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

Michael Hurwitz

The Harvard Graduate School of Education, Gutman Library, 6 Appian Way, Cambridge, MA 02138, United States

ARTICLE INFO

Article history:
Received 28 January 2010
Received in revised form 4 November 2010
Accepted 7 December 2010

JEL classification:
C51
I23

Keywords:
Legacy
College admissions
Selective colleges
Conditional logistic regression
Admissions preferences

ABSTRACT

In this paper, I examine the impact of legacy status on admissions decisions at 30 highly selective colleges and universities. Unlike other quantitative studies addressing this topic, I use conditional logistic regression with fixed effects for colleges to draw conclusions about the impact of legacy status on admissions odds. By doing so, I eliminate most sources of outcome bias by controlling for applicant characteristics that are constant across colleges and college characteristics that are constant across applicants. I estimate that the odds of admission are multiplied by a factor 3.13 due to legacy status. My results also 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, to a lesser extent, the selectivity of the outcome college and the applicant’s academic strength.

© 2010 Elsevier Ltd. All rights reserved.

1. 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 (Espenshade, Chung, & Walling, 2004; Golden, 2006; Shulman & Bowen, 2001). 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 thoroughly 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, even after accounting for important covariates.

In this paper, one goal is to account for bias in estimates of the legacy admissions advantage that has plagued studies utilizing 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 my 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. Using this approach, I eliminate bias in the estimate of legacy status impact due to applicant characteristics that are invariant across the multiple institutions, including those which are challenging or impossible to control for using basic logistic regression (e.g. strength of teacher recommendations, student disposition, and essay quality). In addition, I also control for the relative selectivity of the sampled colleges by including a vector of college-level indicator variables. By removing these two sources of important variability in the admissions outcome, I isolate the impact of legacy status on admissions decisions.
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 to illustrate the magnitude of bias generated from this more common analytical approach.

2. Background and context

2.1. The admissions question: whom to admit?

During the past decade, the heightened competition to win undergraduate admission to America’s elite postsecondary 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 decreasing importance of distance between the applicant and college (Hoxby, 2009), the increasing numbers of applications submitted per applicant (Spivack, 2009), the expanding international applicant pools, and increases in financial aid (Heller, 2006), the continually decreasing acceptance rates have transformed the college-admissions landscape. 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, & Long, 2009).

Given the fairly stable supply of spots at the nation’s most selective institutions (Hoxby, 2009), the increasing demand for these spots has made the question of whom to admit even more difficult. Though the applicant might perceive an equitable admissions process to be one in which only concrete academic characteristics, such as SAT scores and high school GPA are considered, colleges generally realize that such a one-dimensional approach is not actually equitable, acknowledging the diversity of backgrounds and opportunities among applicants (Orfield, 1999). Nevertheless, colleges, like students, place a premium on academic characteristics, recognizing the importance of these characteristics vis-à-vis their influence on student peer effects as well as relative positions in college ranking magazines like US News and World Report (Monks & Ehrenberg, 1999; Winston & Zimmerman, 2004).

Highly selective colleges face the challenge of maximizing the academic profile of their student bodies, with the understanding that sacrificing some academic talent now will enable the college to preserve or improve its selectivity in the future. In other words, from the college’s perspective, an exclusive focus on academics in the admissions process is not sustainable. Tradeoffs necessary to maintain future excellence manifest in the allocation of financial aid at the most selective colleges. For example, focusing the entire financial aid budget on students with the highest SAT scores might temporarily enhance selectivity, but such a move would negatively impact student body diversity, ultimately compromising the college’s desirability in the future. With this in mind, many of the most-selective colleges focus only on a student’s ability to pay, rather than his academic prowess relative to other admitted students when awarding financial aid (McPherson & Schapiro, 2006). Other tradeoffs might include relaxing admissions standards for early decision applicants to decrease acceptance rates and increase yield rates, and consequently to appear more selective (Jensen & Wu, 2010). Or they might involve admitting academically lackluster star athletes to maintain the successful sports teams that encourage alumni giving (Holmes, 2009; Meer & Rosen, 2009). Relatives of alumni (legacies) offer enthusiasm and familiarity to colleges, and the special treatment awarded to them in the admissions process helps to preserve generational ties that also are intended to motivate financial generosity.

