the model is re-run for non-rated officers only, the point estimate for legacy status only drops by 0.01, but it loses its statistical significance. Another alternative is to run the model for all graduates, but to control for career field. Adding Rated and Tech_Job results in a better fit (0.2359 vs. 0.0659 Pseudo R^2), but the marginal effect of legacy status drops to 4.5 percent. In this version, that effect is still statistically significant. So legacy status could still be associated with longer service, but probably not as much as suggested by Table 4-10.

The rank model is perhaps the weakest in this paper because of the lack of data. Ideally, the class of 1995 could be included, but the data do not reflect the latest promotions; only 9 of 941 graduates have the rank of major. There should be more majors based on the time in service field. Still, no changes to the model specification result in a significant effect for legacy status. In fact, very few variables are statistically significant, and the pseudo R^2 is very low in all variations of the model. The small sample size creates large standard errors, so it is not possible to accurately describe the relationship between the admissions data and Air Force rank.

Applicability

The results are based on data from the United States Air Force Academy, which is not representative of most universities. The structure and rigor of the Academy and Air Force service may exaggerate the impact of legacy status. The information or motivation provided by alumni parents may be more (or less) significant for service in the Air Force relative to other career choices. Still, legacy status does appear to contain some information on the future Air Force success of Academy graduates similar to the results of Laband and Lentz (1992) with lawyers.

As far as other universities are concerned, post-educational success of graduates is more difficult to identify and may not be as great a concern. The most common measures are advanced degrees and earnings. The former may be best associated with this study (i.e., students with PhD parents may be more likely to go on to get PhDs). The earnings measure may help a school recruit applicants, but there is no reason to think legacy status has a significant impact unless the focus is on a specific professional school within a university, such as a medical school or law school.

Future Research

This study is limited to looking at the impact of legacy status on students who graduate from the Academy. The easiest way to extend the analysis is to obtain the full admissions data for all Academy classes. Unfortunately, it does not appear that the admissions office has such data, and trying to compile it on a case-by-case basis would be prohibitively expensive. An equally difficult extension would be to identify the career of each cadet's parents. This may be a better indicator of future career than simply using legacy status.

One data weakness that may be easier to resolve is the study of rank. Rather than simply looking at rank attained, it could be possible to investigate the relationship between legacy status and line numbers, the order in which ranks are assigned at each promotion board.

Another intriguing question that cannot be resolved because of data limitations is following up on non-graduates at other colleges and in careers outside the Air Force. If it were possible to track these students, one could determine if legacy status at the Academy is a significant influence on graduation from another college or on career earnings.

Conclusions

Legacy issues are often as hotly debated as affirmative action. Many schools use legacy status as a consideration when looking at student applications. Proponents of such policies argue for the increased donations from alumni parents, while opponents claim such policies are inherently discriminatory and contrary to a merit-based system. Neither side directly addresses the use of legacy status as a signal of student performance.

Admissions data from the classes of 1994 to 2005 are used to test the assertion that legacy status provides some information about a student's future performance in the Air Force. Multinomial logistic models are used to predict the probability of graduates attaining engineering or scientific degrees and the probability of graduates going on to rated or technical careers. Logit models are used to predict the probability of graduates staying beyond eight years of service and attaining the rank of major. Only control variables available to the admissions board are considered in order to evaluate the effectiveness of legacy status as a signal of future performance.

Legacy status has no effect on academic majors but is positively correlated with career field and time in service. Legacy graduates are roughly 9 percentage points more likely to be rated officers and nearly 11 percentage points more likely to serve beyond 8 years. There is no statistically significant relationship between legacy status and Air Force rank. Extending the data set back to 1982 shows that military performance at the Academy is at least ten times as important as grades in predicting time in service and rank.

A surprising result, which follows the same return on investment logic of legacy status, is the impact of intercollegiate athletic participation. Graduates who were athletes are 5.8 percentage points less likely to have engineering degrees, 10.8 points less likely to be rated officers, and 6.4 points less likely to serve at least 8 years. While these numbers may suggest the Air Force Academy should accept fewer athletes, it could be that the benefits of athletes are not reflected in the measures used in this paper. McCormick and Tinsley (1987) show that a university's athletic performance leads to a greater number of applications and greater average SAT scores for incoming students.

Several robustness tests are performed. The impact of legacy status is independent of the other control variables and not very sensitive to model specification. It is possible, however, that legacy status is picking up the effects of other student characteristics. If these other variables are not observed or used in the admissions process, then the use of legacy status to capture these other variables is good policy.

Table 4-1. Expected Effects

	Major	**Career**	**Time**	**Rank**
SAT_Score	+/+	?/+	?	?
Math_Ratio	+/+	?/+	?	?
PAR_Score	+/+	?/+	?	?
Intercollegiate	-/-	?/?	?	?
Prior	?/?	?/?	+	+
Legacy	+/+	+/+	+	+
Other_Academy	?/?	?/?	+	+
Military_Background	?/?	?/?	+	+

Table 4-2. Summary Statistics for Relevant Variables, c/o 1994-2005

Variable	**Obs**	**Mean**	**Std. Dev.**	**Min**	**Max**
Engineer	10705	0.2777	0.4479		Binary
Scientist	10705	0.1879	0.3906		Binary
Rated	10705	0.4497	0.4975		Binary
Technical_Job	10705	0.1253	0.3310		Binary
8_Years	3524	0.7798	0.4144		Binary
Major_Rank_94	974	0.5893	0.4922		Binary
Female	10705	0.1519	0.3589		Binary
Asian	10705	0.0403	0.1966		Binary
Black	10705	0.0557	0.2293		Binary
Hispanic	10705	0.0642	0.2451		Binary
Indian	10705	0.0104	0.1013		Binary
Unknown	10705	0.0037	0.0610		Binary
SAT_Score	10705	1301.87	96.96	860	1600
Math_Ratio	10705	1.0377	0.1124	0.7125	1.9714
PAR_Score	10705	661.03	90.67	354	809
Intercollegiate	10705	0.2383	0.4261		Binary
Prior	10705	0.1298	0.3360		Binary
Legacy	10705	0.0354	0.1848		Binary
Other_Academy	10705	0.0160	0.1254		Binary
Military_Background	10705	0.1727	0.3780		Binary

Notes:

- Table is based on graduates from the Air Force Academy classes of 1994 to 2005.
- The 398 records identified as "bad data" are not included.
- The 8_Years variable only includes data for 1994-1997.
- Major_Rank_94 is the probability that graduates from the class of 1994 attain the rank of major.
- SAT_Score is either (i) the sum of a student's math and verbal scores, using recentered scores for high school classes prior to 1996 or (ii) the converted composite ACT score based on formulas from *The College Board*.

Table 4-3. Summary Statistics for Relevant Variables, c/o 1982-1993

Variable	Obs	Mean	Std. Dev.	Min	Max
10_Years	11180	0.6111	0.4875		Binary
15_Years	8269	0.4475	0.4973		Binary
20_Years	3591	0.3988	0.4897		Binary
Lt Col	7323	0.3083	0.4618		Binary
Col	5473	0.0356	0.1854		Binary
Female	11180	0.1177	0.3223		Binary
Asian	11180	0.0320	0.1761		Binary
Black	11180	0.0640	0.2448		Binary
Hispanic	11180	0.0431	0.2031		Binary
Indian	11180	0.0059	0.0766		Binary
AFA_GPA	11180	2.86	0.4549	2	3.99
AFA_MPA	11180	2.92	0.2891	2.032	3.856

Notes:

- Table is based on graduates from the Air Force Academy classes of 1982 to 1993, except: 10_Years includes 1982-1995, 15_years includes 1982-1990, 20_Years includes 1982-1985; Lt Col includes 1982-1989; Col includes 1982-1987
- The 641 records identified as "bad data" are not included.

Table 4-4. Filters Applied to Identify Bad Data

	Number of Records		
Type of Error	**1982-1993**	**1994-2005**	**All Data**
HS State	198	17	215
HS Year	n/a	24	24
HS Size	n/a	12	12
No SAT/ACT	n/a	4	4
No PAR Score	n/a	6	6
AFA GPA	6	1	7
AFA MPA (too low)	7	3	10
AFA MPA (too high)	1	0	1
Service Commitment	292	197	489
2Lt Service	54	127	181
1Lt Service	7	26	33
Capt Service	105	n/a	105
No Race	0	2	2
Total Bad	641	398	1039
Total	11821	11103	22924

Notes:
- See "Data" section for a description of each type of error.

Table 4-5. Legacy Distribution of Academy Major

AFA_Major	Non-legacy	Legacy	Total
Count			
0 (Other)	5510	211	5721
1 (Engineer)	2874	99	2973
2 (Scientist)	1942	69	2011
Total	10326	379	10705
Percentage			
0 (Other)	53.36	55.67	53.44
1 (Engineer)	27.83	26.12	27.77
2 (Scientist)	18.81	18.21	18.79
Total	100	100	100

Table 4-6. Marginal Effects for Academy Major

	Engineer	**Scientist**
Female	-0.1115 (0.0114)***	0.0692 (0.0121)***
Black	-0.0118 (0.0231)	0.0315 (0.0225)
Hispanic	0.0091 (0.0198)	0.0061 (0.0179)
Indian	0.1083 (0.0503)**	-0.0789 (0.0333)**
Asian	0.0145 (0.0231)	0.0195 (0.0197)
Unknown	0.0804 (0.0830)	0.0363 (0.0735)
SAT_Score	0.00077 (0.000060)***	0.00057 (0.000050)***
Math_Ratio	0.8405 (0.0419)***	0.3252 (0.0358)***
PAR_Score	0.00057 (0.000050)***	0.00046 (0.000050)***
Intercollegiate	-0.0584 (0.0114)***	0.0013 (0.0103)
Prior	0.0161 (0.0160)	-0.0783 (0.0115)***
Legacy	-0.0132 (0.0243)	-0.0080 (0.0204)
Other_Academy	0.0190 (0.0364)	0.0198 (0.0312)
Military_Background	-0.0233 (0.0119)*	0.0208 (0.0107)*
Observations	10705	
Pseudo R^2	0.0930	

Notes:

- Standard errors are given in parentheses.
- Model includes dummies for Academy class year.
- For dummy variables, marginal effect is for discrete change from 0 to 1.

* Significant at the 10% level; ** Significant at the 5% level; *** Significant at the 1% level

Table 4-7. Legacy Distribution of Air Force Career

AF_Job	Non-legacy	Legacy	Total
Count			
0 (Other)	3567	106	3673
1 (Rated)	4623	191	4814
2 (Technical)	1301	40	1341
Total	9491	337	9828
Percentage			
0 (Other)	37.58	31.45	37.37
1 (Rated)	48.71	56.68	48.98
2 (Technical)	13.71	11.87	13.64
Total	100	100	100

Table 4-8. Marginal Effects for Air Force Career

	Rated	**Technical**
Female	-0.3060 (0.0129)***	0.0190 (0.0105)*
Black	-0.1848 (0.0230)***	0.0494 (0.0203)**
Hispanic	-0.1164 (0.0216)***	0.0111 (0.0159)
Indian	-0.0388 (0.0510)	0.0445 (0.0410)
Asian	-0.1566 (0.0255)***	0.0218 (0.0190)
Unknown	-0.2342 (0.0829)***	-0.1097 (0.0281)***
SAT_Score	0.000057 (0.000070)	0.00029 (0.000040)***
Math_Ratio	0.1306 (0.0490)***	0.3018 (0.0311)***
PAR_Score	-0.00017 (0.000060)***	0.00019 (0.000040)***
Intercollegiate	-0.1083 (0.0136)***	0.0084 (0.0095)
Prior	-0.0930 (0.0170)***	0.0202 (0.0126)
Legacy	0.0929 (0.0289)***	-0.0168 (0.0185)
Other_Academy	0.0672 (0.0431)	-0.0410 (0.0243)*
Military_Background	-0.0124 (0.0144)	-0.0024 (0.0094)
Observations	9828	
Pseudo R^2	0.0640	

Notes:

- Standard errors are given in parentheses.
- Model includes dummies for Academy class year.
- Sample size is smaller than Table 6 because there is no AFSC data for the class of 2005.
- For dummy variables, marginal effect is for discrete change from 0 to 1.

* Significant at the 10% level; ** Significant at the 5% level; *** Significant at the 1% level

Table 4-9. Legacy Distribution of Time in Service

8_Years	Non-legacy	Legacy	Total
Count			
0 (No)	764	12	776
1 (Yes)	2658	90	2748
Total	3422	102	3524
Percentage			
0 (No)	22.33	11.76	22.02
1 (Yes)	77.67	88.24	77.98
Total	100	100	100

Table 4-10. Marginal Effects for Time in Service

	8_Years
Female	-0.1930 (0.0251)***
Black	-0.0932 (0.0343)***
Hispanic	0.0318 (0.0280)
Indian	-0.0199 (0.0745)
Asian	-0.0050 (0.0404)
SAT_Score	0.00017 (0.000090)*
Math_Ratio	0.0559 (0.0640)
PAR_Score	0.00018 (0.000090)**
Intercollegiate	-0.0641 (0.0191)***
Prior	-0.0293 (0.0233)
Legacy	0.1099 (0.0294)***
Other_Academy	-0.0387 (0.0711)
Military_Background	0.0332 (0.0168)**
Observations	3498
Pseudo R^2	0.0513

Notes:

- Standard errors are given in parentheses.
- Model includes dummies for Academy class year.
- For dummy variables, marginal effect is for discrete change from 0 to 1.

