Differences By Race and Gender in Persistence Rates After Taking Introductory Classes in Economics and Computer Science

Sakthi Ponnuswamy*

University of California, Berkeley

May, 2019

1 Introduction

Are there systematic differences by race and gender in terms of grades received in introductory classes and persistence rates after completing introductory classes? Why do differences in persistence rates after completing introductory classes exist?

In the United States, there is a disproportionately small number of female and underrepresented minority (URM) students graduating with STEM majors (Astorne-Figari, 2017). There have been numerous studies conducted to understand why the gender and racial gap in major choice exists. Reasons include: male and female students valuing factors like social status, wages, and approval of parents differently (Zafar, 2005), URM students feeling more pressure to succeed in STEM classes (Porter, 2006), and

*I am grateful to my advisor, Martha Olney, for her continuous support, guidance, encouragement, and integral feedback. Thank you to Zachary Bleemer for piquing my interest in studying student outcomes and mentorship for the past three years. All errors are my own.

STEM classes having non-conducive learning environments for female and URM students (Shapiro, 2011). In general, undergraduate students' major choice is highly influenced by their perceived ability in a certain major, which is often informed by grades in their introductory classes (Astorne-Figari, 2017). If certain demographic groups are more likely to get lower grades it can quickly lead to a racial and gender imbalance in certain departments.

This paper attempts to frame the major choice process in a way that pinpoints when and potential reasons for why students are discouraged from pursuing their major by asking the following questions: (1) Are there systematic differences by race and gender in the grades received in introductory classes? (2) Are there systematic differences by race and gender in persistence after completing the introductory class? (3) Where does the biggest drop-off in persistence take place? (4) Can some of the systematic differences by race and gender be explained by section demographics? (5) What factors eventually account for the number of classes a student takes in a department? This paper finds that there are systematic differences by race and gender in persistence rates after taking introductory classes after controlling for the grade that a student receives in an introductory class. This paper hypothesizes that the demographic breakdown in smaller learning environments play a pivotal part in whether a student decides to persist in their respective department, but there is a myriad of other explanations for why the systematic difference in persistence rates exist.

The data studied comes from University of California, Berkeley students who graduated between 2011-2016 and who took introductory classes in Economics and Computer Science for a letter grade. The Economics and Computer Science departments will serve as case studies for two reasons. First, in order to major in Economics or Computer Science, students must receive a GPA of 3.0 in certain prerequisite classes. This allows us to observe how influential grades are in a setting where not everyone is able to declare a major. Second, both introductory classes in Computer Science and Economics have large lectures with 500-1000 students, which means that a lot of the application of material happens in smaller section or lab settings. This allows us to better understand how a student's section, which is where students spend most of their time actively engaging with the material, can influence persistence rates.

Completion of the analysis indicates that there are systematic differences by race and gender in grades received in introductory classes and in persistence rates; however, the magnitude and significance of these results vary based on whether the student is in the Economics or Computer Science department. While persistence rates are influenced by grade received in introductory classes in both Economics and Computer Science, in the Economics department there is data that suggests female students are more responsive to grades than males. In trying to understand why these systematic differences exist, section demographic data is looked at. Students that have peers that are similar in race and gender are more likely to persist after taking their first introductory class in Economics; this effect is especially seen with White, female students. The section demographic breakdown does not have as large of an effect in the Computer Science department. Lastly, gender, race, and grade in introductory class affect how many classes students take in both the Computer Science and Economics department.

2 Literature Review

Modeling a student's major choice is quite difficult since the decision process is dynamic and includes many factors that are often times unobservable. Earlier research relied on collecting qualitative data via focus groups and surveys (Zafar, 2010; Porter, 2006; Arcidiacano, 2010; Staniec, 2004; Hastings, 2015). An overwhelming theme from this set of literature is that a student's perceived ability, often informed by early grades, was a data point that students used to decide whether or not to pursue a major. Furthermore, URM and female students continuously felt as if they needed to hold themselves to a higher standard in comparison to their peers to pursue a more quantitative major (Porter, 2006). This body of research informed the idea of testing how influential grades in introductory classes were for different demographic groups at UC Berkeley. If there are differential responses to grades, then that could explain some of the gender and racial imbalance seen in STEM or other competitive majors.

While grades are one facet of a student's decision making process, the research also indicates other reasons for switching majors including: negative interactions with teaching assistants and professors, ineffective high school preparation, decreased sense of safety in the department of choice, lack of intellectual identity in STEM major, and a negative learning environment (Mullen, 2013; Shapiro, 2011). While some of these insights could explain some of the differences in persistence rates, many of these components cannot be rigorously tested in a sample size of 7000. For example, it would be impossible to aggregate all student-teacher interactions for all the students in the dataset. However, negative learning environment was a common theme that emerged in the research as one of the reasons students switched away from STEM majors. This led to the study of the section demographic breakdown to better understand if similar peers could contribute to a more positive learning environment and thereby affect persistence rates and who pursues a certain major.

There are three studies that investigate the same topic as this paper in a similar manner. Admission and graduation data for three public universities in Texas were pooled together to understand if certain students (male vs. female, non-URM vs. URM) were more likely to intend to major in certain subjects before starting undergraduate college (Dickson, 2010). A student was categorized as switching a major if their intended major was different than their declared major. After looking at switch rates, Dickson concludes that URM and female students were more likely to switch away from STEM majors. A comparable, but more detailed study looks at data from the National Longitudinal Survey of Youth 1997 (Astorne-Figari, 2017). For four years, they collected student's GPA and their major or intended major at the time to build a panel data set of about 9000 students. When a student reported a different major than their intended major, it was coded as a switch and the time and GPA of that year was noted. Astorne-Figari outlines that female and URM students are more likely to switch away from their intended major – this was especially seen if they intended to pursue a STEM major. Semesters in which students receive low grades is an accurate predictor for a student switching their major the following semester; in fact, the lower the grade, the "larger" the switch away from a student's current major. Both of these datasets are not as recent as the data collected at UC Berkeley. Additionally, the data did not have specific grades for the classes students took, so there was no way to attribute the switch in majors to grades in introductory or intermediate classes, which is what this paper aims to do.

The most similar study focuses on a fictional college, Adams College. The study looked at students who took introductory economics, their grades in the course, and how many of them declared the Economics major afterwards (Goldin, 2015). Goldin concludes that female students are more responsive to grade increases and decreases than their male counterparts in Economics. For example, if a female and male student both get a B in introductory Economics, the female student is more likely to not complete the major. The methodology in this paper serves as the foundation for what this paper aims to do with the data collected at UC Berkeley. This paper differs since it also looks at students in the Computer Science Department to understand if the same effects are seen at a larger magnitude since Computer Science is strictly considered a STEM major. Additionally, UC Berkeley has a unique setting since there is a prerequisite GPA that students must achieve, colloquially called the GPA cap, for declaring the Economics and Computer Science major. Female and URM students feel more pressure to do better in their classes, so it will be interesting to note if the external pressure to meet a GPA cap increases the magnitude of differential grade responses (Shapiro, 2011). Additionally, the analysis done on peer effects in section settings is an extension of the current literature and aims to explain some of the differences in persistence rates.

