# ESSAYS IN APPLIED MICROECONOMICS

A Dissertation

by

ADAM BESTENBOSTEL

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

| | |
|---|---|
| Chair of Committee, | Mark Hoekstra |
| Committee Members, | Kalena Cortes |
| | Jonathan Meer |
| | Steven Puller |
| Head of Department, | Steven Puller |

May 2022

Major Subject: Economics

ABSTRACT

This research consists of three separate essays each focused on information, how it affects human preferences and decision-making, and its effects on marginalized groups.

In the first essay, I test a common hypothesis for the gender gap in STEM, which is believed to drive much of the gender gap in earnings. Specifically, it is believed that women are more responsive to the negative grade signals that are common in STEM. I test this by applying a regression discontinuity design to the underlying numerical scores of more than 21,000 university students in feeder courses in STEM and economics. Results indicate that letter grade thresholds (with no plus/minus modifiers in this context) have no effect on STEM major choice for either women or men. This is true even for susceptible subgroups, within particular majors, or at particular grade thresholds.

The second essay is motivated by the racially charged rhetoric which often surrounds layoff events, with specific minorities blamed for the loss of "American jobs." A coauthor and I examine whether information about impending mass layoffs causes racial animus. Our data consist of information on mass layoff notices linked to Google Search Trends and FBI Hate Crime Statistics. We compare outcomes across areas that vary in the timing of news of impending layoffs. Results indicate an increase in both racist internet searches (1.5 percent) and hate crimes (23 percent) following layoff notices.

In the third essay, which is also coauthored, we remark how media ownership has become more concentrated in recent years, leading to concerns over media integrity and pluralism as well as the nature of the information being passed on to the public. In this paper, we study the impact of broadcast television ownership consolidation on ideological preferences. To do so, we use a difference-in-differences design to examine the impact of within-market consolidation on election outcomes. Results show that within-market consolidation shifted vote share towards Democrats by 3-4 percentage points for both presidential and senate elections, and that this effect persists for at least 12 years.

# DEDICATION

To Mom and Dad, for your undying support, love, and belief in me.

ACKNOWLEDGMENTS

First and foremost I wish to thank my advisor, Mark Hoekstra, for teaching me about causal inference and academic research. Your time and patience are valued and appreciated. My gratitude also goes out to the other members of my dissertation committee for their feedback and advice over the years: Jonathan Meer, Steve Puller, and Kalena Cortes. Additionally, I thank the generous faculty in my department for the incredibly supportive graduate school environment. I consider myself fortunate to have been a student here during this time. And of course none of this would have been possible without the amazing support from staff members Chelsi Bass, Kurt Felpel, Mary Owens, and Teri Tenalio.

I am grateful to my fellow graduate students, especially those who have always offered and continue to offer support and advice with endless patience: Abi Peralta, Brittany Street, Carly Will Sloan, and Meradee Tangvatcharapong. I don't know if I would have gotten this far without you all.

Thank you to all of the STEM instructors who were willing to assist with data for my research in Section 2. You chose to help me and it means a great deal.

Finally, I appreciate the patience and support of all of my friends and family who are not economists. You tolerated me talking endlessly about economics – but importantly, you were also there to talk with me about things that are not economics.

CONTRIBUTORS AND FUNDING SOURCES

## Contributors

This work was supervised by a dissertation committee consisting of Professor Mark Hoekstra, primary advisor, as well as Professors Jonathan Meer and Steven Puller of the Department of Economics and Professor Kalena Cortes of the Bush School of Government and Public Service.

Section 3 is joint work with Abi Peralta, and Section 4 is joint work with Meradee Tangvatcharapong. All other work conducted for the dissertation was completed by the student independently.

## Funding Sources

Graduate study was supported by an assistantship from Texas A&M University and a summer fellowship from the Private Enterprise Research Center at Texas A&M University.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

TABLE Page

# 1. INTRODUCTION

In this research, I use quasi-experimental methods to address causal research questions in applied microeconomics. The purpose is to both understand aspects of human behavior and inform policy. These essays focus on information and how it can shape preferences and economic decision making, and in particular how this affects marginalized groups.

In Section 2, I consider one potential information contributor to the gender wage gap. Specifically, I test the hypothesis that college students are sensitive to letter grade signals in challenging STEM courses where desirable grades are scarce. It is commonly believed that men and women respond differently to these signals such that it drives a gender gap in STEM majors. To answer this question, I collected student-level data from twenty instructors teaching nine core, introductory STEM and economics courses which lead to majors in biology, chemistry, engineering, and other major STEM fields. Then I had these data matched with administrative data so that I have final number grades as well as letter grades, major, gender, and other background variables. This allows me to use a regression discontinuity design to identify effects at letter grade thresholds. Despite the hypothesis that grades matter, I find that they do not for either men or women. This null result is consistent for a variety of alternative specifications and subgroups. This suggests that grades are not a factor for the gender gap in STEM and therefore the wage gap, and highlights the importance of understanding other avenues.

In Section 3, coauthored with Abi Peralta, we consider the role of information in a different context. Specifically, we investigate how negative news about the economy can spark racial animus. We use the timing and location of information on impending mass layoffs data to identify effects on Google Search Trends for racial slurs and also on hate crime. We find positive effects on both outcomes in the same month the layoff notice is issued, suggesting that people are responding to the information rather than the layoff events themselves. Interestingly, we find that some of the racial animus measured by internet searches is directed at Blacks, while much of the alt-right rhetoric surrounding layoff events typically targets Hispanic immigrants or outsourcing to Asia. It is something of an open question as to how or why these racist ideas propagate, but one theory is biased media.

Finally, in Section 4, written with Meradee Tangvatcharapong, we are interested in how information and media can shape preferences. We consider the effect of local media ownership consolidation in broadcast TV stations and measure ideological preferences using election outcomes at the county level. The idea here is that consolidation can lead to cost savings in terms of local and syndicated news, or changing coverage of federal or state elected officials and therefore changing levels of accountability. To identify our estimates, we use a difference-in-differences design and find that this kind of consolidation leads to a 3-4 percentage point shift towards Democrats for both presidential and senate elections.

# 2. DO GRADE SIGNALS DRIVE THE GENDER GAP IN STEM? EVIDENCE FROM A REGRESSION DISCONTINUITY

## 2.1 Introduction

In 2017, U.S. women earned 80 cents to the dollar relative to men [Fontenot, Semega, and Kollar, 2017]. While there are several potential explanations for this, gender differences in occupational choice are estimated to be responsible for 44 percent of the wage gap [Goldin, Kerr, Olivetti, and Barth, 2017]. For example, women held 14 percent of full-time architecture and engineering jobs and 25 percent of full-time computer and math jobs in 2016, which pay 25 and 36 percent more than the national average for bachelor's degree graduates, respectively [Bureau of Labor Statistics, 2017, 2019].

The gender disparity in STEM occupations can be traced back to gender differences in college majors. While women make up 58 percent of all bachelor's degree recipients, only 36 percent of bachelor's degrees in STEM are conferred to women [National Center for Education Statistics, 2019]. Indeed, Speer [2021] attributes more than half of the STEM occupation gap to differences in college major, and Card and Payne [2021] suggest these differences in major explain up to a fifth of the wage gap. This is especially striking given that female students outperform their male peers in the relevant STEM courses at all levels of education [O'Dea, Lagisz, Jennions, and Nakagawa, 2018]. A critical question, then, is why women are so much less likely to select into a STEM field. One recent hypothesis is that women are more deterred by the grade environment typical of STEM courses in college, where desirable grades are relatively scarce. For example, Goldin [2015] writes that "Grades in Principles are extremely important in determining whether females major in economics. But that is far less the case for males." The relative aversion to STEM majors for women could be driven by documented gender differences in preferences with respect to risk or competition, as STEM majors are typically viewed as both more difficult/higher risk and where good grades are harder to earn. Indeed, Niederle and Vesterlund [2010] suggest that "evidence of a large gender gap in mathematics performance at high percentiles in part may be explained by the differential manner in which men and women respond to competitive test-taking environments." The purpose of this paper is to document the extent to which grade signals drive the gender gap in STEM.

I begin by descriptively examining the correlation between grades and the likelihood of majoring in STEM, and whether this gradient is different for men and women. However, this approach suffers from obvious omitted variable bias concerns, since major choice is likely affected by student dimensions other than grades. To address this issue, I then use a regression discontinuity design that compares the college major decisions of men and women just above and below the thresholds that determine final letter grades. I use administrative data from a large, public university linked to final number grade. This setting provides several important advantages for implementing this analysis. First, I am able to focus on only those courses where there is little scope for grade manipulation. For example, exams are multiple choice or final grades have been preemptively verified, and in all cases grading policies are strictly enforced and instructors have specifically stated they do not allow arbitrary revision of final grades. Second, I am able to test for differences across a wide range of courses, instructors, and cutoffs to see if grades matter anywhere. Third, I am able to do so in a setting where only full letter grades are given, without pluses or minuses. This generates much larger discontinuities than in the presence of grade modifiers. The identifying assumption of this approach is that all other determinants of major besides letter grade vary smoothly across the letter grade cutoff.

In order to implement the regression discontinuity design, I use administrative data on more than 21,000 student-course observations covering a total of 7 distinct STEM courses taught by 16 different professors at a large university. These data include final number grade (e.g. 0 to 100), letter grade, gender, major, and other individual student characteristics such as prior academic performance, standardized test scores, first generation, transfer, and international status, as well as whether a student has applied for financial aid. These data come from a variety of core, compulsory, first and second-year STEM courses including biology, calculus, computer programming, economics,[1] engineering statics, and organic chemistry. Importantly, I include only those courses for which instructors explicitly declared that they did not allow for or engage in any manipulation of students around the grade thresholds.[2]

The major threat to identification in this context is manipulation around the cutoff. For example, a problematic behavior would be if instructors were to grade exams subjectively so that students judged to be different on unobservable characteristics (such as being more or less deserving) would land on the other side of the threshold. There are three reasons why I believe this is unlikely to have occurred. First, these are all very large classes with an average enrollment of 166, which leaves little opportunity for professors to know students or their unobservables. Second, because of these large class sizes, almost all the exams are multiple choice and machine-graded, which allows few opportunities for the type of subjective grading that would be problematic. Third, and perhaps most importantly, I screened for this behavior when I recruited instructors for their numerical scores. Specifically, I asked whether instructors would bump up scores for students they thought were particularly deserving of a higher grade. Nearly all said they did not; three who suggested they might do so were excluded from the sample.[3] A related threat to identification is if instructors were to draw the thresholds such that the students just above the threshold were unobservably (to the researcher) different than those just below the threshold. While it is clear both from the data and my conversations with the instructors that they do choose the thresholds – many said they place them where there are large gaps in the numerical distribution, rather than at predetermined cutoffs – none stated that they did so based on who the students were. Again, in these large classes, it is unlikely for instructors to know students well. I also note that for many of these courses, instructors met with others to agree on a common grade distribution across sections, limiting the flexibility of any one instructor choosing letter thresholds. Nonetheless, I show that

[1] I do not include economics in the main results since the major is not classified as STEM at this university, although I do use it for later analysis which has identical results. These economics data add an approximate 6,000 observations from 2 courses and 4 more instructors.

[2] As discussed later in section 2.3, instructors often choose to place thresholds where there are large gaps in the distribution and not at predetermined cutoffs such as 90.00%.

[3] One said, "I would never do that because when word got out, there would be a never-ending line of students at my door or emailing me to beg for a higher grade."

results are robust to an exercise where I exclude observations within up to 2 points of the threshold.

Results from a simple analysis of the relationship between grades and STEM major shows a positive correlation, but no difference in this gradient between men and women. However, unobservable differences across students could be biasing these results. For example, it could be true that low grades deter women from STEM, but that women also have more perseverance or grit than men on average. This could offset the deterrent effect of negative grade signals, particularly at lower grade levels. To address these concerns, I should compare students who are identical in all aspects including perseverance and grit, but with as-good-as-random differences in letter grade only. Therefore, I also use a regression discontinuity design at letter grade cutoffs. Results indicate that despite the widespread belief that grades matter differently for men and women's major choice, there is no evidence that letter grade signals affect graduating with a STEM major for either gender. This result is robust to a variety of specifications and holds true even for susceptible subgroups, within particular majors, or at individual grade thresholds. Importantly, estimates enable me to rule out that a higher letter grade reduces the likelihood of graduating with a STEM major by more than 3.2 percentage points (6.0 percent) for women. This is smaller than other input factors to female STEM participation.[4]

In assessing the role of grade signals on college major, this study is most related to two others that examine the likelihood of majoring in economics across grade thresholds. Owen [2010] uses a similar methodology and data but is limited to data from one economics course with 1,300 students, and thus examines only majoring in economics as an outcome. Main and Ost [2014] also use a regression discontinuity on a sample of 2,126 students in micro and macro principles courses and find no effect. Their context involves grade modifiers, so the distinction between an A- and a B+ may not be strong enough to induce an effect. My paper differs from these in two key ways. First, my sample includes many more students and instructors. With more than 21,000 student observations, this is 10 to 16 times larger than the samples examined in these studies. Additionally

[4]Carrell, Page, and West [2010] find having all female instructors for math and science increase graduation with a STEM major by 15.5 p.p. ($\sim$40%) for women with above average math SAT scores, and Porter and Serra [2020] find that exposure to a "successful and charismatic" female role model can induce women to major in economics by 8 p.p. ($\sim$100%).

and perhaps most importantly, I examine behavior beyond economics to cover science, technology, engineering, and mathematics. As a result, the main contribution of this paper is that it is the first to my knowledge to evaluate the impact of letter grades on the gender gap in STEM majors.

In addition, the paper also contributes to a small but growing literature providing evidence outside the laboratory setting on gender responses to environmental factors such as feedback signals. Johnson and Helgeson [2002] find that women were more likely to agree with employee performance evaluations, and their self-esteem was affected to a greater degree than that of men by both positive and negative reviews. Additionally, Mayo, Kakarika, Pastor, and Brutus [2012] show that women in an MBA program are more likely than men to align their self-ratings with those from peers. Lastly, Kugler, Tinsley, and Ukhaneva [2017] use a switching model based on a selection-on-observables identification strategy to estimate that female students choose to switch out of STEM not solely because of grades, but only when they face multiple signals of "lack of fit" such as poor grades in addition to being in a male-dominated field or external stereotypes. My paper complements this literature by demonstrating that women do not adjust their major choices away from STEM as the result of performance feedback from instructors in the form of letter grades.

In short, this paper demonstrates that letter grade signals do not seem to explain any of the gender gap in STEM majors, and thus do not explain any of the gender wage gap that exists due to occupational differences between men and women. This also suggests that while there may indeed be behavioral responses that generate the gender gap in majors, it is not a result of grade signals. In the following sections, I will outline my emprical approach, data, and results.

### 2.2 Research Design

I evaluate the relationship between grades and college major in two ways. First, I evaluate the correlation between both number and letter grades for men and women. Specifically, I use a naive regression of the probability of majoring in STEM on grade to estimate the slope, and then test whether these slopes are equal between genders. However, a concern with this approach is omitted variable bias, since there are many factors related to grades that also affect a student's likelihood of majoring in STEM.

Indeed, it is difficult to assess the impact of grade signals on college major because they are likely correlated with many unobserved factors such as preferences or determination, which impact major themselves. To overcome these concerns, I utilize a regression discontinuity approach in the context of letter grade cutoffs. For the main analysis I utilize a multiple-threshold "stacked" regression discontinuity design following Pop-Eleches and Urquiola [2013]. The advantage of using the stacked approach in this context is greater statistical power to detect significance in results. Formally, I estimate the following model:

$$y_{iz} = \alpha + \beta \cdot \widetilde{\text{score}}_{iz} + \gamma \cdot 1[\widetilde{\text{score}}_{iz} \geq 0] + \delta \cdot \widetilde{\text{score}}_{iz} \cdot 1[\widetilde{\text{score}}_{iz} \geq 0] + \text{Cutoff}_z + X_i + \epsilon_{iz} \quad (2.1)$$

Here $y_{iz}$ represents student $i$'s outcome for cutoff $z$, and $\widetilde{\text{score}}_{iz} = \text{score}_i - \text{score}_z$ is the standardized running variable. $\text{Cutoff}_z$ represents a set of three dummy variables, one for each letter grade cutoff, and $X_i$ represents a vector of individual-specific control variables. I estimate the model with a local linear regression and uniform kernel. The coefficient of interest is $\gamma$, which indicates the magnitude of the discontinuity at the stacked threshold. Here, it can be interpreted as the effect of a higher letter grade in core STEM courses on the likelihood of a student graduating with a STEM major. Standard errors are two-way clustered at the instructor and term levels to account for correlation both within instructor and within terms.

There are a number of reasons to believe a final letter grade can be thought of as a signal which contains information, including that pertaining to student aptitude for a given subject. First, letter grades affect GPA which matters to potential employers and students know this. In some cases, such as for medical school applicants, letter grades for particular courses can matter even more. The students' internal decisions about college major should therefore be affected by their knowledge of the letter grade and its value as an external signal. Second, it stands to reason that the final letter grade and not the precise final number grade is more easily recalled by the student as a result of availability bias. Third, the letter grade can be thought of as validation of effort and knowledge over the entire semester. No matter what thoughts are in the minds of those students who just missed the higher letter grade cutoff, they are certainly different from those students who just *made* the same cutoff and are likely quite pleased with the result of all their hard work.

The identifying assumption of this approach is that all other determinants of major besides letter grade vary smoothly across the letter grade cutoff. The major threat to identification is the possibility that students or professors precisely manipulate student grades around the cutoff according to some unobserved determinant of the outcome. There are several reasons why this identifying assumption is likely to hold in this context. The first is that while students can affect their grade through additional effort, they cannot precisely control the exact number grade. More practically, they likely do not even know the exact letter grade thresholds since instructors often adjust them at the end of the semester in order to yield a certain letter grade distribution, or in order to draw cutoffs where there are gaps in the numerical score distribution. A more significant concern, however, is that individual students could approach the instructor after final grades have been released, seeking to improve their score to achieve a better letter grade. For example, a student who is "more motivated" could be more likely to visit the instructor after final grades are released, and since this attribute is correlated with both desire to achieve a higher letter grade as well as persistence in STEM, this would bias estimates. Similarly, instructors could theoretically adjust scores in order to assign what they believed to be the appropriate grades to those near the threshold. Importantly, it will not invalidate my research design if instructors shift the entire grade distribution to bump up letter grades overall - only if they alter individual numerical scores. Knowing these issues in advance, I carefully selected courses and instructors based on reputation and then spoke with the instructors individually.

In interviews, instructors stated explicitly that they did not allow manipulation of grades. Three professors were unable to make this assertion or mentioned changing individual student scores to help them across grade thresholds - each of these are thus excluded from all data and analysis in this paper. There are also two specific facts gleaned from these interviews that mitigate concerns over grade manipulation. First, no instructor in my sample allows special extra credit opportunities for students close to grade cutoffs, which is one potential pathway for manipulation. Second, the vast majority of the courses in my sample use multiple choice exams, which are objectively machine-graded.[5] Most of the other assignments for the courses in my sample are also computer-graded through online learning management systems. Although this does not prevent a student coming to seek regrades, it does mitigate concerns if most grades are objective and machine-graded with little room for flexibility.

Overall, instructors reported that students seldom if ever even attempt to manipulate their way to a higher letter grade each semester. This is possibly a reflection of their reputation as being strict with respect to grading policies. Additionally, though I did not ask this question in interviews, three of the instructors reported that they take the extra precaution of identifying students close to letter grade thresholds prior to releasing them, and preemptively regrade major assignments for these students. This is so that if students do come to complain about grades, nothing will change. It is also possible for students to contest grades at the department or college level, although that is rare in general and would not be a concern in this context since it would only affect the final letter grade, not the number grade.

In these discussions, many instructors said that while they did not move numerical scores, they would often look for natural gaps in the grade distribution to draw the actual letter grade cutoff. The reasons for this are typically twofold. First, a certain letter grade distribution is often targeted to match that of other classes in the same or previous semesters. Second, instructors know if the next highest score below a letter grade cutoff is further away, students will be less likely to complain about being close to it and less likely to plead for a higher letter grade. Looking for a break in the numerical distribution that yields the desired portion of each letter grade is a solution to both problems. This will likely lead to a density histogram which is not uniform due to a drop in frequency of observations close to the cutoff. Importantly, this does not invalidate the research design as long as they are not drawing the cutoff where based on some unobservable-

[5] Many use a full multiple choice exam including biology, calculus, computer programming, and economics. However, one calculus instructor reported not using a multiple choice exam, and one biology instructor reported using a partial multiple choice exam for the honors section.

to-the-researcher characteristics, e.g. where there are "less motivated to succeed" students with the lower letter grade and "more motivated" students with the higher one. I address this concern in Section 2.4 by performing a "donut" RD that omits those observations within up to 2 grade percentage points of the threshold. It is also helpful that class sizes for my sample are typically quite large as is true with many core classes at public universities, so instructors are less likely to know individual students well.

I also perform the standard empirical tests of the identifying assumption. I plot a histogram of observations near the stacked cutoff in Figure 2.2 and check for abnormal heaping. I regress individual covariates in table A.2. And using student characteristics including pre-existing academic ability to predict whether a student has a STEM major, I test for and find evidence of smoothness through the letter grade threshold in Figure 2.3. These tests are discussed in greater detail in section 2.4 with other results.

### 2.3 Background and Data

Data for this project come from a large, public university in the U.S. and include observations on students who were enrolled in core STEM courses including biology, calculus, computer programming, economics, engineering statics, and organic chemistry. The courses were selected because interviews with advisors and professors indicated that they were challenging first- or second-year classes critical to the curriculum of popular STEM majors including biology, pre-med, computer science, and engineering. In addition, each course selected featured significant enrollment of female students.

At this university, students are admitted with a declared major and are allowed to request a change of major at any time. This means that most students in my sample data are taking these classes to satisfy the requirements of a major they are already pursuing. However, they can also switch to a different major if they wish. Change of major requests are generally allowed as long as the student has met the degree requirements for the program they wish to switch into, though the decision to grant such requests can also depend on space and demand. Typically this means it is relatively easy to switch from a STEM major to a non-STEM major since the GPA and course requirements tend to be less strict. To transfer into another major, a student would have had to have already completed some of the core coursework (e.g. courses in my sample data) and achieved a minimum GPA.

A critical yet unique aspect of the data I use is that in addition to containing final letter grades, they also contains final numerical scores used to determine the letter grade. I collected these data by meeting with individual instructors and coordinating the secure uploading of these scores to the administrative office. That office then matched these scores to university records including final letter grade, major, gender, and other background characteristics. All data were then carefully de-identified to preserve anonymity. In most cases, I observe unique class groups, but for some observations I must assume this using instructor-term groups. I do not observe the actual letter grade thresholds.

For my main analysis, I use 21,533 student records from 16 instructors representing seven different courses and numerous STEM majors including engineering and pre-med as well as biology, chemistry, and math. For each student-by-course, I observe final numerical grade, final letter grade, gender, and major as of graduation. Additionally, the data include information on prior academic ability such as SAT scores,[6] high school rank, and prior post-secondary GPA, student status as a first generation, transfer, or international student, and an indicator for whether the student applied for financial aid. The data span 2004–2019, depending on when each individual instructor was active in the course. Data on major only exist for students who have graduated. To account for this, all analyses are restricted to students for whom the six-year graduation rate can be determined, which leaves the sample of 21,533 observations. Additionally, the main outcome is defined as "graduated with a STEM degree within six years."[7]

In later analyses, I use an additional sample from two economics courses, principles of and intermediate microeconomics. This makes for a total of twenty instructors, nine courses, and 27,572 observations. These data are excluded from the main STEM analysis because the courses

[6]For cases where I observe ACT scores, I use official concordance tables to convert to SAT scores and then take the maximum of the ACT concordance score or the SAT score, if the latter exists.