* Significant at the 10% level; ** Significant at the 5% level; *** Significant at the 1% level

Table 4-11. Marginal Effects for Time in Service Using Academy Performance

	10_Years	**15_Years**	**20_Years**
Female	-0.2242 (0.0136)***	-0.1593 (0.0162)***	-0.1356 (0.0240)***
Black	-0.0517 (0.0187)***	-0.0608 (0.0235)**	-0.0456 (0.0347)
Hispanic	0.0168 (0.0208)	-0.0310 (0.0281)	0.0146 (0.0423)
Indian	0.0140 (0.0539)	-0.1300 (0.0729)*	-0.2390 (0.0839)***
Asian	-0.0120 (0.0247)	0.0022 (0.0320)	0.0887 (0.0497)*
AFA_GPA	-0.0034 (0.0111)	0.0302 (0.0138)**	0.0421 (0.0201)**
AFA_MPA	0.1937 (0.0172)***	0.2271 (0.0217)***	0.2735 (0.0333)***
Observations	13095	8269	3591
Pseudo R^2	0.0356	0.0289	0.0337

Notes:

- Standard errors are given in parentheses.
- Model includes dummies for Academy class year.
- Classes of 1982-1995 are considered for 10 years; 1982-1990 for 15 years; 1982-1985 for 20 years.
- For dummy variables, marginal effect is for discrete change from 0 to 1.

* Significant at the 10% level; ** Significant at the 5% level; *** Significant at the 1% level

Table 4-12. Legacy Distribution of Majors for Class of 1994

MAJ94	**Non-legacy**	**Legacy**	**Total**
Count			
0 (No)	392	8	400
1 (Yes)	557	17	574
Total	949	25	974
Percentage			
0 (No)	41.31	32	41.07
1 (Yes)	58.69	68	58.93
Total	100	100	100

Table 4-13. Marginal Effects for Air Force Rank

	MAJ94
Female	-0.2178 (0.0494)***
Black	-0.1169 (0.0838)
Hispanic	-0.1028 (0.0670)
Indian	-0.2316 (0.1734)
Asian	0.0756 (0.0844)
SAT_Score	0.000070 (0.00020)
Math_Ratio	0.2154 (0.1482)
PAR_Score	0.00019 (0.00020)
Intercollegiate	-0.0749 (0.0428)*
Prior	-0.1238 (0.0541)**
Legacy	0.1115 (0.0946)
Other_Academy	-0.2940 (0.1517)*
Military_Background	0.0451 (0.0410)
Observations	974
Pseudo R^2	0.0395

Notes:

- Standard errors are given in parentheses.
- Model only includes data for the class of 1994.
- For dummy variables, marginal effect is for discrete change from 0 to 1.

* Significant at the 10% level; ** Significant at the 5% level; *** Significant at the 1% level

Table 4-14. Marginal Effects for Air Force Rank Using Academy Performance

	Lt Col	**Col**
Female	-0.0954 (0.0149)***	-0.00045 (0.0060)
Black	-0.0820 (0.0209)***	0.0019 (0.0097)
Hispanic	-0.0336 (0.0259)	-0.0162 (0.0076)**
Indian	-0.1036 (0.0638)	0.0069 (0.0335)
Asian	-0.0160 (0.0298)	0.0066 (0.0131)
AFA_GPA	0.0660 (0.0130)***	0.0067 (0.0045)
AFA_MPA	0.2423 (0.0210)***	0.0692 (0.0089)***
Observations	7323	3591
Pseudo R^2	0.0756	0.1881

Notes:
- Standard errors are given in parentheses.
- Model includes dummies for Academy class year.
- Classes of 1982-1989 are considered for Lt Col; 1982-1985 for Col. The results for Lt Col are not sensitive to the last year, but for Col they are. Still, the marginal effect of MPA is always ten times that of GPA.
- For dummy variables, marginal effect is for discrete change from 0 to 1.

* Significant at the 10% level; ** Significant at the 5% level; *** Significant at the 1% level

CHAPTER 5
FORMAL THEORY AND POTENTIAL BIAS

This chapter builds on previous empirical work on legacy status by developing a theoretical model of the admissions process and evaluating possible sources of bias. The model formalizes the three ways legacy status might affect the process: a direct impact on graduation probability, a selection impact through enrollment, and a signaling effect for unobserved student characteristics. These effects cannot be estimated separately, so empirical results measure the overall impact of legacy status, which is the correct measure to evaluate the admissions policy. The model suggests a technique for testing the optimality of the admissions process, but requires data on all applicants. The additional data are also required to examine other potential sources of bias in the empirical work.

General Theory

This section develops a general theory for admission to the Air Force Academy using legacy status. While the model is general, it is necessarily simplified and does not account for all the steps of the process (see the "Enrollment Selection" section below).

Students

A potential student is characterized by three types of variables: observable characteristics (x_O), unobservable characteristics (x_U), and legacy status (L). While all of these are known to the student, only the observable characteristics and legacy status are observed by the Academy and other universities. Observable characteristics include things like standardized test scores, high school grades, high school class rank, etc. Legacy status is a binary variable equal to one if either (or both) of the student's parents graduated from the Academy, and equal to zero otherwise. The unobservable characteristics are more difficult to define. These can include nebulous traits such as motivation, maturity, and knowledge of the Academy or the military.

Assumption 1. The joint probability density of potential students, $f(x_O, x_U, L)$, is continuous in x_O and x_U.

Assumption 2. All potential students submit applications to the Academy.

This simplification removes the first decision step from the student in order to simplify the analysis. The assumption is not unreasonable because the Academy can recruit students it wants and encourage them to apply.

While an individual student is identified by all three variables (x_O, x_U, L), the Academy and other universities can only see a student as an (x_O, L)-type. Therefore, marginal and conditional density functions must be defined to convert from a student's perspective to the Academy (or another school's) perspective. By assumption, the Academy knows these functions.

The marginal density of observable characteristics and legacy status of potential students is given by

$$f_O(x_O, L) \equiv \int f(x_O, x_U, L)\, dx_U \tag{5-1}$$

and is continuous in x_O (by Assumption 1).

The conditional density function for unobservable characteristics given observable characteristics and legacy status,

$$h(x_{\mathrm{U}} \mid x_{\mathrm{O}}, L) \equiv \frac{f(x_{\mathrm{O}}, x_{\mathrm{U}}, L)}{f_{\mathrm{O}}(x_{\mathrm{O}}, L)} \tag{5-2}$$

is continuous in x_{U} (by Assumption 1).

The probability that a student will enroll at the Academy if accepted is denoted by $R(x_{\mathrm{O}}, x_{\mathrm{U}}, L)$. This probability means little to the Academy admissions office because it cannot observe x_{U}. Therefore, let $R_{\mathrm{O}}(x_{\mathrm{O}}, L)$ denote the probability of an (x_{O}, L)-type student enrolling if accepted.

Student utility for graduating from the Academy is given by $U^{\mathrm{AFA}}(x_{\mathrm{O}}, x_{\mathrm{U}}, L)$. The expected utility from the student's best alternative to the Academy is given by $U^{\mathrm{A}}(x_{\mathrm{O}}, x_{\mathrm{U}})$. Note the difference in definitions here. The alternative is an expected utility, so it incorporates the probability of graduation from the alternate school. This definition is used to simplify the model, because graduation from another school is not the focus. Also, note that the alternative is not a function of legacy status. There is no reason to expect a student's legacy status at the Academy to have an impact on the student's alternatives.

Let $G(x_{\mathrm{O}}, x_{\mathrm{U}}, L)$ denote the probability of graduation for an enrolled student of type $(x_{\mathrm{O}}, x_{\mathrm{U}}, L)$. Note that the student's decision to stay once enrolled has been incorporated into this function, thus removing another step from the process in the previous section. Using this notation,

$$G_{\mathrm{O}}(x_{\mathrm{O}}, L) \equiv \int G(x_{\mathrm{O}}, x_{\mathrm{U}}, L)\,dx_{\mathrm{U}} \tag{5-3}$$

corresponds to one of the performance measures considered in Chapter 3.[1]

Assumption 3. A student who attends the Academy but does not graduate receives utility zero.

This simplification is possible by simply rescaling the student's utility to ensure the alternative available after not graduating is equal to zero. Given Assumption 3, students who are expected utility maximizers will decide to enroll in the Academy if

$$G(x_{\mathrm{O}}, x_{\mathrm{U}}, L) \cdot U^{\mathrm{AFA}}(x_{\mathrm{O}}, x_{\mathrm{U}}, L) \geq U^{A}(x_{\mathrm{O}}, x_{\mathrm{U}}) \tag{5-4}$$

This condition defines a continuum of enrollment constraints, one for each (x_{O}, x_{U}, L)-type student. If the condition in (5-4) holds, then $R(x_{\mathrm{O}}, x_{\mathrm{U}}, L) = 1$; otherwise, $R(x_{\mathrm{O}}, x_{\mathrm{U}}, L) = 0$.

Let $X_{\mathrm{U}}(x_{\mathrm{O}}, L)$ define the set of unobserved characteristics for which an (x_{O}, L)-type student will enroll. That is,

$$X_{\mathrm{U}}(x_{\mathrm{O}}, L) \equiv \left\{ x_{\mathrm{U}} : G(\cdot) \cdot U^{\mathrm{AFA}}(\cdot) \geq U^{A}(\cdot) \right\} \tag{5-5}$$

Therefore, $R_{\mathrm{O}}(x_{\mathrm{O}}, L)$, the probability that an (x_{O}, L)-type student will enroll, can be written

$$R_{\mathrm{O}}(x_{\mathrm{O}}, L) \equiv \int_{X_{\mathrm{U}}(x_{\mathrm{O}}, L)} h(x_{\mathrm{U}} \mid x_{\mathrm{O}}, L)\, dx_{\mathrm{U}} \tag{5-6}$$

The condition in (5-6) is illustrated in Figure 5-1. The conditional density function for unobservable characteristics given observable characteristics and legacy status for students who enroll is given by

[1] Chapter 3 estimated $G_{\mathrm{O}}(x_{\mathrm{O}}, 1) - G_{\mathrm{O}}(x_{\mathrm{O}}, 0) = 0.10$.

$$h_R(x_{\mathrm{U}} \mid x_{\mathrm{O}}, L) \equiv \begin{cases} \dfrac{h(x_{\mathrm{U}} \mid x_{\mathrm{O}}, L)}{R_{\mathrm{O}}(x_{\mathrm{O}}, L)} & \text{for all } x_{\mathrm{U}} \in X_U(x_{\mathrm{O}}, L) \\ 0 & \text{otherwise} \end{cases} \tag{5-7}$$

Academy

For each (x_{O}, L)-type application the Academy receives, it admits the student with probability $A(x_{\mathrm{O}}, L)$. Alternatively, $A(x_{\mathrm{O}}, L)$ could be viewed as the proportion of (x_{O}, L)-type students that are admitted. Therefore, the marginal density of observable characteristics and legacy status for students enrolled at the Academy is given by

$$a(x_{\mathrm{O}}, L) = R_{\mathrm{O}}(x_{\mathrm{O}}, L) A(x_{\mathrm{O}}, L) f_{\mathrm{O}}(x_{\mathrm{O}}, L) \tag{5-8}$$

The number of students attending the Academy can be computed by

$$k = \sum_L \int_{x_{\mathrm{O}}} R_{\mathrm{O}}(x_{\mathrm{O}}, L) A(x_{\mathrm{O}}, L) f_{\mathrm{O}}(x_{\mathrm{O}}, L) dx_{\mathrm{O}} \tag{5-9}$$

Assumption 4. The Academy faces an exogenously determined, fixed capacity constraint of K students that can be enrolled in each class year.

This assumption is realistic since class size for the Academy is mandated by Congress rather than decisions at the Academy level.