3 Data Summary and Summary Statistics

The initial dataset comes from University of California - Office of the President. The anonymized data is from 2011-2016 and includes: student's race and gender, incoming year, graduation year, every course a student has taken, grades in respective classes, section number, intended majors, and declared majors. Since this paper focuses on the Computer Science and Economics department, only students who were enrolled in Computer Science 61A (introductory level Computer Science class) and Economics 1 or 2 (introductory level Economics classes seen as equivalent) were included in the initial dataset. This immediately took out any transfer students, since as juniors they only have to take intermediate classes to declare their major. In order to understand if there were differences in grade by race and gender only students who took the class for a letter grade were kept in the dataset. Some students decided not to report their race or gender and were excluded from the dataset. Additionally, there was a small number of Pacific-Islander and Native Americans in the dataset. Since they were less than 5% of the population, they were excluded since the results would be noisy. Therefore, the racial groups that are studied include: White, Black, and URM (Hispanic and Black) students. Lastly, international and domestic students have cultural differences as well as different motivations when choosing their major. Therefore, international students were excluded from the dataset.

The following table outlines the initial data set and how many students were left after every stage of filtering.[1]

[1]Electrical Engineering and Computer Science (EECS) students were omitted from the initial set of students who took Computer Science 61A. EECS students are already admitted into the School of Engineering and do not have to reach a prerequisite GPA to declare their major. There are inherent differences between Computer Science and EECS students, which include: different levels of pressure to reach a GPA minimum, resources available (counselors, reserved seating for classes), etc. Therefore, it is unrealistic to group them together.

Table 1: Filtering of Data Set

Filter	No. of Economics Students	No. of CS Students
Enrolled	8333	9395
Grade	5426	5157
Race	4403	4431
Gender	4402	4431
International	3382	3819

4 Grade Differences by Race and Gender

The first set of regressions are analyzed in order to understand if there are systematic differences by race and gender in the grades that students receive in introductory classes. Research shows that a leading reason for how undergraduates choose their major is based on perceived ability. Grades in early classes are a signal to students and is a way to predict whether or not a student will switch majors (Figari, 2017). Therefore, the following model was used to understand if and how large the differences are by race and gender.

Model 1: Grade by Race and Gender

$$Grade_i = \beta_0 + \beta_1 Gender_i + \beta_2 Race_i + \beta_3 Race_i * Gender_i + \epsilon_i \tag{1}$$

GPA is noted as the grade that students received in their introductory classes. Gender is a binary variable and Race is a categorical variable.

Table 2 and 6 show the output of the regression using the model described above. The constant represents the Asian, Male student.[2] The summary statistics below (Economics

[2]The Asian, Male student was chosen as the base variable for two reasons. First, after running the regressions with Male-White, Female -White, Male - Asian, Female as the base, the results stayed the same. Therefore, there was no clear reason to choose one over the other. Second, Asian, Male students have the highest rate of persistence in both Computer Science and Economics. Assuming that a majority of the students are pushed out for external reasons and not by choice, both departments should want high persistence rates. This assumption is not valid in all cases, but switching away from a major for personal reasons cannot be observed in the dataset. This paper does not claim that the Asian, Male is

Tables 2-4 and Computer Science Table 6-8) indicate the average GPA of each demographic subgroup and whether it is significant in comparison to the different base groups. For example, in Economics a female student receives an average GPA of 2.99 which is not significantly different than the male student's GPA of 2.97.[3]

Table 2: Grade Difference by Gender in Introductory Economics

Group	Average GPA	* Male
Female	2.99	
Male	2.97	

Table 3: Grade Difference by Race in Introductory Economics

Group	Average GPA	* Asian	* White
Asian	2.99		Yes
URM	2.97	Yes	Yes
White	2.99	Yes	

Table 4: Grade Difference by Race and Gender in Introductory Economics

Group	Average GPA	* Asian, M	* White, M	* URM, M	*White, F	*URM, F
Asian, M	3.10		Yes	Yes	Yes	Yes
Asian, F	3.10		Yes	Yes	Yes	Yes
URM, M	2.53	Yes	Yes		Yes	
URM, F	2.50	Yes	Yes		Yes	
White, M	2.97	Yes		Yes		Yes
White, F	3.02	Yes		Yes		Yes

the ideal student. The summary statistics can be looked at to see how changing the base altered the significance of the results.

[3]An f-test was conducted to see if the sum of coefficient was statistically significant. A result was marked as statistically significant if the p-value was less than or equal to .1.

Table 5: Econ 1 and Econ 2 Grades by Race and Gender

	Dependent variable:		
	Grade in Econ 1 or Econ 2		
	(1)	(2)	(3)
Female	0.013		0.002
	(0.029)		(0.037)
URM:Female			-0.094
			(0.082)
White:Female			0.046
			(0.065)
URM		-0.614***	-0.567***
		(0.041)	(0.058)
White		-0.110***	-0.128***
		(0.032)	(0.043)
Constant	2.975***	3.101***	3.100***
	(0.020)	(0.018)	(0.027)
Observations	3,382	3,382	3,382
R^2	0.0001	0.063	0.064
Adjusted R^2	-0.0002	0.062	0.062
Residual Std. Error	0.834	0.807	0.807
F Statistic	0.214	113.306***	45.791***
Note:	*p<0.1; **p<0.05; ***p<0.01		

Table 6: Computer Science Grade by Gender

Group	Average GPA	* Male
Female	2.85	Yes
Male	3.11	

Table 7: Computer Grade by Race

Group	Average GPA	* Asian	* White
Asian	3.05		Yes
URM	2.59	Yes	Yes
White	3.12		

Table 8: Computer Science Grade by Race and Gender

Group	Average GPA	* Asian, M	* White, M	* URM, M	*White, F	*URM, F
Asian, M	3.13		Yes	Yes	Yes	Yes
Asian, F	2.89	Yes		Yes	Yes	Yes
URM, M	2.70	Yes	Yes		Yes	Yes
URM, F	2.64	Yes	Yes	Yes	Yes	
White, M	3.18			Yes	Yes	Yes
White, F	2.94	Yes		Yes		Yes

Since there is no statistical difference between the grade male and female students

Table 9: CS61A Grades by Race and Gender

	Dependent variable:		
	Grade in CS61A		
	(1)	(2)	(3)
Female	-0.261***		-0.240***
	(0.044)		(0.052)
URM:Female			-0.064
			(0.163)
White:Female			-0.025
			(0.112)
URM		-0.455***	-0.430***
		(0.079)	(0.098)
White		0.074	0.052
		(0.049)	(0.057)
Constant	3.113***	3.045***	3.130***
	(0.025)	(0.025)	(0.031)
Observations	1,564	1,564	1,564
R^2	0.022	0.024	0.044
Adjusted R^2	0.021	0.023	0.041
Residual Std. Error	0.820	0.819	0.812
F Statistic	34.669***	19.536***	14.471***
Note:			*p<0.1; **p<0.05; ***p<0.01

receive in Introductory Economics, there is no gender gap. On the contrary, there is a gender gap in the grades received by students in the Computer Science department. A female student receives .261 (se = 0.04) GPA points lower than their male counterparts. There are differences by race in both the Economics and Computer Science department. Looking at column 2 of Table 5, URM students receive .614 (se = 0.04) GPA points lower and White students receive .110 (se = 0.02) GPA points lower than their Asian counterparts. Even after changing the base, the racial differences are still significant. In Computer Science, the difference between White and Asian students is insignificant. Looking at column 2 of Table 9, URM students receive .455 (se = 0.08) GPA points lower than their Asian counterparts. The largest difference in GPA points is seen when comparing URM, female students in both Economics and Computer Science to their Asian, male student counterparts. URM, female students receive roughly a half GPA point lower than their Asian, male student counterparts in both department. This difference is equivalent from to going from an A to an A-/B+.

