THE UNIVERSITY OF CHICAGO

# ESSAYS IN THE ECONOMICS OF HIGHER EDUCATION

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE SOCIAL SCIENCES

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

KENNETH C. GRIFFIN DEPARTMENT OF ECONOMICS

BY

SIDHARTH SAH

CHICAGO, ILLINOIS

JUNE 2024

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# ACKNOWLEDGMENTS

I thank Michael Dinerstein, Christina Brown, and Jack Mountjoy for their invaluable guidance. Chapter 1 of this dissertation would not have been possible without the assistance of Robert Dixon, Stephanie Estrada, Bill Hayward, Ben Ost, Gabriela Valencia, and others at the University of Illinois Chicago. I am grateful to Stephane Bonhomme, Manasi Deshpande, Tom Hierons, Ali Hortacsu, Derek Neal, Evan Rose, and the participants of the Public/Labor Advising Group and the Student Applied Micro Lunch for useful comments and feedback.

I am, of course, deeply indebted to my parents for their support throughout the long educational journey that led to this point and to my partner Dana for her love.

# ABSTRACT

This dissertation is comprised of two chapters regarding the economics of post-secondary education. The first concerns effects of peer gender: Large gender differences exist in the take-up and completion of college majors across academic fields. The degree of gender concentration within fields tends to increase over time spent in college. In this chapter, I investigate how the gender composition of peers in first-semester classes impacts women's and men's academic outcomes and major choices. I find that a larger proportion of male peers hurts female academic achievement and decreases female persistence in majors, relative to men in the same classes.

The second chapter regards collusion in the determination of financial aid: Collusion is an important economic phenomenon that may play a role in many industries. This paper focuses on a case of alleged collusion in the determination of financial aid offers among a set of elite US universities known as the "Overlap Group." A Justice Department suit alleged that collusion led to an increase in the average effective price of attendance for students at these universities. A simple model of price competition between schools suggests that this would be the case. Applying conventional difference-in-difference and synthetic controls methods to the case of the Overlap Group suggests that the cessation of collusion did lead to a reallocation of university budgets towards financial aid.

# CHAPTER 1

# PEER GENDER COMPOSITION AND UNDERGRADUATE ACHIEVEMENT AND MAJOR CHOICE

## 1.1 Introduction

While women attend and graduate from college at comparable or higher rates than men, large gender differences persist in the take-up and completion of different college majors. Gender concentration within fields of study has important implications for both equity and efficiency. Given that men dominate many of the majors associated with the highest wages, such as engineering and economics, major choice may be an important contributing factor to the gender pay gap (Brown and Corcoran [1997], Gemici and Wiswall [2014], Patnaik et al. [2020]). Moreover, if there is reason to believe that gendered sorting to majors, and the occupations associated with those majors, does not reflect sorting to comparative advantage, this "friction" may dampen overall economic production (Hsieh et al. [2019]).

A substantial amount of work has gone into understanding the sources of gender differences in college major choice, with proposed factors ranging from high school preparedness (Card and Payne [2017], Aucejo and James [2021]), to preferences over non-pecuniary aspects of major-related occupations (Zafar [2013], Wiswall and Zafar [2017]), to competitiveness (Buser et al. [2014]), to the availability of role models (Carrell et al. [2010]). This paper considers whether peer gender composition within college classes contributes to this gap. In equilibrium, peer composition may reinforce the gendered sorting to fields that stems from other sources if students tend to choose majors related to the classes where they have more same-gender peers.

I investigate this question using administrative data from the University of Illinois Chicago (UIC), a large, public research university. The data provides information on all class registrations, class outcomes, major declarations, and semesters of graduation for two entering cohorts of undergraduate students. I focus on the classes that students enroll in during their first semester at the university based on the idea that incoming students have not yet met their peers and thus cannot coordinate or intentionally select into specific classes based on their gender compositions.

For my primary empirical strategy, I estimate how the gap between female and male outcomes evolves with the class-level gender ratio. I control for course-specific female fixed effects, utilizing variation in peer composition across different lecture times within courses. Doing so accounts for any potential gender differences in course-specific tastes or academic preparation. Given that I am estimating an effect on the *gap* between women and men, I am also able to include fixed effects for each specific lecture time in each semester for each course, in order to control for any gender-neutral sorting or shocks to these specific offerings of the courses. I find that when first-semester classes have more men, women tend to receive worse grades, are less likely to graduate, and are less likely to choose majors associated with those classes, all relative to men attending the same lectures. Placebo tests that use pre-college variables as outcomes show that these results are not driven by observable characteristics of students, such as pre-college academic attainment.

The aforementioned results characterize female outcomes *relative* to male ones. However, it is not immediately clear whether those patterns are driven by the behavior of women, the behavior of men, or both. By omitting the controls for specific iterations of courses, I am able to separately look at the effects of peer composition on male students and on female students. Generally, it seems as though women have worse achievement, in terms of grades and eventual graduation likelihood, in the presence of more men, while men are relatively unaffected by gender ratio for these outcomes. When it comes to major choice, however, it is the case both that women are less likely to declare majors related to male-heavy classes and that men are more likely to declare such majors. If anything, the effects on choice of major are driven mostly by male students.

I next consider what mechanisms might plausibly underlie my results on major choice. Theoretical work on major choice treats the decision as a dynamic problem in which students come into college facing uncertainty regarding their own field-specific abilities and tastes, learn about themselves during early classes, and then decide on a major (Arcidiacono [2004], Arcidiacono et al. [2012], Stinebrickner and Stinebrickner [2014], Arcidiacono et al. [2016]). Under such a framework, peer gender could enter into major choice decisions in two broad ways: students whose grades are affected by peer gender may update their beliefs about their field-specific ability or peer gender may influence beliefs about field-specific tastes. I find both that women receive worse grades in more male-heavy classes and that they are subsequently less likely to opt into majors related to those classes. This suggests that peer composition may affect female major choice indirectly through grades, although it is hard to rule out the possibility of tastes being an alternative or complementary mechanism. On the other hand, it is difficult to explain men's persistence in majors with male-dominated classes via beliefs about ability, as men do not receive better grades in more male-dominated classes.

If grades do not drive the positive effect of male peers on men's major choice, this implies that tastes play a role. I provide suggestive evidence that formation of within-major friendships may provide a plausible "taste-based" mechanism. Specifically, I consider students who enroll in both parts of various two-course sequences that are required for popular majors at UIC. When men have more male peers in their first class in one of these sequences, they tend to have more repeat peers from the first class in their second class in the sequence. It appears that this result reflects a behavioral effect, as the estimates I get for this test are large compared to a distribution of coefficients generated by simulations of random movement into classes. Thus, it may be that case that men form more friendships in male-heavy classes and then choose to study with those friends in the future. This may encourage the choice to declare majors related to those classes.

These results connect to an existing literature on the effects of peer gender in educational contexts. There has been a great deal of work focusing on primary and secondary school contexts. This strand of the literature has generally found that male peers are worse than female peers for the academic performance of all students, although whether this is found to matter more to girls or boys varies across studies and contexts (Hoxby [2000], Whitmore [2005], Lavy and Schlosser [2011], Black et al. [2013], Hu [2015], Gong et al. [2021]).

Some more recent work has considered the impact of peer gender in post-secondary education. Focusing on achievement, De Giorgi et al. [2010] find non-linear effects of class-level gender composition on grades for all students, with a roughly equal gender balance being optimal. Oosterbeek and van Ewijk [2014] find little evidence that the gender composition of workgroups within a class has any impact on outcomes. Hill [2017] uses cross-cohort variation at US universities, in the spirit of Hoxby [2000], to provide evidence that a higher proportion of females in an overall freshman cohort modestly increases male graduation rates. Looking at doctoral programs in STEM fields, Bostwick and Weinberg [2018] find that having more women in a cohort increases degree completion for other women.

The previous papers that are most comparable to this one are Griffith and Main [2019] and Zolitz and Feld [2020]. Griffith and Main [2019] exploit random assignment to an introductory class for students entering an undergraduate engineering program at a US university. They find that having more female students in a class increases both grades and persistence beyond the first year of the program for male and female students. Zolitz and Feld [2020] similarly make use of random assignment in the context of compulsory, introductory classes at a Dutch business school. They find that having classes with more female peers increases the likelihood that both men and women select into majors that are more dominated by their own gender.

I contribute to this existing body of work by considering a context where students enter college outside of any particular program and can consider a full array of majors. This allows me to consider heterogeneity across academic fields: I find that both the negative effects of male peers on women and the positive effects of male peers on men are strongly concentrated among male-majority departments. This implies that peer gender effects may be particularly important for many STEM and many highly paid fields. I also explicitly consider whether the the short-term outcome of grades mediates effects on the longer-term outcome of major choice. This prompts my finding that grades do not appear to be the mechanism linking male peers to greater male persistence in majors and that increased friendship formation may be a plausible alternative.

The rest of the paper proceeds as follows. Section 1.2 describes the administrative data and provides institutional context regarding UIC. Section 1.3 describes my empirical strategy and presents the results of balance tests used to validate the strategy. Section 1.4 presents and discusses the results as well as various robustness checks. Section 1.5 concludes.

## 1.2 Data and Institutional Context

My analysis utilizes an administrative dataset from the University of Illinois Chicago, a large, public research university. UIC is one of three universities in the University of Illinois System, and enrolls approximately 21,000 undergraduate students per year. The data covers all undergraduate students who first enrolled at UIC in the Fall of 2015 or Fall of 2016 semesters, including transfers, for a total of 9,797 students.

My identification strategy relies on the assumption that incoming first-semester students do not sort into classes based on the peer composition of those classes, as they have not yet had a formal opportunity to meet their peers. At UIC, incoming students (both freshmen and transfers) are required to attend an on-campus orientation session prior to their first semester of classes. In both summer 2015 and 2016, students attended one of fourteen sessions offered over the course of the summer, registering for their preferred session on a first-come, first-serve basis. At the summer sessions, students met with academic advisors and received course recommendations based on their stated academic interests, prior credits (from either AP/IB examinations or previous college enrollments), and their performance on placement tests taken prior to the orientation. After this meeting, students were free to sign up for classes, subject to capacity constraints.

For the students in the dataset, I observe all class enrollments and outcomes, including grades and class withdrawals, for every semester the student is enrolled at UIC through six or seven years post-matriculation (for students entering in 2015 or 2016, respectively). Here and subsequently I use the term “class" to refer to a specific offering of a “course" that is uniquely identified by a particular semester, lecture time, and instructor. For instance, I would refer to Economics 101 as a “course" and the Fall of 2015, 9 A.M. lecture time for Economics 101 as a “class." I observe the name of the instructor for each class in the set of student-class observations, and, if the class has any associated discussion, laboratory, or experiential sections, the names of the teaching assistants (TAs) that lead those sections. While I do not directly observe characteristics of the instructors or TAs, I infer gender from names.[1] For each student in the sample, the administrative data provides information on several background characteristics, including race, ethnicity, gender, and pre-college academic achievement in the form of high school GPA and a composite ACT score. I also observe all major declarations for all students in the sample. Major information is by semester, meaning that I observe when each student first declares a major and if and when they switch majors.[2] If a student graduates from UIC during the covered period, the data also shows their semester of graduation.

As previously mentioned, my analysis focuses on incoming first-semester students. I thus restrict my main sample to student-class-level observations corresponding to classes taken during a student's first semester. My primary estimating equation, equation (1.1), includes

1. I predict gender using the R package *gender*, which utilizes Social Security data (Mullen [2021]).

2. At UIC, the large majority of students start out “undeclared" and first declare a major some time after their initial semester.

both course-specific female fixed effects and class fixed effects. Estimation of my parameter of interest - the differential effect of class-level male proportion on females compared to males - thus requires observations from courses with multiple classes, each of which has at least one female and one male student. I thereby eliminate observations from all courses and classes that do not meet this requirement from the main sample. This eliminates about 6.6 percent of first-semester observations. The remaining sample consists of 43,351 student-class observations.

Table 1.1 reports descriptive statistics for the main sample. Panel A reports background characteristics broken down by gender. Entering UIC, men and women look fairly similar, with comparable ethnic compositions and standardized test scores (the differences in means for most of these variables are statistically significant but small in magnitude, compared to the variation within groups). Female students have a statistically significantly advantage in mean high school GPA over their male peers, reflecting the pattern found in the overall population, although, again, the size of the gap is modest (a mean GPA of 3.35 for women compared to 3.22 for men). The overall sample has an average ACT composite score of approximately 24.5, which would place the mean student at approximately the 75$^{\text{th}}$ percentile of test-takers, indicating a student body that is academically above average among college-interested students.[3]

Panel B reports individual-level outcomes for the students in the main sample and panel C reports student-class-level outcomes and characteristics. Women tend to academically outperform men at UIC, being more likely to graduate within six years, earn higher grades, and pass their classes. However, despite being more likely to graduate in general, women are substantially less likely to graduate with a STEM major. Among students who graduate within six years, approximately 63% of male students graduate in STEM fields compared to

3. This is based on the 2022-2023 reporting year statistics from ACT, Inc: https://www.act.org/content/dam/act/unsecured/documents/MultipleChoiceStemComposite.pdf

Table 1.1 – Descriptive Statistics by Gender - Main Sample

| | Women | Men | P-Value of Difference |
|---|---|---|---|
| *Panel A: Student characteristics* | | | |
| Ethnicities | | | |
| White | 0.30 | 0.32 | 0.01 |
| | (0.46) | (0.47) | |
| Asian | 0.20 | 0.22 | 0.05 |
| | (0.40) | (0.41) | |
| Hispanic | 0.34 | 0.33 | 0.32 |
| | (0.47) | (0.47) | |
| African American | 0.10 | 0.07 | 0.00 |
| | (0.30) | (0.25) | |
| High school GPA | 3.35 | 3.22 | 0.00 |
| | (0.37) | (0.39) | |
| ACT composite score | 24.02 | 24.74 | 0.00 |
| | (4.02) | (3.93) | |
| *Panel B: Student outcomes* | | | |
| Graduate within 6 Years | 0.70 | 0.64 | 0.00 |
| | (0.46) | (0.48) | |
| Graduate with a STEM major within 6 Years | 0.34 | 0.40 | 0.00 |
| | (0.47) | (0.49) | |
| Observations | 4960 | 4594 | |
| *Panel C: Student-class outcomes and characteristics* | | | |
| Grade (GPA value) | 3.07 | 2.90 | 0.00 |
| | (1.05) | (1.14) | |
| Grade of B or higher | 0.81 | 0.78 | 0.00 |
| | (0.39) | (0.41) | |
| Passed class | 0.96 | 0.94 | 0.00 |
| | (0.20) | (0.24) | |
| Dropped class | 0.05 | 0.05 | 0.47 |
| | (0.22) | (0.22) | |
| Course only offers one class per year | 0.18 | 0.21 | 0.00 |
| | (0.39) | (0.41) | |
| Class has an associated section | 0.41 | 0.42 | 0.00 |
| | (0.49) | (0.49) | |
| Observations | 22162 | 21189 | |

Notes: This table reports the means and standard deviations, by gender, and p-values of t-tests of differences in means of variables, as estimated within the identifying set of the main estimating equation (as described in Section 1.2) for which the relevant variable is observed.

about 49% of women, highlighting the gender differences in choice of field of study.[4]

Panel C also shows that about 40% of student-class observations belong to courses with an associated section, where "section" refers to any additional discussion, laboratory, or practical experience session, typically led by a TA. This subsample, with some additional sample restrictions, is used for a robustness check which relies upon variation in the proportion of male peers in sections within classes, rather than variation in classes within courses. About 20% of student-class observations are from courses for which only one lecture time is offered per year. This subsample is used for a robustness check where the variation in class-level peer composition is based largely on between-cohort variation. More details on each of these alternate specifications is provided in Section 1.4.2.

## 1.3 Empirical Strategy

My aim is to estimate the differential effect of peer composition on the outcomes of women compared to men. I do so by exploiting variation in peer composition across classes within a course, where a "course" is defined by a department and course number, such as Economics 101, while a "class" is a specific offering of a course with a unique semester, lecture time, and instructor combination. I estimate how the difference in female and male outcomes evolves with the gender ratio across different offerings of the course. For each course, this may then involve variation between lecture times within a semester, say 9 AM compared to 11 AM, and may involve variation between students taking the course in the Fall of 2015 compared to the Fall of 2016.[5]

4. "STEM" is defined according to the 2022 Department of Homeland Security STEM Designated Degree Program List: https://www.ice.gov/doclib/sevis/pdf/stemList2022.pdf. The overall proportion of STEM graduates I observe is somewhat high in part due to the inclusion of certain majors that are not designated as STEM by other definitions, including psychology, economics, and certain pre-health majors.

5. For the results presented in the main body of the paper, I pool across both within-semester and across-semester variation in classes to maximize power. However, each of these two sources of variation introduces separate concerns for identification. I thus also present results using only within- and only across-semester variation. The version that is only within-semester is presented in Appendix A.1. The version that is only across-semesters is discussed in Section 1.4.2 and presented in Appendix A.2. Both sets of results are similar

I focus solely on first-semester students. By doing so, I ensure that the peer composition of classes was not observable to students when they initially enrolled. Students register for first-semester classes in the summer prior to matriculation, before they have had a formal opportunity to meet their peers (other than the relatively small subset who attend their same orientation session). Thus, the students in my sample were not directly selecting into peer groups when they chose classes.

By comparing within courses, I allow men and women to differ in course-specific aptitudes or preferences, accounting for the fact that men and women may have received different kinds of education or may have formed dissimilar interests prior to entering college. Moreover, because my primary empirical strategy estimates how the *gap* between women and men changes with peer composition, I am also able to account for class-level fixed effects. The class fixed effects allow for any kind of gender-neutral sorting to classes, within courses, or class-specific shocks. For instance, if more motivated students tended to take morning classes, as opposed to afternoon ones, this would be picked up by the class fixed effects, so long as the sorting behavior was similar between the male and female populations.

The remaining threat to identification stems from the possibility of differential sorting between men and women to classes. For instance, if women were both more likely to register for morning classes *and* the difference between morning-class and afternoon-class women was larger than the difference between morning-class and afternoon-class men, this would bias my estimates. I argue that this concern is minimal using balance tests, which I describe later in this section. I also perform multiple robustness checks intended to minimize the extent to which differential sorting may drive my results. I describe these alternative approaches in greater detail in Section 1.4.2.

I operationalize my identification strategy via the following econometric model of student-

to my main results, although naturally less precise. This suggests that neither form of variation is solely driving the results.

class-level outcomes, $Y_{i,r,c}$:

$$Y_{i,r,c} = \alpha_0 + \alpha_1 \times Fem_i \times MP_c + \gamma_r \times Fem_i \times I_r + \delta_c \times I_c + X'_{i,r,c}\beta + u_{i,r,c} \tag{1.1}$$

where students are indexed by $i$, courses by $r$, and classes by $c$.[6] $Fem_i$ is an indicator variable that takes a value of one if student $i$ is female, while $I_r$ and $I_c$ are indicator variables that take on values of one if outcome $Y_{i,r,c}$ is associated with course $r$ or class $c$, respectively. $MP_c$ denotes the proportion of male students in class $c$, $X_{i,r,c}$ contains a vector of observable student and student-class observables, and $u_{i,r,c}$ is an unobservable error term.[7]

The parameter of interest, $\alpha_1$, measures how the gap between female and male outcomes evolves with the class-level male proportion.[8] Given the focus on an interaction term, I am able to include both course-specific female fixed effects, $\gamma_r$, and general class fixed effects, $\delta_c$.[9] The course-specific female fixed effects restrict the identifying variation to be within-course. The class fixed effects pick up any gender-neutral sorting or shocks associated with specific classes, within a course.

In order to identify my parameter of interest, I assume that the residual variation in $Fem_i \times MP_c$, conditional on the fixed effects and controls, is orthogonal to the residual variation in the error. Essentially, I assume that, within a course, students sort to classes in such a way that certain types of male or female students are not more likely to end up

6. Classes are nested within courses. Thus, any variable with a $c$ subscript could alternatively be denoted with a double $r, c$ subscript. I omit the course subscripts in these cases for readability.

7. For regressions taking the form of equation (1.1), $X_{i,r,c}$ is generally composed of student underrepresented minority status (a dummy for Black, Hispanic, and native students) and instructor-student gender match. The specific set of controls used in each specification is described in the notes for the table reporting the corresponding results.

8. Specifically, this parameter measures how the gap between female and male outcomes evolves *linearly* with class-level male proportion. Appendix A.3 explores potential non-linearities in the effects. In general, it seems that both the absolute effects of male peers and the differences in effects of male peers on women compared to men are concentrated among the most male-dominated classes.

9. The focus on a gap between groups of students and use of fixed effects bears some resemblance to Fairlie et al. [2014], although that paper makes use of a combination of individual fixed effects and class fixed effects.

in more male-dominated classes.[10] The residual, identifying variation may come from a variety of sources, including capacity constraints on classes, student scheduling constraints, between-cohort variation in the numbers of men and women interested in each course, and idiosyncratic preferences for time slots. These sources may create noise in the class-level peer composition, which I argue is uncorrelated with the differences in male and female characteristics.

My fixed effect strategy addresses many forms of potential endogeneity. As previously mentioned, the remaining threat to identification stems from the possibility of differential sorting between men and women to classes. While I cannot fully rule out the possibility of such differential sorting, I do test for it by estimating equation (1.1) for pre-college attributes that may be predictive of college outcomes: standardized ACT score, high school GPA, and underrepresented minority status.[11] I present the results of this exercise in Table 1.2.

The first three columns show that there is little systematic association between class male proportion and the differences in male and female pre-college academic aptitude or ethnicity, conditional on the full set of fixed effects. As it may be difficult to interpret the magnitudes of these estimates, I also form predicted course grades by regressing the GPA value of grades on the three pre-college attributes that I tested. I then use these background characteristic-predicted grades as an additional outcome, reported in the fourth column of Table 1.2. The point estimate suggests that a woman in a 100 percent male class would only be expected to receive a grade that is worth 0.08 fewer GPA points than a women in a 0

10. Here, "sorting" refers both to within-semester selections of a specific class time slot and to between-cohort variation. For the between-cohort variation, I still need to assume that students are not deciding whether or not to register for a *course* based on knowledge of the gender composition of that course in that year.

11. Although I use high school GPA and ACT scores as outcomes in the placebo tests, I do not use them as controls in the specifications reported in the main body of the paper, as these variables are missing for a substantial portion of the sample. In Appendix A.4, I report results excluding any individual controls and results that include high school GPA and ACT scores as controls. The former set of results are nearly always very similar to those reported in the main text. The latter results generally have similar point estimates to those of my preferred specification, but are less precise due to the curtailed sample sizes.

Table 1.2 – Associations Between Class Male Proportion and Pre-College Characteristics

| | Standardized ACT score | High school GPA | Underrepresented minority student | Predicted grade (GPA value) |
|---|---|---|---|---|
| Female student X | 0.026 | -0.073 | -0.045 | -0.079 |
| class male % | (0.110) | (0.057) | (0.058) | (0.070) |
| Outcome Mean | 0.007 | 3.306 | 0.523 | 2.969 |
| Outcome SD | 1.000 | 0.392 | 0.499 | 0.410 |
| Observations | 31674 | 31702 | 43351 | 23018 |

Notes: This table reports the results of regressions of pre-college characteristics, at a student-class observation level, on an interaction between a female-student dummy and the class-level male proportion. Each column corresponds to a separate regression. Each regression includes course-specific female fixed effects, class fixed effects, and a control for instructor-student gender match. The predicted grade outcome is based on another (unshown) regression of the GPA value of grades on ACT score, high school GPA, and minority status among students in the main estimation sample. The predicted grade regression only includes the observations used to form the predicted grades: student-class observations from graded classes that had information on both ACT scores and high school GPA. Standard errors are clustered at the class level. * indicates significance at a level of 0.1, ** at a level of 0.05, and *** at a level of 0.01.

percent male class, based on observable characteristics. Along with not being statistically significant, this estimate is absolutely small, less than one tenth of the difference between an A and a B, and relatively small compared to the estimated effects of peer gender on grades that I report in Section 1.4.1. I interpret this as evidence in favor of the necessary assumption of no differential sorting.

When selecting classes, the notable characteristics that students observe are the time period and, for most classes, the instructor. Thus, one particularly salient potential source of differential sorting is instructor gender. If students were more likely to sort into classes with same-gender instructors and students either performed better in the presence of same-gender instructors or only a certain type of student sorted into classes with same-gender instructors, this would bias my results.[12] I investigate this specific threat directly in Table 1.3, showing that, within courses, there is no meaningful degree of sorting into or out of

12. There is recent literature suggesting that female instructors may improve the achievement of female college students (see for example Hoffman and Oreopoulos [2009] and Carrell et al. [2010]), although there are mixed findings regarding how female instructors affect the future course and major selections of female students (see for example Bettinger and Long [2005] and Price [2010]).

Table 1.3 – Sorting by Students to Female Instructors (Within Course)

| | Female student | Male student |
|---|---|---|
| Female instructor | 0.000* | -0.000 |
| | (0.000) | (0.001) |
| Outcome Mean | 0.511 | 0.486 |
| Observations | 43351 | 43351 |

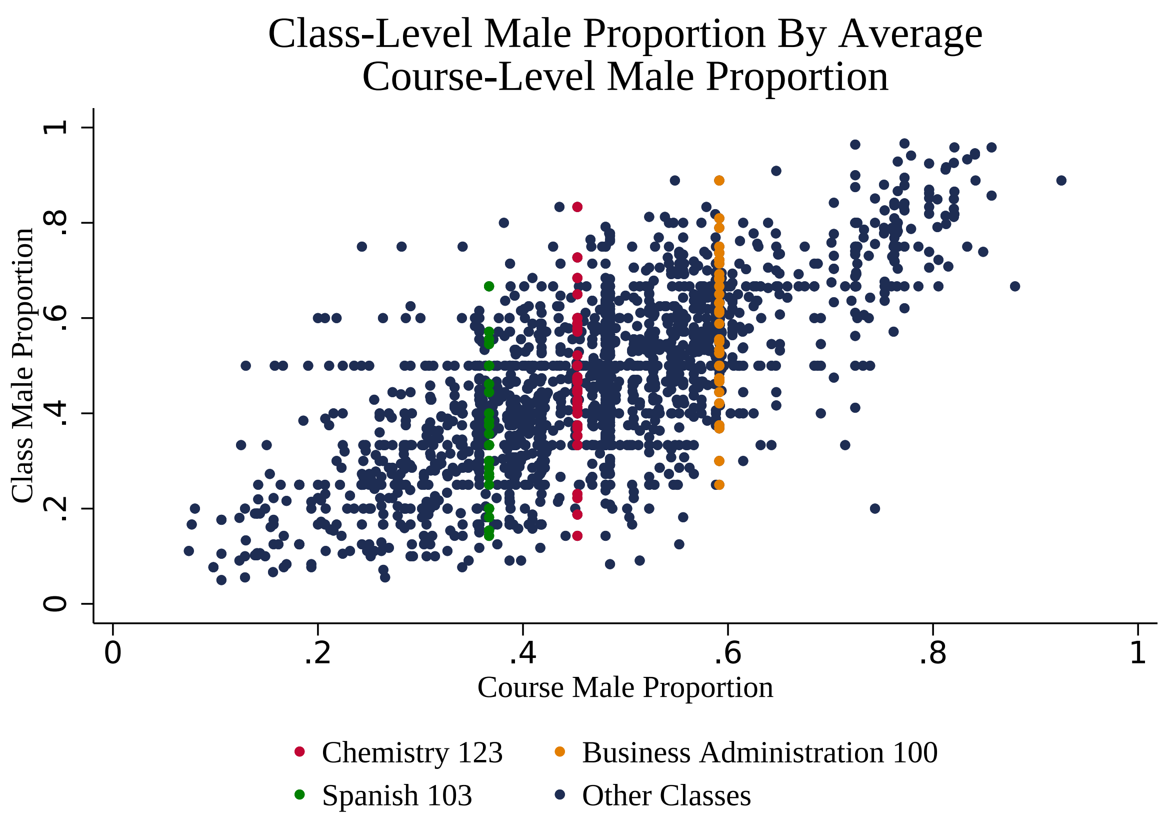
Notes: This table reports the results of regressions of dummies for student gender, defined at a student-class observation level, on a dummy for the class being taught by a female instructor. Each column corresponds to a separate regression. Each regression includes course fixed effects. Standard errors are clustered at the class level. * indicates significance at a level of 0.1, ** at a level of 0.05, and *** at a level of 0.01.

female-taught classes by either female or male students. Nonetheless, I include a control for instructor-student gender match in all specifications taking the form of equation (1.1).