Assumption 5. Success for the Academy is defined as graduation of a cadet, which results in a new officer for the Air Force.[2]

Including the probability of graduation and the conditional density function for unobservable characteristics for enrolled students in (5-9) provides an expression for the density of graduates

---

[2] The quality of the graduates is also important, but is an unnecessary complication for the purposes of this model.

$$g(x_U) = \sum_L \int_{x_O} G(x_O, x_U, L) h_R(x_U \mid x_O, L) R_O(x_O, L) A(x_O, L) f_O(x_O, L) dx_O \quad \text{(5-10)}$$

Integrating this distribution over unobservable characteristics determines the expected number of graduates, hence, the Academy's objective function. The admissions process at the Academy can be written as follows:

$$\max_{A(x_O, L)} \sum_L \int_{x_U} \int_{x_O} G(x_O, x_U, L) h_R(x_U \mid x_O, L) R_O(x_O, L) A(x_O, L) f_O(x_O, L) dx_O dx_U \quad \text{(5-11)}$$

subject to a feasibility constraint

$$A(x_O, L) \in [0,1] \;\; \forall \;\; (x_O, L) \quad \text{(5-12)}$$

and a capacity constraint

$$\sum_L \int_{x_O} R_O(x_O, L) A(x_O, L) f_O(x_O, L) dx_O \le K \quad \text{(5-13)}$$

Substituting (5-6) and (5-7) allows the optimization problem to be rewritten:

$$\max_{A(x_O, L)} \sum_L \int_{X_U(x_O, L)} \int_{x_O} G(x_O, x_U, L) h(x_U \mid x_O, L) A(x_O, L) f_O(x_O, L) dx_O dx_U \quad \text{(5-14)}$$

$$\text{s.t.} \quad A(x_O, L) \in [0,1] \;\; \forall \;\; (x_O, L) \quad \text{(5-15)}$$

$$\sum_L \int_{X_U(x_O, L)} \int_{x_O} h(x_U \mid x_O, L) A(x_O, L) f_O(x_O, L) dx_O dx_U \le K \quad \text{(5-16)}$$

Optimal Admissions Policy

Proposition 1. If a proper subset of legacy (non-legacy) students are admitted to the Academy, then there is a marginal (x_O, L)-type the Academy is indifferent about admitting. The marginal student type is identified by the ratio of the probability of enrolling and graduating to the probability of enrolling being equal to the shadow price of capacity. Any (x_O, L)-type with a ratio that exceeds this constant will be admitted with probability 1, and those who are below will not be admitted.

Proof: Note that (5-15) can be broken into two conditions: $A(x_{\mathrm{O}}, L) \geq 0$ and $A(x_{\mathrm{O}}, L) \leq 1$. The former can be ignored because it will be accounted for in the Kuhn-Tucker analysis of first-order conditions. The latter is accounted for in the lagrangian for the optimization problem:

$$\begin{aligned}\ell = &\sum_L \int_{X_{\mathrm{U}}(x_{\mathrm{O}},L)} \int_{x_{\mathrm{O}}} G(x_{\mathrm{O}}, x_{\mathrm{U}}, L) h(x_{\mathrm{U}} \mid x_{\mathrm{O}}, L) A(x_{\mathrm{O}}, L) f_{\mathrm{O}}(x_{\mathrm{O}}, L) dx_{\mathrm{O}} dx_{\mathrm{U}} \\ &- \lambda \left[\sum_L \int_{X_{\mathrm{U}}(x_{\mathrm{O}},L)} \int_{x_{\mathrm{O}}} h(x_{\mathrm{U}} \mid x_{\mathrm{O}}, L) A(x_{\mathrm{O}}, L) f_{\mathrm{O}}(x_{\mathrm{O}}, L) dx_{\mathrm{O}} dx_{\mathrm{U}} - K \right] \\ &- \mu \left[A(x_{\mathrm{O}}, L) - 1 \right] \end{aligned} \tag{5-17}$$

The first-order conditions are found by taking derivatives with respect to $A(x_{\mathrm{O}}, L)$, λ and μ. For the $A(x_{\mathrm{O}}, L)$ case, it is evaluated at a particular (x_{O}, L), which drops the summation and the integral over x_{O}. The conditions are:

$$\begin{aligned}\ell_{A(x_{\mathrm{O}},L)} = &\int_{X_{\mathrm{U}}(x_{\mathrm{O}},L)} G(x_{\mathrm{O}}, x_{\mathrm{U}}, L) h(x_{\mathrm{U}} \mid x_{\mathrm{O}}, L) f_{\mathrm{O}}(x_{\mathrm{O}}, L) dx_{\mathrm{U}} \\ &- \lambda \left[\int_{X_{\mathrm{U}}(x_{\mathrm{O}},L)} h(x_{\mathrm{U}} \mid x_{\mathrm{O}}, L) f_{\mathrm{O}}(x_{\mathrm{O}}, L) dx_{\mathrm{U}} \right] - \mu \leq 0, \text{ with equality if} \\ &A(x_{\mathrm{O}}, L) > 0 \end{aligned} \tag{5-18}$$

$$\begin{aligned}\ell_{\lambda} = &\sum_L \int_{X_{\mathrm{U}}(x_{\mathrm{O}},L)} \int_{x_{\mathrm{O}}} h(x_{\mathrm{U}} \mid x_{\mathrm{O}}, L) A(x_{\mathrm{O}}, L) f_{\mathrm{O}}(x_{\mathrm{O}}, L) dx_{\mathrm{O}} dx_{\mathrm{U}} - K \geq 0, \\ &\text{with equality if } \lambda > 0 \end{aligned} \tag{5-19}$$

$$\ell_{\mu} = A(x_{\mathrm{O}}, L) - 1 \leq 0, \text{ with equality if } \mu > 0 \tag{5-20}$$

Condition (5-19) simply says $\lambda > 0$ if the capacity constraint is binding and $\lambda = 0$ otherwise. Condition (5-18) can be simplified because $f_{\mathrm{O}}(x_{\mathrm{O}}, L)$ is a positive constant and can be factored out (changing the scale of the lagrangian and μ):

$$\ell^*_{A(\bullet)} = \int_{X_{\mathrm{U}}(x_{\mathrm{O}},L)} G(x_{\mathrm{O}}, x_{\mathrm{U}}, L) h(x_{\mathrm{U}} \mid x_{\mathrm{O}}, L) dx_{\mathrm{U}} - \lambda \left[\int_{X_{\mathrm{U}}(x_{\mathrm{O}},L)} h(x_{\mathrm{U}} \mid x_{\mathrm{O}}, L) dx_{\mathrm{U}} \right] - \mu^* \leq 0 \quad (5\text{-}21)$$

The first term of (5-21) is equal to the probability that an (x_{O}, L)-type applicant will enroll and go on to graduate, $R_G(x_{\mathrm{O}}, L)$. From equation (5-6), the term in brackets in (5-21) is equal to $R_{\mathrm{O}}(x_{\mathrm{O}}, L)$, the probability that an (x_{O}, L)-type student will enroll. By construction $R_G(x_{\mathrm{O}}, L) \leq R(x_{\mathrm{O}}, L)$; it is not possible for the proportion that enroll and graduate to be larger than the proportion that simply enroll. The multiplier λ, is the shadow price of capacity and can also be considered the opportunity cost of enrollment. In economics terms, $R_G(x_{\mathrm{O}}, L)$ can be viewed as the marginal benefit of accepting an (x_{O}, L)-type student, and $R_{\mathrm{O}}(x_{\mathrm{O}}, L)$ is the marginal cost (to the capacity).

First, consider the trivial case in which the enrollment constraint does not bind. If this were true, (5-19) implies $\lambda = 0$ and the second term of (5-21) drops out leaving

$$\ell^*_{A(\bullet)} = R_G(x_{\mathrm{O}}, L) - \mu^* \leq 0, \text{ with equality if } A(x_{\mathrm{O}}, L) > 0 \quad (5\text{-}22)$$

As long as there is some positive probability of graduating, $\mu^* > 0$ is required for the inequality to hold, which means $A(x_{\mathrm{O}}, L) = 1$ because of (5-20). This result makes sense because if there were no capacity constraint, the Academy would simply admit every applicant.

If the capacity constraint does bind, (5-19) implies $\lambda > 0$, and (5-21) can be written:

$$\ell^*_{A(\bullet)} = R_G(x_O, L) - \lambda R_O(x_O, L) - \mu^* \leq 0, \text{ with equality if } A(x_O, L) > 0 \quad (5\text{-}23)$$

Assume $A(x_O, L) = 0$ (i.e., the (x_O, L)-type will not be admitted). This means (5-23) is strictly less than zero. Now assume $A(x_O, L) \in (0,1)$ (i.e., the Academy is indifferent in admitting the (x_O, L)-type student). From (5-20), $\mu^* = 0$ and the relationship in (5-23) is an equality. The equality of (5-23) also holds if $A(x_O, L) = 1$, but in this case, $\mu^* > 0$ so $R_G(x_O, L) > \lambda R_O(x_O, L)$. These results are summarized as follows:

$$\text{If } \ell_{A(\bullet)} < 0, \text{ then } A(x_O, L) = 0 \quad (5\text{-}24)$$

$$\text{If } \ell_{A(\bullet)} = 0, \text{ then } A(x_O, L) \in (0,1) \quad (5\text{-}25)$$

$$\text{If } \ell_{A(\bullet)} > 0, \text{ then } A(x_O, L) = 1 \quad (5\text{-}26)$$

Another way to summarize the optimal admissions policy is to focus on the ratio $R_G(x_O, L)/R_O(x_O, L)$:

$$\text{If } R_G(x_O, L)/R_O(x_O, L) < \lambda, \text{ then } A(x_O, L) = 0 \quad (5\text{-}27)$$

$$\text{If } R_G(x_O, L)/R_O(x_O, L) = \lambda, \text{ then } A(x_O, L) \in (0,1) \quad (5\text{-}28)$$

$$\text{If } R_G(x_O, L)/R_O(x_O, L) > \lambda, \text{ then } A(x_O, L) = 1 \quad (5\text{-}29)$$

A simple way to prioritize applicants is to sort them by increasing $R_G(x_O, L)/R_O(x_O, L)$, a sort of benefit to cost ratio. Those with the highest values are accepted with probability one, until the capacity constraint is reached. The last group of (x_O, L)-types accepted will have a proportion less than one to keep from violating the capacity constraint. QED

Testing the Model

Proposition 1 provides a simple test for the general theory developed in the previous section. If it is possible to identify the marginal legacy and non-legacy students, then their predicted probability of success (i.e., $G(x_O, L)$) should be the same. If they are not the same, then either the model is incorrect or the Academy is not using an optimal admissions policy.[3]

Unfortunately, identifying the marginal student is not possible with the available data. The marginal student should be the one with the minimum estimated graduation probability, but this value is very sensitive to model specification. Trying to reduce the sensitivity by looking at the average of the lowest 5 or 10 percent of the predictions is not a statistically sound technique because it produces a biased estimate of the bottom of the distribution.

A visual examination of the data demonstrates the problem with identifying the marginal student. Figures 5-2 and 5-3 show histograms for the predicted graduation probabilities from two models. The first uses a single probit with state fixed effects, just like the one used in Chapter 3. The latter uses dual probits, one for legacy and one for non-legacy students, and does not use state fixed effects (because of sample size issues in the legacy probit). While both cases clearly show legacy students with higher expected graduation probability on average, the marginal students are very different. In Figure 5-2, it appears that the

[3] There are several simplifications that would suggest problems with the model rather than the Academy. First, there is no consideration for the quality of graduates. The Academy also must balance anticipated academic majors among an incoming class. In addition, there are geographic constraints placed on the Academy because all cadets must have a Congressional appointment. That means an applicant from one region may be offered an appointment over a student with a higher predicted probability of success from another region.

marginal legacy student is much better than the marginal non-legacy student, suggesting the Academy is not admitting enough legacy students. The exact opposite result is shown in Figure 5-3.

If data were available on all applicants, it would be possible to use maximum likelihood estimation to identify marginal students.[4] Let m be the probability that the marginal applicant will graduate. Define $p_{a,i}$ as the admissions office's estimate that applicant i will graduate and $p_{e,i}$ as the econometrician's estimate of the same, where

$$p_{a,i} = p_{e,i} + \varepsilon_i \tag{5-30}$$

and ε_i ~ N(0, σ^2). Using this notation, applicant i is admitted if

$$p_{a,i} \geq m \tag{5-31}$$

Substituting (5-30) gives

$$p_{e,i} + \varepsilon_i \geq m \tag{5-32}$$

Therefore, the probability that applicant i is admitted is equal to the probability that

$$\varepsilon_i \geq m - p_{e,i} \tag{5-33}$$

which can be found using the cumulative normal distribution, $F(\cdot)$.

Let Λ be the set of applicants who are accepted and Ω be the set who are not accepted. The logarithm of the likelihood function is given by

$$\sum_{i\in\Omega} \ln\left[F(m - p_{e,i})\right] + \sum_{i\in\Lambda} \ln\left[1 - F(m - p_{e,i})\right] \tag{5-34}$$

[4] This technique could also be used to estimate all the parameters rather than using a probit model.

