The impact of legacy status on undergraduate admissions at elite colleges and universities

Michael Hurwitz

The Harvard Graduate School of Education

December, 2009

Acknowledgements

This qualifying paper would not have been possible without the help of my colleagues and my advisory committee at the Harvard University Graduate School of Education. First, I would like to thank John Willett and Bridget Terry Long for helping me to transform my ideas into a polished, final research paper. I relied greatly on their substantive and methodological expertise, and I'm immeasurably appreciative for the support they have offered across many dimensions. Second, my colleagues – Ann Birk, Kristine Dillon, and Steve Minicucci – have been invaluable assets in helping me to refine my ideas and, most importantly, to secure the data that serves as a foundation for the subsequent analyses.

Table of Contents

The impact of legacy status on undergraduate admissions at elite colleges and universities

Abstract

In this paper, I examine the impact of legacy status on the undergraduate admissions process for first-year, American citizens applying for entry in the fall of 2007 to 30 highly-selective colleges and universities Unlike other quantitative studies addressing this topic, I use conditional logistic regression with fixed effects for colleges, rather than basic logistic regression, to draw conclusions about the impact of legacy status on admissions odds. Through this methodological technique, I eliminate most sources of outcome bias by controlling for all applicant characteristics that are constant across colleges and all college characteristics that are constant across applicants. I estimate that the odds of admission for applicants with legacy status are 3.13 times the odds for those without legacy status. Moving beyond the previous literature, however, the results suggest that the magnitude of this legacy admissions advantage depends greatly on the nature of the familial ties between the applicant and the outcome college, and the selectivity of the outcome college. In contrast, I do not find a clear relationship between an applicant's academic strength and the admissions advantage granted to the legacy applicant.

The impact of legacy status on undergraduate admissions at elite colleges and universities

Introduction

Recent public attention drawn to the influence of legacy status in undergraduate college admissions has provoked both qualitative and quantitative research addressing this topic (Shulman and Bowen, 2001; Espenshade, Chung and Walling, 2004; Golden, 2006). These studies arrive at the same conclusion almost universally – legacy status matters. Previous research has been influential in laying a foundation for understanding this topic, yet most of these studies have generally failed to account for the many ways that legacy students differ from non-legacy students. That is, applicants with familial ties to an institution may also differ from other applicants in important ways unrelated to their legacy status.

In this paper, one goal is to account for bias in estimates of legacy admissions advantage present in the findings of studies that have applied more traditional analytic methods, such as simply comparing acceptance rates between legacy and non-legacy students or using basic logistic regression to estimate the legacy admissions advantage. The structure of the data set, in which student applications to multiple highly-selective colleges and universities are observed, allows me to apply conditional logistic regression analysis to account for the fixed effects of a particular applicant. In addition, I also control for the fixed effects of colleges. Using this approach, I eliminate bias in the estimate of the impact of legacy status that is due to applicant characteristics that are invariant across the multiple institutions to which an individual applies (e.g. high school GPA, extracurricular activities, rigor of high school coursework) as well as the relative selectivity of the sampled colleges. By removing these two types of important variability in the admissions outcome, I isolate the impact on admissions of legacy status, an applicant-varying characteristic that depends on both the college and the applicant.

In addition to estimating the overall legacy admissions advantage across 30 highly selective colleges, this paper, unlike previous literature, demonstrates that the *type* of familial connection that exists between the applicant and the school impacts the admissions advantage conferred to the legacy applicant greatly. Moreover, the legacy admissions advantage is largely independent of the applicant's academic characteristics; it is larger at schools with more competitive admissions processes; and the legacy admissions advantage is largest when legacies apply through early admissions programs. Finally, contending that the results from the conditional logistic regression analyses provide the best estimates of the legacy admissions advantage, I also present corresponding estimates of the legacy admissions advantage from parallel basic logistic regression analyses that also contain the fixed effects of colleges. Where relevant, I compare the estimates obtained by the different analytic techniques, and these results suggest that, on average, basic logistic regression underestimates the true impact of legacy status on the odds of admission. This finding implies that the estimates of the legacy admissions advantage found in other studies from basic logistic regression may be biased downwards.

Background and Context

The Landscape of Undergraduate College Admissions

During the past decade, the heightened competition to win undergraduate admission to America's elite post-secondary institutions has resulted in increasing numbers of academically talented high school seniors facing rejection from their top-choice schools. Fueled partially by the echo of the baby-boom, the increasing numbers of applications submitted per applicant (Spivack, 2009), the expanding international applicant pools, and increasingly generous financial aid packages, the continually decreasing acceptance rates have transformed the college-admissions landscape (Bound, Hershbein, and Long, 2009). Applicants who might have been shoe-ins at America's most selective institutions a decade ago are now finding themselves on expansive waiting lists.

Because of the unpredictability of admissions decisions at these institutions, students (and their families) work hard to send signals of academic achievement and extracurricular excellence to their choice colleges (Bound, Hershbein, and Long, 2009). Helping applicants to design their high school curriculum strategically and to craft compelling admissions essays, independent college counselors have capitalized on this admissions frenzy—charging upwards of $40,000 for their "expert" advice (Berfield and Tergesen, 2007).

In general, a student's academic record, including standardized test scores, is the most important component of his college application, and some colleges use these academic measures as primary determinants of admission (Bunzel, 1996). Opponents of this one-dimensional approach to admissions contend that a student's non-academic attributes bring interesting and diverse talents and experiences that create an atmosphere where students can learn effectively from each other inside and outside of the classroom (Orfield, 1999). Traditionally, a variety of attributes, including musical talent, geographic residence, and the potential to make sizeable financial donations can impact whether or not a student is deemed worthy of admission (Klitgaard, 1985; Duffy and Goldberg, 1998; Greene and Greene, 2009). However, the three categories receiving the most attention are race/ethnicity, athletic status, and familial ties to a college (legacy status).

In their sample of 13 highly selective post-secondary institutions Bowen, Kurzweil and Tobin (2005) estimate that the minority admissions advantage is roughly 27.7 percentage points. Kane (1998) offers support of the minority admissions advantage by pointing out that Black students at several highly selective post-secondary institutions lag behind their White peers by over 100 points. Preferences by race/ethnicity have also drawn public scrutiny. In 2003, the debate over racial affirmative action came to a head when the U.S. Supreme Court ruled that race-conscious admissions were permissible in *Grutter v. Bollinger (2003)*. The sister case, *Gratz v. Bollinger (2003)*, upheld the Supreme Court's 1978 decision in *University of California v. Bakke (1978)* that quota systems cannot be used in college admissions. Although these two court cases were heralded as a disappointment for the staunch critics of affirmative action for minority applicants, they served as a reminder to colleges that their admissions practices were not immune from judicial supervision.

There has also been extensive analysis on the preference exercised for college athletes. Shulman and Bowen in *The Game of Life: College Sports and Educational Values* (2001) analyzed preferences for athletes and found that high school athletes capable of competing at the collegiate level are given a sizeable boost in the admissions process. At one non-scholarship institution in Shulman and Bowen's study, athletes enjoyed a 50 percent admissions advantage over non-athletes, after controlling for SAT scores. Compared to the admissions advantage Shulman and Bowen found for minorities

(19 percent) at this particular institution, the enormous admissions boost offered to athletes provides insight into the institution's athletic prioritization.

The Legacy Question

Preferences for legacy candidates have existed far longer than the other two major categories. At the beginning of the 20th Century, admissions committees at America's most selective post-secondary institutions may have embraced legacy status as a mechanism for curbing the growing Jewish population in the Ivy League (Karabel, 2005). Concurrent with this anti-Semitic exclusion was the realization that maintaining an academically and "socially" excellent institution required money. Appeasing wealthy alumni meant accepting their relatives and sustaining the family traditions that motivated financial donations (Karabel, 2005). Though the enveloping cloud of anti-Semitism has since lifted, the college admissions process is still very much influenced by alumni financial power (Golden, 2006). In addition, new demands on colleges to keep alumni happy have emerged from rankings magazines like *U.S. News and World Report*. The alumni satisfaction measure in *U.S. News and World Report* reports the percentage of alumni who donate to their alma maters, without regard to the total amount of money donated. Because the alumni satisfaction measure is used to determine an institution's overall score and final rank, efforts to encourage alumni donations- even a few dollars – are particularly important (Golden, 2007).

Empirical evidence on legacy admissions preferences confirms that these students are looked upon favorably by admissions committees. According to the *Stanford Daily* (Weiss, 2005), Princeton accepts 35 percent of children of alumni, while the comparable acceptance rates hover around 40 percent for the University of Pennsylvania and

Harvard. These legacy acceptance rates are up to 4 times larger than the overall acceptance rates at these three Ivy League institutions. Bowen, Kurzweil, and Tobin (2005) found that the legacies in their sample received an admissions advantage of 19.7 percentage points – slightly smaller than for minorities and athletes. However, at one highly selective post-secondary institution highlighted by Shulman and Bowen (2001), legacies received an admissions advantage of nearly 24 percentage points.