[7]The sample with a six-year graduation rate does include some observations as recent as 2019 - likely from students who just needed to complete that one STEM course to complete their graduation requirements.

are commonly taken by many non-STEM majors, and because this university's economics major is not STEM according to the Department of Homeland Security definition.

Summary statistics are shown in Table 2.1. Women make up approximately 46% of my sample, which includes courses that are typically male-dominated, such as engineering, as well as those which are often female-dominated, such as biology. Along most characteristics, men and women look similar on average. Interestingly, women seem to have a better high school rank at 48.2 versus men at 64.6 (lower is better) but worse SAT scores with 1198 versus 1249. Women also graduate at a slightly higher rate within six years (90% vs. 86% for men) and are less likely to do so with a STEM major (50% vs. 66%), both of which are consistent with other literature.

The exact cutoffs between letter grades typically vary both across instructors and even within an instructor across semesters, in part because instructors say they can change the grade thresholds based on the overall distribution of grades and the gaps in the distribution. To determine the grade thresholds, I use an objective algorithm to determine the cutoff rule for each instructor-course-semester group of students separately. For each class, I find the lowest A and highest B. Then, for test cutoffs at each 0.01 numerical grade increment between these two scores, I regress a binary indicator for receiving the higher letter grade on a binary indicator for whether a numerical grade is equal to or greater than the incremented "cutoff." I set the RD cutoff for the class group equal to the test cutoff at the increment that generates the best fit $R^2$. I repeat this procedure for the B/C and C/D cutoffs.[8]

While my data contain scores, letter grades, and major information for each student enrolled in the course, there are cases where I am missing data for certain covariates. Table A.1 shows to what extent missing variables exist in the data. Roughly half of all observations are missing at least one variable, and the most commonly missing variables are high school rank (30%), transfer hours (12%), and prior college GPA (10%). All other variables have missing data for fewer than 10% of observations. In the main analysis, I use mean substitution and also include a set of indicators for missing variables in order to include these observations in specifications with these control

[8] I use instructor-term groups in those cases where I do not know exact class groups. The multiple grade thresholds in these groups can lead to some fuzziness in the effect of threshold crossing on letter grade.

variables and keep the sample consistent throughout.[9]

## 2.4 Results

### 2.4.1 Descriptive analysis

I begin by examining the raw data. The relationship between STEM major and both number and letter grades are shown in Figure 2.1. There is a clear, positive correlation between grades and the probability of majoring in STEM. Visually, it appears that the slope is similar for men and women using either number or letter grades. I estimate this relationship for both types of grades and show the results in Table 2.2. I estimate a positive gradient for men ranging from 1.10–1.45 percentage points using number grade and 10.6–14.5 p.p. with letter grade. Interestingly, the slope for women is slightly shallower, suggesting that women are less responsive to grades if anything, but this difference is not statistically significant.

The main issue with this descriptive analysis is one of selection. Specifically, women in the lower tail of Figure 2.1 could have more perseverance or grit, which are unobservable to the researcher. This leads them to persist in STEM even though on average, women are deterred from majoring in these fields as the result of negative grade signals. As a result, I would estimate no difference between men and women here. To address these concerns, I use a regression discontinuity to compare otherwise similar students at letter grade thresholds.

### 2.4.2 Tests of the identifying assumption

With this regression discontinuity design, the identifying assumption is that all other determinants of major vary smoothly through letter grade thresholds. While I believe the institutional features described above indicate there is little reason to doubt this assumption *ex ante*, I also perform the standard statistical tests of this assumption. First, I present frequency histograms in Figure 2.2. Overall, the number of observations appears smooth across the threshold, which supports the

[9] I also plot the number of missing variables per observation as well as the percent of observations missing at least one variable, show in A.1. The binned data are smooth (if noisy) across the threshold, suggesting that the appearance and number of missing variables is unconcerning. I also demonstrate robustness of my main results using both a subsample of data containing no missing values as well as with the full sample using multiple imputation (Tables A.4 and A.5, respectively). I discuss this robustness exercise in greater detail in section 2.4.

notion that there is no bunching above the threshold which could occur due to manipulation of the running variable. There is a dip near the cutoff for men and women, but this dip is roughly equal both above and below the cutoff. This is because many instructors look for a natural break or gap in the final grade distribution and draw letter grade cutoffs there, which would only invalidate my design if instructors are also choosing cutoffs to push specific students above or below a cutoff. I return to this issue separately in the robustness section below, where I show results are similar when I exclude students up to 2 percentage points from the cutoff.

Additionally, I test whether observed covariates are smooth across the threshold. Specifically, I regress each variable on the left-hand side of equation 2.1 to verify there is no discontinuity at letter grade cutoffs. These include gender, transferred credit hours, prior GPA, high school rank, SAT scores, whether a student applied for financial aid, whether a student is new to the university or any college, same-semester GPA excluding courses in the sample, student classification (e.g. freshman), first-generation status, international student status, and transfer status. I plot these variables around the threshold in Figure A.2 and estimate them in Table A.2. While my main analysis focuses on men and women separately (panels B and C of the table, respectively), I also estimate for the pooled sample of all students together (panel A). I do this using the optimal bandwidth for the main outcome of the sample used in each panel. By and large, the visual evidence and point estimates show no discontinuity across the threshold, with a few exceptions. Across the 46 estimates in the table, 3 estimates are significant at the 10% level (gender, freshman, and sophomore all for the pooled sample), 2 at the 5% level (SAT scores for the pooled sample and transfer for women), and 1 at the 1% level (junior for women). Visually there is some evidence of a discontinuity for SAT max for both genders, and junior, senior, and transfer for women, but not for any of the other variables that appear significant in the table. This is roughly what one should expect to see due to random chance.

Relatedly, I also examine whether predicted outcomes vary smoothly across the threshold. Instead of considering individual covariates separately, this is a linear combination of these factors where the relative weights are chosen to best predict the outcome of interest. Specifically, I predict

STEM major using all predetermined characteristics and plot the predicted outcomes in Figure 2.3. The figure shows that predicted outcomes vary smoothly through the letter grade threshold, which is consistent with the identifying assumption.

### 2.4.3 Effect of threshold crossing on letter grade

Following these tests of the identifying assumption, I transition to the effect of numerical grade crossing the threshold on letter grade. This result is plotted in Figure 2.4. There is clear, visually compelling evidence that threshold crossing is associated with a higher letter grade, and that this effect is close to 1. I believe the main reason this discontinuity is not equal to 1 is that some instructors have multiple sections within the same semester that I do not observe. In those cases, it is possible that there are different thresholds across the sections, and since I do not observe the sections in these cases and group them as one class, it generates a "fuzzy" discontinuity. It is also possible that in rare cases, letter grades (but not number grades) are revised at the department or college level, or that there are administrative errors in the data.

### 2.4.4 Graduation

Graduation rate is an interesting outcome by itself, but it is also important because I only observe major for students who have graduated. It is possible that students respond to feedback signals at letter grade cutoffs in a way that impacts their probability of graduating. This could occur through choice of major or be independent from it. For example, students may be less likely to graduate overall as a result of missing the higher letter grade cutoff due to failing the class, losing a scholarship, or as the result of a poor GPA.

Results for six-year graduation rates are shown in Figure 2.5, and it does not appear that there is any effect on graduation for women. Results indicate there is no evidence of any effect on graduation for women. The figure for men is perhaps less clear, but I still do not find compelling evidence of a discontinuity.

### 2.4.5 STEM major

The main outcome of interest is whether students are more or less likely to graduate with a STEM major as the result of being just above or below a letter grade threshold. This result is plotted for men and women in Figure 2.6. This figure provides visual evidence that for any reasonable bandwidth selection or functional form, the estimated effect appears to be close to zero for both genders.

I formally estimate the effect of letter grade on propensity to major in STEM in Table 2.3. Estimates use a local linear regression and uniform kernel, with the optimal bandwidth determined according to procedures outlined by Calonico et al. [2014]. Column (1) is the base specification with the optimal bandwidth, which controls only for a local linear coefficient of running variable. In column (2) I add instructor and term fixed effects, and introduce additional student-specific control variables in column (3). The added controls include SAT scores, high school rank, and prior post-secondary GPA, student status as a first generation, transfer, or international student, and an indicator for whether the student applied for financial aid. As one might expect, adding control variables does not result in significant changes to the point estimates. In columns (4–6), show that the estimates are robust to alternative bandwidths. It is also worth noting that estimates are still not significant even with the added precision that wider bandwidths allow.

Estimates in Table 2.3 range from 1.8 to 3.0 percentage points for men and -0.8 to 1.5 p.p. for women, and none are statistically significantly different from zero. To interpret one estimate from column (2) as an example, a man achieving a higher letter grade at the margin in a core STEM course is 3.04 percentage points (4.2 percent) more likely to graduate with a STEM major within six years - although this is not statistically significant.[10]

While I do not find significant effects of letter grade on persistence in STEM majors for either men or women, I also find no evidence that results are different between genders. The point estimates for men are higher than for women across all columns of Table 2.3, but so are the baseline

[10]I also verify that results are similar when using a 4 and 8 year graduation rate. Estimates range from 0.1 – 3.1 and -0.7 – 2.1 percentage points for men and women, respectively, and none are statistically significant.

means. Looking again at column (2), the 95% confidence interval for men is (-1.3, 7.4) percentage points (-1.8%, 10.3%). This means the point estimate for women, 0.5 p.p. (0.9%), falls squarely within the confidence interval of the estimate for men.

### 2.4.6 Robustness

In this analysis, there might be some concern that the small number of instructors (sixteen in the main sample) means clustering in that dimension can lead to incorrect statistical inference. To address this possibility, I estimate wild bootstrap p-values and include them in Table 2.3 in the footer of each panel [Cameron, Gelbach, and Miller, 2008]. For men the p-values range from .063 to .417 and only one is significant at the 10% level. The p-values of estimates for women are higher, ranging from .404 to .998. It is clear from this analysis that the estimates are still not statistically significantly different from zero.

A second potential concern is that while instructors do look for a gap at which to draw grade cutoffs, the threshold may be drawn in part because of some student characteristic that is unobservable to the researcher. It is helpful to note that many of the classes in my sample have a large number of students and therefore it is less likely that instructors know individual students as well. Moreover, if the did this it would likely be limited to a student or two, which in this size sample is unlikely to generate bias of a meaningful magnitude. However, to address this issue, I perform a robustness exercise where I drop observations that are close to the threshold on either side, since these are the students who would be affected by such an issue. Specifically, I omit from the regression observations within 0.4, 0.8, 1.2, 1.6, and 2.0 percentage points of the threshold. Results are shown in Table A.3. Estimates range from -2.2 to 4.3 percentage points for men and -0.2 to 1.7 p.p. for women. Only one estimate is significant out of the 30 in this table, and that is a negative effect for men in column (7). I note this is inconsistent with all other estimates in this paper and even within this table, and given the number of coefficients is likely due to chance. Overall, estimates from this table are consistent with the main results in that neither men nor women are impacted by grade thresholds. This suggests that my main results are not affected by any issue of instructors selectively including or excluding certain students at letter grade thresholds due to unobservable characteristics.

I also show that my main results are robust using two alternative solutions to the missing variables in addition to mean substitution, although this only matters where controls are included since these are the only variables that are sometimes missing. Specifically, I estimate first using only the subsample of observations with no missing data in Table A.4, and second using multiple imputation in Table A.5. These exercises show that regardless of how missing control variables are handled, estimates still reflect that grades do not affect propensity to major in STEM.[11]

### 2.4.7 Heterogeneity

While previous results show there is no effect on average for either men or women, it is possible effects could be nonzero for alternative cutoffs, subgroups, or outcomes. To address this, I first consider letter grade cutoffs separately rather than the stacked threshold. Results are shown in Figure A.3, with a wide range on the horizontal axis to show how estimates might look with different bandwidths. While most figures show no significant effect – including all of the graphs for women – there is some suggestive evidence that men are sensitive to letter grades at the C/D cutoff. Since this is the cutoff that represents the pass/fail mark for required-in-major courses at this university, it would make sense that this "hard" cutoff should matter more than others for mechanical reasons if not behavioral ones.

Corresponding estimates for each cutoff are shown in Table A.6. In column (8) of this table, which includes a local linear regression as well as instructor and term fixed effects, I estimate that men earning a C versus D at the margin are 10.8 percentage points (20.2 percent) more likely to graduate with a STEM major within six years. I do also estimate an effect of 4.1 percentage points (4.7 percent) for men at the A/B cutoff, though the visual evidence appears to be much less compelling. Additionally, given the large number of estimates reported in Table A.6, I note that

[11] With the multiple imputation command I cannot compute standard errors using two-way clustering as above. Table A.5 uses standard errors that are only clustered in one dimension by instructor. With larger bandwidths in columns 4 and 5, there are two statistically significant estimates for men that letter grades affect STEM participation by 3 p.p., but results for the main outcome in Figure 2.6 do not look visually compelling. Additionally, a check of my main results with one-way clustering suggests that this can lead to standard errors that are roughly 17% larger than two-way clustering, so p-values in Table A.5 may be too small and potentially misleading. Importantly, results for women are consistently close to zero regardless of how I handle the missing control variables.

some will be statistically significant due to chance. Importantly, there is no evidence of an effect for women at any individual letter grade cutoff, as estimates range from -3.5 to 4.2 p.p., none of which are significant. This suggests that women are not deterred from STEM by poor letter grades even at the crucial C/D threshold.

The null result also holds for a variety of subgroups where an effect seems most likely. I show results for these groups in Figure A.4 and Table 2.4, where the first column simply repeats the main specification from above for comparison. First, I look at courses taught by female and male instructors separately, since the interaction of instructor gender with student gender at letter grade thresholds might be more meaningful. I do not find compelling visual evidence of a discontinuity for men or women taught by either male or female instructors, although the estimate of 5.6 percentage points (8.5%) in column 2 is marginally significant for women in STEM courses taught by women. Estimates for men are 2.9 - 3.4 p.p. for both instructor genders, and -0.6 p.p. for women in classes taught by men, though none of these are significant.

Next, I consider only individuals who are first generation college students in their family, since this is typically a more vulnerable group which may be more sensitive to letter grade signals. Again, there is no compelling visual evidence that grades matter for this group. However, for male first generation students, there is marginal significance for the estimate that a higher grade makes them 4.5 percentage points (8.2%) more likely to major in a STEM field. The estimate for women here is -2.7 p.p but not significant. Again, I interpret even the marginally significant coefficient cautiously given the number of tests shown for different cutoffs and different groups.

I also consider freshman students who should have a lower cost to switch majors, and may therefore be more sensitive to letter grades. This subsample consists of a greater proportion of observations from biology and calculus classes,[12] with less representation from traditionally second-year courses such as organic chemistry or computer programming. The figure shows no evidence of an effect for either men or women, and the estimated results are not significant at 2.9 and -1.2 p.p. for men and women, respectively.

[12] About 90% of the observations in this subsample come from biology, calculus I, and calculus II.

The gender composition of economics majors and the factors affecting this is perhaps of particular interest to economists. While my previous analysis has focused on STEM courses and majors as defined by the Department of Homeland Security, I also collected data from both micro principles and intermediate microeconomics. For these courses, I show the data around the letter grade threshold in Figure A.5, with the outcome being graduating with an economics major (subfigure a) or business major (subfigure b). Business majors are interesting because many of them take economics courses but also because anecdotally it is fairly common for students to transfer into the business program if they achieve a high GPA i.e. good letter grades. This implies that higher letter grades in economics courses might actually lead to a lower probability of majoring in economics. However, I do not find visually compelling evidence that letter grades matter for either economics or business majors from these figures. Corresponding estimates are shown in Table A.7, most of which are not statistically different from zero. One result is marginally significant for men having a business major in column 4 (6.8 percentage points or 22 percent), but I do not consider this a meaningful result given the lack of compelling visual evidence and the fact that the significance (and effect size) go away with the addition of control variables. Overall, the evidence indicates that letter grades in economics classes do not impact women's decisions to major in economics or business.

Finally, I consider the impact of letter grades for other subgroups on other majors as well as estimated annual earnings in Figure A.6 and Table A.8. In subfigure (a) and column (2), I expand the course pool to include economics classes and look at the effect of letter grade on the probability of graduating with a STEM or economics degree within six years. In subfigure (b)/column (3), I use only science courses including biology, calculus, and chemistry, and look at the impact on majoring in a science field. Then I use only engineering courses including calculus, computer programming, and statics and look at the impact on majoring in engineering in subfigure (c)/column (4). Estimates in the table range from -3.3 to 2.9 percentage points for men and 0.4 to 4.1 p.p. for women but are largely not statistically significant. The only estimate significantly different from zero are for women in science courses majoring in science (4.1 p.p.). Visually, there may be a discontinuity for women in science, but I interpret this cautiously given the number of different results by subgroup shown in this section. There is no compelling visual for any of the other cases described in this exercise.

It is possible that students transfer to different majors that are less competitive but still considered STEM fields, in which case I would estimate a zero with my main specification. In an attempt to address this concern, I match average annual earnings by major to my data [Carnevale et al., 2015].[13] I use the STEM courses from the main sample (without economics) and show this exercise in subfigure (d)/column (5) of Figure A.6 and Table A.8. I estimate the impact of letter grade thresholds on national median earnings by major (in 2013 dollars) is $1,156 for men and $500 for women. This estimate is significant for men, but I do not find compelling visual evidence of an effect. Notably, the estimate for women is not statistically significant.

In summary, I show that by and large letter grades do not affect participation in STEM, even among most subgroups and at most grade thresholds. While there are some exceptions – for example, men at the A/B and C/D thresholds, women in courses taught by women, male first generation college students, or women in science courses majoring in science – most of these results are not visually compelling and need to be interpreted cautiously given multiple inference concerns. Thus, my overall conclusion is there is little compelling evidence that letter grade signals deter either men or women from STEM majors across a wide range of different grade thresholds, subgroups, and outcomes.

### 2.5 Discussion and Conclusion

In this paper, I provide the first causal evidence of the impact of letter grades on STEM major. My data contains 21,000 observations from seven distinct, core STEM courses at a large university and includes final numerical as well as letter grade. This allows me to utilize a regression discontinuity at letter grade thresholds. The identifying assumption of this design is that all other determinants of college major besides letter grade vary smoothly through grade cutoffs. With this

[13] It is difficult to match less common majors that may not exist at this university or do not exist in national average income estimates.

in mind while collecting data, instructors were interviewed individually and each stated that they did not allow manipulation across grade thresholds. Additionally, a rich set of covariates including those representing prior academic ability allows me to test for and find evidence in support of the identifying assumption.

Results indicate that negative letter grade signals do not contribute to the gender gap in STEM. Neither men nor women are deterred by poor letter grades relative to their peers on the other side of the cutoff who are otherwise similar. These results are robust to a variety of alternative specifications. Furthermore, these results are relatively precise. Using baseline means and point estimates from column (2) in Table 2.3, I can rule out that a higher letter grade affects STEM participation for women by more than 6.0 percent (3.2 percentage points).

Importantly, I can rule out effects of the magnitude found by researchers in other contexts. For example, Owen [2010] estimates that women who received an A at the margin in principles of economics courses are 15 percentage points ($\sim$167%) more likely to major in economics. This is nearly five times higher than the maximum percentage point effect I can rule out for generalized STEM courses, and even higher when compared in percent terms. Even when I restrict my sample to only consider economics courses and majors, I rule out an effect bigger than 6.2 percentage points (9.8%) for women, which is still significantly smaller than the effect Owen finds. I can also rule out effects that are similar to the effect of same-gender faculty on the STEM major decisions of females. Carrell, Page, and West [2010] find that women taught by all-female faculty in their math and science courses are 15.5 percentage points ($\sim$40%) more likely to graduate with a STEM degree, which is more than four times the upper bound of the 95% confidence interval from my study. I can also rule out that letter grades matter as much as the impact of exposing students to a female role model. Porter and Serra [2020] study what happens when a "charismatic career woman who majored in economics at the same university" talks to principles of economics classes. They find that the visits increased the likelihood of women majoring in economics by 8 percentage points or a roughly 100 percent increase from the baseline mean.

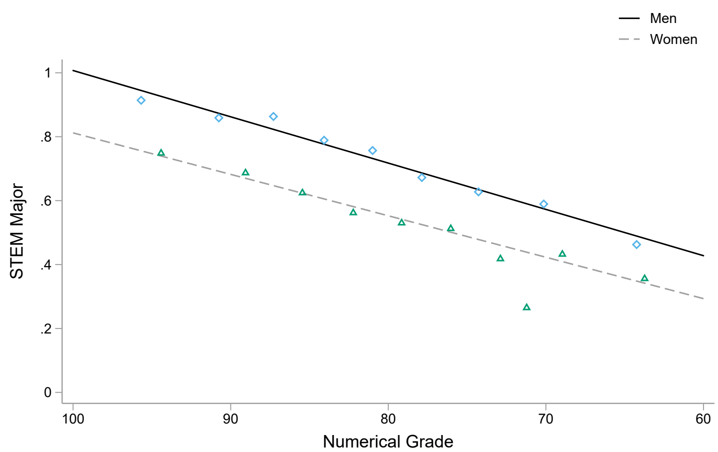
In this paper, I also look at effects across various types of subgroups. These include female or male instructors separately, first generation college students, and freshman. I then consider the impact of grades in economics courses on economics or business majors, science courses on science majors, engineering courses on majoring in engineering, as well as the effect of letter grade thresholds on estimated annual earnings. I find no compelling evidence that letter grades matter in these cases, except perhaps for suggestive evidence for women in science courses. I also look at each letter grade threshold alone, and find no effect for women even at the C/D cutoff which represents the difference between passing and failing within a STEM major at this university.

Overall, these results suggest that negative grade signals - including those at the most important grade thresholds and within the most susceptible groups - are not responsible for the large gender gap in STEM. This highlights the importance of understanding other factors that are important in the gender gap in STEM majors and the corresponding gender gap in earnings.