5 Persistence Rates Difference by Race and Gender

Knowing that grades are the most common and first signal for younger undergraduate students, it is important to see if the systematic differences seen in grades received carry over to persistence rates and thereby influencing who declares the major. The following tables indicate the raw number of students in each demographic group enrolled in each class, the percentage share each group made up of the class, and the percentage of students who persisted in comparison to their introductory classes.[4]

Table 10: Demographic Breakdown of 4-Year Domestic Students in Each Economics Core Class From 2011-2016

Race	Gender	Intro Econ	Micro or Macro	Micro and Macro	Metrics	Completion Rate
Asian	Male	874	468	283	153	17.5
Asian	Female	1074	543	349	185	17.2
URM	Male	244	86	53	17	7.0
URM	Female	245	66	34	16	6.5
White	Male	569	265	160	92	16.2
White	Female	376	149	90	48	12.8

Table 11: Percentage Share of 4-Year Domestic Students in Each Economics Core Class by Race and Gender From 2011-2016

Race	Gender	Intro Econ	Micro or Macro	Micro and Macro	Metrics
Asian	Male	26	30	29	30
Asian	Female	32	34	36	36
URM	Male	7	5	5	3
URM	Female	7	4	4	3
White	Male	17	17	17	8
White	Female	11	9	9	9

[4] In Economics, all students must take their classes in the following order: Introductory Economics, Intermediate Economics, and Econometrics. In Computer Science the order is as followed: CS61A, CS61B, CS61C. Intended Computer Science majors must also take Computer Science 70, which can be taken at any time or simultaneously with another class. Therefore, the tables for Computer Science are coded as second, third, or fourth class since the order varies from student to student.

Table 12: Percentage of 4-Year Domestic Students who Persist in Relation to Introductory Economics by Race and Gender From 2011-2016

Race	Gender	Intro Econ	Micro or Macro	Micro and Macro	Metrics
Asian	Male	100	54	32	18
Asian	Female	100	51	32	17
URM	Male	100	35	22	7
URM	Female	100	27	14	7
White	Male	100	47	28	16
White	Female	100	40	24	13

Table 13: Demographic Breakdown of 4-Year Domestic Students in Each Computer Science Core Class From 2011-2016

Race	Gender	CS61A	Second	Third	Fourth	Completion Rate
Asian	Male	1929	1696	1477	1075	55.7
Asian	Female	757	573	462	330	43.5
URM	Male	196	151	115	65	33.0
URM	Female	90	42	26	15	16.6
White	Male	665	525	432	310	46.6
White	Female	182	109	87	50	27.4

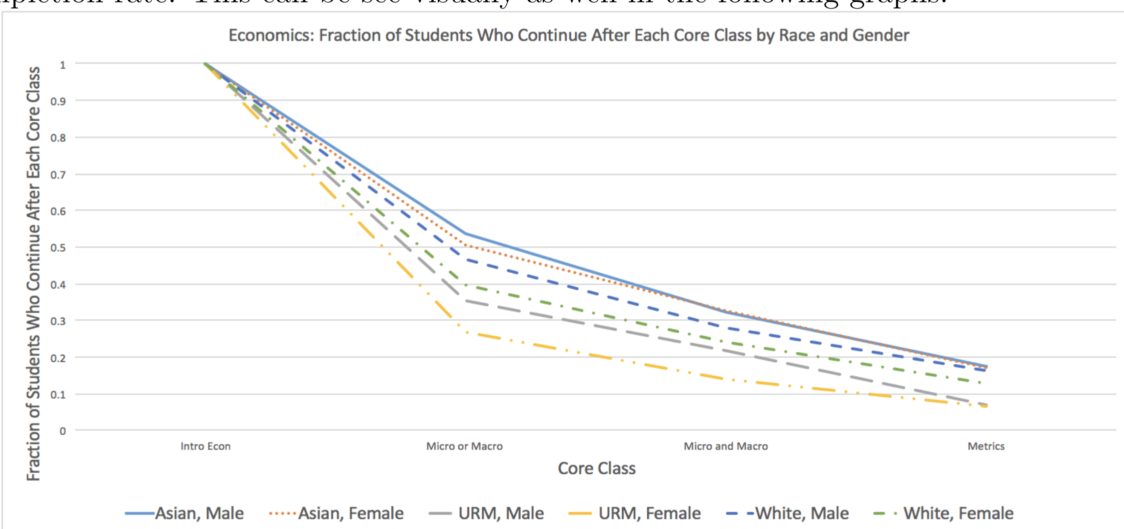
Table 14: Percentage Share of 4-Year Domestic Students in Each Computer Science Core Class by Race and Gender From 2011-2016

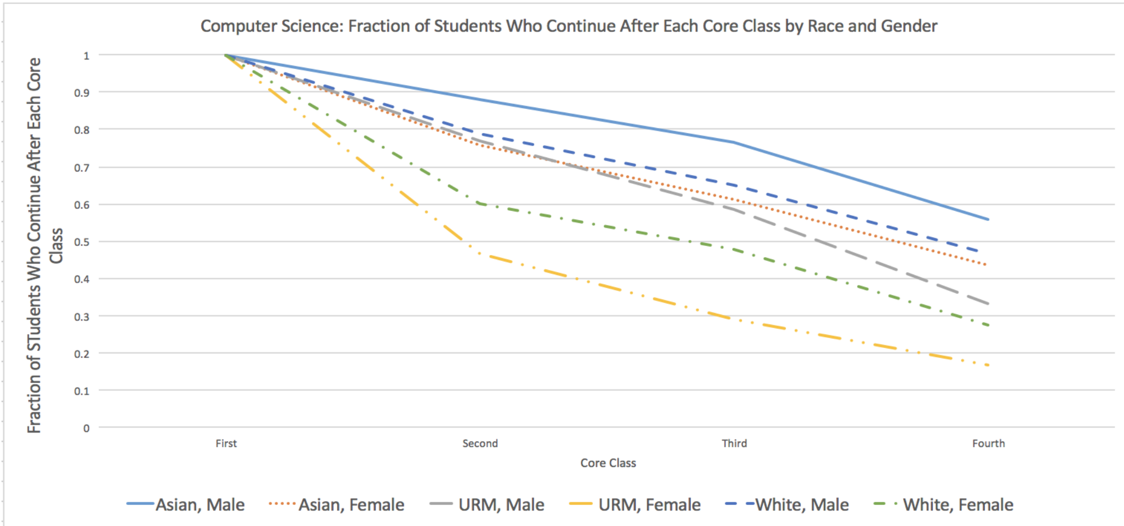
Race	Gender	CS61A	Second	Third	Fourth
Asian	Male	51	55	57	58
Asian	Female	20	19	18	18
URM	Male	5	5	4	4
URM	Female	2	1	1	1
White	Male	17	17	17	17
White	Female	5	4	3	3