Espenshade, Chung, and Walling (2004) discuss the legacy admissions advantage as an odds-ratio, rather than a probability. Examining 124,374 applications to ten different selective colleges in the 1980s and 1990s, these authors found that the estimated odds of admission for legacy applications were 3.13 times those of non-legacy applications. After accounting for SAT scores, gender, ethnicity and U.S. citizenship, the authors found that the admission advantage was marginally eroded, and the estimated odds-ratio fell slightly to 2.91 but remained statistically different from zero ($p<0.001$). This decline in the odds of admission for legacy applicants after controlling for a vector of reasonable covariates may suggest that acceptance rate differences between legacies and non-legacies overstate the true legacy admissions advantage.

The focus on legacy students is not limited to the admissions realm. Using data from the *National Longitudinal Survey of Freshmen* (NLSF) and controlling for a large set of academic preparation measures and demographic characteristics, Massey and Mooney (2007) show that legacies, unlike athletes and minority students, tend to earn lower grades in college when their SAT scores are below the institutional mean. A more recent study conducted at Duke University by Martin and Spenner (2009) showed that Duke legacy students lagged behind their peers in the classroom during the first semester of college. They also found that legacies tend to major in humanities, which may have more generous grading curves, and shy away from coursework in natural sciences and engineering (Jaschik, 2008; Martin and Spenner, 2009). The evidence from these two studies suggests that legacy admits do not fare as well as their peers after enrolling in college.

Given the financial returns associated with attending an elite post-secondary institution (Hoxby and Terry, 1999), as well as the increased probability of attending an elite graduate school (Eide, Brewer and Ehrenberg, 1998), the allocation of undergraduate spots to applicants who might be deemed undeserving would compromise equity. As stated by former democratic presidential candidate, John Edwards, these legacy preferences represent, "a birthright out of 18^{th} -century British aristocracy, not 21^{st}-century American democracy" (Lexington, 2004).

Furthermore, giving preferences to legacy candidates may also negatively impact an institution's goal to enhance campus diversity (Rimer, 2007). Because legacies at America's most selective post-secondary institutions are disproportionately White (Howell and Turner, 2004), awarding preference to children or close relatives of alumni could pose an impediment to racial diversity. The heightened admissions rates of legacy candidates among applicants at Ivy League institutions coupled with the fact that, as of 15 years ago, nearly 96 percent of Ivy League alums were White (Megalli, 1995) could help to preserve racial homogeneity within these schools. Among the admitted legacy applicants at Harvard in 2002, only 7.6 percent came from an underrepresented minority group (Golden, 2003). Of course, as Howell and Turner (2004) suggest, the increasing numbers of minority students enrolling in America's most selective post-secondary institutions will affect the racial composition of the legacy applicant pools in future generations. Nevertheless, in the near future, legacy preference will not help colleges to achieve a level of socio-economic and racial diversity that mirrors that of the overall American population.

Though affirmative action for minority students, and to a much lesser extent, athletic recruiting, have been under attack for decades, attention directed towards legacy preferences in college admissions is more recent (Leef, 2008). In 2003, Senator Kennedy (D-Mass) lobbied for legislation requiring colleges to publicly provide extensive data on the admissions profiles of legacy students. In the *Price of Admission* (2006), Golden profiles legacy applicants to the University of Notre Dame. Children of Fighting Irish alumni constitute almost one-quarter of the Notre Dame undergraduate population-much larger than the 10-15 percent figure at many Ivy League institutions. The story depicted by Golden (2006), in which already socio-economically and educationally advantaged applicants are given an additional admissions boost to Notre Dame is fodder for critics of legacy preferences.

Critiquing the Literature on Preferences

How often are putatively more qualified applicants passed up for less qualified ones? Thwarted applicants often perceive that their spots were given to arguably less-qualified students admitted for non-academic reasons (Kane, 2003). The reality is that so many academically exceptional applicants are rejected by the nation's most selective post-secondary institutions that removing non-academic characteristics from the admissions process would be unlikely to change the number of rejection letters received by any given applicant (Thomas and Shepard, 2003). In fact, the holistic admissions approach taken by these institutions (Hernandez, 1997; Steinberg, 2002) means that it is impossible to create a rank-order of applicants based on a composite of academic and non-academic characteristics. Generally speaking, students with strong secondary school records and high SAT scores stand a better shot at gaining admission; however, there are no guarantees in the college admissions game. For example, in 2007, Harvard rejected 1100 applicants with a perfect math SAT score, and Princeton said “no” to thousands of high school students maintaining perfect GPAs (Dillon, 2007).

Absence of a concrete admissions formula makes the interpretation of anecdotes particularly tricky, as characteristics unavailable to the researcher but available to the admissions officer (e.g. personal qualities, leadership potential) may propel an applicant from the waitlist pile to the accept pile, rather than her legacy, athlete or minority status. Discussing the injustices of a non-egalitarian admissions system by pointing to specific cases is of limited value without access to the student’s complete admissions package, including the teacher and guidance counselor recommendations, application essay, etc. For example, between two applicants, the seemingly more qualified candidate with a higher SAT score and high school grades may have been less engaged academically than the second applicant with lower quantifiable characteristics. These non-quantifiable attributes might have been conveyed through teacher recommendations, for instance. An outsider without access to the applicants’ teacher recommendations might be surprised by the admissions outcomes of these two high school students, and might search through a string of observable characteristics (e.g. legacy, athlete, or minority status) to explain this perplexing scenario. However, the reality of college admissions is far more complex.

Individual decisions can rarely be boiled down to one attribute, and attempting to identify the *cause* of an individual decision will generally yield spurious conclusions.

Another complication in the research is that legacies are markedly different from non-legacies on multiple criteria important to admissions committees, such as SAT scores, underrepresented minority status, and wealth. Between-group differences in measurable criteria suggest the existence of between-group differences in characteristics that *cannot* be adequately measured. Consequently, estimates of the legacy admissions advantage based on raw acceptance rates or basic logistic regression estimates obtained with controls for measurable characteristics (discussed below) will likely be biased.

The previously mentioned quantitative research on legacy preferences have used basic logistic regression to estimate the legacy advantage (Shulman and Bowen, 2001; Espenshade, Chung and Walling, 2004; Bowen, Kurzweil and Tobin, 2005). Similarly, Grodsky and Kalogrides (2005) used basic logistic regression to show that the consideration of race in undergraduate admissions has diminished since the mid-1990's. Finally, in the context of racial preferences in law school admissions, Wightman (1997) used basic logistic regression, with controls for LSAT and undergraduate GPA, to show that ignoring race in law school admissions would result in a drastic decline in minority law students.

Extending the Literature: The Contributions of this Work

The above studies represent important contributions to the literature on admissions preferences, yet they all share a common drawback. The authors were unable to control for *all* of the applicant-level characteristics, other than legacy status, that may have been relevant to the admissions outcome. To overcome this obstacle, this paper utilizes the conditional logistic regression model to reduce the omitted variable bias present in other studies of admissions preferences. Moreover, this paper extends the literature by probing the mechanism through which legacy status functions across four dimensions. First, I explore whether the nature of the familial connection between the applicant and the college plays an important role in the magnitude of the legacy admissions boost. While some institutions only grant admissions preferences to children of alumni, others maintain a more expansive definition of legacy status, including other relatives, like siblings of current students, under the legacy umbrella (Steinberg, 2002). I also test if the selectivity of the outcome college influences the legacy admissions advantage, as the high acceptance rates at modestly selective institutions may suggest that such institutions would need a less aggressive preference policy to cater to their alumni. Third, the academic strength of the applicant could impact the legacy admissions advantage in that admissions staff may be content nudging academically strong applications with a legacy connection from the waitlist pile to the accept pile. For weaker applicants, the legacy connection may be insufficient to catapult the application from the rejection pile to acceptance pile. This mechanism is echoed by one Ivy League admissions officer who claims that, "For students who present strong and competitive applications, the legacy status can serve as a 'tip factor' in the decision to admit the student." (Perret, 2008). Finally, admissions committees look favorably upon applicants who express special interest in their institutions by applying through early admission programs (Avery, Fairbanks, and Zeckhauser, 2003). I test whether applying through early admissions programs augment the legacy admissions advantage.

This paper's scrutiny of the undergraduate college admissions process for legacy applicants extends the literature on this topic by adopting a methodological technique aimed at reducing estimate bias and by revealing that not all legacy applicants are treated equally. To expose the magnitude of bias, I contrast the results from the preferred analytic strategy (conditional logistic regression) to those that would have been obtained from basic logistic regression. Marked differences between results emerge from the application of the different analytic strategies, and these differences reinforce the notion that legacy applicants differ from non-legacy applicants across many dimensions relevant to the college admissions process.