Given that identification is only possible via variation in class-level male proportion within courses, it is of interest how much such variation exists in the data. Figure 1.1 plots class-level male proportion against average course-level male proportion. While there is naturally a high degree of correlation between the two, there is meaningful variation between classes within course, particularly for courses that are closer to the center of the distribution and have many different observed classes. I plotted the three non-general requirement courses that have the most classes in separate colors, revealing that all three have classes spanning much of the range of possible gender ratios. This is the case in spite of the on-average male domination of Business Administration 100 and on-average female domination of Spanish 103.

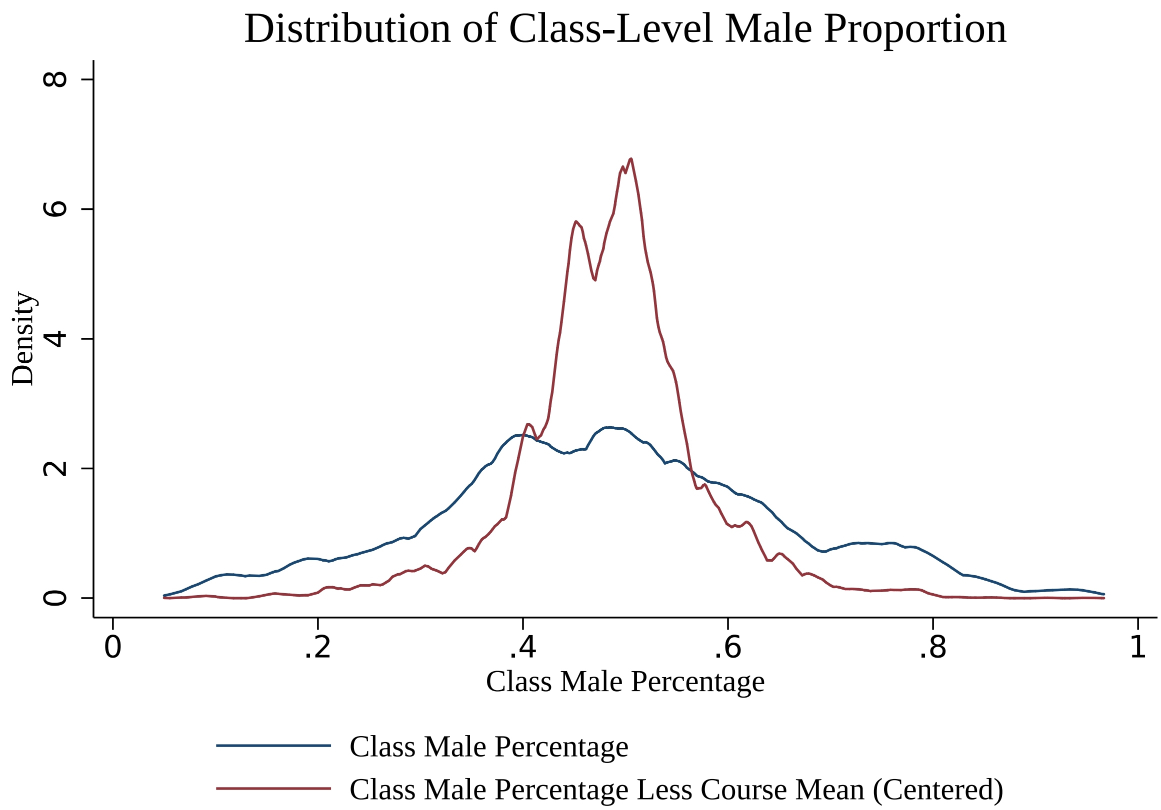
The exact variation that I exploit is plotted in Figure 1.2. The blue line gives the raw distribution of class-level male proportion in the main estimation sample while the maroon line displays the residual variation conditional on course fixed effects. As would be expected, taking out course-level variation substantially condenses the distribution. The remaining variation is concentrated within a span of about twenty percentage points.

Figure 1.1



Class-Level Male Proportion By Average
Course-Level Male Proportion
Class Male Proportion
0
.2
.4
.6
.8
1
Course Male Proportion
0
.2
.4
.6
.8
1
Chemistry 123
Business Administration 100
Spanish 103
Other Classes

Figure 1.2



Distribution of Class-Level Male Proportion
Density
0
2
4
6
8
0
.2
.4
.6
.8
1
Class Male Percentage
Class Male Percentage
Class Male Percentage Less Course Mean (Centered)

### *1.3.1 Within-Gender Empirical Strategy*

Estimating equation (1.1) can establish whether or not peer gender composition creates a separation in the outcomes of male and female students. However, it does not reveal whether this is driven by the outcomes of students of a particular gender or by the simultaneous behavior of both genders. In order to estimate how the female and male populations each separately respond to peer gender composition, I estimate several within-gender regressions, of the forms

$$Y_{i,r,c}^{F} = \alpha_0^{F} + \alpha_1^{F} \times MP_c + \gamma_r^{F} \times I_r + (X^{F})'_{i,r,c}\beta^{F} + u_{i,r,c}^{F} \tag{1.2}$$

$$Y_{i,r,c}^{M} = \alpha_0^{M} + \alpha_1^{M} \times MP_c + \gamma_r^{M} \times I_r + (X^{M})'_{i,r,c}\beta^{M} + u_{i,r,c}^{M} \tag{1.3}$$

where equation (1.2) is estimated only on the set of female students and equation (1.3) is estimated only on the set of males.

By necessity, these specifications exclude class fixed effects. Thus, if there is any kind of absolute sorting to classes that have more male students among students of either gender, this will threaten the validity of my results. While this provides a weaker argument for identification, I again perform the placebo test of putting pre-college characteristics (and the grades predicted by those pre-college characteristics) on the left-hand side of equations (1.2) and (1.3). The results of this test are presented in Table 1.4.

The placebo test shows no strong evidence of students of either gender sorting to more male-dominated classes, within course, in terms of observable background characteristics. Focusing on the predicted grade outcome, which has the most interpretable magnitude, the results indicate that moving from an all-female to an all-male class would be expected to shift a female student's grade by 0.035 GPA points and a male student's grade by 0.006 GPA points, based on the average association between class male percentage and student observables. These predictions are again small not only in an absolute sense but also relative to the estimated effect of male peers on female grades that I report and discuss in Section

Table 1.4 – Associations Between Class Male Proportion and Pre-College Characteristics, By Gender

| | Standardized ACT score | High school GPA | Underrepresented minority student | Predicted grade (GPA value) |
|---|---|---|---|---|
| *Only women:* | | | | |
| Class male % | 0.007 | -0.052 | -0.063 | -0.035 |
| | (0.074) | (0.036) | (0.041) | (0.044) |
| Outcome Mean | -0.080 | 3.371 | 0.506 | 3.005 |
| Outcome SD | 1.005 | 0.379 | 0.500 | 0.411 |
| Observations | 16383 | 16393 | 22162 | 12166 |
| *Only men:* | | | | |
| Class male % | -0.116 | -0.033 | 0.002 | 0.006 |
| | (0.087) | (0.041) | (0.042) | (0.047) |
| Outcome Mean | 0.102 | 3.235 | 0.542 | 2.929 |
| Outcome SD | 0.985 | 0.394 | 0.498 | 0.406 |
| Observations | 15291 | 15309 | 21189 | 10852 |

Notes: This table reports results of regressions of pre-college characteristics, at a student-class observation level, on class-level male proportion. Each cell corresponds to a separate regression with the outcome given by the column header. Top row results are estimated only on female students and bottom row results only on male students. Each regression includes class fixed effects and a control for instructor gender. The predicted grade outcome is based on another (unshown) regression of the GPA value of grades on ACT score, high school GPA, and minority status among students in the main estimation sample. The predicted grade regression only includes the observations used to form the predicted grades: student-class observations from graded classes that had information on both ACT scores and high school GPA. Standard errors are clustered at the class level. * indicates significance at a level of 0.1, ** at a level of 0.05, and *** at a level of 0.01.

1.4.3.

## 1.4 Results

### *1.4.1 Main Results*

I now turn to the estimation of how the gap between female and male outcomes evolves with peer gender composition. Regression coefficients for the interaction between being a female student and class-level male proportion on class and college outcomes are reported in Table 1.5. I report results from a variety of specifications with differing fixed effects. Results from my preferred specification, characterized by equation (1.1), are displayed in column 4.

In addition, the table presents results when including no fixed effects (column 1), including course and course-specific female fixed effects but no class fixed effects (column 2), and class fixed effects but no course-specific female fixed effects (column 3). Controls that are made redundant by fixed effects are excluded from the relevant specifications. Standard errors are clustered at the class level.

I estimate each model for six outcomes, exploring a range of short- and long-term effects. As an immediate outcome, I consider the GPA point value of the grade received in the class. The intermediate outcomes include dummy variables for whether or not a student is observed to take any future classes in the same academic department as the given class, whether the student switches major to another department (conditional on having declared a major in the department of the given class in the first semester), and whether the student goes on to declare a major in the same department.[13] Because only a small minority of students have declared majors in their first semester at UIC, the switching major outcome is estimated on only a small subset of students, and is consequently imprecise. The longest term outcomes are dummy variables for whether or not a student graduates within six years of enrollment and whether or not a student graduates with a declared major in the same academic department as the given class, again within six years of enrollment.

Column 1 of Table 1.5 reveals a general pattern of women having relatively worse academic performance and lower likelihood of persistence in more male-dominated academic departments. Introducing the full set of fixed effects in column 4 reveals the extent to which this pattern is explained by the causal effect of class-level peer gender composition. Starting with the immediate impact of male peers, the best estimate suggests that going from a class with no males to a class with all males would drive down the average female grade by a third of a GPA point, relative to men in the same class. This decrease is approximately one third

13. "Departments" are defined according to the UIC academic catalogue: https://catalog.uic.edu/ucat/degree-programs/degree-minors/. These departments can span multiple majors, although the number of majors is usually small. For instance, the English Department houses both the English major and the Teaching of English major.

Table 1.5 – Estimated Effect of Class Male Proportion on Student Outcomes

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| *Grade (GPA value) [Mean = 2.99]* | | | | |
| Female student X class male % | -0.539*** | -0.125 | -0.280*** | -0.309** |
| | (0.110) | (0.143) | (0.090) | (0.151) |
| Observations | 33071 | | | |
| | | | | |
| *Take a future course in same department [Mean = .64]* | | | | |
| Female student X class male % | -0.013 | -0.024 | -0.118*** | -0.067 |
| | (0.068) | (0.044) | (0.038) | (0.047) |
| Observations | 43351 | | | |
| | | | | |
| *Switch major to another department [Mean = .14]* | | | | |
| Female student X class male % | 0.372*** | -0.034 | 0.054 | 0.031 |
| | (0.056) | (0.111) | (0.064) | (0.146) |
| Observations | 4750 | | | |
| | | | | |
| *Declare a major in same department [Mean = .05]* | | | | |
| Female student X class male % | -0.018 | -0.053** | -0.024 | -0.047* |
| | (0.025) | (0.021) | (0.016) | (0.025) |
| Observations | 37153 | | | |
| | | | | |
| *Graduate within six years [Mean = .68]* | | | | |
| Female student X class male % | -0.208*** | -0.092* | -0.022 | -0.106* |
| | (0.038) | (0.052) | (0.033) | (0.056) |
| Observations | 43351 | | | |
| | | | | |
| *Graduate with a major in same department [Mean = .13]* | | | | |
| Female student X class male % | -0.417*** | -0.047* | -0.003 | -0.046* |
| | (0.075) | (0.025) | (0.024) | (0.028) |
| Observations | 43351 | | | |
| | | | | |
| *Fixed effects* | | | | |
| Course and course-female | No | Yes | No | Yes |
| Class | No | No | Yes | Yes |
| *Controls* | | | | |
| Student gender | Yes | No | Yes | No |
| Instructor gender | Yes | Yes | No | No |
| Class male % | Yes | Yes | No | No |

Notes: This table reports results of regressions of student-class outcomes on an interaction between a female-student dummy and class male proportion. Outcomes are given by row headers and fixed effects and controls specified by column feet. Standard errors are clustered at the class level. * indicates significance at a level of 0.1, ** at a level of 0.05, and *** at a level of 0.01.

of the difference between an A and a B.[14]

Considering other outcomes reveals that peer gender influences outcomes throughout the college career. My preferred specification suggests that having a first-year class with more men decreases the relative likelihood that women will go on to declare a major related to that class, graduate from college, and graduate with a major related to that class, all relative to male peers in the same class. All of these estimates are statistically significant and large relative to the mean likelihoods of these outcomes. These results suggest that peer gender composition influences the decision making of college students in ways that go beyond the impact within a specific class. Whether the longer-term outcomes are a direct function of the short-term grade outcome or not is a question that I will turn to in Section 1.4.4.

The results are generally similar, in terms of sign, across different specifications. Under the assumption that the results of the specification used in column 4 are the "truth," the high degree of similarity between those results and the column 2 results, which exclude class fixed effects, might be taken as evidence that the exclusion of class fixed effects is not a major threat to identification. This is reassuring for the interpretation of the within-gender results, which are estimated without class fixed effects and are presented in Section 1.4.3.

### *1.4.2 Robustness Checks*

The primary threat to the validity of the results presented in the prior section comes from the possibility of differential sorting, whereby the difference between men and women in a class is systematically related to the class gender ratio, within a course. In order to assuage these concerns, I report results from two sets of alternative specifications that may offer stronger arguments against differential sorting. The first set, reported in Appendix A.2, restricts to the set of courses for which only one class is offered per year. That is, if Economics 101

14. Greater context on the magnitudes of the estimates is provided in Section 1.4.3, discussing the implied effect sizes given the range of actually observed class-level male proportions. This discussion is postponed as it is easier to think about the magnitudes of effects on students of each gender than the magnitude of the effect on the gap between genders.

only had one lecture time in each of Fall of 2015 and Fall of 2016, Economics 101 would be included in this subset. For these courses, students in a given cohort do not have the ability to select between class offerings in ways that may be correlated with class gender makeup. The only way students could react to class characteristics so as to create differential sorting would be on the extensive margin of whether or not to take the course at all (during their first semester). Under the assumption that year-specific class characteristics do not have a large impact on course take-up, this specification largely relies on cohort-level variation. The results using this subsample of courses are generally similar to my main results, although less precise.

The other alternative specification focuses on peer gender composition in laboratory, discussion, or practical experience sections that are associated with classes. Because many classes have multiple associated sections, using section-level variation allows for the inclusion of class-specific female fixed effects, which would account for any kind of differential sorting to classes between men and women. The sorting to sections within classes might be considered “more idiosyncratic" than the higher level sorting to classes, as sections have tighter capacity constraints, may have less observable information than classes because some sections do not provide information on the TA in charge (and TAs may have less information available about them than faculty), and students may prioritize class selections over section selections, resulting in sections being subject to greater scheduling constraints (if class times are chosen “first" by students). The section-level analysis provides some supporting evidence for the results reported in the main paper, although the results are generally not statistically significant in the overall sample. Interestingly, it seems that section-level peer effects may be strongly concentrated in sections for female-majority classes. Appendix A.5 further discusses the section-level analysis and reports the results.

### *1.4.3 Results Within Gender*

I now consider whether the effects I find in Section 1.4.1 are driven more by the behavior of female or male students. Estimates of how class-level male proportion affects outcomes within gender, using regressions of the forms of equations (1.2) and (1.3), are presented in Table 1.6. It appears that having more males in a class may be harmful to both the grades received and the likelihood of taking a future course in the same department for female students. For male students, the estimated effects are negative, but small in magnitude and not statistically significant. The divergence between the two groups thus stems from the fact that women are more negatively affected by male peers than men are, rather than male students providing an academic benefit to one another. This finding is consistent with studies from both primary and secondary school contexts, which have generally found that more male-heavy academic environments are worse for the academic outcomes of all students, compared to more female-heavy ones (Hoxby [2000]; Lavy and Schlosser [2011]; Gong et al. [2021]). A variety of mechanisms have been proposed for this pattern, including disruptive behavior (Lavy and Schlosser [2011]) and teacher responses to class composition (Gong et al. [2021]). In the context of UIC, it is the case that that the male population has a lower average high school GPA than the female population. Thus, it might be suspected that the observed effect of male peers is really a function of low-ability peers. However, controlling for peer ability, in the form of average high school GPA, only increases the estimated negative effect of males on grades, for both men and women.[15] To the extent that ability is well captured by this measure, it appears that the effect of male students on achievement is not driven by ability and likely reflects other attributes of males or male-dominated environments.

There is a divergence in the signs of the effects of male peers on men versus women for longer term outcomes. In spite of the result on grades, it appears that, if anything, having more male peers makes male students more likely to graduate within six years. Female

15. Within-gender results including the peer ability control are reported in Appendix A.6.

Table 1.6 – Estimated Effect of Class Male Proportion on Student Outcomes, By Gender

| | Grade (GPA value) | Future course in dept. | Switch major out of dept. | Declare major in dept. | Graduate within 6 years | Graduate in dept. |
|---|---|---|---|---|---|---|
| *Only women:* | | | | | | |
| Class male % | -0.171 | -0.071** | -0.064 | -0.021 | -0.070** | -0.005 |
| | (0.106) | (0.029) | (0.065) | (0.016) | (0.034) | (0.018) |
| Outcome Mean | 3.068 | 0.663 | 0.141 | 0.060 | 0.707 | 0.135 |
| Outcome SD | 1.051 | 0.473 | 0.348 | 0.237 | 0.455 | 0.341 |
| Observations | 17284 | 22162 | 2537 | 18977 | 22162 | 22162 |
| *Only men:* | | | | | | |
| Class male % | -0.046 | -0.047 | -0.027 | 0.031** | 0.022 | 0.042** |
| | (0.129) | (0.036) | (0.095) | (0.015) | (0.039) | (0.019) |
| Outcome Mean | 2.901 | 0.625 | 0.135 | 0.049 | 0.647 | 0.119 |
| Outcome SD | 1.143 | 0.484 | 0.341 | 0.216 | 0.478 | 0.324 |
| Observations | 15787 | 21189 | 2213 | 18176 | 21189 | 21189 |

Notes: This table reports results of regressions of student-class outcomes on class-level male proportion. Each cell corresponds to a separate regression, with outcome given by the column header. Top row results are estimated only on female students and bottom row results only on male students. Each regression includes course fixed effects and controls for student underrepresented minority status and instructor gender. All regressions are estimated on the observations within the identifying set of the main specification (as described in Section 1.2) for which the relevant outcome is observed. Standard errors are clustered at the class level. * indicates significance at a level of 0.1, ** at a level of 0.05, and *** at a level of 0.01.

students, on the other hand, are significantly less likely to graduate when they have more male peers in their first-semester classes. As with grades, another achievement outcome, the effect on the gap between men and women appears to be driven primarily by a negative effect of male peers on female students.

Results on choice of major, however, appear to be driven more by men. While female students may be somewhat less likely to declare majors corresponding to their more male-dominated classes, there is a larger, positive effect of male peers on the likelihood of male students choosing a given major. Similarly, I estimate a null effect of male peers on female likelihood of graduating with a related major, contrasting with a significant, positive effect on the same outcome for male students. Taking these results together suggests that male students harm the academic achievement of female students while having more ambiguous effects on the achievement of other males. However, it seems that the presence of male students encourages other male students to persist in majors while having a more moderate impact on the choices of female students.

The estimates in Table 1.6 are linear in class-level male proportion. Thus, the reported numbers compare outcomes between classes with no men and classes with all men, which is an extreme comparison given the distribution of classes students are likely to take. Due to the inclusion of course fixed effects, these results are estimated using only within-course variation. In the estimation sample, going from a class at the 10th percentile of class male proportion to the 90th percentile, within a course, would correspond to a shift in the class-level male proportion of about 0.2.[16] The estimates in Table 1.6 would thus imply that going from a 10th percentile class to a 90th percentile class, within a course, would decrease female grades by 0.03 GPA points, reduce female likelihood of graduation by 1.4%, and increase the likelihood of men graduating in a related department by 0.8%, on average.

Of course, when considering the full range of classes both within *and* across courses, students are exposed to more extreme variation in peer composition. Across all classes in the estimation sample, going from a a class at the 10th percentile of class male proportion to the 90th percentile would imply a change in class male proportion of 0.46. If my estimates are externally valid to how peer composition affects outcomes when comparing any two classes, rather than just two classes within a course, then going from a 10th percentile class to a 90th percentile class would be expected to decrease female grades by 0.08 GPA points and increase the likelihood of men graduating in a related department by 1.9%. However, it is difficult to gauge the extent to which my estimates may be valid for these kinds of comparisons.

The results presented in Tables 1.5 and 1.6 are also aggregated across courses from all academic departments at UIC. It may be of interest how the effects vary across different departments. Appendix A.7 presents results broken down by whether a class is in a male-

16. The within-course variation in class male proportion is plotted in Figure 1.2.

majority or female-majority department. The results indicate that both the negative effects of male peers on women and the positive effects of male peers on men are strongly concentrated among male-majority departments. Indeed, it appears that there is little if any effect of peer gender in classes within female-majority departments. Given that most STEM majors and most highly paid majors are male-dominated, this suggests that peer gender effects may be particularly relevant to policy-makers who care about increasing female representation in STEM or about the role of major choice in perpetuating the gender pay gap.

### *1.4.4 Relationship Between Short- and Long-Term Outcomes*

My analysis has thus far considered a range of outcomes that span a student's college career, ranging from the immediate outcome of grade in a first-semester class to later outcomes like choice of major and graduation. It is interesting to consider how the short- and long-term outcomes interact. In a dynamic model of major choice, as in Arcidiacono [2004], students enter college considering multiple majors while facing uncertainty about both their major-specific ability levels and their major-specific tastes, learn through experimentation, and eventually make a final major choice. First-semester classes may provide important information about both ability and tastes in a way that could be influenced by peer composition. The preceding results suggest that male peers affect grades. If a student, naive to the influence of peer gender on grades, receives a poor grade in a male-heavy class for a given major, they may negatively update their belief about their ability in that major. Peer gender may also influence beliefs about tastes in multiple ways. If students care directly about major-level gender composition, for instance if they dislike being in a gender minority, they may update their beliefs about the overall gender composition of a major based on the gender composition of the first class for that major. Even if a student does not have explicitly think about gender composition, peer gender may influence her beliefs about taste for a major if class gender composition influences classroom environment or how many friends she makes in a class. Thus, peers may influence major choice both indirectly through the impact of grades on beliefs about ability or directly through beliefs about taste.

Prior work finds that women are more likely to opt out of a major in response to a poor grade than men, which might suggest that women would be more susceptible to an indirect effect of male peers on major choice through grades (Rask and Tiefenthaler [2008], Ahn et al. [2019]). The within-gender results, presented in Table 1.6, could be seen as broadly concordant with these findings. Women appear to be more likely to receive a bad grade in a class with more male peers and are subsequently less likely to take future courses in the same department and may be modestly less likely to declare a major within that department. Men, on the other hand, experience little to no effect of male peers on grades, but are more likely to persist in a major when exposed to more male peers. For women it thus seems plausible that any peer effects on major choice flow through grades and beliefs about abilities, although it is not possible to rule out beliefs about tastes as a complementary or alternative mechanism. For men it seems as though effects on major choice must come from a channel other than grades and their impact on beliefs about ability. In order to further tease apart how peer gender affects major choice, I consider the effects of peer composition in finer subsets of the overall sample.

I first consider effect heterogeneity by ability level, as measured by high school GPA, with the results reported in Table 1.7. Looking at the effects on grades, it appears as though the negative effect of male students on grades is concentrated among the lower half of the ability distribution, for both women and men. For women, effects on major choice and graduation are also concentrated in the lower half of the ability distribution. Women who face stronger effects on grades also having stronger effects on major choice is consistent with the notion that the latter effect is a function of the former. For men, however, the positive effect of male peers on major choice is concentrated in the lower half of the ability distribution. Thus, the male students who are harmed more by male peers in terms of grades are also more likely

Table 1.7 – Estimated Effect of Class Male Proportion on Student Outcomes, By Gender and Pre-College Academic Ability

| | Grade (GPA value) | Future course in dept. | Switch major out of dept. | Declare major in dept. | Graduate within 6 years | Graduate in dept. |
|---|---|---|---|---|---|---|
| *Women with above average HS GPAs:* | | | | | | |
| Class male % | -0.182 | -0.122*** | -0.195 | 0.006 | -0.037 | -0.004 |
| | (0.132) | (0.040) | (0.259) | (0.018) | (0.046) | (0.021) |
| Outcome Mean | 3.298 | 0.668 | 0.394 | 0.053 | 0.759 | 0.060 |
| Outcome SD | 0.927 | 0.471 | 0.489 | 0.224 | 0.428 | 0.238 |
| Observations | 6897 | 9260 | 386 | 8750 | 9260 | 9260 |
| *Men with above average HS GPAs:* | | | | | | |
| Class male % | -0.110 | 0.075 | 0.276 | 0.018 | -0.054 | 0.042* |
| | (0.178) | (0.056) | (0.556) | (0.023) | (0.062) | (0.024) |
| Outcome Mean | 3.237 | 0.650 | 0.309 | 0.042 | 0.736 | 0.059 |
| Outcome SD | 0.961 | 0.477 | 0.463 | 0.200 | 0.441 | 0.235 |
| Observations | 4612 | 6529 | 314 | 6101 | 6529 | 6529 |
| *Women with below average HS GPAs:* | | | | | | |
| Class male % | -0.278 | 0.009 | -0.400 | -0.030 | -0.131** | 0.023 |
| | (0.193) | (0.052) | (0.401) | (0.026) | (0.065) | (0.027) |
| Outcome Mean | 2.748 | 0.600 | 0.349 | 0.052 | 0.523 | 0.051 |
| Outcome SD | 1.175 | 0.490 | 0.478 | 0.222 | 0.500 | 0.220 |
| Observations | 5283 | 7133 | 232 | 6701 | 7133 | 7133 |
| *Men with below average HS GPAs:* | | | | | | |
| Class male % | -0.188 | -0.035 | -0.677 | 0.030 | 0.054 | 0.051** |
| | (0.214) | (0.055) | (0.439) | (0.020) | (0.066) | (0.020) |
| Outcome Mean | 2.593 | 0.589 | 0.282 | 0.043 | 0.495 | 0.050 |
| Outcome SD | 1.218 | 0.492 | 0.451 | 0.203 | 0.500 | 0.218 |
| Observations | 6269 | 8780 | 301 | 8210 | 8780 | 8780 |

Notes: This table reports results of regressions of student-class outcomes on class-level male proportion. Each cell corresponds to a separate regression, with outcome given by the column header and the subset of students used for estimation given by the row header. Each regression includes course fixed effects and controls for student underrepresented minority status and instructor gender. All regressions are estimated on the observations within the identifying set of the main specification (as described in Section 1.2) for which the relevant outcome is observed. Standard errors are clustered at the class level. * indicates significance at a level of 0.1, ** at a level of 0.05, and *** at a level of 0.01.

to persist in majors where they have more male peers. This pattern would be surprising if grades are the mechanism driving the effect of male peers on male major choice.

