

Maes, Daniel P.
2018 Mathematics & Statistics Thesis

Title: Assessing Critical Mass at UC-Berkeley: Creating Predictive Models
for Affirmative Action Policies in Undergraduate Admissions in the
United States

Advisor: Chad Topaz

Advisor is Co-author: None of the above

Second Advisor:

Released: release now

Contains Copyrighted Material: No

Assessing Critical Mass at UC-Berkeley
Creating Predictive Models for Affirmative Action Policies in
Undergraduate Admissions in the United States

by
Daniel Paul Maes

Chad Topaz, Advisor

A thesis submitted in partial fulfillment
of the requirements for the
Degree of Bachelor of Arts with Honors
in Mathematics



WILLIAMS COLLEGE
Williamstown, Massachusetts

Defended May 7, 2018

ACKNOWLEDGEMENTS

This project could not have been accomplished without the support of many wonderful people and organizations along the way.

Thank you to the Andrew W. Mellon Foundation and the Office of Special Academic Programs for the financial and academic support I received through the Mellon Mays Undergraduate Fellowship (MMUF). This support allowed me to pursue this amazing project without the need to worry about how to provide for myself and others over the past roughly two years. I would especially like to thank Molly Magavern and Bob Blay for doing an awesome job leading and putting together this fellowship at Williams. Thanks also go to Ashley Drake for the help with editing; if this paper is at all coherent, it is due greatly as a result of her recommendations. Additionally, my MMUF cohort has been instrumental in my thesis process by providing me with much needed emotional support throughout, for which I am greatly thankful for.

I would also like to thank the Department of Mathematics and Statistics for supporting my journey through the world of applied math and stats. Special thanks go to Chad Topaz for being the absolute best thesis advisor I could ever have asked for, and to Julie Blackwood for being my unofficial math advisor since sophomore year as well as my second reader. Thank you also to Professors Bernhard Klingenberg and Daniel Turek for being my MMUF advisors while I was still trying to narrow down my research question and finding data to facilitate exploring such questions.

Lastly, I would like to thank all of my wonderful family members and friends who have seen this project blossom from a litigation literature review into the applied math thesis it is today. Whether it was listening to a practice presentation, commiserating with me, or putting up with me while stressed out in a Lee's booth, I am grateful to many of my friends who have been especially supportive during my thesis writing process. To list all of them would be far too long, but rest assured that any of you who took out the time to read this thesis would most likely have been on that list.

ABSTRACT

Affirmative action refers to hiring and recruiting practices designed to combat discrimination against members of certain demographic groups. These policies are used in settings ranging from federal contractors to local employment to public education, and can focus on race, ethnicity, or gender. Due to the broad applicability and direct social impact of affirmative action, substantial effort has gone into monitoring the necessity and effectiveness of these policies. This effort has manifested itself in frequent legal and state action over the past forty years. In this thesis, we focus on the application of race-based affirmative action policies in public undergraduate college admissions in the United States, specifically through a case study of admissions to the University of California, Berkeley. We make three primary contributions. First, we introduce a quantitative framework through which to interpret a key concept used in contemporary affirmative action litigation: "critical mass." Historically, policies have been assessed using retroactive data studies. However, using predictive models would reduce the monetary and temporal costliness of such studies. Thus, second, we construct a predictive model of college admissions demographics using Markov Chains. Third, we bring together the two previous contributions, using our quantified version of the critical mass criterion as a benchmark for assessing the outcomes of the predictive model. The type of mathematical model we construct can be modified for use at other universities or for affirmative action applied to other axes, including gender.

TABLE OF CONTENTS

Acknowledgements	iii
Abstract	v
Table of Contents	vii
Chapter I: Introduction	1
Chapter II: Affirmative Action Litigation and Terms	3
2.1 Supreme Court Cases	4
2.2 State Action: California	9
2.3 Reshaping the Interpretation of Critical Mass	10
Chapter III: Modeling Diversity in Academia	13
3.1 Modeling Structures	13
3.2 Model Data	16
Chapter IV: Markov Chain Model Construction	19
4.1 Markov Chain Theory	19
4.2 States for the Model	21
4.3 Initial Conditions for Individual Markov Chains	23
4.4 Transition Probabilities	24
4.5 The Transition Matrix and Finding \mathbf{N} , \mathbf{t} , and \mathbf{B}	30
Chapter V: Model Results	33
5.1 Markov Chain Model Predictions	33
5.2 Quantifying Critical Mass for the Future	36
5.3 Assessment Criterion and Results	39
Chapter VI: Conclusion	41
Bibliography	43
Appendices	45

Chapter 1

INTRODUCTION

Affirmative action has been at the forefront of the American social consciousness in recent decades. This is because such policies can influence the hiring practices for organizations ranging from local plants to federal contractors to admissions in public education. Proponents see affirmative action as a societal necessity to pave the way for equality, while opponents view it as reverse discrimination. Since these policies can affect representation with regards to gender, race, ethnicity, and other axes, they are widely influential. With such a wide possible impact on society, it is important to understand to what extent these policies should remain in place, a debate which forms the basis of much affirmative action litigation.

One factor in this debate is how to assess the efficacy of affirmative action policies. Historically, they have been assessed through retroactive data studies looking at how minority representation changes after the implementation of a new policy. Such assessments are costly in regards to time, money, and energy. All these costs are shared between the organization where the policy takes place, those hired to assess the policy, and litigators (if the policy results in court action). This thesis creates an easier, objective, and effective way of assessing the performance of a specific policy.

Because the applications of affirmative action are numerous, we narrow our focus to one specific form: the use of racial/ethnic affirmative action policies for undergraduate college admissions in the United States (U.S.). We focus on *public* undergraduate institutions in the U.S. because data regarding admissions to these types of institutions are more readily available. We will investigate the admissions and enrollment dynamics at the University of California, Berkeley (UC-Berkeley). This choice has to do with many facets, the most important being the high level of competition in applications to UC-Berkeley and the regular involvement of the state of California in affirmative action litigation and state action. The competitiveness in admissions guarantees that there is a large pool of qualified applicants, and we can therefore assume that there are enough qualified applicants from a given group to be admitted to the university based on merit alone. As we will see in the following chapter, California's history with affirmative action has lasted since at least 1978 manifesting action in the supreme court and on the state level.

The goals of this thesis are to create a predictive model and a new assessment method that

can be used together to assess the impacts of affirmative action on undergraduate admissions at UC-Berkeley. The assessment method will rely on quantifying critical mass as it may ideally be interpreted for the representation at UC-Berkeley. We create a simple model so that it can be generalizable to other contexts, however, it will still be complex enough to capture the dynamics we see in real-life. Additionally, such a model could potentially be modified to assess other types of minority representation such as gender. However, before building the model, we must understand the history and role that affirmative action has played in the U.S. and California so that we can make more effective assessments and sensible recommendations for future affirmative action policies at UC-Berkeley.

*Chapter 2***AFFIRMATIVE ACTION LITIGATION AND TERMS**

The debate over the acceptance and effectiveness of affirmative action policies within the United States has been ongoing since its inception in 1961. Several questions that arose in the initial conception of the policy are still relevant for understanding its place in today's society. One of these questions is how to interpret the effectiveness of a specific institution's policy. Another is how to understand when affirmative action has accomplished its goal—or even whether such an assessment is even possible.

Affirmative action aims to help a particular minority group achieve a "critical mass" of representation within an organization. Critical mass is a very recent term in the long history of affirmative action, and a better understanding of it will make implementing and assessing affirmative policies easier. This is because if critical mass can prove to be more strongly and narrowly defined in its purpose, then that makes evaluating the effectiveness of specific policies less complicated. Due to its undefined nature, and the variety of possible interpretations and implementations from institution to institution, recent uses of critical mass do not offer an easily replicable method of evaluating the efficacy of a given policy. A better understanding of racial/ethnic critical mass would allow for the application of a more nuanced interpretation across different affirmative action contexts. Instead of racial/ethnic critical mass in undergraduate admissions, one could modify the definition to work for graduate admissions or employment, and apply it in that context. This strategy could also allow for adaptation to other forms of affirmative action, such as gender; for example, a similar model to racial/ethnic undergraduate critical mass could be adapted for gender critical mass in undergraduate admissions, graduate admissions, or employment.

Transitioning to a clearer notion of critical mass would also provide for more productive discourse in the future, simultaneously allowing litigation to better handle cases focused on affirmative action policies at undergraduate institutions. After the Regents of the University of California v. Bakke case in 1978, affirmative action moved away from being seen as an aid for groups that have been historically discriminated against, instead being more understood as a way of establishing equal opportunity between groups. While both interpretations are similar in their results, the first interpretation focuses the rationale for such policies on repatriation for past discrimination, while the second focuses on restructuring as a result of present inequalities in society. The newer interpretation of affirmative action leaves out

the emphasis on repaying past injustices and moves toward making the present more equal. That new interpretation forgets—or at least does not inherently suggest—that the current state of inequality is due to those past events that warrant repatriation. This is why many people will now say that affirmative action is used to "create diversity" or to "get more minorities involved," as opposed to saying it is to make up for the discrimination faced by such groups in the past. Yes, interpretations can change over time, but in this instance, it appears that the original version has just been masked, not actually changed. Therefore, by making critical mass more accessible, the affirmative action debate can refocus on the real reason for why it was created in the first place.

While affirmative action can apply to a variety of entities, we will focus on public undergraduate universities as a specific case study in analyzing the efficacy of updating the term critical mass. We will gain a better understanding of the history of affirmative action litigation in order to assess the contemporary workability of such a process. The following sections will outline landmark Supreme Court cases which transformed the use of affirmative action at public undergraduate universities in the U.S., specific action taken by the state of California regarding affirmative action, and how we can change the term critical mass to help during policy assessment.

2.1 Supreme Court Cases

While the Executive Branch (President) and Legislative Branch (Congress) can create new laws pertaining to affirmative action, the Judicial branch holds the power of deciding its constitutionality in practice. More specifically, the Supreme Court of Appeals has ruled on many influential cases that determined the constitutionality of the admissions policies at several institutions. After the first major affirmative action case (*Regents of the University of California v. Bakke*), states would look at subsequent appellate decisions in determining their own affirmative action policies. This section highlights the most relevant appellate cases to the concepts of critical mass in affirmative action and its implementation on a state level, especially as it pertains to undergraduate admissions. This understanding will enable us to make sensible and legal recommendations during policy assessment.

Regents of the University of California v. Bakke (1978)

The first landmark case for affirmative action in public undergraduate institutions was *Regents of the University of California v. Bakke* in 1978. Alan Bakke, a 35-year-old, Caucasian male, applied twice to the University of California at Davis' (UC-Davis) Medical School and was denied admission both times. At the time, UC-Davis reserved 16 out of 100 of their admitted spots for "qualified" minorities. Bakke's scores were higher than

any of the minority students in both years of applying, thus Bakke claimed he was denied admissions unfairly, and solely based on his race [4].

The court was set to decide whether UC-Davis' Medical School violated the Equal Protection Clause of the Fourteenth Amendment¹, as well as the Civil Right Act of 1964, in using this specific affirmative action policy, therein resulting in the wrongful rejection of Bakke's application. The court was divided on many issues pertaining to the case, thereby ruling in the plurality opinion. They decided 8-1 in favor of Bakke that Title VI of the Civil Right Act of 1964 gave Bakke a cause of action in going to court. Additionally, they ruled 5-4 in favor of Regents that Title VI of the Civil Rights Act of 1964² does not prohibit their race-based admissions program. Next, the court voted 5-4 in favor of Regents that the Equal Protection Clause permits race to be a factor in admissions—among other factors. Lastly, there was a 5-4 decision in favor of Bakke that the Equal Protection Clause prohibits the university's specific race-based admissions policy, and that Bakke should be admitted [4].

This case established that the use of any racial quotas (i.e. saving spots for minorities) is a direct violation of the Equal Protection Clause. Later in 1995, President Bill Clinton wrote a White House memorandum that precisely denounced the use of quotas in affirmative action policies. This memorandum addressed the evaluation of affirmative action programs and urged those programs which took race, ethnicity, or gender into consideration to eliminate or reform their policy if it “creates a quota, creates preference for unqualified individuals, creates reverse discrimination, or continues even after its equal opportunity purposes have been achieved” [6]. This action echoed the Bakke decision, while displaying a backing from the President. This banning of quotas means that during policy assessment, we will not be able to recommend any policies that directly save spots for one group over another.

Gratz v. Bollinger (2003)

Another landmark case was Gratz v. Bollinger in 2003. The University of Michigan's (U-M) Undergraduate Admissions Office considered many important factors (such as grades, test scores, class difficulty, relationship with alumni, geography, and leadership qualities) when deliberating their admissions decisions for incoming freshman. They also considered race in their admissions decisions and would admit students that are “underrepresented minorities” at a higher rate than those that are not. In 1998, the office began using a point based system which added 20 points³ to a student's profile if they were considered an underrepresented minority. The following year (1999), U-M created an Admissions Review Committee to

¹For more information on this Clause, check the following reference: [20]

²For more information on this Title, check the following reference [23]

³This was on a 150-point scale, where 100 points guaranteed admission

add an extra level of consideration to the process [2].

In 1995, Jennifer Gratz and Patrick Hamacher—both Caucasian residents of Michigan—applied for admission to U-M. Each student was denied admission from the university and told that, although they were qualified, that their applications were not competitive enough. In 1997, they filed a class action lawsuit against the U-M et al.⁴ on the grounds that they were discriminated against because of their race, directly violating the Equal Protection Clause of the Fourteenth Amendment and Title VI of the Civil Rights Act of 1964. The lower district court held that a racially and ethnically diverse campus could be beneficial, but that the admissions policies between 1995 and 1998 were problematic in their application because they amounted to “holding seats” for certain minorities. As a result, the Supreme Court judged for the petitioners in relation to the admissions policies for 1995-1998 and for the respondents regarding the point-based policy that began in 1999 [2].

The Supreme Court wanted to know whether U-M’s use of racial preference in their undergraduate admissions policy violated the Equal Protection Clause of the Fourteenth Amendment and Title VI of the Civil Rights Act of 1964. The court decided 6-3 in favor of Gratz that the Office of Undergraduate Admissions policy was not narrow enough to meet “strict scrutiny” [2]. Strict scrutiny is a standard in judicial review which is used to determine the constitutionality of a policy. To pass strict scrutiny, a policy must serve a compelling government interest and be narrowly tailored enough to serve that interest [7].

This policy was deemed unconstitutional because it did not look at a student’s profile individually, instead, assuming that every applicant from a specific “underrepresented minority” was from a similar background. This specific policy at Michigan failed because it did not focus enough on the individual level, but rather, on a general stereotype of those from a specific background [2]. This decision helped establish the use of a strict scrutiny standard for evaluating affirmative action policies in public undergraduate institutions.

Grutter v. Bollinger (2003)

Another landmark decided in 2003 was Grutter v. Bollinger. In this case, Barbara Grutter, a Caucasian resident of Michigan, applied to the University of Michigan’s (U-M) Law School in 1997. She had a 3.8 undergraduate GPA and an LSAT score of 161 but was denied admission. The Law School does state that they use race as a factor in their admissions decisions because it serves “a compelling interest in achieving diversity among its student body.” The district court tried to enjoin the university’s goal of the basis that

⁴The lawsuit also included Michigan’s College of Literature, Science, and the Arts, Lee Bollinger (the respondent), and James Duderstadt (the President of the University at the time)

achieving diversity was not a compelling interest, but the Supreme Court held that Regents of the University of California v. Bakke set a precedent for establishing diversity as a compelling argument and is sufficient under strict scrutiny to justify the use of race in admissions decisions. The Supreme Court also rejected the district court's finding that the Law School's "critical mass" was the functional equivalent of a quota. The court wanted to know whether U-M Law School's use of racial preferences in student admissions violated the Equal Protection Clause of the Fourteenth Amendment or Title VI of the Civil Rights Act of 1964. They ruled in a 5-4 decision that the Equal Protection Clause does not prohibit the Law School's narrowly tailored use of race in the admissions process [3].

They upheld the admission policy used at the Law School because they thought that its highly individualized application review meant that no acceptance was based on one sole factor (i.e. race). In addition, this type of policy ensures that all other sorts of factors that help with achieving diversity are accounted for as well. This case upheld that achieving diversity was a compelling reason (first stated in the Bakke case) and that the use of a "critical mass" was not equal to a quota system. Additionally, this case showed that a highly individualized use of race in admissions was constitutional, while a wider use of race would not be considered the same [3]. Thus, we can take away from this case and Gratz v. Bollinger that policy recommendations must also be able to meet a level of strict scrutiny so that they are constitutional in their application.

Fisher v. University of Texas (2016) – Fisher II

The most recent landmark case for affirmative action in public undergraduate institutions was Fisher v. University of Texas in 2016. This case was originally tried in 2013, but was remanded because the Supreme Court did not believe that the lower courts sufficiently applied strict scrutiny to the policy. Thus, the case was tried again in 2016.