Research Design

Sample

This paper's sample contains of 307,643 domestic[1], first-year applications for undergraduate admission to 30 of the nation's most selective colleges and universities for entry in the fall of 2007. These applications were submitted by 133,236 unique applicants, nearly half of whom submitted applications to two or more of the sampled colleges. Among these multiple-application applicants, the average number of submitted applications was 3.58. This large sample size provides sufficient power to detect an increase in the log odds of admissions of 0.005 at nominal levels of Type I error.[2]

Table 1 summarizes the characteristics of the 30 sampled colleges and universities. Clearly, they do not mirror those of the average American four-year, post

[1] Non-citizens were excluded from this analysis because the admissions process for these applicants varies widely across sampled colleges.
[2] The effect size was determined using computer software produced by Lenth (2006), for a type I error of 0.05 and a power of 0.80.

secondary institution. However, they typify the kinds of institutions that practice selective admissions and garner attention based on assertions about their use of preference. A comparison of these 30 sampled colleges to the top-ranked liberal arts colleges and research universities reveals that the 30 sampled colleges are fairly reflective of the top ranked post-secondary institutions in terms of tuition and fees, percentage of underrepresented minority students and percentage of students receiving Pell grants. In contrast, the 30 sampled colleges and universities are considerably wealthier, less diverse, and more expensive than the typical post-secondary institution. The commonalities between the typical, top-tier post-secondary institution and the 30 sampled colleges suggest that this paper's findings may be generalizeable to a broader set of top tier, selective post secondary institutions.

<< Insert Table 1>>

The sampled colleges not only boast larger endowments and wealthier students than does the typical American postsecondary institution, as noted above, the admissions processes of the sampled colleges are atypically selective. In Table 2, I present the average acceptance rates and SAT scores for legacy applications submitted during the early decision, early action, and regular decision application processes. A testament to the exceptional academic caliber of the legacies and non-legacies in the sample, the average application in the sample boasted math and verbal SAT scores near the 93rd percentile.[3] Furthermore, legacy applications surpassed non-legacy applications in mean SAT critical reading (SAT CR) scores by 10 points and on the SAT mathematics (SAT M) section by an average of 6 points. Despite the relatively modest differences in SAT scores across the

---

[3] According the College Board, a score of 680 on the critical reading section of the SAT is at the 93rd percentile, while a score of 690 on the mathematics section of the SAT places a student at the 93rd percentile. (SOURCE: http://professionals.collegeboard.com/profdownload/sat_percentile_ranks_2008.pdf)

three legacy categories in Table 2, marked differences exist in the estimated probability of admission between these categories. For example, on average, the estimated odds of admission for an applicant whose parent attended one of the sampled colleges as an undergraduate was 3.01 times that of non-legacy at that sampled school. [4]

<< *Insert Table 2*>>

Because conditional logistic regression analysis requires students incorporated into the sample to have submitted applications to multiple colleges, I have removed from the sample students who were admitted through early decision processes at the sampled colleges.[5] In contrast to students who applied through the non-binding admissions procedures of early action and regular decision, the students admitted through early decision programs are required to withdraw all other applications (Ehrenberg, 2000). The result is that the regular decision applications of this subset of students were never evaluated, and it is not possible to predict the outcomes of these applications. While it is not the primary goal of this paper to unearth the admissions advantages of early decision, it is noteworthy that early decision applications have a higher probability of acceptance and lower SAT scores across each of the three legacy categories in *Table 2* than do regular decision and early action applications. The apparent dissimilarities between admissions programs and the exclusion of early decision admittees means that the results of this paper are confined to regular decision and early action applicants.

[4] This value was calculated as the odds of acceptance for an applicant whose parent attended the average sample school as an undergraduate (43.7/56.3) divided by the odds of acceptance for an applicant without any familial connections to that sample school (20.5/79.5).

[5] Estimates in conditional logistic regression are generated from students who submitted multiple applications. As such, computer software discards single application students when fitting conditional logistic regression models. However, I intentionally kept single-application students in the sample because these cases are used in the fitting of basic logistic regression models.

Measures

Outcome:

- *ADMIT*. A dichotomous variable that is coded 1 if the applicant is admitted to the institution to which the application was addressed (0, otherwise).

Question Predictors:

- *ANYLEGACY*. A dichotomous variable that is coded 1 if the applicant has any familial ties to the outcome college (0, otherwise).[6]

- *PRIMARYLEGACY*. A dichotomous variable that is coded 1 if the applicant has a parent who attended the outcome college as an undergraduate (0, otherwise).

- *SECONDARY LEGACY*. A dichotomous variable that is coded 1 if the applicant is not included in the primary legacy category, but has at least one parent who attended the outcome college as a graduate student, or has a grandparent, aunt, uncle, or sibling who has attended the outcome school as either an undergraduate or a graduate student (0, otherwise). Applicants are also placed into this category if the institution does not know whether the parent attended the institution as an undergraduate or graduate student, or does not know how many relatives attended the outcome school.

---

[6] The category of "any familial ties" includes: a connection through which one or more parents, grandparents, aunts, uncles, or siblings attended the outcome school as an undergraduate or graduate student.

Interaction Predictors:

- *SATCAT.* A vector of 14 dichotomous variables that represents the sum of an application's SAT critical reading and SAT math scores.[7] With the exception of *SATCAT1* and *SATCAT14*, each dichotomy spans 50 composite SAT points. For example, *SATCAT1* is coded 1 if a student received a 1600 SAT composite score. *SATCAT2*=1 if a student received a score between 1550 and 1590. *SATCAT3*=1 if a student received a score between 1500 and 1540, and so on. *SATCAT14* =1 if a student received an SAT composite score less than 1000.

- *TIER1* (most selective)*, TIER2, TIER3, TIER4* (least selective). Four dichotomous variables (e.g. *TIER1*) with each coded to 1 if *COLLEGE* is in the specified selectivity tier (e.g. *TIER1*) and zero otherwise. [8]

- *POVERTY.* A dichotomous variable coded 1 if the individual resides in a high-poverty zip-code, defined as one in which the median household income in year 2000 was less than or equal to twice the year 2000 poverty threshold for a family of four ($17,603) (0, otherwise).

[7] I created this variable to obviate the assumption that the odds of admissions are linearly related to a student's composite SAT score. Of the 294,457 regular-decision, early-action applications, and early-decision rejected/deferred applicants, 258,280 (or 87.7%) had valid SAT scores.

[8] To obtain a selectivity metric for each school in the data, I add the normalized rejection rate, the yield rate, and the mean SAT verbal and math scores for applicants to a college for entry in the fall of 2006 and the fall of 2007, weighting each component equally. I choose to examine the SAT scores for the applicants rather than the matriculants because applicant data provide information about the relative academic strength of the applicant pool. While often leading to a "high-scoring" student body, the decision to weigh SAT scores heavily in the admissions process is an institutional priority that does not fully reflect the academic attributes of the applicant pool from which the university can choose. Four tiers emerge from this selectivity analysis, and the selectivity rank order of the schools is aligned closely to that found in other analyses that order schools by selectivity/desirability like Avery, Glickman, Hoxby, and Metrick's revealed preference study (2005) and *U.S. News and World Report Best College Rankings*.

- *URM.* A dichotomous variable coded 1 if the individual is an underrepresented minority applicant, including students who identify as African-American, Native-American, or Hispanic (0, otherwise).
- *URSTATE.* A dichotomous variable coded 1 if the individual resides in an underrepresented state, 0 otherwise. I define underrepresented states as those in which fewer than 1,000 students submitted applications to the sampled colleges. These include AK, AL, AR, DC, DE, HI, IA, ID, KS, KY, LA, ME, MS, MT, ND, NE, NM, NV, OK, PR, RI, SC, SD, UT, VT, WY, WV. I also classify Guam and Puerto Rico as underrepresented states.

Grouping Predictor:

- *COLLEGEID*. A vector of 30 dichotomous variables that distinguishes the thirty sampled colleges. The *i*'th component of the *COLLEGEID* vector is coded 1 if the application was addressed to the *i*'th college (0, otherwise).

Control Predictor:

- *EAAPPLICANT.* A dichotomous variable coded 1 if applicant *j* was an early action applicant at college *i*, regardless of the admissions decision (0, otherwise).

Data Analyses

In order to reduce the omitted variable bias resulting when more conventional analytic techniques are used to address this topic, I utilize *conditional logistic regression* (CLR) *analysis* to quantify the impact of legacy status on the log-odds of admission to the sampled colleges. I apply this technique to eliminate the variability in the outcome attributable to all observed and unobserved applicant characteristics that do *not* differ across applications within applicant, (McFadden, 1973; Chamberlain, 1980; Allison, 2005).[9] In addition, I eliminate variability in the outcome attributable to differences in the observed and unobserved characteristics of colleges that differ neither across applicant nor application (e.g. admissions selectivity) by including in all models the fixed effects for colleges. Removing these two types of variation in the outcome allows me to isolate the impact of legacy status, while controlling for all application-invariant attributes of candidate and college, on the odds of admission.

If the outcome in this analysis had been continuous, rather than binary, I could have achieved the same aims analytically by incorporating as predictors in the statistical models a vector of dummy variables to represent the fixed effects of applicant and a second vector of dummy predictors to represent the fixed effects of colleges. This would have effectively controlled the outcome for variability in all applicant-level characteristics that do not differ across schools and for variability in college-level characteristics that do not differ across applicants. However, as Allison (2006) notes, when the outcome is binary, maximum likelihood estimators of logistic regression slope parameters become biased if the number of parameters in the model increases as the sample size increases. Adding the fixed effects of *applicants* would yield this type of bias because each new applicant would require the inclusion of an additional fixed effect. In contrast, in this paper's analyses, incorporating the fixed effects of colleges does not introduce similar bias because the number of colleges in the sample remains constant at thirty even when additional applicants are added to the sample. For this reason, I use

[9] Examples of such characteristics include SAT scores, high school grades, teacher recommendations, and extracurricular activities. CLR obviates controlling for each of these characteristics.

conditional logistic regression analysis, with *strata* distinguishing the applicant, to control for all observed and unobserved variability in the outcome attributable to applicant-level characteristics that do not differ across colleges. I continue to include a vector of dummy predictors representing the fixed effects of the sampled colleges.