I also directly consider how longer-term outcomes are mediated by the intermediate grade outcome. Specifically, I compare effects between the subset of students who received an A and those who received a worse grade. Given that nearly half of first-semester grades in the sample are As, it might be reasonable to think that UIC students would consider an A to be a "good" grade and anything else to be "bad."[17] As would be predicted if beliefs about ability was an important mechanism, women who receive As exhibit little if any effect of male peers on major choice, while women who receive worse grades in a class appear to avoid the major associated with that class. For men, on the other hand, the positive effect of male peers on major choice appears to be concentrated among men who received "bad" grades. This further suggests that male peers do not influence male major choice decisions indirectly through beliefs about ability, as men receiving inferior signals about their major-specific ability sort into the majors associated with their male-heavy classes at a higher rate.

### *1.4.5 Evidence on the Effect of Class Gender Composition on Friendship Formation*

For women, it is plausible that if peer gender influences major choice, it does so indirectly through grades, although it is not possible to rule out the existence of other mechanisms. For men it does not seem that grades are an intermediary connecting peer gender to major choice. It thereby remains to consider other ways in which peer gender may influence male major choice. Prior work has found that having at least one peer choose a given major may increase the likelihood of take-up of that major, with one potential explanation being the direct utility value of studying with a friend (De Giorgi et al. [2010]). If men are more likely to form friendships with other men, then having a more male-dominated first class in a major may increase the expected number of friends who are interested in the same major. This could then translate into an increased willingness to declare and persist in that major.

17. Approximately 41% of grades in the main estimation sample are As.

Table 1.8 – Estimated Effect of Class Male Proportion on Student Outcomes, By Gender and Realized Course Grade

| | Future course in dept. | Switch major out of dept. | Declare major in dept. | Graduate within 6 years | Graduate in dept. |
|---|---|---|---|---|---|
| *Women who received As:* | | | | | |
| Class male % | -0.043 | -0.045 | -0.019 | 0.010 | -0.015 |
| | (0.041) | (0.072) | (0.027) | (0.048) | (0.031) |
| Outcome Mean | 0.714 | 0.102 | 0.082 | 0.833 | 0.182 |
| Outcome SD | 0.452 | 0.302 | 0.274 | 0.373 | 0.386 |
| Observations | 7454 | 1073 | 6231 | 7454 | 7454 |
| | | | | | |
| *Men who received As:* | | | | | |
| Class male % | 0.047 | 0.096 | -0.000 | 0.001 | 0.023 |
| | (0.059) | (0.103) | (0.025) | (0.054) | (0.030) |
| Outcome Mean | 0.715 | 0.083 | 0.068 | 0.817 | 0.190 |
| Outcome SD | 0.451 | 0.276 | 0.253 | 0.387 | 0.392 |
| Observations | 5874 | 949 | 4731 | 5874 | 5874 |
| | | | | | |
| *Women who did not receive As:* | | | | | |
| Class male % | -0.060 | 0.071 | -0.059** | -0.017 | -0.052* |
| | (0.049) | (0.116) | (0.027) | (0.059) | (0.029) |
| Outcome Mean | 0.649 | 0.132 | 0.051 | 0.666 | 0.135 |
| Outcome SD | 0.477 | 0.339 | 0.220 | 0.472 | 0.342 |
| Observations | 9830 | 1180 | 8273 | 9830 | 9830 |
| | | | | | |
| *Men who did not receive As:* | | | | | |
| Class male % | -0.036 | 0.145 | 0.051** | 0.072 | 0.031 |
| | (0.054) | (0.144) | (0.024) | (0.062) | (0.028) |
| Outcome Mean | 0.634 | 0.145 | 0.049 | 0.594 | 0.111 |
| Outcome SD | 0.482 | 0.352 | 0.216 | 0.491 | 0.314 |
| Observations | 9913 | 973 | 8499 | 9913 | 9913 |

Notes: This table reports results of regressions of student-class outcomes on class-level male proportion. Each cell corresponds to a separate regression, with outcome given by the column header and the subset of students used for estimation given by the row header. Each regression includes course fixed effects and controls for student underrepresented minority status and instructor gender. All regressions are estimated on the observations within the identifying set of the main specification (as described in Section 1.2) for which the relevant outcome is observed. Standard errors are clustered at the class level. * indicates significance at a level of 0.1, ** at a level of 0.05, and *** at a level of 0.01.

While I cannot directly observe friendships in the data, I provide suggestive evidence using taking classes together as a proxy. Specifically, I consider all two-course sequences that are required for at least one of the twenty most popular majors at UIC and are commonly taken by first year students. A two-course sequence is defined as a set of two courses with one being a pre-requisite for the other. For instance, General Chemistry I and General Chemistry II is a two-course sequence that is required for the Biology and Chemistry majors, among others, and is recommended as first-year coursework for students interested in those majors. For students who take the first course in one of these sequences during their first semester and take the second course in a later semester, I define as outcomes the number of peers, of either gender, who were in the same class for both the first and second course and the number of peers who were in the same laboratory or discussion section for both the first and second course.[18]

I estimate the association between class-level male proportion and the number of same-subsequent-class peers using equations (1.2) and (1.3). Similarly, I look at the link between section-level male proportion and same-subsequent-section peers using the estimating equations

$$Y^F_{i,r,c,s} = \eta^F_0 + \eta^F_1 \times MP_s + \theta^F_c \times I_c + \varepsilon^F_{i,r,c,s} \tag{1.4}$$

$$Y^M_{i,r,c,s} = \eta^M_0 + \eta^M_1 \times MP_s + \theta^M_c \times I_c + \varepsilon^M_{i,r,c,s} \tag{1.5}$$

where sections are indexed by $s$, $MP_s$ denotes the proportion of male students in section $s$, $\eta^g_1$ measures how the number of repeated peers evolves with $MP_s$ for students of gender $g$, $\theta^g_c$ are class fixed effects, and $\varepsilon^g_{i,r,c,s}$ is the error term. Equation (1.4) is estimated only

18. This necessarily means that students who only take the first course in a sequence and never take the second are excluded. It also means that each peer pair is "double-counted" in the sense that any pair of students taking the same class for both courses in a sequence will be reflected in the outcome variable of both students.

on the sample of female students and equation (1.5) only on male students. I utilize eleven total sequences between the class and section analysis.[19]

It is almost certainly the case that some students will enroll in the same class twice in a row by chance. However, male students being systematically more likely to take future classes with the same peers when their initial class has more men may represent an increased likelihood of forming friends and coordinating future enrollment with them.

The results of this exercise are presented in Table 1.9. While the estimated effects of peer composition on the number of same-subsequent-class peers are imprecise, the results suggest that more male-dominated classes tend to result in more continued peers in the next class for both men *and* women. Similarly, having more men in the section of a first class increases the likelihood of having any same-subsequent-section peers for both men and women, even within a given class. The observed results may reflect homophily in friendship formation among men with some complementary explanation for women, some greater overall degree of friendship formation in more male-dominated environments, or other explanations. In any case, if men are more likely to form same-major friends in more male-dominated environments, this offers one plausible mechanism for greater male persistence in majors associated to male-dominated classes, in spite of male peers seemingly not providing academic benefits to male students.

Table 1.9 reports statistical significance based on the null hypotheses of coefficients equalling zero. However, it is not clear that the absence of intentional sorting to classes with prior classmates implies a zero coefficient. For instance, the point estimates in Table

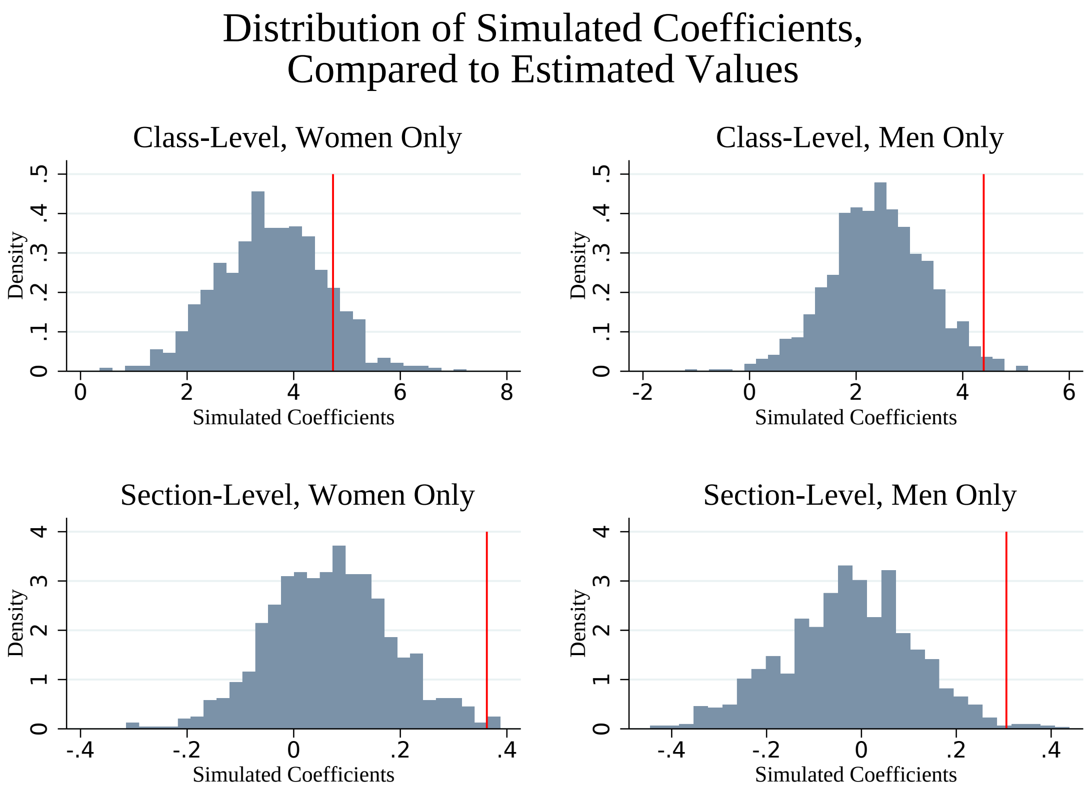
19. For class-level results, I use all sequences for which there are at least two classes for the first course in the sequence in both Fall 2015 and Fall 2016: General Chemistry I and General Chemistry II; Introduction to Psychology and Introduction to Research in Psychology; Introduction to UIC and Professional Development and Business Professional Development II; Calculus I and Calculus II; Calculus II and Calculus III; General Physics I and General Physics II; Principles of Microeconomics and Microeconomics: Theory and Applications; Principles of Macroeconomics and Macroeconomics in the World Economy: Theory and Applications; and Biology of Cells and Organisms and Biology of Populations and Communities. For the section-level results, I use the same set of sequences, excluding Introduction to UIC and Professional Development and Business Professional Development II as they have no associated sections and including Program Design I and Program Design II and Introduction to Criminology, Law, and Justice and Foundations of Law and Justice as, although there is not sufficient variation in these sequences for class-level analysis, there is enough for section-level analysis.

1.6 would suggest that male peers encourage male persistence in major to a greater extent than they discourage female persistence. Taking these coefficients literally would imply that more male-heavy classes would have greater net persistence into related majors, which could mechanically generate the patterns seen in Table 1.9 (if students are more likely to take the second course in a sequence if they are also persisting in a related major). To address this, I benchmark the estimates I get from the data against simulated coefficients.

For the simulations, I use the same sample of students used for estimation in Table 1.9. I leave fixed their first-course enrollments and simulate fully random movement into classes for the second course in each sequence, preserving the observed sizes of the second classes. That is, if a student is observed taking the Fall 2015, 9 AM General Chemistry I and the Spring 2016, 9 AM General Chemistry II, in a simulation, I "leave" them in the Fall 2015, 9 AM General Chemistry I but randomly "place them" in a class for General Chemistry II, such that each simulated General Chemistry II class has the same number of students as is observed in the data. I then estimate the same regression models reported in Table 1.9 using the simulated data. This exercise fully conditions on which students choose to take both courses in a sequence. The comparison between my estimates from the data and the estimates from the simulations is thus based only on selection of class, conditional on choosing to complete the sequence.

The results are presented in Figure 1.3, with the distributions of simulated coefficients (from 1,000 simulations) appearing in blue and the estimates from the original data appearing in red. It appears that, for both women and men, random enrollments for second courses would not yield coefficients of zero on class-level male proportion. However, it is clear that the estimate magnitudes I observe in the data are unlikely to occur based on purely random enrollment. Focusing on the results for men, the class-level coefficient lies at the 98th percentile of simulated coefficients and the section-level coefficient lies at the 96th percentile. It is thus reasonable to conclude that men (and women) are more likely to take classes and

Figure 1.3




sections with former classmates when their initial classes are more male-heavy, even fully accounting for the extensive margin of course selection. This may reflect a higher degree of friendship formation in more male-dominated classes, which would provide a plausible explanation for the positive effect of male peers on male persistence in majors.

## 1.5 Conclusion

Gender differences in college major take-up and completion are of interest to policy-makers, largely due to the substantial labor market implications. I investigate the role of peer composition in driving gender differences in college student achievement and major choice using a large administrative dataset that provides information on two cohorts of students at the University of Illinois Chicago over the courses of their college careers. I focus on incoming first-semester students, who enroll for courses without knowledge of peer enrollments, in order to minimize the risk of intentional sorting into peer groups. A rich set of fixed effects accounts for gender differences in performance at a course level, utilizing variation in lecture times within a course, and accounts for gender-neutral sorting or shocks associated with specific class times. I find that when a class has more male students, women receive worse grades, are less likely to declare majors associated with that class, and are less likely to graduate, all relative to men attending the same lectures. My identification strategy is supported by balance tests showing that I am comparing across classes that have very similar students in terms of observable characteristics. The findings are further validated by two robustness checks: one only uses cross-cohort variation to avoid issues of sorting to different classes within a semester and the other allows for arbitrary gendered sorting to specific classes and instead focuses on variation across TA-led sections within a class.

The aforementioned results concern how men and women diverge based on peer composition. I further estimate student responses by gender in order to determine whether men or women are responding more to peer composition. I find that women receive worse grades and are less likely to eventually graduate when they have more male peers, while male achievement is not substantially affected by peer gender. On the other hand, when a first-semester class has more men, men are more likely to declare a related major, while the effect on women's major choice is negative but more modest in magnitude than the effect on men.

In a simple model of college major choice, as in Arcidiacono [2004], we can think of students as entering college with uncertainty about their major-specific abilities and tastes, learning about themselves during classes taken early in college, and making major decisions based on what they learned. The fact that I observe women receiving worse grades in male-dominated classes and subsequently being less likely to pursue related majors implies that male peers may affect women's major choice via beliefs about ability, although it is impossible to rule out taste as a mechanism. However, the lack of a positive effect of male peers on men's grades makes it unlikely that the positive effect of male peers on men's major choice is explained by beliefs about ability. One alternative explanation could be that men form more friends in more male-dominated classes, which encourages them to continue taking courses in the same field in order to continue studying with those friends. I find that men in more male-dominated classes take future classes with more repeated peers than men in less male-dominated classes, suggesting that friendship formation is a plausible mechanism for the effect of male peers on male major choice. Future work would do well to further consider mechanisms of peer gender on student outcomes, potentially leveraging different kinds of data to directly study friendship networks.

My results suggest that increasing the proportion of students of a given gender in a class yields positive results for the other students of the same gender. One policy implication could be that any policy that increases the representation of gender-minority students in an academic field will have a greater than anticipated impact. The presence of more gender-minority students should encourage other students of that gender to perform well and persist in the field, creating a total effect larger than the direct impact of the policy. While my estimates are necessarily based only on the relatively modest amount of variation in gender composition across classes within a course, future work on the impact of larger differences in gender composition would be useful. This would be informative about how much peer effects may matter for the efficacy of policies that change class gender ratios by large amounts.

Table 1.9 – Estimated Effect of Class and Section Male Proportion on the Number of Peers Taking Future Classes Together

| | Number of peers in the same subsequent class | Number of peers in the same subsequent section |
|---|---|---|
| *Only women:* | | |
| Class male % | 4.738 | - |
| | (3.679) | - |
| Section male % | - | 0.362* |
| | - | (0.218) |
| Outcome Mean | 9.278 | 0.348 |
| Outcome SD | 10.665 | 0.754 |
| Observations | 1027 | 880 |
| | | |
| *Only men:* | | |
| Class male % | 4.394 | - |
| | (2.654) | - |
| Section male % | - | 0.306 |
| | - | (0.207) |
| Outcome Mean | 8.459 | 0.550 |
| Outcome SD | 9.442 | 0.970 |
| Observations | 1319 | 1179 |

Notes: The first column of this table reports results of regressions of number of peers from a student's first class in a two-course sequence who take the same second class on class male proportion, conditional on course fixed effects. The second column reports results of regressions of number of peers from a student's first section in a sequence who take the same second section on section male proportion, conditional on class fixed effects. Each cell corresponds to a separate regression. All regressions are estimated on the set of students who take the first course in one of the listed two-courses sequences during their first semester and take the second course in a subsequent semester. Top panel regressions are only estimated on female students and bottom panel regressions only on male students. Standard errors are clustered at the first-class level. * indicates significance at a level of 0.1, ** at a level of 0.05, and *** at a level of 0.01.

# CHAPTER 2

# COLLUSION AND FINANCIAL AID DETERMINATION IN HIGHER EDUCATION

## 2.1 Introduction

In the late 1980's the US Justice Department launched an investigation into the allegedly collusive behavior of a group of elite US universities known as the "Overlap Group."[1] For several years prior to the beginning of the investigation, the Overlap Group had been meeting to discuss matters related to tuition and financial aid, including sharing their intended financial aid offers for individual students. The Justice Department's suit alleged a form of price fixing, accusing the Group of "illegally conspiring to restrain price competition on financial aid"(MIT News [1992]). The government claimed that this behavior resulted in elevation of both the price paid by high-income, highly valued students and the average price paid across all students (Carlton et al. [1995]).

The Overlap Group and the case against them, which has received scant attention in the economics literature, provides an opportunity to explore multiple questions. There is an immediate question of whether collusion among higher educational institutions is bad and should be prevented by policy makers. This is particularly salient given that there is currently another, ongoing lawsuit against several elite universities (including some members of the original Overlap Group) that has similar allegations to those brought in the case against the Overlap Group (Saul and Hartocollis [2022]). This question can itself be broken down into whether collusion suppresses financial aid generosity and what happens to the money that would otherwise be spent on aid, if it does. The first question of whether or not

1. The full list of schools in the group is as follows: Amherst College, Barnard College, Bowdoin College, Brown University, Bryn Mawr College, Colby College, Columbia University, Cornell University, Dartmouth College, Harvard University, Massachusetts Institute of Technology, Middlebury College, Mount Holyoke College, Princeton University, Smith College, Trinity College, Tufts University, University of Pennsylvania, Vassar College, Wellesley College, Wesleyan University, Williams College, and Yale University.

collusion affects the average price paid by students is not immediately obvious.[2] While most of the Overlap Group settled with the Justice Department, MIT fought the case, arguing, in part, that the Group's operations allowed for the redistribution of financial aid away from higher-income students towards lower-income students, which might imply no effect on the average price. If collusion does allow schools to spend less on financial aid, it is then important to determine where this money goes. Certain alternative streams of spending, such as administrative salaries and sports programs, may be perceived as having relatively little social value, based on critical reporting in popular media (Krupnick and Marcus [2015], Hobson and Rich [2015]). On the other hand, spending on research and improved educational offerings are more likely to be universally seen as creating social benefit, as well as potential benefit to students.

Questions about the impact of collusion also tie to deeper questions about the incentives of non-profit universities. Models of university behavior typically assume that universities compete on "prestige," which is assumed to rise in the quality of students admitted and the quality of the educational experience provided, while also maximizing "profit" (Epple et al. [2003], Epple et al. [2006], Fu [2014], and Blair and Smetters [2021]). On the other hand, the MIT argument in the suit would suggest that universities also directly value allowing for the enrollment of lower-income students (Matlock [1994]). The behavior of the Overlap Group provides a context in which to consider what objectives universities allocate money towards in different circumstances.

In order to consider these questions, I develop a simple model in which schools choose prices to charge to different types of students, with each type offering different value to the schools. Schools care simultaneously about prestige generated by student quality, prestige generated by other sources (such as research or faculty quality), and a nebulous "profit" term. The model suggests that collusion will increase the cost of attendance for nearly all

2. Other papers discussing the impact of specific factors on university pricing include Hoxby [1997], Cellini and Goldin [2014], Dinerstein et al. [2015], and Turner [2017].

types of students admitted, with the largest increases for the (admitted) types who offer the least value, and potential minor effects on enrollment. The model also sustains two types of competitive equilibria: equilibria in which any additional income would be spent on "profit" and equilibria in which at least some additional income would be spent on non-student prestige.

In an empirical analysis, I use both conventional difference-in-differences and synthetic control methods to compare the Overlap Group schools with a set of comparable universities before and after the cessation of the Group's annual meetings, using data from the Integrated Postsecondary Education Data System. Assuming that any collusive behavior among the Overlap Group stopped after the cancellation of the annual meetings (as was agreed to by a majority of the participating schools) and that the control group were not colluding either before or after the beginning of the Justice Department investigation, such a comparison will allow for the identification of effects of collusion on university behavior.

Both methods provide evidence that the cessation of Overlap meetings led to a rise in the proportion of university expenditures that went toward financial aid, suggesting that collusive behavior may have been reducing the total share of overall expenditures on financial aid by as much as 1.4%. In 1990, the year that the Overlap Group held their final annual meeting, the average share of expenditures on financial aid was 10%, which would suggest that collusion was reducing the amount of aid by over a tenth compared to the typical amount of the time. The evidence on where this money was redirected from in the absence of collusion is less clear, but is suggestive that the savings on financial aid from colluding was spent on research and instruction and not high-level administrative salaries (which is treated as one possible interpretation of "profit"). These results are broadly consistent with the price-setting model in the case of a competitive equilibrium in which additional income would be spent in part on non-student prestige. This result suggests that, although the Justice Department appears to have been correct about the effect of collusion on average financial aid offerings, there are unclear welfare implications to non-profit university collusion that depend on the weights a social planner would place on research and educational quality versus student surplus.

The remainder of the paper proceeds as follows: Section 2.2 provides additional background about the Overlap Group and the Justice Department investigation. Section 2.3 describes the theoretical model of financial aid determination. Section 2.4 describes the data and empirical strategy and Section 2.5 describes the results. Section 2.6 concludes.

## 2.2 Background

The Overlap Group commenced operations in the 1950s when members of the Ivy League began meeting to discuss the idea of not engaging in bidding wars over promising student athletes who had been admitted to multiple member institutions - students who “overlapped" across members. According to the founding schools, the purpose of the group was to focus the then limited and relatively new practice of financial aid provision on needier students in order to enable them to access the otherwise unaffordable elite institutions. By the 1970’s the Group had grown to its full contingent of 23 member institutions, including the entirety of the Ivy League, MIT, and several elite liberal arts colleges that are also concentrated in the Northeastern United States (Carlton et al. [1995]).[3]

By the 1980’s, the Overlap Group was performing multiple functions for its consituent schools. They shared information on planned tuition rates, faculty salaries, and the financial situations of common applicants (Dodge [1989]). However, the activity of the most interest both to the Justice Department and this paper was the explicit comparison of the amount of aid being offered to specific, individual students.

The full financial expense of college attendance can be broken into three categories: grants, “self-help,” and “family contributions.” Grants are sums of gift aid that need not be repaid by the student and are typically provided either by the government or universities.

3. See Footnote 1 for a full list of member institutions.

"Self-help" includes both loans (which are repaid) and student employment agreements, the wages of which are assumed to offset costs of attendance. "Family contributions" refer to the remainder of the total costs of attendance that are not covered by the preceding categories, which are expected to be paid directly by the student and their family (University of Washington [2022]).

Overlap Group schools would directly compare the family contributions that each institution was planning to charge to commonly admitted students (in other words the sticker price of the university less outside grants and self-help less the planned institutional grants and self-help, or total institutional financial aid). Each institution would create a list with the amount of family contribution planned for each admitted student. At a yearly spring meeting, typically held at Wellesley College, the schools would compare the intended family contributions for the set of students who had been admitted by more than one member institution, a number that frequently exceeded 10,000 students. For students that had widely discrepant numbers across member institutions, officials from the institutions in question would discuss the individual student's circumstances in an attempt to narrow the gap between their intended family contributions. After this meeting, each institution would privately finalize the financial aid package they were going to offer each student, comprised of grants and self-help, implicitly setting the family contributions for each student (Dodge [1989]). While all Overlap schools participated in the process of comparing family contributions, some of the member institutions would further discuss the composition of aid across grants and self-help they intended to offer to students (Carlton et al. [1995]).

In 1989, the Justice Department launched an investigation into the Overlap Group, alleging that they were engaged in an illegal form of price-fixing through their joint financial aid considerations, shortly after the Overlap Group met for their last full meeting in the spring of 1989 (MIT News [1992]). For the meeting in spring of 1990, Yale University and Barnard College did not attend, although the other members did, and the meeting in the

Spring of 1991 was cancelled altogether (Chira [1991]). Meanwhile, in the spring of 1991, the Justice Department followed up their investigation with a suit against the Overlap Group. While MIT fought the suit, the eight Ivy League universities settled, agreeing to stop sharing financial information and cease holding the spring meetings, effectively suspending the operations of the Group (DePalma [1991]).

The Justice Department settled with MIT in December of 1993. The settlement allowed MIT and other non-profit colleges to engage in some of the activities the Overlap Group had engaged in prior to the investigation and suit, including corroborating student financial data (prior to decisions about aid) and comparing retrospective aid data. However, the settlement explicitly forbid discussion of intended family contributions to be made by individual students as well as grants and self-help to be offered to individual students (MIT News [1994]).[4] This settlement was later codified into law, Section 568 of the Improving America's Schools Act, which maintained the illegality of discussions regarding financial aid packages intended for individual students (568 Presidents Group [2007]). While it is unclear if the Overlap Group resumed meeting in the more limited capacity allowable by law following the settlement, the 568 Group, which can be seen as a successor organization, began meeting in 1998. This new group included many members of the original Overlap Group as well as some comparable schools that were not in the original.

Between voluntary cancellations of meetings, the settlement agreement, and the new legislation, the Overlap Group most recently (legally) met as a full body to discuss individual financial aid packages in 1989 and last met with partial membership to do the same in 1990. As such 1991 will be treated as the first year of non-collusion in the empirical section of this paper.

4. The settlement also forbid the discussion of prospective tuition rates and faculty salary levels.

## 2.3 Theory

I now present a simple model in which schools determine prices for students based on a desire to maximize the quality of the student body, the non-student prestige of the university, and a "profit" term.