2.2. The legacy question

The 20th century realization that maintaining an academically and “socially” excellent institution required money is at least partially responsible for the special admissions treatment enjoyed by legacy applicants. Appeasing wealthy alumni meant accepting their relatives and sustaining the family traditions that motivated financial donations (Karabel, 2005). 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, even after accounting for “measurable” differences between legacies and non-legacies, such as SAT scores, gender, ethnicity and U.S. citizenship (Bowen, Kurzweil, & Tobin, 2005; Espenshade et al., 2004; Shulman & Bowen, 2001). The focus on legacy students, however, is not limited to the admissions realm. Researchers have also examined the college outcomes of legacies, compared to non-legacies. These studies identify some evidence of legacy underperformance in college courses, as well the tendency of legacies to choose college majors that are often perceived to be less rigorous (Martin & Spenner, 2009; Massey & Mooney, 2007).

Given the financial returns associated with attending an elite postsecondary institution (Hoxby & Long, 1999), as well as the increased probability of attending an elite graduate school (Eide, Brewer, & Ehrenberg, 1998), from the perspective of society, allocation of undergraduate spots to applicants who might be deemed undeserving would compromise equity. Furthermore, giving preferences to legacy candidates may also negatively impact the institution’s goal to enhance campus diversity. In light of the demographic shift among college-going students (Perfetto, 2010), such diversity may be necessary to create an atmosphere friendly to underrepresented students, a group whom colleges will need to court increasingly in the interest of maintaining or improving selectivity. Moreover,
student-body diversity is noted as an essential element in establishing the enriching educational environment expected by students and professors (AAUP, 2000). Because legacies at America’s most selective postsecondary institutions are disproportionately White (Howell & Turner, 2004), awarding preference to children or close relatives of alumni could pose an impediment to racial diversity (Megalli, 1995).

2.3. 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 postsecondary 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 (Hernández, 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.

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.

A major complication in the empirical research is that legacies are different from non-legacies on multiple criteria important to admissions committees, such as SAT scores, underrepresented minority status, and wealth. The contrast in measurable criteria between legacies and non-legacies suggests 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 will likely be biased. Basic logistic regression in the studies of admissions preferences has been used to chip away at some bias by controlling for measurable criteria, such as standardized test scores (Bowen et al., 2005; Espenshade et al., 2004; Grosky & Kalogrides, 2008; Shulman & Bowen, 2001; Wightman, 1997). However, numerous applicant-level characteristics (e.g. quality of teacher recommendations, essay quality and content) that cannot be controlled for continue to introduce bias in these studies’ estimates.

2.4. 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 failure to control for all applicant-level characteristics. This paper shows how conditional logistic regression can achieve the goal of reducing omitted variable bias, thus overcoming the obstacles of previous studies. 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). Second, 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. 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. Finally, admissions committees look favorably upon applicants who express special interest in their institutions by applying through early admission programs (Avery, Fairbanks, & Zeckhauser, 2003). I test whether applying through early admissions programs augments the legacy admissions advantage.

For each of this paper’s research questions, 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.
3. Research design

3.1. Sample

This paper’s sample contains of 307,643 domestic,1 first-year applications for undergraduate admission in the fall of 2007 to 30 highly selective, private colleges and universities existing within a consortium that has served as the data source for recent studies of admissions and financial aid processes (e.g. Hill & Winston, 2010). Of these applications, 294,457 are incorporated into this paper’s analyses and this group includes non-binding early action, regular decision, and deferred early decision applications. Among my sample’s 133,236 unique applicants, nearly 47% (61,962) submitted applications to two or more of the sampled colleges, and among these multiple-application applicants, the average number of submitted applications was approximately 3.6. In addition to information on admissions outcome, school to which the application was submitted and legacy status, each admissions record contains information on the applicant’s gender, hometown, race, athlete status and SAT scores.

Table 1 summarizes the characteristics of the 12 private liberal arts colleges and 18 private research universities in this paper’s sample. The commonalities between the typical, top-tier postsecondary institution and the 30 sampled colleges suggest that this paper’s findings may be applicable only to the top tier, selective postsecondary institutions.