This paper's analyses will not be the first application of CLR in the field of higher education. Recommended by Manski and Wise (1983) to study college choice, Long (2004) used CLR to model the probability that an applicant chose a particular college, given a set of college choices. In Long's (2004) analysis, college options were nested within the applicant, and subject to the constraint that the sum of the outcomes equaled 1 for each applicant. This constraint is logical as an applicant can only choose to attend one college, and must reject his or her other choices. As previously mentioned, my analyses also rely on a clustered design in which applications are nested within applicant; however, unlike Long (2004), the constraint that the sum of outcomes equals 1 for each individual is relaxed (Chamberlain, 1980).

RQ1: What is the admissions advantage granted to legacy applications, on average, controlling for all applicant-level characteristics that do not vary across institutions? To address this question, I fit the following conditional logistic regression model:

$$(1) \qquad \log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \alpha_j + \boldsymbol{\gamma}' COLLEGEID_i + \beta_1 ANYLEGACY_{ij}$$

For applicant j's application to college i: Parameters α_j are applicant-specific intercepts, a consequence of the strata that distinguish individual applicants j under the conditional approach, γ is a vector of parameters representing the fixed effects of college, and the antilog of slope parameter β_1 is an odds-ratio that represents the population admissions advantage attributable to legacy status (Allison, 2005). A statistically significant and positive estimated value of parameter β_1 will indicate that legacy status provides applicants with an admissions advantage. In additional analyses, I also fit *Model #1* using a basic logistic regression model, in which the α_j values are set equal to the single intercept parameter α for all applicants, and I compare the fitted value of parameter β_1 with the estimate of the corresponding parameter obtained under the CLR approach. This comparison reveals whether controlling for all applicant invariant characteristics through the CLR model results in a larger estimated legacy admissions advantage than failing to control for these applicant invariant characteristics in the basic logistic regression model.

RQ2: Is the legacy advantage for children of alumni greater than that of legacy applicants with another type of familial connection to the college? I address the second research question by replacing predictor *ANYLEGACY* in Model #1 with predictors *PRIMARYLEGACY* and *SECONDARYLEGACY* to provide Model #2, as follows:

(2)

$$\log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \alpha_j + \boldsymbol{\gamma}' COLLEGEID_i + \beta_2 PRIMARYLEGACY_{ij} + \beta_3 SECONDARYLEGACY_{ij}$$

The antilogs of regression parameters β_2 and β_3 are odds ratios describing the population admissions advantages attributable to having a parent who attended the outcome school as an undergraduate (β_2) and having a non-parent-undergraduate familial connection to the outcome school (β_3). If the estimate of β_2 exceeds that of β_3, I will conclude that the legacy admissions advantage is larger for applicants with a parent who attended the outcome school as an undergraduate than for other legacy applicants. I also fit Model #2 using basic logistic regression to compare the fitted values of β_2 and β_3 with those obtained under CLR.

RQ3: If there is a legacy advantage, is it larger among applicants who are more academically able? I address this research question by replacing the *ANYLEGACY* main effect in Model #1 with a vector of interaction terms between *ANYLEGACY* and *SATCAT* (Model #3), and by replacing the *PRIMARYLEGACY* and *SECONDARYLEGACY* main effects in Model #2 with a vector of interaction terms between *SATCAT* and both *PRIMARYLEGACY* and *SECONDARYLEGACY* (Model #4)[10], as follows

$$(3) \qquad \log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \alpha_j + \gamma' COLLEGEID_i + \vec{\beta} SATCAT_{ij} * ANYLEGACY_{ij}$$

The antilogs of the parameters in vectors $\vec{\beta}$ and $\vec{\varphi}$ expose the population legacy admissions advantage in each of the 14 *SAT* categories, revealing the legacy admissions advantages across the spectrum of academic abilities, as measured by composite SAT scores. I fit these models using both CLR and basic logistic regression to compare estimates across methods.

RQ4: Does the legacy admissions advantage differ by school selectivity, with more selective schools granting a larger admissions boost? To answer this research question, I extend the analyses performed under RQ1 and RQ2 by replacing the main effect of *ANYLEGACY* in Model #1 with a vector of interaction terms between the four *TIER*

[10] Model #4 is not shown. It's structure parallels that of Model #3.

dummy variables and *ANYLEGACY* (Model #5). Similarly, I replace the main effects of *PRIMARYLEGACY* and *SECONDARYLEGACY* in Model #2 with a vector of interaction terms between the four *TIER* indicator variables and both *PRIMARYLEGACY* and *SECONDARYLEGACY* (Model #6) [11], as follows:

(5)

$$\log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \alpha_j + \boldsymbol{\gamma}' COLLEGEID_i + \beta_1 ANYLEGACY_{ij} * TIER1_{ij} + \beta_2 ANYLEGACY_{ij} * TIER2_{ij} + \beta_3 ANYLEGACY_{ij} * TIER3_{ij} + \beta_4 ANYLEGACY_{ij} * TIER4_{ij}$$

The antilogs of the β regression coefficients in Model #5 and Model #6 reveal the admissions advantages, by legacy type and selectivity tier.

RQ5: *Is the legacy admissions advantage larger when a student applies through a non-binding early action process rather than a regular decision application process?* To answer this question, I add to Model #1 the main effects of *EAAPPLICANT* and interaction terms between *EAAPPLICANT* and *ANYLEGACY* (Model #7) and interaction terms between *EAAPPLICANT* and the *PRIMARYLEGACY* and *SECONDARYLEGACY* dummy variables (Model #8), as follows.

$$\log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \alpha_j + \boldsymbol{\gamma}' COLLEGEID_i + \beta_1 ANYLEGACY_{ij} + \beta_2 EAAPPLICANT_{ij} + \beta_3 EAAPPLICANT_{ij} * ANYLEGACY_{ij} \tag{7}$$

[11] Model #6 is not shown. It's structure parallels that of Model #5.

(8)

$$\log(\frac{p_{ij}}{1-p_{ij}}) = \alpha_j + \boldsymbol{\gamma}'COLLEGE_i + \beta_4 PRIMARYLEGACY_{ij} + \beta_5 SECONDARYLEGACY_{ij} + \beta_6 EAAPPLICANT_{ij} + \beta_7 EAAPPLICANT * PRIMARYLEGACY_{ij} + \beta_8 EAAPPLICANT * SECONDARYLEGACY_{ij}$$

Taking the antilog of the sum of regression parameters β_1 and β_3 in Model #7, I present the admissions advantage attributable to legacy status among early action applications. I then compare the legacy admissions advantage for primary legacies using early action (antilog of sum of β_4 and β_7) with the legacy admissions advantage for secondary legacies using early action (antilog of sum of β_5 and β_8).

Findings

In Table 3, I present the parameter estimates and standard errors associated with the *ANYLEGACY*, *PRIMARYLEGACY* and *SECONDARYLEGACY* predictors in Models #1 and #2 obtained by both conditional logistic regression (CLR) and basic logistic regression (LR). Antilogging the parameter fitted by CLR associated with *ANYLEGACY* status in Model #1 (1.14), I find that the fitted odds of admission for all legacies are 3.13 times the odds of admission for those without legacy status, suggesting that legacy applicants do, indeed, have an admissions advantage over their non-legacy peers. Henceforth, I will refer to such estimated odds ratios as the estimated *legacy admissions advantage*.

As previously mentioned, parameter estimates obtained from fitting models with LR suffer from omitted variable bias. This analytic technique does not account for the clustering of applications within applicants. I present these estimates in Table 3 to illustrate the magnitude of bias that arises from failing to control for all observed and unobserved applicant-level characteristics that do not differ across sampled colleges. For example, if I had been unable to cluster applications within an applicant, I would have estimated that the odds of admission for a legacy would be 2.05 times the odds for a non-legacy. [12]

<<Insert Table 3 here>>

In Table 3, I show that the estimated legacy advantages for primary and secondary legacies are 7.63 and 2.07, respectively. Both of these estimates are larger than the *biased* legacy advantage estimates obtained from fitting Model #2 by basic logistic regression analysis. In this latter case, the corresponding estimates are 3.24 for primary legacies and 1.66 for secondary legacies.

In Table 4, I present the CLR and LR parameter estimates and standard errors associated with the interaction between *SATCAT* and the *ANYLEGACY*, *PRIMARYLEGACY* and *SECONDARYLEGACY* predictors in Models #3 and #4. I also present the CLR and LR estimates and standard errors associated with the interaction between *TIER* and the *ANYLEGACY*, *PRIMARYLEGACY* and *SECONDARYLEGACY* variables in Models #5 and #6. As shown in this table [13], there is surprisingly little variation in the estimated legacy admissions advantage across the academic spectrum.[14] The largest estimated legacy admissions advantage (OR=3.64) is enjoyed by applicants with an SAT (M+CR) score of 1600, while the smallest legacy admissions advantage exists for applicants with SAT scores between 1350-1390 and 1500-1540 (OR=3.31).