To allow for different admission criteria for legacy and non-legacy applicants, let m_ℓ be the probability that the marginal legacy admit will graduate and m_n be the probability that the marginal non-legacy admit will graduate. Now (5-34) can be re-written

$$\sum_{i\in\Omega_\ell}\ln\left[F(m_\ell - p_{e,i})\right] + \sum_{i\in\Lambda_\ell}\ln\left[1 - F(m_\ell - p_{e,i})\right] + \sum_{i\in\Omega_n}\ln\left[F(m_n - p_{e,i})\right] + \sum_{i\in\Lambda_n}\ln\left[1 - F(m_n - p_{e,i})\right] \tag{5-35}$$

Maximizing (5-35) by choosing m_ℓ, m_n, and σ (and the parameters of $p_{e,i}$) yields the maximum likelihood estimates of all the parameters. That is, the technique computes parameter values that are most likely, given the observed data. These parameter estimates are unbiased. Furthermore, the estimates have minimum variance as the sample size tends to infinity, so they are best for large samples. In this case, however, the technique cannot be used without data on all applicants.

Ideally, the data set should contain all information submitted by all applicants and fields denoting which applicants are accepted by the Academy and which enrollees go on to graduate. Of course, to test the impact of legacy status on other performance measures (GPA, MPA, majors, etc.), these data fields must also be included in the data set. The Academy may also be interested in knowing the impact of legacy status on yield, i.e., the percentage of accepted students who decide to enroll. If so, this information must also be collected. It may be difficult to incorporate some of the data from the subjective portion of an application. As much as possible, these data fields should be quantified. For example, binary variables could be created for yes/no questions (e.g., "Are you an Eagle Scout?"). Writing samples could be assigned a numerical score, preferably assigned by the admissions office prior to an acceptance decision. While the ideal data set may not be available now, the admissions office could start collecting this information now in anticipation of future studies.

Applying the MLE technique with a standard statistical package such as STATA will also provide the standard errors of the parameters. With these estimates, it is then possible to test whether $m_\ell = m_n$ using a simple t-test. The statistical package can also perform this test. Similarly, a t-test could also be used to determine whether corresponding parameters for legacy and non-legacy students are the same. These tests could be used to determine if legacy students are more (or less) likely to graduate. Minor changes to the model can shift the focus from graduation to other performance measures: yield, GPA, MPA, etc.

There are a couple of weaknesses to the MLE approach as presented in this section (although not to MLE in general). On the technical side, the derivation of the model does not guarantee that $p_{a,i}$ will be a probability (i.e., lie in the [0,1] interval). Although (5-35) could be modified to take this into account, it is simpler to run the model as is and then check whether $p_{a,i}$ is a probability or not.[5] More importantly, (5-30) assumes a random normally distributed error term between the admissions office's graduation prediction and the econometrician's prediction. This could be explained by random noise added by admissions officers. If there is a *known* systematic difference between the estimates, that can easily be added to the model. If the difference is caused by omitted variables

---

[5] This is similar to using OLS to predict GPA which is technically bound on the [0,4] interval. If the predicitons remain in the interval, there is no need to complicate the model.

(i.e., something the admissions office has access to that the econometrician does not), however, this approach will not work. See the "Omitted Variables" section below.

Direct vs. Indirect Effect

The model developed in this chapter illustrates how legacy status (or any other observable characteristic) can impact the admissions process and in turn affect $R_G(x_O, L)/R_O(x_O, L)$. There are three distinct ways legacy status enters the objective function in (5-11). These show direct and indirect effects of legacy status, which could be interpreted as a source of bias in empirical work if the effects cannot be estimated separately.

First, L enters directly into the probability of graduation. This situation could occur if legacy students are simply better (or worse) than non-legacy students. Another explanation could be that legacy students have more motivation beyond the typical motivation used as an unobserved characteristic. The motivation could be caused by the parents of a legacy admit not allowing the student to quit. In that case, for a given (x_O, x_U)-type student, $G(x_O, x_U, 1) > G(x_O, x_U, 0)$. This is the direct (or independent) causal effect of legacy status. It is the usual focus of econometric work.

The second way legacy status could affect the process is through information content. That is, legacy status could be a signal for unobserved characteristics through the conditional distribution $h(x_U \mid x_O, L)$. In this case, a causal relationship between legacy status and graduation probability is not important as long as legacy is correlated with some unobserved characteristic that does impact the probability of graduation. Awarding extra points to legacy students would be justified if $h(x_U \mid x_O, L)$ possesses stochastic dominance in terms of L and $G(x_O, x_U, L)$ is increasing in x_U. That is, the distribution of x_U for non-legacy students is to the left of the distribution for legacy students, and greater values of x_U lead to greater probability of graduation. Another way to explain stochastic dominance is to say that higher values of x_U are more likely to be associated with legacy students relative to non-legacy students. Unfortunately, because x_U is unobservable (by definition), it is not possible to isolate the impact of legacy on $h(x_U \mid x_O, L)$ from the effect on $G(x_O, x_U, L)$.

The third way legacy status enters the admissions process described in this model is through the student's enrollment decision. In (5-11), this impact is captured by $R_O(x_O, L)$. The alternative specification in (5-14) captures the selection issue by changing the bound on the second integral with $X_U(x_O, L)$. If the enrollment decision is made differently between legacy and non-legacy students, it is possible that the distribution of unobserved characteristics also differs. As with the case of $h(x_U \mid x_O, L)$, it is impossible to separate the impact on enrollment from the impact on observed graduation probabilities.

Schools that award extra points to legacy applicants are indicating that they believe $R_G(x_O, L)/R_O(x_O, L)$ is increasing in L for a particular x_O (i.e., a legacy student who enrolls is more likely to graduate than an equally qualified non-legacy student who enrolls). Note that this is not the typical ideal of normal econometric studies that want to show causality. A traditional economic study would seek to find the independent effect of legacy status on graduation for an (x_O, x_U)-type:

$$G_L \equiv G(x_{\mathrm{O}}, x_{\mathrm{U}}, 1) - G(x_{\mathrm{O}}, x_{\mathrm{U}}, 0) \tag{5-36}$$

Given the fact that some of the variables are unobservable, however, the best that can be measured is the effect of legacy status on an (x_{O})-type:

$$G_{\mathrm{O}L} \equiv G_{\mathrm{O}}(x_{\mathrm{O}}, 1) - G_{\mathrm{O}}(x_{\mathrm{O}}, 0) \tag{5-37}$$

where

$$G_{\mathrm{O}}(x_{\mathrm{O}}, L) = \int_{X_U(x_{\mathrm{O}}, L)} G(x_{\mathrm{O}}, x_{\mathrm{U}}, L) h(x_{\mathrm{U}} \mid x_{\mathrm{O}}, L) dx_{\mathrm{U}} \tag{5-38}$$

From (5-38), it is again possible to see all three impacts of legacy status. $G_{\mathrm{O}}(x_{\mathrm{O}}, L)$ is the probability that an (x_{O}, L)-type student will graduate if enrolled. This is exactly what is estimated in Chapter 3 and is the same measure that drives the optimal admissions policy because

$$\frac{R_G(x_{\mathrm{O}}, L)}{R_{\mathrm{O}}(x_{\mathrm{O}}, L)} = \frac{\Pr[\text{Grad \& Enroll}]}{\Pr[\text{Enroll}]} = \frac{\Pr[\text{Grad} \mid \text{Enroll}]\Pr[\text{Enroll}]}{\Pr[\text{Enroll}]} = \Pr[\text{Grad} \mid \text{Enroll}] \tag{5-39}$$

Therefore, the work of Chapter 3 is an estimate of the overall effect of legacy status but not of the direct (causal) effect of legacy status.

Omitted Variables

A potential problem with the empirical results on legacy status is that there may be observable characteristics used by the admissions office that are not included in the data set. For example, subjective criteria such as student essays and teacher evaluations are not included. If these characteristics are correlated with unobservable characteristics (x_{U}) or with legacy status, the results could be biased.

A simulation of the effect of omitted data can be seen in Table 5-1, which shows the results of three different probit models, each adding successively more information about a student's high school performance: no high school data, high school GPA, and PAR score (which combines GPA with class standing and other measures). The PAR_Score column is the same model estimated in Chapter 3 with a couple of differences. First, the sample size is smaller because an additional filter is applied to keep high school GPA in the [2,5] interval. The model estimated in Table 5-1 also does not use splines for simplicity in interpreting the results.

The table illustrates how adding additional data can change the marginal effect of each explanatory variable. Some have a lesser impact and others become more prominent as data is added. In the case of legacy status, the marginal effect increases, but by less than 10 percent, rising from 0.0910 with no high school data to 0.0987 with the most data. While this shows legacy status to be fairly stable, it is not necessarily indicative of what would happen if other omitted data were added.

There are two ways to investigate this possible source of bias, but both require additional data. The simplest way is to add all other observable data that the admissions office has on enrolled students. This could prove difficult since much of the omitted data are subjective measures. An alternative requires an expanded data set that includes all applicants, not just enrolled students. A model could be estimated to determine if the observable data used in Chapter 3 does a good job of predicting the probability of acceptance. If so, the omitted observable characteristics are not very important, so the potential of bias is low.

Enrollment Selection

A different type of bias could follow from the fact that only enrollment data is used to evaluate an admissions policy. From the general model, a student's enrollment decision is captured by the $X_U(x_O, L)$ set. While the impact of legacy status on this choice cannot be separated from $h(x_U \mid x_O, L)$ or $G(x_O, x_U, L)$, it is possible to model the enrollment decision in more detail to discover possible ways in which legacy and non-legacy applicants make different choices. It is possible that these decisions lead to different proportions of legacy and non-legacy students who enroll compared to those who apply. In addition, the observable (and unobservable) characteristics of the enrolled students may differ from those of the applicants.

It is useful to discuss the overall process by which a student graduates from a particular university. There is a specific sequence of events that must occur. First, the student must decide to apply to the university. Most students apply to multiple schools in order to have backup plans or to pick the school that offers the best financial aid package. Each school reviews its applications and offers admission to a subset based on the school's objectives. The student receives updated information based on the results of these school decisions (i.e., the alternatives are more clearly defined). If accepted, the student must then decide whether to enroll in the school. If the student does enroll, information is updated again since the perceived benefits or costs could change based on first-hand experience. The student can decide to stay or to leave the school and pursue another alternative. If the student stays, there is some probability of successful completion (graduation) based primarily on student characteristics.[6]

The sequence of events involves several opportunities for the student to make decisions. If these decisions are made differently by different types of students, then the difference between the characteristics of the different types of enrolled students will not reflect the differences between the applicants. For example, enrolled legacy students could systematically have larger values of x_U than non-legacy students, but this difference may not be present in legacy and non-legacy applicants. If that is the case, then using enrollment data to evaluate a legacy admissions policy is not valid.

Figure 5-4 shows a representation of the selection process. The rectangle represents the set of all prospective students. The vertical line divides this set into legacy and non-legacy students. The horizontal lines divide the set based on the selection process. The shaded area denotes the set of all enrolled students at the Academy. This area is the focus of Chapter 3. The lowest horizontal line divides the set of enrollees into those who graduate and those who do not. The slope of this line is greater than the enrollment line because a greater proportion of legacy students graduate. Since all the previous lines are flat, Figure 5-4 shows legacy and non-legacy students make the same decisions (and are equally accepted) based on population proportions.

Table 5-2 uses some numbers to quantify the point of the figure. The numbers are manufactured to illustrate the point and are not based on the scale

[6] Other contributing factors (changing family circumstances, economic conditions, natural disasters, etc.) are not considered in this paper.

of the figure. They show the basic result of Chapter 3, the ten point difference in graduation probability based on enrollment, but the numbers are not based on the data set used in Chapter 3. In the case displayed in Figure 5-4, there is no selection bias. While the percentage point increase in graduation probability for legacy admits drops from 0.10 to 0.06 when looking at all admits, the actual percentage increase is the same, 15 percent. This shows the result of Chapter 3 does generalize to all applicants if there is no selection bias in the enrollment process.

Figure 5-5 shows cases where the selection bias could exaggerate or negate the results from Chapter 3. The figure on the left shows non-legacy students consistently less likely to decide to apply, get accepted to, and enroll in the Academy. The figure on the right shows the opposite. The second two columns in Table 5-2 correspond to these figures. In the first case, the result is exaggerated when looking at all admits instead of just enrolled cadets: legacy applicants are 44 percent more likely to graduate compared to only 20 percent of legacy enrollees. The opposite is true for the figure where non-legacy students are consistently more likely to decide in favor of the Academy. Here the legacy advantage observed in enrolled cadets (13 percent more likely to graduate) is nearly nonexistent from an applicant's perspective (2 percent).

These are dramatic examples to illustrate the potential problem. Since admissions offices consider the set of applicants, the findings of empirical studies based on enrollment data may not apply.

Conclusions

This chapter builds on the empirical work investigating legacy status. While the previous chapters conclude that legacy status is a valid signal of future performance, they have potential bias introduced by selection issues because they rely exclusively on enrollment data. This chapter presents a theoretical framework for college admissions that explicitly accounts for legacy status in order to examine these issues.

The general model derives an optimal admissions policy for the Academy to maximize the expected number of graduates. This model allows legacy status to impact the process directly through graduation probability, in addition to a selection effect through enrollment and a signaling effect through the conditional distribution of unobserved student characteristics. The optimal policy suggests that the marginal legacy and non-legacy students admitted should have the same predicted probability of graduation. A maximum likelihood estimator is derived to identify the marginal student, but the technique requires data on all applicants, not just enrollees.