The utility function of school $i$, for all $i \in \{1, ..., m\}$ is:

$$U_i = r(e_i) + \beta f_i + \sum_{\Theta} q_{i,j}\theta_j \tag{2.1}$$

$$s.t.\ e_i + f_i = d + \sum_{\Theta} q_{i,j}(p_{i,j} - c) \tag{2.2}$$

Non-student prestige (henceforth research prestige) is a function, $r(\cdot)$, of expenditures on research, $e_i$. $r(\cdot)$ is a strictly concave, twice differentiable, and increasing function with $r'(0) = \infty$ and $\lim_{e_i \to \infty} r'(e_i) = 0$. $f_i$ represents "profit." The profit term can be taken to represent any socially undesirable spending of money. In the empirical section below, this will be operationalized as the salaries of high-level administrative officials, although in practice the quality of such individuals likely affects the educational and research missions of the universities. Prestige from student quality is the amount of students of each type $j$ who attend school $i$, $q_{i,j}$, multiplied by the value of type $j$ in the eyes of schools, $\theta_j$, summed across all types.[5] When indifferent between multiple actions, schools will choose the action that maximizes the overall enrollment of students.

The schools are subject to a budget constraint stating that the sum of endowment income, $d$, and net revenue from enrolling students must equal the sum of research expenditure and profit.[6] The net revenue from enrolling students is equal to the price charged to students

5. $\theta_j$ can contain any priorities regarding student body composition that schools may have, including pre-college academic achievement, diversity concerns, athletic ability, or any other characteristic of interest to universities.

6. This assumption is made for simplicity, but variation in endowments is likely an important aspect of heterogeneity across universities. Epple et al. [2006], for instance, uses this variation to identify a structural model of university admissions and pricing behavior.

of type $j$, $p_{i,j}$, less the constant marginal cost of student enrollment, $c$, multiplied by the number of students of type $j$ enrolled by school $i$, summed across all types. $p_{i,j}$ can be thought of as the family contribution that schools will leave students with, after financial aid. Schools will be allowed to set this to any non-negative value, so we can think of the sticker price of the school as simply being the highest finite price that is set for any type, and all lower prices as reflecting the offering of the appropriate financial aid.

Potential students each have a type $\theta_j \in \Theta$.[7] There are measure 1 students of each type. All students have an identical maximum willingness-to-pay for school attendance, $v$, and it is assumed that $v > c$. Students will attend the school that offers them the lowest price that is less than or equal to $v$. If multiple schools tie for lowest price, students randomize between the tying schools. If no school offers a price less than or equal to $v$, students choose their outside option.

The schools engage in a game of simultaneous price setting, where they must charge a uniform, non-negative price to all students of a given type but are allowed to differentiate prices across types. Equilibria exist at sets of prices, $p^*_{i,j}$, research expenditures, $e^*_i$, and profits, $f^*_i$, where no school has incentive to deviate. I focus on the set of symmetrical equilibria.[8] Call the common equilibrium research expenditure $e^*$. There can exist two types of equilibria: those in which $r'(e^*) = \beta$ and those in which $r'(e^*) > \beta$. There can be no equilibria with $r'(e^*) < \beta$, as any school would deviate to decrease research spending and allocate it to profit in such a case.

7. The nature of $\Theta$ differs across the two cases of equilibrium considered below.

8. It is not immediately obvious whether or not asymmetrical equilibria might exist in either of the cases considered below.

### 2.3.1 Profit-Seeking Equilibria

I consider equilibria with $r'(e^*) = \beta$ first. Call the value of research expenditure that satisfies the above condition $\overline{e}$. For this type of equilibrium, potential students each have a type

$$\theta \in \Theta = \{\theta_1, ..., \theta_n\}$$

where $\theta_1 < \theta_2 < ... < \theta_n$. Thus, there are $n$ ordered types of students. For the the discussion of this form of equilibrium, I maintain the following assumptions:

- For type sets $\tilde{J} = \{j | \beta(c - v) \leq \theta_j \leq \beta c\}$ and $\overline{J} = \{j | \theta_j > \beta c\}$:

$$-\sum_{\tilde{J}} \frac{\theta_j}{\beta} - \sum_{\overline{J}} c \geq m(\overline{e} - d)$$

  This indicates that the group of students who end up being admitted to the university are not "too good" such that schools end up spending "too much" trying to recruit them. If this is violated, it pushes the schools into the other type of equilibrium, discussed in the next subsection.

- At least one of the following statements is true:

$$\theta_{\overline{k}} \leq \beta(d - \overline{e}) - \frac{1}{m} \sum_{j \neq \overline{k}} \theta_j$$
$$\beta c \leq \beta(d - \overline{e}) - \frac{1}{m} \sum_{j \neq \overline{k}} \theta_j$$

  This assumption restricts how expensive the most valuable type of student will be. Note that this assumption is not necessarily required for an equilibrium to exist, it merely greatly simplifies the nature of the pricing schedule. If this assumption was violated, it would likely be the case that an equilibrium of the profit-seeking type could be found, with an alternative pricing schedule.

With these assumptions maintained, the set of student types can be divided into four subsets. Students with $\theta_j \in (-\infty, (c-v)\beta)$ have negative value to the school and are sufficiently undesirable that schools would only be willing to admit them at a price above their willingness-to-pay. We can think of schools as setting $p_{i,j} = \infty$ to represent not admitting students who fall below a certain attractiveness. All other students are admitted, meaning charged a price that is at or below willingness to pay. Specifically, the pricing schedule will be:

$$p_k^* = \begin{cases} c - \frac{\theta_k}{\beta} \; \forall \; k \in \{k | \theta_k \in [(c-v)\beta, 0) \cup (\beta, \beta c]\} \\ c \; \forall \; k \in \theta_j \in [0, \beta] \\ 0 \; \forall \; k \in \{k | \theta_k > \beta c\} \end{cases}$$

Because schools are operating at such a point that $e^* = \bar{e}$, all income beyond $e^*$ is spent on research. This allows all types to be priced such that schools are precisely indifferent between enrolling the student or not. For negative-valued types, this means that price is high enough for schools to offset the disutility of enrolling the students with spending on profit. For positive-valued types, schools are willing to enroll the students at a price below marginal cost, “spending" money on students such that the amount of profit foregone is compensated for by the value the student brings in prestige. For sufficiently highly valued students, the non-negativity constraint on prices binds, and these students are admitted for free (“full-ride”). No school is willing to deviate in either direction for any student.[9] Because all schools match each other’s prices, each school shares $\frac{1}{m}$ of each type of admitted student.

If the $m$ schools in a market featuring this type of equilibrium decided to collude, their best course of action, if they wanted to maintain symmetrical outcomes, would be to simply charge each student at exactly their willingness-to-pay. They would maintain the exact same student body and total enrollment (the non-admitted types would remain undesirable by

9. A derivation of this equilibrium is presented in Appendix B.1.

the nature of the non-admittance criteria) while generating greater revenue. All additional revenue would be spent on profit. Nearly all students would experience an increase in effective price of admission, with the price increase being the greatest for the highest value students.[10,11]

While this model is highly stylized in general, the most obvious deviation between it and the actual situation of the Overlap Group is the absence of competitors who are not part of the collusive agreement. In reality, there were comparably ranked schools that were not part of the Overlap Group, and, presumably, not party to any collusive agreements. While it is certainly not the case that these schools did not compete directly with the Overlap Group, there were some reasons why the Overlap Group might have had some unique market power. For students who care about the "name brand" of their school, for instance, the Overlap Group contained the entirety of the Ivy League and the majority of the "Little Ivies," and so might be considered substantially superior in this dimension. The Overlap schools are also all located in the Northeast of the US while the large majority of the comparison schools are not, potentially allowing for a locational preference.

### *2.3.2 Research-Seeking Equilibria*

There can also be equilibria for which $r'(e^*) < \beta$.[12] In these equilibrium, while some amount of money is spent on research, no money is spent on "profit," as any small quantity of additional money would be preferred to be spent on research than on profit. As these equilibria are harder to characterize, I consider a restricted type space in which there are

10. It was undoubtedly not the case that all students are charged an identical price, even during the operation of the Overlap Group. Perhaps the most obvious "culprit" for this, which is not in the model, is heterogeneous willingness-to-pay. Allowing for heterogeneous willingness-to-pay might allow for such cases as high-value types with willingness-to-pay below $c$, who schools might be willing to subsidize the attendance of, given a high enough budget. This would effectively capture the MIT argument in their defense against the Justice Department suit.

11. Price would remain the same for a type with a value exactly equal to $(c - v)\beta$.

12. The case in which there is no profit motive, so $\beta = 0$, is a special case of this type of equilibrium where $\overline{e} = \infty$.

only two types that the university will choose to admit.[13] Refer to the two relevant types as $\theta_L < 0 < \theta_H$. Along with this definition of the type space, the following assumptions are maintained for the characterization of this type of equilibrium:

- At least one of the following is true:

$$|\theta_H| \geq |\theta_L|$$

  or

$$r'(\bar{e}) > \frac{-1}{m}\frac{\theta_H + \theta_L}{\bar{e} - d}$$

  This assumption prevents the overall student body from being so "unattractive" that schools end up not spending enough money on recruiting them and moving back into the preceding type of equilibrium.

- Assume $\bar{e} > d$ and that:

$$r'(e^*) \geq \max\{\frac{\theta_H}{d}, \frac{-\theta_L}{v - c}, \frac{\theta_H}{c}, \frac{\frac{1}{m}\theta_H - \frac{m-1}{m}\theta_L}{\bar{e} - e^*}\}$$

  These assumptions are not necessarily required for an equilibrium to exist. They create bounds on the prices of each type in order to manage the types of deviations away from equilibrium it is possible for schools to make, which greatly simplifies the identification of possible equilibrium prices. It is likely that equilibria could be found if these assumptions were violated, given greater care in the determination of prices.[14]

13. For any set of conditions, negative types that are sufficiently undesirable that they will never be enrolled can always be found. There can be an arbitrarily large number of these types, and the price for them can be set to $\infty$ to indicate non-admission, as before.

14. It is difficult to define this conditions based solely on the parameters. As discussed in Appendix B.1, the exact value of $e^*$ varies directly with the exact shape of $r(\cdot)$ as well as $m$, $\theta_H$, $\theta_L$, and $d$, while remaining independent of all other parameters. Thus, this condition could be interpreted as saying that the pricing schedule proposed below is guaranteed to work for a certain set of values for the vector $(v, c, \theta_H)$ given a specific $r(\cdot)$ and specific value of $(m, \theta_L, d)$.

With these assumptions in place, it is possible to show that there is at least one set of prices that will create an equilibrium.[15] For instance, the prices:

$$p_H^* = \frac{-\theta_H}{r'(e^*)} + c$$
$$p_L^* = \frac{-\theta_L}{r'(e^*)} + c$$

will both be possible and disincentive any school from deviating away from equilibrium. As before, the negatively-valued type pays above marginal cost while the positively-valued type pays below marginal cost. At the suggested prices, the gain (loss) from enrolling the marginal student of the negatively-valued (positively-valued) type is such that the additional spending on research is equal to the utility cost of enrolling that student. As all schools set identically prices, schools again each enroll $\frac{1}{m}$ of the students of each admitted type.

If schools in this market were to collude and wanted to maintain symmetry, they again would set price to $v$ for both types of students, if they continued to admit both types, leading to an increase in the effective price for both types, with the high-type facing the larger change. As $e^* < \bar{e}$, at least some proportion of the additional revenue generated by collusion will go towards non-student prestige. Overall enrollment will weakly decrease. If there were any types that were not admitted in the competitive equilibrium, those types will still not be admitted, as the condition that precluded their enrollment previously will still hold when then marginal value of money decreases due to the increase in revenue. On the other hand, if the increase in revenue generated from charging the higher type $v$ is large enough, it is possible that the decline in the marginal value of money could create a situation in which it is no longer desirable to admit the low-type students.

Thus, the model predicts that, compared to any form of competitive equilibrium, collusion will lead to an increase in prices for all types and either no or negative effects on enrollment.

15. The full derivation of the equilibrium is in Appendix B.1.

However, what the additional revenue from collusion would be spent on depends on the nature of the competitive equilibrium, which is itself partly a function of how low the satiation point for research expenditures is.

## 2.4 Empirical Strategy and Data

In order to empirically assess how the operations of the Overlap Group impacted university behavior, I compare the behavior of Overlap and other elite schools before and after the cessation of Overlap meetings, using both convention difference-in-differences (DID) and synthetic controls methods.

For the duration of the paper, the period of 1984-1990 will be considered the pre-period while the period of 1991-1997 will be considered the post-period, as 1990 was the year of the final annual Overlap Group meeting. This may be slightly imprecise for a couple of reasons. Two schools, Yale University and Barnard College did not attend the 1990 spring Overlap Group meeting and so might be considered "untreated" in 1990 itself. Moreover, the schools that did attend may have behaved differently from years prior due to the scrutiny from the ongoing Justice Department investigation. However, as an Overlap meeting did take place with a large majority of member schools present, 1990 will be considered as an Overlap year. On the other hand, there may have been meetings by the Overlap group in the years following the 1993 settlement. However, as any meetings that did take place would not (legally) have involved discussions of individual financial aid offers, these years will be considered non-Overlap years. The period from 1998 on is excluded due to the existence of the 568 group, which, although restricted in the same way, included members of both the original Overlap Group and schools that will be considered control schools later in the paper. Due to any relevant changes in behavior this may induce, this period is not considered.

The group of control schools in the DID analyses and the set of potential donor schools in the synthetic controls analyses is comprised of a set of institutions ranked comparably to the

Overlap schools. Specifically, the control group is composed of all research universities that were ranked within the top 20 of the widely referenced US News and World Report rankings from 1989 (all research universities in the Overlap Group are ranked within the top 20, with the lowest being the University of Pennsylvania at 15) as well as all liberal arts colleges that were ranked within the top 25 of the US News and World Report rankings (all liberal arts colleges in the Overlap Group are ranked within the top 25, with the lowest being Vassar College at 22) (Los Angeles Times [1989], US News & World Report [2022]).[16]

### *2.4.1 Data*

I use data on university financial figures and other characteristics from the Integrated Post-secondary Education Data System (IPEDS). IPEDS offers yearly financial and other data for all of the schools in the Overlap Group, as well as several comparable schools, from 1984 through 2020.[17] This covers six years of Overlap Group operation and all seven years following the cessation of Overlap meetings and before the advent of the 568 group.

Over this entire period, IPEDS reports all scholarship expenditures in one of two streams - "restricted" and "unrestricted." Restricted scholarships will typically have more specific criteria or earmarking for specific populations while unrestricted scholarships will have more opaque criteria and discretion in allocation by the university. Per the Harvard University Committee on General Scholarships, for instance, "Restricted Scholarships are individual funds, usually endowment funds, created by alumni and other donors to support specific populations of students within the University" (Harvard University [2022]).

16. The full set of control schools is: Stanford University, University of Chicago, California Institute of Technology, Rice University, Johns Hopkins University, Northwestern University, Washington University in St. Louis, Duke University, University of Notre Dame, Georgetown University, University of Virginia, Swarthmore College, Pomona College, Claremont McKenna College, Carleton College, Davidson College, Grinell College, Hamilton College, Haverford College, Colgate College, University of Richmond, Bates College, and Washington and Lee College.

17. The acquisition of endowment stock and income data from The National Association of College and University Business Officers is currently in progress.

Table 2.1 – Summary Statistics For Full Sample of Schools in 1990

| Variable | Overlap Schools | | Control Schools | |
|---|---|---|---|---|
| | Mean | Std. Deviation | Mean | Std. Deviation |
| Share of Expend on Scholarships | 0.109 | 0.037 | 0.090 | 0.047 |
| Total Enrollment | 3573.39 | 2246.56 | 5578.73 | 6188.05 |
| Tuition (Sticker Price) | 15222.50 | 506.91 | 12548.82 | 2932.54 |
| Undergraduate Minority % | 0.108 | 0.038 | 0.100 | 0.050 |
| Research Expend (in Mills of $) | 52.5 | 71.6 | 62.6 | 85.8 |
| Instruction Expend (in Mills of $) | 76.4 | 89.4 | 109 | 118 |

Table 2.1 presents summary statistics for both the Overlap and the control schools from the year 1990. The two groups appear generally similar across most dimension, although it is notable that the Overlap schools are smaller and more expensive (in terms of sticker price). It is also interesting that the Overlap schools have a strikingly low variance in tuition, potentially suggesting that shared information regarding planned tuition increases induced similar tuition setting among the Group schools.

### *2.4.2 Difference-in-Differences*

The primary DID estimating equation used through the paper is:

$$Y_{it} = \beta \text{OGC}_{it} + \alpha_t + \delta_i + \varepsilon_{it} \tag{2.3}$$

where $Y_{it}$ is the outcome of interest, for instance, expenditures on scholarships as a share of total expenditures, $OGC_{it}$ is an indicator for members of the Overlap Group in the years after 1990, $\alpha_t$ is a year fixed effect, $\delta_i$ is a school fixed effect, and $\varepsilon_{it}$ is an error term. This estimating equation will produce unbiased estimates of the effect of the Overlap Group disbanding if, along with parallel trends, it is also the case that there was no anticipation of the Justice Department investigation before it began, there was no collusion among the control schools before or after the beginning of the investigation, and if there was no effect of Overlap collusion on financial aid offers by other schools. This last point may be particularly questionable if financial aid offers by some schools have effects in the broader market for prospective college students. We might expect, however, that non-colluding schools would generally “move with" colluding schools, increasing financial aid offers after the cessation of collusion in response to more competitive offers from former colluders. If this was the case, it would bias the results against an increase in financial aid after the beginning of the investigation, suggesting the results may be conservative.

It is also worth emphasizing that $\beta$ is the effect of the cessation of potential collusive activity. Thus, it represents the negation of the effect of collusion itself.

### *2.4.3 Synthetic Controls*

In a setting such as this, where the treatment of interest (the cessation of potential collusion) is applied to a relatively small number of units - 23 universities - conventional difference-in-differences estimation may be a poor choice. DID imposes a relatively unstructured process on the weighting of the control units that are compared to each treated unit which may result in a failure of the assumption of parallel trends. For instance, there appears to be a visible failure of parallel trends for the proportion of university expenditures on restricted scholarships, even in the pre-period, as seen in Figure 2.3.

The synthetic control method initially developed in Abadie and Gardeazabal [2003] and Abadie et al. [2010] may help to rectify this problem by creating individual synthetic controls for each treated unit from averages of control units weighted so as to replicate predictors of the outcome variable in the pre-period. In principal, this method enables for balancing on unobservables as well as observables, to the extent that unobservables are part of the data-generating processe for the predictor variables (which in this case includes lags of the outcome of interest). This method has been applied in a wide variety of contexts, including the study of the effects of right-to-carry laws on crime (Donohue et al. [2019]), legalized prostitution (Cunningham and Shah [2018]), and immigration policy (Bohn et al. [2014]).

Specifically, the synthetic control method provides a framework for considering the effect that treatment, in this case the cessation of collusion, will have on each school in the Overlap Group, $i \in \{1, ..., I\}$, in each period it is treated, $t \in \{1991, ..., 1997\}$:

$$\hat{\tau}_{it} = Y_{it} - \hat{Y}_{it}$$

where $\hat{Y}_{it}$ represents the synthetic control for unit $i$ formed as a weighted average of the control units $j \in \{1, ..., J\}$:

$$\hat{Y}_{it} = \sum_{j=1}^{J} w_j Y_{jt} \tag{2.4}$$

The pool of potential donors is the same as the control group used in the DID specification so as to provide a donor pool that is broadly qualitatively comparable to the treated units.

The weights in Equation 2.4 are chosen to minimize the distance between a vector of predictor variables for each treated unit and all potential control units. The specific predictors used in the selection procedure include three lags of the outcome of interest (values for the years 1985, 1987, and 1989 specifically), total undergraduate enrollment, tuition, undergraduate minority percentage, tuition revenues, research expenditure, and instructor expenditures.[18] The optimization further requires specification of a matrix of weights that place relative priority over the predictor variables.[19] This matrix was selected via a nested procedure that searches among the set of viable weights on both predictors and controls so as to minimize the prediction error between each treated unit and the corresponding synthetic

18. This specific set of predictor variables allowed for good performance in the sense that there was a unique optimal selection of weights for each Overlap school for each of the three primary outcomes of interest: total scholarship expenditures, unrestricted scholarship expenditures, and restricted scholarship expenditures. There was also a sparsity of donor schools for each treated school for each of the three outcomes listed above, in the sense defined by Abadie [2021] - fewer donors received positive weights than there are predictor variables.

19. The specific object of optimization is an expression $\sqrt{(X_i - X_c W)' V (X_i - X_c W)}$, where $V$ is a diagonal matrix of weights placed upon the predictors, $W$ is a vector of non-negative weights that sum to one, $X_i$ is a vector containing the predictor variables for each treatment unit, and $X_c$ is a matrix containing the values of the predictor variables for all possible control units.

control, as proposed by Abadie and Gardeazabal [2003] and Abadie et al. [2010].[20]

Inference in the context of synthetic controls is typically done via some form of randomization. I follow the example of Cavallo et al. [2013] and perform an exact inference technique. This method proceeds by forming a synthetic control for each control unit using the set of other available control units, following the same procedure as in the creation of the synthetic controls for the treated units. A large number, specifically 1,000,000, permutations are then formed, with each permutation assigning one of the control units in place of one of the treated units. The average estimated effect, by year after treatment, on the treated units is then ranked among the similar estimates from each of the permutations. The proportion of permutations that possesses a lower estimated effect than that of the treated units is then reported as a p-value, with a separate value being assigned for each individual year following the year of the last meeting.

## 2.5 Results

### *2.5.1 Scholarship Spending*

I begin by assessing the effect of the Overlap Group on scholarship spending and, implicitly, the effective price faced by students. Both the Justice Department and the simple model predicted increased scholarship spending in the absence of collusion.

Figures 2.1-2.3 present plots of the differences in total, unrestricted, and restricted scholarship spending between the Overlap and control schools over the period of interest, broken down by year, with 1990 being the leave out year. The plots for total and unrestricted scholarship shares demonstrate reasonably parallel trends in the pre-period, providing some degree of confidence that the crucial parallel trends assumption may hold after the begin-

20. All synthetic control construction in this paper was done via the Stata package synth developed by Alberto Abadie, Alexis Diamond, and Jens Hainmueller as well as synth_runner (Galiani and Quistorf [2017]) and allsynth (Wiltshire [2021]).

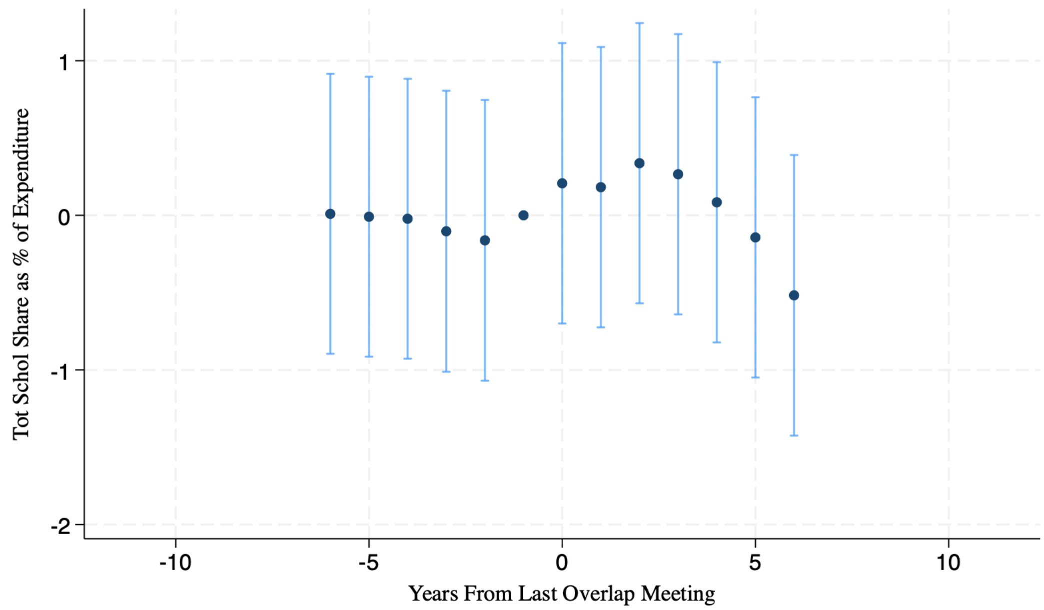



Figure 2.1: DID Between Overlap Schools and Controls - Total Scholarship Spending

ning of the investigation. There also appears to be a consistent upward trend after 1990, particularly for the unrestricted stream, after the investigation begins, suggesting a possible effect of the Overlap Group, although the yearly estimates are not significant at 5%, which is perhaps unsurprising given the small number of universities in consideration. The restricted stream plot, being considerably noisier, and seeming to lack parallel trends in the pre-period, is harder to draw any meaningful information from.