Table 15: Percentage of 4-Year Domestic Students who Persist in Relation to CS61A by Race and Gender From 2011-2016

Race	Gender	CS61A	Second	Third	Fourth
Asian	Male	100	88	77	58
Asian	Female	100	76	61	18
URM	Male	100	77	59	4
URM	Female	100	47	29	1
White	Male	100	79	65	17
White	Female	100	60	48	3

In economics, URM students make up the smallest percentage of students in Introductory Economics and have the lowest completion rate. In Computer Science, URM and White female students make up the smallest percentage of the class and have the lowest completion rate. This can be see visually as well in the following graphs.





Economics has a much steeper drop-off between Introductory Economics and the first intermediate economics class in comparison to CS61A and the second class in Computer Science. This can be explained by the fact that many departments have Economics 1 or 2 as a prerequisite to declaring the majors; examples include: Global Studies, Public Health, Business and Political Economy. Many students have no intention of taking another Economics class when they sign up for Introductory Economics; therefore, focusing on the drop-off rates after the first intermediate economics class is more informative. Even after completing both intermediate economics classes there are many students who no longer persist. Most notably, almost half the students who completed Microeconomics and Macroeconomics do not continue on to take Econometrics. This is confirmed by the summary statistics seen in Table 12. Additionally, the Asian and White, male students have the highest persistence rates. Visually there is only a small difference between the three groups.

In comparison, Asian, male students have the highest persistence rate in each of the Computer Science classes. URM, female students have the lowest persistence rate with more than half of them dropping out after CS61A. White, female students are not far behind as 40% of them drop out after taking CS61A. Since White and URM female students received significantly lower grades in CS61A it indicates that grades can predict persistence rates. There is a difference between predicting and influencing persistence rates. The following regressions models the differences in persistence by race and gender.

Model 2: Persistence Rates by Race and Gender

$$P_i = \beta_0 + \beta_1 Gender_i + \beta_2 Race_i + \beta_3 Grade + \beta_4 Race_i * Gender_i + \beta_5 Gender_i * GPA$$
$$+ \beta_6 Race_i * GPA + \beta_7 Gender_i * Race_i * GPA + \epsilon_i \tag{2}$$

Persistence is a binary variable and coded as a 1 if a student decides to take one more class after the introductory class.[5] Grade refers to the numeric grade received in the introductory class. The summary statistics can be used to understand if there are systematic differences for persistence rates by race and gender.

Table 16: Persistence Rates in Economics by Gender

Group	Persistence Rate (pp)	* Male
Female	65.8	
Male	68.1	

Table 17: Persistence Rates in Economics by Race

Group	Persistence Rate (pp)	* Asian	* White
Asian	70.9		Yes
URM	62.7	Yes	Yes
White	60.7	Yes	

[5]Since persistence is a binary variable a probit regression was originally run. There were no major differences between the output from the probit and OLS regression. The OLS regressions are presented for ease of interpretation.

Table 18: Persistence Rates in Economics by Race and Gender

Group	Persistence Rate (pp)	* Asian, M	* White, M	* URM, M	*White, F	*URM, F
Asian, M	73.1		Yes	Yes	Yes	Yes
Asian, F	69.1	Yes	Yes	Yes	Yes	Yes
URM, M	63.5	Yes				
URM, F	61.9	Yes				
White, M	62.0	Yes				
White, F	58.7	Yes				

Table 19: Persistence Rates in CS by Gender

Group	Persistence Rate (pp)	* Male
Female	91.6	Yes
Male	95.1	

Table 20: Persistence Rates in CS by Race

Group	Persistence Rate (pp)	* Asian	* White
Asian	95.3		Yes
URM	92.0	Yes	
White	91.4	Yes	

Table 21: Persistence Rates in CS by Race and Gender

Group	Persistence Rate (pp)	* Asian, M	* White, M	* URM, M	*White, F	*URM, F
Asian, M	96.3			Yes	Yes	Yes
Asian, F	92.5	Yes			Yes	Yes
URM, M	92.3	Yes	Yes			
URM, F	91.0	Yes	Yes	Yes	Yes	
White, M	92.3	Yes			Yes	Yes
White, F	88.0	Yes				Yes

The difference in persistence rates by gender is only significant in the Computer Science department. There are differences in persistence rates by race in both the Computer

Science and Economics department. The following regression output gives us a better understanding of persistence rates and controls for the grade received in introductory classes.