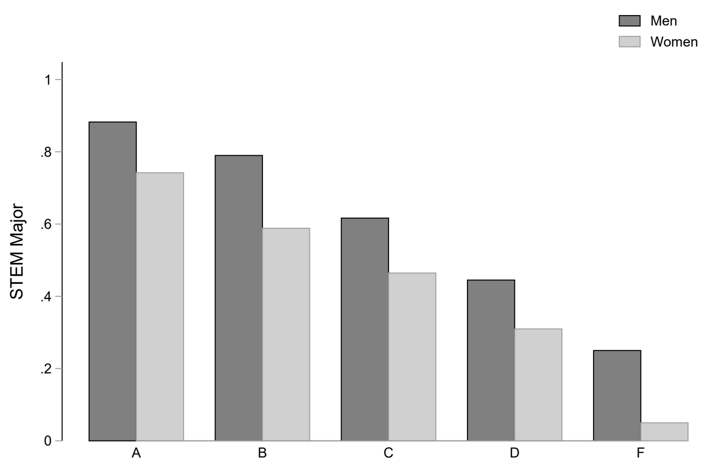
## 2.6 Figures

Figure 2.1: Descriptive Analysis

(a) Probability of STEM major by number grade

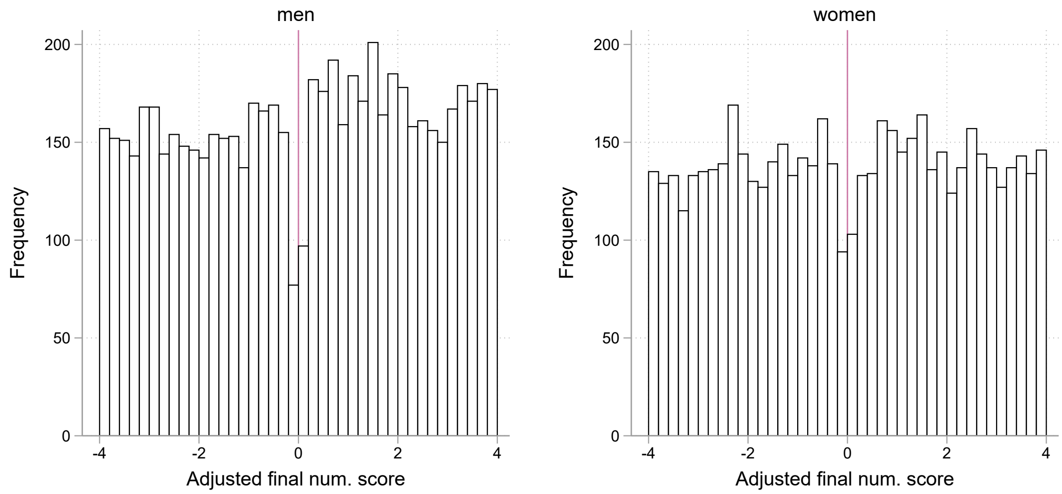



(b) Probability of STEM major by letter grade



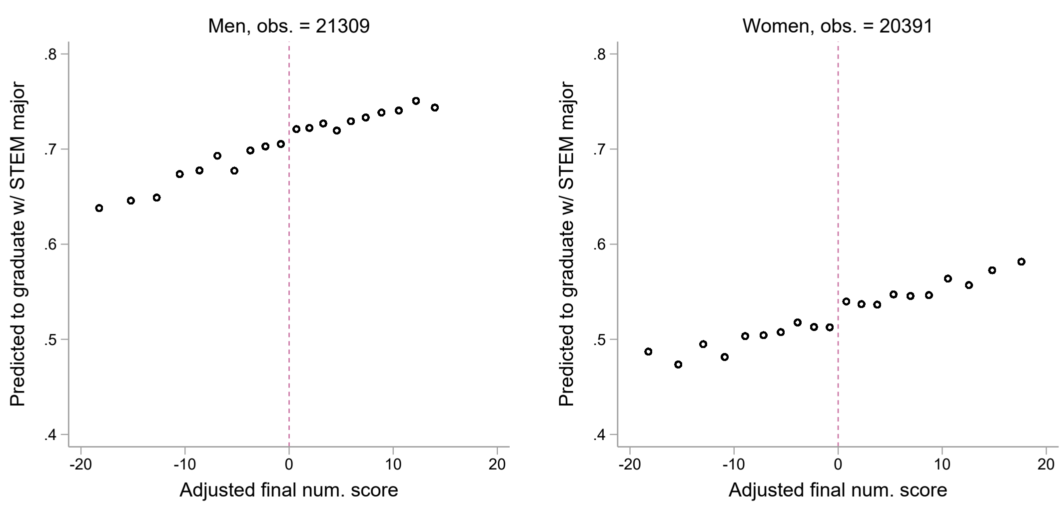

This figure shows the correlation between grades and STEM major using both number and letter grades. I estimate this relationship in Table 2.2. Note that higher grades are on the left. For this figure, I restrict to observations where the student earned a numerical score of at least 60.

Figure 2.2: Observations Histogram



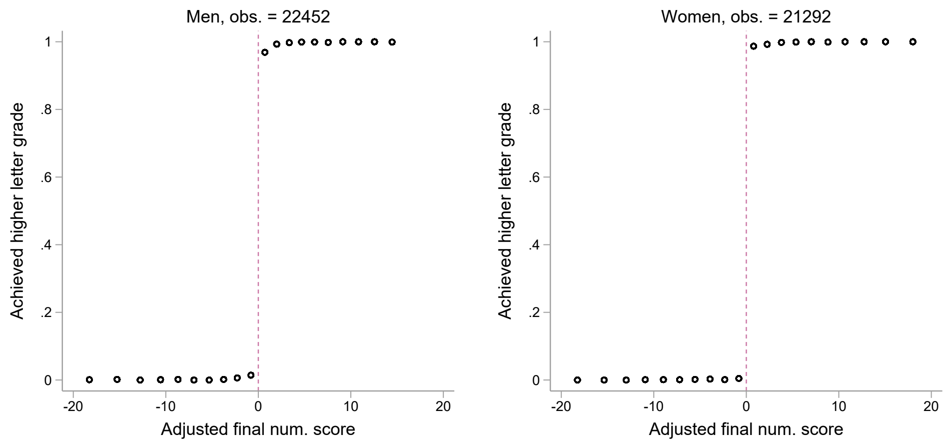

This figure shows the frequency (count) of observations close to the stacked regression discontinuity threshold for men and women in the sample STEM courses. The figure shows no unusual heaping of observations just above the threshold for either men or women, which if present could be a concern. There is a dip in observations very close to the threshold, which is consistent with what we know about instructors looking for a natural gap in the grading distribution at which to draw letter grade cutoffs.

Figure 2.3: Test of Identifying Assumption: Predicted to graduate with a STEM major within six years



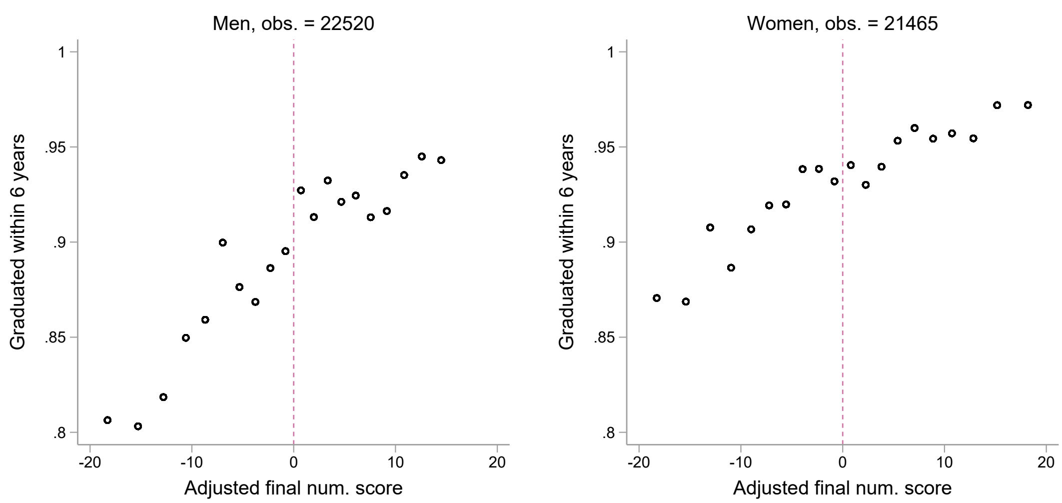

This figure shows predicted outcome for the main outcome of interest, graduated with a STEM major within six years. As a test of the identifying assumption, I predict the main outcome using only pre-determined characteristics. This predicted outcome varies smoothly through the stacked letter grade cutoff for both men and women, in support of the identifying assumption. Each point represents an equal number of observations.

Figure 2.4: Effect of Number Grade Crossing Threshold on Receiving Higher Letter Grade



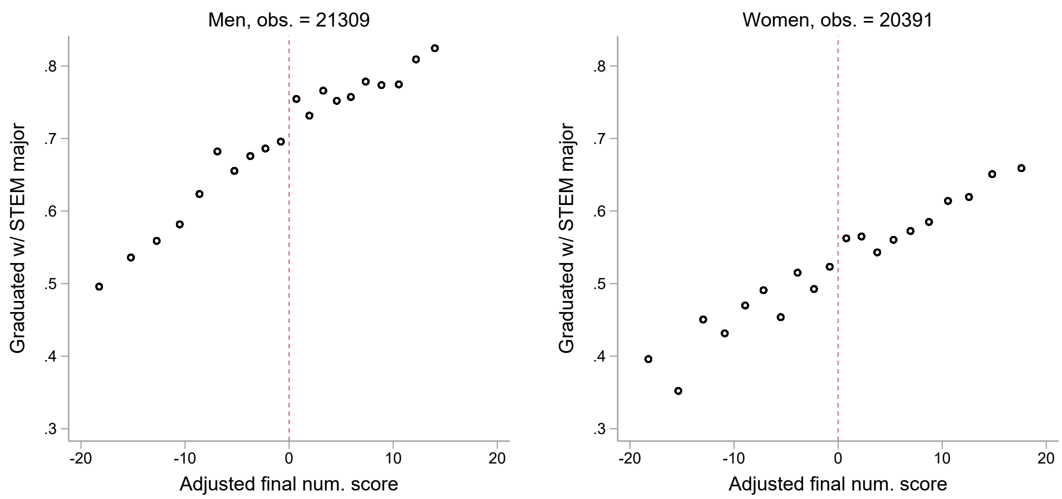

This figure shows effect of final numerical grade threshold crossing on final letter grade. Each point represents an equal number of observations.

Figure 2.5: Effect of Letter Grade Thresholds on Six-year Graduation Rate




This figure shows six-year graduation rate for students in the main sample. Graduation is an important outcome by itself, but especially so with these data because I only observe major for students who graduated. Visually there does not appear to be any significant effect of letter grade thresholds on the probability of graduation for either gender. Each point represents an equal number of observations.

Figure 2.6: Effect of Letter Grade Thresholds on Graduating With a STEM Major




This figure shows the main outcome of interest, graduated with a STEM major within six years. The point of this figure is to show the data and what treatment effect might exist, if any, for different bandwidths or functional forms. It is clear that for any reasonable bandwidth, there is no effect for either men or women. I estimate effects formally in Table 2.3. Each point represents an equal number of observations.

## 2.7 Tables

Table 2.1: Summary Statistics

| | All students | Men | Women |
|---|---|---|---|
| female | 0.46 | 0.00 | 1.00 |
| HS stu. rank | 56.39 | 64.64 | 48.19 |
| max(SAT score, ACT conc. score) | 1,223.05 | 1,249.63 | 1,198.56 |
| prior gpa | 2.37 | 2.43 | 2.34 |
| transfer hours | 24.80 | 25.04 | 24.71 |
| appl. fin. aid | 0.58 | 0.58 | 0.57 |
| 1st gen. stu. | 0.22 | 0.21 | 0.22 |
| freshman | 0.24 | 0.22 | 0.25 |
| sophomore | 0.44 | 0.44 | 0.44 |
| transfer stu. | 0.07 | 0.09 | 0.06 |
| intl. stu. | 0.02 | 0.02 | 0.01 |
| grad. in 6 yrs | 0.88 | 0.86 | 0.90 |
| grad. w/ STEM major | 0.58 | 0.66 | 0.50 |
| Observations | 21,533 | 10,790 | 9,188 |

This table shows summary statistics for all students in the main sample, and also broken down for men and women separately. Because many students do not take both the SAT and ACT, I use official concordance tables to convert ACT scores to comparable SAT scores, then take the max of SAT score and ACT concordance score where both exist. (See Table A.1 for more information on missing variables.)

Table 2.2: Descriptive Analysis: Probability of STEM major and grade

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Number grade | 0.0145*** | 0.0110*** | | |
| | (0.0014) | (0.0012) | | |
| Female × Number grade | –0.0015 | –0.0007 | | |
| | (0.0015) | (0.0011) | | |
| Letter grade | | | 0.1446*** | 0.1060*** |
| | | | (0.0174) | (0.0115) |
| Female × Letter grade | | | –0.0039 | –0.0059 |
| | | | (0.0160) | (0.0108) |
| Observations | 16,334 | 16,334 | 16,326 | 16,326 |
| Control vars. | | Y | | Y |

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ . This table shows results a naive regression model estimating the relationship between grades and STEM major using both number and letter grades. Standard errors are in parentheses and clustered at the instructor level. For this table, I restrict to observations where the student earned a numerical score of at least 60.

Table 2.3: Effect of Letter Grade Thresholds on Graduating With a STEM Major

Panel A: Men

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Above letter grade cutoff | 0.0270 | 0.0304 | 0.0260 | 0.0300 | 0.0296 | 0.0177 |
| | (0.0298) | (0.0224) | (0.0272) | (0.0221) | (0.0171) | (0.0144) |
| Observations | 8,654 | 8,654 | 8,654 | 10,698 | 12,586 | 15,965 |
| Outcome mean | .717 | .717 | .717 | .719 | .718 | .713 |
| Wild bootstrap p-value | .417 | .232 | .284 | .122 | .0627 | .125 |
| Instructor & term FEs | | Y | Y | Y | Y | Y |
| Other control vars. | | | Y | Y | Y | Y |
| Opt. Bandwidth = 5.72 | 1x | 1x | 1x | 1.25x | 1.5x | 2x |

Panel B: Women

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Above letter grade cutoff | 0.0154 | 0.0045 | –0.0001 | 0.0032 | 0.0020 | –0.0078 |
| | (0.0174) | (0.0146) | (0.0158) | (0.0141) | (0.0104) | (0.0097) |
| Observations | 10,078 | 10,078 | 10,078 | 12,246 | 14,267 | 17,778 |
| Outcome mean | .528 | .528 | .528 | .528 | .529 | .529 |
| Wild bootstrap p-value | .433 | .771 | .998 | .809 | .823 | .404 |
| Instructor & term FEs | | Y | Y | Y | Y | Y |
| Other control vars. | | | Y | Y | Y | Y |
| Opt. Bandwidth = 7.80 | 1x | 1x | 1x | 1.25x | 1.5x | 2x |

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ . This table shows estimates of the effect of letter grade on the main outcome of interest, graduation with a STEM major within six years. Column (1) is the base specification using 1x the optimal bandwidth. Column (2) includes instructor and term fixed effects. Column (3) further includes available student-specific control variables. In columns (4–6), I increase the bandwidth. Estimated with a local linear regression and uniform kernel. Standard errors are in parenthesis and clustered at the instructor and term level. I also include wild bootstrap p-values in the footer of each panel since the number of instructors is low (sixteen), but results are similar.

Table 2.4: Results for Subgroups Where Effect Seems Most Likely (Outcome: Graduated with a STEM major)

Panel A: Men

| | (1) Main result | (2) Female instr | (3) Male instr | (4) First gen | (5) Freshmen |
|---|---|---|---|---|---|
| Above letter grade cutoff | 0.0304 | 0.0344 | 0.0292 | 0.0454* | 0.0292 |
| | (0.0222) | (0.0237) | (0.0228) | (0.0254) | (0.0270) |
| Observations | 8,654 | 4,289 | 10,904 | 3,775 | 3,185 |
| Outcome mean | .717 | .813 | .498 | .556 | .567 |
| Instructor & term FEs | Y | Y | Y | Y | Y |
| Opt. Bandwidth = | 5.72 | 9.35 | 6.97 | 9.58 | 10.6 |

Panel B: Women

| | (1) Main result | (2) Female instr | (3) Male instr | (4) First gen | (5) Freshmen |
|---|---|---|---|---|---|
| Above letter grade cutoff | 0.0045 | 0.0556* | –0.0064 | –0.0270 | –0.0124 |
| | (0.0139) | (0.0273) | (0.0138) | (0.0170) | (0.0173) |
| Observations | 10,078 | 1,834 | 8,568 | 2,227 | 3,239 |
| Outcome mean | .528 | .653 | .393 | .405 | .401 |
| Instructor & term FEs | Y | Y | Y | Y | Y |
| Opt. Bandwidth = | 7.8 | 7.49 | 5.79 | 5.97 | 10.74 |

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ . This table shows estimates for alternative samples where an effect seems most likely. Namely I look at only courses taught by female and female instructors (columns 2 and 3, respectively), and only students that are first generation or freshmen (columns 4 and 5). I plot these data in Figure A.4. All models include instructor and term fixed effects. Estimated with a local linear regression and uniform kernel. Standard errors are in parenthesis and clustered at the instructor and term level.

# 3. ECONOMIC WARNINGS: THE IMPACT OF NEGATIVE NEWS ON RACIAL ANIMUS

## 3.1 Introduction

Economic pressure brought by events such as the Great Recession [Yagan, 2019], NAFTA [Hakobyan and McLaren, 2016], rising import competition from China [Acemoglu, Autor, Dorn, Hanson, and Price, 2016], and globalization in general [Ebenstein, Harrison, McMillan, and Phillips, 2014] have resulted in some U.S. companies reorganizing their operations, or shutting down entirely. These decisions affect a large fraction of their workforce, often resulting in mass layoffs and plant closings. News of the sudden unemployment of a large number of workers presents a significant disruption to the fabric of affected local communities. In addition, racially charged rhetoric surrounding such events often points to the shifting of “American jobs” to foreign workers (e.g. Preston, 2015; May, 2020).

In this context, linking information about impending mass layoffs to racial animus seems straightforward: as communities learn of workers losing their source of income and social status, local tensions and animosity toward certain groups rise [Falk, Kuhn, and Zweimüller, 2011]. Similar responses to news have been found in criminal sentencing (e.g., Philippe and Ouss, 2018; Eren and Mocan, 2018). However, more closely related work on negative economic shocks arising from differential impacts of the Great Recession [Anderson, Crost, and Rees, 2020], and trade competition (Ortega, Di Fruscia, and Louise, 2021 and DiLorenzo, 2021), find mixed results. Anderson, Crost, and Rees [2020] and Ortega, Di Fruscia, and Louise [2021] both find that layoffs increase animus, as measured by either internet searches for racial slurs, or hate crimes against Blacks. On the other hand, DiLorenzo [2021] only finds effects on the number of hate crime groups, but not hate crimes.

The purpose of this paper is to determine the extent to which information about impending mass layoffs and plant closings lead to animosity toward racial minorities. We overcome several obstacles to answering this research question. First, aggregate economic data, such as county-level employment rates, is often not granular enough to reflect job loss shocks. In addition, disentangling the information content from the experienced job loss is difficult, because local employment data typically reflect changes in current or past local labor market situations. Second, racial animosity is also hard to measure, largely because people are understandably hesitant to reveal "socially unacceptable attitudes" [Stephens-Davidowitz, 2014]. Lastly, causal identification can be hampered by endogenously different labor market conditions across communities and by the difficulty of attributing crimes and other actions to racial animosity. To address these issues, we assemble a dataset containing a broad information set of notices about impending mass layoffs, and link this to two different measures of racial animus.

Specifically, we leverage the timing and distribution of Worker Adjustment and Retraining Notification (WARN) Act Notices of mass layoffs and plant closings to identify their impact on racial animus. Using these data has several advantages. First, we are able to determine the location and timing of impending layoff events with great detail. This is useful because it allows us to separate the impact of the information from the impact of the layoff itself, as done by Carlson [2015] in his study of the health effects of WARN Notices. Second, because some mass layoff events may affect only a small fraction of the local labor force, our data can pick up smaller labor shocks than may be seen in county-level unemployment rates. Lastly, Krolikowski and Lunsford [2020] show that WARN Notices are a strong leading indicator of local labor market indicators, such as unemployment insurance claims. We assemble data on all available WARN Notices of mass layoffs and plant closings for 23 states.[1] Our compiled data contains over 75,000 notices of mass layoff events that happened between the years 2004 and 2020. Although there is some dispersion, the average layoff event affects 100 workers.

Measuring racial animus is complicated by issues of underreporting and attribution. We draw on two data sources to address these difficulties: Google Trends data on searches for racial slurs and FBI Hate Crime Statistics. Google Trends data were first used to measure racial animus by Stephens-Davidowitz [2014], who showed that internet searches for the most commonly used anti-

[1]These are 60-day advance notices that large firms (>100 employees) are required to give to workers ahead of a layoff event affecting more than 50 workers at a single employment site.

Black slur were negatively correlated with voting for President Obama in 2008. In this paper, we exploit the ability of Google Trends to also report data on internet searches for other racial slurs, specifically those commonly used to refer to Asians, Hispanics, as well as two placebo search terms. Google Trends reports an index of search activity for specified search terms over time at the Designated Market Area (DMA) level. This is useful because the DMA level is likely where information about mass layoff notices also propagates through broadcast media or other community pathways.[2]

Although we have reason to believe that online behavior can be both harmful on its own and also reflect behavior in other spheres [Müller and Schwarz, 2020], we also corroborate our findings using data from the FBI Hate Crime Statistics program. The data contain incident-level information on crimes that are determined to have been motivated by bias against specific groups. Using these data allows us to provide further evidence of the increased harm associated with racial animus, and capture a broad picture of any potential increase in animus resulting from mass layoffs. We view these animosity measures to be linked, as recent studies suggest that anti-minority sentiment online is predictive of hate crimes against minorities (Müller and Schwarz, 2020 and Müller and Schwarz, 2021). In using both internet searches and hate crimes, we aim to measure both racially targeted harmful rhetoric and criminal behavior.

Our empirical strategy is to compare how the values of our animosity measures change in counties where mass layoff notices occur relative to other counties, controlling for month-year and county fixed effects. Because mass layoff notices can occur several times for each county during our study period, we mainly focus on the the contemporaneous effect of mass layoff notices on animosity. Our approach is similar in spirit to that taken by other studies that examine the effect of news shocks on a variety of outcomes (e.g. Eren and Mocan, 2018; Philippe and Ouss, 2018 – criminal sentencing, and Carlson, 2015 – birth outcomes). We also show that the contemporaneous effect captures most of the treatment effect, as effects dissipate within two months. The identifying assumption is that absent mass layoffs, counties would have experienced changes in these measures

[2]We have made the editorial decision to not write out the racial slurs used in this paper. The slurs we use come from Anderson and Lepore [2013].

of animosity similar to what other counties experienced. We estimate falsification exercises to show evidence for this assumption, demonstrating that future layoffs do not affect past outcomes. In other specifications, we also account for state-month shocks and county-specific time trends.

Estimates indicate a significant increase in racial animosity following WARN Notices of impending mass layoffs. In particular, we estimate a 1.4 percent increase in the search rate for racial slurs. While this outcome does not necessarily imply animosity targeted at some group, we also estimate a 23 percent increase in the number of hate crimes committed within a month of the WARN Notice. Thus, we interpret our results to mean that there is a sharp increase in racial animosity with real, harmful effects.

Exploiting the ability of our data to identify the target of the animosity, we also estimate effects on animosity toward different racial minorities. We find that our results are driven by increased animosity against Blacks and Hispanics specifically, rather than Asians. Importantly, we do not find any evidence of reverse causality – future layoffs do not affect animosity measures today. We also do not find any evidence of increased internet searches for placebo terms like "weather," or a slur commonly used to refer to rural Whites. Finally, we do not find any evidence of increased hate crimes against Whites, LGBT, or people with disabilities.

Taken together, these reinforce our interpretation that it is the mass layoff notices that cause the increase in racial animosity toward minorities, and that we are not simply picking up increases in general animosity across groups. Our results are all the more noteworthy because the WARN data we leverage exists, by definition, to soften the blow of mass layoffs. Yet we still estimate an increase in internet searches for racial slurs. Combined with our hate crimes estimates, we conclude that information about impending mass layoffs cause an increase in harmful rhetoric and criminal behavior toward minorities. Our estimates are robust to including controls for state-month shocks, time-varying demographic characteristics, and to allowing counties to follow different trends over time.