Table 2.2 presents the results of estimating Equation 2.3 on the full set of schools, both with robust standard errors and standard errors that are clustered at the university level. The regression results provide some evidence that total and unrestricted scholarship shares increased in Overlap schools after 1990, with both streams having positive results and being highly significant, although clustering the standard errors removes all significance. The results also possess a fair degree of economic significance. In 1990, the average share of expenditures on total scholarships among all universities in the sample was 10.2% while the corresponding share for unrestricted scholarships was 5.7%. The point estimates would then

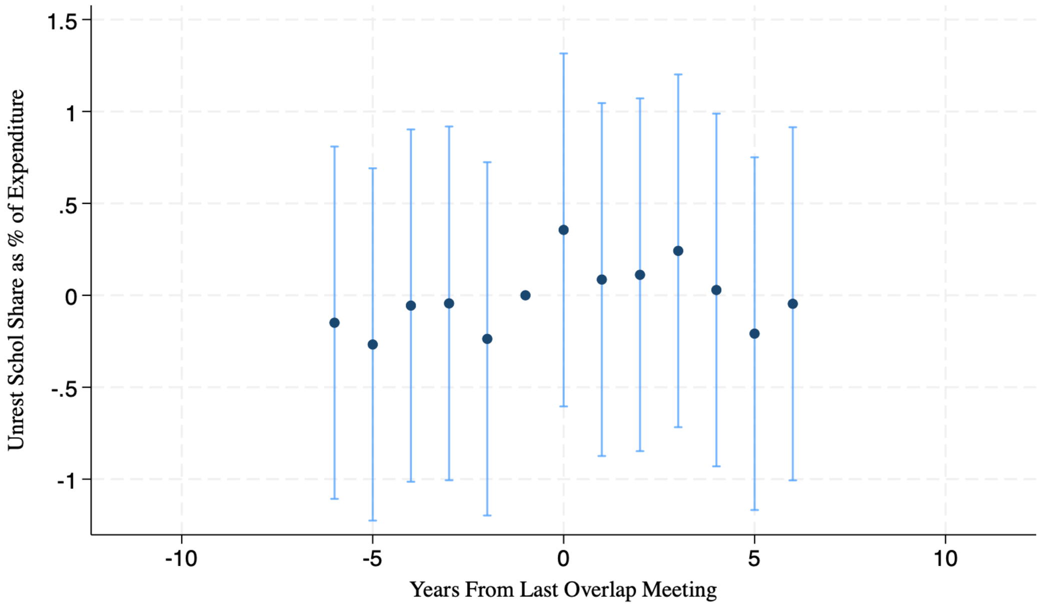



Figure 2.2: DID Between Overlap Schools and Controls - Unrestricted Scholarship Spending

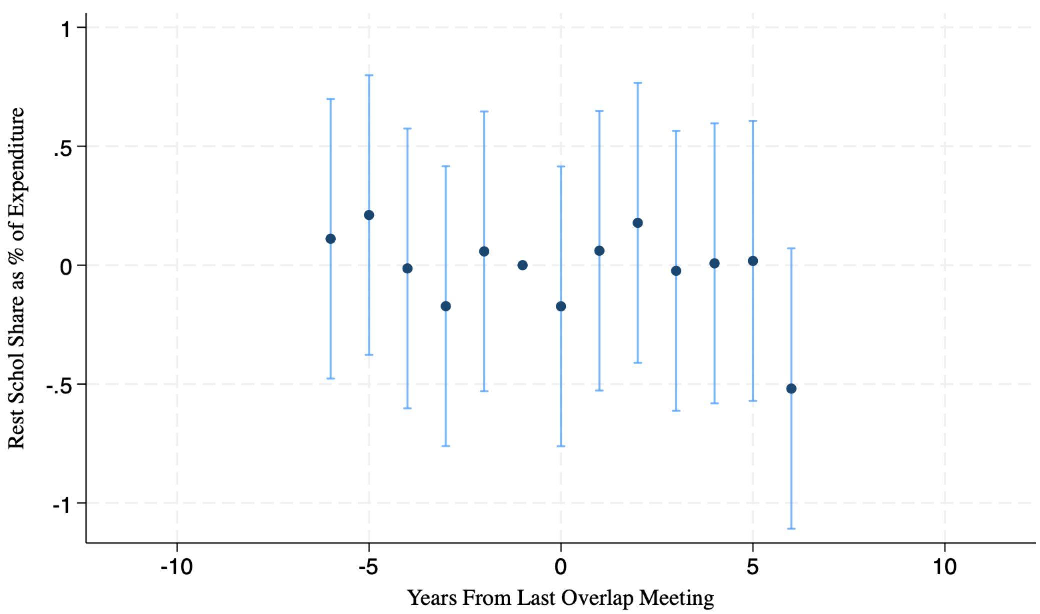



Figure 2.3: DID Between Overlap Schools and Controls - Restricted Scholarship Spending

Table 2.2 – DID Estimates of Scholarship Spending

| Total Scholarships | Unrestricted Scholarships | Rest Scholarships |
|---|---|---|
| 0.623 | 0.733 | -0.123 |
| (0.223)*** | (.226)*** | (0.149) |
| [0.574] | [0.573] | [0.300] |

Notes: Reported coefficients are in terms of the number of percentage points of overall university expenditures comprised by each stream of scholarship spending. Robust standard errors are reported in parentheses. Clustered standard errors (clustered at university level) are reported in brackets. * indicates significance at a level of 0.1, ** indicates significance at a level of 0.05, *** indicates significance at a level of 0.01.

imply that the cessation of Overlap meetings led to increases of 6% and 13% relative to the average scholarship expenditures shares of the time. The restricted scholarship stream, on the other hand, while estimated to decrease in Overlap schools after 1990 is estimated too imprecisely to make any meaningful inference.

Figures 2.4-2.6 plot the average total, unrestricted, and restricted scholarship spending for both the Overlap schools and their full set of synthetic controls, with a line at 1990 indicating the year of the final Overlap meeting. All three figures exhibit a high degree of fit between the averages of the treated units and the synthetic controls, demonstrating the potential advantages of the synthetic control method in creating appropriate counterfactuals for the treated units. The total spending appears to show a clear trend break after the year of the final Overlap meeting, with the Overlap schools noticeably moving substantially above the synthetic controls at that time. In contrast to the DID analysis, the unrestricted Overlap averages appear to track the synthetic ones quite well while the restricted numbers seem to hold steady while the synthetic controls noticeably fall. Similar plots of the differences in the averages appear in Appendix B.2.

The average yearly estimated effect of the cessation of Overlap meetings on scholarship spending, based on the synthetic controls, are reported in Table 2.3, along with p-values from the exact inference technique described in Section 2.4.3. The coefficients are largely in line with what is visible in Figures 2.4-2.6. The estimated coefficients are generally positive across

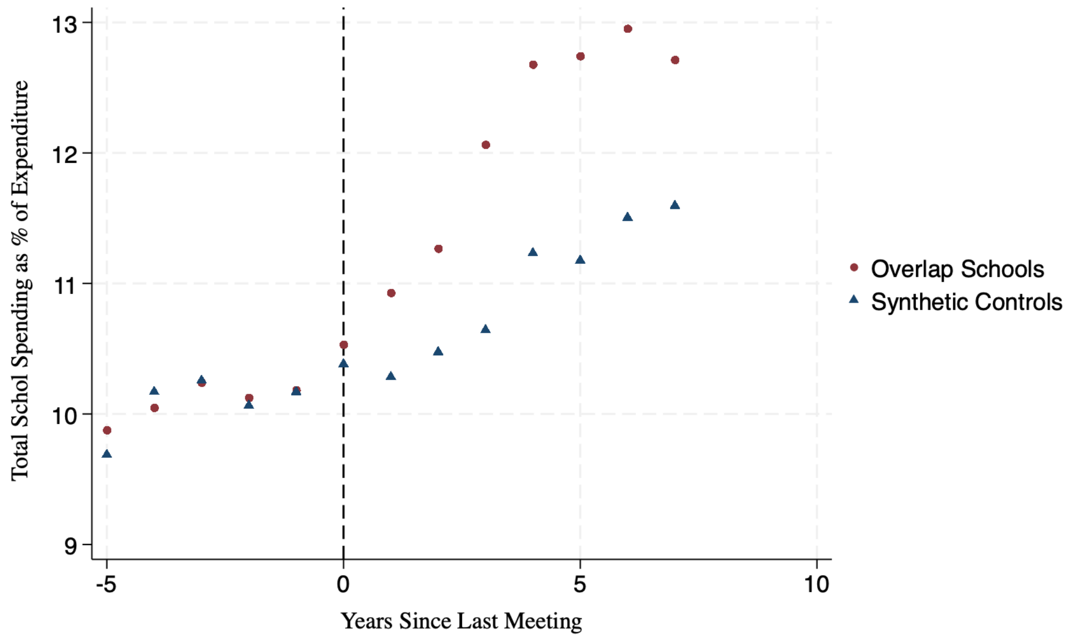



Figure 2.4: Overlap Schools Compared to Synthetic Controls - Total Scholarship Spending

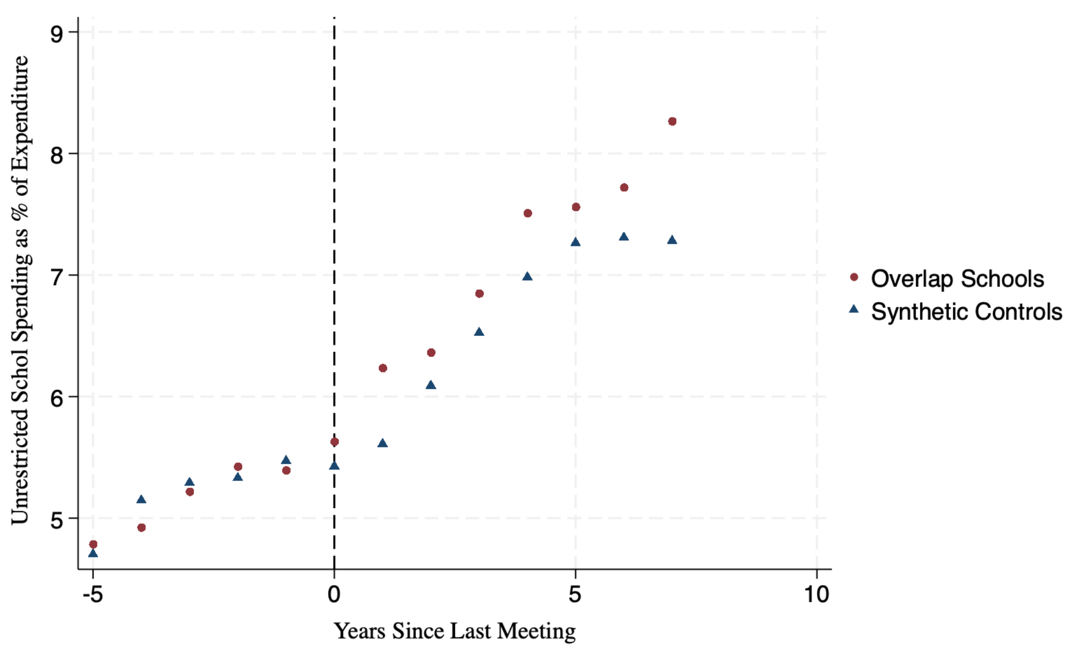



Figure 2.5: Overlap Schools Compared to Synthetic Controls - Unrestricted Scholarship Spending

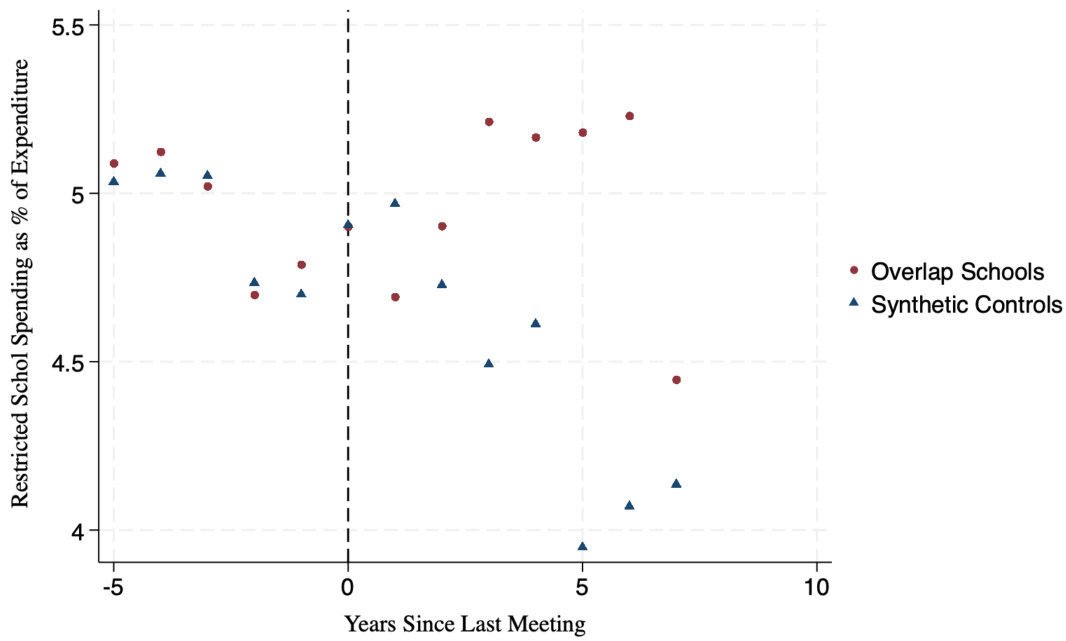



Figure 2.6: Overlap Schools Compared to Synthetic Controls - Restricted Scholarship Spending

all years and all three streams of scholarship expenditure, with those for total and restricted scholarships often being highly significant. The coefficients for those two outcomes are also economically significant. The estimated effect on total scholarship expenditure reaches a maximum of 1.4% four years after the cessation of overlap meetings. Given an average total scholarship expenditure of 10% of all expenditures in 1990, this represents a major effect. The effect on restricted scholarships also reach a maximum of 1.4%, compared to an average of 4.5% in 1990. If this estimate is taken literally, it would suggest that participating in the Overlap Group may have been suppressing restricted scholarship expenditure share by nearly a third of the typical mean share of the time.

Taking the DID and synthetic controls results together appears to offer substantial evidence that, at the very least, the Overlap Group members were offering less total financial aid before the group dissolved. This would suggest that the basic prediction of the model that collusion was reducing average financial aid spending, and thereby increasing average

Table 2.3 – SC Estimates on Scholarship Expenditure Shares by Year After Last Meeting

| | Total Scholarships | Unrestricted Scholarships | Restricted Scholarships |
|---|---|---|---|
| 1 | 0.634 | 0.651 | -0.276 |
| | (0.019) | (0.001) | (0.170) |
| 2 | 0.728 | 0.265 | 0.176 |
| | (0.024) | (0.364) | (0.455) |
| 3 | 1.253 | 0.260 | 0.681 |
| | (0.039) | (0.473) | (0.014) |
| 4 | 1.367 | 0.435 | 0.537 |
| | (0.036) | (0.203) | (0.059) |
| 5 | 1.354 | 0.182 | 1.181 |
| | (0.101) | (0.819) | (0.014) |
| 6 | 1.331 | 0.389 | 1.108 |
| | (0.074) | (0.653) | (0.028) |
| 7 | 1.164 | 0.982 | 0.261 |
| | (0.095) | (0.098) | (0.606) |

Notes: Reported coefficients are in terms of the number of percentage points of overall university expenditures comprised by each stream of scholarship spending. P-values are reported in parentheses. P-values are derived from an exact inference technique derived in Carvallo et al. (2012).

prices, is correct in this case.

### *2.5.2 Other Expenditures*

If the proportion of university expenditures on scholarships increases for Overlap schools following the end of the annual meetings, that money must be being displaced from other types of spending. IPEDS reports total expenditures broken into several categories. This section reports the DID and synthetic control estimated effects on spending on each of the non-scholarship IPEDS spending categories.

Table 2.4 contains the DID estimates and Table 2.5 contains the synthetic controls estimates.[21,22] The results from these exercises are less clear than those regarding scholarship spending, and hard to draw any overly strong conclusions from. However, they might provide some suggestive evidence of where money saved via collusion was spend during the Overlap period.

Across both tables, the largest point estimates are for instructional spending, which is estimated to fall using both methods, although it is not statistically significant in the DID specification or in all but one year of the synthetic control estimation. Research spending is also estimated to fall in the DID specification, although the magnitude is relatively small and insignificant. In the synthetic controls estimation, research is also estimated to fall in the initial years after the cessation of Overlap meetings, with the estimates being significant for the second and third years after dissolution.

Instruction and research expenditures seem to be among the most obvious components of what might be considered non-student prestige for universities. Faculty compensation and qualifications directly enter into the US News and World Report University rankings, which students and schools have been shown to be responsive to (Morse and Brooks [2021]).[23] Models of student application and enrollment behavior also commonly assume either that schools directly maximize quality of education or that students respond directly to quality of education, which presumably rises in spending on instruction.[24] The results on spending on instruction and research might then be taken as suggestive evidence that displaced scholarship money is spent on non-student prestige, suggesting that the competitive equilibrium

21. Table 2.5 excludes the categories of Hospital and Independent Operations spending to conserve space. Both of these categories can be seen as outside the primary operations of the university (Independent Operations refers to things like federally funded, non-educational research centers). They also pose particular empirical challenges, as these categories are either zero or very small for many schools throughout the entire period of interest. The synthetic controls results for these categories are generally similar to those in Table 2.4, being both small in magnitude and insignificant.

22. Figures in the style of Figures 2.1-2.3 and Figures 2.4-2.6 for the variables discussed in this section are available in Appendix B.3.

23. See Meredith [2004] and Bastedo and Bowman [2010]

24. See Epple et al. [2003], Epple et al. [2006], and Arcidiacono [2005].

is of a research-seeking type.

While the concept of "profit" was left fairly nebulous in the model, an easy and cynical interpretation that can be loosely mapped to the IPEDS categorization would be that profit could include the salaries for high-level administrative officials at universities. To the extent that such individuals influence budgetary decisions at university, they could accrue excess revenues as compensation. These salaries would be included in the Institutional Support and Academic Support categories in IPEDS, with the highest-level individuals generally being included in the former. Strikingly, both of these quantities are estimated to increase following the cessation of Overlap meetings. This is in direct contradiction to the prediction of what would happen if the competitive equilibrium was of a profit-seeking type, assuming that these types of salaries could indeed be thought of as part of a profit term. This would seem to provide evidence against either the competitive equilibrium being such that excess income is spent on profit or the interpretation of executive salaries as profit for non-profit institutions, although definitive interpretation is made difficult due to the crudity of the IPEDS reporting categories, which obfuscates the precise destinations of the spending.[25]

The results on where displaced spending from increased scholarship provision comes from, while not overly strong, would generally seem to agree with the idea that at least some of the money goes towards non-student prestige for universities. As this goal includes activities that are relatively non-controversial in terms of their social benefit, such as education quality and research, this complicates the welfare implication of collusion in financial aid setting. While it does seem that collusion dampens aid generosity, decreasing student surplus, whether this is good or bad depends on the weights placed on student surplus compared to educational quality and research output (which of course may also impact student surplus.).

25. One theory that could explain why these categories might increase in expenditure share would be that the cessation of collusion increases the importance of internal decisions regarding admissions and financial aid offerings. While this could be the case, both admissions activity and financial aid determination are meant to be included in the Student Services category of IPEDS.

Table 2.4 – DID Estimates on Non-Scholarship Expenditures After Last Meeting

| | Instruction | Research | Public Service | Academic Support | Student Services | Institutional Support |
|---|---|---|---|---|---|---|
| | -0.594 | -0.168 | 0.001 | 0.387 | 0.039 | 0.444 |
| | (0.368) | (0.186) | (0.002) | (0.278) | (0.124) | (0.203)** |
| | [0.810] | [0.471] | [0.004] | [0.692] | [0.185] | [0.398] |
| Observations | 598 | 589 | 598 | 598 | 593 | 598 |
| | Operation of Plant | Educational Transfers | Auxiliary | Hospitals | Industrial Operations | |
| | -0.004 | 0.005 | -0.297 | -0.002 | -0.006 | |
| | (0.002)** | (0.006) | (0.245) | (0.004) | (0.004) | |
| | [0.004] | [0.011] | [0.589] | [0.009] | [0.008] | |
| Observations | 598 | 521 | 598 | 598 | 598 | |

Notes: Reported coefficients are in terms of percentage points of overall university expenditures comprised by each stream of spending. All regressions include university and year fixed effects. Robust standard errors are reported in parentheses. Clustered standard errors (clustered at university level) are reported in brackets. * indicates significance at a level of 0.1, ** indicates significance at a level of 0.05, *** indicates significance at a level of 0.01.

Table 2.5 – SC Estimates on Non-Scholarship Expenditure Shares by Year After Last Meeting

| Year | Instruction | Research | Public Service | Academic Support | Student Services | Institutional Support | Operation of Plant | Auxiliary |
|---|---|---|---|---|---|---|---|---|
| 1 | -0.772 | -0.340 | -0.324 | 0.802 | -0.289 | 0.163 | -0.829 | -0.270 |
| | (0.750) | (0.152) | (0.052) | (0.032) | (0.209) | (0.728) | (0.000) | (0.387) |
| 2 | -1.227 | -0.607 | -0.147 | 0.689 | -0.466 | 0.399 | -0.223 | 0.0701 |
| | (0.366) | (0.029) | (0.401) | (0.098) | (0.040) | (0.161) | (0.194) | (0.849) |
| 3 | -1.264 | -0.669 | -0.107 | 0.760 | 0.285 | 0.754 | 0.261 | -0.475 |
| | (0.260) | (0.005) | (0.512) | (0.029) | (0.100) | (0.035) | (0.425) | (0.104) |
| 4 | -1.136 | 0.478 | -0.0327 | 0.743 | 0.128 | 0.194 | 0.230 | -0.286 |
| | (0.194) | (0.022) | (0.845) | (0.089) | (0.533) | (0.447) | (0.483) | (0.344) |
| 5 | -1.196 | 0.290 | -0.0420 | 1.033 | -0.00563 | 0.137 | 0.463 | 0.209 |
| | (0.275) | (0.288) | (0.784) | (0.044) | (0.971) | (0.647) | (0.218) | (0.572) |
| 6 | -1.036 | 0.573 | 0.00662 | 0.671 | 0.168 | 0.245 | 0.368 | 0.129 |
| | (0.295) | (0.210) | (0.962) | (0.150) | (0.352) | (0.383) | (0.245) | (0.758) |
| 7 | -2.091 | -0.0343 | -0.148 | 1.435 | 0.0787 | 1.104 | 0.0493 | -0.344 |
| | (0.025) | (0.949) | (0.316) | (0.004) | (0.752) | (0.000) | (0.888) | (0.606) |

Notes: Reported coefficients are in terms of percentage points of overall university expenditures comprised by each stream of spending. P-values are reported in parentheses. P-values are the result of an exact inference technique derived in Carvallo et al. (2012).

## 2.6 Conclusion

Collusion in the determination of financial aid appears to decrease average levels of aid across students, at least for the most elite schools in the country. This result is consistent with a simple model of price-setting among universities that care about student quality as well as other objectives.

However, the welfare implication of this result is complicated by at least two concerns. One is that money not spent on financial aid during a period of collusion appears to be redirected towards objectives such as instruction and research, which are likely to be seen as beneficial to society and the students in question. This would be consistent with a set of circumstances such that the preeminent concern for schools, outside of student quality, is their non-student prestige, which includes these objectives.

Another concern is the distribution of financial aid between students. Given the lack of granularity in the data covering the time period of interest, it is hard to empirically assess how different types of students were affected by the change in overall aid generosity. A simple model would predict that the students who offer the greatest value to the school would experience the greatest decline in aid received. However, it is both unclear how schools value different types of students and how heterogeneity in student willingness-to-pay would affect this prediction. Future work might more carefully consider the distribution of aid across student types in different market settings.

## REFERENCES

568 Presidents Group. A brief summary and analysis of section 568. 2007.

Alberto Abadie. Using synthetic controls: Feasibility, data requirements, and methodological aspects. *Journal of Economic Literature*, 59(2):391–425, 2021.

Alberto Abadie and Javier Gardeazabal. The economic costs of conflict: A case study of the basque country. *American Economic Review*, 93(1):113–132, 2003.

Alberto Abadie, Alexis Diamond, and Jens Hainmueller. Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program. *Journal of the American Statistical Association*, 105(490):493–505, 2010.

Thomas Ahn, Peter Arcidiacono, Amy Hopson, and James R. Thomas. Equilibrium grade inflation with implications for female interest in stem majors. *NBER Working Papers*, 2019.

Peter Arcidiacono. Ability sorting and the returns to college major. *Journal of Econometrics*, 121(1-2):343–375, 2004.

Peter Arcidiacono. Affirmative action in higher education: How do admission and financial aid rules affect future earnings? *Econometrica*, 73(5):1477–1524, 2005.

Peter Arcidiacono, V. Joseph Hotz, and Songman Kang. Modeling college major choices using elicited measures of expectations and counterfactuals. *Journal of Econometrics*, 166:3–16, 2012.

Peter Arcidiacono, Esteban Aucejo, Arnaud Maurel, and Tyler Ransom. College attrition and the dynamics of information revelation. *NBER Working Papers*, 2016.

Esteban Aucejo and Jonathan James. The path to college education: The role of math and verbal skills. *Journal of Political Economy*, 129(10):2721–2993, 2021.

Michael N. Bastedo and Nicholas A. Bowman. U.s. news & world report college rankings: Modeling institutional effects on organizational reputation. *American Journal of Education*, 116:163–183, 2010.

Eric P. Bettinger and Bridget Terry Long. Do faculty serve as role models? the impact of instructor gender on female students. *American Economic Review*, 95(2):152–157, 2005.

Sandra E. Black, Paul J. Devereux, and Kjell G. Salvanes. Under pressure? the effect of peers on outcomes of young adults. *Journal of Labor Economics*, 31(1):119–153, 2013.

Peter Q. Blair and Kent Smetters. Why don't elite colleges expand supply. *NBER Working Papers*, 2021.

Sarah Bohn, Magnus Lofstrom, and Steven Raphael. Did the 2007 legal arizona workers act reduce the state's unauthorized immigrant population. *Review of Economic and Statistics*, 96(2):258–269, 2014.

Valerie K. Bostwick and Bruce A. Weinberg. Nevertheless she persisted? gender peer effects in doctoral stem programs. *NBER Working Papers*, 2018.

Charles Brown and Mary Corcoran. Sex-based differences in school content and the male-female wage gap. *Journal of Labor Economics*, 15(3):431–465, 1997.

Thomas Buser, Muriel Niederle, and Hessel Oosterbeek. Gender, competitiveness, and career choices. *The Quarterly Journal of Economics*, 129(3):1409–1447, 2014.

David Card and A. Abigail Payne. High school choices and the gender gap in stem. *NBER Working Papers*, 2017.

Dennis W. Carlton, Gustavo E. Bamberger, and Roy J. Epstein. Antitrust and higher education: Was there a conspiracy to restrict financial aid? *The RAND Journal of Economics*, 26(1):131–147, 1995.

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

Eduardo Cavallo, Sebastian Galiani, Ilan Noy, and Juan Pantano. Catastrophic natural disasters and economic growth. *The Review of Economics and Statistics*, 95(5):1549–1561, 2013.

Stephanie Riegg Cellini and Claudia Goldin. Does federal student aid raise tuition? new evidence on for-profit colleges. *American Economic Journal: Economic Policy*, 6(4):174–206, 2014.

Susan Chira. 23 colleges won't pool fiscal data. *The New York Times*, 1991.

Scott Cunningham and Manisha Shah. Decriminalizing indoor prostitution: Implications for sexual violence and public health. *Review of Economic Studies*, 85(3):1683–1715, 2018.

Giacomo De Giorgi, Michele Pellizzari, and Silvia Redaelli. Identification of social interactions through partially overlapping peer groups. *American Economic Journal: Applied Economics*, 2(2):241–275, 2010.

Anthony DePalma. Ivy universities deny price-fixing but agree to avoid it in the future. *The New York Times*, 1991.

Michael F. Dinerstein, Caroline M. Hoxby, Jonathan Meer, and Pablo Villanueva. Did the fiscal stimulus work for universities? In Jeffrey R. Brown and Caroline M. Hoxby, editors, *How the Financial Crisis and Great Recession Affected Higher Education*. University of Chicago Press, Chicago IL, 2015.

Susan Dodge. Overlap group makes aid process fairer, targets of inquiry argue. *The Chronicle of Higher Education*, 1989.

John J. Donohue, Abhay Aneja, and Kyle D. Weber. Right-to-carry laws and violent crime: A comprehensive assessment using panel data and a state-level synthetic control analysis. *Journal of Empirical Legal Studies*, 16(2):198–247, 2019.

Dennis Epple, Richard Romano, and Holger Sieg. Peer effects, financial aid and selection of students into colleges and universities: An empirical analysis. *Journal of Applied Econometrics*, 18:501–525, 2003.

Dennis Epple, Richard Romano, and Holger Sieg. Admission, tuition, and financial aid policies in the market for higher education. *Econometrica*, 74(4):885–928, 2006.

Robert W. Fairlie, Florian Hoffman, and Philip Oreopoulos. A community college instructor like me: Race and ethnicity interactions in the classroom. *The American Economic Review*, 104(8):2567–2591, 2014.

Chao Fu. Equilibrium tuition, applications, admissions, and enrollment in the college market. *Journal of Political Economy*, 122(2):225–281, 2014.

Sebastian Galiani and Brian Quistorf. The synth runner package: Utilities to automate synthetic control estimation using synth. *The Stata Journal*, 17(4):834–849, 2017.

Ahu Gemici and Matthew Wiswall. Evolution of gender differences in post-secondary human capital inverstments: College majors. *International Economic Review*, 55(1):23–56, 2014.

Jie Gong, Yi Lu, and Hong Song. Gender peer effects on students' academic and noncognitive outcomes. *Journal of Human Resources*, 56(3):686–710, 2021.

Amanda L. Griffith and Joyce B. Main. First impressions in the classroom: How do class characteristics affect student grades and majors? *Economics of Education Review*, 69: 125–137, 2019.