Potential sources of bias in the empirical work are identified. These include causal effects, omitted variables, and enrollment selection issues. The first results from the fact that the causal effect of legacy status cannot be separated from the indirect effects. The empirical work estimates the overall impact of legacy status, which is not the typical focus of econometric analysis. Fortunately, the overall effect of legacy status is the correct measure for evaluating the admissions policy.

The other sources of bias can preclude the use of previous results to evaluate the legacy admissions policy. The only way to determine if these sources cause a problem is to expand the data set. The empirical models need to be re-run with any omitted variables included. Alternatively, the existing variables could be used to predict acceptance decisions to determine how important the omitted variables are. Data on all applicants are also required to determine if there is bias introduced by different enrollment decisions between legacy and non-legacy students. Without addressing these issues, prior empirical results for legacy status may not be useful to the Academy admissions office.

Table 5-1. Marginal Effects for Graduation Probability

	No HS Data	**HS_GPA**	**PAR_Score**
Female	-0.0054 (0.0111)	-0.0198 (0.0114)*	-0.0295 (0.0116)***
Black	0.0070 (0.0193)	0.0123 (0.0191)	0.0205 (0.0187)
Hispanic	-0.0330 (0.0184)*	-0.0298 (0.0183)*	-0.0276 (0.0182)
Indian	-0.0837 (0.0401)**	-0.0798 (0.0400)**	-0.0686 (0.0394)*
Asian	-0.0101 (0.0211)	-0.0134 (0.0213)	-0.0141 (0.0213)
Unknown	-0.1147 (0.0756)	-0.1123 (0.0754)	-0.0984 (0.0746)
SAT_Score	0.00023 (0.000047)***	0.00017 (0.000047)***	0.000092 (0.000048)*
Math_Ratio	0.1145 (0.0363)***	0.0928 (0.0364)**	0.0821 (0.0364)**
HS_GPA		0.1004 (0.0114)***	
PAR_Score			0.00063 (0.000046)***
Intercollegiate	-0.0486 (0.0105)***	-0.0387 (0.0105)***	-0.0243 (0.0104)**
Prior	0.0043 (0.0161)	0.0352 (0.0154)**	0.0267 (0.0154)*
Legacy	0.0910 (0.0187)***	0.0964 (0.0183)***	0.0987 (0.0180)***
Other_Academy	0.1030 (0.0270)***	0.1053 (0.0267)***	0.1081 (0.0262)***
Military_Background	0.0164 (0.0107)	0.0174 (0.0106)	0.0195 (0.0106)*
Observations	12196	12196	12196
Pseudo R2	0.0268	0.0325	0.0404

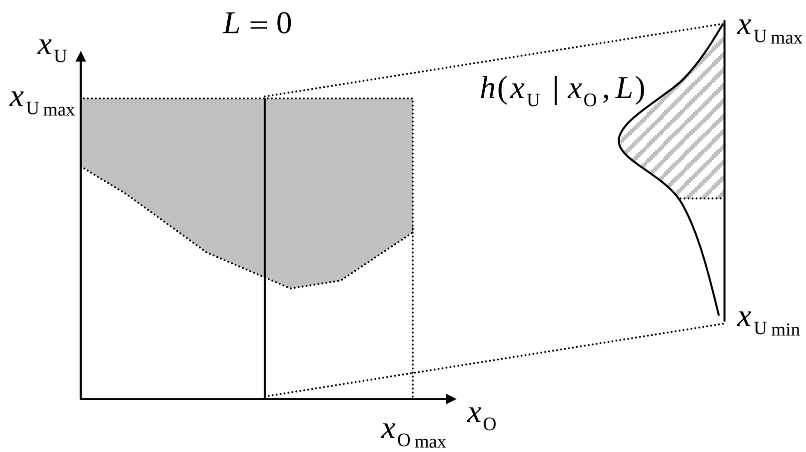
Notes:

- Standard errors are given in parentheses.
- Model includes dummies for high school state and Academy class year.
- Sample size is smaller than Table 3-3 because an additional filter is used to ensure HS_GPA $\in$ [2,5].
- For dummy variables, marginal effect is for discrete change from 0 to 1.

* Significant at the 10% level; ** Significant at the 5% level; *** Significant at the 1% level

Table 5-2. Numerical Examples Illustrating Potential Bias From Enrollment Data

	No Bias		Exaggerate		Negate	
	% of Population		**% of Population**		**% of Population**	
	Non-leg	**Legacy**	**Non-leg**	**Legacy**	**Non-leg**	**Legacy**
Apply	0.50	0.50	0.40	0.50	0.60	0.50
Accepted	0.40	0.40	0.30	0.40	0.50	0.40
Enroll	0.30	0.30	0.20	0.30	0.40	0.30
Graduate	0.20	0.23	0.10	0.18	0.30	0.255
Grad as % enroll	0.67	0.77	0.50	0.60	0.75	0.85
Grad as % apply	0.40	0.46	0.25	0.36	0.50	0.51
	Difference in		**Difference in**		**Difference in**	
	% Pnts	**%**	**% Pnts**	**%**	**% Pnts**	**%**
Enrollees	0.10	0.15	0.10	0.20	0.10	0.13
Applicants	0.06	0.15	0.11	0.44	0.01	0.02



Set of applicants who enroll; defines $X_{\mathrm{U}}(x_{\mathrm{O}}, L)$ for $L = 0$

$R_{\mathrm{O}}(x_{\mathrm{O}}, L) = \int_{X_{\mathrm{U}}(x_{\mathrm{O}}, L)} h(x_{\mathrm{U}} \mid x_{\mathrm{O}}, L)\,dx_{\mathrm{U}}$

This area can be inflated to define the conditional distribution of unobserved characteristics given observed characteristics and legacy status for students who enroll (whereas $h(x_{\mathrm{U}} \mid x_{\mathrm{O}}, L)$ is for all applicants):

$$h_R(x_{\mathrm{U}} \mid x_{\mathrm{O}}, L) = \begin{cases} \dfrac{h(x_{\mathrm{U}} \mid x_{\mathrm{O}}, L)}{R_{\mathrm{O}}(x_{\mathrm{O}}, L)} & \text{for all } x_{\mathrm{U}} \in X_U(x_{\mathrm{O}}, L) \\ 0 & \text{otherwise} \end{cases}$$

Figure 5-1. Conditional Distributions of Unobserved Characteristics

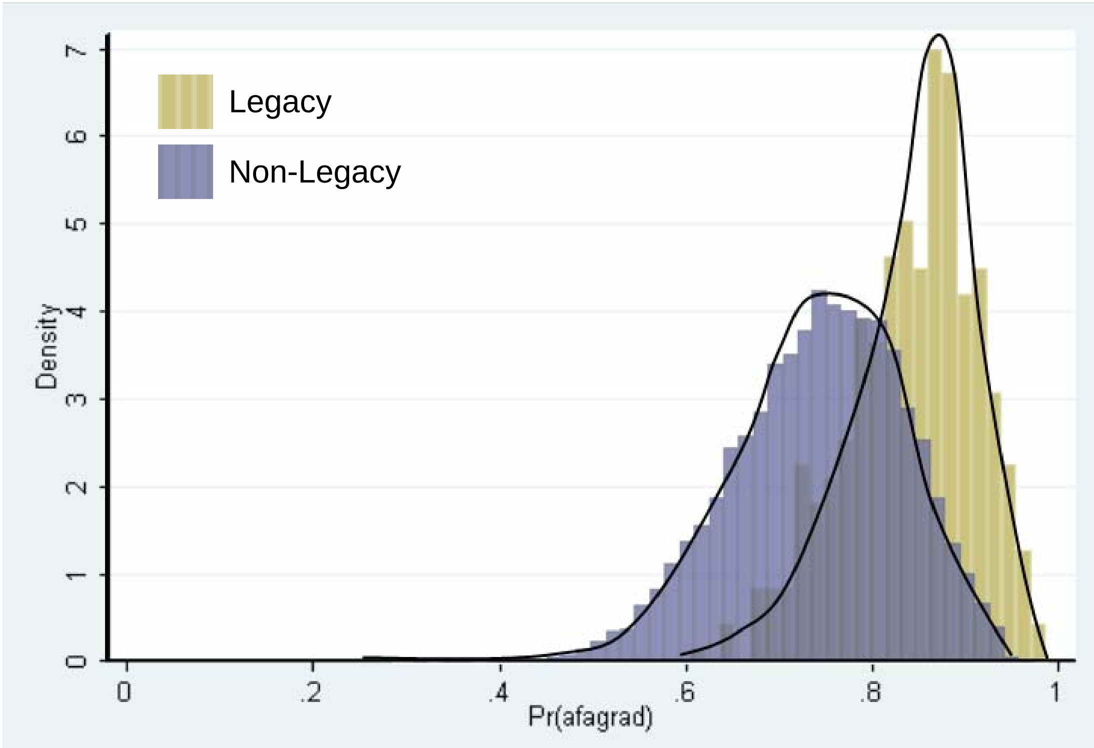


Figure 5-2. Predicted Probability of Graduation–Single Probit with State Fixed Effects

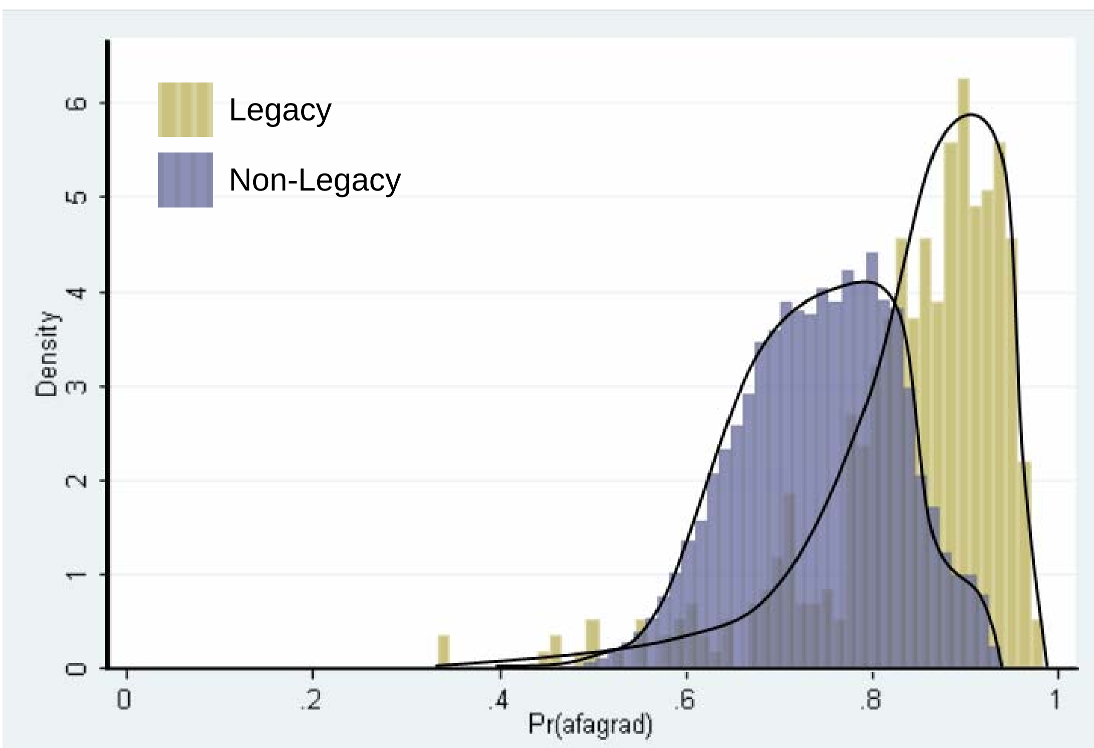


Figure 5-3. Predicted Probability of Graduation–Dual Probits without State Fixed Effects

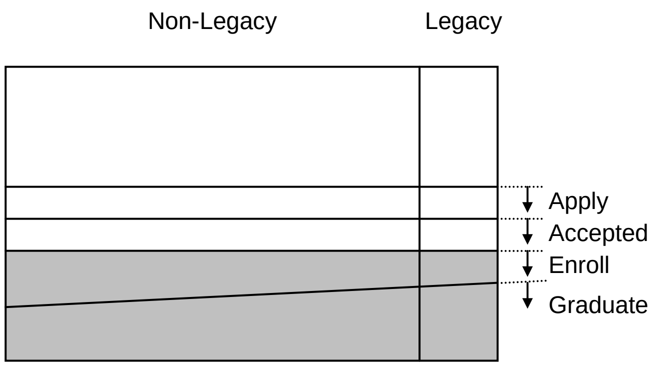


Figure 5-4. No Selection Issues

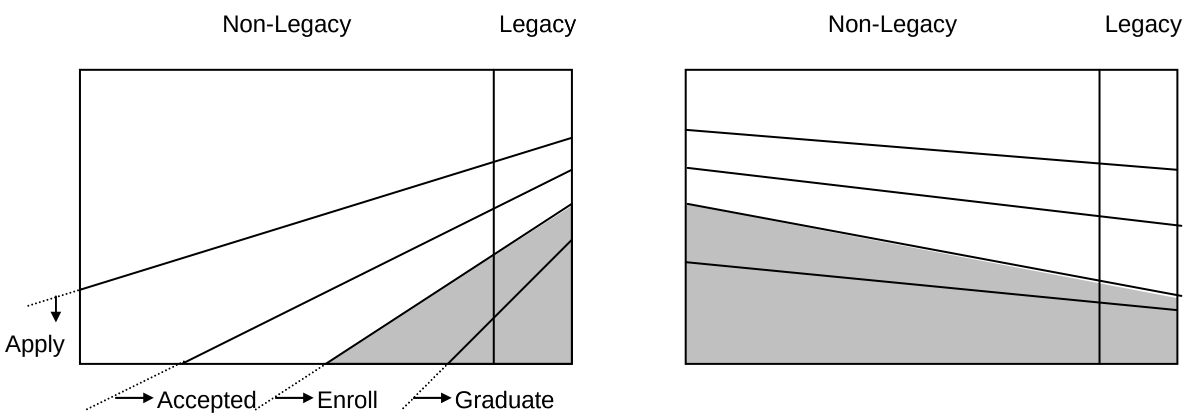


Figure 5-5. Selection Issues and Exaggerate or Negate Results from Enrollment Data

CHAPTER 6
CONCLUSIONS

Legacy issues are often as hotly debated as affirmative action. Many schools use legacy status as a consideration when looking at student applications. Proponents of such policies argue for the increased donations from alumni parents, while opponents claim such policies are inherently discriminatory and contrary to a merit-based system. Neither side directly addresses the use of legacy status as a signal of student performance.