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. As shown in Table 1, students matriculating at the sampled colleges boast markedly higher SAT scores than matriculants at the typical American postsecondary institution. The academically exceptional student bodies at the sampled colleges are selected from talented and large applicant pools. In Table 2, I present the average acceptance rates and SAT scores for legacy applications submitted during the early decision (offered by 24 colleges), early action (offered by 6 colleges), and regular decision application (offered by 30 colleges) 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.2 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 three legacy categories in Table 2, marked differences exist in the estimated probability of admission between these categories. For example, across all applications to the 30 sampled colleges, the estimated odds of admission for an application indicating that a parent attended the sample college as an undergraduate was 3.01 times that of a non-legacy application.3 Finally, early decision applicants are more likely to be legacies than are regular decision/early action applicants, with 14.1% of early decision applicants classified as legacies compared to 6.3% of regular decision/early action applicants.

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.4 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 in subsequent analyses means that the results of this paper are confined to regular decision and early action applicants.

3.2. Data analyses

In order to reduce the omitted variable bias resulting when more conventional analytic techniques are used to

---

1 Non-citizens were excluded from this analysis because the admissions process for these applicants varies widely across sampled colleges.
2 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).
3 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).
4 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 (who were not accepted by a sampled college through early decision) in the sample because these cases are used in the fitting of basic logistic regression models, and a researcher unaware of CLR or unable to cluster applications within applicants most certainly would not have discarded these single-application regular decision/early action students. In this paper, I focus on the differences between the basic logistic regression estimates and the CLR estimates, continually noting the magnitude of bias in the estimates obtained from basic logistic regression. The astute reader might question whether characteristic differences between the single-application and multiple-application students are responsible for the evident bias in estimates obtained from basic logistic regression. However, basic logistic regression parameter estimates associated with legacy variables remain virtually unchanged when single-application cases are discarded, discounting this scenario. For example, the biased legacy advantages from basic logistic regression when single-application students are excluded is 2.32, essentially identical to the value of 2.31 when single-application students are included. The biased legacy advantages for primary and secondary legacies change from 3.89 to 3.83 and from 1.82 to 1.86, respectively, when single-application students are removed.
Table 1
Comparison of 30 sample schools to selected postsecondary institutions.

<table>
<thead>
<tr>
<th>Sample of colleges and universities</th>
<th>U.S. News and World Report top 50 liberal arts colleges and national universities</th>
<th>Public and private not-for-profit 4-year postsecondary institutions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average endowment value in millions</td>
<td>5850</td>
<td>1751</td>
</tr>
<tr>
<td>($ June 30, 2008)</td>
<td>[30]</td>
<td>[94]</td>
</tr>
<tr>
<td>Average percentage of undergraduate</td>
<td>12.9%</td>
<td>13.3%</td>
</tr>
<tr>
<td>students receiving Pell Grants</td>
<td>[30]</td>
<td>[100]</td>
</tr>
<tr>
<td>Average tuition and fees for 2007</td>
<td>$35,197</td>
<td>$32,293</td>
</tr>
<tr>
<td>academic year</td>
<td>[30]</td>
<td>[98]</td>
</tr>
<tr>
<td>Average six year graduation rate</td>
<td>91.5%</td>
<td>86.37%</td>
</tr>
<tr>
<td>[30]</td>
<td>[102]</td>
<td>[1794]</td>
</tr>
<tr>
<td>Average undergraduate enrollment</td>
<td>4936</td>
<td>7352</td>
</tr>
<tr>
<td>[30]</td>
<td>[102]</td>
<td>[2019]</td>
</tr>
<tr>
<td>Average percentage of undergraduates who are underrepresented minority students</td>
<td>14.9%</td>
<td>12.5%</td>
</tr>
<tr>
<td>[30]</td>
<td>[102]</td>
<td>[2019]</td>
</tr>
<tr>
<td>Average acceptance ratea</td>
<td>24.3%</td>
<td>35.0%</td>
</tr>
<tr>
<td>All applicants</td>
<td>[30]</td>
<td>[102]</td>
</tr>
<tr>
<td>Average</td>
<td>1405</td>
<td>1337</td>
</tr>
<tr>
<td>Median SAT (M + CR) of matriculants</td>
<td>[30]</td>
<td>[97]</td>
</tr>
</tbody>
</table>