Harvard University. Frequently asked questions. *Harvard University Committee on General Scholarships*, 2022.

Andrew J. Hill. The positive influence of female college students on their male peers. *Labour Economics*, 44:151–160, 2017.

Will Hobson and Steven Rich. Why students foot the bill for college sports, and how some are fighting back. *The Washington Post*, 2015.

Florian Hoffman and Philip Oreopoulos. A professor like me: The influence of instructor gender on college achievement. *The Journal of Human Resources*, 44(2):479–494, 2009.

Caroline Hoxby. Peer effects in the classroom: Learning from gender and race variation. *NBER Working Papers*, 2000.

Caroline M. Hoxby. How the changing market structure of u.s. higher education explains college tuition. *NBER Working Papers*, 1997.

Chang-Tai Hsieh, Erik Hurst, Charles I. Jones, and Peter J. Klenow. The allocation of talent and u.s. economic growth. *Econometrica*, 87(5):1439–1474, 2019.

Feng Hu. Do girl peers improve your academic performance? *Economics Letters*, 137:54–58, 2015.

Matt Krupnick and Jon Marcus. Think university administrators' salaries are high? critics say their benefits are lavish. *The Hechinger Report*, 2015.

Victor Lavy and Analia Schlosser. Mechanisms and impacts of gender peer effects at school. *American Economic Journal: Applied Economics*, 3(2):1–33, 2011.

Los Angeles Times. Yale, swarthmore ranked no. 1 by news magazine. *Los Angeles Times*, 1989.

Thao P. Matlock. The overlap group: A study of nonprofit competition. *Journal of Law & Education*, 23(4):523–547, 1994.

Marc Meredith. Why do universities compete in the ratings game? an empirical analysis of the effects of the u.s. news and world report college rankings. *Research in Higher Education*, 45:443–461, 2004.

MIT News. A brief history of overlap and the antitrust suit. *MIT News*, 1992.

MIT News. Settlement allows cooperation on awarding financial aid. *MIT News*, 1994.

Robert Morse and Eric Brooks. A more detailed look at the ranking factors. *US News & World Report*, 2021.

Lincoln Mullen. *gender: Predict Gender from Names Using Historical Data*, 2021. URL `https://github.com/lmullen/gender`. R package version 0.6.0.

Hessel Oosterbeek and Reyn van Ewijk. Gender peer effects in university: Evidence from a randomized experiment. *Economics of Education Review*, 38:51–63, 2014.

Arpita Patnaik, Matthew J. Wiswall, and Basit Zafar. College majors. *NBER Working Papers*, 2020.

Joshua Price. The effect of instructor race and gender on student persistence in stem fields. *Economics of Education Review*, 29:901–910, 2010.

Kevin Rask and Jill Tiefenthaler. The role of grade sensitivity in explaining the gender imbalance in undergraduate economics. *Economics of Education Review*, 27(6):676–687, 2008.

Stephanie Saul and Anemona Hartocollis. Lawsuit says 16 elite colleges are part of price-fixing cartel. *The New York Times*, 2022.

Ralph Stinebrickner and Todd R. Stinebrickner. A major in science? initial beliefs and final outcomes for college major and drop. *Review of Economic Studies*, 81:426–472, 2014.

Lesley J. Turner. The economic incidence of federal student grant aid. Working Paper, 2017.

University of Washington. Glossary of terms for award notifications. 2022.

US News & World Report. National liberal arts colleges rankings. *US News & World Report*, 2022.

Diane Whitmore. Resource and peer impacts on girls' academic achievement: Evidence from a randomized experiment. *American Economic Review*, 95(2):199–203, 2005.

Justin Wiltshire. allsynth: Synthetic control bias-corrections utilities for stata. *2021 Stata Conference 15, Stata Users Group*, 2021.

Matthew Wiswall and Basit Zafar. Preference for the workplace, investment in human capital, and gender. *The Quarterly Journal of Economics*, 133(1):457–507, 2017.

Basit Zafar. College major choice and the gender gap. *The Journal of Human Resources*, 48 (3):545–595, 2013.

Ulf Zolitz and Jan Feld. The effect of peer gender on major choice in business school. *IZA Institute of Labor Economics Discussion Paper Series*, 2020.

# APPENDIX A

# CHAPTER 1: SUPPLEMENTARY MATERIAL

## A.1 Only Utilizing Within-Semester Variation

The results presented throughout the main body of the text utilize variation in classes within courses both between the Fall of 2015 and the Fall of 2016 and across lecture times within each semester. However, each form of variation may introduce separate identification concerns. I present here results that only utilize within-semester variation. Appendix A.2 presents results for only across-semester variation. For the results here, I estimate equation (1.1), but instead of using course-specific female fixed effects, I use course-by-year-specific female fixed effects. That is, in lieu of a fixed effect for women in Econ 101, I include one fixed effect for women in Econ 101 in Fall of 2015 and another fixed effect for women in Econ 101 in Fall of 2016. The results are presented in Table A.1. The point estimates are quite similar to what is seen in Table 1.5 in the main text, although they are naturally somewhat less precise.

Table A.1 – Estimated Effect of Class Male Proportion on Student Outcomes, Using only Within-Semester Variation

| | Grade (GPA value) | Future course in dept. | Switch major out of dept. | Declare major in dept. | Graduate within 6 years | Graduate in dept. |
|---|---|---|---|---|---|---|
| Female student X | -0.227 | -0.064 | -0.083 | -0.031 | -0.078 | -0.032 |
| class male % | (0.169) | (0.051) | (0.231) | (0.024) | (0.061) | (0.027) |
| Outcome Mean | 2.988 | 0.644 | 0.138 | 0.054 | 0.677 | 0.127 |
| Outcome SD | 1.099 | 0.479 | 0.345 | 0.227 | 0.467 | 0.333 |
| Observations | 33071 | 43351 | 4750 | 37153 | 43351 | 43351 |

Notes: This table reports the results of regressions of student-class outcomes on an interaction between a female-student dummy and class male proportion. Each regression includes course-specific female fixed effects, class fixed effects, and controls for student underrepresented minority status and instructor-student gender match. All regressions are estimated on the observations within the identifying set of the main specification (as described in Section 1.2) for which the relevant outcome is observed and the given course only has only class in each of Fall of 2015 and Fall of 2016. Standard errors are clustered at the class level. * indicates significance at a level of 0.1, ** at a level of 0.05, and *** at a level of 0.01.

## A.2 Estimation Using Courses with Only One Class Per Year

Table A.2 – Estimated Effect of Class Male Proportion on Student Outcomes, For Courses that Offer One Class Per Year

| | Grade (GPA value) | Future course in dept. | Switch major out of dept. | Declare major in dept. | Graduate within 6 years | Graduate in dept. |
|---|---|---|---|---|---|---|
| Female student X | -0.463 | -0.131 | 0.168 | -0.064 | -0.322* | -0.133 |
| class male % | (0.291) | (0.150) | (0.181) | (0.078) | (0.183) | (0.101) |
| Outcome Mean | 3.057 | 0.561 | 0.110 | 0.065 | 0.716 | 0.239 |
| Outcome SD | 1.051 | 0.496 | 0.313 | 0.246 | 0.451 | 0.426 |
| Observations | 6758 | 8481 | 1986 | 5848 | 8481 | 8481 |

Notes: This table reports the results of regressions of student-class outcomes on an interaction between a female-student dummy and class male proportion. Each regression includes course-specific female fixed effects, class fixed effects, and controls for student underrepresented minority status and instructor-student gender match. All regressions are estimated on the observations within the identifying set of the main specification (as described in Section 1.2) for which the relevant outcome is observed and the given course only has only class in each of Fall of 2015 and Fall of 2016. Standard errors are clustered at the class level. * indicates significance at a level of 0.1, ** at a level of 0.05, and *** at a level of 0.01.

### A.3 Exploring Non-Linearity of Effects

Throughout the main body of the paper, I report results from specifications estimating linear effects of class-level male proportion. However, it is reasonable to think that important non-linearities in the effects of peer gender composition may exist. In this appendix, I investigate this possibility. Specifically, I estimate regressions of the forms

$$Y_{i,r,c}^{F} = \alpha_0^{F,NP} + \sum_{j=2}^{6} \alpha_j^{F,NP} I_j + \gamma_r^{F,NP} \times I_r + (X^{F,NP})'_{i,r,c} \beta^{F,NP} + u_{i,r,c}^{F,NP} \tag{A.1}$$

$$Y_{i,r,c}^{M} = \alpha_0^{M,NP} + \sum_{j=2}^{6} \alpha_j^{M,NP} I_j + \gamma_r^{M,NP} \times I_r + (X^{M,NP})'_{i,r,c} \beta^{M,NP} + u_{i,r,c}^{M,NP} \tag{A.2}$$

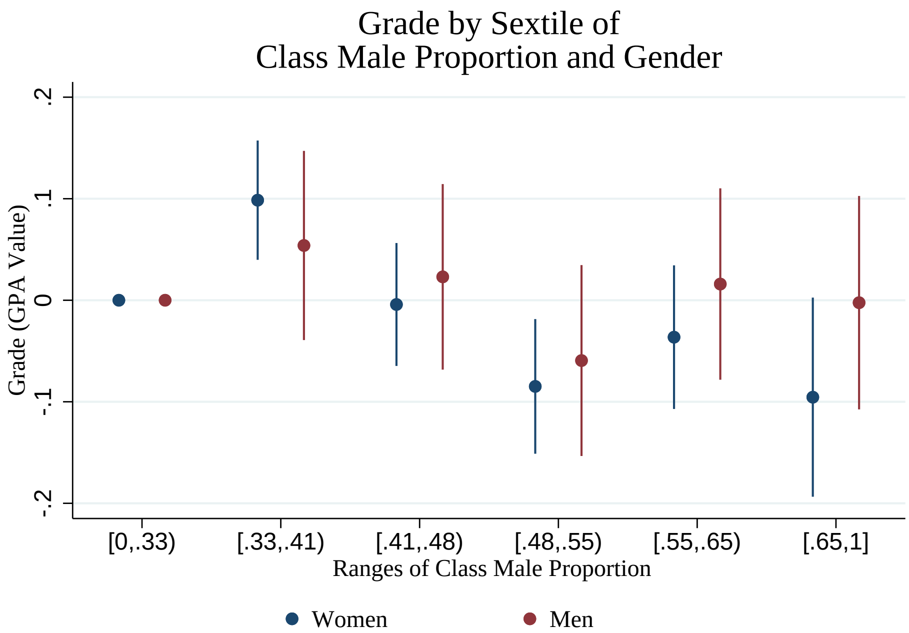
These regressions are nearly identical to equations (1.2) and (1.3), except, rather than estimating linear effects of class-level male proportion, they estimate coefficients on indicators of being in the second through sixth sextiles of class-level male proportion, $I_2, ..., I_6$ (with the first sextile being the omitted category). This will flexibly capture the form of non-linearities in the effects of peer gender. These regressions all include course fixed effects and controls for student underrepresented minority status and instructor gender, as with the within-gender specifications described in the main text. The results of this exercise are presented in Figures A.1-A.6, for each of the six main outcomes considered throughout the paper. For each outcome, equations (A.1) and (A.2) are estimated separately, but the estimates are plotted on the same graphs for concision and ease of comparison.

The results generally indicate that both the absolute effects of male peers and the differences in effects on women compared to men are concentrated among the most male-dominated classes, conditional on course. Looking at the effects on grade and graduation, there seems to be a generally declining pattern of female achievement with increasing male proportion, with the most negative point estimates being on the top sextile of class-level male proportion, for both outcomes. This is also where the differences in the point estimates of the effects on men and women are largest (although the differences are not significant).

Similarly, there seems to be an increasing likelihood of male declaration of a major with the male proportions of associated first-semester classes. The largest positive point estimate on this outcome for men, and the biggest difference relative to women, is for the top sextile of class male proportion. The same holds for the top sextile of class male proportion and graduation within a department associated to a class as an outcome.

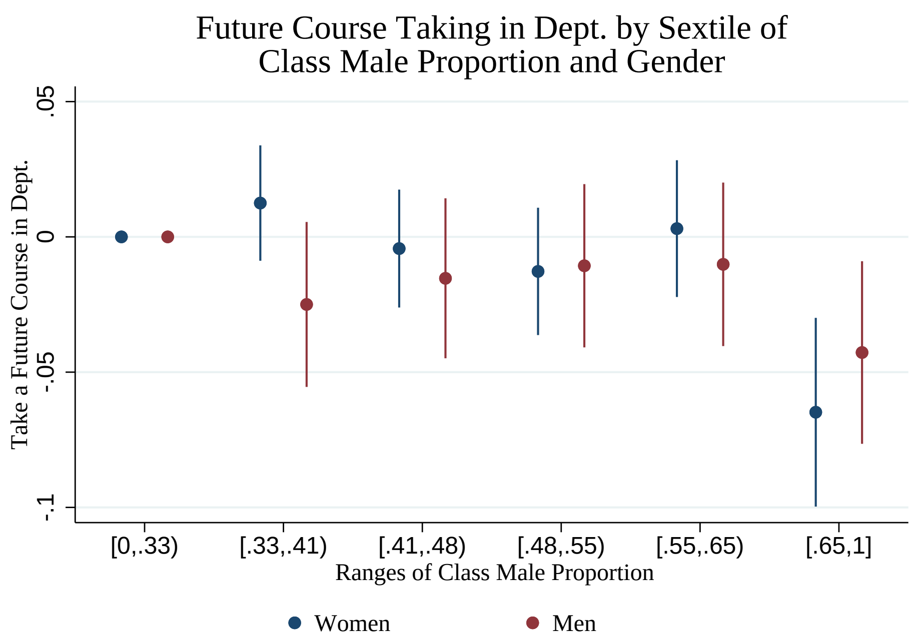
Many of the most male-dominated majors, both nationally and at UIC, are STEM majors, among the most highly paid majors on average, or both. If policy makers have particular interest in representation in STEM and the gender wage gap, a concentration of the effect of peer gender in the most heavily male-dominated environments underscores the importance of peer gender to real-world outcomes of interest. It is precisely in many STEM majors and many highly paid majors where we would expect peer gender to matter most, based on these results.

Figure A.1



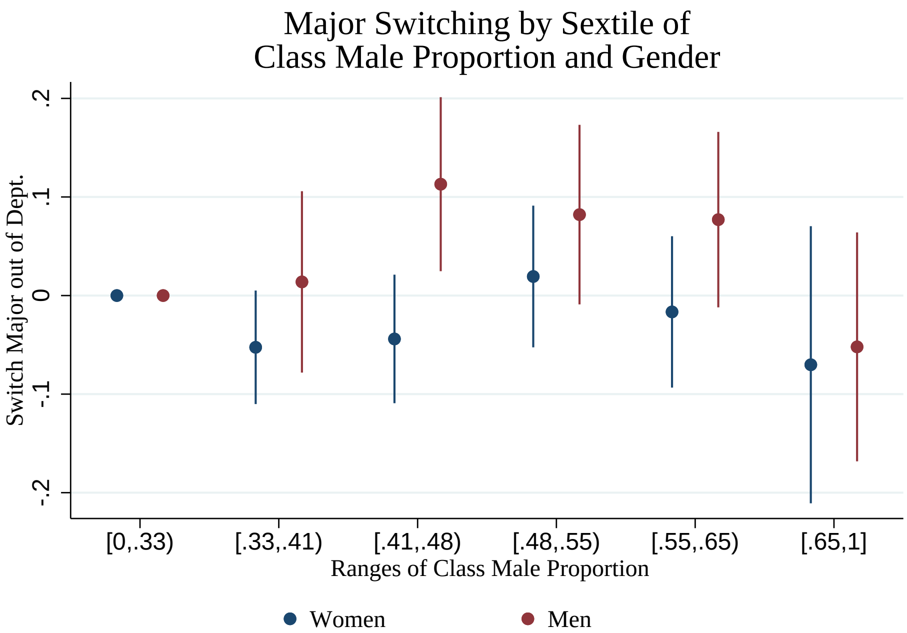

*Notes:* This figure presents the coefficients and 95% confidence intervals from regressions of the GPA value of grades on indicators for being in each sextile of class-level male proportion, conditional on course fixed effects and controls for instructor gender and student minority status. Coefficients for men and women are estimated separately. The first sextile is the omitted category.

Figure A.2



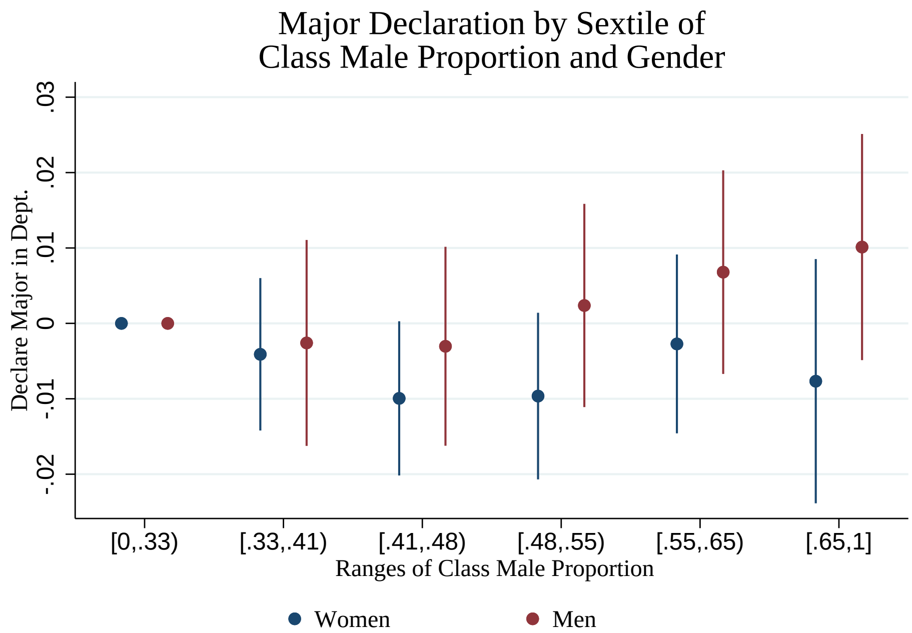

*Notes:* This figure presents the coefficients and 95% confidence intervals from regressions of an indicator for taking any future class in the same department on indicators for being in each sextile of class-level male proportion, conditional on course fixed effects and controls for instructor gender and student minority status. Coefficients for men and women are estimated separately. The first sextile is the omitted category.

Figure A.3



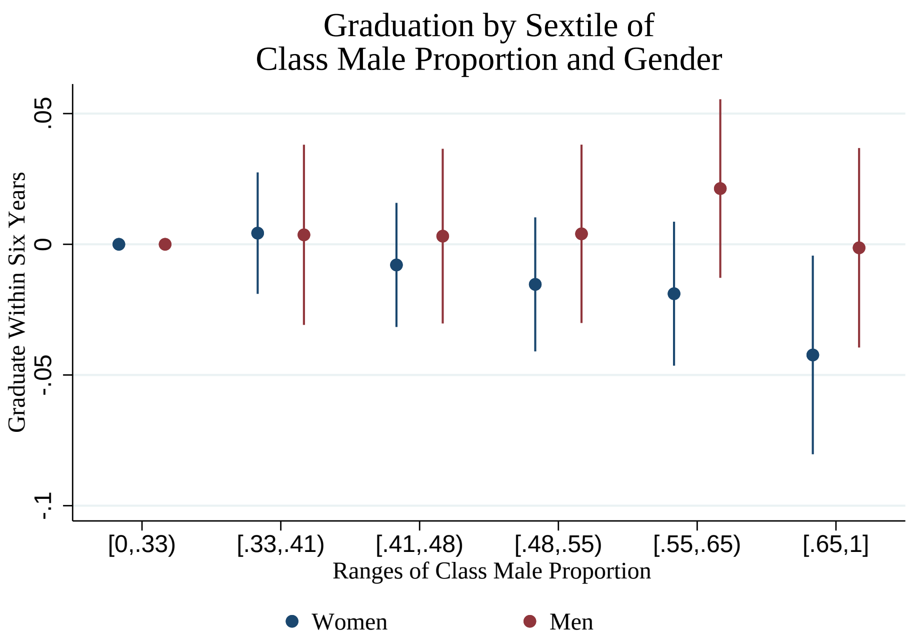

*Notes:* This figure presents the coefficients and 95% confidence intervals from regressions of an indicator of switching out of a major on indicators for being in each sextile of class-level male proportion, conditional on course fixed effects and controls for instructor gender and student minority status. Coefficients for men and women are estimated separately. The first sextile is the omitted category.

Figure A.4



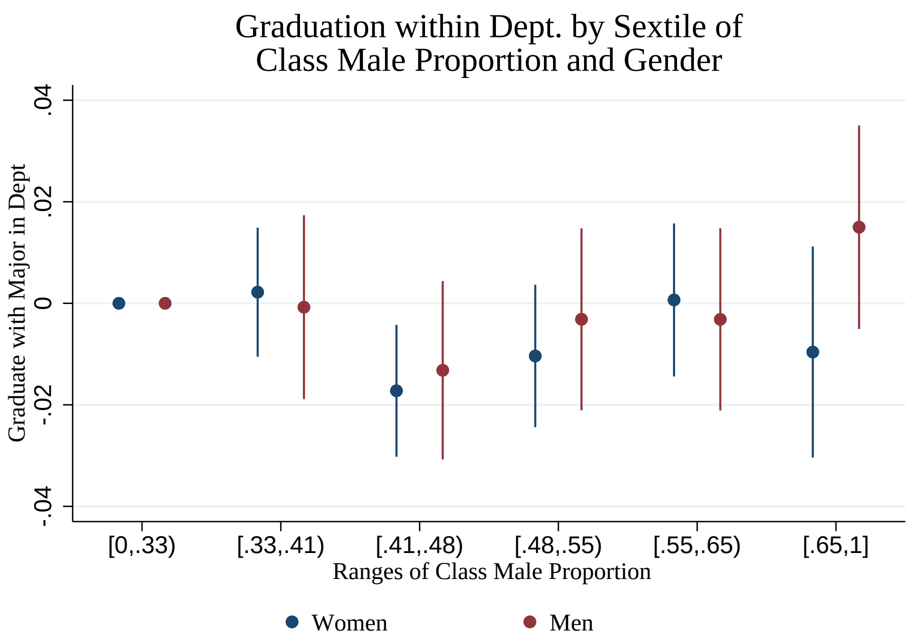

*Notes:* This figure presents the coefficients and 95% confidence intervals from regressions of an indicator for ever declaring a major in the same department on indicators for being in each sextile of class-level male proportion, conditional on course fixed effects and controls for instructor gender and student minority status. Coefficients for men and women are estimated separately. The first sextile is the omitted category.

Figure A.5




*Notes:* This figure presents the coefficients and 95% confidence intervals from regressions of an indicator for graduating within six years on indicators for being in each sextile of class-level male proportion, conditional on course fixed effects and controls for instructor gender and student minority status. Coefficients for men and women are estimated separately. The first sextile is the omitted category.

Figure A.6




*Notes:* This figure presents the coefficients and 95% confidence intervals from regressions of an indicator for graduating within six years with a declared major in the same department on indicators for being in each sextile of class-level male proportion, conditional on course fixed effects and controls for instructor gender and student minority status. Coefficients for men and women are estimated separately. The first sextile is the omitted category.

## A.4 Alternative Control Schemes

Table A.3 reports the results of estimating equation (1.1) while including both course-specific female fixed effects and class fixed effects but varying the sets of additional student and class-student controls. The specifications in column 1 include no controls beyond the fixed effects, column 2 specifications include controls for student minority status and student-instructor gender match (just as in column 4 of Table 1.5 in the main body of the paper), and column 3 includes the same controls as column 2 in additional to controls for high school GPA and ACT composite score. Table A.3 reveals that the set of controls used has very little impact on the estimated effects of class male proportion on student outcomes. Importantly, including controls for individual pre-college academic aptitude has little impact on the qualitative interpretation of the results. It does however curtail the precision of the results, due to the required exclusion of the portion of the sample that is missing information on either high school GPA, ACT score, or both.

Table A.3 – Estimated Effect of Class Male Proportion on Student Outcomes Under Varying Sets of Controls

| | (1) | (2) | (3) |
|---|---|---|---|
| *Grade (GPA value) [Mean = 2.99]* | | | |
| Female student X class male % | -0.319** | -0.309** | -0.227 |
| | (0.150) | (0.151) | (0.175) |
| Observations | 33071 | 33071 | 23018 |
| *Take a future course in same department [Mean = .64]* | | | |
| Female student X class male % | -0.067 | -0.067 | -0.079 |
| | (0.047) | (0.047) | (0.055) |
| Observations | 43351 | 43351 | 31645 |
| *Switch major to another department [Mean = .14]* | | | |
| Female student X class male % | 0.018 | 0.031 | 0.255 |
| | (0.149) | (0.146) | (0.649) |
| Observations | 4750 | 4750 | 1232 |
| *Declare a major in same department [Mean = .05]* | | | |
| Female student X class male % | -0.047* | -0.047* | -0.049* |
| | (0.025) | (0.025) | (0.028) |
| Observations | 37153 | 37153 | 29710 |
| *Graduate within six years [Mean = .68]* | | | |
| Female student X class male % | -0.111* | -0.106* | -0.104 |
| | (0.057) | (0.056) | (0.066) |
| Observations | 43351 | 43351 | 31645 |
| *Graduate with a major in same department [Mean = .13]* | | | |
| Female student X class male % | -0.047* | -0.046* | -0.052* |
| | (0.028) | (0.028) | (0.029) |
| Observations | 43351 | 43351 | 31645 |
| *Controls* | | | |
| Student underrepresented minority status | No | Yes | Yes |
| Student-instructor gender match | No | Yes | Yes |
| High school GPA | No | No | Yes |
| ACT composite score | No | No | Yes |

Notes: This table reports results of regressions of student-class outcomes on an interaction between a female-student dummy and class male proportion. Outcomes are given by row headers and fixed effects and controls specified by column feet. Standard errors are clustered at the class level. * indicates significance at a level of 0.1, ** at a level of 0.05, and *** at a level of 0.01.

## A.5 Section-Level Analysis

The results reported in the main body of the paper exploit variation in peer composition across classes within a course. This strategy assumes that there is no systematic, gendered sorting to the classes, within courses, that have more male students. This assumption is supported by the fact that first-semester students enroll in classes with little or no access to information about peer enrollments and by the highly comparable observable characteristics of students across classes with differing gender ratios, as reported in Table 1.2. However, there could still be reasonable concern that sorting to class characteristics, such as time slot or instructor traits, may induce differential sorting on unobservables between men and women, which would bias my results.

For classes that have associated laboratory, discussion, or practical experience sections, there is an additional level of variation to exploit: peer composition across sections, within classes. Utilizing this variation lets me to allow for any arbitrary pattern of sorting to classes by students. I then assume that, within a class, there is not meaningful sorting to sections that have more men. There are reasons why sorting to sections within a class may be “more" exogenous than the initial sorting to classes: sections have tighter capacity constraints, differences between sections may be less visible or salient than differences between classes, and section choices may be more constrained by schedules if choice of preferred classes is prioritized by students over choice of preferred sections. However, focusing on sections, rather than classes, introduces the notable drawback of reducing the available portion of the sample, as not all classes have sections, and reducing useful variation, as no classes have as many different sections as there are different classes within certain courses. Moreover, estimation using section-level variation implies a different parameter than estimation using class-level variation. It may be that class peers and section peers matter differently, depending on what the mechanisms are via which peers influence outcomes.