This dissertation studies the effects of legacy status on educational outcomes at the Air Force Academy and post-educational outcomes in the Air Force. Data from the classes of 1994 to 2005 are used to verify the assertion that legacy status provides some information about a student's future performance above and beyond the information contained in traditional measures such as high school academic performance.

A probit model is used to predict the probability of graduation as a function of admissions data and legacy status. Ordinary Least Squares models are run using the same control variables to predict student GPA, MPA, and graduation order of merit. Multinomial logistic models are used to predict the probability of graduates attaining engineering or scientific degrees and the probability of graduates going on to rated or technical careers. Logit models are used to predict the probability of graduates staying beyond eight years of service and attaining the rank of major. Only control variables available to the admissions board are considered in order to evaluate the effectiveness of legacy status as a signal of future performance.

Legacy status has no significant effect on GPA, order of merit, academic majors or Air Force rank. All other measures have statistically significant relationships with legacy status. Legacy admits are 10 percentage points more likely to graduate, and those legacy graduates have 0.04 points higher MPA. The increase in graduation probability comes mainly from a reduction in the likelihood that a legacy admit will voluntarily quit the Academy. The effect on probability of graduation increases as the academic qualifications of the students decrease. That means legacy status is more important for those students for whom the additional points awarded by a legacy policy are most beneficial.

Legacy status is positively correlated with career field and time in service. Legacy graduates are roughly 9 percentage points more likely to be rated officers and nearly 11 percentage points more likely to serve beyond 8 years. Extending the data set back to 1982 shows that military performance at the Academy is at least ten times as important as grades in predicting time in service and rank.

Theoretically, legacy status can impact the university process directly through graduation probability, indirectly by a selection effect through enrollment, and via a signaling effect through the conditional distribution of unobserved student characteristics. A model is developed to expand the selection theory, which, combined with numerical examples, demonstrates that empirical conclusions based on enrollment data do not necessarily generalize to admissions data. If that is the case, the results of this dissertation may not be useful to the Academy admissions office. This issue can only be resolved by further empirical work that looks at all applicants, not just enrolled students.

APPENDIX A
DATA SUMMARY

Each record contains the following fields (listed in alphabetical order):

ACT_Eng	Student's score on the English portion of the ACT exam.
ACT_Math	Student's score on the mathematics portion of the ACT exam.
ACT_Read	Student's score on the reading portion of the ACT exam.
ACT_Scir	Student's score on the science reasoning portion of the ACT exam.
AFA_Class	1994-2005. Student's class year at the Air Force Academy. There are no records missing this information.
AFA_Class_Size	Number of cadets who graduate from each Academy class. This is equal to the largest value for order of merit for each class. There are no records missing this information.
AFA_GPA	Final grade point average either before disenrolling or upon graduation. There are no records missing this information although 1,285 records have 0 GPA, possibly indicating cadets who left the Academy before the end of their first semester.
AFA_Grad	Graduated or Not Graduated. There are no records missing this information.
AFA_Major	Cadet's declared (non-graduates) or awarded (graduates) academic major. There are no records missing this information, although there are 2,492 records with "No Major." Of these only two are graduates (who probably did not meet the requirements for their declared major at the end of their last semester).
AFA_MPA	Final military performance average either before disenrolling or upon graduation. There are no records missing this information although 1,412 records have 0 MPA, possibly

indicating cadets who left the Academy before the end of their first semester.

AFA_OM — Order of merit for each cadet who graduates. This combines academic, military, and athletic scores. Records for non-graduates list a zero, which is replaced with a period to denote missing data in STATA. There are 28 graduates with zero order of merit, possibly because they graduated late.

AF_Rank — 2LT, 1LT, CAPT, MAJ, Lt Col, COL, BGEN. Current or last rank held in the Air Force as of July 2005. This information is missing for 1,019 graduates.

AF _Years — Number of years service in the Air Force. There are 882 graduates who are missing this information.

AFSC — Air Force Specialty Code. Designator for each officer's career field in the Air Force. There are 1,426 graduates who are missing this information. There are another 36 who have invalid AFSCs.

Athlete — Student's intercollegiate status at time of admission.

A	Blue Chip Athlete (Endorsed by Athletic Recruiting)
D	Coach loses interest
M	Monitored athletes
R	Recruited athletes

Based on discussions with the Academy's Plans and Analysis Division, the best proxy for intercollegiate athletic status are those cadets who have an "A" or "R" in this field. This is not a perfect measure because there can be recruited athletes who do not play on a team, just as there can be people who walk on to teams. Since other records (non-athletes) have blanks for this field, it is impossible to determine if there is any missing data for athletic status.

Entry_Age — Age of student when entering the Air Force Academy. There are no records missing this information.

Gender — Male or female. There are no records missing this information.

HS_GPA — Student's grade point average from high school. There are 1,832 records missing this field. Worse than missing data is the possibility of corrupt data. The values range from 0.04

to 9.98. There are 83 records between 0 and 2 and 370 records above 5.

HS_GPA_Scale The grading scale used at the student's high school. Unfortunately, this field is only available for the class of 2002 and later. Of the 4,986 records for 2002-2005, this field is missing for 476 of them and is less than the recorded GPA for 735 of them.

HS_Name Name of student's high school. There are only two records missing this field.

HS_Rank Student's graduating rank from high school. There are 2,798 records missing this field.

HS_Size Size of student's high school class. There are 2,675 records missing this field.

HS_State State from which the student graduated high school. The field includes postal abbreviations for all 50 states plus DC and the following:[1]

AA APO or FPO (Asia)
AE APO or FPO (Europe)
AP APO or FPO (Pacific)
AS Pago Pago Samoa
GU Guam
MP Mariana Islands
PR Puerto Rico
VI Virgin Islands
ZZ Overseas Address

The overseas military addresses (APO/FPO) are combined into a single location. The U.S. territories are also combined into a single location. Another location ("Missing") is created for a total of 55 locations: 50 states, DC, APO, Territory, Overseas, and Missing. There are 18 records in the Missing category.

HS_Year Year in which student graduated from high school. There are 26 records missing this field.

[1] There are also codes for Caroline Islands and Marshall Islands, but there are no records with these codes.

HS_ZIP	ZIP code for the student's high school. There are 124 records that are either blank or have a ZIP code of 0.
PAR_Score	Academic composite score awarded by Air Force Academy admissions. Only nine records are missing this field.
Parent_Academy	Indicates which service academy the student's parent attended: A U.S. Air Force Academy C U.S. Coast Guard Academy K U.S. Merchant Marine Academy M U.S. Military Academy (aka West Point) N U.S. Naval Academy (aka Annapolis) Since other records have blanks for this field, it is impossible to determine if there is any missing data.
Parent_Branch	Denotes parent's branch of military service: Army, Air Force, Coast Guard, Marines, or Navy. Since other records have blanks for this field, it is impossible to determine if there is any missing data.
Parent_Service	Denotes parent's military status 0 None (civilian) 1 Active duty 2 Active duty Reserve 3 Reserve 5 Retired from active duty 6 Deceased while on active duty 8 National Guard 9 Retired from Reserve 11 Retired from National Guard 12 Separated 13 Retired, not active duty There are no records missing this field.
PID	Primary key for the Air Force Academy database. This is a unique number assigned to each record.
Prior_Service	Denotes student's military status prior to entering the Academy. The codes are similar to Parent_Service except the only values are 0, 1, 3, and 8. There are no records missing this field.

Race	Asian, Black, Caucasian, Hispanic, Indian, Other, or Unknown. For the time period in question, there is no significant linear trend for any racial group. Other and Unknown are combined in order to ensure sufficient observations. After this adjustment, there are at least six members of each racial group in each class year. Only two records are missing this field.
SAT_Math	Student's score on the mathematics portion of the SAT.
SAT_Verb	Student's score on the verbal portion of the SAT.

Dummy variables are created for gender, race, Academy class, and high school state. The following fields are computed based on the data available:

8_Years	1 if AF_Years >= 8; only defined for AFA_Class between 1982 and 1997.
10_Years	1 if AF_Years >= 10; only defined for AFA_Class between 1982 and 1995.
15_Years	1 if AF_Years >= 15; only defined for AFA_Class between 1982 and 1990.
20_Years	1 if AF_Years >= 20; only defined for AFA_Class between 1982 and 1985.
ACT_Math_Ratio	ACT_Math divided by the average of ACT_Eng and ACT_Read to emulate SAT_Math_Ratio. See Appendix B.
ACT_Score	Recentered SAT scores are converted into composite ACT scores using tables from *The College Board*. After combining scores, there are only six records missing a standardized test score.
AFJob	2 if officer is in a technical field (see TechJob); 1 if officer is rated (see Rated); 0 for all other AFSCs.
AFA_Major	2 if major is science related (see Scientist); 1 if major is engineering related (see Engineer); 0 for all other majors.
AFA_OMp	AFA_OM divided by AFA_Class_Size. Academy order of merit as a percentage of class size so that order of merit can be compared between classes.

COL	1 if AF_Rank is "COL" or "BGEN" for AFA_Class between 1982 and 1987.
Comp_ACT	ACT composite score. Average of ACT_Eng, ACT_Math, ACT_Read, ACT_Scir for all records that have all four individual ACT scores (6,498 records).
Dropout	1 if AFA_GPA = 0 and AFA_Grad = Not Graduated; assumes student left the academy before the end of the first semester. There are 1,284 students with AFA_GPA = 0.
Engineer	1 if AFA_Major is an engineering field. These include: AeroEngr AstroEngr CivEngr CivEngrEnv CompEngr ElEngr Engr EngrMech EngrSci EnvEngr GenEngr MechEngr SpaceOps There are 3,062 records that meet this criterion.
Grad_Fail_Quit	0 for graduates; 1 if non-graduate with AFA_GPA between 0 and 2 (fail); 2 if non-graduate with AFA_GPA = 0 or ≥ 2 (quit)
HS_Rankp	HS_Rank divided by HS_Size. High school order of merit as a percentage of class size so that class standings can be compared between schools.
High_Math_Ratio	0 if Math_Ratio ≤ 0.97; Math_Ratio – 0.97 if Math_Ratio > 0.97. This is the upper portion of the spline, which allows the linear relationship between Math_Ratio and graduation rate to change for higher levels of Math_Ratio.
High_PAR	0 if PAR_Score ≤ 600; PAR_Score – 600 if PAR_Score > 600. This is the upper portion of the spline, which allows the linear relationship between PAR_Score and graduation rate to change for higher levels of PAR_Score.