Notes: The data are not weighted by school enrollment as the goal of this table is to provide the reader with descriptive statistics on the typical institution in each category. Therefore, the data in Table 1 represent averages across institutions, with each institution weighted equally. 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 2006–2007 school year by the full-time undergraduate enrollment in the fall of the 2006–2007 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 postsecondary 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. The IPEDS enrollment variable used in this analysis is “Grand_total_EF2007A_ALL_student.” SAT and acceptance rate data are for the fall of 2007 and are calculated using IPEDS data. One of the sampled schools is missing SAT data on IPEDS, and I replaced the missing data for this school with an estimate calculated from this paper’s data file. IPEDS provides the inter-quartile SAT ranges for postsecondary institution matriculants, and I estimated the median SAT for an institution as the midpoint point between the 25th and 75th percentiles.

a The acceptance rate of 24.3% (for example) represents an average of the acceptance rates across the 30 sampled schools. This differs from the percentage of the total applications at the 30 sampled schools that were offered admission (21.5%; see Table 2).

Table 2
Sample acceptance rates and mean application SAT scores and standard deviation, by legacy status.

<table>
<thead>
<tr>
<th>All applications</th>
<th>No legacy</th>
<th>Any legacy</th>
<th>Primary legacy</th>
<th>Secondary legacy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean SAT critical reading</td>
<td>679 (83)</td>
<td>679 (84)</td>
<td>689 (76)</td>
<td>699 (73)</td>
</tr>
<tr>
<td>Mean SAT math</td>
<td>686 (84)</td>
<td>685 (84)</td>
<td>691 (76)</td>
<td>694 (74)</td>
</tr>
<tr>
<td>Percent admitted</td>
<td>21.5</td>
<td>20.5</td>
<td>35.2</td>
<td>43.7</td>
</tr>
<tr>
<td>Observations</td>
<td>307,643</td>
<td>286,478</td>
<td>21,165</td>
<td>6523</td>
</tr>
<tr>
<td>Mean SAT critical reading</td>
<td>673 (79)</td>
<td>672 (80)</td>
<td>682 (73)</td>
<td>693 (70)</td>
</tr>
<tr>
<td>Mean SAT math</td>
<td>680 (81)</td>
<td>679 (82)</td>
<td>686 (73)</td>
<td>692 (70)</td>
</tr>
<tr>
<td>Percent admitted</td>
<td>41.1</td>
<td>39.0</td>
<td>53.7</td>
<td>56.8</td>
</tr>
<tr>
<td>Observations</td>
<td>22,068</td>
<td>18,963</td>
<td>3105</td>
<td>1265</td>
</tr>
<tr>
<td>Mean SAT critical reading</td>
<td>680 (83)</td>
<td>679 (84)</td>
<td>690 (77)</td>
<td>700 (74)</td>
</tr>
<tr>
<td>Mean SAT math</td>
<td>686 (84)</td>
<td>686 (84)</td>
<td>692 (77)</td>
<td>695 (75)</td>
</tr>
<tr>
<td>Percent admitted</td>
<td>20.0</td>
<td>19.2</td>
<td>32.0</td>
<td>40.6</td>
</tr>
<tr>
<td>Observations</td>
<td>285,575</td>
<td>267,515</td>
<td>18,060</td>
<td>5258</td>
</tr>
</tbody>
</table>

Source: Admissions data of 30 sampled colleges.
Notes: Standard deviations are included in parentheses. To be included in the Any Legacy category, an applicant would have 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 have 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. Six of the sampled schools use early action and twenty-four of the sampled schools use early decision.

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 (Allison, 2005; Chamberlain, 1980; McFadden, 1973). In addition, I eliminate variability in the outcome

5 Examples of such characteristics include SAT scores, high school grades, teacher recommendations, and extracurricular activities. CLR
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 sources 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. Like in other fixed effects models, standard errors associated with conditional logistic regression estimates are amplified slightly compared to the basic logistic regression analogs. However, this tradeoff in the context of the forthcoming analyses is not overly problematic due to this study’s large sample size.