In this case, Abigail (Abby) Fisher, a Caucasian female, applied for admission to the University of Texas in 2008 but was denied. She did not qualify for Texas' Top Ten Percent Plan which guarantees admissions to the top ten percent of students in every public high school class. The remaining in-state admissions decisions are chosen based on a holistic process, which uses race as a factor in its decision. Fisher sued the university, arguing that the use of race as an admissions consideration was a violation of the Equal Protection Clause [1]. The court wanted to know whether the University of Texas' use of race as a consideration in the admissions process violated the Equal Protection Clause. The court ruled 4-3 in favor of the University of Texas that the race-conscious admissions program in use at the time of the suit was legal under that clause [1].

The university was allowed to use race as a factor in its admission policy so long as they would study how diversity was being achieved or maintained in their academic environment. In a similar manner to *Grutter v. Bollinger*, their use was upheld because it established a more individualized assessment of an application, using race “as a factor of a factor of a factor” instead of a blanket improvement. The policy also met strict scrutiny because the precedent was set on diversity being a compelling interest for universities to achieve. The court determined that the university had concrete enough goals, along with reasoned explanations for how to achieve them and thoughtful consideration as to why previous plans may not have helped. They believed that “the University of Texas’ plan was also narrowly tailored to serve this compelling interest because there are no other available and workable alternatives for doing so” [1].

The rhetoric and implementation for affirmative action in undergraduate admissions is almost as ineffective and problematic as when it was first initiated. As the years have progressed, and numerous court cases have come and gone, the constitutional implementation of affirmative action has changed but has not become any more useful. The first implementation was a quota, but quotas were deemed unconstitutional in the *Bakke* case. Therefore, the terms used in affirmative action litigation needed to become broader in order for its use to persist. Eventually, we have arrived at the term used most in recent court cases, the notion of a ‘critical mass.’ Critical mass has been the least specific term used thus far. The *Fisher II* case stated that “critical mass is neither some absolute number of African-American or Hispanic students nor the percentage of African-Americans or Hispanics in the general population of the State. . . and [the] term remains undefined” [19]. This lack of specificity is the biggest problem with contemporary implementations of critical mass, and is not an isolated incident. Rather, any undergraduate institution can dictate their own interpretation of critical mass, and since diversity is already considered a compelling government interest, almost any definition of critical mass will pass the standard of strict scrutiny.

This leads to the second largest problem, which is that the term has changed from a blanket statement for every institution to follow, and moved toward becoming more nuanced and defined by each institution. Changing the workability for critical mass from a federal level to an institutional one has made appellate cases more difficult to judge, as well as making the implementation and effectiveness of policies harder to assess. For this reason, it is necessary to impose more specifications on the future use of critical mass in order to increase the workability of affirmative action in undergraduate institutions. This will simultaneously

make the effectiveness of such policies easier to analyze. Prior to attempting to create a working definition for critical mass, we must narrow our scope further, to California, in order to produce a better understanding of the context in which it will be used.

2.2 State Action: California

Outside of federal policies and decisions, states and institutions have a say in how and to what extent their public systems use affirmative action policies. Because of this case study's focus on the University of California system—and more specifically modeling students at the UC-Berkeley campus—this section highlights the changes in affirmative action policies that California has had in relation to public undergraduate admissions. The choice of California was very deliberate, as California has proven to be highly involved with the development of affirmative action policies, not only on a local or state level, but also influencing policies on a regional and/or federal level (as seen in the previous section).

First, in 1995, the Regents of the University of California (UC) passed SP-1, a resolution which eliminated the use of race, ethnicity, and gender in admissions decisions for institutions in the UC system. In immediately prior years to the passing of this resolution, the enrollment of underrepresented minorities (including African Americans and Hispanics/Latinos) dropped. However, since 1998, enrollment for those underrepresented minorities has been steadily increasing [16].

In the following year, California passed Proposition 209 which prohibited state and local agencies from giving preferential treatment to individuals or groups on the basis of their race, sex, ethnicity, or national origin in public education, employment, or contracting. Under this proposition, schools in the UC system are not allowed to consider the use of affirmative action programs in their admissions policies [16].

In 1999, California implemented a state-wide initiative that guaranteed admissions to the University of California for top students in public schools. For those that graduated in the top four percent of a California public high school's graduating class, or in the top four percent statewide, they were guaranteed admittance to one school in the UC system. These students admitted under the "Four Percent Plan" do not get to choose the institution to which they are admitted. Later, in 2012, this plan expanded to nine percent but held the same regulations as delineated in 1999 [16].

The Regents voted in 2001 to rescind SP-1. Along with rescinding that resolution, the Regents decided on several actions that UC system schools should follow. One action was that all students would be treated equally in the admissions process regardless of their race, sex, color, ethnicity, or national origin. Another action was that each campus should seek

to enroll a student body which demonstrates a high academic achievement or talent level, as well as encompassing the broad diversity of backgrounds represented in the state of California [26]. Although SP-1 was rescinded by the Regents, affirmative action policies were still not allowed at any UC institution under Proposition 209.

Since affirmative action policies are banned in California, policy recommendations will need to focus on aspects which do not directly influence the acceptance rate of some racial/ethnic groups over others. This means, for example, that admissions offices can try to recruit more applicants from a specific demographic group⁵, but cannot change their admissions policy to admit students from one group more than another to achieve the same outcome. Lastly, we will consider the lack of an affirmative action policy a policy in itself. Thus, for the rest of the paper, any reference to the current affirmative action policy in place at UC-Berkeley really means the lack of one.

2.3 Reshaping the Interpretation of Critical Mass

With all of this history in mind, we will now shift to discussing ways to change the current use of critical mass that will make affirmative action policy implementation and assessment easier. Briefly, I will readdress the reasons why this change is beneficial for the contemporary use of affirmative action.

First, increased specification of the terms used in affirmative action discourse will help to deter from the hyper-nuanced definitions that persist today. This will make affirmative actions a bit more streamlined for general use. The goal is to specify critical mass enough so that it can be applied to a wide range of contexts, but loosely enough to allow for some institutional input and specification. This leads to the next benefit from a better outlined critical mass, the ability to apply such a policy to contexts outside the scope of this paper. While I will be focusing solely on racial affirmative action in undergraduate admissions, if defined correctly, the same notion of critical mass could also be used—or slightly altered—for gender affirmative action or for other practices, such as in graduate admissions or employment hiring. This would make the dialogue surrounding affirmative action more productive by shifting the conversation from figuring out how to use affirmative action in contemporary contexts, and instead, toward discussing why we are using it and to what extent we should continue to do so. For all these reasons, I believe that the most productive and beneficial work to be done in affirmative action public policy is to increase the specificity of critical mass.

⁵This may be because one group is applying to the school at a much lower rate than others, even though there is a sufficient pool of qualified applicants that could apply each year from the under-applying group.

The most recent way in which critical mass has been used in litigation was in the Fisher II case from 2016. According to the court announcement, the University of Texas described critical mass as “an adequate representation of minority students so that the educational benefit that can be derived from diversity can actually happen.” This iteration leaves many aspects loosely defined on purpose. Words such as “adequate” and “actually happen” [19] imply a perceivable amount necessary for change to occur, but the definition does not provide what such a level could be. My current reworking of critical mass is the following: *the enrollment level necessary to attain a reflective demographic structure for a university, given its location and available applicant pool*. While still being vague enough for workability in a variety of contexts, there are some modifiers that make it a more rigorous definition than the one from the Fisher II case. For instance, instead of an “adequate representation,” I specify it as the necessary level of enrollment for the university to be reflective of the demographics relevant to their applicant pool. The addition of adding “given its location and available applicant pool” is necessary in order to allow for its application to a case where there is some sort of bias in admissions—i.e. in-state versus out-of-state admissions for public universities, minority admissions for Historically Black Colleges and Universities (HBCUs), or gendered admissions for single gender institutions. However, I still leave the goal as obtaining a necessary enrollment level to allow for a slight use of nuance.

This new definition is more specific than what has typically been used by undergraduate institutions, and I seek to make the implementation of critical mass even more transparent. While UT-Austin, or another university, may not want to define critical mass as an actual number or percentage, what they imply in their use of it is that there is a value in mind that will signify when their goal for diversity has been achieved. Critical mass can be an actual value, without running the risk of becoming a quota. The derived number or percentage is simply an ideal amount of representation in mind to achieve diversity. A quantification of critical mass will depend upon a few key factors relevant to demographics. Because public institutions have biased enrollment for in-state students, and since many state demographics are not reflective of the country at large, we need to find estimates of current demographics for both the state in question and the U.S. as a whole.

To see how this quantification can work, we will go through an example for finding one quantification of critical mass at UT-Austin⁶. The in-state demographics for Texas can be seen in Table .1 [22] and the U.S. demographics can be seen in Table .2 [22]. Since UT-Austin is coeducational and not an HBCU, we will assume that they have no bias in admitting students from those specific racial/ethnic backgrounds. However, since UT-

⁶All tables references in this paragraph will be put in Appendix A

Austin is a public university, many more students who attend are in-state residents of Texas, this disparity can be seen in Table .3 [27]. By using this pre-existing knowledge, we can create a critical mass breakdown at UT-Austin that would reflect fairly for all races/ethnicities. This can be done by summing the products of entries in Tables .1 and .2 with Table .3, respectively. Doing so, we can obtain the following estimation for critical mass: 41.3% for White/Caucasian students, 12.0% for African-American students, 4.6% for Asian-American students, and 36.0% for Hispanic/Latino students. However, if we compare these percentages to the existing data for the incoming freshman class in Fall 2016, we see that the breakdown is actually 42.5% for White/Caucasian students, 4.2% for African-American students, 20.7% for Asian-American students, and 22.6% for Hispanic/Latino students [27]. Therefore, for this specific quantification of critical mass, we have an over-enrollment (actual > predicted) of White/Caucasian and Asian-American students, and an under-enrollment (actual < predicted) of African-American and Hispanic/Latino students. Of course, this is only one potential interpretation and quantification of critical mass, which means that there can be different estimations that can be just as valid as this one. This is why it is important to continue working on better defining critical mass in order to limit the interpretability, while simultaneously keeping the definition workable.

While quantification on its own is a valid goal for making critical mass more workable, we aim to use this new quantification in an applied mathematical and statistical context. Almost all of the work done thus far in affirmative action policy assessment uses retroactive data analysis. However, retroactive data analysis does not make assessment of current policies possible. That means that up until now, the most prevalent way to assess these policies has been to wait several years until we could figure out if a new policy's implementation made an impact or not on the demographics.

This is where we change the mode of response for policy makers. We will create a predictive Markov chain model to assess current affirmative action policy and observe what its effect could be 5 or 10 years in the future. By using the predicted quantification of critical mass as a guideline, we can assess the future of a policy by comparing the model's results to this quantification. Therefore, using such methods for the prediction of enrollment—rather than focusing on past enrollment numbers—can help provide brand new insights on policies regarding their effectiveness in admitting underrepresented students according to a more specified notion of critical mass.

Chapter 3

MODELING DIVERSITY IN ACADEMIA

There is an inherent trade off of complexity for interpretability that comes with the mathematical model building process. Good mathematical models tend to be sufficiently simple while still adequately capturing the qualitative behavior of the process being described. The first step in building such a model for our problem is to identify the foundational modeling structures that can capture how we see students move throughout the college process. More specifically, this project focuses on tracking the movement of students—starting as seniors in high school—throughout each term in their undergraduate college experience. We construct our model to follow students through each stage¹ that occurs in the college process up to their departure, either through graduation or attrition. Two modeling structures which can be used for such dynamics are Ordinary Difference Equations and Markov chains.

The first section of this chapter examines past literature that also focuses on understanding diversity in academia. This is done in order to understand how to apply such modeling structures to our own problem. Specifically, we explore the application of ordinary difference equations and Markov chains. The second part of this chapter details the data we gather to help us complete our model.

3.1 Modeling Structures

Difference equations are useful for our model dynamics because we can discretize the stages a student goes through in their undergraduate college process. Since we are working with discrete time stages rather than continuous time, difference equations will be more useful in our model than differential equations. Therefore, from now on in this chapter, ODEs will stand for Ordinary *Difference* Equations, rather than differential equations.

At a Snail's Pace: Time to Equality in Simple Models of Affirmative Action Programs

In this paper, Feinberg examined how affirmative action policies could help achieve a better racial composition at a single occupational group or organization. One motivator for this study was the decision on United Steelworkers of America, AFL-CIO-CLC v. Weber. This case upheld the constitutionality of an affirmative-action-based training program used by the United Steelworkers of America and the Kaiser Aluminum and Chemical Corporation. This program aimed to increase the representation of black skilled craft workers. Their

¹These stages are described in Section 4.2

policy reserved half of the eligible positions in the training programs for black workers. Brian Weber, a white laboratory assistant at Kaiser, was not selected for the program and claimed reverse discrimination due to the affirmative action policy in place. The court wanted to know whether or not such a program violates Title VII of the Civil Rights Act of 1964, which prohibits discrimination based on race. The court upheld that the policy did not violate Title VII since it sought to eliminate historical patterns of racial segregation and did not prohibit white employees from advancing in the company [5]. The court also decided that since Kaiser's objective was to make their minority representation reflect the local population, once this goal was achieved, then the policy must be ended [12].

Using this court case as its basis for analysis, Feinberg created a simple model to examine the time needed to reach this criterion, while operating the policy under different conditions. Overall, Feinberg uses the affirmative action policy's structure and possible alterations to come up with the system of ODEs used in the model. The system starts with modeling numbers of black and white workers and then the ODEs are transformed to track proportions for each group in the training program. Specific conditions are then imposed on the model to see how the qualitative behavior differs with each change. For example, it looks at what occurs with the presence or absence of the affirmative action program while holding the number of positions constant for black workers. Feinberg also explores what occurs when a program is in place, and there is an option for the number of positions for black workers either to increase or decrease over time [8].

This paper is useful to understand one approach of modeling a given policy in place at an organization. Feinberg starts by taking an affirmative action policy resulting from litigation and evaluated how altering the policy would change the representation of black workers. We similarly begin with litigation as our starting point. However, we only seek to understand how the current (lack of) affirmative action policy at UC-Berkeley will change the racial/ethnic demographics for future incoming classes.

Leaks in the pipeline: separating demographic inertia from ongoing gender differences in academia

In this paper, ODEs are used in a similar context to Feinberg. Shaw and Stanton focus on understanding how gender demographics in varying stages in academia have shifted from year to year for many different subjects. They modeled several disciplines—mostly in the hard sciences, and especially mathematics and computer science—which have historically been comprised mostly of males. While there have been several hypotheses presented as to why this has been the case, the gender disparity has appeared to diminish (but not

entirely disappear) over time. The focus of this paper is to understand how much of this demographic shift is the consequence of "demographic inertia" and how much is due to gender discrimination. In this context, demographic inertia relates to the time lag that occurs as a consequence of the length of several stages in academia. In other words, while policies may be put in place to promote the inclusion of women in a specific discipline and/or career stage, we must wait some time for the lag of that career stage and others after it to assess parity [17].

To understand the effect of demographic inertia, Shaw and Stanton create a simple model consisting of a system of ODEs which represent the five different career stages in academia that they consider: undergraduate studies, graduate studies, postdoctoral fellowships, assistant professorships, and tenured professorships. They call it a "leaky pipeline" since the route into academia is often described analogously to a pipeline and its leakiness comes from the fact that there are several opportunities to exit the model over the years. The authors used data from the National Science Foundation (NSF) to estimate their transition parameters from 1979 to 2006. Additionally, they looked at varying disciplines in the sciences, engineering, and social sciences to see how demographic inertia affects disciplines with varying gender disparities [17].

To do this, they first create their model using a system of 10 ODEs to model the number of students progressing or dropping out of the model at various stages. Next, they create four more equations which model the proportion of females in each pool. The actual assessment of whether the change in female representation is hindered by gender discrimination starts with the calculation of Inertial Effect (IE). This is done by finding the ratio between the proportion of females within a class and discipline based on the data, to the predicted proportion from the model. An IE of 1 means that the actual female representation within a discipline matches the prediction exactly. However, a value less than 1 means there is less female representation than expected and a value greater than 1 implies more female representation than expected [17]. For our model, we will create an assessment criterion related to critical mass, and this paper is useful in understanding how to create a meaningful criterion and use it for model assessment.

Markov chains are also useful for the dynamics we seek to model for similar reasons as ODEs. Specifically, the states in a Markov chain pair well with the discrete stages we follow for undergraduates students, and the movement between these stages are easily modeled with Markov chains using transition probabilities. For these reasons, our model will employ

Markov chains and the theory behind them will be touched on in subsequent chapters dealing with model construction.

Analytical Models for Minority Representation in Academic Departments

In this paper by Wiley, Markov chains are used to look at the representation of minorities on an academic department level. The author specifically investigates whether unbiased hiring from a mixed minority and majority pool results in a deviation from the expected value of hires for minority populations. In this instance, a minority was defined as a member from any given subset or small fraction of the general population, while an individual in the majority was defined as someone who was not a minority [11].