Our results contribute to a nascent literature that seek to causally identify the effects of various local labor shocks on racial animus in the US (Anderson, Crost, and Rees, 2020; DiLorenzo, 2021; Ortega, Di Fruscia, and Louise, 2021). These papers each focus on different reasons for layoffs. Anderson, Crost, and Rees [2020] focus on layoffs due to the Great Recession, exploiting cross-state variation in pre-Recession sector composition. Ortega, Di Fruscia, and Louise [2021], who focus on import exposure to China, and DiLorenzo [2021], who uses the universe of layoffs reported to the Trade Adjustment Assistance Database, both focus on layoffs due to trade. Our paper complements these studies by drawing on a wider and more inclusive set of mass layoffs and by separating the timing of the information about impending layoffs from the layoff events.

Through our use of internet searches from Google Trends to corroborate our hate crime results, we also join an emerging literature that leverage novel internet and social media data to measure individual attitudes, and how they relate to observed behavior. Anderson, Crost, and Rees [2020] use internet searches for a commonly used anti-Black slur, from Google Trends, to measure animus. Similarly, a number of papers have used data from Twitter to examine the impacts of social media personalities and access on political behavior (e.g., Fujiwara, Müller, and Schwarz, 2021; Giavazzi, Iglhaut, Lemoli, and Rubera, 2020; Müller and Schwarz, 2020) and Facebook access on hate crimes [Müller and Schwarz, 2021].

This paper proceeds as follows. Section 3.2 presents our data on layoff notices and racial animus. Section 3.3 lays out our empirical strategy and placebo exercises. In Section 3.4, we present our estimates of the effect of mass layoff notices on our two measures of racial animus. We also show placebo estimates of the effect of mass layoff notices on anti-LGBT, anti-disabilities, and anti-White animus. Section 3.5 concludes.

## 3.2 Data

### 3.2.1 Layoff Data from WARN Notices

The Worker Adjustment and Retraining Notification (WARN) Act of 1988 provides protections against sudden termination for employees of large firms. Its main provision is the requirement of a 60-day notice to workers, local chief executive, and State Dislocated Worker Unit, ahead of a planned mass layoff or plant closing. The objective of the advance notice is to give workers time to find another job or to enter retraining programs [Bartell, 2001]. When these filings are made, the local community is also made aware of the impending mass layoffs by the various news media outlets that pick up the story.[3]

We construct our data from the WARN Notices from 23 states whose State Dislocated Worker Units maintain historical records.[4]. The notices include employer locations, WARN Notice date, and for some states, the date of the layoff or plant closing and the number of affected workers. However, data availability for some variables is inconsistent, so we focus on the location and notice date. The assembled data contain WARN notices for 75,574 mass layoff events, affecting an average of 105 workers each event. On average, counties experience around 6 mass layoff events during the study period, with the affected workers representing a small fraction of the local labor force, 0.3 percent. We do not observe whether the mass layoffs actually take place, or whether they involve fewer workers than originally stated in the WARN Notice. We take 60 days after the WARN Notice to be the layoff date, and consider the number of affected workers stated in the WARN Notice to be the number of laid off workers.

### 3.2.2 Google Trends

Google Trends is a tool that provides an index of search activity for specified search terms. We follow Stephens-Davidowitz and Varian [2014] and Anderson, Crost, and Rees [2020] in using searches for a commonly used anti-Black slur as a measure of racial animus. We also obtain internet search data for the slurs used to refer to Asians, Hispanics, Whites, and the word "weather."The last two are placebo search terms where we do not expect to see an effect from mass layoffs. Since

[3] A quick internet search using the keywords "WARN Act mass layoff news" will show many news articles based on WARN Act notice filings made by firms, for example: https://www.pennlive.com/news/2022/03/zulily-informs-state-that-504-people-will-be-laid-off-when-pa-facility-closes.html; https://www.postcrescent.com/story/money/2022/03/10/waupaca-elevator-employees-still-facing-mass-layoffs-after-second-warn-notice/6986906001/; https://www.wmbfnews.com/2022/03/12/hundreds-jobs-danger-upstate-fulfillment-center-plans-close/; https://www.siouxlandproud.com/news/local-news/121-tur-pak-foods-employees-to-be-laid-off-with-closure-of-sioux-city-facility/; https://www.gainesville.com/story/news/2021/12/20/gainesville-medical-device-plant-lay-off-more-than-500-employees/8975211002/

[4] A list of coordinators for these offices can be found at: https://www.dol.gov/agencies/eta/layoffs/contact

the smallest geographic level available from Google Trends is the DMA (Designated Market Area), we extract data at the DMA-month level. We then link the DMA-month internet searches data to the county-level layoffs data. Because DMAs are larger than counties, we assign the internet search activities observed at the DMA level to all counties within that DMA, and then weight by county-level population.

The data that we obtain from Google Trends are an index – for a chosen geography, it measures the fraction of searches that include the specified search term relative to the total search volume at that time. In essence, the raw data from Google Trends is a relative, rather than absolute, measure of search activity. To be able to use this data in our analysis, we follow previous studies and use the logarithm of the reported search rate [Anderson, Crost, and Rees, 2020].

The data have a few limitations, which we account for in this paper. First, Google Trends data are drawn from a sample of all Google searches, using searches that are cached each day. The reported search index are also averaged to the nearest integer. We account for these potential issues by pulling data from Google Trends five times on separate days, and then taking the average of the reported search indices. Second, Google Trends data have a privacy threshold – if search volume for a given area during the specified time period is below the threshold, Google Trends will report a zero. This is one reason we limit to the more commonly-used racial slurs. Finally, Google Trends reports the search activity for racial slurs, but we do not directly observe the intent behind the searches. Using FBI data on hate crimes allow us to speak to race-motivated behavior.

### 3.2.3 FBI Hate Crime Statistics

Our second measure of animus is hate crime data from the FBI Hate Crime Statistics, part of the Uniform Crime Reporting program. Hate crimes are defined according to the Hate Crime Statistics act as crimes motivated by bias based on "race, gender or gender identity, religion, disability, sexual orientation, or ethnicity." Hate crimes reported to the FBI include both violent and property crimes. The most prevalent hate crimes tend to be destruction/damage/vandalism, intimidation, and simple assault [Masucci and Langton, 2017].[5]

Incident-level hate crime data, reported by law enforcement agencies to the FBI, are available for offenses that occurred from 1991 onward. The data contain the date of the incident, the nature and motivation behind the offense, and information about the offender and victim. Because the data contain the motivation behind each hate crime, we are able to measure effects for hate crimes against different races separately. We also use hate crimes that are not racially motivated to conduct placebo exercises. Specifically we focus on hate crimes against Asians, Blacks and Hispanics in our main results, and use hate crimes against Whites, LGBT, and people with disabilities as placebo outcomes. We use the Law Enforcement Agency Identifiers Crosswalk to link the hate-crime data to the county-level layoffs data.[6]

Our use of hate crimes to corroborate our results from using internet searches is in line with their use with previous studies on the determinants and correlates of hate crimes.[7]. Like previous studies, we acknowledge the potential limitations of hate crime data: that classifying crimes as motivated primarily by hatred is difficult, and consequently, may not be reported across jurisdictions consistently. This would pose a problem if local law enforcement agencies become more likely to push for a hate crime classification for crimes committed immediately after a WARN Notice, even in the absence of a true increase in "real" hate crimes. In using hate crime as a secondary outcome, we only seek to provide further context for our primary animosity measure, internet searches for racial slurs. As we will show in Section 3.4, we estimate consistent results across subgroups for both internet searches for racial slurs and for hate crimes, which alleviates our concerns about using hate crimes as an animosity measure.

[5]More information about hate crimes, including the types of biases considered, can be found at: `https://www.fbi.gov/services/cjis/ucr/hate-crime`

[6]`https://www.icpsr.umich.edu/web/NACJD/studies/35158`

[7]e.g., Kaushal et al. [2007], Ryan and Leeson [2011], Ryan and Leeson [2011], Anderson et al. [2020]

### 3.3 Empirical Strategy

We estimate the effect of mass layoffs on the various measures of animosity using the following general specification:

$$y_{it} = \alpha_i + \alpha_t + \beta \cdot X_{it} + \gamma_0 \cdot \text{Treated}_{it} + \gamma_1 \cdot \text{Treated}_{i,t-1} + \gamma_2 \cdot \text{Treated}_{i,t-2} + \epsilon_{it} \tag{3.1}$$

The variable $y_{it}$ takes the value of the animosity measure for county $i$ in month $t$, $\alpha_i$ and $\alpha_t$ are county and month fixed effects, respectively, and $X_{it}$ are other control variables. The coefficient of interest is $\gamma_0$, the coefficient on the indicator variable Treated$_{it}$, which takes on a value of 1 if area $i$ had a WARN Notice in month $t$, and 0 otherwise. The coefficients $\gamma_1$, and $\gamma_2$, on the indicator variables Treated$_{i,t-1}$ and Treated$_{i,t-2}$, are also of interest. The indicator variable Treated$_{i,t-1}$ takes on a value of 1 if area $i$ had a WARN Notice in month $t-1$, while Treated$_{i,t-2}$, takes on a value of 1 if area $i$ had a WARN Notice in month $t-2$. Including these indicator variables allows us to determine whether a WARN Notice that occurred in period $t$ also had lagged effects over the next two months. Since WARN Notices are given 60 days ahead of a mass layoffs, $\gamma_1$ would capture the effect during the layoff month while $\gamma_2$ would capture the effect one month after the layoff.

As discussed in Section 3.2, we measure animosity using both internet searches for racial slurs and hate crimes. For the analysis using internet searches for racial slurs, we define $y_{it}$ to be the log of 1 plus the search index for any racial slur.[8] The search trends data, when queried at the month-level, is given from the first of each month. Therefore our analysis is at the county and month level, but from the first of each month, not from the exact date of a layoff notice. For the analysis using hate crimes, we define $y_{it}$ to be the number of hate crimes in county $i$ in that month.

The county fixed effects $\alpha_i$ account for unobserved, time-invariant, differences across counties that may drive differences in hate crimes. The month fixed effects $\alpha_t$ account for unobserved, time-varying shocks that affect all counties similarly. In most specifications, we also account for state-month shocks using a set of state by month fixed effects, which subsume the month fixed

[8]The set of racial slurs we consider are a common slur for Blacks, a slur for Asians, and two slurs for Hispanics or Latinos.

effects. The addition of state by month fixed effects account for time-varying shocks that may be affecting counties within each state similarly. When all these fixed effects are included in the estimation, our identifying assumption is that the change in animosity measures observed in the unaffected counties in the same state provide a valid counterfactual for the change that would have been experienced in the counties affected by the mass layoff, had the mass layoff not occurred. All standard errors are clustered at the county level. For analyses using internet searches from Google Trends, where we spread DMA-level outcome values to the component counties, we also weight observations by county population.

We also conduct a falsification exercise to demonstrate that future layoffs do not affect past animosity measures. The equation we estimate is:

$$y_{it} = \alpha_i + \alpha_t + \beta \cdot X_{it} + \gamma_0 \cdot \text{Treated}_{it} + \gamma_1 \cdot \text{Treated}_{i,t-1} + \gamma_2 \cdot \text{Treated}_{i,t-2} + \sum_{n=1}^{6} \zeta_n \cdot \text{Treated}_{i,t+n} + \epsilon_{it} \quad (3.2)$$

where the coefficients $\zeta_n$ measure the effect of future layoffs, in periods $t + n$, on animosity measures in period $t$. Our empirical strategy requires that our estimated $\zeta_n s$ be zero.

Finally, we estimate Equation 3.1 on placebo search terms and hate crimes not motivated by race. If racial animus against minorities drives our results, we should not estimate significant effects on Whites, LGBT, or people with disabilities. Estimating positive effects for these subgroups would imply that WARN Notices result in a general increase in animosity and propensity to commit hate crimes.

## 3.4 Results

### 3.4.1 Falsification Check

We first estimate a falsification-type exercise. Evidence against the validity of our approach would be if we see that future layoffs are predictive of animosity "today." In Figures 3.1 and 3.2, we report the estimated coefficients from this falsification exercise, on internet searches and hate crimes, respectively. Figure 3.1 shows, in particular, layoffs that will occur 6 months, or even as soon as 1 month from "today" have no effect on searches containing racial slurs "today." We show a similar pattern in Figure 3.2, where the animosity measure is hate crimes.

### 3.4.2 Internet Searches for Racial Slurs

We then move to estimating the effect of mass layoffs on animosity measures. Table 3.1 reports the estimated coefficients $\gamma_0$, and the lagged effects $\gamma_1$ and $\gamma_2$, where the animosity measure used is the log of the search index for any racial slur. Column 1 reports estimates from a base specification that includes only state and month fixed effects. In Column 2, we include state-by-month fixed effects, which subsume the month fixed effects. In Columns 3 and 4, we add county linear time trends and time-varying demographic controls to the specification in Column 2. Our preferred specification is this final specification, which allows for state-month shocks as well as for counties to follow different trends over time. Across all four columns, we consistently estimate approximately a 1.3 percent increase in internet searches for racial slurs in the same month as the mass layoff notice.

We base our definition of treatment timing on the WARN notice date, which is 60 days before the actual mass layoff. Our estimates here do not show the effects persisting beyond the month of the WARN layoff notice, indicating that the increase in animus is driven by the information about and the anticipation of impending mass layoffs. By the time the mass layoff event happens, the racial animus has dissipated.

In Columns 5 and 6 of Table 3.1, we estimate Equation 3.1 again but use placebo search terms as dependent variables. Specifically, we examine whether mass layoffs affect internet searches for the placebo search terms "weather" and the racial slur for Whites. Since the search term "weather" is very unlikely to be affected by news of mass layoffs, we should not expect to find effects on this outcome. Similarly, given that much of the animus in the US is targeted at racial minorities, we should also not expect to find effects on the rate of searches for a slur typically used to refer to Whites. That we report precisely estimated null effects on the searches for both of these words provides further evidence that our main results identify the effects of mass layoffs.

In Table 3.2, we break down the treatment effect by the specific racial slur used. Column 1 reports the effect of mass layoff notices on searches for the common anti-Black slur, Column 2 reports the effect on searches for an anti-Asian slur, and Columns 3 and 4 report the effects on searches for two anti-Hispanic slurs. Table 3.2 tells us that most of the effects we see in Table 3.1 are driven by increases in animus against Black and Hispanic people. We do not find evidence of increased animosity toward Asians, as the main effect is small and precisely estimated.

### 3.4.3 Hate Crimes

To corroborate our findings on internet searches for racial slurs, we also estimate Equation 3.1 using hate crimes as the outcome variable. Table 3.3 reports the estimates of $\gamma_0$, and the lagged effects $\gamma_1$ and $\gamma_2$ from Equation 3.1, where the animosity measure used is the number of hate crimes. In Table 3.3, Columns 1 to 4 use the same specifications as in Table 3.1. Estimates from our preferred specification in Column 4 indicate that the notice of mass layoffs leads to a statistically significant 0.05 increase in the number of hate crimes in that county and month, and that this effect persists for 2 months. With a baseline number of hate crime incidents per county-month of 0.22 (approximately one every five months), this represents a 23 percent increase.

Columns 5 to 8 break down the hate crime estimates by race. Column 5 shows the effect of mass layoff notices on hate crimes against Blacks, Column 6 shows the effect on anti-Asian hate crimes and Column 7 shows the effect on anti-Hispanic hate crimes. As with Table 3.2, we estimate increases in hate crimes against Blacks and Hispanics but not against Asians.

Unlike in Table 3.1, where we do not see effects past the month of the WARN Notice, in some specifications here we still estimate an increase in hate crimes one month later. However, the effect is smaller, and dissipates by the next month. Moreover, when we focus on hate crimes against specific racial minorities, we do not estimate any lagged effects. Therefore, evidence still points toward much of the effect being driven by the WARN Notice, rather than the layoff event.

We then turn to estimating effects on hate crimes motivated by other biases. Columns 8 to 10 present estimates from these placebo specifications. Column 8 shows that mass layoff notices do not lead to increases in racially motivated hate crimes against Whites. Column 9 shows that mass layoff notices do not affect hate crimes against LGBT people, while Column 10 shows a similar novel effect on hate crimes against people with disabilities.

Taken together, our estimates show that mass layoff notices lead to immediate increases in racial animus, that then dissipate within two months. Moreover, the increase in animus causes direct harm to racial minorities, as evidenced by the estimated increases in hate crime. We also show that the increase in racial animus is largely directed toward Blacks and Hispanics and not toward Asians and Whites. Because we also show null effects for hate crimes against LGBT and people with disabilities, we take our results to mean that mass layoff notices cause a contemporaneous increase in animus against racial minorities, rather than a general increase in hate.

### 3.5 Discussion and Conclusion

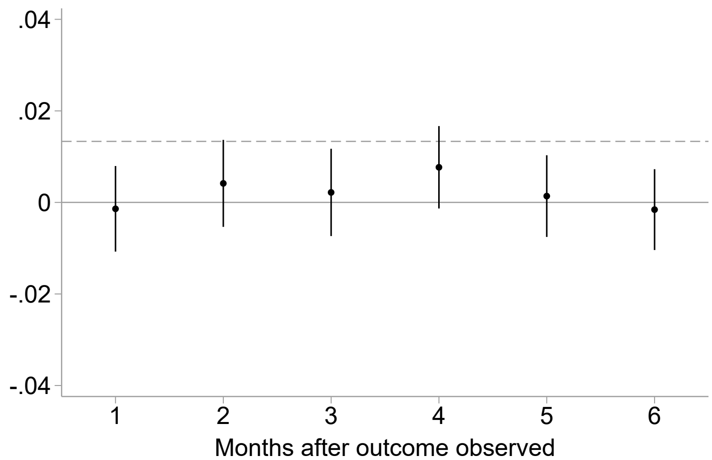
This paper presents evidence of sharp increases in racial animosity following mass layoffs notices. By exploiting the advance notice that the WARN Act requires firms to provide to affected workers, we are able to disentangle the effect of the notice of the mass layoff from the layoff event itself. Our findings show that mass layoff notices lead to an increase in internet searches for racial slurs, indicating heightened racial animus. We put this finding in context by showing a similar effect on hate crime incidents. Thus, while some of the estimated effect on online behavior might just reflect people expressing harmless racist sentiments, or looking up racial slurs, our estimates of the effect on hate crimes indicate that at least part of what we pick up is harmful racial animus directed at minorities. With mass layoff events often a small fraction of a locality's labor force, we interpret our results to mean that the information content in mass layoff notices has its own effect on racial animus, separate from the effect of the layoff event itself.

We explicitly do not attempt to conduct a cost-benefit analysis of providing the type of advance notice that the WARN Act requires, nor do we focus on its unintended consequences. Rather, we only exploit the timing and information content in the WARN Act layoff notices in our empirical strategy. Other studies document effects of WARN Act notices on educational choice [Acton, 2020] and health outcomes Carlson [2015]. We note that although the WARN Notices allow us to separate the information component from the actual layoff event, it is possible that in a counterfactual setting where mass layoffs occur without warning, racial animosity will increase around the time of the layoff event, as would seem to be implied by related studies on economic downturns (Anderson, Crost, and Rees, 2020; Ortega, Di Fruscia, and Louise, 2021; DiLorenzo, 2021).

Our contribution is that we show that information about impending mass layoffs by itself is enough to generate increases in racial animosity. We show that this increase in racial animosity manifests itself both in broad online behavior, and in observed criminal behavior. In an era where information is transmitted much faster than before, and through ever-changing networks and media, it is important to be able to anticipate second and third-order effects of negative economic shocks. This knowledge, combined with the advance notice that WARN Notices provide, can allow local communities to prepare safeguards for their minority populations during times of economic distress.

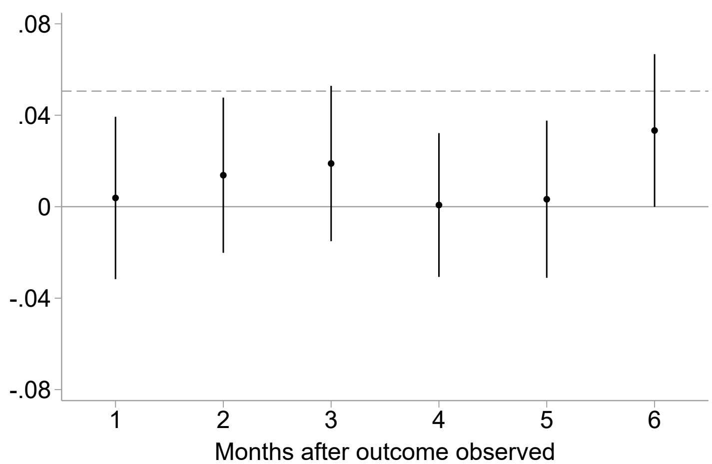
### 3.6 Figures

Figure 3.1: The Effect of Mass Layoff Notices in the “Future” on Internet Searches for Racial Slurs “Today”




This figure shows a falsification exercise estimating the impact of layoffs in periods $t+1,...,t+6$ on outcomes in period $t$. I.e. the effect of layoffs ‘tomorrow’ on hate crimes ’today.’ The dashed line represents the estimated treatment main effect (from layoffs in period $t$ on outcomes in period $t$) from column 4 of the corresponding table, and estimates in each month are shown with 95% confidence intervals.

Figure 3.2: The Effect of Mass Layoff Notices in the "Future" on Hate Crimes "Today"




This figure shows a falsification exercise estimating the impact of layoffs in periods $t + 1, ..., t + 6$ on outcomes in period $t$. I.e. the effect of layoffs 'tomorrow' on hate crimes 'today.' The dashed line represents the estimated treatment main effect (from layoffs in period $t$ on outcomes in period $t$) from column 4 of the corresponding table, and estimates in each month are shown with 95% confidence intervals.

### 3.7 Tables

Table 3.1: The Impact of Mass Layoff Notices on Internet Searches for Racial Slurs and Placebo Words

| | Any racial slur | | | | Placebo searches | |
|---|---|---|---|---|---|---|
| | | | | | 'redneck' | 'weather' |
| | 1 | 2 | 3 | 4 | 5 | 6 |
| Main effect | 0.0162*** | 0.0136*** | 0.0132*** | 0.0133*** | 0.0060 | 0.0033 |
| | (0.0060) | (0.0048) | (0.0043) | (0.0044) | (0.0054) | (0.0020) |
| 1 mo. lagged | 0.0046 | 0.0020 | 0.0017 | 0.0018 | –0.0055 | 0.0015 |
| | (0.0063) | (0.0058) | (0.0061) | (0.0062) | (0.0058) | (0.0021) |
| 2 mo. lagged | –0.0054 | –0.0052 | –0.0053 | –0.0052 | 0.0077 | 0.0006 |
| | (0.0055) | (0.0051) | (0.0052) | (0.0052) | (0.0060) | (0.0022) |
| Observations | 220,158 | 220,158 | 220,158 | 220,158 | 220,158 | 220,158 |
| State-by-month FE | | Y | Y | Y | Y | Y |
| County linear time trend | | | Y | Y | Y | Y |
| Demographic controls | | | | Y | Y | Y |

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
This table shows estimates for the effect of a WARN notice on searches for any racist term, including a slur for blacks known colloquially as the n word, a term for Asians, and two terms for Hispanics. Standard errors clustered at the county level where treatment is assigned.