Table 22: Persistence after Introductory Economics by Race and Gender

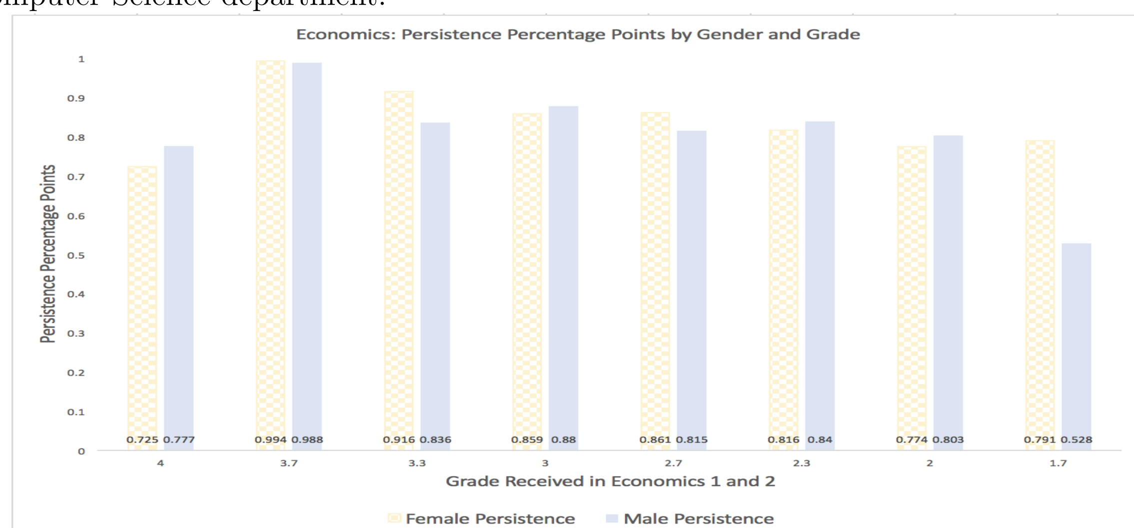
	Dependent variable:					
	Taking One More Class After Introductory Economics					
	(1)	(2)	(3)	(4)	(5)	(6)
URM	-0.082^{***}		-0.083^{***}	-0.134^{***}		-0.210^{***}
	(0.024)		(0.024)	(0.047)		(0.066)
White	-0.102^{***}		-0.107^{***}	-0.070^{**}		-0.053
	(0.019)		(0.019)	(0.031)		(0.041)
Grade			0.011^{**}	0.009	0.087^{***}	0.010
			(0.005)	(0.006)	(0.012)	(0.009)
Female:URM						0.158^{*}
						(0.095)
Female:White						-0.054
						(0.062)
Female:Grade					0.035^{**}	0.002
					(0.018)	(0.012)
Female		-0.023	-0.034^{**}		-0.155^{**}	-0.045
		(0.016)	(0.016)		(0.058)	(0.034)
URM:Grade				0.024		0.053^{**}
				(0.019)		(0.026)
White:Grade				-0.015		-0.027^{*}
				(0.012)		(0.016)
Female:URM:Grade						-0.062
						(0.038)
Female:White:Grade						0.029
						(0.023)
Constant	0.709^{***}	0.681^{***}	0.751^{***}	0.728^{***}	0.712^{***}	0.752^{***}
	(0.011)	(0.011)	(0.017)	(0.017)	(0.019)	(0.025)
Observations	3,400	3,400	3,400	3,400	3,400	3,400
R^2	0.010	0.001	0.013	0.013	0.002	0.016
Adjusted R^2	0.010	0.0003	0.012	0.011	0.001	0.013
Residual Std. Error	0.468	0.471	0.468	0.468	0.470	0.468
F Statistic	17.378^{***}	2.019	11.079^{***}	8.789^{***}	2.561^{*}	4.920^{***}
Note:						$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

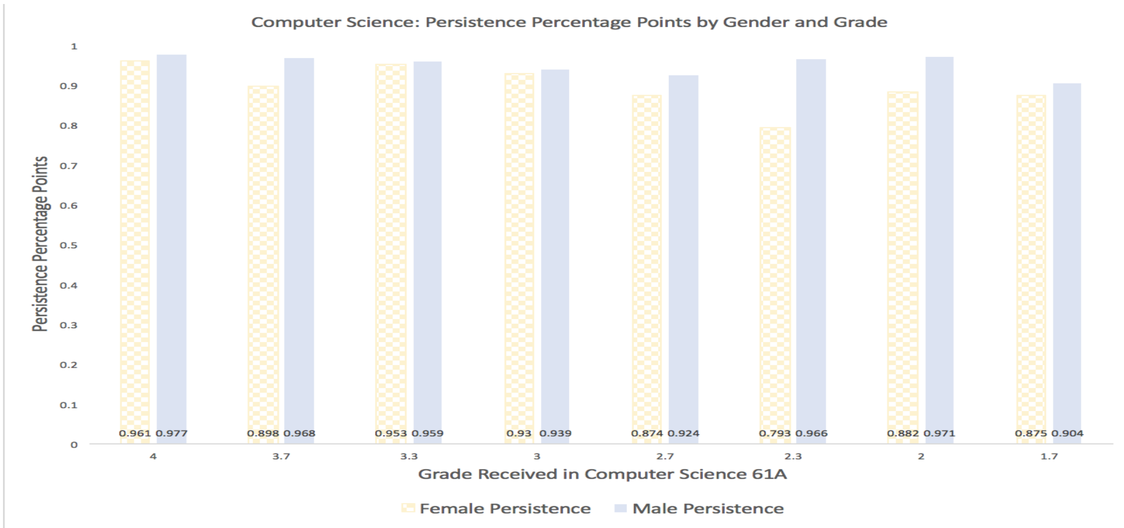
In both Economics and Computer Science there is a small number of URM students (Economics: 489 students, Computer Science: 286 students), which leads to a lot of noise in the regressions seen in Column 6 of Table 19 and Table 22. Therefore, column 5 is focused on. In Economics, a female student is 15.5 percentage points (se = 0.058) less likely to persist that their male counterparts controlling for grade. Perhaps the most interesting insight that is uncovered is the grade elasticity, or differential response by gender to the same grade, which can be seen through the analysis of the interaction between the variables: female and grade. For every GPA point increase, female students are 3.5 percentage points (se = 0.018) more likely to persist than their male counterparts who received the same grade. Therefore, female students are more grade elastic than their male counterparts in the Economics department. The same effect is seen in Computer

Table 23: Persistence after Introductory Computer Science by Race and Gender

	Dependent variable:					
	Taking One More Class After CS61A					
	(1)	(2)	(3)	(4)	(5)	(6)
URM	-0.033^{**}		-0.031^{**}	-0.055^{*}		-0.049
	(0.015)		(0.015)	(0.029)		(0.032)
White	-0.039^{***}		-0.041^{***}	-0.031^{**}		-0.037^{**}
	(0.009)		(0.009)	(0.014)		(0.015)
Grade			0.003	0.004	0.003	0.002
			(0.002)	(0.003)	(0.002)	(0.003)
Female:URM						-0.032
						(0.071)
Female:White						0.037
						(0.039)
Female:Grade					-0.003	-0.001
					(0.006)	(0.006)
Female		-0.035^{***}	-0.036^{***}		-0.027^{*}	-0.034^{**}
		(0.009)	(0.009)		(0.015)	(0.017)
URM:Grade				0.010		0.005
				(0.011)		(0.013)
White:Grade				-0.004		-0.002
				(0.005)		(0.006)
Female:URM:Grade						0.025
						(0.029)
Female:White:Grade						-0.019
						(0.014)
Constant	0.953^{***}	0.951^{***}	0.969^{***}	0.960^{***}	0.956^{***}	0.968^{***}
	(0.005)	(0.004)	(0.007)	(0.007)	(0.006)	(0.008)
Observations	3,831	3,831	3,831	3,831	3,831	3,831
R^2	0.005	0.004	0.011	0.007	0.005	0.012
Adjusted R^2	0.005	0.004	0.010	0.005	0.004	0.009
Residual Std. Error	0.234	0.234	0.234	0.234	0.234	0.234
F Statistic	10.187^{***}	16.931^{***}	10.409^{***}	5.132^{***}	6.459^{***}	4.273^{***}
Note:						*p<0.1; **p<0.05; ***p<0.01

Science, but to a smaller magnitude. The grade elasticity coefficient is 2.7 percentage points (se $= 0.015$). The regressions above and visual graphics below confirm what Goldin saw in the Economics department at the fictional Adams College and extends it to the Computer Science department.