The model's states are the number of minorities in the department at a given time. The assumption made in this model is that the number of minorities at the next hire can either stay the same or change by 1. The author then goes on to describe the resulting structure of the transition matrix. Next, the long-term behavior is described and the process of finding this behavior is explained [11]. Since the model in this case is not an absorbing Markov chain (ours will be, and this type of Markov chain will be explained in a later chapter), we will not use any of these techniques. Additionally, other strategies for assessment are introduced and discussed in the paper, but it does not appear these methods will be very relevant for the model structure we have. Regardless, seeing one way in which a Markov chain can be applied to models focused on minority representation in academia gives some idea on how to begin our model creation process.

3.2 Model Data

In any mathematical model, there is the need to construct parameters that control the model's behavior. In our model's construction, we need to find data relevant to determining how many students will be in each stage for a given time, as well as the transition rates between each stage. We selected data from the National Center for Education Statistics, the University of California, and UC-Berkeley's Office of Planning and Analysis. Our analysis concludes that only 9 pieces of data are needed to estimate all of the necessary parameters in the model. Chapter 4 will detail how we use this data to derive the parameters we need in the model.

The National Center for Education Statistics

Since the model begins with high school students moving through the college application process, it is helpful to determine how many high school students graduate each year in the United States. Because our model specifically tracks the movement of students directly from

high school to college, we will not include students that take a gap year, transfer students, or students that re-enroll in college after previously enrolling and not finishing. Therefore, the only information we need are totals of how many high school seniors are graduating per year, as they are the only students we assume will be enrolling in college the following year.

The National Center for Education Statistics² (NCES) publishes annual data on the number of nationwide high school graduates in their Digest of Education Statistics. This data is provided in Tables 219.10 and 219.30 under Chapter 2, Section 219: High School Completers and Dropouts [13]. The first table, Table 219.10, contains data from academic years 1869-70 through 2026-27 (although not contiguously) on the number of high school graduates by year, sex, and control of school. Rows for years 2013-14 until 2026-27 are projections made by the NCES [14]. We will use these projections instead of imputing our own data points for this time span. The pertinent information for this thesis are the raw totals of high school graduates by year. The second table, Table 219.30, contains data from academic years 1998-99 through 2026-27 on public high school graduates by race/ethnicity; and once again, for the academic years 2013-14 through 2026-27 the NCES reports their own projections. For our simple model, we will assume that the racial/ethnic makeup of public schools are similar enough to private schools so that we can use data from this table to generalize for the entire high school graduating class of each academic year. With this assumption in mind, we can use the columns from Table 219.30 that give the percentage distribution of graduates for the following racial/ethnic categories: White, Black, Hispanic, and Asian/Pacific Islander [15]. While there are two more categories included in the table, these four groups make up the majority of college racial/ethnic representation, therefore, our simple model will only track how students in these four groups progress through college.

The University of California

The next step in constructing our model will be to determine the rates of high school students progressing through the college application process. Therefore, it was necessary to find how many students apply to, are admitted to, and enroll at UC-Berkeley. The University of California publishes an undergraduate admissions summary each year. In this summary, they detail how many students apply to, get accepted to, and enroll to each of their campuses for a given year. This data is provided from 1994 through 2016. Not only do they provide the total numbers, but they also provide these numbers by race/ethnicity. The UC tracks more categories than the NCES does with high school graduates, but the UC

²The NCES is run by the Institute for Education Sciences (IES) and is part of the U.S. Department of Education. The IES is an independent, non-partisan organization with a mission to provide scientific evidence for which education practices and policies can be addressed [10].

similarly provides numbers for African American, Asian American, Chicano/Latino, and White students. Although the names for these groups that the NCES and UC each specify by race/ethnicity are not exactly the same, they are similar enough for us to consider them to represent the same students in similarly titled groups for each respective organization. For example, we will assume that a student who would be reported as Hispanic by NCES would also most likely self-identify as Chicano/Latino for the UC's data. We can also find similar pairings for the other three groups we will be modeling. Additionally, these undergraduate admissions summaries provide specific data, such as the type of school students attended (e.g. California public or private, out-of-state, etc.) or their residency status (e.g. California resident, out-of-state resident, etc.) [24]. In order to keep our model as simple as possible, we will not use these more detailed groupings, rather, we will focus solely on modeling the racial/ethnic groups. While we may not use data on residency explicitly in our model, this data will be used in Chapter 5 when we find our critical mass benchmarks for policy assessment. Lastly, the UC provides data for three additional racial/ethnic groups: American Indian, unknown, and international. While these groups make up roughly 15-20% of the data, attempting to model these groups in our model will overcomplicate the process, so we will therefore focus on the four racial/ethnic groups reflected in the NCES data set.

UC-Berkeley, Office of Planning and Analysis

The final pieces of information we need to construct the model parameters are provided by UC-Berkeley. Most institutions of higher education include an office that looks at their own institutional data to analyze patterns in attendance, hiring, students' choice of major, or any information they would like to understand as it changes annually. The specific office that conducts this work at UC-Berkeley is the Office of Planning and Analysis. The relevant data they provide for our model are freshman retention rates and undergraduate graduation rates. The Office of Planning and Analysis provides freshman retention rates from incoming freshman classes for Fall 2004 through Fall 2014. Additionally, they have data on their freshman four-, five-, and six-year graduation rates. The four-year graduation rate data spans the incoming classes from 2004 through 2011, the five-year data spans from 2004 through 2010, and the six-year data is provided for 2004 through 2009 [25]. UC-Berkeley also includes rates at which transfer students graduate, but we will restrict our model to freshman for simplicity. These rates will help model the movement of students at UC-Berkeley as they progress term by term until they eventually graduate or do not finish. While all the pieces of data described in this chapter are necessary in building our model, they are not sufficient in its completion. That is, we will discuss in the next chapter how to derive all the necessary parameters for our model, using these provided pieces of data.

Chapter 4

MARKOV CHAIN MODEL CONSTRUCTION

In this chapter, we will construct the Markov chain model for predicting racial/ethnic demographics at UC-Berkeley over the next 10 years. First, it is necessary to introduce the theory behind Markov chains. Then we begin to describe the states of our model. After, we establish general methods for deriving initial conditions and transition probabilities for our model. Lastly, we will show how to find the asymptotic behavior of our Markov chains.

4.1 Markov Chain Theory

In this section we will introduce the theory behind Markov chains, what components make up a Markov chain, what an absorbing Markov chain is, and how to answer questions related to the long-term behavior of such Markov chains.

What is a Markov Chain?

Markov chains are important because they are a "memoryless" model. This means that the past stages of the model do not necessarily influence future ones, that is, only the current state will influence what occurs in the subsequent state. The basic pieces of a Markov chain are *states*, and we can put them into a set $S = \{s_1, s_2, \dots, s_m\}$. These states can represent almost anything such as, the type of weather on a given day, the number of people in an area, the stops on a journey, really the possibilities are endless. A Markov chain operates on the ability to move between states, one at a time, where one movement is called a *step*. We say that the probability of moving from a state s_i to a state s_j occurs with probability p_{ij} , these probabilities are called *transition probabilities*. These transition probabilities allow for individuals to move throughout the stages in our model. We allow the ability to stay in a given state in the next time step and denote this transition probability as p_{ii} . Starting states can be specified by creating a vector, \mathbf{u} , consisting of probabilities defined on the set S . Typically this is done by specifying a specific starting state, s_i , where the i th element of the vector is 1 and the remaining entries are all 0. Otherwise, we have that each element $0 \leq u_i < 1$ and $\sum_{i=1}^m u_i = 1$ [9].

In our model, we have 24 states which model a student's progression throughout the college experience: a high school student in the fall, spring, or summer (1-3); a first-, second-, third-, fourth-, or fifth-year college student in their fall, spring, or summer term (4-18); a sixth-year college student in their fall or spring term (19-20); graduating after four, five, or

six years of attending college (21-23); or dropping out (24).

Using a Transition Matrix

Once we know all the transition probabilities, p_{ij} , we can put them into a single m -by- m array called a *transition matrix* where the ij th entry is the transition probability p_{ij} . It can be shown that given a transition matrix \mathbf{P} for a Markov chain, the ij th entry $p_{ij}^{(n)}$ of the matrix \mathbf{P}^n gives the probability of going from state s_i to state s_j in n steps. Next, we can define a vector \mathbf{u} to be the probability vector representing the starting distribution for the Markov chain. It can be shown that given this starting vector \mathbf{u} and the transition matrix \mathbf{P} , the probability that the chain is in state s_i after n steps is the i th entry in the vector $\mathbf{u}^{(n)} = \mathbf{u}\mathbf{P}^n$. These facts will be useful in eventually determining the long-term behavior of a given Markov chain [9]. For the model we will be creating, this will result in a 24-by-24 matrix, \mathbf{P} , of transition probabilities, and \mathbf{P}^n will tell us the probability of a student going from state s_i to state s_j in n steps where our time span for a single step is 4 months.

Absorbing Markov Chains

Sometimes in Markov chains the probability of leaving a given state is 0, meaning the transition probability is $p_{ii} = 1$ and the remaining probabilities are $p_{ij} = 0 \forall i \neq j$. When this is the case, such a state is called *absorbing*. We call any state that is not an absorbing state a *transient* state. Additionally, if a Markov chain has at least one absorbing state and it can be reached in finite steps, then the Markov chain is also called *absorbing* [9]. There are many processes that can be modeled with an absorbing state, and the type of process we are modeling is one. Our model is specifically an absorbing Markov chain because we assume that after a student graduates (either in 4, 5, or 6 years) or drops out, they do not re-enroll at the college.

Canonical Form for Absorbing Markov Chains

For an absorbing Markov chain, we can separate the states into two distinct sets $T = \{s_1, s_2, \dots, s_t\}$ and $G = \{s_{t+1}, s_{t+2}, \dots, s_{t+g}\}$, where the original set of states $S = \{T, G\}$ contains all $t + g = m$ states in the Markov chain. We will define T to be the set containing all t transient states and G to be the set containing all g absorbing states. We can then put the transition matrix, \mathbf{P} , into *canonical form*, by which we have

$$\mathbf{P} = \begin{matrix} & \begin{matrix} T & G \end{matrix} \\ \begin{matrix} T \\ G \end{matrix} & \begin{pmatrix} \mathbf{Q} & \mathbf{R} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \end{matrix}$$

where \mathbf{Q} is a t -by- t matrix, \mathbf{R} is a nonzero t -by- g matrix, $\mathbf{0}$ is an g -by- t zero matrix, and \mathbf{I} is an g -by- g identity matrix. Next, we can try to understand what form the matrix \mathbf{P}^n will have in this new canonical form since we need to use it to find the probability of being in state s_j after n steps. Basic matrix algebra gives us that

$$\mathbf{P}^n = \begin{matrix} & T & G \\ T & \mathbf{Q}^n & * \\ G & \mathbf{0} & \mathbf{I} \end{matrix}$$

where the asterisk, $*$, is a placeholder for some t -by- g matrix in the upper right-hand block matrix which results after \mathbf{P} is multiplied by itself n times. We can see that \mathbf{Q}^n gives us the probabilities of being in a transient state after starting in a transient state n steps prior. It can be shown that in an absorbing Markov chain, the probability of absorption is 1 or in other words $\mathbf{Q}^n \rightarrow \mathbf{0}$ as $n \rightarrow \infty$ [9].

The Fundamental Matrix

For a given absorbing Markov chain, we can find what we call the *fundamental matrix*

$$\mathbf{N} = \mathbf{I} + \mathbf{Q} + \mathbf{Q}^2 + \dots = (\mathbf{I} - \mathbf{Q})^{-1}$$

It can be shown that $\mathbf{I} - \mathbf{Q}$ is never singular and therefore always has an inverse. This matrix is useful because the ij th entry, n_{ij} , of \mathbf{N} is the expected number of times the process is in the transient state s_j after starting in the transient state s_i .

A question we may want to investigate is: given that the Markov chain starts in a transient state s_i , what is the expected number of steps before absorption? It can be shown that $\mathbf{t} = \mathbf{N}\mathbf{c}$, where \mathbf{c} is a column vector of all 1s, is a vector that gives this information. This means that the i th entry of \mathbf{t} , t_i , gives the expected number of steps before the chain is absorbed, given we start in transient state s_i .

Next, we can tackle the question of what the probability of absorption to a specific absorbing state is from a given transient state. It can be shown that the t -by- g matrix $\mathbf{B} = \mathbf{NR}$, where \mathbf{R} is defined as in the canonical form of \mathbf{P} , gives us those probabilities. That is, the ij th entry of \mathbf{B} , b_{ij} , is the probability of being absorbed into an absorbing state s_j after starting at a transient state s_i [9].

4.2 States for the Model

Prior to finding our initial conditions and transition rates, it will be beneficial to first describe the states for our Markov chain and how individuals will move throughout the

model. Since this process is specifically one that uses an absorbing Markov chain, our set of all states $\mathbf{S} = \{s_1, s_2, \dots, s_m\}$ can be split into distinct sets \mathbf{T} and \mathbf{G} , where $\mathbf{T} = \{\forall s_i \in \mathbf{S} : s_i \text{ is a transient state}\}$ and $\mathbf{G} = \{\forall s_j \in \mathbf{S} : s_j \text{ is an absorbing state}\}$.

As mentioned in section 4.1, we will have 24 states which model a student's progression throughout their time in college: a high school student in their fall, spring, or summer (1-3); a first-, second-, third-, fourth-, or fifth-year college student in their fall, spring, or summer term (4-18); a sixth-year college student in their fall or spring term (19-20); graduating after four, five, or six years of attending college (21-23); or dropping out (24). Since there are so many states and they are somewhat long to describe, we will create shorthands for each stage. To do this, we will pair a letter with a number to describe the 23 stages that have to do with a student's academic standing in the college process (excluding not finishing college). We will use a number from the set $\{0, 1, 2, 3, 4, 5, 6\}$ to describe the i th year the student is enrolled in college. This means that a high school student would be in their 0th year of college, freshman would be in their 1st year, and so on. Next, we will use the a letter from the set $\{A, B, C, G\}$ to describe what term a student is in or what standing a student has. Specifically, we will say that A corresponds to the fall term, B corresponds to the spring term, C corresponds to the summer term, and G corresponds with graduating. We can then combine one number with one letter to act as a shorthand for a stage in the model. For example, a high school senior in their summer would be described using $0C$, a third-year college student in their fall term would be $3A$, and a college student graduating in their sixth year would be $6G$. This will describe all of the first 23 stages described above, while the 24th stage will be denoted by DNF , meaning they did not finish their collegiate progression¹. This gives us that

$$\mathbf{S} = \{0A, 0B, 0C, 1A, 1B, 1C, \dots, 5A, 5B, 5C, 6A, 6B, 4G, 5G, 6G, DNF\}$$

Since we will assume that once a student graduates or drops out they do not re-enter the applicant pool, we can separate these states into our transient and absorbing states

$$\mathbf{T} = \{0A, 0B, 0C, 1A, 1B, 1C, 2A, 2B, 2C, 3A, 3B, 3C, 4A, 4B, 4C, 5A, 5B, 5C, 6A, 6B\}$$

$$\mathbf{G} = \{4G, 5G, 6G, DNF\}$$

With this in mind, we can now find the transition rates between these states and put them into the transition matrix (\mathbf{P}) of canonical form.

The transition matrix is one way to observe the dynamics of a Markov chain, but another way is to create a graphical representation of the process with nodes representing the states

¹Possibilities for moving to this class include not applying to, not getting admitted to, not enrolling at, or dropping out from the university being modeled.

from the model and directed edges representing the transition rates between states. A schematic of the directed graph for our model is seen as Figure 4.1². An individual begins their process in state 0A, a graduating high school student. That individual will then move throughout the graph according to the transition probabilities on directed edges. Following the blue path of edges represents progression through the transient states of the model. This means that an individual will continue to follow the blue path until they either move to a red path leading to the absorbing state *DNF*—signifying that they have decided to discontinue their collegiate progression—or on a green path leading to the absorbing state 4G, 5G, or 6G, signifying the individual has graduated in four, five, or six years. Now that this schematic has been developed, we can move to completing the numerical derivations for our initial conditions and the transition probabilities.