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 produce this type of bias because each new applicant would require the inclusion of an additional dummy variable. 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 conditional logistic regression analysis, with strata \( \alpha_j \) 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. Unlike in other studies within the higher education literature, I do not incorporate college-level characteristics (e.g., enrollment size, cost) into the conditional logistic regression models (Avery & Hoxby, 2003; Long, 2004; Niu & Tienda, 2008) because it is not this paper’s goal to explore how these characteristics influence admission. Instead, I include a vector of dummy predictors \( \text{COLLEGEID} \) 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 odds that an applicant chose a particular college, given a set of college choices. In Long’s (2004) analysis (for example), 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).

4. Data analysis

4.1. Research question 1

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:

\[
\log \left( \frac{p_{ij}}{1 - p_{ij}} \right) = \alpha_j + \gamma \text{COLLEGEID}_i + \beta_1 \text{ANYLEGACY}_{ij} \quad (1)
\]

For applicant \( j \)'s application to college \( i \): parameters \( \alpha_j \) are applicant-specific intercepts, a consequence of the strata that distinguish individual applicants \( j \) under the conditional approach, \( \gamma \) is a vector of parameters representing the fixed effects of college \( \text{COLLEGEID} \), and the antilog of slope parameter \( \beta_1 \) is an odds ratio that represents the population admissions advantage attributable to legacy status \( \text{ANYLEGACY} \) (Allison, 2005). A statistically significant and positive estimated value of parameter \( \beta_1 \) will indicate that legacy status provides applicants with an admissions advantage. In additional analyses, I also fit Model #1 (and all subsequent models) using basic logistic regression with additional covariates representing race, gender, SAT (M+CR) composite category \( \text{SATCAT} \), and athlete status \( \text{ATHLETE} \). In the basic logistic regression model, \( \alpha_j \) values are set equal to the single intercept parameter \( \alpha \) for all applicants, and I compare the fitted value of parameter \( \beta_1 \) 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 controlling only for the convenient and traditionally included applicant invariant characteristics (e.g., Espenshade et al., 2004) in the basic logistic regression model.

4.2. Research question 2

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 the \( \text{ANYLEGACY} \) term in Model #1 with indicator variable

\[6\text{ RACE includes indicator variables for BLACK, HISPANIC, WHITE, ASIAN, and OTHER (including unknown race as the reference category). SATCAT includes a vector of 15 dichotomous variables that represents the sum of an application’s SAT critical reading and SAT math scores. With the exception of SATCAT1, SATCAT14 and SATCAT15, each dichotomy spans 50 composite SAT points. For example, SATCAT1 is coded 1 if a student received a 1600 SAT composite score. SATCAT2 = 1 if a student received a score between 1550 and 1590. SATCAT2 = 1 if a student received a score between 1500 and 1540, and so on. SATCAT14 = 1 if a student received an SAT composite score less than 1000. SATCAT15 = 1 if the application’s SAT score is missing.}\]
predictors PRIMARYLEGACY and SECONDARYLEGACY (Model #2). The antilogs of the regression parameters associated with PRIMARYLEGACY and SECONDARYLEGACY are odds ratios describing the population admissions advantages attributable to having a parent who attended the outcome school as an undergraduate and having a non-parent-undergraduate familial connection to the outcome school, respectively.

4.3. Research question 3

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).

The antilogs of the parameters associated with the ANYLEGACY × SATCAT interaction terms expose the population legacy admissions advantage in each of the 15 SAT categories, revealing the legacy admissions advantages across the spectrum of academic abilities. Similarly, I replace the PRIMARYLEGACY and SECONDARYLEGACY main effects in Model #2 with a vector of interaction terms between SATCAT and both PRIMARYLEGACY and SECONDARYLEGACY.

4.4. Research question 4

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 indicator variables and ANYLEGACY (Model #4).

4.5. Research question 5

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, a dummy variable indicating whether or not student j applied to applied to college i via early action, and interaction terms between EAAPPLICANT and ANYLEGACY (Model #5) and interaction terms between EAAPPLICANT and the PRIMARYLEGACY and SECONDARYLEGACY indicator variables (Model #6). The antilog of the sum of regression parameters associated with EAAPPLICANT and EAAPPLICANT × ANYLEGACY represents the admissions advantage attributable to legacy status among early action applicants.