---

[12] See column 4, row 1.
[13] See column 2, rows 1-8
[14] Approximately 81% of applications with SAT scores had scores greater than or equal to 1250.

The overall difference in legacy admissions advantage between primary and secondary legacies shown in Table 3 persists across the academic spectrum. Table 4 shows that, among primary legacies, the largest estimated legacy admissions advantages [15] are granted to applicants with SAT scores of 1600 (OR=10.91) and applicants with SAT scores between 1550-1590 (OR=10.93). The smallest estimated legacy admissions advantage among primary legacy applicants exists for those with SAT scores between 1400-1450 (OR=7.28). In each SAT category, the estimated legacy admissions advantage is smaller for secondary legacies than for primary legacies. Furthermore, across the academic ability spectrum, I find less variation in the estimated legacy admissions advantages for secondary legacies[16] than for primary legacies. Among secondary legacies, the largest estimated admissions advantage occurs for applicants with SAT scores between 1250-1290 (OR=2.45), while the smallest estimated admissions advantage occurs for applicants with SAT scores between 1550-1590 (OR=1.83).

In Table 4, I also present the LR legacy parameter estimates from fitting Models #3 and #4 [17] to highlight two phenomena. First, there is considerably more variation in the parameter estimates associated with all three legacy variables when these models are fit with LR rather than CLR. For example, among all legacy applicants, the biased estimates of legacy admissions advantage range from a high of 12.33 among applicants with an SAT score of 1600 to a low of 0.73 among applicants with SAT scores between 1250-1290. The second reason I show the biased legacy admissions advantage estimates in Table 4 is closely related to the first. Unlike the relatively unbiased estimates obtained from CLR, the biased legacy admissions advantage estimates convey a distinct trend in

[15] See column 4, rows 1-8
[16] See column 6, rows 1-8
[17] See columns 8, 10 , 12, rows 1-8

which the estimate values decrease with SAT scores. Consequently, LR over predicts the admissions advantage for the academically stronger students in the sample and under predicts the admissions advantage for the weaker students.

Finally, I present in Table 4 the admissions advantages granted to legacies across each of the selectivity tiers from fitting Model #5 and Model #6 with CLR.[18] In the most selective colleges (Tier 1), the estimated odds of admission are multiplied by a factor of 5.19 as a result of legacy status. This relatively large admissions advantage for legacies in Tier 1 is driven by primary legacies, who enjoy an estimated legacy admissions advantage in Tier 1 of 14.60. The estimated admissions advantage for secondary legacies in tier 1 colleges (OR= 2.09) is similar to the estimated secondary legacy admissions advantage in tier 2 colleges (OR=2.11) and in tier 3 colleges (OR=2.06).

All legacy admissions advantage estimates provided thus far in the paper have included the sample of non-binding early action and regular decision applications. When applicants accepted through non-binding early action are eliminated from the sample, the estimated primary legacy admissions advantage obtained from fitting Model #2 is 6.31.[19] This estimate is smaller than that obtained when early action admits are included in the sample (OR=7.63; Table 3). From this finding, I hypothesized that the legacy admissions advantage differs by the chosen application route. To test this hypothesis, I fit Models #7 and #8. The estimated regression coefficients on the *EAAPPLICANT*ANYLEGACY* (Model #7) and the *EAAPPLICANT*PRIMARYLEGACY* (Model #8) interaction terms are both highly significant ($p<0.001$). In contrast, the estimated regression coefficient on the *EAAPPLICANT*SECONDARYLEGACY* (Model #8) interaction term is not

[18] See columns 2, 4, 6; rows 12-15

[19] Results not shown in tables. Contact the author for tables summarizing the legacy admissions advantages among the sample excluding early-action admits.

significant at the 0.05 level ($p = 0.62$).[20] The implication of these findings is that the added legacy admissions advantage associated with applying early action is granted only to primary legacy applicants.

In Figure 1, I plot the estimated legacy admissions advantage for legacy applications submitted via the early action application route and legacy applications submitted via the regular decision application route. The horizontal axis in Figure 1 defines the application route and the vertical axis conveys the estimated legacy admissions advantage odds ratio. While primary legacy status leads to the odds of admission being multiplied by 5.5 for regular decision applications, the estimated legacy admissions advantage is 15.5 for early action applications. In contrast, the estimated secondary legacy admissions advantage among early action applications is nearly identical for early action applications (OR=1.9) and regular decision applications (OR=2.0).

Discussion

In this paper, I have shown that, among the sampled colleges, the odds of admission for a legacy student are 3.13 times that of a non-legacy student. This estimated legacy admissions advantage is similar in magnitude to that obtained by Espenshade, Chung, and Walling (2004). My analyses show that the nature of the familial connection between the applicant and the alumna/alumnus impacts the legacy admissions advantage, with parental connections being the most advantageous type of legacy connection. This finding is explained by the fact that, at some colleges, non-parental alumni connections are not as influential as parental alumni connections in the admissions process.

[20] Results not shown in tables. Contact the author for output.

On average, the students in this paper's sample are strong academically. However, variation in academic abilities is present in this sample, and previous research suggests that the legacy admissions advantage is not constant across the spectrum of academic abilities (Espenshade, Chung, and Walling, 2004). Due to the pronounced demographic differences between legacies and non-legacies, the SAT is likely an imperfect tool to compare the academic prowess of these two groups of applicants (Steele and Aronson, 1995; Croizet and Claire, 1998). Presumably, admissions committees contextualize SAT scores against an applicant's background in order to avoid penalizing the educationally-disadvantaged applicant. However, among a more demographically and socioeconomically homogenous group (e.g. legacies), using the SAT to draw inferences about academic talent is more justifiable.

Using conditional logistic regression, I find considerably less variation in the estimated legacy admissions advantage across the academic abilities spectrum than did Espenshade, Chung and Walling (2004). The narrow range of estimated legacy admissions advantages strongly contradicts the widely held notion that legacy status serves as a tip factor, only helping academically exceptional applicants on the borderline between acceptance and rejection. Examining the *biased* legacy admissions advantage estimates in this paper, I note that the trend in which these estimates decrease with SAT scores somewhat resembles that found in Espenshade, Chung and Walling (2004) and in Bowen, Kurzweil and Tobin (2005), an unsurprising finding given that both sets of estimates were obtained by basic logistic regression.

Conflicting results from the CLR and LR analyses reinforce that legacies and non-legacies differ on characteristics relevant to admissions committees that cannot easily be controlled for. Furthermore, these differences may also shed light on how admissions staff perceive legacies of varying academic strength. In this paper's analyses, I show that the direction of bias in the legacy admissions advantage is related to the applicant's academic strength, with LR overestimating the legacy admissions advantage for students with the highest SAT scores (>1500) and underestimating the legacy admissions advantage for students with lower SAT scores (<1450). This finding suggests that a student with a 1600 SAT score and a legacy at sample school 1 would be more likely to gain admission to sample school 2 (where he is not a legacy) than would a student with a 1600 SAT score, but no legacy connection at the sampled schools. Conversely, a student with a 1300 SAT score and a legacy at school 1 would be less likely to gain admission to school 2 (where he is not a legacy) than would a student with a 1300 SAT score, but no legacy connection at the sampled schools.

One plausible explanation for this phenomenon is that having a legacy connection at a highly selective college means that the applicant likely will have had ample educational and extracurricular opportunities. Such applicants with modest SAT scores might be easily categorized by admissions staff as BWRK (bright well-rounded kids), who, without a legacy connection, are unable to differentiate themselves from the average applicant. These BWRK applications are almost always rejected (Toor, 2001). Acing the SAT allows an applicant to distinguish himself, and a having a parent (or other close relative) who is familiar with elite college admissions by virtue of having attended an elite college means that the applicant likely has had access to the resources necessary to maximize his application's appeal. An applicant with similarly exceptional SAT scores and parents who attended a non-selective postsecondary institution may lack the capital necessary to synthesize a standout application. The result is that the high-scoring legacy applicant has a higher probability of acceptance at the sampled colleges (even those to which he is not a legacy) than the equally high-scoring non-legacy applicant.

The absence of variability in the estimated legacy admissions advantages across a wide range of SAT scores raises the question of whether the sampled colleges with varying selectivity offer applicants similar legacy admissions advantages. On one hand, high acceptance rates at modestly selective institutions may suggest that such institutions would need a less aggressive preference policy to cater to their alumni. On the other hand, one could argue that pressure to favor alumni relatives would be larger at less selective institutions, as these colleges are often under enormous pressure to beef up endowments (Winston, 1999). With the exception of the most selective tier of sampled colleges, which offer primary legacies an atypically large admissions boost, I find little variation in the legacy admissions advantages across the four selectivity tiers defined in this paper.

Generalizeability of findings to early decision applicants

Avery, Fairbanks, and Zeckhauser (2003) suggest that early admissions processes function differently than do regular decision processes. By applying through a binding early decision (ED) plan, a student sends the outcome college a clear signal that it is the student's top choice. The student is rewarded for her commitment, and Avery and Levin (2009) indicate that an ED application is associated with a 31 to 37 percentage point increase in admissions probability for an applicant with average values on their model covariates. Although early action (EA) is a non-binding early admissions process, the average applicant also appears to be granted an admissions advantage (17 to 20 percentage points) by choosing to apply through EA rather than regular decision (RD) (Avery and Levin, 2009).