High_SAT	0 if SAT_Score ≤ 1280; SAT_Score – 1280 if SAT_Score > 1280. This is the upper portion of the spline, which allows the linear relationship between SAT_Score and graduation rate to change for higher levels of SAT_Score.
Intercollegiate	1 if Athlete = "A" or "R." There are 3,808 records that meet this criterion.
Legacy	1 if Parent_Academy = "A." There are 466 (3%) records that meet this criterion.
Low_Math_Ratio	Math_Ratio if Math_Ratio ≤ 0.97; 0.97 if Math_Ratio > 0.97. This is the lower portion of the spline.
Low_PAR	PAR_Score if PAR_Score ≤ 600; 600 if PAR_Score > 600. This is the lower portion of the spline.
Low_SAT	SAT_Score if SAT_Score ≤ 1280; 1280 if SAT_Score > 1280. This is the lower portion of the spline.
LTC	1 if AF_Rank is "Lt Col" or "COL" or "BGEN" for AFA_Class between 1982 and 1989.
MAJ94	1 if AFA_Class is 1994 and AF_Rank is "MAJ." There are 575 majors among the 1,024 graduates from the class of 1994 (56%).
MAJ95	1 if AFA_Class is 1995 and AF_Rank is "MAJ." There are 9 majors among the 993 graduates from the class of 1995 (1%).
Math_Ratio	Combines ACT_Math_Ratio and SAT_Math_Ratio. Since the Academy only keeps the best score, this field captures the ratio for whichever exam the student took.
Military_ Background	1 if Parent_Service > 0 and Parent_Academy is blank. This captures military backgrounds for non-legacy admits. There are 2,575 records that meet this criterion.
New_SAT_Math	*The College Board* recentered SAT scores in 1995 to account for differences in score distributions between 1947 and 1990. SAT_Math scores are converted to recentered scores for all students who graduated high school prior to 1996. The year is chosen by assuming students take the

	SAT in the spring of their junior year or fall of their senior year (i.e., class of 1996 took the SAT in 1995).[2]
New_SAT_Verb	SAT_Verb converted to recentered score for all students who graduated high school prior to 1996.
Other_Academy	1 if Parent_Academy is not blank or "A" (i.e., any service academy other than the Air Force Academy). There are 209 records that meet this criterion.
Prior	1 if Prior_Service > 0 (i.e., any form of military service). Unfortunately, there is no way to tell the difference between actual enlisted service in the military and people who simply attended the Air Force Academy Prep School. There are 2,044 records that meet this criterion.
Rated	1 if AFSC starts with 11 (pilot), 12 (navigator), or 92T (pilot or navigator trainee). There are 4,898 records that meet this criterion.
SAT_Math_Ratio	New_SAT_Math divided by New_SAT_Verb based on Maloney and McCormick (1993). See Appendix B.
SAT_Score	Composite ACT scores are converted to equivalent recentered SAT scores using tables from *The College Board*. After combining scores, there are only six records missing a standardized test score.
Scientist	1 if AFA_Major is a science related field. These include: BioChem Biology Chem ChemGen CompSci CompSciIA CompSciSci CompSciSys GeogMet Math MathAM MathMA MatlSci Meteor

[2] The results do not change significantly if using 1995 or 1997 as the cutoff.

	OpsRsch Physics PhysicsApl PhysicsATM PhysicsSpa
	There are 2,467 records that meet this criterion.
TechJob	1 if AFSC starts with:
	13A Astronaut 13S Space and Missiles 15 Weather 32 Civil Engineer 61 Scientist 62 Developmental Engineer
	There are 1,050 records that meet this criterion.
Total_SAT	Adds SAT_Math and SAT_Verb for all records that have both SAT scores (8,572 records).

APPENDIX B
SAT AND ACT CONVERSIONS

Recentering is done on SAT scores for all students who graduated from high school prior to 1996. Table B-1 shows how the mean and standard deviation for SAT scores change. Figure B-1 shows how the recentered scores appear much closer in distribution to the scores for students who graduated in 1996 or later.

Because the Academy only records an applicant's highest standardized test score, many students have an SAT score, but not an ACT score, and vice versa. In order to have a single test score for the models in this dissertation, a conversion from *The College Board* is used to turn ACT scores into comparable recentered SAT scores. Table B-2 and Figure B-2 show the distribution of SAT scores is not changed dramatically by converting composite ACT scores to recentered SAT scores.

Following Maloney and McCormick (1993), a math ratio is computed in order to account for skewed test scores where students perform better (or worse) on the quantitative section versus the verbal section. For SAT scores, the ratio is simply SAT_Math/SAT_Verb. For ACT scores, the math score is divided by the average of the English and reading scores: ACT_Math/(ACT_Eng + ACT_Read)/2. Table B-3 and Figure B-3 show the distributions of the two ratios are nearly identical. There are only three observations for SAT-based ratios that are above the ACT-based maximum of 1.6.[1]

[1] The figures in this appendix omit the 730 records identified as bad data, but the results are very similar if that data is included.

Table B-1. Summary Statistics for Recentered SAT Scores

	Obs	Mean	Std. Dev.	Min	Max
<1996	4015	1227.33	99.79	890	1590
<1996 Recentered	4015	1296.53	92.64	990	1600
1996 or Later	4105	1285.80	104.05	860	1600

Table B-2. Summary Statistics for SAT Scores from Converted ACT Scores

	Obs	Mean	Std. Dev.	Min	Max
SAT Only	8120	1291.14	98.71	860	1600
With ACT	14340	1297.92	98.59	860	1600

Table B-3. Summary Statistics for SAT and ACT Based Math Ratios

	Obs	Mean	Std. Dev.	Min	Max
SAT	8120	1.0420	0.1087	0.6471	1.9714
ACT	6226	1.0291	0.1194	0.7059	1.6000
Combined	14340	1.0363	0.1136	0.6471	1.9714

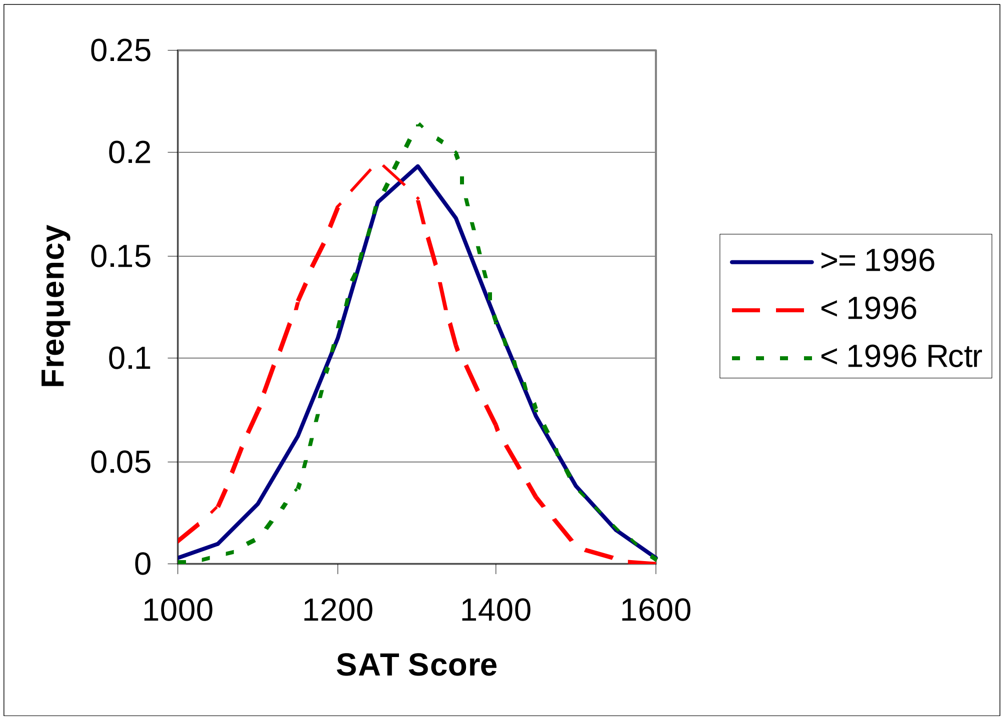


Figure B-1. Distributions of Regular and Recentered SAT Scores

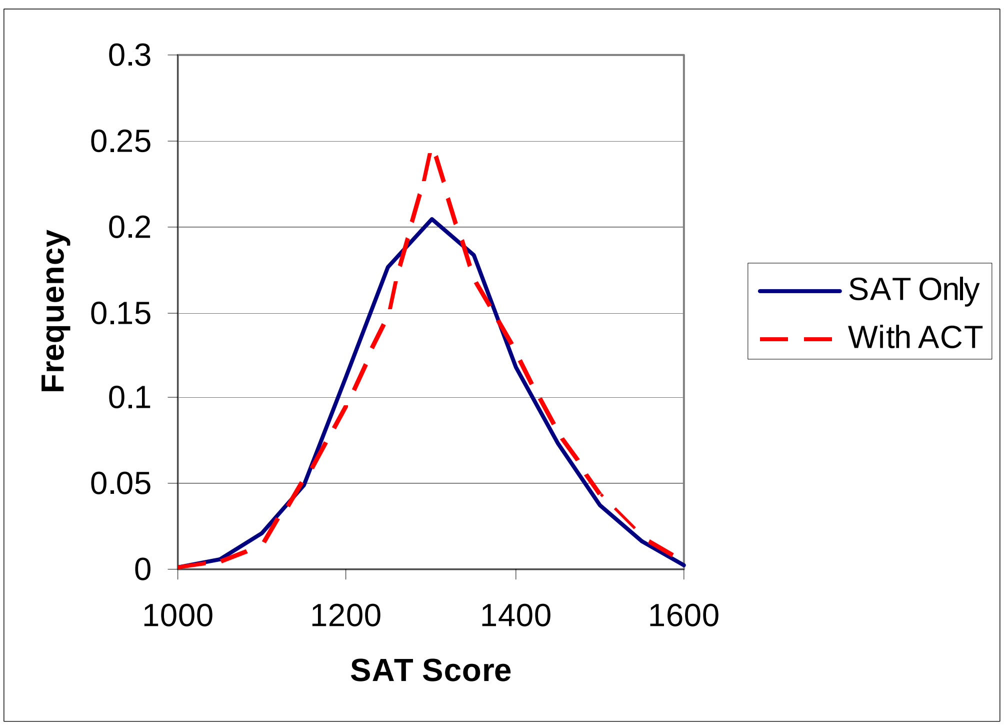


Figure B-2. Distributions of Recentered and Converted SAT Scores

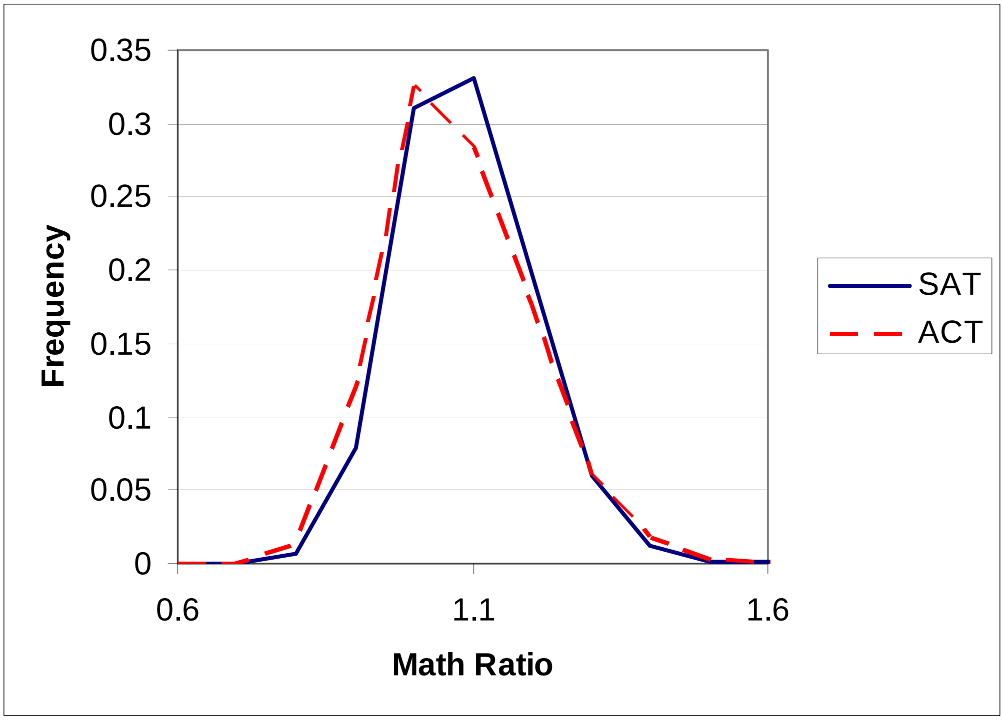


Figure B-3. Distributions of SAT and ACT Based Math Ratios

LIST OF REFERENCES

Altonji, Joseph G. and Charles R. Pierret. "Employer Learning and Statistical Discrimination." *The Quarterly Journal of Economics*. Vol. 116, No. 1 (Feb 2001), 293-312.

Attiyeh, Gregory and Richard Attiyeh. "Testing for Bias in Graduate School Admissions." *The Journal of Human Resources*. Vol. 32, No. 3 (Summer 1997), 524-548.

Autor, David H. and David Scarborough. "Will Job Testing Harm Minority Workers?" NBER Working Paper 10763 (Sep 2004).

Becker, Gary S. and Nigel Tomes. "Human Capital and the Rise and Fall of Families." *Journal of Labor Economics*. Vol. 4, No. 3 (Jul 1986), 1-39.