5. Results

5.1. Research question 1: the overall legacy admissions advantage

In Table 3, I present the antilogged parameter estimates (odds ratios) and associated 95% confidence intervals associated with the ANYLEGACY predictor as described in Model #1 obtained by both conditional logistic regression (CLR) and basic logistic regression (LR). Across all applicants to the 30 sampled colleges, I find that the fitted odds of admission for legacies are 3.13 times the odds of admission for applicants without legacy status, confirming 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 legacy admissions advantage. As noted, the estimate of 3.13 represents an average across all applicants to the 30 sampled colleges. Of course, the number of applicants differs across schools, so one might expect the legacy admissions advantage at the typical (median) school to differ somewhat from 3.13. Replacing the ANYLEGACY variable in Model #1 with a vector of interaction terms between ANYLEGACY and each of the 30 colleges, I find that the median institution awards a legacy admissions advantage of 2.93. The largest legacy admissions advantage among sampled colleges is 15.69, and the smallest is 0.74 (not different from 1.00 at the 0.05 level of significance).

Although odds ratios are useful in the context of this paper’s comparisons, the lingua franca among researchers and practitioners of college admissions is acceptance rates, or the probability of admission multiplied by a factor of 100. Unfortunately, the relationship between odds and probabilities is not linear, so a large odds ratio admis-
Table 3 also illustrates the difference in legacy admissions advantage between primary and secondary legacies (specification 2). The estimated odds admissions advantage granted to primary legacies (7.63) is more than three times the estimated advantage for secondary legacies (2.07), indicating a significant legacy admissions advantage in favor of primary legacies.

5.2. Research question 2: the primary and secondary legacy admissions advantages

The inclusion of “deferred” early decision applications explains the difference between 19.0 and the estimate of 19.2 in Table 2, which was estimated by excluding early decision deferrals. The odds of admissions for the typical non-legacy application is 0.235, which implies a base acceptance rate of 43.7%. Including deferred early decision applications, the acceptance rate increases to 47.7%, explaining the difference in the estimates between Tables 2 and 3. For example, if I had been unable to cluster applications within an applicant, I would have estimated a legacy admissions advantage of only 2.31. This biased estimate is smaller than the unbiased estimate of 3.13. Different estimates obtained from these two analytic approaches reveal the existence of unobservable (in this paper’s dataset) characteristics positively correlated with legacy status, which prove disadvantageous in the college admissions process. Such a list of unobservable characteristics is lengthy and might include anything from differences in family wealth between legacies and non-legacies to differences in the quality of college essays between the two groups. The limited number of variables in this paper’s dataset prohibits an exhaustive search for these unobservable characteristics.
times the odds admissions advantage granted to secondary legacies (2.07). 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. Moreover, basic logistic regression underestimates the legacy admissions advantages for both secondary legacies and primary legacies, indicating the presence of unobservable characteristics among both groups of legacies that are negatively related to admissions odds (specification 4).

5.3. Research question 3: the relationship between student academic strength and legacy admissions advantages

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 et al., 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 (Croizet & Claire, 1998; Steele & Aronson, 1995). 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.

In Table 4, I present the CLR and LR odds ratios and 95% confidence intervals associated with the interaction between SATCAT and the ANYLEGACY, PRIMARYLEGACY and SECONDARYLEGACY predictors (specifications 1–4). As shown in this table, there is a relatively weak positive relationship between legacy status and academic prowess. The largest estimated legacy admissions advantage (OR = 3.74) is enjoyed by applicants with an SAT (M + CR) score of 1600, while the smallest legacy admissions advantage (OR = 3.74) is enjoyed by applicants with an SAT score of 1250–1290 (OR = 2.87).