In Computer Science, at all grade levels male students are more likely to persist than female students. In Economics, at the majority of grade levels male students persist at a marginally higher rate. Unsurprisingly, there is a large difference between the number of students who persist between the 3.7 and 4.0 grade level. Most students who receive a 4.0 in Economics 1 or 2 end up majoring in Business Administration.

When accounting for race, gender, and grade the systematic differences by gender disappear in both Economics and Computer Science. Additionally, the differences in grade elasticity by gender are no longer significant, which suggests that the results Goldin saw at the fictional Adams College are not happening to the same degree at UC Berkeley. Since the data is noisy when accounting for race, it should not serve as a basis for trying to understand if the grade elasticity stems from race instead of gender. Regardless, when looking at just race (Economics: Table 17, Computer Science: Table 20), URM and White students persist at a lower rate than their Asian counterparts.

In both departments, the grade coefficient is insignificant in Column 5 and 6 (Table 19 and 23). This entails that even though there are systematic differences in grades received by students it does not affect whether a student decides to take an additional class after the initial one. Therefore, there are other factors that are coming into play that are causing the systematic differences in persistence by race and gender.

6 Results - Persistence Rate Differences by Race and Gender Adjusted for Peer Effects

Research with focus groups hypothesize some ideas on why persistence rates differ by race and gender including: teacher-student interactions, learning environment for students, and previous preparation in high school (Shapiro, 2011). In order to better understand if the learning environment influences a student's persistence rate, I focus on the section dynamics. Since both Economics and Computer Science Introductory lectures are over 500 people, most of the active learning takes place in labs or sections. Part of having a conducive learning environment is feeling safe and being comfortable vocalizing opinions (Shapiro, 2011). I focus on whether having a similar peer (based on race and gender) increases persistence rates for students. Therefore, the following model was used to see if peer effects could explain some of the differences in persistence rates.

Model 3 Persistence Rates by Race and Gender Adjusted by Peer Effects

$$\begin{aligned} P_i = \beta_0 + \beta_1 Gender_i + \beta_2 Race_i + \beta_3 Grade + \beta_4 Race_i * Gender_i + \beta_5 Gender_i * GPA \\ + \beta_6 Race_i * GPA + \beta_7 Gender_i * Race_i * GPA \\ + \beta_8 SimilarPeer_i + \beta_9 SimilarPeer_i * Race_i \\ + \beta_{10} SimilarPeer_i * Gender_i + \beta_{11} SimilarPeer_i * Gender_i * Race_i + \epsilon_i \end{aligned} \tag{3}$$

In determining if a student had a similar peer in their section, international students were re-added into the section dataset. Peer effects is only studied on the domestic four-year students who took their introductory class for a letter grade. The descriptive tables below outline the section demographics for both Computer Science and Economics.[6]

[6] This dataset was created by looking at unique section IDs provided in the initial dataset. Each section's racial and gender breakdown is reported below.

Table 24: Economics Section Data

Race	Gender	No. of Students	No. of Sections
Asian	Male	0	29
Asian	Male	1	37
Asian	Male	2	33
Asian	Male	3	58
Asian	Male	3+	129
Asian	Female	0	31
Asian	Female	1	29
Asian	Female	2	25
Asian	Female	3	40
Asian	Female	3+	161
URM	Male	0	114
URM	Male	1	93
URM	Male	2	44
URM	Male	3	27
URM	Male	3+	8
URM	Female	0	113
URM	Female	1	93
URM	Female	2	50
URM	Female	3	21
URM	Female	3+	9
White	Male	0	58
White	Male	1	63
White	Male	2	51
White	Male	3	44
White	Male	3+	70
White	Female	0	85
White	Female	1	82
White	Female	2	54
White	Female	3	32
White	Female	3+	33
Total			286

Table 25: Computer Science Section Data

Race	Gender	No. of Students	No. of Sections
Asian	Male	0	28
Asian	Male	1	30
Asian	Male	2	27
Asian	Male	3	19
Asian	Male	3+	186
Asian	Female	0	38
Asian	Female	1	49
Asian	Female	2	45
Asian	Female	3	39
Asian	Female	3+	119
URM	Male	0	140
URM	Male	1	84
URM	Male	2	44
URM	Male	3	15
URM	Male	3+	7
URM	Female	0	182
URM	Female	1	77
URM	Female	2	21
URM	Female	3	6
URM	Female	3+	4
White	Male	0	55
White	Male	1	59
White	Male	2	57
White	Male	3	42
White	Male	3+	77
White	Female	0	132
White	Female	1	98
White	Female	2	45
White	Female	3	13
White	Female	3+	2
Total			290

Looking at the raw numbers it becomes clear that there are notably fewer sections in which there are more than 2 URM students in the same section. Therefore, when looking at peer effects, a student was said to have a similar peer as long as there was one other student in the section with the same gender and race. The following set of regressions look at whether or not having a similar peer in a student's section increases likelihood of persisting.

Table 26: Persistence after Introductory Economics Adjusted for Peer Effects by Race and Gender

	Dependent variable:					
	Taking One More Class After Introductory Economics					
	(1)	(2)	(3)	(4)	(5)	(6)
Gender	−0.026		−0.036**		0.061	−0.110
	(0.016)		(0.016)		(0.054)	(0.138)
URM		−0.020	−0.021	0.039		−0.012
		(0.026)	(0.026)	(0.080)		(0.111)
White		−0.090***	−0.095***	−0.0004		−0.177
		(0.019)	(0.019)	(0.082)		(0.117)
SimilarPeer	0.123***	0.107***	0.106***	0.170**	0.172***	0.142
	(0.028)	(0.030)	(0.030)	(0.070)	(0.040)	(0.098)
Grade	0.054***	0.053***	0.053***	0.053***	0.054***	0.053***
	(0.009)	(0.010)	(0.010)	(0.010)	(0.009)	(0.010)
URM:SimilarPeer				0.059		0.033
				(0.084)		(0.118)
White:SimilarPeer				0.094		0.081
				(0.084)		(0.120)
Gender:URM						0.104
						(0.158)
Gender:White						0.234**
						(0.164)
Gender:URM:SimilarPeer						0.058
						(0.167)
Gender:White:SimilarPeer						0.351**
						(0.169)
Gender:SimilarPeer					0.095*	0.062
					(0.056)	(0.140)
Constant	0.411***	0.444***	0.465***	0.383***	0.367***	0.438***
	(0.039)	(0.044)	(0.045)	(0.076)	(0.047)	(0.102)
Observations	3,428	3,428	3,428	3,428	3,428	3,428
R^2	0.016	0.022	0.024	0.023	0.017	0.027
Adjusted R^2	0.016	0.021	0.022	0.021	0.016	0.024
Residual Std. Error	0.466	0.465	0.464	0.465	0.466	0.464
F Statistic	19.006***	19.583***	16.726***	13.263***	14.983***	7.992***
Note:						*p<0.1; **p<0.05; ***p<0.01

In both the Economics and Computer Science department the grade coefficient, which was the grade received in introductory classes, is significant. Looking at Column 6, an additional GPA point increased persistence by 5.3 percentage points (se = 0.010) in Economics and 2.9 percentage points (se = 0.004) in Computer Science. The similar peer coefficient was significant and positive across all regressions, but alone does not mean that much since Asian students are more likely to have similar peers in their section and also have the highest rate of persistence. Looking at the interaction variables between race, gender, and similar peer an interesting insight is gleaned from the Economics department.