For my analysis of the effects of section peers on student outcomes, I emphasize two estimating equations. Analogous to my primary estimating equation for class-level analysis, equation (1.1), I estimate

$$Y_{i,r,c,s} = \pi_0 + \pi_1 \times Fem_i \times MP_s + \rho_c \times Fem_i \times I_c + \upsilon_s \times I_s + X'_{i,r,c,s}\tau + \varepsilon_{i,r,c,s} \quad \text{(A.3)}$$

where sections are indexed by $s$, $MP_s$ denotes the proportion of male students in section $s$, $I_s$ is an indicator variable that take on a value of one if outcome $Y_{i,r,c,s}$ is associated with section $s$, and $\varepsilon_{i,r,c,s}$ is the error term. $\pi_1$ captures how the gap between female and male outcomes evolves with $MP_s$. This specification includes both class-specific female fixed effects, $\rho_c$, section fixed effects, $\upsilon_s$, and controls for student and student-section characteristics, $\tau$. This allows for both any kind of sorting to classes and gender-neutral sorting to sections, meaning that differential sorting to sections within class by men compared to women is the only threat to identification.

In practice, the results of estimating equation (A.3) are too imprecise to be interpretable. I therefore focus on a specification without section fixed effects,

$$Y_{i,r,c,s} = \eta_0 + \eta_1 \times Fem_i \times MP_s + \theta_c \times Fem_i \times I_c + \iota_c \times I_c + X'_{i,r,c,s}\kappa + \varepsilon_{i,r,c,s} \quad \text{(A.4)}$$

This specification still includes class-specific female fixed effects, here denoted $\theta_c$, but foregoes section fixed effects for class fixed effects, $\iota_c$. Identification with this model relies upon an assumption of no sorting, absolute or differential, to sections within a class. This is similar in spirit to the within-gender, class-level analysis presented in the main text.

I present the results of estimating equations (A.3) and (A.4) on pre-college characteristics, and the grades predicted by those characteristics, in Table A.4. The top panel displays the results of estimating equation (A.3) and the bottom the results for equation (A.4). The estimates show that the differences between observable male and female characteristics do not change systematically with section-level male proportion, whether conditioning on section

Table A.4 – Associations Between Section Male Proportion and Pre-College Characteristics

| | Standardized ACT score | High school GPA | Underrepresented minority student | Predicted grade (GPA value) |
|---|---|---|---|---|
| *With section fixed effects:* | | | | |
| Female student X | -0.140 | 0.063 | -0.015 | 0.093 |
| section male % | (0.185) | (0.076) | (0.082) | (0.099) |
| Outcome Mean | 0.000 | 3.282 | 0.517 | 2.783 |
| Outcome SD | 1.000 | 0.385 | 0.500 | 0.452 |
| Observations | 13604 | 13612 | 19106 | 11870 |
| *With class fixed effects:* | | | | |
| Female student X | -0.000 | 0.004 | 0.007 | 0.024 |
| section male % | (0.110) | (0.042) | (0.047) | (0.055) |
| Outcome Mean | 0.000 | 3.282 | 0.517 | 2.783 |
| Outcome SD | 1.000 | 0.385 | 0.500 | 0.452 |
| Observations | 13604 | 13612 | 19106 | 11870 |

Notes: This table reports the results of regressions of pre-college characteristics, at a student-class observation level, on an interaction between a female-student dummy and the section-level male proportion. Each cell corresponds to a separate regression. Top panel regressions include section fixed effects. Bottom panel regressions include class fixed effects. All regressions include class-specific female fixed effects. The predicted grade outcome is based on another (unshown) regression of the GPA value of grades on ACT score, high school GPA, and minority status among students in the main estimation sample. The predicted grade regressions only include the observations used to form the predicted grades: student-class observations from graded classes that had information on both ACT scores and high school GPA. Standard errors are clustered at the class level. * indicates significance at a level of 0.1, ** at a level of 0.05, and *** at a level of 0.01.

fixed effects or not. Importantly, observable differences between students predict no change in grade differential between men and women when going from a wholly female to a wholly male section, again whether conditioning on section fixed effects or not. I interpret this as evidence that the identifying assumptions for both specifications are satisfied in this sample.

Table A.5 presents the estimated effects of section peer composition on student outcomes, across different specifications. Column 1 includes no fixed effects and documents a strong negative association between female academic performance and major choice with section-male percentage, matching the pattern seen in the class-level analysis. Column 4 reports the results of estimating equation (A.3), yielding results that, as previously mentioned, are too imprecise to make meaningful inference from. Column 2 reports results from equation (A.4), which I consider the preferred specification for the section-level analysis. Here, although no results are significant at conventional levels, there is suggestive evidence of a similar pattern to the results of the class-level analysis: worse female academic achievement in the face of more male peers. The strongest evidence of a section peer effect is for grades, with relatively meager evidence of effects on major choice.

I conclude my discussion of the section-level analysis with one interesting note on the distribution of estimated effects of section peer gender. It seems that any negative effect male section peers have on women, relative to men, is concentrated among female-majority classes. Table A.6 reports results on the subset of classes that are female-majority while Table A.7 does the same for classes that are male-majority. Focusing on the second column of Table A.6, the point estimates of the effects of male section peers among female-majority classes are quite similar to the estimated effects of male class peers that are reported in the main body of the paper. The same is not true among male-majority classes, as seen in Table A.7. In particular, coefficients on the differential effects of male peers on grades, major declarations, and graduation in a department are all negative in Table A.6 and much larger in magnitude in Table A.6 than in Table A.7. It would be interesting to see if future work, which may have more data on sections, documents a similar pattern. Considering why the effects are concentrated among certain classes could help shed light on the mechanisms by which peers influence outcomes.

Table A.5 – Estimated Effect of Section Male Proportion on Student Outcomes

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| *Grade (GPA value) [Mean = 2.83]* | | | | |
| Female student X section male % | -0.619*** | -0.206 | -0.196 | -0.083 |
| | (0.126) | (0.178) | (0.134) | (0.215) |
| Observations | 16939 | | | |
| *Take a future course in same department [Mean = .64]* | | | | |
| Female student X section male % | -0.178*** | -0.052 | -0.190*** | 0.002 |
| | (0.062) | (0.059) | (0.048) | (0.076) |
| Observations | 19106 | | | |
| *Switch major to another department [Mean = .15]* | | | | |
| Female student X section male % | 0.412*** | -0.060 | 0.174 | 0.002 |
| | (0.088) | (0.191) | (0.131) | (0.394) |
| Observations | 2052 | | | |
| *Declare a major in same department [Mean = .08]* | | | | |
| Female student X section male % | -0.034 | -0.015 | -0.054* | -0.013 |
| | (0.031) | (0.029) | (0.030) | (0.040) |
| Observations | 16292 | | | |
| *Graduate within six years [Mean = .67]* | | | | |
| Female student X section male % | -0.228*** | -0.048 | -0.012 | 0.074 |
| | (0.041) | (0.067) | (0.054) | (0.086) |
| Observations | 19106 | | | |
| *Graduate with a major in same department [Mean = .14]* | | | | |
| Female student X section male % | -0.498*** | -0.029 | -0.071** | 0.015 |
| | (0.076) | (0.038) | (0.033) | (0.044) |
| Observations | 19106 | | | |
| *Fixed effects* | | | | |
| Class and class-female | No | Yes | No | Yes |
| Section | No | No | Yes | Yes |
| *Controls* | | | | |
| Student gender | Yes | No | Yes | No |
| TA gender | Yes | Yes | No | No |
| Section male % | Yes | Yes | No | No |

Notes: This table reports results of regressions of student-class outcomes on an interaction between a female-student dummy and section male proportion. Outcomes are given by row headers and fixed effects and controls specified by column feet. Standard errors are clustered at the class level. * indicates significance at a level of 0.1, ** at a level of 0.05, and *** at a level of 0.01.

Table A.6 – Estimated Effect of Section Male Proportion on Student Outcomes, Among Female Majority Classes

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| *Grade (GPA value) [Mean = 2.83]* | | | | |
| Female student X section male % | -0.240 | -0.330 | -0.184 | -0.056 |
| | (0.193) | (0.233) | (0.255) | (0.317) |
| Observations | 8236 | | | |
| *Take a future course in same department [Mean = .64]* | | | | |
| Female student X section male % | -0.156 | -0.116 | -0.148 | -0.127 |
| | (0.097) | (0.082) | (0.096) | (0.101) |
| Observations | 9552 | | | |
| *Switch major to another department [Mean = .15]* | | | | |
| Female student X section male % | 0.534*** | 0.416 | 0.016 | 0.399 |
| | (0.172) | (0.308) | (0.156) | (0.637) |
| Observations | 998 | | | |
| *Declare a major in same department [Mean = .08]* | | | | |
| Female student X section male % | -0.028 | -0.034 | -0.112** | -0.054 |
| | (0.049) | (0.043) | (0.051) | (0.057) |
| Observations | 8221 | | | |
| *Graduate within six years [Mean = .67]* | | | | |
| Female student X section male % | -0.130* | 0.086 | 0.058 | 0.182 |
| | (0.073) | (0.088) | (0.092) | (0.113) |
| Observations | 9552 | | | |
| *Graduate with a major in same department [Mean = .14]* | | | | |
| Female student X section male % | -0.280*** | -0.101* | -0.114** | -0.055 |
| | (0.076) | (0.057) | (0.051) | (0.062) |
| Observations | 9552 | | | |
| *Fixed effects* | | | | |
| Class and class-female | No | Yes | No | Yes |
| Section | No | No | Yes | Yes |
| *Controls* | | | | |
| Student gender | Yes | No | Yes | No |
| TA gender | Yes | Yes | No | No |
| Section male % | Yes | Yes | No | No |

Notes: This table replicates Table A.5 among female-majority classes.

Table A.7 – Estimated Effect of Section Male Proportion on Student Outcomes, Among Male Majority Classes

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| *Grade (GPA value) [Mean = 2.83]* | | | | |
| Female student X section male % | -0.374* | -0.010 | -0.099 | -0.069 |
| | (0.192) | (0.264) | (0.209) | (0.284) |
| Observations | 8703 | | | |
| *Take a future course in same department [Mean = .64]* | | | | |
| Female student X section male % | -0.072 | 0.044 | 0.026 | 0.156 |
| | (0.092) | (0.093) | (0.087) | (0.115) |
| Observations | 9554 | | | |
| *Switch major to another department [Mean = .15]* | | | | |
| Female student X section male % | -0.080 | -0.911* | 0.041 | -0.253 |
| | (0.211) | (0.497) | (0.273) | (0.569) |
| Observations | 1054 | | | |
| *Declare a major in same department [Mean = .08]* | | | | |
| Female student X section male % | 0.123*** | 0.006 | 0.055 | 0.033 |
| | (0.046) | (0.039) | (0.049) | (0.054) |
| Observations | 8071 | | | |
| *Graduate within six years [Mean = .67]* | | | | |
| Female student X section male % | -0.296*** | -0.155* | -0.128 | -0.044 |
| | (0.066) | (0.094) | (0.087) | (0.123) |
| Observations | 9554 | | | |
| *Graduate with a major in same department [Mean = .14]* | | | | |
| Female student X section male % | -0.280*** | 0.046 | 0.003 | 0.091 |
| | (0.097) | (0.046) | (0.057) | (0.060) |
| Observations | 9554 | | | |
| *Fixed effects* | | | | |
| Class and class-female | No | Yes | No | Yes |
| Section | No | No | Yes | Yes |
| *Controls* | | | | |
| Student gender | Yes | No | Yes | No |
| TA gender | Yes | Yes | No | No |
| Section male % | Yes | Yes | No | No |

Notes: This table replicates Table A.5 among male-majority classes.

## A.6 Within-Gender Peer Effects, When Controlling for Peer Ability

Table A.8 – Estimated Effect of Class Male Proportion on Student Outcomes, By Gender And Controlling for Average Peer Ability

| | Grade (GPA value) | Future course in dept. | Switch major out of dept. | Declare major in dept. | Graduate within 6 years | Graduate in dept. |
|---|---|---|---|---|---|---|
| *Only Women:* | | | | | | |
| Class male % | -0.235* | -0.078** | -0.146 | -0.021 | -0.111*** | -0.009 |
| | (0.130) | (0.033) | (0.217) | (0.017) | (0.041) | (0.019) |
| Outcome Mean | 3.068 | 0.663 | 0.141 | 0.060 | 0.707 | 0.135 |
| Outcome SD | 1.051 | 0.473 | 0.348 | 0.237 | 0.455 | 0.341 |
| Observations | 12180 | 16393 | 618 | 15451 | 16393 | 16393 |
| *Only Men:* | | | | | | |
| Class male % | -0.103 | -0.002 | -0.234 | 0.031** | 0.014 | 0.055*** |
| | (0.159) | (0.041) | (0.337) | (0.016) | (0.047) | (0.018) |
| Outcome Mean | 2.901 | 0.625 | 0.135 | 0.049 | 0.647 | 0.119 |
| Outcome SD | 1.143 | 0.484 | 0.341 | 0.216 | 0.478 | 0.324 |
| Observations | 10881 | 15309 | 615 | 14311 | 15309 | 15309 |

Notes: This table reports the results of regressions of student-class outcomes on class-level male proportion. Each cell corresponds to a separate regression, with outcome given by the column header. Top row results are estimated only on female students and bottom row results only on male students. Each regression includes course fixed effects and controls for student underrepresented minority status, instructor gender, and average peer HS GPA. All regressions are estimated on the observations within the identifying set of the main specification (as described in Section 1.2) for which the relevant outcome is observed. Standard errors are clustered at the class level. * indicates significance at a level of 0.1, ** at a level of 0.05, and *** at a level of 0.01.

## A.7 Heterogeneity of Effects Between Female- and Male-Majority Departments

Table A.9 – Estimated Effect of Class Male Proportion on Student Outcomes, By Gender and Department-Level Male Proportion

| | Grade (GPA value) | Future course in dept. | Switch major out of dept. | Declare major in dept. | Graduate within 6 years | Graduate in dept. |
|---|---|---|---|---|---|---|
| *Women in female-majority-department classes:* | | | | | | |
| Class male % | -0.067 | -0.044 | -0.016 | -0.016 | -0.037 | 0.002 |
| | (0.117) | (0.030) | (0.073) | (0.016) | (0.040) | (0.019) |
| Outcome Mean | 3.148 | 0.674 | 0.141 | 0.058 | 0.710 | 0.146 |
| Outcome SD | 1.016 | 0.469 | 0.348 | 0.233 | 0.454 | 0.353 |
| Observations | 12266 | 15203 | 2016 | 12704 | 15203 | 15203 |
| *Men in female-majority-department classes:* | | | | | | |
| Class male % | -0.054 | -0.063 | 0.036 | 0.009 | -0.010 | 0.008 |
| | (0.159) | (0.040) | (0.125) | (0.013) | (0.050) | (0.017) |
| Outcome Mean | 2.950 | 0.604 | 0.171 | 0.039 | 0.628 | 0.099 |
| Outcome SD | 1.131 | 0.489 | 0.377 | 0.194 | 0.483 | 0.299 |
| Observations | 8503 | 10757 | 1006 | 9417 | 10757 | 10757 |
| *Women in male-majority-department classes:* | | | | | | |
| Class male % | -0.501** | -0.160** | -0.329*** | -0.039 | -0.176*** | -0.029 |
| | (0.233) | (0.072) | (0.125) | (0.042) | (0.066) | (0.045) |
| Outcome Mean | 2.873 | 0.637 | 0.138 | 0.063 | 0.698 | 0.111 |
| Outcome SD | 1.109 | 0.481 | 0.345 | 0.243 | 0.459 | 0.314 |
| Observations | 5018 | 6959 | 521 | 6273 | 6959 | 6959 |
| *Men in male-majority-department classes:* | | | | | | |
| Class male % | -0.038 | -0.015 | -0.122 | 0.076** | 0.087 | 0.105** |
| | (0.223) | (0.075) | (0.147) | (0.037) | (0.063) | (0.043) |
| Outcome Mean | 2.845 | 0.647 | 0.104 | 0.060 | 0.666 | 0.140 |
| Outcome SD | 1.154 | 0.478 | 0.306 | 0.237 | 0.472 | 0.347 |
| Observations | 7284 | 10432 | 1207 | 8759 | 10432 | 10432 |

Notes: This table reports the results of regressions of student-class outcomes on class-level male proportion. Each cell corresponds to a separate regression, with outcome given by the column header and the subset of students used for estimation given by the row header. Each regression includes course fixed effects and controls for student underrepresented minority status and instructor gender. All regressions are estimated on the observations within the identifying set of the main specification (as described in Section 1.2) for which the relevant outcome is observed. Standard errors are clustered at the class level. * indicates significance at a level of 0.1, ** at a level of 0.05, and *** at a level of 0.01.

# APPENDIX B

# CHAPTER 2: SUPPLEMENTARY MATERIAL

## B.1 Derivation of Equilibrium

### *B.1.1 Profit-Seeking Equilibria*

Consider equilibria with $r'(e^*) = \beta$. For any $j$ such that $\theta_j \in [0, \beta]$, there will be equilibrium price $p_j^* = c$. If equilibrium price was any higher, any school would deviate to a slightly lower price, thereby capturing that entire student population and gaining utility both from increased enrollment and from money to be kept as profit. At any lower price, any school would deviate to a higher price, enroll no students of type $j$, and would gain by keeping the money that subsidized the $j$-type attendance as profit.

Consider $k$ such that $\theta_k < 0$. $p_k^* > c$, as otherwise any school would deviate to a higher price. If a school follows the equilibrium price, and students are willing to pay it, they receive:

$$U_i = r(\bar{e}) + \frac{1}{m}\theta_k + \frac{1}{m}\sum_{j \neq k}\theta_j + \beta(d - \bar{e} + \frac{1}{m}(p_k^* - c) + \frac{1}{m}\sum_{j \neq k}(p_j^* - c))$$

If a school deviates up in price, they will not enroll any students of type $k$ and receive:

$$U_i = r(\bar{e}) + \frac{1}{m}\sum_{j \neq k}\theta_j + \beta(d - \bar{e} + \frac{1}{m}\sum_{j \neq k}(p_j^* - c))$$

Schools will choose prices to maximize the enrollment of students when indifferent, so schools will not want to deviate up from $p_k^*$ if:

$$\begin{aligned} \frac{1}{m}\theta_k + \frac{\beta}{m}(p_k^* - c) \geq 0 \\ \Rightarrow p_k^* \geq c - \frac{\theta_k}{\beta} \end{aligned}$$

If the school deviates down in price, they receive:

$$U_i < r(\bar{e}) + \theta_k + \frac{1}{m}\sum_{j \neq k} \theta_j + \beta(d - \bar{e} + p_k^* - c + \frac{1}{m}\sum_{j \neq k}(p_j^* - c))$$

Schools will not deviate down in price if:

$$\begin{aligned} \frac{m-1}{m}\theta_k + \beta\frac{m-1}{m}(p_k^* - c) \leq 0 \\ \Rightarrow p_k^* \leq c - \frac{\theta_k}{\beta} \end{aligned}$$

Combining the two conditions implies that $p_k^* = c - \frac{\theta_k}{\beta}$ is the sustainable equilibrium price for any negative value students. If there are any types such that $\beta(c - v) > \theta_k$, schools are only willing to enroll students of these types at a price that is above the students' willingness to pay, so schools enroll none of these students. We can think of schools setting the price to $\infty$ for these types, to create an analogue of not admitting sufficiently undesirable students.

Now consider types where $\theta_l \geq \beta$ and further assume that, for any one type, $p_l^*$ is sufficiently low that enrolling all students of that type at the equilibrium price will not induce the school to set $f_i = 0$ and $e_i < \bar{e}$. Then, by analogous arguments to the above, the school will not deviate up in price if:

$$p_l^* \geq c - \frac{\theta_l}{\beta}$$

and will not deviate down if:

$$p_l^* \leq c - \frac{\theta_l}{\beta}$$

implying the same pricing rule as for negative-value students: $p_l^* = c - \frac{\theta_k}{\beta}$ This will apply to all students for $\beta \leq \theta_k \leq \beta c$. For any possible types above this upper bound, schools will run into the non-negativity constraint on the price, and set price to 0.

Ensuring that this an equilibrium now requires checking that the budget constraint, with $e_i = \bar{e} \ \forall \ i$ works and that the assumption of no positive-value types having a price that is "too high" for the assumption from the preceding paragraph holds. For the sake of the budget constraint, define the sets of types $\tilde{J} = \{j | \beta(c - v) < \theta_j \leq 0 \cup \beta < \theta_j \leq \beta c\}$ and $\overline{J} = \{j | \theta_j > \beta c\}$. Then, in order for the equilibrium to exist such that $e^* = \bar{e}$, it must be the case that:

$$\begin{aligned} d + \frac{1}{m} \sum_{\tilde{J}} (p_j^* - c) - \frac{1}{m} \sum_{\overline{J}} c \geq \bar{e} \\ - \sum_{\tilde{J}} \frac{\theta_j}{\beta} - \sum_{\overline{J}} c \geq m(\bar{e} - d) \end{aligned}$$

which is satisfied by assumption.

Finally, putting a maximum on the highest type maintains the earlier assumption that enrolling all students of any one type will not push the school to the point of not being able to afford $\bar{e}$. Specifically, it must be the case that, for $\overline{k} = \{k | \theta_k = \sup \Theta\}$

$$d + \frac{1}{m} \sum_{j \neq \overline{k}} (p_j^* - c) + p^*_{\overline{k} - c \geq \bar{e}}$$

This is satisfied by either of the following two conditions:

$$\theta_{\overline{k}} \leq \beta(d - \overline{e}) - \frac{1}{m}\sum_{j \neq \overline{k}} \theta_j$$
$$\beta c \leq \beta(d - \overline{e}) - \frac{1}{m}\sum_{j \neq \overline{k}} \theta_j$$

with the latter indicating that the non-negativity of price precludes a problematically high price for any type. Note that this condition is not a necessary condition for equilibrium, but merely a condition that greatly simplifies the characterization of equilibria. It is likely the case that higher types can exist, with suitable adjustments to the pricing regime.

### *B.1.2 Research-Seeking Equilibrium*

Consider equilibria for which $r'(e^*) = \beta$, with two relevant types as $\theta_L < 0 < \theta_H$.

- Assume that at least one of $|\theta_H| \geq |\theta_L|$ or $r'(\overline{e}) > \frac{-1}{m}\frac{\theta_H + \theta_L}{\overline{e} - d}$ is true
- Assume $\overline{e} > d$ and that:

$$r'(e^*) \geq \max\{\frac{\theta_H}{d}, \frac{-\theta_L}{v - c}, \frac{\theta_H}{c}, \frac{\frac{1}{m}\theta_H - \frac{m-1}{m}\theta_L}{\overline{e} - e^*}\}$$

A difference for consideration of this equilibrium, compared to the preceding variety, is that deviation in the price charged to one type might induce further deviation in the price charged to the other type. For $e^* = \overline{e}$, given the assumption about the upper limit on the quality of the highest type, price for any one type was based on comparison to the fixed alternative of spending money on profit, which has a constant marginal return. Behavior regarding each type could therefore be considered separately. With $e^* < \overline{e}$, any increase or decrease in money generated/lost from sales to students will change the marginal rate of return to research, as research is a concave function of expenditures. Thus, deviation on either price in either direction will result in multiple potentially desirable outcomes.

I make multiple suppositions here regarding the nature of the equilibrium prices, which will simplify the consideration of returns to deviation. First, presume that $p_H^* \geq 0$ and $p_L^* \leq v$. Similar to the assumption of a maximum type for the $e^* = \bar{e}$ equilibria, I now assume that prices are such that, at any possible enrollment strategy, schools will never have enough money that:

$$d + \sum_{j \in \{H,L\}} q_{i,j}(p_j^* - c) > \bar{e}$$

such that schools will always set profit to 0 in any scenario considered below. I further assume that $p_H^*$ is such that enrolling all of the high-type students and none of the low-types will not push the school into negative remaining budget:

$$d + p_H^* - c > 0$$

These assumptions are made for convenience, and it is likely the case that an equilibrium could be found without them with suitable adjustments to the pricing regime.

In equilibrium schools receive:

$$U_i = (d + \frac{1}{m} \sum_{j \in \{H,L\}} (p_j^* - c)) + \frac{1}{m}(\theta_H + \theta_L)$$

For $p_H^*$, deviation up will yield utility:

$$U_i = (d + \frac{1}{m}(p_L^* - c)) + \frac{1}{m}\theta_L$$

As $p_H^*$ must be below $c$, this will reduce the marginal value of research expenditures. Because $p_L^*$ must be above $c$, this could hypothetically make it attractive to deviate further to increase the price charged to the low-type, and enroll no students:

$$U_i = r(d)$$

Thus, avoiding deviations upward on high-type price necessitates two conditions. For simplicity in defining these conditions, I denote $\hat{p}_H^* \equiv p_H^* - c$ and $\hat{p}_L^* \equiv p_L^* - c$:

$$\frac{1}{m}\theta_H \geq r(e^* - \frac{1}{m}\hat{p}_H^*) - r(e^*) \tag{B.1}$$

$$\frac{1}{m}(\theta_H + \theta_L) \geq r(e^* - \frac{1}{m}(\hat{p}_H^* + \hat{p}_L^*)) - r(e^*) \tag{B.2}$$

By similar reasoning, avoiding deviation down from $p_H^*$ necessitates that neither capturing all high-types nor capturing all high- *and* low-types is preferred to equilibrium. This yields two more conditions:

$$r(e^*) - r(e^* + \frac{m-1}{m}\hat{p}_H^*) > \frac{m-1}{m}\theta_H \tag{B.3}$$

$$r(e^*) - r(e^* + \frac{m-1}{m}(\hat{p}_H^* + \hat{p}_L^*)) > \frac{m-1}{m}(\theta_H + \theta_L) \tag{B.4}$$

Deviating up on $p_L^*$ can entail foregoing enrollment of any low types or further deviating to forego all enrollment. Deviation is undesirable if condition (6) holds as well as

$$r(e^*) - r(e^* - \frac{1}{m}\hat{p}_L^*) \geq \frac{-1}{m}\theta_L \tag{B.5}$$

Finally, deviating down on $p_L^*$ will be avoided if schools prefer equilibrium to both capturing all low-types and capturing all low- and high-types. This occurs if condition (8) holds along with:

$$\frac{m-1}{m}\theta_L > r(e^* + \frac{m-1}{m}\hat{p}_L^*) - r(e^*) \tag{B.6}$$

This leaves a total of 6 conditions to be satisfied for equilibrium to hold, (5)-(10).

Given the assumptions made, there will exist at least one set of prices that satisfies all 6 conditions. Set $\hat{p}_H^* = \frac{-\theta_H}{r'(e^*)}$ and $\hat{p}_L^* = \frac{-\theta_L}{r'(e^*)}$, assuming for now that such a point is possible (given that $e^*$ depends on the prices). Increasing research expenditures from $e^*$ to $e^* + \frac{1}{m}\frac{\theta_H}{r'(e^*)}$ will increase research utility by less than $\frac{1}{m}\theta_H$ given that $r(\cdot)$ is continuous and strictly concave. This satisfies (5). Similarly, decreasing research expenditures from $e^*$ to $e^* - \frac{m-1}{m}\frac{\theta_H}{r'(e^*)}$ will decrease research utility by more than $\frac{m-1}{m}\theta_H$, satisfying (7). Appealing to the continuity and strict concavity of $r(\cdot)$ will similarly satisfy (9) and (10) at the price given above. Conditions (6) and (8) will be satisfied trivially if $|\theta_H| = |\theta_L|$ at the given prices. If not, both are also satisfied due to the strict convexity of the function.