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 are granted to applicants with SAT scores of 1250–1290 (OR = 11.72), which, incidentally, translates into a 49.1 percentage point increase in the admission rate for this group. The smallest estimated legacy admissions advantage among primary legacy applicants exists for those with SAT scores between 1250 and 1290 (OR = 5.71). 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 than for primary legacies. Among secondary legacies, the largest estimated admissions advantage occurs for applicants with 1600 SAT scores (OR = 2.40), while the smallest estimated admissions advantage occurs for applicants with SAT scores between 1550 and 1590 (OR = 1.92). Finally, the positive relationship between academic prowess and legacy advantage found for primary legacies is absent among secondary legacies, suggesting that the presence of this relationship for the overall group of legacies is driven exclusively by primary legacies.

Not surprisingly, the formula for admissions differ across the nation’s most selective colleges, and directors of admission are loath to reveal the inner-workings of their unique processes. Occasionally, however, glimpses of insight are provided by admissions directors that shed light on how legacy status, for example, is considered in college admissions. Quotes from directors of admission at selective colleges suggest that legacy status serves as a tip factor, only helping academically strong applicants on the borderline between acceptance and rejection (e.g. Perret, 2008). If this were the case, one might expect the entire legacy advantage to be concentrated among the academically strongest students. Though a relationship exists between academic prowess and the magnitude of the legacy admissions advantage, particularly among primary legacies, the presence of odds admissions advantages greater than 2 in the 1250–1290 SAT score range suggests that even relatively weak applicants may enjoy special preference.

In Table 4, I also present the LR legacy odds ratios to emphasize that the unobservable characteristics found among legacies, which bias LR estimates of the legacy admissions advantage are present across a wide range of academic aptitude. Each CLR estimate in Table 4 (specifications 1–4) is larger than the corresponding LR estimate. Perhaps this finding implies that these unobservable, bias-inducing characteristic differences between legacies and non-legacies are non-academic in nature.

5.4. Research question 4: the relationship between college selectivity and legacy admissions advantages

The presence of some variability in the estimated legacy admissions advantages across a wide range of SAT scores in conjunction with differences in the mean academic profiles of the applicants across the sampled colleges raises the question of whether sampled college selectivity is related to the 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). The results from conditional logistic regression demonstrate that the largest legacy admissions advantages occur at the most selective
### Table 4
Fitted parameter estimates (as odds ratios) and predicted percentage point increases in admissions probability describing the legacy admissions advantage, with 95% confidence intervals in parentheses.

<table>
<thead>
<tr>
<th>Specification</th>
<th>Conditional logistic regression</th>
<th>Basic logistic regression</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>[1]</td>
<td>[2]</td>
</tr>
<tr>
<td></td>
<td>Any legacy</td>
<td>Primary legacy</td>
</tr>
<tr>
<td>SAT: 1600</td>
<td>3.74***</td>
<td>29.8</td>
</tr>
<tr>
<td>SAT: 1550–1590</td>
<td>3.52***</td>
<td>30.2</td>
</tr>
<tr>
<td>SAT: 1500–1540</td>
<td>3.33***</td>
<td>29.1</td>
</tr>
<tr>
<td>SAT: 1450–1490</td>
<td>3.49***</td>
<td>28.6</td>
</tr>
<tr>
<td>SAT: 1400–1450</td>
<td>3.38***</td>
<td>25.7</td>
</tr>
<tr>
<td>SAT: 1350–1390</td>
<td>3.06***</td>
<td>21.1</td>
</tr>
<tr>
<td>SAT: 1300–1340</td>
<td>3.27***</td>
<td>21.4</td>
</tr>
<tr>
<td>SAT: 1250–1290</td>
<td>2.87***</td>
<td>17.5</td>
</tr>
<tr>
<td>Obs.</td>
<td>294,457</td>
<td></td>
</tr>
<tr>
<td>AIC</td>
<td>54,632.5</td>
<td></td>
</tr>
</tbody>
</table>