Table 3.2: The Impact of Mass Layoff Notices on Internet Searches for Specific Racial Slurs

| | Slur for Blacks | Slur for Asians | Slur for Hispanics | Alternative Slur for Hispanics |
|---|---|---|---|---|
| | 1 | 2 | 3 | 4 |
| Main effect | 0.0197*** | 0.0058 | 0.0257*** | 0.0021 |
| | (0.0064) | (0.0100) | (0.0095) | (0.0104) |
| 1 mo. lagged | –0.0012 | –0.0149 | 0.0030 | 0.0204* |
| | (0.0066) | (0.0094) | (0.0108) | (0.0121) |
| 2 mo. lagged | 0.0058 | –0.0206** | –0.0017 | –0.0043 |
| | (0.0074) | (0.0094) | (0.0069) | (0.0111) |
| Observations | 220,158 | 220,158 | 220,158 | 220,158 |
| State-by-month FE | Y | Y | Y | Y |
| County linear time trend | Y | Y | Y | Y |
| Demographic controls | Y | Y | Y | Y |

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
This table shows estimates for the effect of a WARN notice on searches for any racist term, including a slur for blacks known colloquially as the n word, a term for Asians, and two terms for Hispanics. Standard errors clustered at the county level where treatment is assigned.

Table 3.3: The Impact of Mass Layoff Notices on Hate Crime

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| | | all hate crime | | | anti-Black | anti-Asian | anti-Hispanic | anti-White | anti-LGBT | anti-disability |
| Main effect | 0.0478 | 0.0425 | 0.0505** | 0.0505** | 0.0169** | 0.0010 | 0.0046* | 0.0045 | 0.0027 | −0.0005 |
| | (0.0403) | (0.0349) | (0.0229) | (0.0230) | (0.0077) | (0.0017) | (0.0027) | (0.0046) | (0.0055) | (0.0014) |
| 1 mo. lagged | 0.0346 | 0.0308 | 0.0392** | 0.0393** | 0.0071 | 0.0008 | 0.0045 | 0.0050 | 0.0049 | −0.0012 |
| | (0.0397) | (0.0332) | (0.0189) | (0.0190) | (0.0093) | (0.0014) | (0.0029) | (0.0043) | (0.0042) | (0.0014) |
| 2 mo. lagged | 0.0205 | 0.0142 | 0.0224 | 0.0224 | 0.0098 | −0.0006 | 0.0022 | 0.0012 | 0.0010 | 0.0010 |
| | (0.0430) | (0.0373) | (0.0195) | (0.0196) | (0.0080) | (0.0013) | (0.0029) | (0.0044) | (0.0048) | (0.0015) |
| Observations | 220,158 | 220,158 | 220,158 | 220,158 | 220,158 | 220,158 | 220,158 | 220,158 | 220,158 | 220,158 |
| Outcome mean | .221 | .221 | .221 | .221 | .0673 | .00441 | .0139 | .022 | .0378 | .00269 |
| State-by-month FE | | Y | Y | Y | Y | Y | Y | Y | Y | Y |
| County linear time trend | | | Y | Y | Y | Y | Y | Y | Y | Y |
| Demographic controls | | | | Y | Y | Y | Y | Y | Y | Y |

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Standard errors clustered at the county level where treatment is assigned.

# 4. DOES IT MATTER WHO OWNS THE MEDIA? EVIDENCE FROM WITHIN-MARKET MEDIA OWNERSHIP CONSOLIDATION

## 4.1 Introduction

Since the 1990s, the television broadcasting market has become increasingly concentrated as a result of the deregulation of media ownership rules by the Federal Communications Committee (FCC). While supporters of the deregulation argue that allowing consolidation would increase cost efficiency and the quality of the broadcast, critics worry that it would actually result in the opposite and detrimentally impact viewers in other ways. Of particular concern, in addition to broadcast quality, are media integrity, the diversity of viewpoints, and the nature of the information being passed on to the public. For example, by owning multiple TV stations, the owners can more easily impose their views on a larger portion of the population if they so choose.

Previous research has found that media ownership consolidation can impact coverage content as well as the political preferences of viewers [Martin and McCrain, 2019, Miho, 2020]. But they have only focused on consolidation under Sinclair Broadcast Group, a large right-wing conglomerate. However, little is known about the effects of other kinds of ownership consolidation. One common type is within-market consolidation, where two or more stations in an area have the same owner. This could be due to both being purchased by a large conglomerate, or simply when two local stations merge. This paper will bridge this gap by looking at the impact of a very common type of consolidation within a media market where an entity becomes an owner of two TV stations.[1]

Specifically, this paper asks whether within-market consolidation affects viewers and changes their preferences in a meaningful way. To answer this question, we estimate the effects of within-market consolidation on ideological preferences, which we assess by looking at election outcomes. We do so by linking broadcast station transaction records to election outcomes. We then identify effects using a difference-in-differences approach comparing viewers' political preferences in

[1]This type of consolidation is commonly referred to as a 'duopoly' in the broadcasting industry and was forbidden by the FCC until 1999. In this paper, we will refer to this type of consolidation as 'within-market' consolidation in order to separate it from the traditional definition of a *duopoly* in economics.

counties that did or did not experience within-market consolidation over time. The identifying assumption here is that the election outcomes of the treated and control counties would have changed in the same way if there had been no consolidation.

Our results indicate that within-market consolidation did shift viewers' ideological preferences. In particular, we find that within-market consolidation led to an increase in vote share of 3.3 percentage points for Democratic candidates along with a corresponding drop for Republicans. Dynamically, we find this effect persists for at least twelve years. For senate elections, which are more localized, results are larger in magnitude especially in the first four years. This suggests that (1) our estimates are consistent across multiple levels of government and (2) that the impacts of media consolidation are potentially stronger for local versus national outcomes.

In providing evidence on the effects of within-market consolidation of local TV stations, this paper contributes to the literature in several ways. First, to our knowledge, we are the first to look at the impact of within-market consolidation of TV stations on the ideological preferences of viewers. One other paper has looked at within-market consolidation using a structural model to focus on the supply side effects and profitability [Stahl, 2016]. Second, this paper complements studies that look at consolidation under Sinclair [Martin and McCrain, 2019, Miho, 2020]. Relative to these, we focus on an entirely different kind of ownership consolidation. We also demonstrate that ideologies change in a broader context where it is less obvious that media slant would shift – unlike under a single, right-wing conglomerate.

Finally, this paper contributes to the large literature on the impact of media on viewers' political preferences. Based on this literature, we have evidence that exposure to slanted media can shift political preferences [DellaVigna and Kaplan, 2007, Martin and Yurukoglu, 2017, Gerber, Karlan, and Bergan, 2009]. Particularly relevant to us is Snyder Jr and Strömberg [2010] which showed that local newspaper coverage increased constituents' recognition of their district's representative. If within-market consolidation changes local coverage, this could be one channel that explains our result. This story is also consistent with the larger and more immediate effects we see in senate elections, which are likely more reliant on local coverage relative to presidential outcomes.

## 4.2 Background

In the US, the television industry is regulated by the Federal Communications Commission (FCC). To ensure diversity of viewpoints and serve the needs of each local market, the FCC has established rules that limit the national share of media ownership, joint ownership of multiple stations, and cross-ownership across media types. In this paper, we will focus on the rules regarding joint ownership of TV stations. The FCC initially adopted a 'TV duopoly rule' in 1941. This rule prohibited an entity from owning more than one TV station in the same local market. In 1964, they amended the rule to also bar an entity from owning more than one TV station in an overlapped area. With this 'TV duopoly rule', owning multiple TV stations in the same local area was effectively banned in the US until 1999, when the FCC loosened the rule.[2]

In 1999, following the Telecommunications Act of 1996, the FCC relaxed the 'TV duopoly rule' to allow 1. joint ownership of two stations with overlapping coverage as long as they are not in the same market, and 2. joint ownership of two stations in the same market (colloquially called a 'duopoly') so long as eight unique station owners remain in the market and the four highest-rated stations remain under separate ownership. In 2003, the FCC loosened the rule even further and allowed joint ownership of up to 3 stations in large broadcast markets. Because of this, since the late 1990s, we have observed a steady increase in within-market consolidation that created a 'duopoly.' Figure 4.1 plots the number of proposed within-market consolidations over the years, and we can see that the number stayed relatively flat until at least 1996 when it started rising rapidly. In theory, we should not observe any joint ownership before 1999, but the FCC did grant exemptions to a handful of TV stations that were failing even before 1996. Additionally, since the FCC relaxed many other rules regarding TV ownership in 1996 through the Telecommunications Act of 1996, it is possible that we observe the number of joint ownerships increased starting in 1996 because the FCC was more lenient in granting exemptions.

[2] People in the broadcasting industry regularly use 'duopoly' to refer to a case where an entity owns two TV stations in the same market. However, we note that what is referred to as a 'duopoly' here is very different from the traditional definition of *duopoly* in economics.

## 4.3 Data

### 4.3.1 TV ownership and transaction data

To identify within-market consolidation, we need information on each TV station's ownership. We, therefore, obtained the data on the transactions of broadcast TV stations from BIA Advisory Services. These data contain the information on proposed buy and sell transactions of each local TV station in the US from 1950 to 2020. Included in the data are the TV station's call sign, affiliate, location, the type of broadcasting license, the date of proposal, the date of transaction, the proposed buyers, and the proposed sellers. We use the information in this data to identify the owner of each TV station and when the ownership was changing hands. Then with this ownership information, we are able to identify within-market consolidation in each period. In our analysis, for simplicity, we restrict our scope to only the stations that are affiliates of ABC, NBC, CBS, and FOX and are the main signal providers, as these stations makeup most of the market.

### 4.3.2 Election Data

For the analysis, we use election outcomes to measure political preferences. We obtain election data from CQ Voting and Elections Collection. The data include county-level presidential election outcomes and county-level senate election outcomes. The presidential election outcomes span 1948–2016 while senate elections data is from 1968–2018. In particular, we are looking at three outcomes to identify the impact of within-market consolidation of TV stations. First, we look at Democratic vote share to gauge whether public opinion has shifted to the left. Then we look at Republican vote share to see whether there is any change in the support of right-wing politicians. Finally, we look at the margin of victory, which would inform whether within-market consolidation has caused the race to become more (or less) competitive.

## 4.4 Empirical Approach

As TV station consolidation was likely driven by profit maximization, we exploit the exogenous timing of ownership changes with a difference-in-differences approach to estimate the impact of media consolidation. Specifically, we compare the outcomes of counties that saw a consolidation(s) of their local TV stations to counties that did not, over time, using the following model:

$$Y_{it} = \alpha_i + \gamma_t + \theta_x X_{it} + \beta\, I[\text{Consolidated county} \times \text{Post Consolidation}]_{it} + u_{it} \tag{4.1}$$

where $Y_{it}$ is the outcome of county $i$ in year $t$. The outcomes that we look at are presidential and senate election results, which we use as a measure of ideological preferences. $\alpha_i$ is county fixed effects, which capture time-invariant county-specific characteristics that contribute to the voting outcome. Year fixed effects, $\gamma_t$, account for common shocks that affect all counties in year $t$. $X_{it}$ is a matrix containing time-varying characteristics of county $i$ in year $t$, such as population and unemployment rate. $I[\text{Consolidated county} \times \text{Post Consolidation}]_{it}$ is a binary variable equal to one if county $i$'s local TV station(s) was consolidated and year $t$ is post-consolidation. The *coefficient of interest* is $\beta$, which measures the effect of consolidation on outcomes. All standard errors are clustered at the media market level where treatment is assigned, allowing for correlation within a market over time. As with any difference-in-differences approach, the underlying assumption here is that, in the absence of TV station consolidation, the outcomes of counties that saw their local stations consolidated (treated counties) and counties that did not (control counties) would have changed similarly over time. We provide support for this assumption by conducting an event study that shows a common trend between the treated and control counties.

We also estimate the dynamic treatment effects of consolidation by splitting up the treatment effects into multiple post-periods. Specifically, we use the model shown in Equation 4.2.

$$Y_{it} = \alpha_i + \gamma_t + \theta_x X_{it} + \sum_{k \geq 1} \beta_k\, I[\text{Consolidated county}_i \times \text{Election } \#k]_{it} + u_{it} \tag{4.2}$$

where $I[\text{Consolidated county}_i \times \text{Election } \#k]_{it}$ is a binary variable equal to one if the election in year $t$ is the $k$-th election post consolidation for county $i$.[3] And $\beta_k$ identifies the effect of consolidation on the $k$-th election post treatment. There are two major reasons for estimating

[3] Since the presidential election happens every four years, the effects of consolidation in the first four years are captured in the first election post-treatment and are represented by $\beta_1$. The effects of consolidation in years 5–8 are captured in the second election and are represented by $\beta_2$ and so on.

the dynamic treatment effects. First, we anticipate that the effect could be dynamic. Political preferences tend to change gradually rather than abruptly. For example, the Pew Research Center [2018] reports that there has been no substantial shift in partisan affiliation in the US in more than two decades. Therefore, it is likely that the effects would be different in the short run and long run. Second, estimating dynamic effects with multiple post-periods allows us to include in the model county-specific linear time trends that are only estimated based on the preexisting trends and not the dynamic response to treatment [Wolfers, 2006]. By including county-specific time trends, we can also verify that our estimates are not driven by pre-existing trends.

## 4.5 Results

### 4.5.1 Effects on Presidential Election Outcomes

We begin by examining the effects of within-market consolidation on ideological preferences. Using county-level election results to measure political preferences, we specifically look at the following three outcomes: Republican vote share, Democratic vote share, and margin of victory.

For each of the outcomes, we first conduct an event study to provide support for the parallel trend assumption required for the difference-in-differences method. We control for county fixed effects and year fixed effects in all the estimations. The estimates obtained here let us know whether the counties that have experienced within-market consolidation and those that have not ever diverged in any period prior to the consolidation. Figure 4.2 plots the dynamic difference-in-differences estimates for Democratic vote share, Republican vote share, and margin of victory in presidential elections. Notably, since our election data begin in 1948 and within-market consolidation only started in the late 1990s, we are able to observe a very long pre-period. Lending support to our identifying assumption, all three panels in Figure 4.2 show strong evidence of common trends between counties that have experienced within-market consolidation and those that have not in the years preceding the consolidation.

Additionally, looking at the post-period estimates shown in Figure 4.2, it appears that within-market consolidation positively impacted Democratic vote share, and conversely, reduced Republican vote share in the presidential election. The figure also suggests that within-market consolidation reduces the margin of victory in presidential elections.

Next, we formally estimate the effects of within-market consolidation on presidential election outcomes and report the results in Table 4.1. We report the estimates from both the simplest difference-in-differences specification with only the basic county and year fixed effects (Columns 1, 3, 5) and the specification that includes demographic controls (Columns 2, 4, 6). The results from both specifications are similar and thus we will focus on the estimates from the specification with characteristic controls. The estimates indicate that within-market consolidation increased Democratic vote share in presidential elections by 3.31 percentage points, and conversely, decreased Republican vote share by 3.33 percentage points. Furthermore, the estimates also indicate an increase in competitiveness in presidential elections as the margin of victory decreased by 3.21 percentage points.

Table 4.2 reports the estimated dynamic treatment effects on presidential election outcomes. The effects on Democratic vote share are reported in Columns 1–3, Republican vote share in Columns 4–6, and margin of victory in Columns 7–9. For each outcome, the first column reports the estimates from a basic difference-in-differences model without controls, the second column includes controls, and the third column adds county-specific time trends.

Starting with short-term effects in the first row, the estimate from the specification with controls in column 2 indicates significant increases of 1.73 percentage points in Democratic vote share in the first four years. The estimates from the specification that allows county-specific time trends are shown in column 3. With this specification, the short-term estimate is still positive but becomes smaller and no longer significant. Looking at medium-term and long-term effects, they are larger and robust across all specifications. Specifically, the estimate from the specification with controls in row 2 indicates that consolidation increased Democratic vote share by 3.68 percentage points in years 5–8. And although the effect size decreases when we include county-specific linear time trends in the model, the estimate remains positive and significant. For the longer-term effects, rows 3 and 4 show that they are similar to medium-term effects reported in row 2 across all specifications. While some estimates become marginally significant, the effect size and direction remain relatively the same as medium-term effects.

Overall, our estimates here suggest that the effects were smaller in the first four years, but then stabilized in the range of 3.68–4.76 percentage points from the fifth year onward. Additionally, the effects on Republican vote share mirror the effects on Democratic vote share. As for the margin of victory, although we estimate significant negative effects of 2.5–4 percentage points in all the periods when using the specification with controls, the estimates are all smaller in magnitude and become statistically insignificant when we allow for county-specific time trends. We, therefore, do not assert that within-market consolidation reduced the margin of victory or made the race more competitive.

### 4.5.2 Effects on Senate Election Outcomes

Next, we look at the impact of within-market consolidation on senate election outcomes. Doing so allows us to see whether within-market consolidation impacts public opinion on each party in a consistent way and across more localized levels of government. The main difference-in-differences estimates on senate election outcomes are reported in Table 4.3. The results here are similar to the results from presidential elections. Specifically, they indicate that within-market consolidation led to increases of 4.43 percentage points in Democratic vote share and decreases of 4.06 percentage points in Republican vote share in senate elections.

We report the dynamic treatment effects in Table 4.4. For consistency, we use four-year blocks which correspond to one presidential election in the analysis above, but here cover two senate elections in the same period. Starting with Democratic vote share, the estimates are robust across all three specifications and indicate significant positive effects both in the short- and long-runs. The effects on the first two senate elections post-consolidation are reported in Row 1. In particular, the coefficient from our preferred specification in column 2 indicates a significant positive effect of 3.81 percentage points. Looking at the mid-term and long-term effects reported in Rows 2–4, we find that they are positive and larger in magnitude than the effects in the first four years. Specifically, the effects range between 4.69 - 5.04 percentage points and stay at this level for at least 12 years. Again, the effects on Republican vote share mirror the effects on Democratic vote share.

All in all, our result on senate elections here is consistent with what we observe in presidential elections. It indicates that within-market consolidation increased Democratic vote share and decreased Republican vote share. Furthermore, in the first four years, the effects on senate elections are twice as large as the effects on presidential elections and more robust. Similar to the presidential elections, we do not have strong results to conclude that there are significant impacts on the margin of victory.

### 4.6 Discussion and Conclusion

We study the impact of within-market consolidation of TV stations by exploiting the location and timing variation of mergers and acquisitions. Using a difference-in-differences approach with county-level election outcomes that we linked with consolidation information, we are able to estimate the effects of within-market consolidation on ideological preferences. Our results indicate that consolidation resulted in increases in Democratic vote share and decreases in Republican vote share in both presidential and senate elections. Although, the effects appear to be larger and more immediate on senate elections than presidential elections, the effects on both elections persist for at least twelve years. These results are robust to the inclusion of controls and county-specific trends, suggesting that our estimates are not driven by the change in demographics or a difference in trends between the treated and control counties.

Our result shows that it does matter who owns TV stations. Within-market consolidation might not change the menu of TV stations available to the viewers, but it does affect the viewers, in particular their ideological preferences, through other channels. For example, one such channel could be that consolidation shifts the slant of the coverage leftward. This could be true if the entities that led the consolidation lean left-wing.

Alternately, another channel could be that consolidation resulted in changes in local coverage, leading to changes in consumer preferences. The consolidation of the newsrooms and the decrease in competition in the local market, seeing as now these two stations are of the same owners, could be the driving cause. Additionally, if the consolidation makes TV stations shift away from local coverage, it would explain why we estimate larger and more immediate effects on senate election as senate candidates rely on local coverage more than presidential candidates. This would also be consistent with Snyder Jr and Strömberg [2010] which found that constituents were less likely to recognize and able to judge congress members who were less covered by local newspapers.

Two channels could potentially explain the larger and more immediate effects on senate elections. First, since senate candidates rely more on regional and local coverage than presidential candidates, if the consolidation of local TV stations changes their coverage, senate candidates will always be more affected by the consolidation than presidential candidates. Second, because senate elections are held twice as often as presidential elections, we are more likely to observe senate elections in the third year and fourth year post-consolidation than presidential elections. Therefore, mechanically, the short-term effects (1–4 years) on senate elections likely reflect the effects in the third and fourth years post-consolidation more than the short-term effects (1–4 years) on presidential elections.

Our result contributes to the literature on the impact of media consolidation. Closest to us are Martin and McCrain [2019] and Miho [2020]. They looked at the effects of acquisitions by Sinclair Broadcast Group and reported that the takeovers shift both the coverage and the political preferences of viewers rightward. Our findings apply to a more general and common case of within-market consolidation (‘duopoly’). They show that even the more general consolidation, not focused on Sinclair, can impact viewers’ preferences and change the political landscape in the long run.

This paper also adds to the evidence of the impact of television. Many studies have looked at the effects of exposure to a particular type of media [Chen and Yang, 2019, Trudeau, 2016, Kearney and Levine, 2015, La Ferrara, Chong, and Duryea, 2012, Kearney and Levine, 2019, Lindo, Swensen, and Waddell, 2020, Cornelson, 2018]. In particular, DellaVigna and Kaplan [2007], Martin and Yurukoglu [2017], Gerber, Karlan, and Bergan [2009] looked at the effects of media slant and found significant effects on voting behaviors. Our result complement these papers and suggest that not only does exposure to certain content matter, but exposure to a less noticeable change from ownership consolidation also impacts viewers in a meaningful way.