It remains to be shown that prices can be set to the given levels for some value of $e^*$. This would imply:

$$\begin{aligned}
e^* &= d + \frac{1}{m}(\hat{p}_H^* + \hat{p}_L^*)\\
e^* &= d + \frac{-1}{m}(\frac{\theta_H}{r'(e^*)} + \frac{\theta_L}{r'(e^*)})\\
e^* - d &= \frac{-1}{m}\frac{1}{r'(e^*)}(\theta_H + \theta_L)
\end{aligned}$$

If $e^* = d$, that implies that $|\theta_H| = |\theta_L|$, and the above is satisfied. If not, the above can be represented

$$r'(e^*) = \frac{-1}{m}\frac{\theta_H + \theta_L}{e^* - d} \tag{B.7}$$

I first note that the signs of the above will be consistent, as $r'(e^*) > 0$ by assumption while

$e^* - d$ and $\theta_H + \theta_L$ must have opposite signs given the pricing schedule (if non-zero). If $e^* < d$, then, on the interval $e \in [0, d]$, $r'(e^*)$ is a continuous, monotonically decreasing function that goes from $\infty$ to some positive number while the RHS of (9) is a continuous, monotonically increasing function that goes from some positive number to $\infty$. Thus, the two will cross at exactly one point, which represents the equilibrium value of $e^*$. If $e^* > d$, then, on the interval $e \in [d, \bar{e}]$ the RHS is a continuous, monotonically decreasing function that goes from $\infty$ to $\frac{-1}{m}\frac{\theta_H+\theta_L}{\bar{e}-d}$ while $r'(e^*)$ is a continuous, monotonically decreasing function that goes from some $a$ to $b$, such that $a > b > \frac{-1}{m}\frac{\theta_H+\theta_L}{\bar{e}-d} > 0$ (with the penultimate inequality following from assumption), ensuring that the two functions cross exactly once, giving the value of $e^*$.

Finally it remains to check the conditions that were placed to simplify the relationships between types and prices. The assumptions of $p_H^* \geq 0$ and $p_L^* \leq v$ restrict attention to types that avoid the constraints on prices. These assumptions are satisfied by:

$$r'(e^*) \geq \frac{\theta_H}{c}$$
$$r'(e^*) \geq \frac{-\theta_L}{v - c}$$

I also assumed that maximizing revenue by enrolling all low-types and no high-types would not allow for research spending greater than $\bar{e}$, while minimizing revenue by enrolling all high-types and no low-types would not create negative income. These conditions are satisfied by:

$$r'(e^*) \geq \frac{-\theta_L}{\bar{e} - d}$$
$$r'(e^*) \geq \frac{\theta_H}{d}$$

All four of these conditions are satisfied by assumption.

## B.2 Synthetic Control Difference Plots for Scholarship Spending

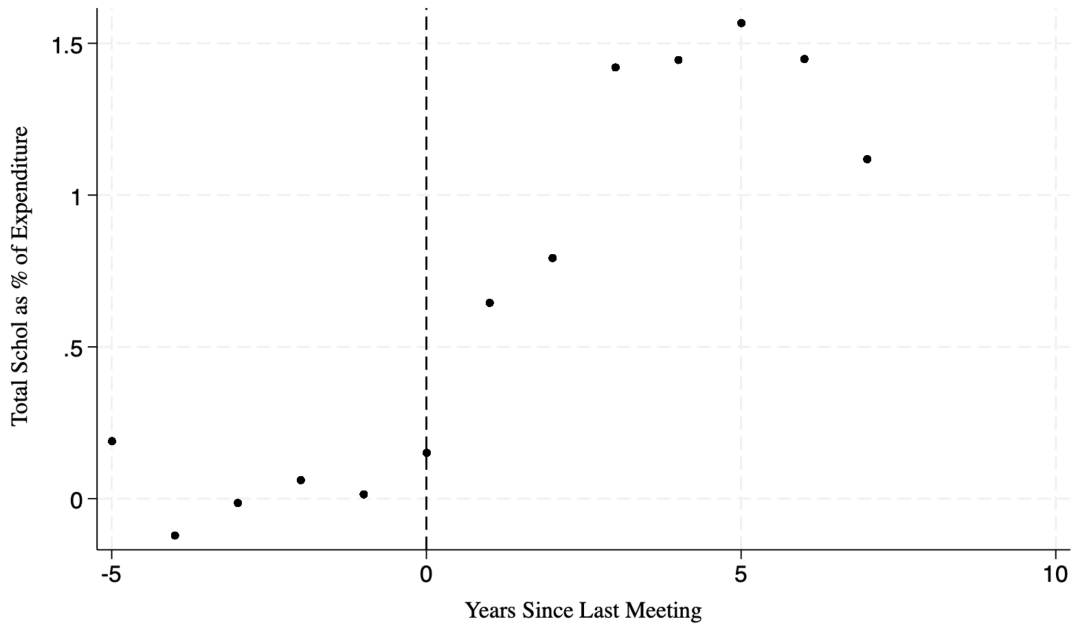



Figure B.1: Differences Between Overlap Schools and Synthetic Controls - Total Scholarship Spending

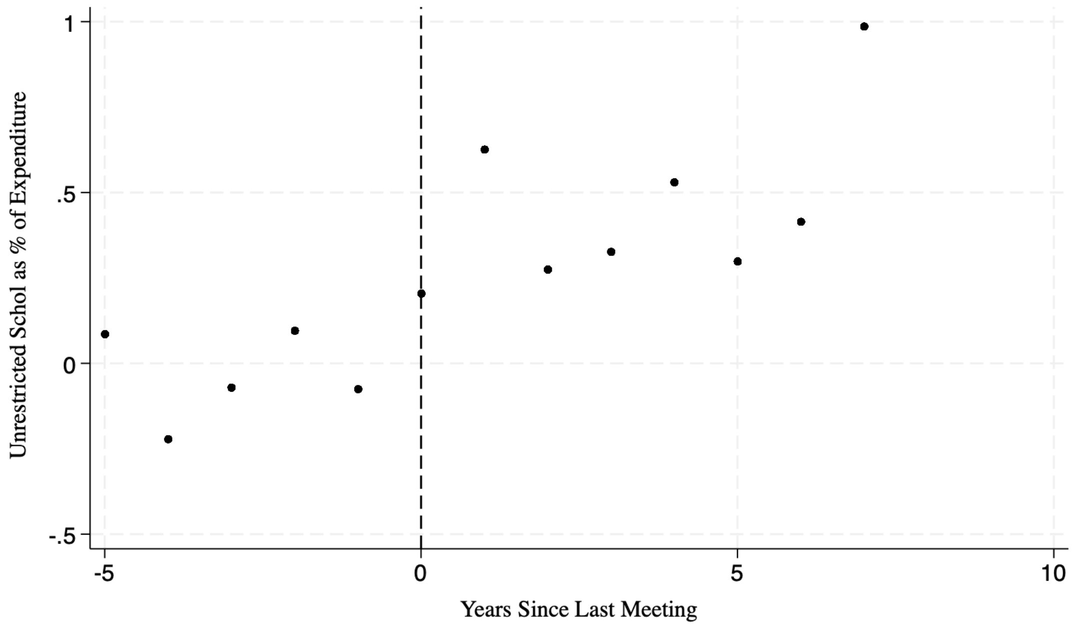



Figure B.2: Differences Between Overlap Schools and Synthetic Controls - Unrestricted Scholarship Spending

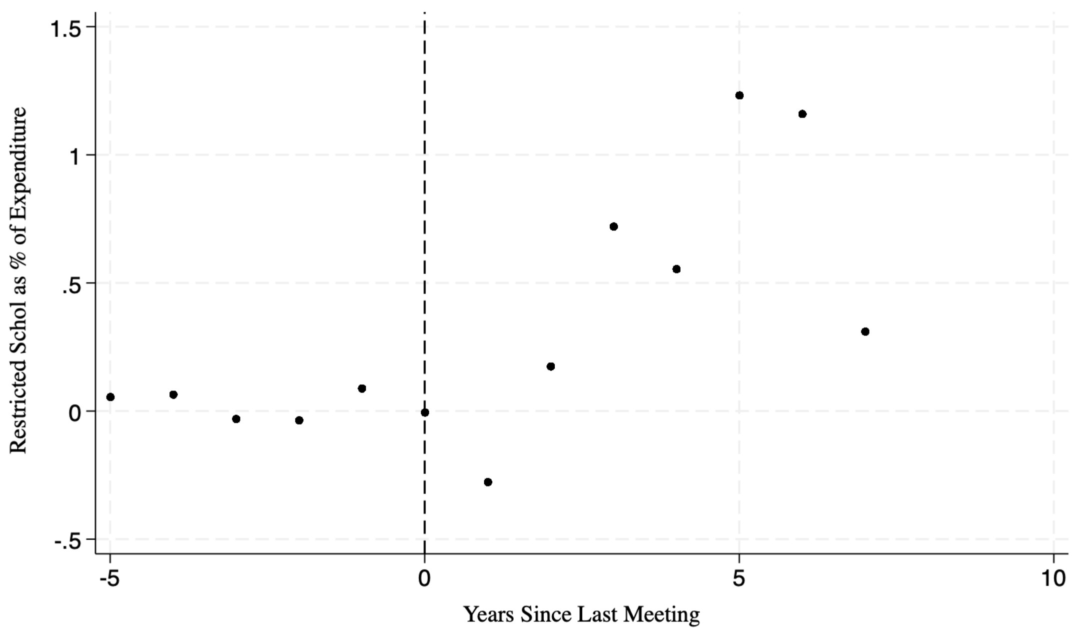



Figure B.3: Differences Between Overlap Schools and Synthetic Controls - Restricted Scholarship Spending

## B.3 DID and Synthetic Controls Plots for Other Expenditures

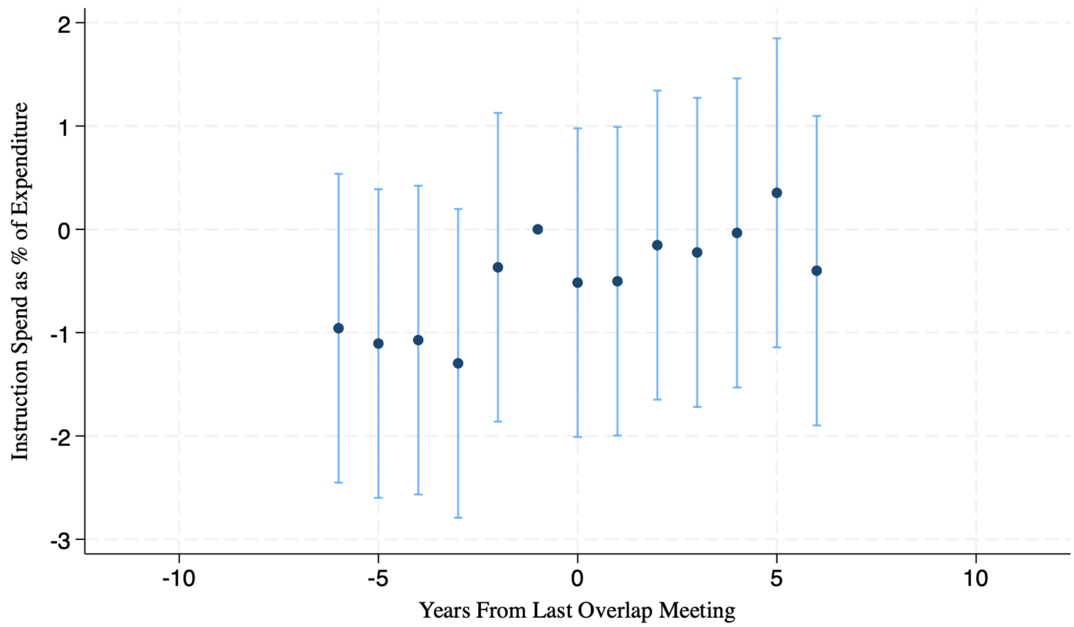



Figure B.4: DID Between Overlap Schools and Controls - Instruction Spending

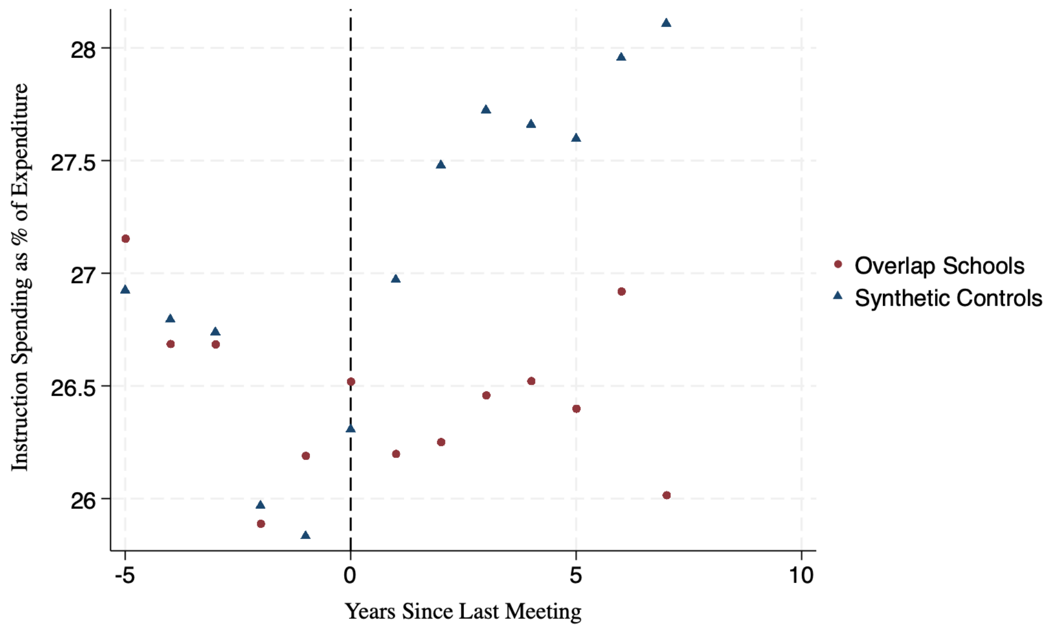



Figure B.5: Overlap Schools Compared to Synthetic Controls - Instruction Spending

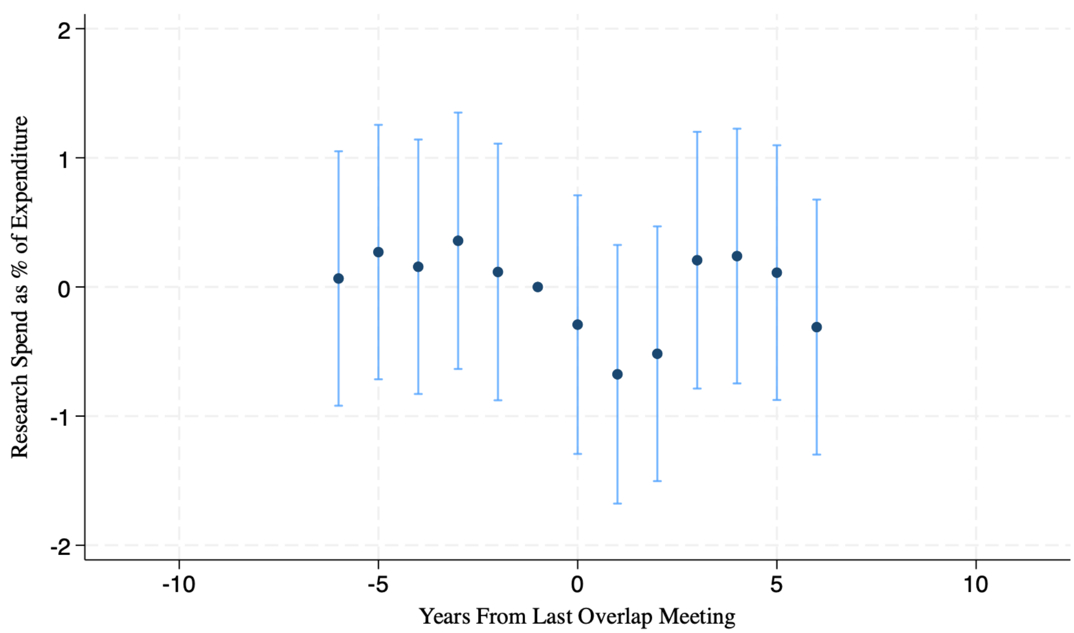



Figure B.6: DID Between Overlap Schools and Controls - Research Spending

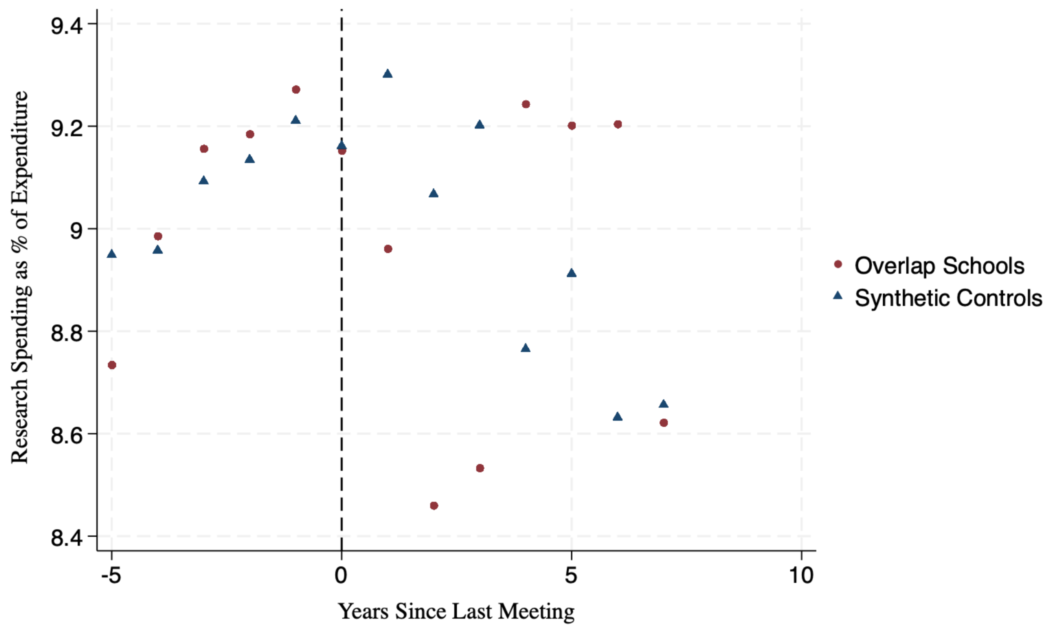



Figure B.7: Overlap Schools Compared to Synthetic Controls - Research Spending

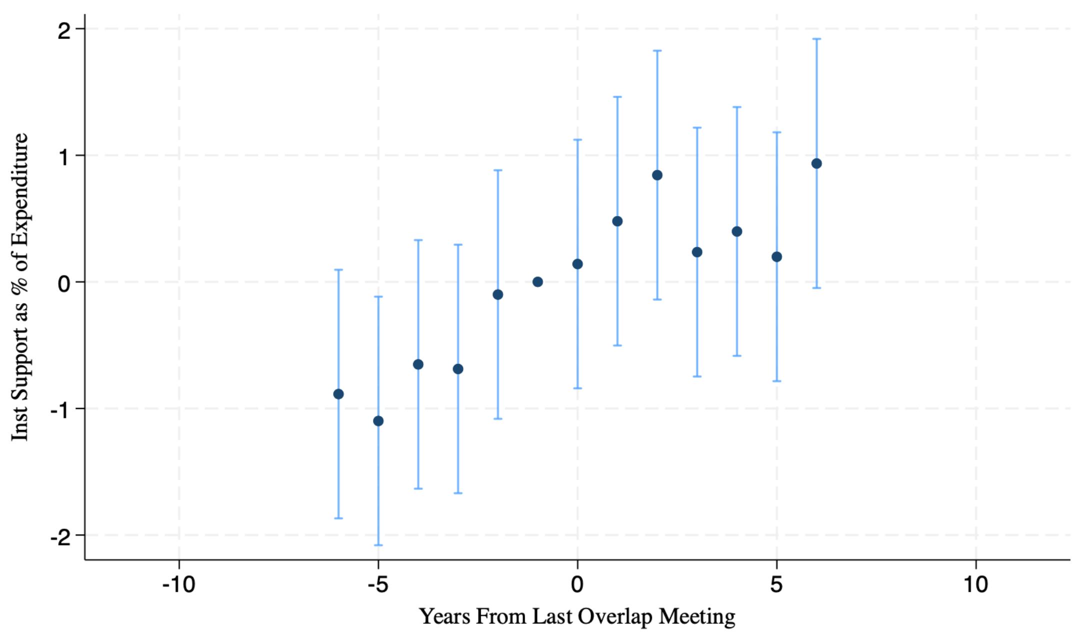



Figure B.8: DID Between Overlap Schools and Controls - Institutional Support

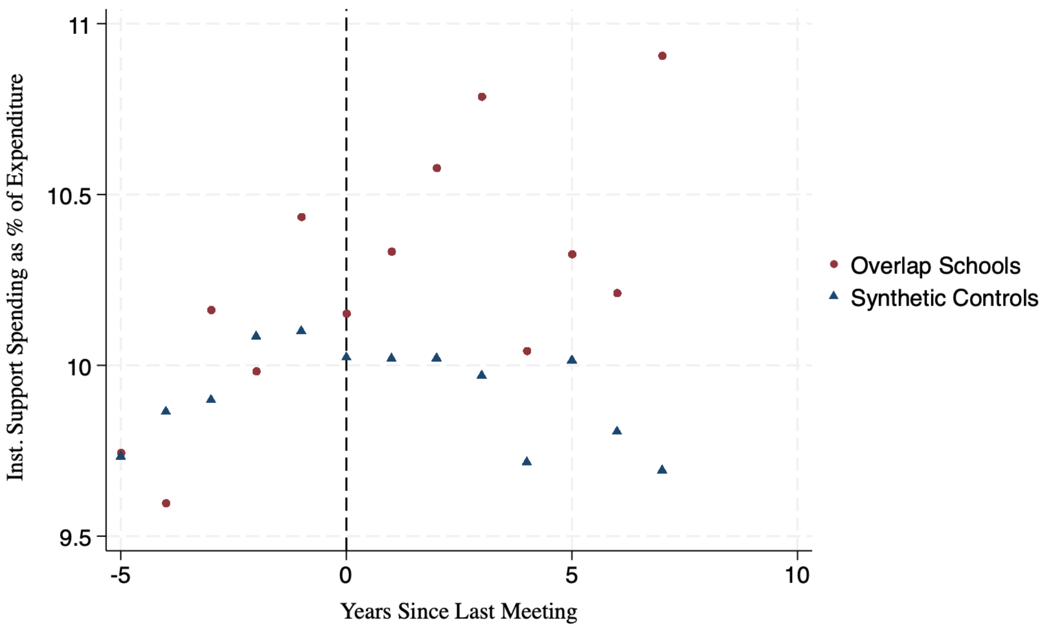



Figure B.9: Overlap Schools Compared to Synthetic Controls - Institutional Support

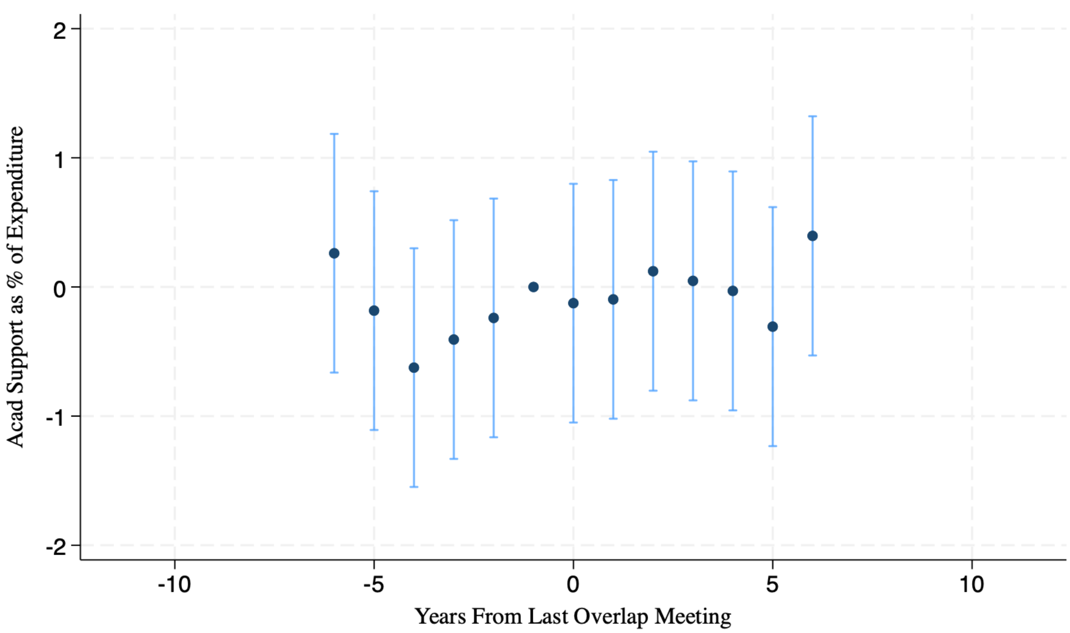



Figure B.10: DID Between Overlap Schools and Controls - Academic Support

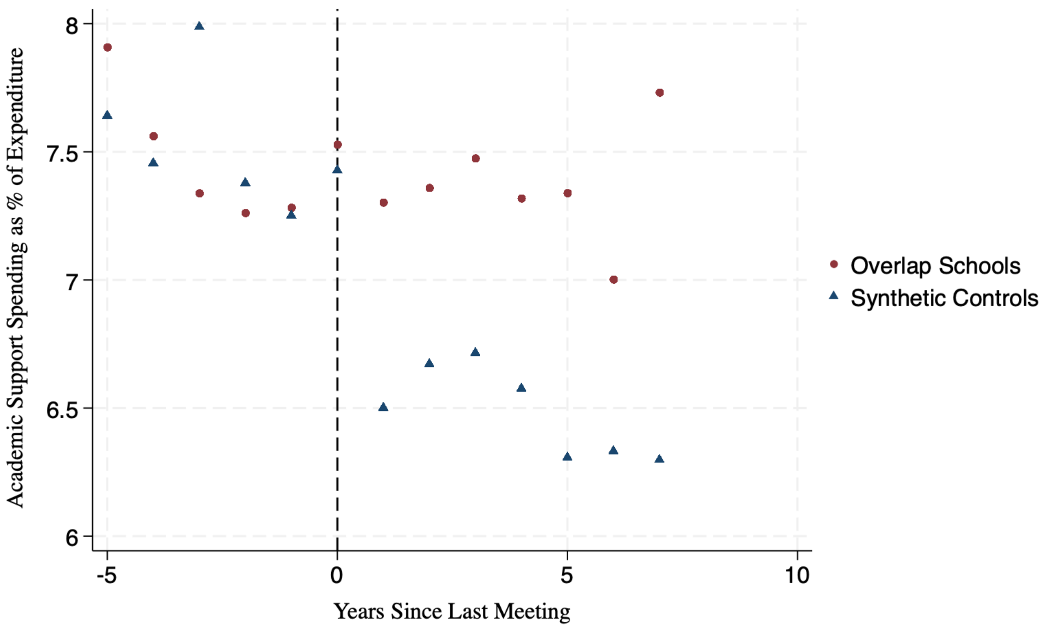



Figure B.11: Overlap Schools Compared to Synthetic Controls - Academic Support