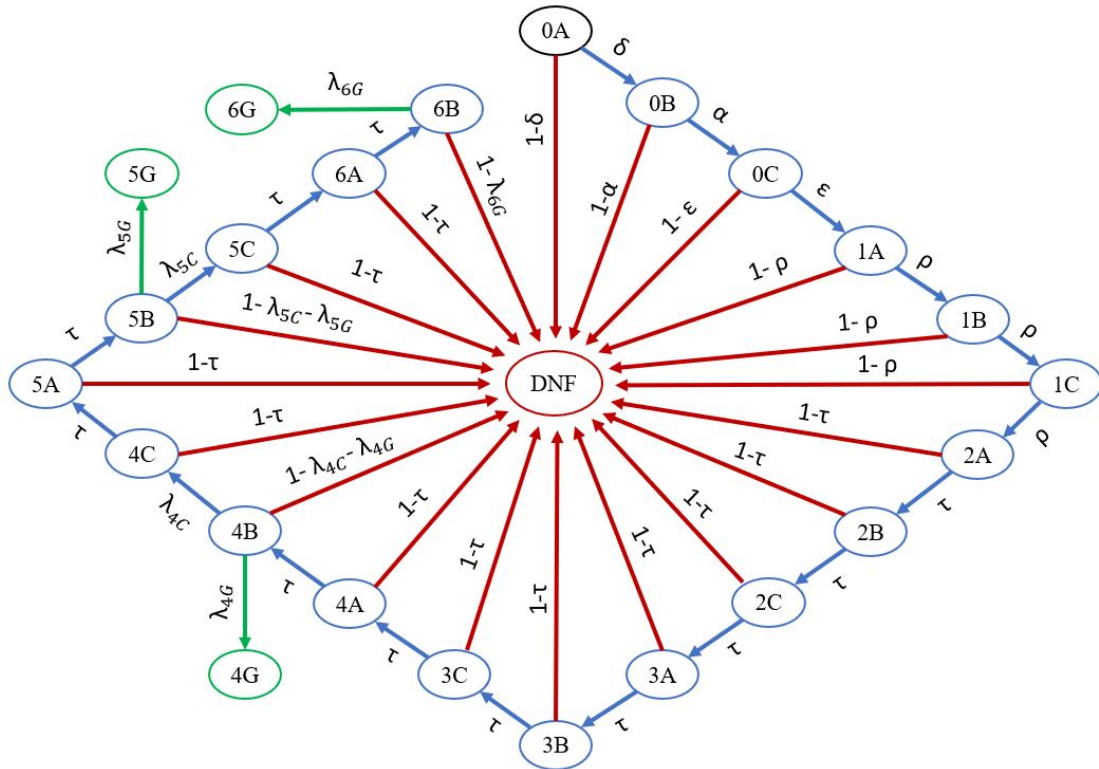


Figure 4.1: Graphical Representation of Markov Chain Model for College Progression

4.3 Initial Conditions for Individual Markov Chains

Using the data described in Chapter 3, we can begin to create the numerical items needed for our model. So that we can compare the demographic structure of each newly enrolled class and the university as a whole, we need to create a Markov chain for each class year i

²An additional model schematic can be seen in Appendix B.

and racial/ethnic group j . One assumption of our model is that the only way to enter the process is as a graduating high school senior (state $0A$). As a result, the vector \mathbf{u} of initial conditions will be of the form

$$\mathbf{u} = \begin{bmatrix} n_{ij} & 0 & 0 & \cdots & 0 & 0 \end{bmatrix}$$

where \mathbf{u} is a vector with 24 elements and n_{ij} is the number of high school graduate for group j in year i . We can find such initial conditions by using the data from the National Center for Education Statistics (NCES). As mentioned previously, this data consists of the number of high school graduates per year in the United States and the percentage that each racial/ethnic groups makes up each graduating class. The groups we will model are African-American students, Asian-American students, Hispanic/Latino students, and White/Caucasian students.

To find these initial conditions, we can simply multiply the total graduates for a given year by the percentage that each racial/ethnic group makes up for that year's graduating class. For example, from the data, we know that there were 3,199,650 graduating seniors in the class of 2007 [14]. We also know that the demographic makeup of this graduating class was 14.5% African-American, 5.4% Asian-American, 14.6% Hispanic/Latino, and 64.6% White/Caucasian [15]. Therefore we can estimate that there were $3199650 \times 0.145 \approx 463949$ graduating African-American students, $3199650 \times 0.054 \approx 172781$ graduating Asian-American Islander students, $3199650 \times 0.146 \approx 467149$ graduating Hispanic/Latino students, and $3199650 \times 0.646 \approx 2066974$ graduating White students in the high school class of 2007. In general, if we let N_i be the overall number of high school graduates and let p_{ij} be the percentage that group j makes up for the graduating class of year i , then the initial condition n_{ij} we need is $n_{ij} = N_i \times p_{ij}$. We seek to model numbers rather than percentages because different academic classes have different incoming sizes and interpreting results for the model using numbers is more intuitive than using percentages. Now that our initial conditions have been derived, we can progress to deriving the transition probabilities.

4.4 Transition Probabilities

Using data from the NCES digest tables and University of California (UC) undergraduate admissions summaries, we can find transition probabilities for movement between states $0A$ to $0B$, $0B$ to $0C$, and $0C$ to $1A$. The rate at which graduating high school students decide to apply to UC-Berkeley represents the movement from $0A$ to $0B$. We will call this rate δ , and find it in the following way: $\delta = \frac{\# \text{ applicants}}{\# \text{ high school graduates}}$. Next, the rate at which high school students are accepted to UC-Berkeley after applying is the movement from $0B$ to $0C$. This transition probability will be called α we find this similarly: $\alpha = \frac{\# \text{ acceptances}}{\# \text{ applicants}}$. Lastly, ϵ , the

rate at which students enroll at UC-Berkeley is the transition probability from 0C to 1A. We also find this in a similar manner: $\epsilon = \frac{\# \text{enrollees}}{\# \text{acceptances}}$. These rates will help us model the movement of high school students throughout their senior year. Next we focus on how to model the movement of first-year college students.

We saw previously that one of the data sets gathered contains the freshman retention rate for an incoming class. This is reported by UC-Berkeley's Office of Planning and Analysis and we will call such a rate ν . This data represents a student's movement after attending the university for a year. We must modify this rate to be used in our model since the time step for our Markov chain is in thirds of a year (4 months) while this data represents a full year's movement. Since Markov chains are "memoryless" processes, we know that progression through consecutive steps in the model entails repeated multiplication of the transition probabilities between states. This means that in order to convert ν into a transition probability that can be used to represent the movement from states 1A to 1B, 1B to 1C, and 1C to 2A, we must find a number that when multiplied by itself three times returns ν . This can be easily found to be the cubed root of ν . We will now call this new transition probability ρ which we can find using the following: $\rho = \sqrt[3]{\nu}$.

Finding the Characteristic Polynomial of Transition and Basic Transition Rate

The next transition rate we will find for our model is τ , the *basic transition rate*. This rate will denote the movement of an enrolled college student that stays in the model for the longest time possible before being forced to go to an absorbing state. This means we would like to find a rate τ which describes the movement from 1A to 6B. For simplicity, while this rate may change in between stages 4B and 5B or 5B and 6B, we will assume such a change is negligible enough that we can use the same τ throughout.

Now that we have these assumptions in mind, we can think about how to find τ . To more easily find such a rate, we will introduce the notation, \circ , which acts as a sort of "multiplication" in terms of how to get from one part of the model to another since the movements from state to state are independent events. Thus, we will use \circ between two states of the model while deriving an expression that analytically shows what percentage of students are in a given state s_j after starting in another state s_i .

To introduce this, we will begin by looking at moving by a single step. We know that to get from one state to a subsequent one, we must multiply the percent in the current state by a transition rate that corresponds with the rate of moving from the initial state to the subsequent one. We will use \circ to act as this multiplication that allows us to model movement between them. For example, if we assume that the entire college population starts in the

state 1A (i.e. the proportion in 1A is 1 for this derivation), and we know that the transition rates in a student's first-year are ρ , then we can write

$$1A \circ 1B = 1 \times \rho = \rho$$

With this \circ notation in mind, we will introduce another piece of notation, \rightarrow . We will use this when we already know the expression that shows the percentage of students in state s_j after starting in state s_i . That is $s_i \rightarrow s_j = \text{some expression}$. For example, now that we know what this expression is for the rate between 1A and 1B, we write that

$$1A \rightarrow 1B = \rho$$

Using both of these notations together will make finding τ easier. To understand how we can do this, we will start with a smaller example of finding the expression for the movement between 1A and 2A. We earlier found that $1A \rightarrow 1B = \rho$. Now, let us think about how to express the transition between 1B and 2A. We know that this process is achieved by moving from 1B to 1C then 1C to 2A. To do this, we would need to multiply the percentage of students in state 1B by the two transition rates between 1B and 2A. We can then see that

$$1B \circ 1C \circ 2A = (\rho)(\rho) = \rho^2 \implies 1B \rightarrow 2A = \rho^2$$

The way in which we can use the \circ notation to find $1A \rightarrow 2A$ is simple now that we have $1A \rightarrow 1B$ and $1B \rightarrow 2A$. We will assume is possible to find the expression that represents the percentage of students in 1B from some starting state prior to finding 1B to 2A. We will call expression $K(\mu)$, where K is some function of parameters in our model and μ is a vector of those parameters. Since we want to start from 1A, not 1B, we assume that $1A \rightarrow 1B = K(\mu)$. We found earlier that $1A \rightarrow 1B = \rho \implies K(\mu) = \rho$, which means

$$1A \rightarrow 2A = (1A \rightarrow 1B) \circ (1B \rightarrow 2A) = (\rho)(\rho^2) = \rho^3 = \nu$$

As we can see, using \circ and \rightarrow in these conventional manners will help create an efficient and understandable shorthand to find the movement between 1A and 6B.

The next part of this route pertains to the 7 transitions made between 2A and the first state with a graduation opportunity, 4B. We would like to find some number x , representing a new transition rate, such that when we multiply it by itself 7 times, we get the percentage of students in state 4B after starting in state 2A. This means that $2A \rightarrow 4B = x^7$ and gives us

$$1A \rightarrow 4B = (1A \rightarrow 2A) \circ (2A \rightarrow 4B) = \nu x^7$$

Now we can think about moving one step forward to state 4C, which will require a bit more work. Since we need the basic transition rate to represent the movement for enrolled

students, not graduated students, we must subtract the percentage of students graduating in 4 years (going to stage 4G) prior to moving one step forward to 4C. In the data provided by UC-Berkeley's Office of Planning of Analysis, we have the four-year graduation rate which we will call γ_4 . To find $1A \rightarrow 4C$, we first subtract the percentage of students graduating in four years from the percentage we know are in 4B, this gives us $\nu x^7 - \gamma_4$. We can then move one step forward to 4C and get that

$$1A \rightarrow 4C = (1A \rightarrow 4B) \circ (4B \rightarrow 4C) = (\nu x^7 - \gamma_4)x$$

Moving from 4C to 5B requires multiplication by the basic transition rate twice, so we get

$$1A \rightarrow 5B = (1A \rightarrow 4C) \circ (4C \rightarrow 5B) = (\nu x^7 - \gamma_4)x^3$$

Next, in a similar manner to finding $1A \rightarrow 4C$, we can find $1A \rightarrow 5C$. We once again need to subtract the percentage of students graduating in 5 years (i.e. going from 5B \rightarrow 5G) from this expression before multiplying by the basic transition rate. We will call the five-year graduation rate γ_5 . This new percentage of students graduating in 5 years will be the difference between those graduating in 5 years and those graduating in 4 years, $\gamma_5 - \gamma_4$. Following the same process as finding $1A \rightarrow 4C$ we get the following expression

$$1A \rightarrow 5C = (1A \rightarrow 5B) \circ (5B \rightarrow 5C) = [(\nu x^7 - \gamma_4)x^3 - (\gamma_5 - \gamma_4)]x$$

Multiplying by x two more times will give us the process of going from 1A to 6B, so

$$1A \rightarrow 6B = (1A \rightarrow 5C) \circ (5C \rightarrow 6B) = [(\nu x^7 - \gamma_4)x^3 - (\gamma_5 - \gamma_4)]x^3$$

We can then simply this polynomial to get

$$\begin{aligned} 1A \rightarrow 6B &= [(\nu x^7 - \gamma_4)x^3 - (\gamma_5 - \gamma_4)]x^3 - (\gamma_6 - \gamma_5) \\ &= (\nu x^7 - \gamma_4)x^6 - (\gamma_5 - \gamma_4)x^3 - (\gamma_6 - \gamma_5) \\ &= \nu x^{13} - \gamma_4 x^6 - \gamma_5 x^3 + \gamma_4 x^3 - \gamma_6 + \gamma_5 \\ &= \nu x^{13} - \gamma_4(x^6 - x^3) - \gamma_5(x^3 - 1) - \gamma_6 \end{aligned}$$

Since we defined τ to be such a number x which satisfies this polynomial, to find τ we simply find the root of this polynomial over the interval $[0, 1]$. This is because τ is a rate and must therefore exist only on this interval.

We will now call this polynomial, $T(x)$, the *polynomial of transition*

$$T(x) = \nu x^{13} - \gamma_4(x^6 - x^3) - \gamma_5(x^3 - 1) - \gamma_6 \quad (4.1)$$

It is easy to show that such a root exists over this interval.

Claim: The polynomial of transition, $T(x) = \nu x^{13} - \gamma_4(x^6 - x^3) - \gamma_5(x^3 - 1) - \gamma_6$, has a real root in the interval $[0, 1]$.

Proof:

We know that the characteristic polynomial is $T(x) = \nu x^{13} - \gamma_4(x^6 - x^3) - \gamma_5(x^3 - 1) - \gamma_6$, where $\nu, \gamma_4, \gamma_5, \gamma_6 \in [0, 1]$ since they are retention and graduation rates. It is known that a polynomial with real coefficients is continuous at every point.

Since ν represents the first-year retention rate, this rate must at least as much as any of the graduation rates for a given class. Additionally, the six-year graduation rate must be at least as much as the five-year graduation rate and the five-year graduation rate must be at least as much as the four-year graduation rate. Thus, will assume that $\nu \geq \gamma_6 \geq \gamma_5 \geq \gamma_6$.

We can see that $T(0) = -\gamma_6$. Since $\gamma_6 \in [0, 1]$, we know that $T(0) \leq 0$.

We can also see that $T(1) = \nu - \gamma_6$. Since $\nu \geq \gamma_6$, we know that $T(1) \geq 0$.

By the Intermediate Value Theorem \implies there must be a root of $T(x)$ on $[0, 1]$. ■

Deriving the Remaining Transition Rates

After finding τ , there still remain eight transition probabilities to find for the model to be complete. The remaining probabilities are related to states $4B$, $5B$, and $6B$ since these are the only states through which students can graduate. Particularly, we do not have transition probabilities for the following: $4B \circ 4G$, $4B \circ 4C$, $4B \circ DNF$, $5B \circ 5G$, $5B \circ 5C$, $5B \circ DNF$, $6B \circ 6G$, and $6B \circ DNF$. Earlier, we defined τ as the basic transition *rate*. This is because for some directed edges in our schematic it is the transition probability, but for $4B \circ 4C$ and $5B \circ 5C$ it will not be. While individuals will still progress from $4B$ to $4C$ and $5B$ to $5C$ at rate τ , due to graduating students leaving to an absorbing state, the actual percentage leaving $4B$ and $5B$, but remaining in the Markov chain, will not be the same as τ .

The process for deriving the remaining 8 transition probabilities relies on conditional probability. We know from conditional probability that $P(B|A) = \frac{P(A \cap B)}{P(A)}$. Intuitively, this is because we are choosing to reduce the probability space from all possibilities to only those occurring in conjunction with event A. Then, the probability that an event B occurs given another event A occurred is equivalent to the proportion of the new probability space where both events A and B occur. This means that conditional probability really just gives us the relative probability of an event occurring given another event occurred. We can use conditional probability to help find the remaining transition probabilities because none of

our states allow for movement to a state that occurred previously. Thus, for any two adjacent states, we know that the only way to get to state B is by coming from state A which implies that $P(B) = P(A \cap B)$, given that A is the state immediately preceding B.

Now, we can combine the definition of conditional probability with this new fact to see that $P(B|A) = \frac{P(B)}{P(A)}$ for any two adjacent states where we have $A \rightarrow B$. We will use this to find the transition probability from 4B to 4G and call this λ_{4G} . Note that the probability of going from 1A to 4G is the definition of the four-year graduation rate γ_4 . Therefore we see that

$$\lambda_{4G} = \frac{\% \text{ in } 4G}{\% \text{ in } 4B} = \frac{1A \rightarrow 4G}{1A \rightarrow 4B} = \frac{\gamma_4}{\nu\tau^7}$$

Similarly we can find λ_{4C} , the transition probability from 4B to 4C

$$\lambda_{4C} = \frac{(\% \text{ in } 4B - \% \text{ in } 4G)\tau}{\% \text{ in } 4B} = \frac{[(1A \rightarrow 4B) - (1A \rightarrow 4G)]\tau}{1A \rightarrow 4B} = \frac{(\nu\tau^7 - \gamma_4)\tau}{\nu\tau^7} = \frac{\nu\tau^7 - \gamma_4}{\nu\tau^6}$$

For the transition probability from 4B to *DNF*, $\lambda_{4B, DNF}$, rather than going through a similar process, from basic probability we can see that since

$$\lambda_{4C} + \lambda_{4G} + \lambda_{4B, DNF} = 1 \implies \lambda_{4B, DNF} = 1 - \lambda_{4C} - \lambda_{4G} = 1 - \frac{\nu\tau^7 - \gamma_4}{\nu\tau^6} - \frac{\gamma_4}{\nu\tau^7}$$

This is because we assume that those who do not graduate nor continue their studies after their fourth year in college drop out and do not re-enroll.