## 4.7 Figures

Figure 4.1: Number of stations that were owned by an entity that also owned at least one more station in the same market (based on proposed transactions)

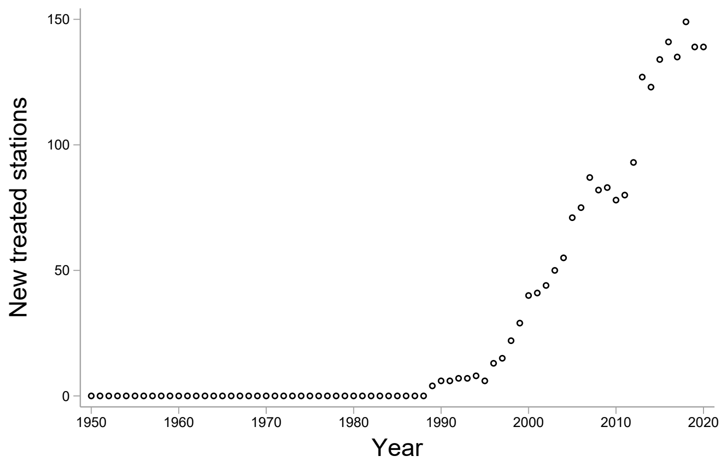
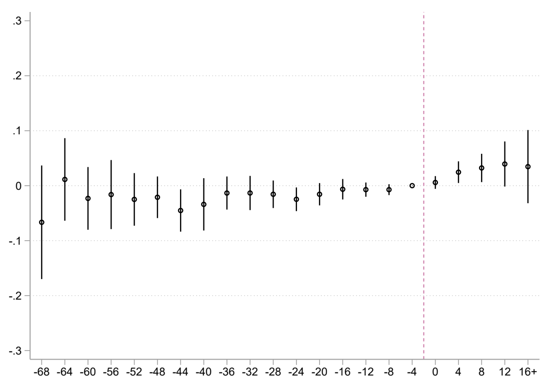
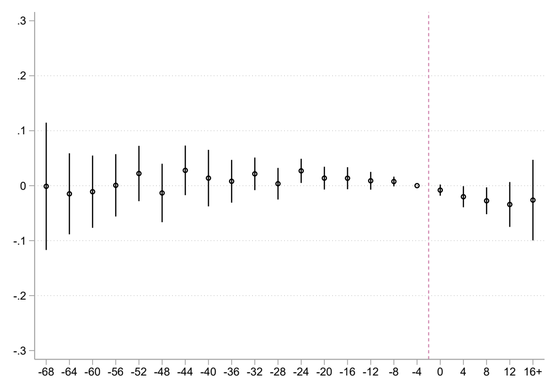


Figure 4.2: Dynamic Difference-in-Differences Estimates on Presidential Election Outcomes

Panel 1: Democratic Vote Share




Panel 2: Republican Vote Share




Panel 3: Margin of Victory

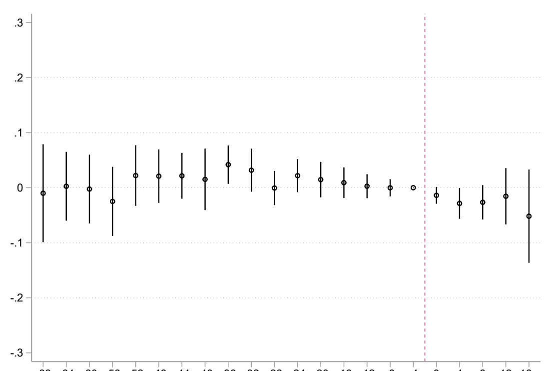
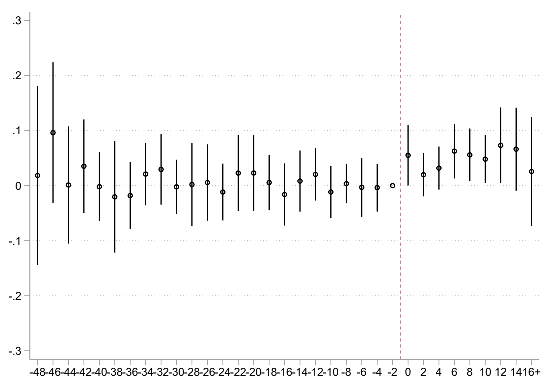
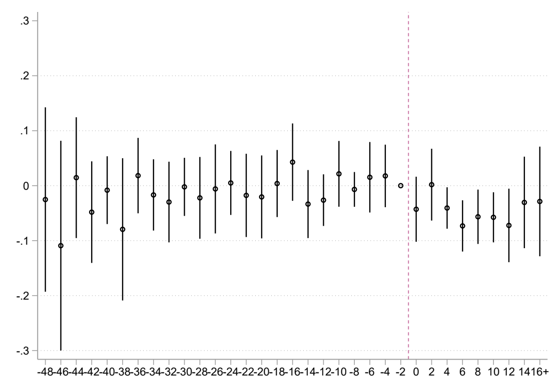


Figure 4.3: Dynamic Difference-in-Differences Estimates on Senate Election Outcomes

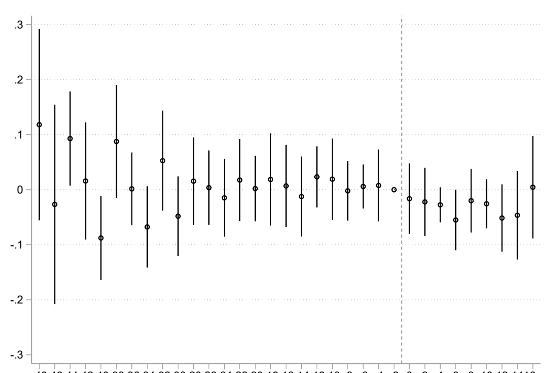
Panel 1: Democratic Vote Share




Panel 2: Republican Vote Share




Panel 3: Margin of Victory



## 4.8 Tables

Table 4.1: Effects of Within-Market Consolidation (Duopoly) on Presidential Election Outcomes

| | Dem. vote share | Dem. vote share | Rep. vote share | Rep. vote share | Margin of victory | Margin of victory |
|---|---|---|---|---|---|---|
| Local Consol. X Post | 0.0385** | 0.0331*** | −0.0301* | −0.0333*** | −0.0356** | −0.0321** |
| | (0.0149) | (0.0113) | (0.0160) | (0.0109) | (0.0162) | (0.0126) |
| Observations | 55,582 | 35,600 | 55,715 | 35,600 | 55,573 | 35,595 |
| Outcome mean | .42 | .401 | .529 | .554 | .247 | .249 |
| Population controls | | Y | | Y | | Y |

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Standard errors in parentheses are clustered at the county level.

Table 4.2: Dynamic Effects of Within-Market Consolidation (Duopoly) On Presidential Election Outcomes

| | Dem. vote share | Dem. vote share | Dem. vote share | Rep. vote share | Rep. vote share | Rep. vote share | Margin of victory | Margin of victory | Margin of victory |
|---|---|---|---|---|---|---|---|---|---|
| Effect 0–3 years | 0.0217** | 0.0173** | 0.0046 | −0.0177 | −0.0212** | −0.0090 | −0.0249* | −0.0235** | −0.0024 |
| | (0.0109) | (0.0085) | (0.0066) | (0.0119) | (0.0085) | (0.0071) | (0.0142) | (0.0117) | (0.0113) |
| Effect 4–7 years | 0.0402*** | 0.0368*** | 0.0225** | −0.0302* | −0.0349*** | −0.0211** | −0.0401** | −0.0386** | −0.0137 |
| | (0.0150) | (0.0126) | (0.0100) | (0.0158) | (0.0120) | (0.0096) | (0.0187) | (0.0159) | (0.0169) |
| Effect 8–11 years | 0.0484*** | 0.0430*** | 0.0225* | −0.0379** | −0.0414*** | −0.0214* | −0.0385** | −0.0333** | −0.0002 |
| | (0.0174) | (0.0143) | (0.0119) | (0.0177) | (0.0137) | (0.0126) | (0.0187) | (0.0164) | (0.0210) |
| Effect 12+ years | 0.0534** | 0.0476** | 0.0294* | −0.0420 | −0.0452** | −0.0266 | −0.0444 | −0.0388* | −0.0145 |
| | (0.0265) | (0.0189) | (0.0156) | (0.0292) | (0.0187) | (0.0184) | (0.0295) | (0.0234) | (0.0350) |
| Observations | 55,582 | 35,600 | 35,600 | 55,715 | 35,600 | 35,600 | 55,573 | 35,595 | 35,595 |
| Outcome mean | .42 | .401 | .401 | .529 | .554 | .554 | .247 | .249 | .249 |
| Population controls | | Y | Y | | Y | Y | | Y | Y |
| Area linear time trend | | | Y | | | Y | | | Y |

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Standard errors in parentheses are clustered at the county level.

Table 4.3: Effects of Within-Market Consolidation (Duopoly) on Senate Election Outcomes

| | Dem. vote share | Dem. vote share | Rep. vote share | Rep. vote share | Margin of victory | Margin of victory |
|---|---|---|---|---|---|---|
| Local Consol. X Post | 0.0434** | 0.0443** | −0.0400** | −0.0406** | −0.0306* | −0.0243 |
| | (0.0193) | (0.0173) | (0.0184) | (0.0164) | (0.0174) | (0.0165) |
| Observations | 48,935 | 44,668 | 50,047 | 45,752 | 49,415 | 45,143 |
| Outcome mean | .465 | .464 | .517 | .518 | .266 | .266 |
| Population controls | | Y | | Y | | Y |

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Standard errors in parentheses are clustered at the county level.

Table 4.4: Dynamic Effects of Within-Market Consolidation (Duopoly) On Senate Election Outcomes

| | Dem. vote share | Dem. vote share | Dem. vote share | Rep. vote share | Rep. vote share | Rep. vote share | Margin of victory | Margin of victory | Margin of victory |
|---|---|---|---|---|---|---|---|---|---|
| Effect 0–3 years | 0.0364** | 0.0381** | 0.0385*** | –0.0207 | –0.0219 | –0.0411** | –0.0212 | –0.0163 | –0.0424* |
| | (0.0180) | (0.0175) | (0.0137) | (0.0182) | (0.0174) | (0.0162) | (0.0228) | (0.0229) | (0.0250) |
| Effect 4–7 years | 0.0468** | 0.0469*** | 0.0338*** | –0.0567*** | –0.0563*** | –0.0713*** | –0.0488** | –0.0441** | –0.0627*** |
| | (0.0187) | (0.0175) | (0.0126) | (0.0175) | (0.0162) | (0.0179) | (0.0199) | (0.0182) | (0.0240) |
| Effect 8–11 years | 0.0489** | 0.0504*** | 0.0509*** | –0.0546** | –0.0555*** | –0.0756*** | –0.0256 | –0.0171 | –0.0390 |
| | (0.0208) | (0.0189) | (0.0139) | (0.0233) | (0.0213) | (0.0210) | (0.0206) | (0.0209) | (0.0350) |
| Effect 12+ years | 0.0468 | 0.0473 | 0.0572** | –0.0416 | –0.0429 | –0.0785*** | –0.0303 | –0.0209 | –0.0516 |
| | (0.0352) | (0.0300) | (0.0226) | (0.0339) | (0.0288) | (0.0288) | (0.0273) | (0.0245) | (0.0450) |
| Observations | 48,935 | 44,668 | 44,668 | 50,047 | 45,752 | 45,752 | 49,415 | 45,143 | 45,143 |
| Outcome mean | .465 | .464 | .464 | .517 | .518 | .518 | .266 | .266 | .266 |
| Population controls | | Y | Y | | Y | Y | | Y | Y |
| Area linear time trend | | | Y | | | Y | | | Y |

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Standard errors in parentheses are clustered at the county level.

REFERENCES


Daron Acemoglu, David Autor, David Dorn, Gordon H Hanson, and Brendan Price. Import competition and the great us employment sag of the 2000s. *Journal of Labor Economics*, 34(S1): S141–S198, 2016.

Riley Acton. Community college program choices in the wake of local job losses. 2020.

D Mark Anderson, Benjamin Crost, and Daniel I Rees. Do economic downturns fuel racial animus? *Journal of Economic Behavior & Organization*, 175:9–18, 2020.

Luvell Anderson and Ernie Lepore. Slurring words. *Noûs*, 47(1):25–48, 2013.

Laura B Bartell. Why warn–the worker adjustment and retraining notification act in bankruptcy. *Bankr. Dev. J.*, 18:243, 2001.

Bureau of Labor Statistics. Women in architecture and engineering occupations in 2016, 2017. URL `https://www.bls.gov/opub/ted/2017/women-in-architecture-and-engineering-occupations-in-2016.htm`.

Bureau of Labor Statistics. Occupational employment statistics, 2019. URL `https://www.bls.gov/oes/current/oes_stru.htm`.

Sebastian Calonico, Matias D Cattaneo, and Rocio Titiunik. Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, 82(6):2295–2326, 2014.

A Colin Cameron, Jonah B Gelbach, and Douglas L Miller. Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics*, 90(3):414–427, 2008.

David Card and A Abigail Payne. High school choices and the gender gap in stem. *Economic Inquiry*, 59(1):9–28, 2021.

Kyle Carlson. Fear itself: The effects of distressing economic news on birth outcomes. *Journal of health economics*, 41:117–132, 2015.

Anthony P Carnevale, Ban Cheah, and Andrew R Hanson. The economic value of college majors. 2015.

Scott E. Carrell, Marianne E. Page, and James E. West. Sex and science: How professor gender perpetuates the gender gap. *Quarterly Journal of Economics*, 125(3):1101–1144, 2010. ISSN 00335533. doi: 10.1162/qjec.2010.125.3.1101.

Pew Research Center. Trends in party affiliation among demographic groups., 2018.

Yuyu Chen and David Y Yang. The impact of media censorship: 1984 or brave new world? *American Economic Review*, 109(6):2294–2332, 2019.

Kirsten Cornelson. Media role models and black educational attainment: Evidence from the cosby show. Technical report, Working paper. University of Notre Dame, 2018.

Stefano DellaVigna and Ethan Kaplan. The fox news effect: Media bias and voting. *The Quarterly Journal of Economics*, 122(3):1187–1234, 2007.

Matthew DiLorenzo. Trade layoffs and hate in the united states. *Social Science Quarterly*, 102(2): 771–785, 2021.

Avraham Ebenstein, Ann Harrison, Margaret McMillan, and Shannon Phillips. Estimating the impact of trade and offshoring on american workers using the current population surveys. *Review of Economics and Statistics*, 96(4):581–595, 2014.

Ozkan Eren and Naci Mocan. Emotional judges and unlucky juveniles. *American Economic Journal: Applied Economics*, 10(3):171–205, 2018.

Armin Falk, Andreas Kuhn, and Josef Zweimüller. Unemployment and right-wing extremist crime. *Scandinavian Journal of Economics*, 113(2):260–285, 2011.

Kayal R Fontenot, Jessica L Semega, and Melissa A Kollar. US Census Bureau, Current Population Reports, P60-259, Income and Poverty in the United States: 2016. *US Government Printing Office, Washington, DC*, (September 2018), 2017.

Thomas Fujiwara, Karsten Müller, and Carlo Schwarz. The effect of social media on elections: Evidence from the united states. Technical report, National Bureau of Economic Research, 2021.

Alan S Gerber, Dean Karlan, and Daniel Bergan. Does the media matter? a field experiment measuring the effect of newspapers on voting behavior and political opinions. *American Economic Journal: Applied Economics*, 1(2):35–52, 2009.

Francesco Giavazzi, Felix Iglhaut, Giacomo Lemoli, and Gaia Rubera. Terrorist attacks, cultural incidents and the vote for radical parties: Analyzing text from twitter. Technical report, National Bureau of Economic Research, 2020.

Claudia Goldin. Gender and the undergraduate economics major: Notes on the undergraduate economics major at a highly selective liberal arts college. *manuscript, April*, 12, 2015.

Claudia Goldin, Sari Pekkala Kerr, Claudia Olivetti, and Erling Barth. The expanding gender earnings gap: Evidence from the LEHD-2000 census. *American Economic Review*, 107(5): 110–114, 2017. ISSN 00028282. doi: 10.1257/aer.p20171065.

Shushanik Hakobyan and John McLaren. Looking for local labor market effects of nafta. *Review of Economics and Statistics*, 98(4):728–741, 2016.

Maria Johnson and Vicki S. Helgeson. Sex differences in response to evaluative feedback: A field study. *Psychology of Women Quarterly*, 26(3):242–251, 2002. ISSN 03616843. doi: 10.1111/1471-6402.00063.

Neeraj Kaushal, Robert Kaestner, and Cordelia Reimers. Labor market effects of september 11th on arab and muslim residents of the united states. *Journal of Human Resources*, 42(2):275–308, 2007.

Melissa S Kearney and Phillip B Levine. Media influences on social outcomes: The impact of mtv's 16 and pregnant on teen childbearing. *American Economic Review*, 105(12):3597–3632, 2015.

Melissa S Kearney and Phillip B Levine. Early childhood education by television: Lessons from sesame street. *American Economic Journal: Applied Economics*, 11(1):318–50, 2019.

Pawel Michal Krolikowski and Kurt G Lunsford. Advance layoff notices and labor market forecasting. 2020.

Adriana D. Kugler, Catherine H. Tinsley, and Olga Ukhaneva. Choice of Majors: Are Women Really Different from Men? *NBER Working Paper No. 23735*, page 38, 2017. URL `http://www.nber.org/papers/w23735`.

Eliana La Ferrara, Alberto Chong, and Suzanne Duryea. Soap operas and fertility: Evidence from brazil. *American Economic Journal: Applied Economics*, 4(4):1–31, 2012.

Jason M Lindo, Isaac D Swensen, and Glen R Waddell. Persistent effects of violent media content. Technical report, National Bureau of Economic Research, 2020.

Joyce B. Main and Ben Ost. The impact of letter grades on student effort, course selection, and major choice: A regression-discontinuity analysis. *Journal of Economic Education*, 45(1):1–10, 2014. ISSN 00220485. doi: 10.1080/00220485.2014.859953.

Gregory J Martin and Joshua McCrain. Local news and national politics. *American Political Science Review*, 113(2):372–384, 2019.

Gregory J Martin and Ali Yurukoglu. Bias in cable news: Persuasion and polarization. *American Economic Review*, 107(9):2565–99, 2017.

Madeline Masucci and Lynn Langton. Hate crime victimization, 2004-2015. *NCJ*, 250653, 2017.

Caroline May. U.s. graduates expect mass layoffs as companies keep hiring h-1b visa workers. *Breitbart*, Mar 2020. URL `https://www.breitbart.com/politics/2020/03/23/u-s-graduates-expect-mass-layoffs-as-companies-hire-more-h-1b-workers/`.

Margarita Mayo, Maria Kakarika, Juan Carlos Pastor, and Stéphane Brutus. Aligning or inflating your leadership self-image? A longitudinal study of responses to peer feedback in MBA teams. *Academy of Management Learning and Education*, 11(4):631–652, 2012. ISSN 1537260X. doi: 10.5465/amle.2010.0069.

Antonela Miho. Small screen, big echo? estimating the political persuasion of local television news bias using sinclair broadcast group as a natural experiment. 2020.

Karsten Müller and Carlo Schwarz. From hashtag to hate crime: Twitter and anti-minority sentiment. *Available at SSRN 3149103*, 2020.

Karsten Müller and Carlo Schwarz. Fanning the flames of hate: Social media and hate crime. *Journal of the European Economic Association*, 2021.

National Center for Education Statistics. Status and trends in the education of racial and ethnic groups, indicator 26: Stem degrees, 2019. URL `https://nces.ed.gov/programs/r` aceindicators/indicator_reg.asp.

Muriel Niederle and Lise Vesterlund. Explaining the gender gap in math test scores: The role of competition. *Journal of Economic Perspectives*, 24(2):129–44, 2010.

Rose E O'Dea, Malgorzata Lagisz, Michael D Jennions, and Shinichi Nakagawa. Gender differences in individual variation in academic grades fail to fit expected patterns for stem. *Nature communications*, 9(1):1–8, 2018.

Alberto Ortega, Ema Di Fruscia, and Bryn Louise. Trade liberalization and racial animus. *Contemporary Economic Policy*, 39(1):194–204, 2021.

Ann L. Owen. Grades, gender, and encouragement: A regression discontinuity analysis. *Journal of Economic Education*, 41(3):217–234, 2010. ISSN 00220485. doi: 10.1080/00220485.2010.486718.

Arnaud Philippe and Aurélie Ouss. "no hatred or malice, fear or affection": Media and sentencing. *Journal of Political Economy*, 126(5):2134–2178, 2018.

Cristian Pop-Eleches and Miguel Urquiola. Going to a better school: Effects and behavioral responses. *American Economic Review*, 103(4):1289–1324, 2013.

Catherine Porter and Danila Serra. Gender differences in the choice of major: The importance of female role models. *American Economic Journal: Applied Economics*, 12(3):226–54, 2020.

Julia Preston. Pink slips at disney. but first, training foreign replacements. *The New York Times*, Jun 2015. URL https://www.nytimes.com/2015/06/04/us/last-task-after-layoff-at-disney-train-foreign-replacements.html?ref=topics&_r=0.

Matt E Ryan and Peter T Leeson. Hate groups and hate crime. *International Review of Law and Economics*, 31(4):256–262, 2011.

James M Snyder Jr and David Strömberg. Press coverage and political accountability. *Journal of political Economy*, 118(2):355–408, 2010.

Jamin Speer. Bye bye ms. american sci: Women and the leaky stem pipeline. 2021.

Jessica Calfee Stahl. Effects of deregulation and consolidation of the broadcast television industry.

*American Economic Review*, 106(8):2185–2218, 2016.

Seth Stephens-Davidowitz. The cost of racial animus on a black candidate: Evidence using google search data. *Journal of Public Economics*, 118:26–40, 2014.

Seth Stephens-Davidowitz and Hal Varian. A hands-on guide to google data. *further details on the construction can be found on the Google Trends page*, 2014.

Jennifer Trudeau. The role of new media on teen sexual behaviors and fertility outcomes—the case of 16 and pregnant. *Southern Economic Journal*, 82(3):975–1003, 2016.

Justin Wolfers. Did unilateral divorce laws raise divorce rates? a reconciliation and new results. *American Economic Review*, 96(5):1802–1820, 2006.

Danny Yagan. Employment hysteresis from the great recession. *Journal of Political Economy*, 127 (5):2505–2558, 2019.

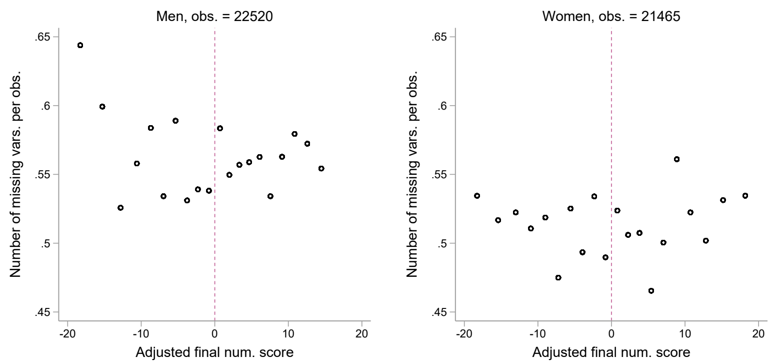
APPENDIX A

# DO GRADE SIGNALS DRIVE THE GENDER GAP IN STEM? EVIDENCE FROM A REGRESSION DISCONTINUITY APPENDIX

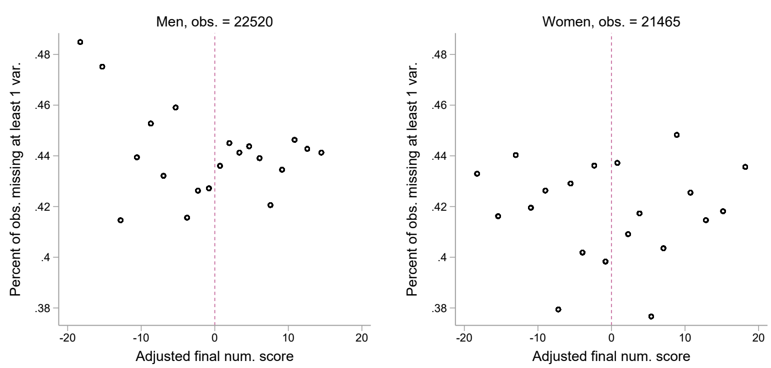
## A.1 Figures

Figure A.1: Missing Variables Analysis

(a) Total number of missing variables for each observation



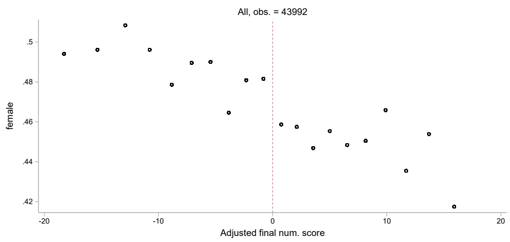

(b) Percent of observations missing at least one variable



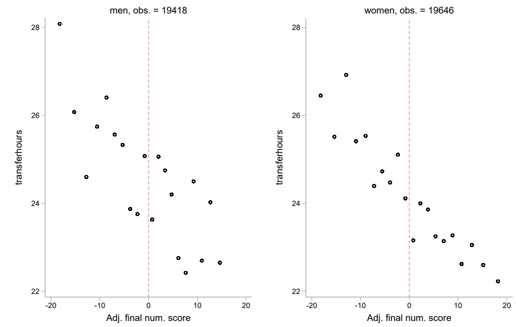

In these data, some observations are missing some variables. To address concerns that these missing characteristics are non-random and could bias those parts of my analysis that include these control variables, I present these two figures. Figure A.1a shows the total number of missing variables per observation and Figure A.1b shows the number of observations missing at least one variable. Both figures show that these measures of missing variables appear smooth through the threshold. Each point represents an equal number of observations.