### Table 4 (continued)

<table>
<thead>
<tr>
<th>Specification</th>
<th>Conditional logistic regression</th>
<th>Basic logistic regression</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>[5]</td>
<td>[6]</td>
</tr>
<tr>
<td></td>
<td>Any legacy</td>
<td>Primary legacy</td>
</tr>
<tr>
<td>B. Legacy type interacted with college selectivity tier</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Selectivity Tier 1</td>
<td>5.19***</td>
<td>26.3</td>
</tr>
<tr>
<td>Selectivity Tier 2</td>
<td>2.94***</td>
<td>18.2</td>
</tr>
<tr>
<td>Selectivity Tier 3</td>
<td>2.16***</td>
<td>16.4</td>
</tr>
<tr>
<td>Selectivity Tier 4</td>
<td>3.41***</td>
<td>28.3</td>
</tr>
<tr>
<td>Obs.</td>
<td>294,457</td>
<td></td>
</tr>
<tr>
<td>AIC</td>
<td>54,526.7</td>
<td></td>
</tr>
</tbody>
</table>

The basic logistic regression includes covariates for SAT (M + CR) score category, gender, race, and athlete status. 95% confidence intervals appear in italics in parentheses. Approximately 12.3%, or 36,177 applications, are missing SAT scores, though these cases are included in all analyses through the inclusion of a separate SAT category for missing values. The base probabilities, as acceptance rates, of non-legacies used to calculate the percentage point increase in admissions probability are as follows: Tier 1 = 9.9%; Tier 2 = 13.8%; Tier 3 = 23.6%; Tier 4 = 46.3%; SAT 1600 = 47.0%; SAT 1550–1590 = 39.2%; SAT 1500–1540 = 32.3%; SAT 1450–1490 = 24.5%; SAT 1400–1440 = 19.9%; SAT 1350–1390 = 16.4%; SAT 1300–1340 = 14.8%; SAT 1250–1290 = 13.6%.

*** p < 0.001
I present in Table 4 the admissions advantages granted to legacies across each of the selectivity tiers.\footnote{See specifications 5 and 6.} 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 of 14.61. The estimated admissions advantage for secondary legacies awarded by Tier 1 colleges (OR = 2.09) is similar to the estimated secondary legacy admissions advantage awarded by Tier 2 colleges (OR = 2.11) and by Tier 3 colleges (OR = 2.07).

5.5 Research question 5: early admissions programs and the legacy advantage

Avery et al. (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 31–37 percentage point increase in admissions probability for an applicant with average values on model covariates. Although early action (EA) is a non-binding early admissions process, the average applicant also appears to be granted an admissions advantage (17–20 percentage points) by choosing to apply through EA rather than regular decision (RD) (Avery & Levin, 2009).

As previously mentioned, some sampled colleges 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, 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.\footnote{Results not shown in tables. Contact the author for tables summarizing the legacy admissions advantages among the sample excluding early-action admits.} 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 #5 and #6, as described in Section 4.5, in which EAAPPLICANT is a dummy variable indicating that the application was submitted through early action. The estimated regression coefficients on the EAAPPLICANT × ANYLEGACY and the EAAPPLICANT × PRIMARYLEGACY interaction terms are both highly significant (p < 0.001). In contrast, the estimated regression coefficient on the EAAPPLICANT × SECONDARYLEGACY interaction term is not significant at the 0.05 level (p = 0.62).\footnote{Results not shown in tables. Contact the author for output.} The implication of these findings is that the added legacy admissions advantage associated with applying early action is granted only to primary legacy applicants.

---

Fig. 1. Estimated legacy admissions advantages (odds ratio) from conditional logistic regression models, by admissions process. Source: Admissions data of 30 sampled colleges. Notes: To be included in the Any Legacy category, an applicant would have had a parent, grandparent, aunt, uncle, or sibling attend the institution as an undergraduate or graduate student. To be included in the Primary Legacy category, an applicant would have 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.
In Fig. 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 Fig. 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 is nearly identical between early action applications (OR = 1.9) and regular decision applications (OR = 2.0). 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 et al., 2003), thus easing the burden on admissions officers who might have reservations about lowering the admissions bar for legacy applicants.

6. Concluding statement

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 colleges. Alumni have helped to grow these endowments for generations (Karabel, 2005) through charitable gifts and contributions to annual funds that channel money to financial aid packages that fund the choice of 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.

Acknowledgements

This 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 past and present colleagues – Ann Birk, Kristine Dillon, Steve Minicucci, and Tony Broh – 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 these analyses. Finally, I would like to acknowledge the assistance of two anonymous reviewers, who also provided exceptional feedback.

Appendix A. Supplementary data


References