We can go through similar processes to find λ_{5G} , λ_{5C} , and $\lambda_{5B, DNF}$.

$$\lambda_{5G} = \frac{\% \text{ in } 5G}{\% \text{ in } 5B} = \frac{1A \rightarrow 5G}{1A \rightarrow 5B} = \frac{\gamma_5 - \gamma_4}{(\nu\tau^7 - \gamma_4)\tau^3} = \frac{\gamma_5 - \gamma_4}{\nu\tau^{10} - \gamma_4\tau^3} \quad \text{and}$$

$$\lambda_{5C} = \frac{(\% \text{ in } 5B - \% \text{ in } 5G)\tau}{\% \text{ in } 5B} = \frac{[(1A \rightarrow 5B) - (1A \rightarrow 5G)]\tau}{1A \rightarrow 5B} = \frac{[(\nu\tau^7 - \gamma_4)\tau^3 - (\gamma_5 - \gamma_4)]\tau}{(\nu\tau^7 - \gamma_4)\tau^3}$$

So we get that

$$\lambda_{5C} = \frac{\nu\tau^{10} - \gamma_4\tau^3 - \gamma_5 + \gamma_4}{\nu\tau^9 - \gamma_4\tau^2} = \frac{\nu\tau^{10} - \gamma_4(\tau^3 - 1) - \gamma_5}{\nu\tau^9 - \gamma_4\tau^2}$$

Once again, we will avoid explicitly finding the transition rate from 5B to *DNF*. Instead, we can find $\lambda_{5B, DNF}$ as follows

$$\lambda_{5B, DNF} = 1 - \lambda_{5C} - \lambda_{5G} = 1 - \frac{\nu\tau^{10} - \gamma_4(\tau^3 - 1) - \gamma_5}{\nu\tau^9 - \gamma_4\tau^2} - \frac{\gamma_5 - \gamma_4}{\nu\tau^{10} - \gamma_4\tau^3}$$

since we are again assuming that any student who does not graduate nor continue their studies after their fifth year will have dropped out.

Lastly, we will repeat this process once more to find λ_{6G} and $\lambda_{6B,DNF}$

$$\lambda_{6G} = \frac{1A \rightarrow 6G}{1A \rightarrow 6B} = \frac{\% \text{ in } 6G}{\% \text{ in } 6B} = \frac{\gamma_6 - \gamma_5}{[(v\tau^7 - \gamma_4)\tau^3 - (\gamma_5 - \gamma_4)]\tau^3} = \frac{\gamma_6 - \gamma_5}{(v\tau^7 - \gamma_4)\tau^6 - (\gamma_5 - \gamma_4)\tau^3}$$

So we get that

$$\lambda_{6G} = \frac{\gamma_6 - \gamma_5}{v\tau^{13} - \gamma_4\tau^6 - \gamma_5\tau^3 + \gamma_4\tau^3} = \frac{\gamma_6 - \gamma_5}{v\tau^{13} - \gamma_4(\tau^6 - \tau^3) - \gamma_5\tau^3}$$

Once more, we will implicitly find our last rate, $\lambda_{6B,DNF}$, as follows

$$\lambda_{6B,DNF} = 1 - \lambda_{6G} = 1 - \frac{\gamma_6 - \gamma_5}{v\tau^{13} - \gamma_4(\tau^6 - \tau^3) - \gamma_5\tau^3}$$

4.5 The Transition Matrix and Finding \mathbf{N} , \mathbf{t} , and \mathbf{B}

Now that we have found all the transition rates needed for our model, we can put together the transition matrix \mathbf{P} . We will put \mathbf{P} into canonical form because then we can easily find the fundamental matrix \mathbf{N} , which will allow for understanding the long term behavior of our model. While our model results will not be dependent on the asymptotic behavior of the Markov chain, providing the framework for observing such behavior will allow others to study aspects we do not delve into for this project. We recall that the canonical form of a Markov chain's transition matrix is of the form

$$\mathbf{P} = \begin{matrix} & \begin{matrix} T & G \end{matrix} \\ \begin{matrix} T \\ G \end{matrix} & \begin{pmatrix} \mathbf{Q} & \mathbf{R} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \end{matrix}$$

where \mathbf{Q} , \mathbf{R} , $\mathbf{0}$, and \mathbf{I} are described as in section 4.1. We can now construct these matrices using the sets T and G described in Section 4.2 and the transition probabilities we derived in Section 4.4. Putting these items together, we can see the associated block matrices for this Markov chain in Appendix C.1.

After putting the transition matrix into canonical form, we can now find the fundamental matrix, \mathbf{N} . The first step in this process is to find $(\mathbf{I} - \mathbf{Q})$, the associated matrix for our Markov chain can be seen in Appendix C.2. Due to the off-diagonal nature of \mathbf{Q} , this results in an upper triangular matrix where every element on the diagonal is 1. The process of finding the inverse of a matrix using elementary row operations, as well as the formula for finding the elements of \mathbf{N} are shown in Appendix D.1-D.3.

Since \mathbf{N} is upper triangular, we can see that it will be of the following form

$$\mathbf{N} = (\mathbf{I} - \mathbf{Q})^{-1} = \begin{pmatrix} 1 & n_{1,2} & n_{1,3} & \dots & n_{1,18} & n_{1,19} & n_{1,20} \\ 0 & 1 & n_{2,3} & \dots & n_{2,18} & n_{2,19} & n_{2,20} \\ 0 & 0 & 1 & \dots & n_{3,18} & n_{3,19} & n_{3,20} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & n_{18,19} & n_{18,20} \\ 0 & 0 & 0 & \dots & 0 & 1 & n_{19,20} \\ 0 & 0 & 0 & \dots & 0 & 0 & 1 \end{pmatrix} \quad (4.2)$$

Now that we have \mathbf{N} , we can find the vector \mathbf{t} whose elements give the expected number of steps before the chain is absorbed, given we start in a state s_i . To do this, we must first set up a vector \mathbf{c} , which is a 20-by-1 vector of 1s. We know that the vector \mathbf{t} is defined as $\mathbf{t} = \mathbf{N}\mathbf{c}$, so we see that

$$\mathbf{t} = \left(t_{0A} \ t_{0B} \ t_{0C} \ t_{1A} \ t_{1B} \ t_{1C} \ \dots \ t_{5A} \ t_{5B} \ t_{5C} \ t_{6A} \ t_{6B} \right)^T \quad (4.3)$$

The elements of \mathbf{t} are long and messy, but are explicitly given in Appendix D.4.

Lastly, we can find the matrix \mathbf{B} whose elements b_{ij} describe the probability of being absorbed into the absorbing state s_j after starting in the transient state s_i . This matrix \mathbf{B} is defined as $\mathbf{B} = \mathbf{NR}$, so we have that

$$\mathbf{B} = \begin{pmatrix} p_{0A,4G} & p_{0A,5G} & p_{0A,6G} & p_{0A,DNF} \\ p_{0B,4G} & p_{0B,5G} & p_{0B,6G} & p_{0B,DNF} \\ p_{0C,4G} & p_{0C,5G} & p_{0C,6G} & p_{0C,DNF} \\ \vdots & \vdots & \vdots & \vdots \\ p_{5C,4G} & p_{5C,5G} & p_{5C,6G} & p_{5C,DNF} \\ p_{6A,4G} & p_{6A,5G} & p_{6A,6G} & p_{6A,DNF} \\ p_{6B,4G} & p_{6B,5G} & p_{6B,6G} & p_{6B,DNF} \end{pmatrix} \quad (4.4)$$

Each elements' subscript represents the transient and absorbing states which the probability relates to. For example, element $p_{3A,5G}$ is the probability of being absorbed into the absorbing state $5G$ after starting in the transient state $3A$. The elements of \mathbf{B} are also messy but are described explicitly in Appendix D.5.

Chapter 5

MODEL RESULTS

With our Markov chain model fully constructed, we can begin to analyze its behavior over the next 10 years. We will use the model to predict incoming enrollment at UC-Berkeley over that span for each of the four racial/ethnic groups we are tracking. We will then construct ideal critical mass projections for the same years. Finally, we will develop a metric to understand how close or far our model predictions are from the ideal critical mass projections in order to assess the efficacy of the current admissions policy at UC-Berkeley.

5.1 Markov Chain Model Predictions

The first step in finding our results is to use the Markov chain model to predict enrollment at UC-Berkeley for the next 10 years.

Data Predictions for Markov Chain Model

Since our model is predictive, we need to forecast the data we gathered from the National Center for Education Statistics (NCES), UC undergraduate admissions summaries, and UC-Berkeley's Office of Planning and Analysis for the next 10 years.

In Figure 5.1 we can see a plot of the data from the NCES on the number of graduating high school seniors per year in the United States. These points span from 1980 through 2027. The most recent reports only give observed data up to the graduating class of 2013, then the NCES has their own projections for the remaining years up to 2027. The blue dashed line in Figure 5.1 represents where the observed data ends and the projected data begins, while the red data points represent the 10 year span our model will be predicting for—the graduating classes of 2018 through 2027. Demographic information provided by the NCES follows the same structure of observed data until 2013 and projected data from 2014 through 2027. Once again, we will use the NCES' projections

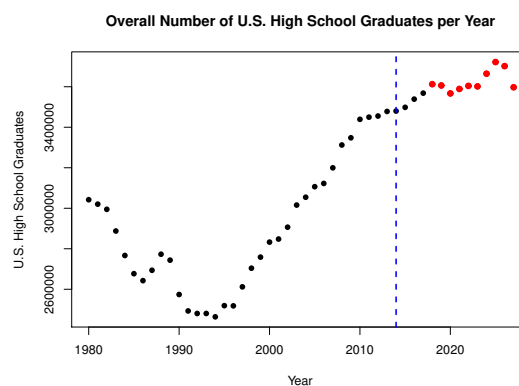


Figure 5.1: U.S. High School Graduates By Year

from 2018 through 2027 rather than creating our own.

Data reported in the UC undergraduate admissions summaries end in 2017. Therefore, we will make our own predictions over the next 10 years using a log-linear regression. Figure 5.2 depicts plots comparing the performance of a regular linear regression and a log-linear regression on annual applicant for UC-Berkeley by race/ethnicity. The black data points are observed values for applicant data from 1994 through 2016. We run a linear regression (red) and log-linear regression (blue) on this data and plot lines for the fitted values from 1994 through 2016 to compare their performances on the observed data. Lastly, we add the most recently observed value (2017) in purple and compare the performance of each regression for the year 2017. Results from these models indicated that the fitted values for the log-linear regression aligned better than for the regular linear regression, and that the performance for 2017 with the log-linear regression is closer than for the regular linear regression. Appendix E1-2 shows the comparisons for data on acceptances and enrollments by year. Comparing both regression across all 3 data sets, we see that the log-linear performs as well or better than the regular linear regression, so we will use a log-linear regression to predict on the UC undergraduate admissions data for the next 10 years.

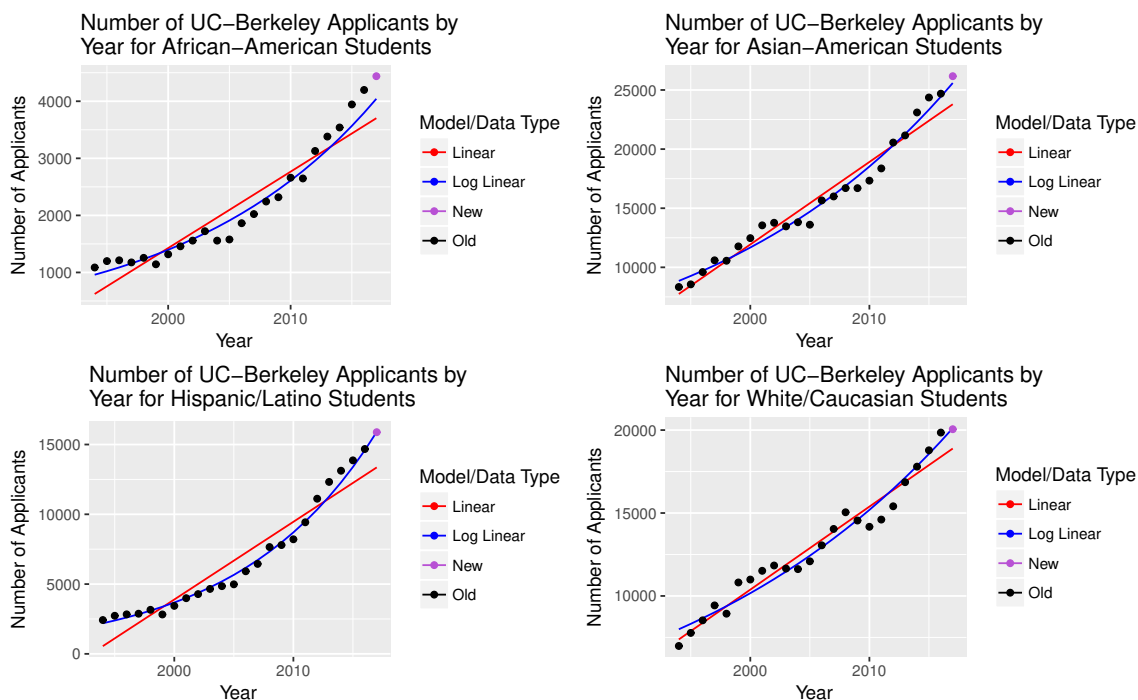


Figure 5.2: Comparing Linear Regression and Log-Linear Regression for UC Data

Lastly, we will use a basic linear regression on the data from UC-Berkeley’s Office of Planning and Analysis to assess the first-year retention rate and graduation rates for incoming

classes. In Figure 5.3, we see the observed rates for these data sets in black and the blue points are predicted values according to a linear regression fit to the respective observed data. Note that the y-axis for each graph is quite narrow (spanning less than 0.05 units) except for the 4-year graduation rate (which spans about 0.25 units). While the sample sizes for all the graduation rate data are less than 10, the fits appear reasonable so we will use these predictions in our Markov chain model.

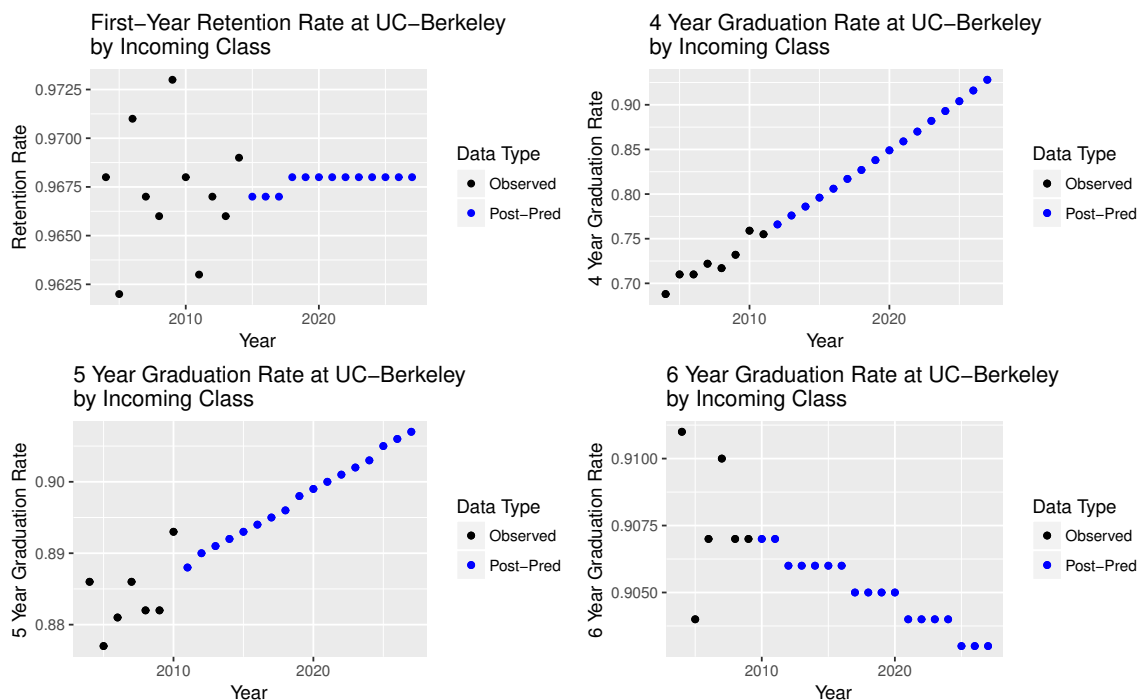


Figure 5.3: UC-Berkeley Retention and Graduation Rate Data with Linear Regression Predictions

Markov Chain Model Results

After predicting all the necessary data for the model, we can run the Markov chains to generate results. To do this, we create individual models for each racial/ethnic group and each year, and then repeatedly multiply the associated transition matrix by itself to iterate forward each 4-month time step. One way to accomplish this modeling structure is by staggering models according to the incoming year they represent, we can then track the demographic structure of the entire university in each time step until every individual in each Markov chain goes to an absorbing state. However, we will follow a simpler approach, which models up to the stage 1A–fall term in the first year of college. Thus, we will only be observing whether the incoming first-year classes are moving towards or away from critical mass rather than the entire universities demographics.

For our model results, we will simulate each Markov chain up to 1A, save the enrollment numbers from each racial/ethnic group for the same year in a vector, and then normalize this vector to find each group’s respective percent makeup of that incoming class to UC-Berkeley. Table 5.1 shows the results we get from performing this process to acquire predictions for the incoming classes for the next 10 years at UC-Berkeley. A more exact output can be seen in Appendix E3. Notice that none of the percentages appear to be shifting more than a few percentage points over this time span. This means that we predict the racial/ethnic demographic structure at UC-Berkeley to be somewhat stagnant over the next 10 years. Now that we have results from our predictive model, we can go through the process of quantifying the ideal critical mass percentages over the next 10 years to compare against these expected results.