Table 27: Persistence after Introductory Computer Science Adjusted for Peer Effects by Race and Gender

	Dependent variable:					
	Taking One More Class After CS61A					
	(1)	(2)	(3)	(4)	(5)	(6)
Female	−0.021**		−0.024***		0.051*	0.033
	(0.009)		(0.009)		(0.028)	(0.066)
URM		0.012	0.009	0.028		0.026
		(0.016)	(0.016)	(0.040)		(0.058)
White		−0.028***	−0.031***	−0.059		−0.103*
		(0.009)	(0.009)	(0.039)		(0.061)
SimilarPeer	0.067***	0.070***	0.062***	0.063*	0.108***	0.091*
	(0.015)	(0.016)	(0.016)	(0.033)	(0.021)	(0.051)
Grade	0.029***	0.031***	0.030***	0.031***	0.029***	0.029***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
URM:SimilarPeer				0.029		0.028
				(0.044)		(0.062)
White:SimilarPeer				0.034		0.076
				(0.040)		(0.062)
Female:URM						0.019
						(0.081)
Female:White						0.069
						(0.079)
Female:URM:SimilarPeer						0.021
						(0.090)
Female:White:SimilarPeer						0.075
						(0.084)
Female:SimilarPeer					0.080***	0.064
					(0.030)	(0.067)
Constant	0.795***	0.785***	0.805***	0.793***	0.757***	0.778***
	(0.020)	(0.021)	(0.022)	(0.035)	(0.024)	(0.052)
Observations	3,875	3,875	3,875	3,875	3,875	3,875
R^2	0.022	0.024	0.026	0.024	0.024	0.029
Adjusted R^2	0.022	0.023	0.024	0.023	0.023	0.026
Residual Std. Error	0.231	0.231	0.231	0.231	0.231	0.230
F Statistic	29.408***	23.451***	20.297***	16.135***	23.904***	9.690***
Note:						*p<0.1; **p<0.05; ***p<0.01

When a white, female student has a similar peer in her section, she is 35.1 percentage points (se = 0.169) more likely to persist in comparison to a white, female student who does not have a similar peer in her section. No similar peer effects are seen in the Computer Science department.

7 Results - Number of Classes

While persistence rates can help explain the initial drop-off that is seen after the introductory class, it does not indicate who has the ability to declare the major. Therefore, the following model aims to understand if there are systematic differences in the number of prerequisite, or core, classes that students take.

Model 4: Number of Core Classes Taken by Race and Gender

$$N_i = \beta_0 + \beta_1 Gender_i + \beta_2 Race_i + \beta GPA + \beta_4 Race_i * Gender_i + \beta_5 Gender_i * GPA \\ + \beta_6 Race_i * GPA + \beta_7 Gender_i * Race_i * GPA + \epsilon_i \quad (4)$$

Classes indicates the number of core classes that a student take. Core classes are needed to declare the major and is a better metric than number of classes since some students decide to double major and do not have time to take more than the required number of classes to graduate. Additionally, many cross-departmental classes are allowed to count for upper division electives. The summary statistics and regressions aim to understand what the systematic differences by race and gender are for the number of core classes.

Table 28: Number of Economics Class by Gender

Group	Number of Classes	* Male
Female	2.23	Yes
Male	2.37	

Table 29: Number of Economics Class by Race

Group	Number of Classes	* Asian	* White
Asian	2.47		Yes
URM	2.05	Yes	Yes
White	2.08	Yes	

Table 30: Number of Economics Class by Race and Gender

Group	Number of Classes	* Asian, M	* White, M	* URM, M	*White, F	*URM, F
Asian, M	2.55		Yes	Yes	Yes	Yes
Asian, F	2.40	Yes	Yes	Yes	Yes	Yes
URM, M	2.21	Yes			Yes	Yes
URM, F	1.90	Yes	Yes	Yes		
White, M	2.17	Yes	Yes		Yes	Yes
White, F	1.94	Yes		Yes		

Table 31: Number of CS Class by Gender

Group	Number of Classes	* Male
Female	3.42	Yes
Male	3.93	

Table 32: Number of CS Class by Race

Group	Number of Classes	* Asian	* White
Asian	3.89		Yes
URM	3.71	Yes	Yes
White	3.50	Yes	

Table 33: Number of CS Class by Race and Gender

Group	Number of Classes	* Asian, M	* White, M	* URM, M	*White, F	*URM, F
Asian, M	4.03		Yes		Yes	Yes
Asian, F	3.56	Yes	Yes	Yes	Yes	Yes
URM, M	3.96		Yes		Yes	Yes
URM, F	3.64	Yes	Yes	Yes	Yes	
White, M	3.65	Yes	Yes	Yes	Yes	Yes
White, F	2.99	Yes	Yes	Yes		

Looking at the summary statistics, there is a gender gap for the number of classes female students take in comparison to male students in both departments. Additionally,

Table 34: Number of Core Classes Taken After Introductory Economics by Race and Gender

	Dependent variable:					
	Number of Core Economics Classes Taken					
	(1)	(2)	(3)	(4)	(5)	(6)
URM	−0.414***		−0.288***	0.090		0.185
	(0.070)		(0.072)	(0.220)		(0.317)
White	−0.388***		−0.394***	−0.778***		−0.819***
	(0.055)		(0.055)	(0.215)		(0.287)
Grade			0.220***	0.210***	0.259***	0.216***
			(0.029)	(0.040)	(0.040)	(0.059)
Female:URM						−0.163
						(0.440)
Female:White						0.036
						(0.440)
Female:Grade					−0.022	−0.011
					(0.057)	(0.081)
Female		−0.148***	−0.196***		−0.086	−0.115
		(0.048)	(0.048)		(0.177)	(0.258)
URM:Grade				−0.150*		−0.158
				(0.078)		(0.111)
White:Grade				0.138**		0.158*
				(0.069)		(0.092)
Female:URM:Grade						−0.003
						(0.157)
Female:White:Grade						−0.047
						(0.140)
Constant	2.469***	2.375***	1.894***	1.817***	1.605***	1.880***
	(0.031)	(0.034)	(0.100)	(0.129)	(0.124)	(0.189)
Observations	3,386	3,386	3,386	3,386	3,386	3,386
R^2	0.020	0.003	0.040	0.039	0.024	0.044
Adjusted R^2	0.019	0.002	0.039	0.037	0.024	0.041
Residual Std. Error	1.391	1.403	1.377	1.378	1.388	1.376
F Statistic	33.690***	9.467***	35.487***	27.173***	28.300***	14.189***
Note:						*p<0.1; **p<0.05; ***p<0.01