As previously mentioned, some sampled schools use the EA form of early admissions instead of ED. Unlike students admitted through ED, students admitted through EA may exercise their rights to apply to and may ultimately choose another school during RD. Keeping the EA admits in the sample would not influence legacy admissions advantage estimates if the legacy advantage were independent of the application route chosen. However, in this paper's analysis, I show that primary legacies who apply through EA are offered an even larger legacy admissions advantage than if they were to have applied through RD. In contrast, virtually no difference exists between the legacy admissions advantage granted to secondary legacies choosing the RD application route and secondary legacies choosing the EA application route. These findings reveal clearly that primary legacies must choose an early application route to realize the full benefit of their admissions advantage.

Because ED involves a level of commitment surpassing EA, I hypothesize that the gap in the legacy admissions advantage between ED and RD applicants would exceed that found between EA legacy applicants and RD legacy applicants. Perhaps this gap arises because a college can rationalize more easily adjusting its admissions criteria to accommodate an applicant if the college is certain that the applicant will matriculate. By choosing early admissions, an applicant expresses a unique interest in the college (Avery, Fairbanks, and Zeckhauser, 2003), thus easing the burden on admissions officers who might have reservations about lowering the admissions bar for legacy applicants.

Threat to the validity of findings

Varying weights attached to admissions criteria across the sampled schools would not pose a problem to the accuracy of this paper's estimates if the average legacy were identical to the average non-legacy on all observable and unobservable characteristics. However, in addition to differing with respect to composite SAT scores, as previously discussed, legacies and non-legacies differ in the demographic characteristics considered important by admissions committees. Applicants with a legacy connection to at least one of the sampled colleges are less likely to reside in a high-poverty zip-code [21], to be identified as an under-represented minority [22], and to reside in an under-represented state.[23] Although these characteristics generally play a secondary role to the most important predictors of college admissions, such as academic and extracurricular talent, they are likely related to the probability of exposure to the rigorous academic curriculum and rich extracurricular activities sought by college admissions committees. These three applicant-level characteristics do not represent an exhaustive list of criteria to which schools assign different weights in the admissions process; however, the literature consistently cites these characteristics as differentially prioritized across colleges. For example, as Kane (1998) highlights, the gap in mean SAT scores between Black and White students differs markedly across the nation's most selective institutions, suggesting

[21] I define a high poverty zip-code as one in which the median household income in year 2000 was less than or equal to twice the year 2000 poverty threshold for a family of four ($17,603). 12.2% of non-legacy applicants live in a high poverty zip-code, compared to 6.7% of legacies.

[22] Under-represented minority applicants include students identified as African-American, Native-American, or Hispanic. 19.1% of non-legacy applicants identify as URM, compared to 10.6% of legacy applicants.

[23] Under-represented states include AK, AL, AR, DC, DE, HI, IA, ID, KS, KY, LA, ME, MS, MT, ND, NE, NM, NV, OK, PR, RI, SC, SD, UT, VT, WY, WV. I also classify Guam and Puerto Rico as underrepresented states. 9.5% of non-legacy applicants live in under-represented states, compared to 7.9% of legacy applicants.

variation in the commitment to diversity across these schools. To enroll students who "live in far away places" (Steinberg, 2002), schools with more national applicant pools may feel less compelled to relax admissions standards for distance applicants than schools with primarily regional appeal. Finally, sampled schools differ in their financial aid policies, implying variability in commitment to socio-economic diversity. Some of the sampled institutions are "need-blind," and make admissions decisions regardless of an applicant's ability to pay (McPherson and Schapiro, 1998), while others maintain "need-aware" admissions policies. Several of the sample schools have even been making concerted efforts to attract low-income students by drastically overhauling financial aid programs, generally replacing loans with grants (Block, 2008).

If the overwhelming majority of legacies at sample school 1 (for example) have a certain feature that is especially appealing to school 1 and not at other sample schools (e.g. coming from a wealthy family), this paper's CLR estimates would overestimate the actual legacy admissions advantage. This bias would stem from attributing these legacy applicants' admissions advantage exclusively to their legacy status, when, in reality, the estimated admissions boost is due to a combination of family wealth and legacy status.

I test the sensitivity of this paper's estimated legacy admissions advantages to varying weights sampled colleges place on the three previously discussed applicant-level characteristics. To perform this sensitivity analysis, I refit all of this paper's conditional logistic regression models with the additional inclusion of three sets of interaction terms between each of the *COLLEGEID* dummy variables and *URSTATE, URM*, and *POVERTY*. After accounting for differences in the weighting of these three important admissions related characteristics at sampled schools, estimates of the legacy admissions advantages changed little. In fact, none of the CLR estimated regression coefficients presented in Tables #3 and #4 differ by more than 8% from the corresponding CLR estimated regression coefficients obtained from fitting these models after the addition of the three demographic*college interaction terms.[24]

Conclusion

A major goal of this analysis is to contribute to the relatively small, but immensely important, existing literature on the impact of legacy status on college admissions. Rather than taking a stance on the issue, I have attempted to describe findings without ascribing labels to them as “good” or “bad.” Although the admissions advantage received by legacy applicants may strike some readers as unacceptably large, I urge readers to consider that donations from alumni are increasingly important to the well-being of this paper’s sampled schools. Among several of these sampled schools, the operating budgets rely more heavily on money drawn from endowments and annual gifts than on tuition revenue. Alumni sustain these endowments through charitable gifts and contribute to annual funds that channel money to financial aid for low-income students. These gifts preserve and grow endowments, ensuring academic excellence for future generations of students. I hope that this point, in conjunction with the previously discussed results, will help individuals to synthesize or refine their own opinions on the justifiability of special admissions preferences for legacy applicants.

[24] Please contact author for output from the CLR models with demographic*college interaction terms.

Tables

Table 1: Comparison of 30 sample schools to selected postsecondary institutions

	Sample of colleges and universities	*U.S. News and World Report* top 50 liberal arts colleges and national universities	Public and private not-for-profit 4-year post-secondary institutions
Endowment value in millions $ (June 30, 2008)	5,850 [30]	1,751 [94]	522 [791]
Average percentage of undergraduate students receiving Pell Grants	12.9% [30]	13.3% [100]	39.3% [1,632]
Average tuition and fees for 2007 academic year	$35,197 [30]	$32,293 [98]	$16,966 [1,914]
Average six year graduation rate	91.5% [30]	86.37% [102]	49.6% [1,794]
Average percentage of undergraduates who are underrepresented minority students	14.9% [30]	12.5% [102]	21.9% [2,019]

Sources: The Chronicle of Higher Education College and University Endowments 2008-2009 database; IPEDS; *U.S. News and World Report Best Colleges 2009*.

Notes: Sample sizes appear in brackets. Ranking ties in *U.S. News and World Report* mean that 102 schools appear in this category, rather than 100. Pell grant percentages for each college are calculated by dividing the total Pell recipients during the 2006-07 school year by the full-time undergraduate enrollment in the fall of the 2006-07 academic year. The average endowment for all public and private not-for-profit 4-year postsecondary institutions is estimated using available data for all 791 post-secondary institutions in *The Chronicle of Higher Education* 2008-2009 database. Students are classified as underrepresented minority students if they are identified as African-American, Hispanic, or Native-American. Although there are 2,318 public and private not-for-profit 4-year institutions listed in the IPEDS database, some institutions do not report data on one or more of the variables used in Table 1.

Table 2: Sample Acceptance rates and mean application SAT scores and standard deviation, by legacy status

	All applications	No Legacy (n=286,478)	Any Legacy (n=21,165)	Primary Legacy (n=6,523)	Secondary Legacy (n=14,642)
All applicants					
Mean SAT Critical Reading	679	679	689	699	684
	(83)	(84)	(76)	(73)	(77)
Mean SAT Math	686	685	691	694	689
	(84)	(84)	(76)	(74)	(77)
Percent Admitted	21.5	20.5	35.2	43.7	31.4
Observations	307,643	286,478	21,165	6,523	14,642
Early decision applicants					
Mean SAT Critical Reading	673	672	682	693	674
	(79)	(80)	(73)	(70)	(74)
Mean SAT Math	680	679	686	692	681
	(81)	(82)	(73)	(70)	(74)
Percent Admitted	41.1	39.0	53.7	56.8	51.5
Observations	22,068	18,963	3,105	1,265	1,840
Regular and early action applicants					
Mean SAT Critical Reading	680	679	690	700	686
	(83)	(84)	(77)	(74)	(77)
Mean SAT Math	686	686	692	695	690
	(84)	(84)	(77)	(75)	(78)
Percent Admitted	20.0	19.2	32.0	40.6	28.5
Observations	285,575	267,515	18,060	5,258	12,802

Source: Admissions data of 30 sampled colleges.

Notes: To be included in the *Any Legacy* category, an applicant would to have had a parent, grandparent, aunt, uncle, or sibling attend the institution as an undergraduate or a graduate student. To be included in the *Primary Legacy* category, an applicant would to have had a parent attend the institution as an undergraduate student. To be included in the *Secondary Legacy* category, an applicant must not be included in the *Primary Legacy* category and must have at least one parent who has attended the target school as a graduate student, or a grandparent, aunt, uncle, or sibling who attended the institution as either an undergraduate or graduate student. Applicants are also placed into this category if the institution does not know whether the parent attended the institution as an undergraduate or graduate student, or how many relatives attended the outcome college.