Figure A.2: Verifying Covariates Are Smooth Through the Letter Grade Threshold

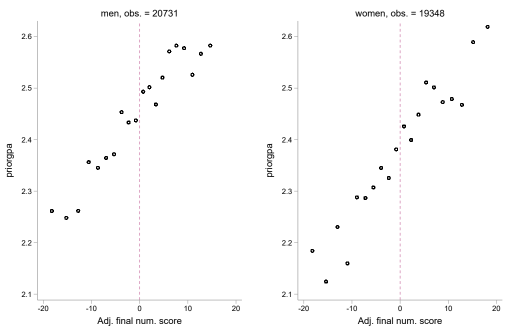
(a) Student Gender



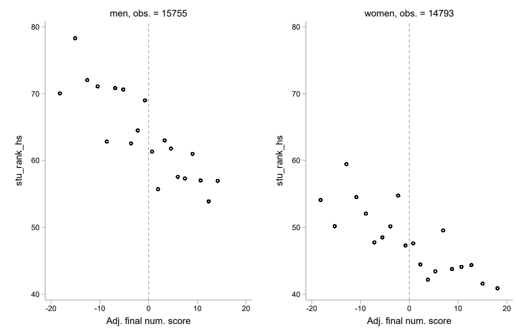

(b) Transfer Hours



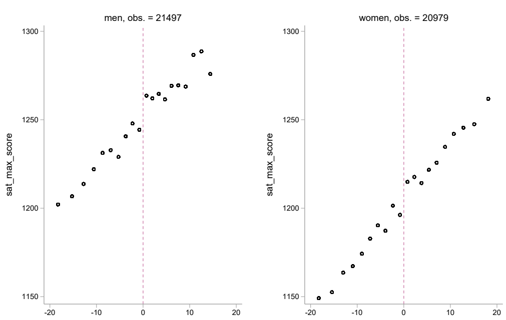

(c) Prior College GPA



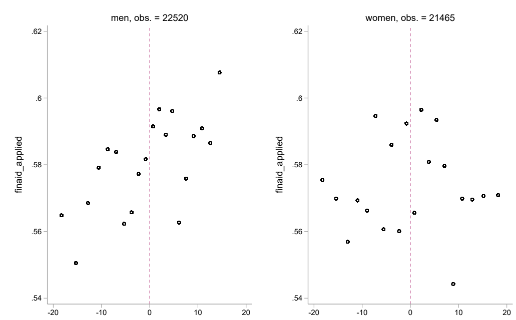

(d) High School Rank



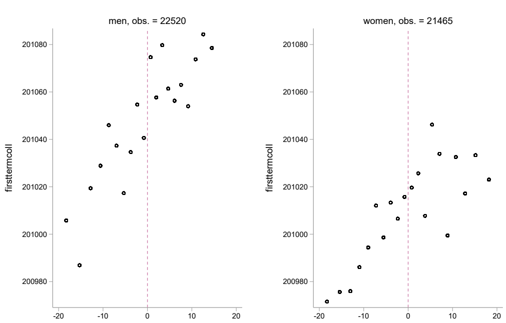

(e) max{SAT, ACT concordance}



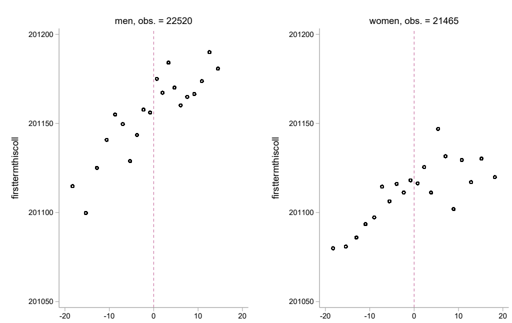

(f) Applied for Financial Aid




(g) First Term Any College

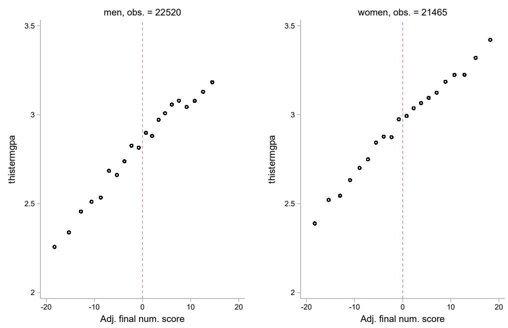



(h) First Term This College

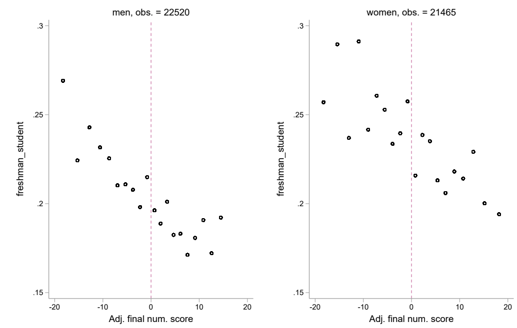



Continued on next page.

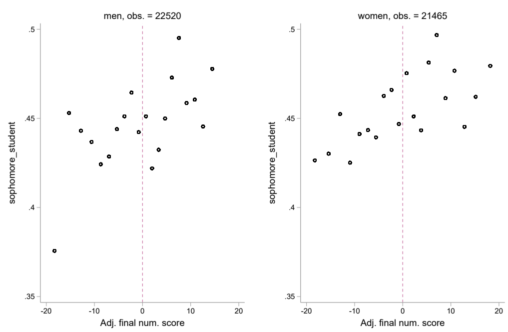
(i) Semester GPA Excl. Sample Courses



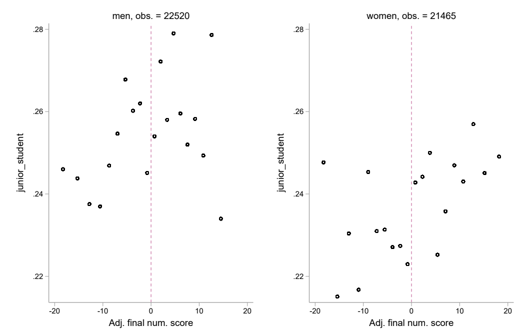

(j) Freshman



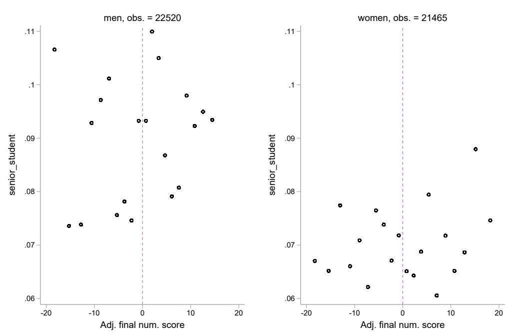

(k) Sophomore



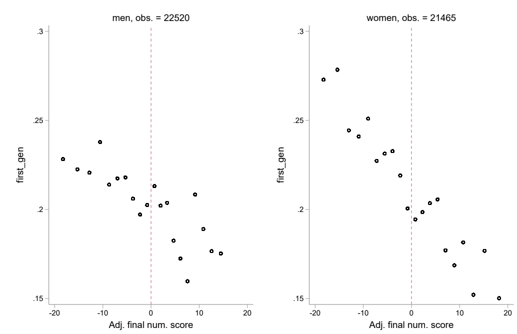

(l) Junior



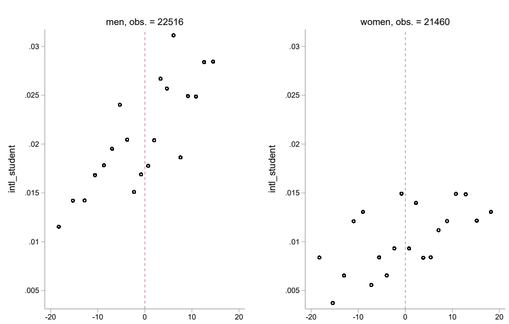

(m) Senior



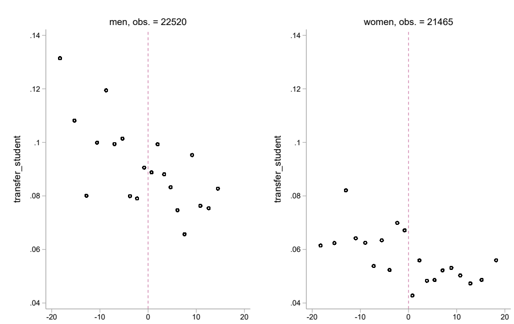

(n) First Generation Student




(o) International Student



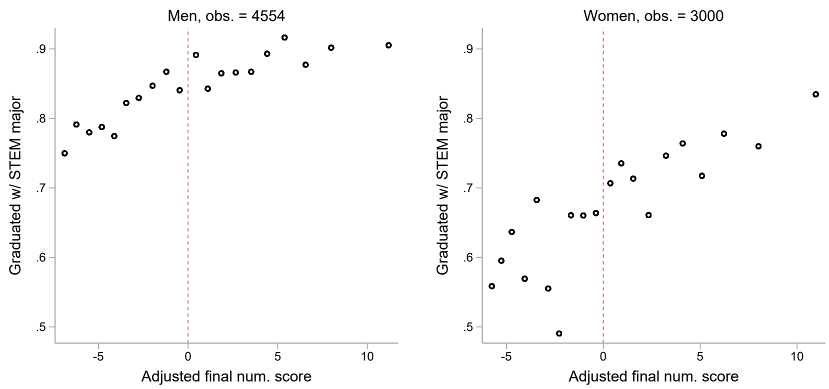

(p) Transfer Student



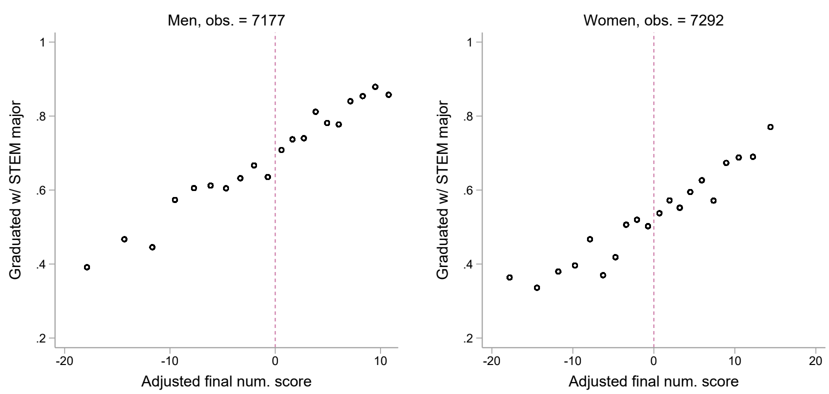

These figures show that all other variables appear smooth through the threshold, which supports the identifying assumption of the regression discontunity. I estimate each variable in Table A.2. Each point represents an equal number of observations.

Figure A.3: Effect at each letter grade cutoff

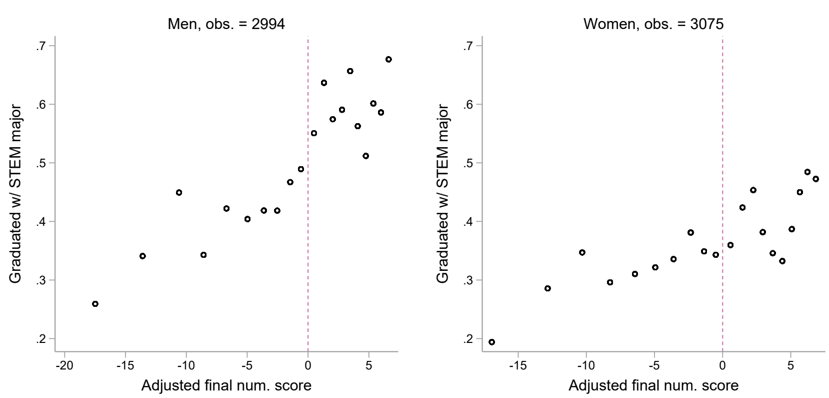
(a) Effect of A/B cutoff on graduation with a STEM major




(b) Effect of B/C cutoff on graduation with a STEM major



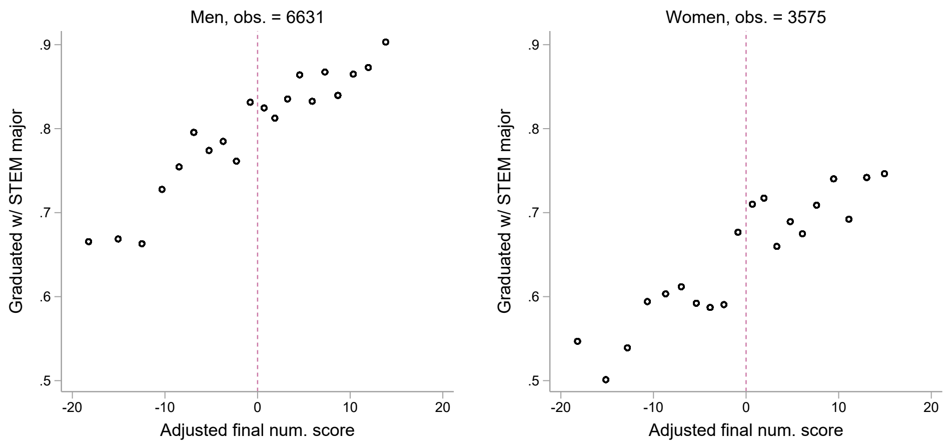

(c) Effect of C/D cutoff on graduation with a STEM major



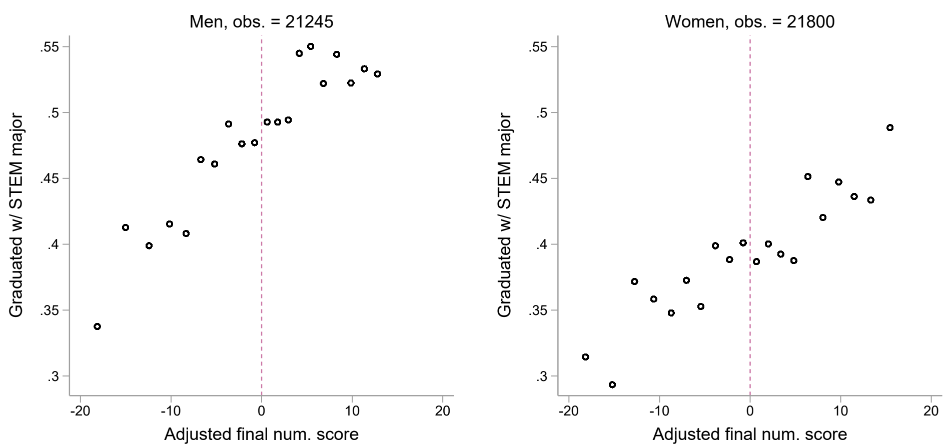

These figures show the main outcome, graduated with a STEM major within six years, for each letter grade cutoff separately. While most figures show no significant effect, there is some suggestive evidence of a discontunity for men at the C/D cutoff. Each point represents an equal number of observations.

Figure A.4: Subgroups Where Effect Seems Most Likely

(a) Only courses taught by female instructor

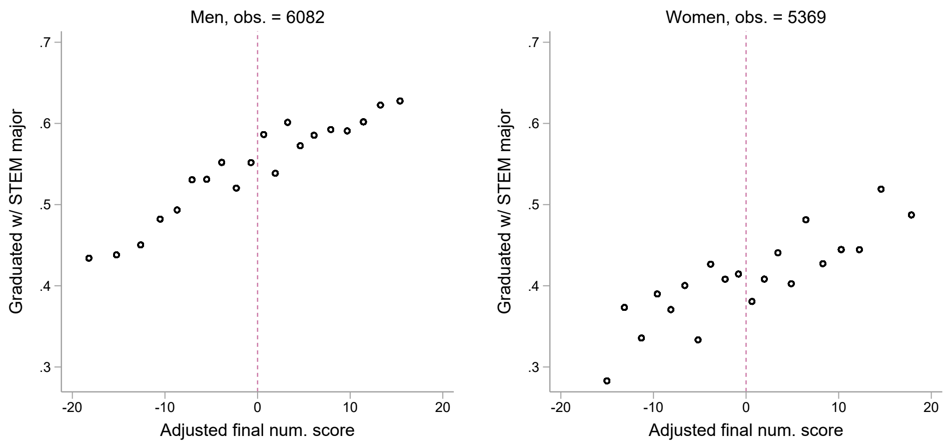



(b) Only courses taught by male instructor

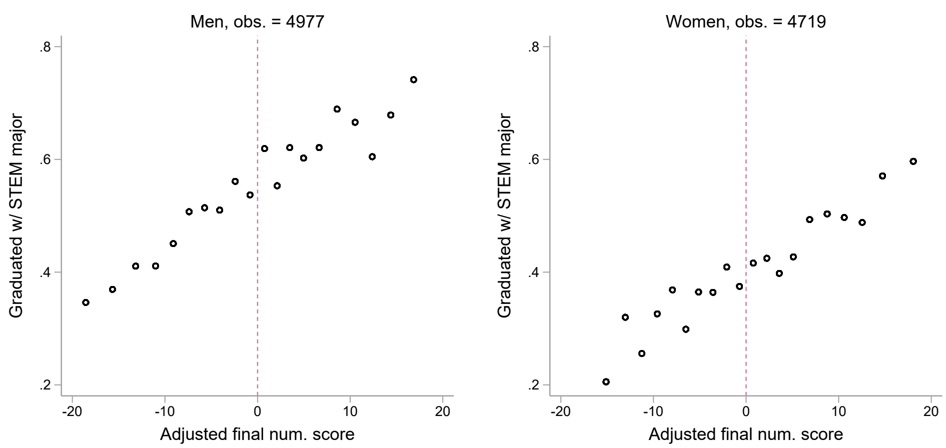



Continued on next page.

(c) Only first generation students



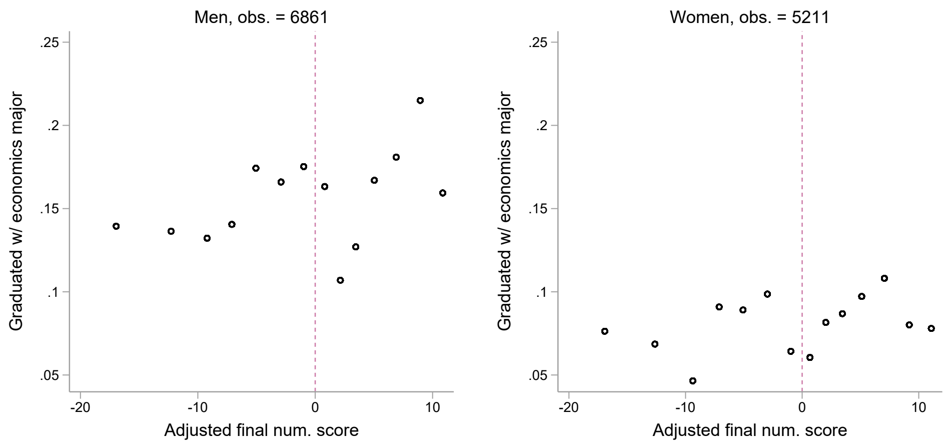

(d) Only freshmen students



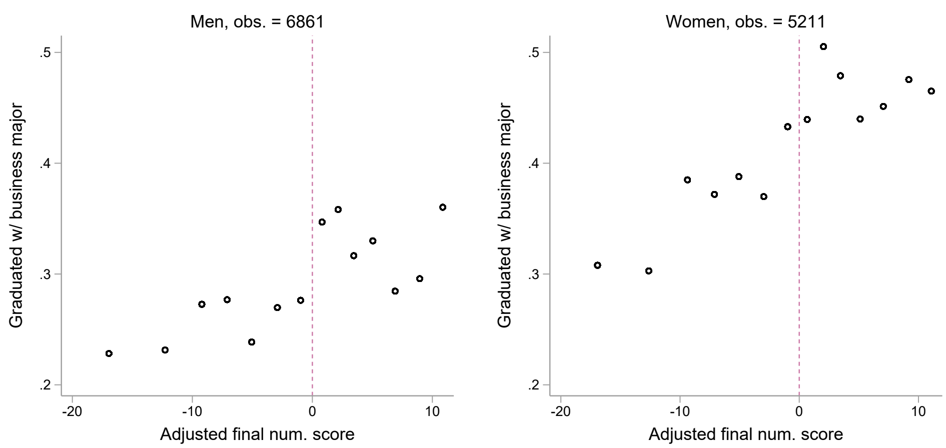

These figures show data around the threshold for subgroups where an effect seems likely. I estimate these regressions in Table 2.4. Each point represents an equal number of observations.

Figure A.5: Effect of letter grades in economics courses

(a) Effect on graduation with an econonmics major



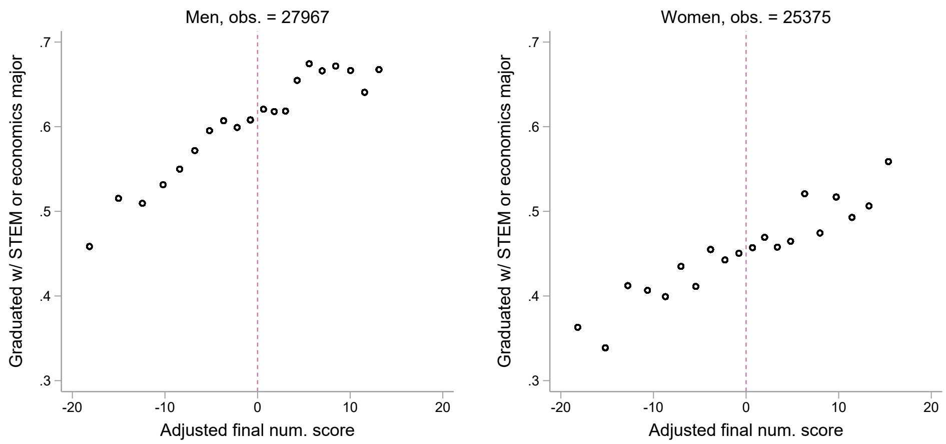

(b) Effect on graduation with a business major



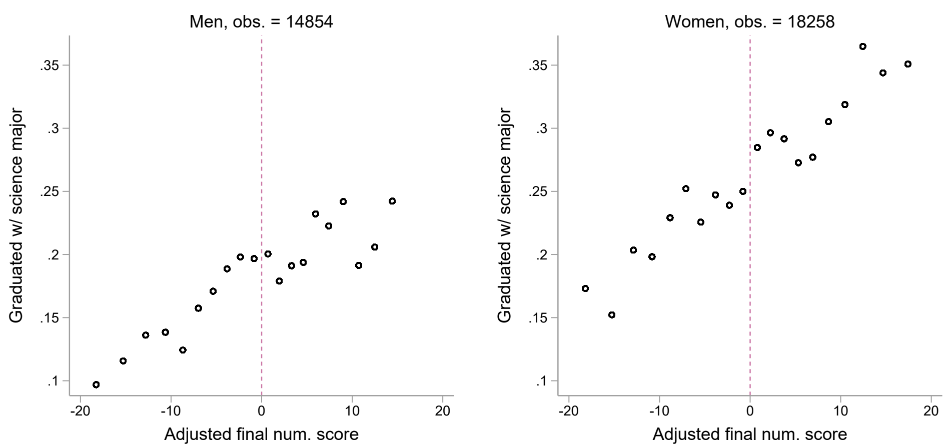

These figures show data using only economics courses, and look at the outcome of majorng in economics and business (subfigures (a) and (b), respectively). I estimate these effects in Table A.7. Each point represents an equal number of observations.