Table 5.1: Markov Chain Model Results for Enrollment Percentage, 2018-2027 Incoming Classes

Race/Ethnicity	Year									
	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027
African-American	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03
Asian-American	0.50	0.50	0.49	0.49	0.49	0.49	0.49	0.49	0.48	0.48
Hispanic/Latino	0.17	0.17	0.18	0.18	0.19	0.19	0.20	0.20	0.21	0.21
White/Caucasian	0.30	0.30	0.30	0.29	0.29	0.29	0.28	0.28	0.28	0.27

5.2 Quantifying Critical Mass for the Future

At the end of Section 2.3, we briefly discussed one method of quantifying critical mass, specifically for the University of Texas, Austin (UT-Austin). To do this, we needed data on Texas’ state demographics, the United States’ demographics, and UT-Austin’s campus demographics. That method was straightforward because we used observed data from 2016 to create the ideal critical mass percentages for each racial/ethnic group, and then compared those values with the actual percentages at UT-Austin. However, since our Markov Chain model is predictive and based on UC-Berkeley, we need to project data over the next 10 years for California state’s demographics, the United States’ demographics, and UC-Berkeley’s campus demographics. Once these projections are found, we can make our quantification of the ideal critical mass at UC-Berkeley and compare against our predictive model’s results.

Data for Critical Mass Projections

The State of California Department of Finance provides state demographic projections from 2016 until 2060 [18]. Additionally, the U.S. Census Bureau provides population projections from 2016 through 2060 [21]. The data from both sources are grouped by county, age, sex, race, and Hispanic origin. For our ideal critical mass projections, we will use the

data that accounts for year, Hispanic origin, and race. We need to take data from each year regarding the number of African-Americans, Asian-Americans, Hispanics/Latinos, and Whites/Caucasians in these projections since the data for our model is partitioned into those racial/ethnic groups.

To create these partitions, we will first reduce each year's data to include only those that include African-American, Asian-American, White/Caucasian, and/or Hispanic race/ethnicity identifications. According to the data, someone is either categorized as Hispanic (H) or not Hispanic ($\neg H$), and then they are categorized as African-American (B), Asian-American (A), or White/Caucasian (W). We can treat these groups as sets H , B , A , and W , whose elements are people. We can see that anyone who is in B , W , or A , but not in H can be categorized as only African-American, Asian-American, or only White/Caucasian. Thus, for our projections we find that our African-American population is $B - H$, our Asian-American population is $A - H$, and our White/Caucasian population is $W - H$. Lastly, we need to find the Hispanic/Latino population using these sets. We can assume that anyone categorized as Hispanic and one of the racial categories will not be in either of the three populations we just found. This means that our overall Hispanic population is $H \cap B + H \cap A + H \cap W$. Therefore, given the sets B , A , W , and H , we can derive the following populations

African-American Population: $B - H$

Asian-American Population: $A - H$

White/Caucasian Population: $W - H$

Hispanic/Latino Population: $H \cap B + H \cap A + H \cap W$

Now that we have the necessary demographic data for California's and the United State's demographics, we only need to find the residency demographics of UC-Berkeley. The UC undergraduate admissions summaries detail an enrolled student's residency status in conjunction with their race/ethnicity and incoming class year [24]. We will use this to determine how many students from each racial/ethnic group come from the in-state applicant pool and the out-of-state applicant pool. While the data provided by California and the U.S. are projections for the years we would like to find critical mass percentages for, the data from the UC undergraduate admissions summaries only contain observed data from 1994 through 2017. Because of this, we will need to create our own projections for the residency breakdown at UC-Berkeley for the incoming classes of 2018 through 2027.

Projection Process for UC-Berkeley Residency Data and Ideal Critical Mass Results

We find it necessary to create our own projections for the in- vs. out-of-state representation of UC-Berkeley's incoming class over the next 10 years. To do this, we will use a stochastic process rather than a deterministic process like we did for the Markov chain model. We choose to use a stochastic process because it shows another method for generating predictions than the deterministic method we used in the Markov chain model. The stochastic process will use the following steps: (1) find the kernel density for each respective observed data set, (2) draw a sample from that density and save it, (3) add that sample to the observed data set, and (4) repeat this process 10 times, since we are predicting for the next 10 years. We repeat this process for each racial/ethnic group and each residency status. This stochastic process will give us how many students we would expect in a given year and for a given race/ethnicity to be a California resident and out-of-state U.S. resident. We can take both these numbers for a given year and normalize between the two of them to find the percentage of students we would expect for each residency status.

Now that we have all our necessary projections, we can think about how to generally construct ideal critical mass projections. Let us say that CA_{ij} is the percent that racial/ethnic group j makes up in year i from the California demographic projections. Next, we will say that US_{ij} is the percent that racial/ethnic group j makes up in year i from the U.S. population projections. We will also say that $UCB_{ij,in-state}$ and $UCB_{ij,out-of-state}$ are the percent that in-state or out-of-state residents respectively make up the incoming class at UC-Berkeley for racial/ethnic group j and year i . We will then use the following equation to construct the ideal critical mass percentage at UC-Berkeley for racial/ethnic group j in year i

$$CM_{ij} = CA_{ij} \times UCB_{ij,in-state} + US_{ij} \times UCB_{ij,out-of-state}$$

After running this process 10,000 times, we get the following means:

Table 5.2: Ideal Critical Mass Projections for 2018-2027 Incoming Classes

Race/Ethnicity	Year									
	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027
African-American	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07	0.07
Asian-American	0.13	0.13	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14
Hispanic/Latino	0.37	0.37	0.37	0.38	0.38	0.38	0.38	0.38	0.39	0.39
White/Caucasian	0.43	0.43	0.42	0.42	0.42	0.41	0.41	0.41	0.41	0.41

A more exact output for Table 5.2 can be seen in Appendix E4, as well as the standard

deviations for these projections in Appendix E5. With these ideal critical mass projections created, we can now compare them to the Markov chain model results.

5.3 Assessment Criterion and Results

After running our Markov chain model and projecting the ideal critical mass percentages for the next 10 years, we can compare both results to understand how much our model predictions differ from our critical mass projections. We will use the following criterion

$$\text{Markov Chain Predictions} \times \left(\frac{100}{\text{Critical Mass Projections}} \right)$$

We would like to compare how many students we predict will enroll at UC-Berkeley for each racial/ethnic group to how many we would ideally see as determined by our critical mass projections. This metric will tell us how many students we *predict* to enroll at UC-Berkeley per 100 students we ideally *project* to enroll. Multiplying the Markov chain prediction for a group by $100/(\text{critical mass projection})$ will allow us to do this. This will normalize the ratio between the Markov chain predictions and ideal critical mass projections to allow for the interpretation to be based off of each group of 100 students enrolling rather than by each student (which is what would occur if we did not multiply the ratio by 100). Figure 5.4 plots this assessment criterion for each racial/ethnic group we are modeling over the 10-year period we are assessing. The exact results used to create this plot can be seen in Appendix E6.

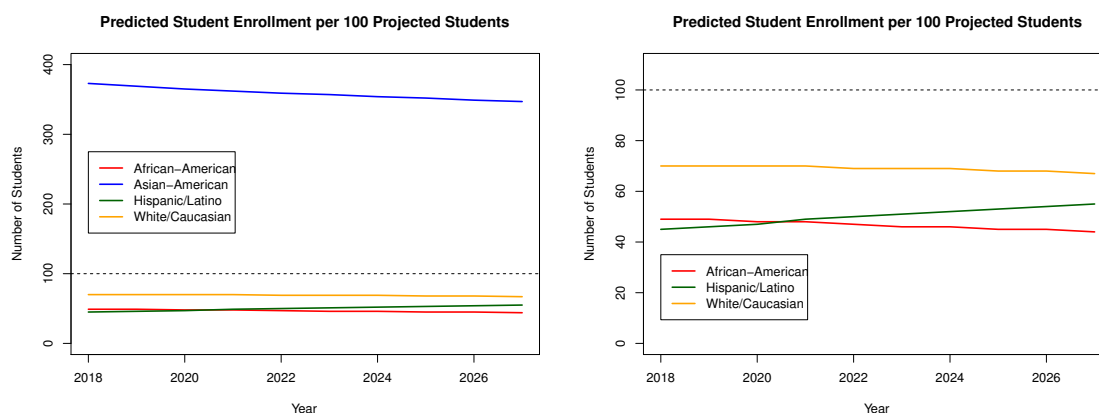


Figure 5.4: Assessment Criterion Results

In Figure 5.4, the horizontal dotted line at 100 represents an ideal outcome. We say an outcome is ideal if the Markov chain model predicts the same percentage that the ideal critical mass projections does. This means that for every 100 students we expect to enroll

based on our critical mass projections, our Markov chain model also predicts 100 students will enroll. Any result greater than 100 corresponds to overrepresentation for a given group and any result less than 100 subsequently corresponds to underrepresentation for a given group. The results determine that the only overrepresented group is Asian-American students, which move from about 375 to about 350. This means that for every 100 Asian-American students we would expect to enroll at UC-Berkeley for a given year, we actually see about 350 students enroll by 2027.

The results for each of our underrepresented groups differ in behavior from each other. We see that White/Caucasian students are predicted to change from about 70 to 65 over the next 10 years. This means that for every 100 White/Caucasian students we would expect to enroll, we actually see about 30 – 35 fewer students. We see a similarly sized decline for African-American enrollment, as it begins at about 50 and decreases to about 45. This means that by the time we reach 2027, for every 100 African-American students would expect to enroll at UC-Berkeley, we would see about 50 – 55 fewer students. Lastly, the representation for Hispanic/Latino students appears to improve over the next 10 years, however, by 2027 we still only anticipate 55 students in attendance per each 100 students that we ideally project would attend. The biggest takeaway from this plot is that none of our lines are approaching 100 in our 10-year time span, nor do they appear to be on pace to approach 100 in a feasible time period. This means that for future incoming classes at UC-Berkeley, improvements can be made for movement towards critical mass .

Chapter 6

CONCLUSION

Assessing affirmative action policies in undergraduate admissions procedures in the United States is important for ensuring educational equity. Specifically, racial affirmative action policies provide the opportunity for students to have the same access to higher education, regardless of racial/ethnic background. The goal of this thesis was to develop a method of predicting how the current (lack of) racial affirmative action policy at UC-Berkeley would effect the racial/ethnic representation of incoming classes over the next ten years. By using a Markov chain model and an ideal quantification of critical mass, our results indicate that students from different racial/ethnic backgrounds at UC-Berkeley are not accessing higher education at equal rates. Our results suggest that African-American students, Hispanic/Latino students, and White/Caucasian students are underrepresented, while Asian-American students are overrepresented.

Importantly, these results imply that different racial/ethnic groups are most likely not applying or matriculating to UC-Berkeley at the same rate. Implementing an affirmative action policy to rectify these differences is not currently feasible due to Proposition 209, which prohibits public education institutions in California from selecting students on the basis of race, sex, or ethnicity. Consequently, no action can legally be taken to influence the admissions rate for any group. However, there are certain strategies that can be employed to affect other rates. For example, an admissions office can try to recruit more students because some racial/ethnic groups may not decide to apply to UC-Berkeley at comparable rates. Another strategy could be for the campus administration to provide more campus resources—be that financial, educational, social, etc.—to increase the matriculation rate for groups that are committing to UC-Berkeley lower than we would anticipate. While explanations for such disparities and exact methods for decreasing them are not within the scope of this thesis, the use of our Markov chain model in conjunction with an assessment criterion for expected enrollment can help identify where these disparities may be occurring.

While this thesis does provide a novel framework for identifying strategies to improve educational equity in college admissions, we must simultaneously take into account the limitations of our methods. Many of these limitations stem from assumptions we had to make along either the data collection or model construction process. One assumption we made was that the racial/ethnic demographics for public school graduates are comparable to private

school graduates. Data from the NCES only lists these demographics for public school graduates. Since it appears that less than 10% of high school graduates are from private schools, we made the assumption that we could use the same racial/ethnic percentages of each graduating class for the total number of graduates even though the number is only reflective of public school graduates. Another assumption we made with the data was that freshman retention rates and graduation rates were equal across all racial/ethnic groups. This was because UC-Berkeley's Office of Planning and Analysis only reported data on the overall rates. Finding data on these rates for each group would make our dynamics more accurately representative of current rates. Other limitations had to do with simplifications of the college process that we made to more easily construct our Markov chain model. Some of these include excluding transfer students, students from other racial/ethnic backgrounds or origins, and excluding off-cycle students¹.

There are several directions this project could take going forward. One first step would be to mitigate some of these limitations. Since this was the first predicted model created for racial affirmative action in undergraduate admissions, the aim was to make it as simple as possible while still managing to retain current social dynamics. However, something to consider is that while increasing model complexity might allow for more accurate results, it could also complicate model explanation/interpretation. Another direction for future investigations would be to apply this model to another university. Initially, this would allow for the comparison of results across universities, but would also allow for confirmation that the model captures the dynamics occurring at other institutions. Lastly, it would be interesting to see if this modeling framework could be manipulated and applied to other areas of affirmative action such as gender representation in undergraduate admissions or racial/ethnic representation in workplaces.

¹Off-cycle students here means students on track to graduate in a term other than spring.

BIBLIOGRAPHY

- [1] Chicago-Kent College of Law at Illinois Tech. *Fisher v. University of Texas*. 2016. URL: <https://www.oyez.org/cases/2015/14-981>.
- [2] Chicago-Kent College of Law at Illinois Tech. *Gratz v. Bollinger*. 2003. URL: <https://www.oyez.org/cases/2002/02-516>.
- [3] Chicago-Kent College of Law at Illinois Tech. *Grutter v. Bollinger*. 2003. URL: <https://www.oyez.org/cases/2002/02-241>.
- [4] Chicago-Kent College of Law at Illinois Tech. *Regents of the University of California v. Bakke*. 1978. URL: <https://www.oyez.org/cases/1979/76-811>.
- [5] Chicago-Kent College of Law at Illinois Tech. *United Steelworkers of America, AFL-CIO-CLC v. Weber*. 1978. URL: <https://www.oyez.org/cases/1978/78-432>.
- [6] Clinton, William J. *Evaluation of Affirmative Action Programs*. Ibiblio, 1995. URL: <https://www.ibiblio.org/pub/archives/whitehouse-papers/1995/Jul/1995-07-19-Presidential-Memorandum-on-Affirmative-Action>.
- [7] Cornell Law School. *Strict Scrutiny*. LII / Legal Information Institute, 2016. URL: https://www.law.cornell.edu/wex/strict_scrutiny.
- [8] Feinberg, Williams E. *At a Snail's Pace: Time to Equality in Simple Models of Affirmative Action Programs*. 1984. URL: https://www.jstor.org/stable/2779331?seq=1#page_scan_tab_contents.
- [9] Grinstead, Charles M. and Snell, J. Laurie. *Introduction to Probability*. 1997. Chap. Chapter 11: Markov Chains. URL: https://www.dartmouth.edu/~chance/teaching_aids/books_articles/probability_book/amsbook.mac.pdf.
- [10] Institute of Education Statistics. *About IES*. 2018. URL: <https://ies.ed.gov/aboutus/>.
- [11] Johnson, Kurt N. and Wiley, J.D. *Analytical Models for Minority Representation in Academic Departments*. 2000. URL: <https://link.springer.com/article/10.1023/A:1007084509600>.
- [12] Justia. *Steelworkers v. Weber, 443 U.S. 193*. 1979. URL: <https://supreme.justia.com/cases/federal/us/443/193/>.
- [13] National Center for Education Statistics. *Most Current Digest Tables*. 2017. URL: https://nces.ed.gov/programs/digest/current_tables.asp.
- [14] National Center for Education Statistics. *Table 219.10. High school graduates, by sex and control of school: Selected years, 1869-70 through 2026-27*. 2017. URL: https://nces.ed.gov/programs/digest/d16/tables/dt16_219.10.asp?current=yes.