Table 3: Fitted Conditional Logistic Regression (CLR) parameter estimates, and antilogged fitted parameter estimates (odds ratios), describing the legacy admissions advantage

	Conditional logistic regression				**Basic logistic regression**			
	Coeff. (SE)	Odds Ratio	Coeff. (SE)	Odds Ratio	Coeff. (SE)	Odds Ratio	Coeff. (SE)	Odds Ratio
	Model #1		Model #2		Model #1		Model #2	
Any Legacy	1.14 (0.03)	3.13***			0.72 (0.02)	2.05***		
Primary Legacy			2.03 (0.06)	7.63***			1.18 (0.03)	3.24***
Secondary Legacy			0.73 (0.04)	2.07***			0.51 (0.02)	1.66***
Obser-vations	294,457		294,457		294,457		294,457	
-2LL	54,593.04		54,221.43		280,631.18		280,278.31	

*** p< 0.001 ** p< 0.01 *p<0.05

Source: Admissions data of 30 sampled colleges.

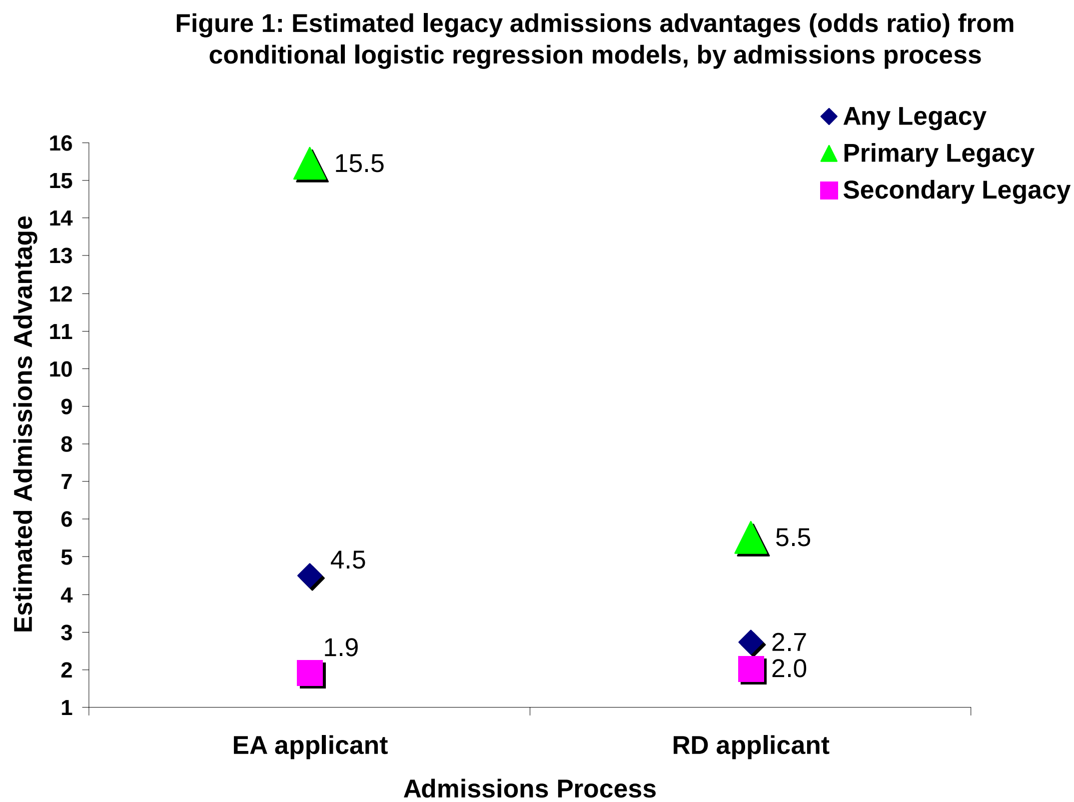
Notes: To be included in the *Any Legacy* category, an applicant would to have had a parent, grandparent, aunt, uncle, or sibling attend the institution as an undergraduate or a graduate student. To be included in the *Primary Legacy* category, an applicant would to have had a parent attend the institution as an undergraduate student. To be included in the *Secondary Legacy* category, an applicant must not be included in the *Primary Legacy* category and must have at least one parent who attended the target school as a graduate student, or a grandparent, aunt, uncle, or sibling who attended the institution as either an undergraduate or graduate student. Applicants are also placed into this category if the institution does not know whether the parent attended the institution as an undergraduate or graduate student, or how many relatives attended the outcome college.

Table 4: Fitted parameter estimates, and antilogged fitted parameter estimates (odds ratios), from selected logistic regression models

	Conditional logistic regression						**Basic logistic regression**					
	Any Legacy		Primary Legacy		Secondary Legacy		Any Legacy		Primary Legacy		Secondary Legacy	
SAT Composite(M+CR) x Legacy Type	Coeff. (SE)	Odds Ratio	Coeff. (SE)	Odds Ratio	Coeff. (SE)	Odds Ratio	Coeff. (SE)	Odds Ratio	Coeff. (SE)	Odds Ratio	Coeff. (SE)	Odds Ratio
	Model #3		Model #4				Model #3		Model #4			
1600	1.29 (0.20)	3.64***	2.39 (0.37)	10.91***	0.84 (0.23)	2.31***	2.51 (0.12)	12.33***	2.88 (0.22)	17.82***	2.35 (0.14)	10.46***
1550-1590	1.20 (0.10)	3.32***	2.39 (0.18)	10.93***	0.60 (0.12)	1.83***	2.16 (0.07)	8.67***	2.72 (0.12)	15.16***	1.88 (0.08)	6.56***
1500-1540	1.20 (0.08)	3.31***	2.12 (0.14)	8.35***	0.78 (0.09)	2.18***	1.69 (0.05)	5.42***	2.23 (0.08)	9.30***	1.43 (0.06)	4.20***
1450-1490	1.23 (0.07)	3.42***	2.21 (0.12)	9.10***	0.74 (0.09)	2.09***	1.35 (0.04)	3.87***	1.88 (0.07)	6.56***	1.10 (0.05)	3.00***
1400-1450	1.21 (0.08)	3.35***	1.98 (0.14)	7.28***	0.82 (0.1)	2.27***	0.83 (0.04)	2.28***	1.24 (0.07)	3.47***	0.63 (0.05)	1.87***
1350-1390	1.20 (0.11)	3.31***	2.11 (0.18)	8.28***	0.81 (0.12)	2.25***	0.29 (0.05)	1.34***	0.65 (0.09)	1.92***	0.13 (0.06)	1.13*
1300-1340	1.24 (0.13)	3.47***	2.04 (0.22)	7.72***	0.89 (0.15)	2.43***	0.04 (0.06)	1.04	0.52 (0.1)	1.68***	-0.18 (0.07)	0.83**
1250-1290	1.21 (0.19)	3.35***	2.04 (0.33)	7.72***	0.90 (0.22)	2.45***	-0.32 (0.07)	0.73***	0.06 (0.13)	1.07	-0.47 (0.09)	0.62***
Observations	258,280		258,280				258,280		258,280			
-2LL	48,471.67		48,104.00				245,751.13		245,457.38			
Tier x Legacy Type	Model #5		Model #6				Model #5		Model #6			
Tier 1	1.65 (0.06)	5.19***	2.68 (0.08)	14.60***	0.73 (0.08)	2.09***	1.12 (0.03)	3.08***	1.77 (0.05)	5.90***	0.55 (0.05)	1.74***
Tier 2	1.08 (0.07)	2.94***	1.53 (0.1)	4.60***	0.75 (0.09)	2.11***	0.69 (0.04)	2.00***	0.85 (0.06)	2.34***	0.56 (0.06)	1.75***
Tier 3	0.76 (0.05)	2.14***	1.03 (0.13)	2.81***	0.72 (0.06)	2.06***	0.52 (0.02)	1.69***	0.74 (0.06)	2.09***	0.48 (0.03)	1.61***
Tier 4	1.13 (0.22)	3.10***	1.35 (0.36)	3.87***	0.99 (0.27)	2.70***	0.78 (0.06)	2.19***	1.05 (0.09)	2.86***	0.54 (0.08)	1.72***
Observations	294,457		294,457				294,457		294,457			
-2LL	54,460.30		54,067.911				280,432.56		280,022.21			

*** p<0.001 ** p<0.01 *p<0.05

Figure



Source: Admissions data of 30 sampled colleges.
Notes: To be included in the *Any Legacy* category, an applicant would to have had a parent, grandparent, aunt, uncle, or sibling attend the institution as an undergraduate or a graduate student. To be included in the *Primary Legacy* category, an applicant would to have had a parent attend the institution as an undergraduate student. To be included in the *Secondary Legacy* category, an applicant must not be included in the *Primary Legacy* category and must have at least one parent who has attended the target school as a graduate student, or a grandparent, aunt, uncle, or sibling who attended the institution as either an undergraduate or graduate student. Applicants are also placed into this category if the institution does not know whether the parent attended the institution as an undergraduate or graduate student, or how many relatives attended the outcome college.