Figure A.6: Alternate subgroups of courses, and alternate outcomes

(a) All courses including economics and effect on STEM or economics major

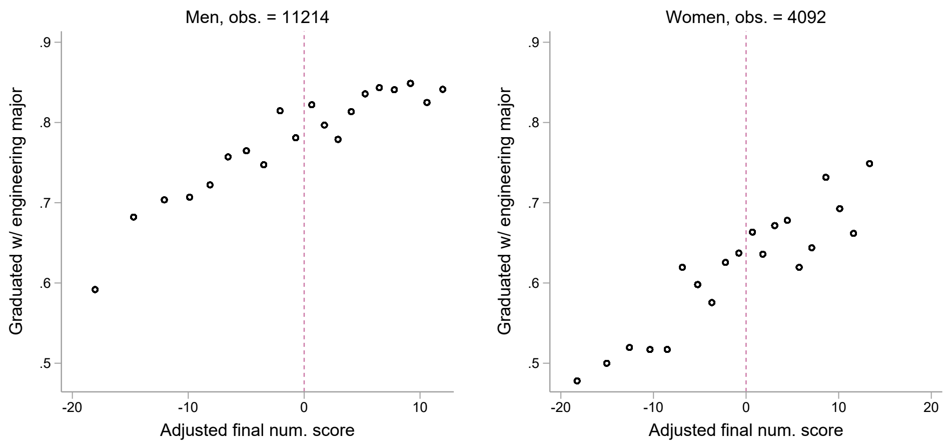



(b) Only science courses (biology, calculus, organic chemistry) and effect on science major

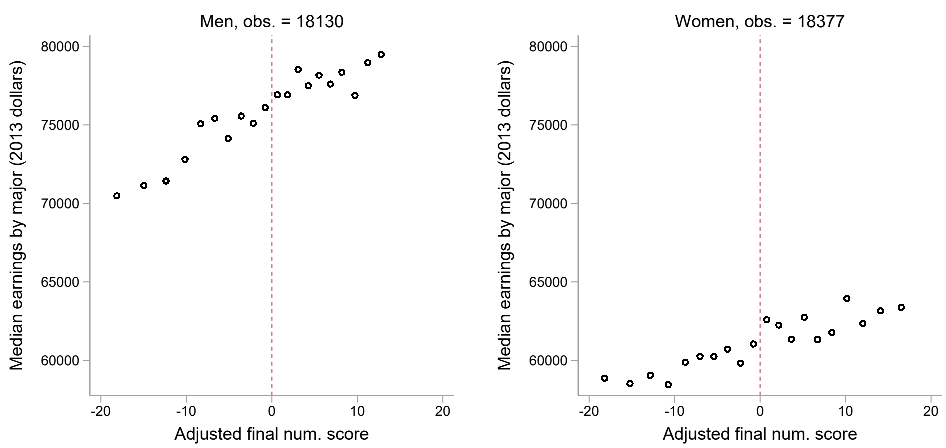



Continued on next page.

(c) Only engineering courses (calculus, computer programming, and statics) and effect on engr. major




(d) Main sample STEM courses and effect on estimated median earnings




These figures show data around the letter grade threshold for alternative subgroups and outcomes. I estimate these regressions in Table A.8. Each point represents an equal number of observations.

## A.2 Tables

Table A.1: Missing variable statistics

| | All students | Men | Women |
|---|---|---|---|
| avg. num. missing vars | 0.63 | 0.58 | 0.54 |
| missing $>=$ 1 var. | 0.49 | 0.45 | 0.44 |
| female | 0.07 | 0.00 | 0.00 |
| HS stu. rank | 0.30 | 0.30 | 0.32 |
| max(SAT score, ACT conc. score) | 0.03 | 0.04 | 0.02 |
| prior gpa | 0.10 | 0.08 | 0.10 |
| transfer hours | 0.12 | 0.15 | 0.09 |
| intl. stu. | 0.00 | 0.00 | 0.00 |
| Observations | 21,533 | 10,790 | 9,188 |

This table shows the average number of missing variables per observation (first row), the percent of observations missing at least one variable (second row), and the percent of observations missing each variable (remaining rows). Because many students do not take both the SAT and ACT, I use official concordance tables to convert ACT scores to comparable SAT scores, then take the max of SAT score and ACT concordance score where both exist. The resulting variable is only missing for 3% of the observations in the sample.

Table A.2: Test of identifying assumption: regression discontinuity estimates for covariates

## Panel A: All students

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) | (13) | (14) | (15) | (16) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | female | transfer hours | prior gpa | HS stu. rank | max(SAT score, ACT conc. score) | appl. fin. aid | 1st term any coll. | 1st term this coll. | GPA excl.sample courses | freshman | sophomore | junior | senior | 1st gen. stu. | intl. stu. | transfer stu. |
| Above letter grade cutoff | −0.0288* | −0.3731 | 0.0184 | −3.4359 | 9.1586** | 0.0005 | 14.9416 | 9.4970 | −0.0356 | −0.0272* | 0.0093* | 0.0140 | 0.0029 | 0.0096 | −0.0007 | −0.0018 |
| | (0.0152) | (0.4318) | (0.0566) | (3.8152) | (3.9876) | (0.0187) | (13.4645) | (14.1628) | (0.0245) | (0.0129) | (0.0045) | (0.0110) | (0.0090) | (0.0125) | (0.0047) | (0.0096) |
| Observations | 14,997 | 14,444 | 14,739 | 11,575 | 15,757 | 16,273 | 16,273 | 16,273 | 16,273 | 16,273 | 16,273 | 16,273 | 16,273 | 16,273 | 16,270 | 16,273 |
| Variable mean | .464 | 24.1 | 2.42 | 55 | 1,228 | .587 | 201,046 | 201,150 | 2.91 | .222 | .457 | .241 | .0779 | .207 | .0151 | .0691 |
| Bandwidth = 5.08 | 1x | 1x | 1x | 1x | 1x | 1x | 1x | 1x | 1x | 1x | 1x | 1x | 1x | 1x | 1x | 1x |

## Panel B: Men

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) | (13) | (14) | (15) | (16) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | transfer hours | prior gpa | HS stu. rank | max(SAT score, ACT conc. score) | appl. fin. aid | 1st term any coll. | 1st term this coll. | GPA excl.sample courses | freshman | sophomore | junior | senior | 1st gen. stu. | intl. stu. | transfer stu. |
| Above letter grade cutoff | | 0.0433 | 0.0032 | −6.8439 | 9.1764 | 0.0194 | 16.3395 | 7.8851 | −0.0331 | −0.0084 | −0.0276 | 0.0166 | 0.0174 | 0.0227 | 0.0028 | 0.0190 |
| | | (0.6826) | (0.0294) | (4.9951) | (8.0086) | (0.0210) | (12.6329) | (14.7383) | (0.0261) | (0.0178) | (0.0178) | (0.0103) | (0.0124) | (0.0149) | (0.0066) | (0.0138) |
| Observations | | 7,716 | 8,223 | 6,378 | 8,577 | 8,970 | 8,970 | 8,970 | 8,970 | 8,970 | 8,970 | 8,970 | 8,970 | 8,970 | 8,967 | 8,970 |
| Variable mean | | 24.3 | 2.46 | 63.5 | 1,253 | .583 | 201,053 | 201,160 | 2.86 | .2 | .448 | .259 | .0882 | .202 | .0212 | .0878 |
| Bandwidth = 5.70 | | 1x | 1x | 1x | 1x | 1x | 1x | 1x | 1x | 1x | 1x | 1x | 1x | 1x | 1x | 1x |

## Panel C: Women

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) | (13) | (14) | (15) | (16) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | transfer hours | prior gpa | HS stu. rank | max(SAT score, ACT conc. score) | appl. fin. aid | 1st term any coll. | 1st term this coll. | GPA excl.sample courses | freshman | sophomore | junior | senior | 1st gen. stu. | intl. stu. | transfer stu. |
| Above letter grade cutoff | | −0.5637 | −0.0123 | −4.5592 | 4.1412 | 0.0024 | 0.1671 | −3.0615 | −0.0313 | −0.0089 | −0.0140 | 0.0270*** | −0.0052 | 0.0028 | −0.0044 | −0.0221** |
| | | (0.4406) | (0.0671) | (3.0534) | (3.5846) | (0.0098) | (11.6287) | (9.6781) | (0.0449) | (0.0131) | (0.0114) | (0.0059) | (0.0081) | (0.0163) | (0.0035) | (0.0081) |
| Observations | | 9,635 | 9,505 | 7,341 | 10,286 | 10,510 | 10,510 | 10,510 | 10,510 | 10,510 | 10,510 | 10,510 | 10,510 | 10,510 | 10,508 | 10,510 |
| Variable mean | | 24 | 2.39 | 47.5 | 1,206 | .581 | 201,018 | 201,120 | 2.97 | .234 | .462 | .233 | .069 | .209 | .00971 | .0559 |
| Bandwidth = 7.82 | | 1x | 1x | 1x | 1x | 1x | 1x | 1x | 1x | 1x | 1x | 1x | 1x | 1x | 1x | 1x |

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ . This table shows estimated coefficients for each observable student characteristic in the data. As a test of the identifying assumption, we expect these estimate to be close to zero to show that student characteristics vary smoothly across the letter grade threshold. I plot these data in Figure A.2. Out of all variables and panels, six estmates are significant, which is roughly what we should expect to see due to random chance. Estimated with a local linear regression and uniform kernel, with bandwidths determing using the same method as for the main results, so that the bandwidths match for men and women. Standard errors are in parenthesis and clustered at the instructor and term level.

Table A.3: “Donut” RD: Graduated with a STEM major: omitting X percentage points of number grade on either side of threshold

Panel A: Men

| | (1)<br>(0.4) | (2)<br>(0.4) | (3)<br>(0.4) | (4)<br>(0.8) | (5)<br>(0.8) | (6)<br>(0.8) | (7)<br>(1.2) | (8)<br>(1.2) | (9)<br>(1.2) | (10)<br>(1.6) | (11)<br>(1.6) | (12)<br>(1.6) | (13)<br>(2.0) | (14)<br>(2.0) | (15)<br>(2.0) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Above letter grade cutoff | 0.0430 | 0.0355 | 0.0329 | 0.0158 | 0.0143 | 0.0132 | −0.0224*** | −0.0163 | −0.0184 | 0.0050 | 0.0095 | 0.0087 | −0.0028 | −0.0036 | −0.0002 |
| | (0.0261) | (0.0221) | (0.0250) | (0.0109) | (0.0139) | (0.0205) | (0.0045) | (0.0172) | (0.0211) | (0.0278) | (0.0263) | (0.0304) | (0.0241) | (0.0219) | (0.0312) |
| Observations | 8,167 | 8,167 | 8,167 | 7,491 | 7,491 | 7,491 | 6,867 | 6,867 | 6,867 | 6,212 | 6,212 | 6,212 | 5,606 | 5,606 | 5,606 |
| Outcome mean | .717 | .717 | .717 | .718 | .718 | .718 | .717 | .717 | .717 | .714 | .714 | .714 | .709 | .709 | .709 |
| Instructor & term FEs | | Y | Y | | Y | Y | | Y | Y | | Y | Y | | Y | Y |
| Other control vars. | | | Y | | | Y | | | Y | | | Y | | | Y |
| Opt. Bandwidth = 5.72 | 1x | 1x | 1x | 1x | 1x | 1x | 1x | 1x | 1x | 1x | 1x | 1x | 1x | 1x | 1x |

Panel B: Women

| | (1)<br>(0.4) | (2)<br>(0.4) | (3)<br>(0.4) | (4)<br>(0.8) | (5)<br>(0.8) | (6)<br>(0.8) | (7)<br>(1.2) | (8)<br>(1.2) | (9)<br>(1.2) | (10)<br>(1.6) | (11)<br>(1.6) | (12)<br>(1.6) | (13)<br>(2.0) | (14)<br>(2.0) | (15)<br>(2.0) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Above letter grade cutoff | 0.0122 | 0.0018 | −0.0019 | 0.0164 | 0.0068 | 0.0030 | 0.0095 | 0.0075 | 0.0026 | 0.0168 | 0.0147 | 0.0107 | 0.0108 | 0.0095 | 0.0044 |
| | (0.0198) | (0.0160) | (0.0182) | (0.0209) | (0.0164) | (0.0196) | (0.0189) | (0.0191) | (0.0226) | (0.0304) | (0.0281) | (0.0316) | (0.0324) | (0.0330) | (0.0359) |
| Observations | 9,627 | 9,627 | 9,627 | 9,049 | 9,049 | 9,049 | 8,505 | 8,505 | 8,505 | 7,924 | 7,924 | 7,924 | 7,407 | 7,407 | 7,407 |
| Outcome mean | .528 | .528 | .528 | .527 | .527 | .527 | .525 | .525 | .525 | .523 | .523 | .523 | .522 | .522 | .522 |
| Instructor & term FEs | | Y | Y | | Y | Y | | Y | Y | | Y | Y | | Y | Y |
| Other control vars. | | | Y | | | Y | | | Y | | | Y | | | Y |
| Opt. Bandwidth = 7.80 | 1x | 1x | 1x | 1x | 1x | 1x | 1x | 1x | 1x | 1x | 1x | 1x | 1x | 1x | 1x |

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ . This table shows estimates from regressions where observations are omitted from the middle on both sides of the threshold, for some x percentage points of number grade. The first three columns show results omitting 0.4 percentage points on either side of the cutoff, the next three columns omit 0.8 p.p., and so on up to 2.0 p.p. This is to address the potential concern that instructors draw letter grade cutoffs in part because of student characteristics that are unobservable to the researcher. This table shows that this is not a concern since the result is still a zero and is remains relatively precise even when excluding those students closest to the threshold. Bandwidths are fixed to match the main results. Estimated with a local linear regression and uniform kernel. Standard errors are in parenthesis and clustered at the instructor and term level.

Table A.4: Graduated with a STEM major: subsample of obs. without missing variables

Panel A: Men

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Above letter grade cutoff | 0.0348 | 0.0507 | 0.0427 | 0.0455 | 0.0470 | 0.0316 |
| | (0.0465) | (0.0316) | (0.0453) | (0.0325) | (0.0472) | (0.0305) |
| Observations | 4,846 | 4,846 | 4,846 | 5,986 | 7,048 | 8,928 |
| Outcome mean | .732 | .732 | .732 | .735 | .734 | .728 |
| Instructor & term FEs | | Y | Y | Y | Y | Y |
| Other control vars. | | | Y | Y | Y | Y |
| Bandwidth = 5.70 | 1x | 1x | 1x | 1.25x | 1.5x | 2x |

Panel B: Women

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Above letter grade cutoff | 0.0258 | 0.0020 | –0.0010 | 0.0036 | 0.0031 | –0.0089 |
| | (0.0239) | (0.0173) | (0.0408) | (0.0216) | (0.0291) | (0.0871) |
| Observations | 5,925 | 5,925 | 5,925 | 7,140 | 8,297 | 10,291 |
| Outcome mean | .532 | .532 | .532 | .535 | .538 | .537 |
| Instructor & term FEs | | Y | Y | Y | Y | Y |
| Other control vars. | | | Y | Y | Y | Y |
| Bandwidth = 7.82 | 1x | 1x | 1x | 1.25x | 1.5x | 2x |

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ . This table repeats the estimates from Table 2.3 but with a subsample of the data containing only those observations for which no variable is missing. This subsample may not be a representative sample if variables are missing nonrandomly, and the smaller sample has less statistical power than the main specification. The point of this table is to show that the main results are robust to alternative methods of addressing the issue of missing variables. Estimated with a local linear regression and uniform kernel. Standard errors are in parenthesis and clustered at the instructor and term level.

Table A.5: Graduated with a STEM major: results using multiple imputation for missing variables in columns 3–6

Panel A: Men

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Above letter grade cutoff | 0.0270 | 0.0304 | 0.0266 | 0.0302 | 0.0300 | 0.0178 |
| | [0.0273] | [0.0206] | [0.0206] | [0.0158]* | [0.0136]** | [0.0106] |
| Observations | 8,654 | 8,654 | 8,654 | 10,698 | 12,586 | 15,965 |
| Outcome mean | .697 | .697 | .697 | .698 | .699 | .695 |
| Instructor & term FEs | | Y | Y | Y | Y | Y |
| Other control vars. | | | Y | Y | Y | Y |
| Opt. Bandwidth = 5.72 | 1x | 1x | 1x | 1.25x | 1.5x | 2x |

Panel B: Women

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Above letter grade cutoff | 0.0154 | 0.0045 | 0.0003 | 0.0034 | 0.0022 | –0.0076 |
| | [0.0157] | [0.0145] | [0.0146] | [0.0135] | [0.0094] | [0.0084] |
| Observations | 10,078 | 10,078 | 10,078 | 12,246 | 14,267 | 17,778 |
| Outcome mean | .522 | .522 | .522 | .519 | .516 | .518 |
| Instructor & term FEs | | Y | Y | Y | Y | Y |
| Other control vars. | | | Y | Y | Y | Y |
| Opt. Bandwidth = 7.80 | 1x | 1x | 1x | 1.25x | 1.5x | 2x |

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ . This table shows estimates of the effect of letter grade on the main outcome of interest, graduation with a STEM major within six years. For this table, I use multiple imputation to account for variables that are missing in some observations. The point estimates here are directly comparable to the main results in Table 2.3, although the standard errors here are clustered at the instructor level because I cannot compute two-way clustered errors with the multiple imputation command. Estimated with a local linear regression and uniform kernel.

Table A.6: Graduated with a STEM major: results by individual letter grade thresholds

Panel A: Men

| | (1) (A) | (2) (A) | (3) (A) | (4) (B) | (5) (B) | (6) (B) | (7) (C) | (8) (C) | (9) (C) |
|---|---|---|---|---|---|---|---|---|---|
| Above letter grade cutoff | 0.0281 | 0.0406*** | 0.0381* | 0.0510 | 0.0345 | 0.0353 | 0.1009* | 0.1077** | 0.0978** |
| | (0.0204) | (0.0122) | (0.0199) | (0.0471) | (0.0356) | (0.0344) | (0.0484) | (0.0407) | (0.0439) |
| Observations | 1,989 | 1,989 | 1,989 | 3,811 | 3,811 | 3,811 | 2,197 | 2,197 | 2,197 |
| Outcome mean | .855 | .855 | .855 | .702 | .702 | .702 | .532 | .532 | .532 |
| Instructor & term FEs | | Y | Y | | Y | Y | | Y | Y |
| Other control vars. | | | Y | | | Y | | | Y |
| Opt. Bandwidth | 3.38 | | | 6.48 | | | 6.48 | | |

Panel B: Women

| | (1) (A) | (2) (A) | (3) (A) | (4) (B) | (5) (B) | (6) (B) | (7) (C) | (8) (C) | (9) (C) |
|---|---|---|---|---|---|---|---|---|---|
| Above letter grade cutoff | 0.0416 | 0.0183 | 0.0077 | −0.0293 | −0.0250 | −0.0347 | 0.0010 | −0.0207 | −0.0228 |
| | (0.0387) | (0.0323) | (0.0359) | (0.0205) | (0.0301) | (0.0329) | (0.0308) | (0.0251) | (0.0273) |
| Observations | 1,827 | 1,827 | 1,827 | 3,619 | 3,619 | 3,619 | 3,716 | 3,716 | 3,716 |
| Outcome mean | .665 | .665 | .665 | .525 | .525 | .525 | .408 | .408 | .408 |
| Instructor & term FEs | | Y | Y | | Y | Y | | Y | Y |
| Other control vars. | | | Y | | | Y | | | Y |
| Opt. Bandwidth | 4.13 | | | 6.80 | | | 10.83 | | |

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ . This table shows estimates of the effect of letter grade on persistance in STEM major for each letter grade threshold separately. Estimated with a local linear regression and uniform kernel. Standard errors are in parenthesis and clustered at the instructor and term level.

Table A.7: Graduated with an economics or business major: data from econonomics courses only

Panel A: Men

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | Grad. with economics major | | | Grad. with business major | | |
| Above letter grade cutoff | –0.0485 | –0.0218 | –0.0207 | 0.0683* | 0.0418 | 0.0281 |
| | (0.0238) | (0.0165) | (0.0117) | (0.0280) | (0.0318) | (0.0304) |
| Observations | 3,555 | 3,555 | 3,555 | 3,322 | 3,322 | 3,322 |
| Outcome mean | .154 | .154 | .154 | .306 | .306 | .306 |
| Instructor & term FEs | | Y | Y | | Y | Y |
| Other control vars. | | | Y | | | Y |
| Opt. Bandwidth = | 6.28 | 6.28 | 6.28 | 5.82 | 5.82 | 5.82 |

Panel B: Women

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | Grad. with economics major | | | Grad. with business major | | |
| Above letter grade cutoff | –0.0059 | 0.0035 | 0.0078 | 0.0373 | 0.0224 | 0.0190 |
| | (0) | (0.0059) | (0.0135) | (0.0356) | (0.0476) | (0.0526) |
| Observations | 3,524 | 3,524 | 3,524 | 3,279 | 3,279 | 3,279 |
| Outcome mean | .0857 | .0857 | .0857 | .433 | .433 | .433 |
| Instructor & term FEs | | Y | Y | | Y | Y |
| Other control vars. | | | Y | | | Y |
| Opt. Bandwidth = | 8.52 | 8.52 | 8.52 | 7.86 | 7.86 | 7.86 |

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ . This table shows results for only economics courses, looking at the effect of letter grade on graduation with an economics or business major within six years (columns 1–3 and 4–6, respectively). Estimated with a local linear regression and uniform kernel. Standard errors are in parenthesis and clustered at the instructor and term level. Note that the number of instructor clusters is quite low, with only four economics instructors in these regressions. This is why I cannot compute the standard error for women in column (1).

Table A.8: Results for alternative sample groups and outcomes: impact on different majors and estimated earnings

Panel A: Men

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Above letter grade cutoff | 0.0293 | 0.0272 | –0.0326 | –0.0146 | 1156.2427** |
| | (0.0222) | (0.0182) | (0.0209) | (0.0228) | (407.7100) |
| Observations | 8,636 | 13,474 | 7,735 | 7,354 | 8,540 |
| Outcome mean | .717 | .62 | .194 | .797 | 76,514 |
| Instructor & term FEs | Y | Y | Y | Y | Y |
| Opt. Bandwidth = | 5.70 | 6.58 | 7.53 | 8.76 | 6.34 |

Panel B: Women

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| Above letter grade cutoff | 0.0051 | 0.0037 | 0.0414** | 0.0118 | 500.0893 |
| | (0.0142) | (0.0128) | (0.0160) | (0.0256) | (627.4462) |
| Observations | 10,098 | 14,385 | 10,077 | 2,412 | 8,841 |
| Outcome mean | .528 | .456 | .263 | .631 | 61,248 |
| Instructor & term FEs | Y | Y | Y | Y | Y |
| Opt. Bandwidth = | 7.82 | 8.64 | 8.78 | 8.41 | 7.37 |

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$ . This table shows, for different subsets of the courses in the data, the impact of letter grades in those course on different outcomes. Column (1) repeats the main result from above using instructor and term fixed effects. Column (2) expands the course pool to include economics classes and looks at the effect of letter grade on the probability of graduating with a STEM or economics degree within six years. Column (3) uses only science courses including biology, calculus, and organic chemistry, and looks at the impact on majoring in a science field. Column (4) uses only engineering courses including calculus, computer programming, and statics and looks at the impact on majoring in engineering. Column (5) uses the STEM courses from the main sample (without economics classes) and looks at the impact of letter grade thresholds on national median earnings by major, in 2013 dollars. Estimated with a local linear regression and uniform kernel. Standard errors are in parenthesis and clustered at the instructor and term level.