Table 35: Number of Core Classes Taken After Introductory Computer Science by Race and Gender

	Dependent variable:					
	Number of Core Computer Science Classes Taken					
	(1)	(2)	(3)	(4)	(5)	(6)
URM	−0.189*		−0.088	−0.388		−0.626*
	(0.099)		(0.100)	(0.271)		(0.333)
White	−0.395***		−0.419***	−0.980***		−1.128***
	(0.063)		(0.062)	(0.232)		(0.268)
Female		−0.508***	−0.502***		−0.798***	−1.005***
		(0.058)	(0.058)		(0.205)	(0.259)
Grade			0.119***	0.095**	0.097***	0.007
			(0.030)	(0.037)	(0.035)	(0.045)
Female:URM						0.816
						(0.570)
Female:White						0.249
						(0.533)
URM:Grade				0.108		0.218*
				(0.098)		(0.117)
White:Grade				0.191***		0.236***
				(0.072)		(0.082)
Female:URM:Grade						−0.417*
						(0.215)
Female:White:Grade						−0.123
						(0.172)
Female:Grade					0.108	0.179**
					(0.066)	(0.082)
Constant	3.899***	3.934***	3.662***	3.599***	3.627***	4.008***
	(0.031)	(0.030)	(0.103)	(0.123)	(0.115)	(0.149)
Observations	3,819	3,819	3,819	3,819	3,819	3,819
R^2	0.011	0.020	0.036	0.019	0.025	0.041
Adjusted R^2	0.010	0.020	0.035	0.018	0.025	0.038
Residual Std. Error	1.589	1.581	1.569	1.583	1.577	1.567
F Statistic	20.322***	77.712***	35.851***	14.850***	33.070***	14.634***
Note:						*p<0.1; **p<0.05; ***p<0.01

in Economics and Computer Science URM and White students take less classes than Asian students. Looking at Column 6, in Economics, the grade elasticity term for White students is significant. This indicates that White students are 15.8 percentage points more likely to persist than their non-white students, conditional on the same GPA point increased.

In Computer Science, many of the coefficients are now significant in comparison to the persistence regressions. The grade elasticities are significant for URM, White, and female students. URM students have a grade elasticity of .218 classes (se = 0.117), White students have a grade elasticity coefficient of .236 classes (se = 0.082) and female students have a grade elasticity of .179 classes (se = 0.082). Since there are larger differences by race and gender in the number of classes taken in the Computer Science department, it indicates that the drop-off happens somewhere in between the second and fourth class.

8 Conclusion

Are there systematic differences by race and gender in the grades received in introductory classes and persistence rates? In Economics, there is a systematic difference in the grade received in introductory classes by race and systematic differences in persistence rates by race and gender for 4-year domestic students. In Computer Science, there are systematic differences in the grades received in introductory classes and in persistence rates by race and gender for 4-year domestic students.

The first significant result is that there is no gender effect in the grade received in Economics, but there is one in Computer Science. One of the reasons that could explain this discrepancy is previous high school coursework. In the state of California, which is where a majority of UC Berkeley students come from, taking an Economics class is mandatory to graduate. On the other hand, Computer Science is an elective that is often offered in richer high school districts. In 2011, which is a relevant time period for the data studied, only 21% of all students who took the AP Computer Science exam nationally were female students and only 29 students who took the AP Computer Science exam were Black (UCLA Idea). This indicates that there are differences in who has access to more technical and advanced classes, such as Computer Science, which indicates that previous coursework could be a reason for there are varied systematic differences in the Computer Science and Economics department.

The second significant result is that there is a systematic difference in persistence rates by race and gender in both the Economics and Computer Science department. Goldin found that in Economics, female students respond to grades differently than males. After accounting for both race and gender, the analysis in this paper suggests that the differences in persistence rates could stem from race and not gender. Unfortunately, there are not enough URM students to smooth out the standard errors. Research that has been done via focus groups indicates that URM students feel as if they have more pressure to succeed and hold themselves to a higher standard when taking quantitative courses (Zafar, 2010). While this study was focusing on the distinction between URM and non-URM students, the same analysis could potentially apply for White and URM students since they make up the smallest proportion of students in the Computer Science and Economics department.

The third significant result is that when a White, female student has a similar peer in her section they are 35 percentage points more likely to persist than a White, female student who does not have a similar peer in her section. There is roughly a 40% drop-off rate with White, female students after Introductory Economics. Programs such as Undergraduate Women in Economics at Berkeley (UWE) and Students of Color in Economics (SOCE) are working towards mitigating the gender gap observed in the department. Since a similar peer has a positive impact for White, female students, these supplementary programs could create another place for them to become more comfortable being part of the Economics department.

The last finding is that in Economics, the systematic differences by race and gender are not as large when tracking the number of classes versus persistence rate. This suggests that once a student takes their first intermediate economics class they are likely to continue taking more classes. Therefore, it is crucial for the Economics department to work on providing the support needed for students in introductory Economics so that they have the incentive to persist afterwards. On the other hand, Computer Science students were more likely to persist after the first class; the drop-off happened between the second and fourth class. It is harder to tell which class is dissuading students to persist or why they drop-out, but equalizing the playing field is essential to ensure equal representation in the students who declare Economics and Computer Science as their major.

This year, the Computer Science created a 2-unit course for students who have never taken Computer Science before. The class aims to prepare students in a non-competitive environment for Computer Science 61A. If students who are currently struggling in CS61A take this class, there is a chance that the racial and gender gap could lessen. While it is too early to tell the effects of this program, if it proves to be scalable, it can be mimicked in the Economics department. The ideas mentioned above are just starting points for how to lessen the racial and gender gap in the Economics and Computer Science department. Future impacts should be analyzed to understand the efficacy of all the programs.

9 Bibliography

Arcidiacono, Peter. 2010. *Modeling College Major Choices Using Elicited Measures of Expectations and Counterfactuals.* 1-5.

Astorne-Figari, Carmen. 2017. *Are Changes of Major Major Changes? The Roles of Grades, Gender, and Preferences in College Major Switching.* 2-25.

Dickson, Lisa. 2010. *Race and Gender Differences in College Major Choice.* 1-17.

Goldin, Claudia. 2015. *Gender and the Undergraduate Economics Major.* 1-20.

Hastings, Justine. 2015. *UnInformed College and Major Choice: Evidence from Linked Survey and Administrative Data.* 1-15.

Mullen, Ann. 2013. *Gender, Social Background, and the Choice of College Major in a Liberal Arts College.*

Porter, Stephen. 2006. *College Major Choice: An Analysis of Person-Environment Fit.* 420-446.

Shapiro, Casey. 2011. *Major Selection and Persistence for Women in Stem.*

Staniec, Farley. 2004. *The Effects of Race, Sex, and Expected Returns on the Choice of College Major.* 549-560.

Wiswall, Matthew. 2014. *Determinants of College Major Choice: Identification using an Information Experiment.* 1-34.

Zafar, Basit. 2005. *College Major Choice and the Gender Gap.* 1-47.