References

Allison, P. (2005). *Fixed Effects Regression Methods for Longitudinal Data Using SAS®*. Cary, NC: SAS Institute Inc.

Allison, P. (2006). Fixed effects regression methods in SAS®. *SUGI conference paper 184-31*. Retrieved October 24, 2008, from http://www2.sas.com/proceedings/sugi31/184-31.pdf.

Avery, C., Fairbanks, A., and Zeckhauser, R. (2003). *The early admissions game: Joining the elite*. Cambridge, Mass.: Harvard University Press.

Avery, C., Glickman, M. Hoxby, C.M., and Metrick, A. (2005, December). A revealed preference ranking of U.S. colleges and universities. *NBER Working Paper 10803*. Retrieved August 4, 2008, from http://www.nber.org/papers/w10803.pdf.

Avery, C. and Levin, J.D. (2009, April). Early admissions at selective colleges. *NBER Working Paper 14844*. Retrieved April 8, 2009, from http://www.nber.org/papers/w14844.pdf.

Berfield, S. and Tergesen, A. (2007, October 22). I can get your kid into an Ivy. *BusinessWeek*. Retrieved August 4, 2008, from http://www.businessweek.com/magazine/content/07_43/b4055063.htm

Block, S. (2008, February 5). Private colleges try to stay affordable. *USA Today.* p. B1.

Bound. J., Hershbein, B., and Long, B.T. Playing the admissions game: Student reactions to increasing college competition. *Journal of Economic Perspectives, 23*(4), 119-146.

Bowen, W. G., Kurzweil, M. A. and Tobin, E. M (2005). *Equity and excellence in American higher education*. Charlottesville, VA: University of Virginia Press.

Bunzel, J. H. (1996). Race and college admissions. *Public Interest,* (122), 49-58.

Chamberlain, G. (1980). Analysis of covariance with qualitative data. *Review of Economic Studies, 47*(146), 225-238.

Croizet, J.C., and Claire, T. (1998). Extending the concept of stereotype threat to social class: The intellectual underperformance of students from low socioeconomic backgrounds. *Personality and Social Psychology Bulletin, 24*(6), 588-594.

Dillon, S. (2007, April 4). A great year for Ivy League schools, but not so good for applicants to them. *The New York Times*. Retrieved August 4, 2008, from http://www.nytimes.com/2007/04/04/education/04colleges.html?_r=1andemandex=117andoref=slogin.

Duffy, E. A., and Goldberg, I. (1998). *Crafting a class: College admissions and financial aid, 1955-1994*. Princeton, N.J.: Princeton University Press.

Ehrenberg, R. G. (2000). *Tuition rising: Why college costs so much*. Cambridge, Mass.: Harvard University Press.

Eide, E., Brewer, D. J., and Ehrenberg, R. G. (1998). Does it pay to attend an elite private college? Evidence on the effects of undergraduate college quality on graduate school attendance. *Economics of Education Review, 17*(4), 371-376.

Espenshade, T. J., Chung, C.Y. and Walling, J. L. (2004). Admission preferences for minority students, athletes, and legacies at elite universities. *Social Science Quarterly, 85*(5), 1422-1446.

Golden, D. (January 15, 2003). Admissions preferences given to alumni children draws fire. *The Wall Street Journal Online*. Retrieved September 2, 2009 from http://online.wsj.com/public/resources/documents/golden3.htm

Golden, D. (2006). *The price of admission: How America's ruling class buys its way into elite colleges--and who gets left outside the gates*. New York: Crown Publishers.

Golden, D. (March 2, 2007). To boost donor numbers, colleges adopt new tricks. *The Wall Street Jornal*, p. A1.

Gratz v. Bollinger. 2003. 123 S. Ct. 2411.

Greene, H. and Greene, M. (2009). *Making it into a top college 2nd Ed.* New York: HarperCollins.

Grodsky, E., and Kalogrides, D. (2008). The declining significance of race in college admissions decisions. *American Journal of Education, 115*(1), 1-33.

Grutter v. Bollinger. 2003. 123 S. Ct. 2325.

Hernández, Michele A. (1997). *A is for Admission: The insider's guide to getting into the Ivy League and other top colleges*. New York: Warner Books, Inc.

Howell, C., and Turner, S. E. (2004). Legacies in black and white: The racial composition of the legacy pool. *Research in Higher Education, 45*(4), 325-351.

Hoxby, C.M. and Terry, B. (1999). Explaining rising income and wage inequality among the college educated. *NBER Working Paper 6873*. Retrieved August 4, 2008, from http://www.nber.org/papers/w6873.pdf.

Jaschik, S. (2008, August 4). Legacy admits: more money, lower scores. *Inside Higher*

Ed. Retrieved September 12, 2008, from http://www.insidehighered.com/news/2008/08/04/legacy

Kane, T.J. (1998). "Racial and Ethnic Preference in College Admissions", in Christopher Jencks and Meredith Phillips (eds.), *The Black-White Test Score Gap.* Washington: Brookings Institution.

Kane, T. J. (2003). The long road to race-blindness. *Science, 302*, 571-573.

Karabel, J. (2005). *The chosen : The hidden history of admission and exclusion at Harvard, Yale, and Princeton*. Boston: Houghton Mifflin Co.

Klitgaard, R. E. (1985). *Choosing elites*. New York: Basic Books.

Leef, G. (2008, February 19). Legacy admissions – affirmative action for the rich?: Should junior get in just because daddy's a graduate? The John William Pope Center for Higher Education Policy. Retrieved August 4, 2008, from http://www.popecenter.org/issues/article.html?id=1966.

Lenth, R. V. (2006). Java Applets for Power and Sample Size [Computer software]. Retrieved December 29, 2008 from http://www.stat.uiowa.edu/~rlenth/Power.

Lexington (2004, January 8). The curse of nepotism. *The Economist*. Retrieved September 2, 2009 from http://www.economist.com/world/unitedstates/displaystory.cfm?story_id=2333345

Long, B. T. (2004) How have college decisions changed over time? An application of the conditional logistic choice model. *Journal of Econometrics*, *121*(1-2), 271-296.

Martin, N.D. and Spenner, K.I. (2009) Capital conversion and accumulation: a social portrait of legacies at an elite university. *Research in Higher Education*, *50*, 623-648.

Manski, C. and Wise, D. (1983) *College choice in America*. Cambridge, MA: Harvard University Press.

Massey, D. S., and Mooney, M. (2007). The effects of America's three affirmative action programs on academic performance. *Social Problems, 54*(1), 99-117.

McFadden, D. (1973). "Conditional logit analysis of qualitative choice behavior." In: Zarembka, P. (Ed.), *Frontiers in Econometrics*: New York, Academic Press.

McPherson, M. S., and Schapiro, M. O. (1998). *The student aid game : Meeting need and* *rewarding talent in American higher education*. Princeton, N.J.: Princeton University Press.

Megalli, M. (1995). So your dad went to Harvard: Now what about the lower board scores of White legacies? *The Journal of Blacks in Higher Education,* (7), 71-73.

Orfield, G. (1999, December 10). Affirmative action works—but judges and policy makers need to hear that verdict. *The Chronicle of Higher Education*.

Perret, A. (2008, April 17). College is twice as likely to admit legacy applicants. *The Dartmouth.com News*, Retrieved March 7, 2009, from http://thedartmouth.com/2008/04/17/news/admissions/

Rimer, S. (2007, May 27). Elite colleges open new door to low-income youths. *The New York Times*. Retrieved September 2, 2009 from http://www.nytimes.com/2007/05/27/education/27grad.html?pagewanted=2

Shulman, J. L., and Bowen, W. G. (2001). *The game of life: College sports and educational values*. Princeton, N.J. : Princeton University Press.

Spivack, M.S. (2009, April 12). The admissions gap. *The Washington Post*. p. W10.

Steinberg, J. (2002). *The gatekeepers : Inside the admissions process of a premier college*. New York: Viking Press.

Steele, C. M., and Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology, 69*(5), 797-811.

Thomas, D. and Shepard, T. (2003, March 14). Legacy admissions are defensible, because the process can't be fair. *The Chronicle of Higher Education: The Chronicle Review*. Retrieved August 4, 2008, from http://chronicle.com/free/v49/i27/27b01501.htm.

Toor, R. (2001). *Admissions confidential : An insider's account of the elite college selection process* (1st ed.). New York: St. Martin's Press.

University of California v. Bakke. 1978. 98 S. Ct. 2733.

U.S. News and World Report Best Colleges 2009 (2008).

Weiss, N. (2005, April 13). Legacy Admissions: Does a family history at the Farm matter? *The Stanford Daily*. Retrieved September 2, 2009, from http://www.stanforddaily.com/cgi-bin/?p=1017471

Wightman, L.F. (1997). The threat to diversity in legal education: An empirical analysis in the consequences of abandoning race as a factor in law school admissions decisions. *New York University Law Review 72*(1). 1-53.

Winston, Gordon G. (1999). Subsidies, Hierarchy and Peers: The Awkward Economics of Higher Education. *Journal of Economic Perspectives 13*(1). 13-36.