- [15] National Center for Education Statistics. *Table 219.30. Public high school graduates, by race/ethnicity: 1998-99 through 2026-27*. 2017. URL: https://nces.ed.gov/programs/digest/d16/tables/dt16_219.10.asp?current=yes.
- [16] National Conference of State Legislatures. *Affirmative Action: Court Decisions*. 2016. URL: <http://www.ncsl.org/research/education/affirmative-action-court-decisions.aspx>.
- [17] Shaw, Allison K. and Stanton, Daniel E. *Leaks in the pipeline: separating demographic inertia from ongoing gender differences in academia*. 2012. URL: <http://www.jstor.org/stable/pdf/41727364.pdf?refreqid=search:d15a7abf3%20afa79c830aaad951d9f7d16>.
- [18] State of California Department of Finance. *Demographic Projections*. 2017. URL: <http://www.dof.ca.gov/Forecasting/Demographics/Projections/>.
- [19] Supreme Court of Appeals for the Fifth District. *Fisher v. University of Texas*. 2016. URL: https://www.supremecourt.gov/opinions/15pdf/14-981_4g15.pdf.
- [20] Temchenko, Eugene. *Equal Protection*. LII / Legal Information Institute, 2016. URL: https://www.law.cornell.edu/wex/equal_protection.
- [21] United States Census Bureau. *Population Projections*. 2017. URL: <https://www.census.gov/programs-surveys/popproj.html>.
- [22] United States Census Bureau. *Quick Facts: Texas*. 2016. URL: <https://www.census.gov/quickfacts/table/RHI125215/48,00>.
- [23] United States Department of Justice. *Title VI of the Civil Rights Act of 1964 42 U.S.C. § 2000d Et Seq*. 2016. URL: <https://www.justice.gov/crt/fcs/TitleVI-Overview>.
- [24] University of California. *Undergraduate admissions summary*. 2017. URL: <https://www.universityofcalifornia.edu/infocenter/admissions-residency-and-ethnicity>.
- [25] University of California, Berkeley. *UC Berkeley Retention and Graduation Rates*. 2018. URL: <https://opa.berkeley.edu/uc-berkeley-retention-and-graduation-rates>.
- [26] University of California Board of Regents. *Regents Policy 4401: Policy on Future Admissions, Employment, and Contracting (Resolution Rescinding SP-1 and SP-2)*. 2001. URL: <http://regents.universityofcalifornia.edu/governance/policies/4401.html>.
- [27] University of Texas, Austin. *UT-Austin 2016-2017 Statistical Handbook*. 2016. URL: <https://utexas.app.box.com/v/SHB16-17Complete>.

Appendices

Appendix A

Table .1: Texas State Demographics (2016)

Race/Ethnicity	Percentage Makeup
African-American	0.126
Asian-American	0.048
Hispanic/Latino	0.391
White/Caucasian	0.426

Table .2: U.S. Demographics (2016)

Race/Ethnicity	Percentage Makeup
African-American	0.133
Asian-American	0.057
Hispanic/Latino	0.178
White/Caucasian	0.613

Table .3: UT-Austin In-State vs. Out-of-State Demographics (2016)

Residency Status	Percentage Makeup
Texas Residents	0.898
Out-of-State	0.049

Table .4: Actual UT-Austin Freshman
Fall 2016 Enrollment

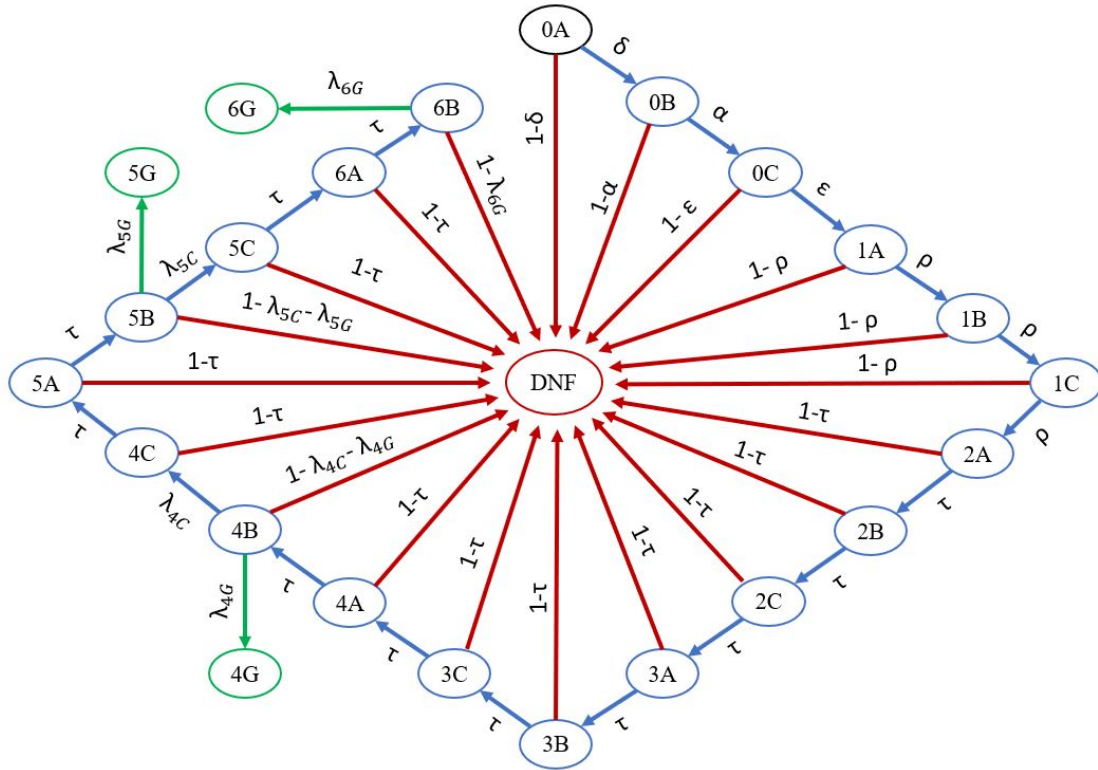
Race/Ethnicity	Percentage Makeup
African-American	0.042
Asian-American	0.207
Hispanic/Latino	0.226
White/Caucasian	0.425

Table .5: Predicted Critical Mass Freshman
Fall 2016 Enrollment

Race/Ethnicity	Percentage Makeup
African-American	0.120
Asian-American	0.046
Hispanic/Latino	0.360
White/Caucasian	0.413

Appendix B

Graphical Representation of the Markov Chain Model for College Progression



Appendix C

As mentioned in Section 4.5, we can find the necessary block matrices for our Markov chain model to put together the canonical form of \mathbf{P} .

C.1 Block Matrices \mathbf{Q} , $\mathbf{0}$, \mathbf{R} , and \mathbf{I}

In our Markov chain model, we have that \mathbf{Q} and $\mathbf{0}$ are

$$\mathbf{Q} = \begin{array}{c} \begin{array}{cccccccccccccccccccc} & 0A & 0B & 0C & 1A & 1B & 1C & 2A & 2B & 2C & 3A & 3B & 3C & 4A & 4B & 4C & 5A & 5B & 5C & 6A & 6B \end{array} \\ \begin{array}{l} 0A \\ 0B \\ 0C \\ 1A \\ 1B \\ 1C \\ 2A \\ 2B \\ 2C \\ 3A \\ 3B \\ 3C \\ 4A \\ 4B \\ 4C \\ 5A \\ 5B \\ 5C \\ 6A \\ 6B \end{array} \end{array} \begin{pmatrix} 0 & \delta & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \alpha & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \epsilon & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \rho & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \rho & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \rho & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \tau & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \tau & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \tau & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \tau & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \tau & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \tau & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_{4C} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \tau & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \tau & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda_{5C} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \tau & 0 \\ 0 & \tau \\ 0 & 0 \end{pmatrix}$$

and

$$\mathbf{0} = \begin{array}{c} \begin{array}{cccccccccccccccccccc} & 0A & 0B & 0C & 1A & 1B & 1C & 2A & 2B & 2C & 3A & 3B & 3C & 4A & 4B & 4C & 5A & 5B & 5C & 6A & 6B \end{array} \\ \begin{array}{l} 4G \\ 5G \\ 6G \\ DNF \end{array} \end{array} \begin{pmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$$

where the parameters δ , α , ϵ , ρ , τ , λ_{4C} , and λ_{5C} are as described in Chapter 4.

Additionally, we have that \mathbf{R} and \mathbf{I} are

$$\mathbf{R} = \begin{matrix} & \begin{matrix} 4G & 5G & 6G & DNF \end{matrix} \\ \begin{matrix} 0A \\ 0B \\ 0C \\ 1A \\ 1B \\ 1C \\ 2A \\ 2B \\ 2C \\ 3A \\ 3B \\ 3C \\ 4A \\ 4B \\ 4C \\ 5A \\ 5B \\ 5C \\ 6A \\ 6B \end{matrix} & \left(\begin{array}{cccc} 0 & 0 & 0 & 1 - \delta \\ 0 & 0 & 0 & 1 - \alpha \\ 0 & 0 & 0 & 1 - \epsilon \\ 0 & 0 & 0 & 1 - \rho \\ 0 & 0 & 0 & 1 - \rho \\ 0 & 0 & 0 & 1 - \rho \\ 0 & 0 & 0 & 1 - \tau \\ 0 & 0 & 0 & 1 - \tau \\ 0 & 0 & 0 & 1 - \tau \\ 0 & 0 & 0 & 1 - \tau \\ 0 & 0 & 0 & 1 - \tau \\ 0 & 0 & 0 & 1 - \tau \\ 0 & 0 & 0 & 1 - \tau \\ \lambda_{4G} & 0 & 0 & \lambda_{4B, DNF} \\ 0 & 0 & 0 & 1 - \tau \\ 0 & 0 & 0 & 1 - \tau \\ 0 & 0 & 0 & 1 - \tau \\ 0 & \lambda_{5G} & 0 & \lambda_{5B, DNF} \\ 0 & 0 & 0 & 1 - \tau \\ 0 & 0 & 0 & 1 - \tau \\ 0 & 0 & 0 & 1 - \tau \\ 0 & 0 & \lambda_{6G} & \lambda_{6B, DNF} \end{array} \right) \end{matrix}$$

$$\mathbf{I} = \begin{matrix} & \begin{matrix} 4G & 5G & 6G & DNF \end{matrix} \\ \begin{matrix} 4G \\ 5G \\ 6G \\ DNF \end{matrix} & \left(\begin{array}{cccc} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array} \right) \end{matrix}$$

where the parameters δ , α , ϵ , ρ , τ , λ_{4G} , $\lambda_{4B, DNF}$, λ_{5G} , $\lambda_{5B, DNF}$, λ_{6G} , and $\lambda_{6B, DNF}$ are as described in Chapter 4.

We can put all of these block matrices together to form our complete transition matrix \mathbf{P} into canonical form. This can will help us find the fundamental matrix \mathbf{N} .

C.2 The Matrix (I - Q), Useful in Finding the Fundamental Matrix

We need to find $\mathbf{I} - \mathbf{Q}$ as the first step to finding \mathbf{N} . We see that $\mathbf{I} - \mathbf{Q}$ is

$$\mathbf{I} - \mathbf{Q} = \begin{pmatrix} 1 & -\delta & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -\alpha & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -\epsilon & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -\rho & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & -\rho & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & -\rho & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & -\tau & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -\tau & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -\tau & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -\tau & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -\tau & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -\tau & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -\lambda_{4C} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -\tau & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -\tau & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -\lambda_{5C} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -\tau & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & -\tau \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

where the parameters $\delta, \alpha, \epsilon, \rho, \tau, \lambda_{4C}$, and λ_{5C} are as described in Chapter 4.

Appendix D

As mentioned in the Section 4.5, we can use elementary row operations to find the inverse of a given matrix. Prior to doing this for our matrix of interest, $\mathbf{I} - \mathbf{Q}$, we will walk through an example.

D.1 Finding the Inverse Using Elementary Row Operations

Let us say that we have the matrix \mathbf{A} , where

$$\mathbf{A} = \begin{pmatrix} 2 & 0 & 1 \\ 3 & 4 & 1 \\ 0 & 9 & 2 \end{pmatrix}$$

We can then create an augmented matrix of the form

$$\left(\mathbf{A} \mid \mathbf{I} \right) \tag{1}$$

Using elementary row operations, we can put \mathbf{A} into reduced row echelon form. All of these operations will also be executed on the Identity matrix augmented to \mathbf{A} . The result is that we will have a new augmented matrix consisting of \mathbf{I} and \mathbf{A}^{-1} as follows

$$\left(\mathbf{I} \mid \mathbf{A}^{-1} \right) \tag{2}$$

This is because performing all of the necessary row operations to reduce \mathbf{A} to \mathbf{I} acts in the same manner as multiplying both parts of (1) by \mathbf{A}^{-1} to get (2).

We will now go through this process to find \mathbf{A}^{-1} .

We can first use $R_1 = \frac{1}{2}R_1$, $R_2 = \frac{1}{3}R_2$, and $R_3 = \frac{1}{9}R_3$ to get

$$\left(\begin{array}{ccc|ccc} 2 & 0 & 1 & 1 & 0 & 0 \\ 3 & 4 & 1 & 0 & 1 & 0 \\ 0 & 9 & 2 & 0 & 0 & 1 \end{array} \right) \rightarrow \left(\begin{array}{ccc|ccc} 1 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 1 & \frac{4}{3} & \frac{1}{3} & 0 & \frac{1}{3} & 0 \\ 0 & 1 & \frac{2}{9} & 0 & 0 & \frac{1}{9} \end{array} \right)$$

Next we will take $R_2 = R_2 - R_1$ to get

$$\left(\begin{array}{ccc|ccc} 1 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 1 & \frac{4}{3} & \frac{1}{3} & 0 & \frac{1}{3} & 0 \\ 0 & 1 & \frac{2}{9} & 0 & 0 & \frac{1}{9} \end{array} \right) \rightarrow \left(\begin{array}{ccc|ccc} 1 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & \frac{4}{3} & -\frac{1}{6} & -\frac{1}{2} & \frac{1}{3} & 0 \\ 0 & 1 & \frac{2}{9} & 0 & 0 & \frac{1}{9} \end{array} \right)$$

Now we will use $R_2 = \frac{3}{4}R_2$ and $R_3 = R_3 - \frac{3}{4}R_2$ to get

$$\left(\begin{array}{ccc|ccc} 1 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & \frac{4}{3} & -\frac{1}{6} & -\frac{1}{2} & \frac{1}{3} & 0 \\ 0 & 1 & \frac{2}{9} & 0 & 0 & \frac{1}{9} \end{array} \right) \rightarrow \left(\begin{array}{ccc|ccc} 1 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 1 & -\frac{1}{8} & -\frac{3}{8} & \frac{1}{4} & 0 \\ 0 & 0 & \frac{25}{72} & \frac{3}{8} & -\frac{1}{4} & \frac{1}{9} \end{array} \right)$$

Lastly, we will take $R_3 = \frac{72}{25}R_3$, $R_1 = R_1 - \frac{72}{50}R_3$, and $R_2 = R_2 + \frac{72}{200}R_3$ to get

$$\left(\begin{array}{ccc|ccc} 1 & 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 1 & -\frac{1}{8} & -\frac{3}{8} & \frac{1}{4} & 0 \\ 0 & 0 & \frac{25}{72} & \frac{3}{8} & -\frac{1}{4} & \frac{1}{9} \end{array} \right) \rightarrow \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & -\frac{1}{25} & \frac{9}{25} & -\frac{4}{25} \\ 0 & 1 & 0 & -\frac{6}{25} & \frac{4}{25} & \frac{1}{25} \\ 0 & 0 & 1 & \frac{27}{25} & -\frac{18}{25} & \frac{8}{25} \end{array} \right)$$

This gives us that if

$$\mathbf{A} = \begin{pmatrix} 2 & 0 & 1 \\ 3 & 4 & 1 \\ 0 & 9 & 2 \end{pmatrix} \implies \mathbf{A}^{-1} = \begin{pmatrix} -\frac{1}{25} & \frac{9}{25} & -\frac{4}{25} \\ -\frac{6}{25} & \frac{4}{25} & \frac{1}{25} \\ \frac{27}{25} & -\frac{18}{25} & \frac{8}{25} \end{pmatrix}$$

One could check by hand that this is in fact the matrix that satisfies $\mathbf{I} = \mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A}$.