Behrman, Jere R., Mark R. Rosenzweig and Paul Taubman. "College Choice and Wages: Estimates using Data on Female Twins." *The Review of Economics and Statistics*. Vol. 78, No. 4 (Nov 1996), 672-685.

Benabou, Roland. "Equity and Efficiency in Human Capital Investment: The Local Connection." *Review of Economic Studies*. Vol. 63 (1996), 237-264.

Betts, Julian R. and Darlene Morell. "The Determinants of Undergraduate Grade Point Average: The Relative Importance of Family Background, High School Resources, and Peer Group Effects." *The Journal of Human Resources*. Vol. 34, No. 2 (Spring 2000), 268-293.

Bishop, John. "The Effects of Public Policies on the Demand for Higher Education." *The Journal of Human Resources*. Vol. 12, No. 3 (Summer 1977), 285-307.

Black, Sandra E., Paul J. Devereux and Kjell G. Salvanes. "Why the Apple Doesn't Fall Far: Understanding Intergenerational Transmission of Human Capital." NBER Working Paper No. 10066 (Oct 2003).

Blanchfield, William C. "College Dropout Identification: An Economic Analysis." *The Journal of Human Resources*. Vol. 7, No. 4 (Autumn 1972), 540-544.

Bound, John, Zvi Griliches and Bronwyn H. Hall. "Wages, Schooling and IQ of Brothers and Sisters: Do the Family Factors Differ?" *International Economic Review*. Vol. 27, No. 1 (Feb 1986), 77-105.

Brewer, Dominic J., Eric R. Eide and Ronald G. Ehrenberg. "Does it Pay to Attend an Elite Private College?" *The Journal of Human Resources*. Vol. 34, No. 1 (Winter 1999), 104-123.

Bush, George W. "President's Remarks to the Unity: Journalists of Color Convention." Washington Convention Center, Washington, D.C. 6 Aug 2004. http://www.whitehouse.gov/news/releases/2004/08/20040806-1.html. Accessed Jun 2005.

Card, David and Alan B. Krueger. "Does School Quality Matter? Returns to Education and the Characteristics of Public Schools in the United States." *Journal of Political Economy*,. Vol. 100, No. 1 (Feb 1992), 1-40.

Cascio, Elizabeth and Ethan Lewis. "Schooling and the AFQT: Evidence from School Entry Laws." IZA Discussion Paper No. 1481 (Jan 2005).

Coelli, Michael B. "Parental Income Shocks and the Education Attendance of Youth." University of British Colombia. Job Market Paper (21 Dec 2004).

Coleman, James S. "Social Capital in the Creation of Human Capital." *The American Journal of Sociology*. Vol. 94 (1988), 95-120.

The College Board. "2000 SATI-ACT Score Comparisons." http://www.collegeboard.com/sat/cbsenior/html/stat00f.html. Accessed Jun 2005.

The College Board. "The Effects of SAT Scale Recentering on Percentiles." May 1999. http://www.collegeboard.com/research/pdf/rs05_3962.pdf. Accessed Jun 2005.

Corazzini, Arthur J., Dennis J. Dugan and Henry G. Grabowsky. "Determinants and Distributional Aspects of Enrollment in US Higher Education." *The Journal of Human Resources*. Vol. 7, No. 1 (Winter 1972), 39-59.

Datcher, Linda. "Effects of Community and Family Background on Achievement." *The Review of Economics and Statistics*. Vol. 64, No. 1 (Feb 1982), 32-41.

Daymont, Thomas and Paul Andrisani. "Job Preference, College Major, and the Gender Gap in Earnings." *The Journal of Human Resources*. Vol. 19, No. 3 (Summer 1984), 408-428.

Dolan, Robert C., Clarence R. Jung, Jr., and Robert M. Schmidt. "Evaluating Educational Inputs in Undergraduate Education." *The Review of Economics and Statistics*. Vol. 67, No. 3 (Aug 1985), 514-520.

Eckstein, Zvi and Kenneth I. Wolpin. "Why Youths Drop Out of High School: The Impact of Preferences, Opportunities, and Abilities." *Econometrica*. Vol. 67, No. 6 (Nov 1999), 1295-1339.

Ehrenberg, Ronald G. "Adam Smith Goes to College: An Economist Becomes an Academic Administrator." *Journal of Economic Perspectives*. Vol. 13, No. 1 (Winter 1999), 99-116.

Epple, Dennis, Richard Romano and Holger Seig. "The Practice and Proscription of Affirmative Action in Higher Education: An Equilibrium Analysis." Carnegie Mellon University. Draft (Jun 2003).

---. "Diversity, Profiling, and Affirmative Action in Higher Education." Carnegie Mellon University. Draft (2006).

Evans, William N. and Robert M. Schwab. "Finishing High School and Starting College: Do Catholic Schools Make a Difference?" *The Quarterly Journal of Economics*. Vol. 110, No. 4 (Nov 1995), 941-974.

Fryer, Roland G., Jr., Glenn C. Loury and Tolga Yuret. "Color-Blind Affirmative Action." Harvard University. Draft (Oct 2003).

Fuller, Winship C., Charles F. Manski and David A. Wise. "New Evidence on the Economic Determinants of Postsecondary Schooling Choices." *The Journal of Human Resources*. Vol. 17, No. 4 (Autumn 1982), 477-498.

Goldstein, Amy. "Bush Hits 'Legacy' College Admissions." *Washington Post*, 7 Aug 2004, A08.

Greene, William H. "Chapter 21: Models for Discrete Choice." *Econometric Analysis*. 5th ed. Upper Saddle River: Prentice Hall (2003), 663-755.

Havemen, Robert and Barbara Wolfe. "The Determinants of Children's Attainments: A Review of Methods and Findings." *Journal of Economic Literature*. Vol. 33, No. 4 (Dec 1995), 1829-1878.

Hungerford, Thomas and Gary Solon. "Sheepskin Effects in the Returns to Education." *The Review of Economics and Statistics*. Vol. 69, No. 1 (Feb 1987), 175-177.

Husted, Thomas A. and Lawrence W. Kenny. "Evidence on the Impact of State Government on Primary and Secondary Education and the Equity-Efficiency Trade-Off." *Journal of Law and Economics*. Vol. 43, No. 1 (Apr 2000), 285-308.

Jones, Ethel and John D. Jackson. "College Grades and Labor Market Rewards." *The Journal of Human Resources*. Vol. 25, No. 2 (Spring 1990), 253-266.

Kane, Thomas J. and Cecilia Elena Rouse. "Labor Market Returns to Two-Year and Four-Year College: Is a Credit a Credit and Do Degrees Matter?" NBER Working Paper No. 4268 (Jan 1993).

Laband, David N. and Bernard F. Lentz. "Self-Recruitment in the Legal Profession." *Journal of Labor Economics*. Vol. 10, No. 2 (Apr 1992), 182-201.

Lassila, Kathrin. "Why Yale Favors Its Own." *Yale Alumni Magazine*. Vol. 68, No. 2 (Nov/Dec 2004), http://www.yalealumnimagazine.com/issues/2004_11/q_a.html. Accessed Feb 2006.

Lazear, Edward. "Academic Achievement and Job Performance: Note." *American Economic Review*. Vol. 67, No. 2 (Mar 1977), 252-254.

Lentz, Bernard F. and David N. Laband. "Why So Many Children of Doctors Become Doctors. Nepotism vs. Human Capital Transfers." *The Journal of Human Resources*. Vol. 24, No. 3 (Summer 1989), 396-413.

Light, Audrey and Wayne Strayer. "Determinants of College Completion: School Quality or Student Ability?" *The Journal of Human Resources*. Vol. 35, No. 2 (Spring 2000), 299-332.

Lott, John R., Jr. and Lawrence W. Kenny. "Did Women's Suffrage Change the Size and Scope of Government?" *Journal of Political Economy*. Vol. 107, No. 6, Pt. 1 (1999), 1163-1198.

Loury, Linda D. and David Garman. "College Selectivity and Earnings." *Journal of Labor Economics*. Vol. 13, No. 2 (Apr 1995), 289-308.

Lowe, John C. and Arthur Viterito. "Differential Spatial Attraction of Private Colleges and Universities in the United States." *Economic Geography*. Vol. 65, No. 3 (Jul 1989), 208-215.

Maloney, Michael T. and Robert E. McCormick. "An Examination of the Role that Intercollegiate Athletic Participation Plays in Academic Achievement." *The Journal of Human Resources*. Vol. 28, No. 3 (Summer 1993), 555-570.

McCormick, Robert E. and Maurice Tinsley. "Athletics versus Academics? Evidence from SAT Scores." *Journal of Political Economy*. Vol. 95, No. 5 (Oct 1987), 1103-1116.

"Naked Hypocrisy: The Nationwide System of Affirmative Action for Whites." *The Journal of Blacks in Higher Education*. No. 18 (Winter 1997/1998), 40-43.

Olmstead, Alan L. and Steven M. Sheffrin. "The Medical School Admission Process: An Empirical Investigation." *The Journal of Human Resources*. Vol. 16, No. 3 (Summer 1981), 459-467.

Oreopoulos, Philip, Marianne E. Page and Ann Huff Stevens. "Does Human Capital Transfer from Parent to Child? The Intergenerational Effects of Compulsory Schooling." NBER Working Paper No. 10164 (Dec 2003).

Orlans, Harold. "Affirmative Action in Higher Education." *The Annals of the American Academy of Political and Social Science*. Vol. 523 (Sep 1992), 144-158.

Pruden, William H. III. Response to Schmidt's paper. *The Chronicle of Higher Education*. Vol. 50, Issue 27 (13 Mar 2004), B21.

Radner, R. and L.S. Miller. "Demand and Supply in Higher Education: A Progress Report." *American Economic Review*. Vol. 60, No. 2 (May 1970), 326-334.

Rothschild, Michael and Lawrence J. White. "The Analytics of Pricing in Higher Education and Other Services in Which Customers are Inputs." *Journal of Political Economy*. Vol. 103 (Jun 1995), 573-586.

Sander, William and Anthony C. Krautmann. "Catholic Schools, Dropout Rates and Educational Attainment." *Economic Inquiry*. Vol. 33, No. 2 (Apr 1995), 217-233.

Sanoff, Alvin P. "Americans See Money for College Somewhere Over the Rainbow." *The Chronicle of Higher Education*. Vol. 50, Issue 34 (30 Apr 2004), B6.

Schaffner, Paul E. "Competitive Admission Practices When the SAT is Optional." *Journal of Higher Education*. Vol. 56, No. 1 (Jan/Feb 1985), 55-72.

Schmidt, Peter. "New Pressure Put on Colleges to End Legacies in Admissions." *The Chronicle of Higher Education*. Vol. 50, Issue 21 (30 Jan 2004), A1.

Shea, John. "Does Parents' Money Matter?" Journal of Public Economics. Vol. 77, No. 2 (Aug 2000), 155-184.

Stata Reference Manual. Release 7. Volume 2 H-P. College Station: Stata Press (2001), 248-259, 358-370, 580-594.

Status of Higher Education Act Bills. 26 Jan 2004. http://www.nyu.edu/ofp/pdf/Summary.pdf. Accessed Jun 2005.

Thornberry, Terence P., Melanie Moore and R.L. Christenson. "The Effect of Dropping out of High School on Subsequent Criminal Behavior." *Criminology*. Vol. 23, No. 1 (Feb 1985), 1-40.

Tuckman, Howard P. "College Presence and the Selection of College." *Land Economics*. Vol. 47, No. 2 (May 1971), 198-205.

Willis, Robert J. and Sherwin Rosen. "Education and Self-Selection." *Journal of Political Economy*. Vol. 87, No. 5 (Oct 1979), 7-36.

Winston, Gordon C. "Subsidies, Hierarchy and Peers: The Awkward Economics of Higher Education." *Journal of Economic Perspectives*. Vol. 13, No. 1 (Winter 1999), 13-36.

Winston, Gordon C. and David J. Zimmerman. "Peer Effects in Higher Education." NBER Working Paper No. 9501 (Feb 2003).

Wise, David. "Academic Achievement and Job Performance." *American Economic Review*. Vol. 65, No. 3 (Jun 1975), 350-366.

BIOGRAPHICAL SKETCH

Len Cabrera was born and raised in Miami, Florida, where he graduated from Southwest Miami High in 1991. He was commissioned as an Air Force officer and received a Bachelor of Science in operations research from the United States Air Force Academy in 1995. The following year, he earned a Master of Science in the same field from Stanford University. From 1996 to 1998, he worked as a survivability analyst for Detachment 1, 31st Test and Evaluation Squadron, Air Combat Command, at Kirtland Air Force Base, New Mexico. Then he was a flight test analyst for the 18th Flight Test Squadron, Air Force Special Operations Command, at Hurlburt Field, Florida, from 1998 to 2001. While there, he earned a Master of Business Administration from the University of West Florida. In 2001, he became an instructor of economics and operations research at the United States Air Force Academy, where he was selected to enter a PhD program in 2003. After graduating from the University of Florida, he will be assigned to the joint staff of the United States Transportation Command at Scott Air Force Base, Illinois.