D.2 Finding \mathbf{N} Using Elementary Row Operations

Now that we have seen an example of how this works, we will use this method to find the inverse of $\mathbf{I} - \mathbf{Q}$ —whose elements are shown explicitly in Appendix C2—in order to find the fundamental matrix \mathbf{N} . Most of this process has already been completed since $\mathbf{I} - \mathbf{Q}$ is already in the form where all the elements of the main diagonal are 1s and we only have non-zero elements on the off-diagonal immediately above the main diagonal. We can see that the last few rows of $\mathbf{I} - \mathbf{Q}$ are

$$\begin{array}{c} \dots & R_{18} & R_{19} & R_{20} \\ \vdots & \left(\begin{array}{ccc|ccc} \ddots & \vdots & \vdots & \vdots & & \\ \dots & 1 & -\tau & 0 & & \\ R_{18} & \dots & 0 & 1 & -\tau & \\ R_{19} & \dots & 0 & 0 & 1 & \\ R_{20} & \dots & 0 & 0 & 0 & 1 \end{array} \right) \end{array}$$

We can then take $R_{19} = R_{19} + \tau R_{20}$ to get

$$\left(\begin{array}{ccc|ccc} \ddots & \vdots & \vdots & \vdots & & \\ \dots & 1 & -\tau & 0 & & \\ \dots & 0 & 1 & -\tau & & \\ \dots & 0 & 0 & 1 & & \end{array} \middle| \begin{array}{ccc|ccc} \ddots & \vdots & \vdots & \vdots & & \\ \dots & 1 & 0 & 0 & & \\ \dots & 0 & 1 & 0 & & \\ \dots & 0 & 0 & 1 & & \end{array} \right) \rightarrow \left(\begin{array}{ccc|ccc} \ddots & \vdots & \vdots & \vdots & & \\ \dots & 1 & -\tau & 0 & & \\ \dots & 0 & 1 & 0 & & \\ \dots & 0 & 0 & 1 & & \end{array} \middle| \begin{array}{ccc|ccc} \ddots & \vdots & \vdots & \vdots & & \\ \dots & 1 & 0 & 0 & & \\ \dots & 0 & 1 & \tau & & \\ \dots & 0 & 0 & 1 & & \end{array} \right) \quad (3)$$

The we will use $R_{18} = R_{18} + \tau R_{19}$ to get

$$\left(\begin{array}{ccc|ccc} \ddots & \vdots & \vdots & \vdots & & \\ \dots & 1 & -\tau & 0 & & \\ \dots & 0 & 1 & 0 & & \\ \dots & 0 & 0 & 1 & & \end{array} \middle| \begin{array}{ccc|ccc} \ddots & \vdots & \vdots & \vdots & & \\ \dots & 1 & 0 & 0 & & \\ \dots & 0 & 1 & \tau & & \\ \dots & 0 & 0 & 1 & & \end{array} \right) \rightarrow \left(\begin{array}{ccc|ccc} \ddots & \vdots & \vdots & \vdots & & \\ \dots & 1 & 0 & 0 & & \\ \dots & 0 & 1 & \tau & & \\ \dots & 0 & 0 & 1 & & \end{array} \middle| \begin{array}{ccc|ccc} \ddots & \vdots & \vdots & \vdots & & \\ \dots & 1 & \tau & \tau^2 & & \\ \dots & 0 & 1 & \tau & & \\ \dots & 0 & 0 & 1 & & \end{array} \right) \quad (4)$$

As we can see, this will eventually give us that \mathbf{N} is upper triangular and looks like

$$\mathbf{N} = \begin{pmatrix} 1 & n_{1,2} & n_{1,3} & \dots & n_{1,18} & n_{1,19} & n_{1,20} \\ 0 & 1 & n_{2,3} & \dots & n_{2,18} & n_{2,19} & n_{2,20} \\ 0 & 0 & 1 & \dots & n_{3,18} & n_{3,19} & n_{3,20} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & n_{18,19} & n_{18,20} \\ 0 & 0 & 0 & \dots & 0 & 1 & n_{19,20} \\ 0 & 0 & 0 & \dots & 0 & 0 & 1 \end{pmatrix}$$

So far, we have found that $n_{18,19} = \tau$, $n_{18,20} = \tau^2$, and $n_{19,20} = \tau$. We can continue with a similar process we used in (3) and (4) to rows $R_{17}, R_{16}, \dots, R_2$, and then R_1 in order to find the remaining entries in \mathbf{N} .

D.3 Explicit Formula for the Elements of \mathbf{N}

We can see that these row operations are directly related to the elements we have on the off-diagonal in $\mathbf{I} - \mathbf{Q}$. We will save a vector of these elements as

$$\mathbf{v}_{OD} = - \left[\delta \quad \alpha \quad \epsilon \quad \rho \quad \rho \quad \rho \quad \tau \quad \tau \quad \tau \quad \tau \quad \tau \quad \tau \quad \tau \quad \tau \quad \lambda_{4C} \quad \tau \quad \tau \quad \lambda_{5C} \quad \tau \quad \tau \right]$$

Since we will only be adding equations to each other in the process of finding \mathbf{N} , we ignore the negative sign and can reduce \mathbf{v}_{OD} simply to \mathbf{v} , where

$$\mathbf{v} = \left[\delta \quad \alpha \quad \epsilon \quad \rho \quad \rho \quad \rho \quad \tau \quad \tau \quad \tau \quad \tau \quad \tau \quad \tau \quad \tau \quad \tau \quad \lambda_{4C} \quad \tau \quad \tau \quad \lambda_{5C} \quad \tau \quad \tau \right]$$

We can see that \mathbf{v} has elements we can denote as \mathbf{v}_k for $k = 1, 2, \dots, 19$. With this in mind, we can now try to relate a formula for the i, j th element of \mathbf{N} to the elements of this vector \mathbf{v} . This formula is as follows

$$n_{i,j} = \prod_{k=i}^{j-1} \mathbf{v}_k, \text{ for } 1 \leq i \leq 19, j > i \quad (5)$$

Using this formula, it is now trivial to find an element of \mathbf{N} analytically.

As an example, we can see that $n_{5,16} = \prod_{k=5}^{15} \mathbf{v}_k = \rho^2 \times \tau^7 \times \lambda_{4C} \times \tau = \rho^2 \tau^8 \lambda_{4C}$

D.4 Explicit Formula for the Elements of \mathbf{t}

Now that we have this formula for the elements of \mathbf{N} , we can do the same for \mathbf{t} , where

$$\mathbf{t} = \left(t_{0A} \quad t_{0B} \quad t_{0C} \quad t_{1A} \quad t_{1B} \quad t_{1C} \quad \dots \quad t_{5A} \quad t_{5B} \quad t_{5C} \quad t_{6A} \quad t_{6B} \right)^T$$

We know that $\mathbf{t} = \mathbf{N}\mathbf{c}$, where \mathbf{c} is a vector of 1s. This means that we will be summing the i th row of \mathbf{N} to get the i th element of \mathbf{t} . We can then find the following formula

$$\mathbf{t}_m = 1 + \sum_{M=m}^{19} \left(\prod_{k=m}^M \mathbf{v}_k \right) \quad (6)$$

$$\begin{aligned} \text{As an example we will find } \mathbf{t}_{11} = \mathbf{t}_{3B} &= 1 + \sum_{M=11}^{19} \left(\prod_{k=11}^M \mathbf{v}_k \right) = 1 + \prod_{k=11}^{11} \mathbf{v}_k + \prod_{k=11}^{12} \mathbf{v}_k + \cdots + \prod_{k=11}^{19} \mathbf{v}_k \\ \implies \mathbf{t}_{11} = \mathbf{t}_{3B} &= 1 + \tau + \tau^2 + \tau^3 + \tau^3 \lambda_{4C} + \tau^4 \lambda_{4C} + \tau^5 \lambda_{4C} + \tau^5 \lambda_{4C} \lambda_{5C} + \tau^6 \lambda_{4C} \lambda_{5C} + \tau^7 \lambda_{4C} \lambda_{5C} \end{aligned}$$

D.5 Explicit Formula for the Elements of \mathbf{B}

Lastly, we can also use the formula for $n_{i,j}$, we can find a formula for the elements of \mathbf{B} . We know that \mathbf{N} is a 20-by-20 matrix and \mathbf{R} is a 20-by-4 matrix, which means that we will get \mathbf{B} , a 20-by-4 matrix. We know that by matrix multiplication, we will have have that

$$b_{x,y} = n_{x,1}r_{1,y} + n_{x,2}r_{2,y} + \cdots + n_{x,19}r_{19,y} + n_{x,20}r_{20,y} = \sum_{l=1}^{20} n_{x,l}r_{l,y} \quad (7)$$

Additionally, we can see that $n_{i,i} = 1$, the first column of \mathbf{R} is all 0s except for $r_{14,1} = \lambda_{4G}$, the second column of \mathbf{R} is all 0s except for $r_{17,2} = \lambda_{5G}$, the third column of \mathbf{R} is all 0s except for $r_{20,3} = \lambda_{6G}$, and since \mathbf{B} needs to be a right stochastic matrix, we see that the fourth element of any row is just $1 - \sum_{y=1}^3 b_{x,y}$. Due to these special properties, we can separate this formula based on cases related to the entries in both matrices.

Case I: $x < 14$ and $y = 1$

$$b_{x,y} = b_{x,1} = \sum_{l=1}^{20} n_{x,l}r_{l,1} = n_{x,14}r_{14,1} = \lambda_{4G} \prod_{k=x}^{13} \mathbf{v}_k$$

Case II: $x = 14$ and $y = 1$

$$b_{x,y} = b_{14,1} = \sum_{l=1}^{20} n_{14,l}r_{l,1} = n_{14,14}r_{14,1} = \lambda_{4G}$$

Case III: $x > 14$ and $y = 1$

$$b_{x,y} = b_{x,1} = \sum_{l=1}^{20} n_{x,l}r_{l,1} = n_{x,14}r_{14,1} = 0 \text{ since } n_{x,14} = 0 \text{ for } x > 14$$

Case IV: $x < 17$ and $y = 2$

$$b_{x,y} = b_{x,2} = \sum_{l=1}^{20} n_{x,l}r_{l,2} = n_{x,17}r_{17,2} = \lambda_{5G} \prod_{k=x}^{16} \mathbf{v}_k$$

Case V: $x = 17$ and $y = 2$

$$b_{x,y} = b_{17,2} = \sum_{l=1}^{20} n_{17,l}r_{l,2} = n_{17,17}r_{17,2} = \lambda_{5G}$$

Case VI: $x > 17$ and $y = 2$

$$b_{x,y} = b_{x,2} = \sum_{l=1}^{20} n_{x,l}r_{l,2} = n_{x,17}r_{17,2} = 0 \text{ since } n_{x,17} = 0 \text{ for } x > 17$$

Case VII: $x < 20$ and $y = 3$

$$b_{x,y} = b_{x,3} = \sum_{l=1}^{20} n_{x,l}r_{l,3} = n_{x,20}r_{20,3} = \lambda_{6G} \prod_{k=x}^{19} \mathbf{v}_k$$

Case VIII: $x = 20$ and $y = 3$

$$b_{x,y} = b_{20,3} = \sum_{l=1}^{20} n_{20,l}r_{l,3} = n_{20,20}r_{20,3} = \lambda_{6G}$$

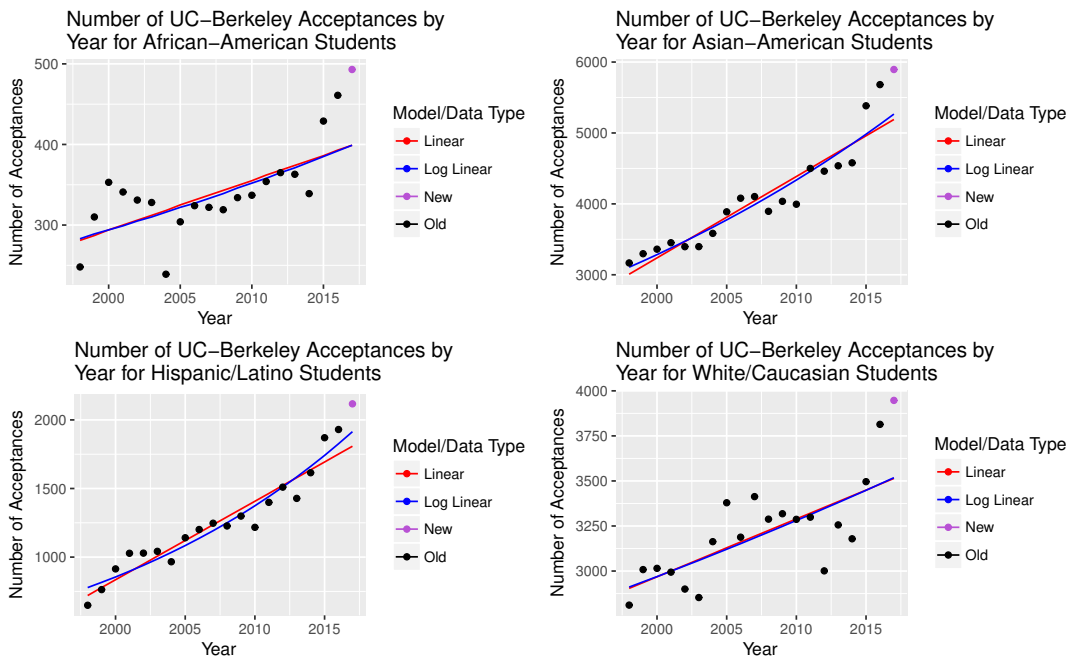
Case IX: $1 \leq x \leq 20$ and $y = 4$

$$b_{x,y} = b_{x,4} = \sum_{l=1}^{20} n_{x,l}r_{l,4} = 1 - b_{x,1} - b_{x,2} - b_{x,3} \text{ where } b_{x,1}, b_{x,2}, b_{x,3} \text{ are described as above}$$

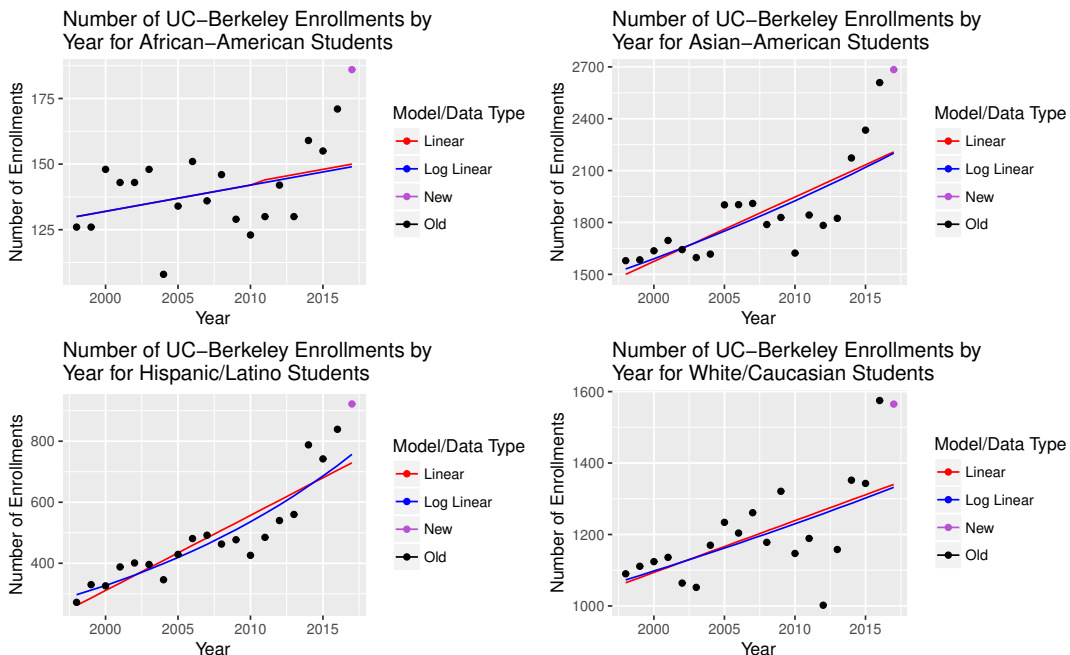
Appendix E

Referenced plots and results from Chapter 5.

E1 Regression Comparison Plots for Number of Acceptances to UC-Berkeley



E2 Regression Comparison Plots for Number of Enrolled Students at UC-Berkeley



E3 Markov Chain Model Predictions

	year.2018	year.2019	year.2020	year.2021	year.2022	year.2023	year.2024	year.2025	year.2026	year.2027
African-American	0.03338	0.03301	0.03273	0.03227	0.03184	0.03129	0.03086	0.03041	0.02997	0.02943
Asian-American	0.50027	0.49698	0.49424	0.49264	0.49112	0.48967	0.48796	0.48623	0.48440	0.48252
Hispanic/Latino	0.16755	0.17152	0.17755	0.18265	0.18783	0.19298	0.19823	0.20362	0.20909	0.21476
white/Caucasian	0.29880	0.29849	0.29548	0.29244	0.28921	0.28607	0.28295	0.27974	0.27654	0.27329

E4 Means from 10000 Iterations of Stochastic Ideal Critical Mass Process

	year.2018	year.2019	year.2020	year.2021	year.2022	year.2023	year.2024	year.2025	year.2026	year.2027
African-American	0.06790	0.06780	0.06770	0.06759	0.06754	0.06745	0.06738	0.06732	0.06726	0.06721
Asian-American	0.13416	0.13479	0.13546	0.13610	0.13665	0.13723	0.13776	0.13824	0.13874	0.13917
Hispanic/Latino	0.36948	0.37192	0.37421	0.37646	0.37846	0.38057	0.38259	0.38453	0.38647	0.38834
white/Caucasian	0.42846	0.42549	0.42264	0.41985	0.41735	0.41475	0.41226	0.40990	0.40754	0.40528

E5 Standard Deviations from 10000 Iterations of Stochastic Ideal Critical Mass Process

	year.2018	year.2019	year.2020	year.2021	year.2022	year.2023	year.2024	year.2025	year.2026	year.2027
African-American	0.00094	0.00092	0.00092	0.00089	0.00090	0.00088	0.00087	0.00085	0.00085	0.00086
Asian-American	0.00113	0.00109	0.00109	0.00104	0.00104	0.00100	0.00097	0.00095	0.00093	0.00093
Hispanic/Latino	0.00286	0.00279	0.00278	0.00266	0.00268	0.00259	0.00253	0.00247	0.00243	0.00244
white/Caucasian	0.00305	0.00297	0.00294	0.00281	0.00282	0.00271	0.00263	0.00256	0.00251	0.00251

E6 Critical Mass Assessment Criterion Results

Race/Ethnicity	Year									
	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027
African-American	49	49	48	48	47	46	46	45	45	44
Asian-American	373	369	365	362	359	357	354	352	349	347
Hispanic/Latino	45	46	47	49	50	51	52	53	54	55
White/Caucasian	70	70	70	70	69	69	69	68	68	67