Title
Essays in Applied Microeconomics

Author
Rury, Derek

Publication Date
2021

Peer reviewed|Thesis/dissertation

Essays in Applied Microeconomics

By

DEREK RURY
DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

Doctor of Philosophy

in

Economics

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Scott Carrell, Chair

Giovanni Peri

Paco Martorell

Committee in Charge

2021

Essays in Applied Microeconomics

Derek Rury (UC Davis)

May 2021

Abstract

In this dissertation, I discuss three papers. The first two, "Fixing the Leaky Pipeline: The Role of Beliefs About Ability in STEM Major Choice" and "Knowing What it Takes: The Effect of Information About Returns to Studying on Study Effort and Achievement", use field experiments to influence student beliefs to study student decision making. These papers demonstrate both the need to focus on students' mental models of decision characteristics when creating education policy, as well as the potential for thoughtful information interventions to influence student behavior. The third paper, "The Economic Impacts of Hurricane Maria", studies how a sudden influx of workers into a labor market influences the local economy. We find that these new workers grow the economy overall, with heterogeneous effects by sector. These three papers each show how rigorous research designs and econometric techniques can help us study complex social phenomenon.

Fixing the Leaky Pipeline: The Role of Beliefs About Ability in STEM Major Choice

Derek Rury[1]

Abstract

Empirical evidence shows that STEM students are more likely to switch to a different major or drop out compared to students of other fields, creating a ”leaky pipeline” out of STEM. Previous research shows that this may be due to students beginning college with incorrect beliefs about their ability to complete a STEM degree. I explore this via a field experiment where I provide students with information that they are above average in their top fields of study. I then measure the effect of this information on persistence in those fields, with a focus on STEM students. I find that STEM students are indeed more likely to switch out of their major compared to students of other fields. I also find evidence that the information I provide increases persistence in STEM fields. Information also increases STEM major choice across all students by 12% of a standard deviation, as non-STEM students increase switching into STEM as a result of learning they are above average in that field. These effects are all largest for first-generation students. Lastly, I find suggestive evidence for the role of confirmation bias in major choice, as treatment effects appear larger for students who had higher beliefs about their ability before the experiment, particularly STEM students.

[1]Address: Department of Economics, University of California Davis, One Shields Avenue, email:drury@ucdavis.edu; I would like to thank my advisors Scott Carrell, Paco Martorell, Giovanni Peri and Michal Kurlaender. I also like to thank Geoffrey Schnorr, Diana Moreira, Marianne Bitler, David Rapson, Marianne Page and other participants of the UC Davis applied micro brown bag series for helpful comments. I also want to thank panelist at the 2020 Western Economic Association annual conference, 2019 American Education Finance and Policy conference, participants at the 2019 All-California Labor Economics Conference, the 2020 Bogota Experimental Eonomics Conference as well as attendees at the 2019 International Workshop on Applied Economics of Education. I would also like to thank the Center for Educational Effectiveness at UC Davis for all of their support with this project. This project was pre-registered using the AEA RCT registry under the title ”Beliefs, Ability and College Major Success” and RCT ID: AEARCTR-0003432

1 Introduction

Students make numerous consequential decisions throughout their academic careers. These include whether to go to college, what institution to attend, how much to study and which major to choose. A large literature has demonstrated that these decisions can have profound causal effects on their later life outcomes such as income, family formation, social mobility and mortality (Kirkeboen et al., 2016; Zimmerman, 2014; Chiappori et al., 2017; Chetty et al., 2020; Zimmerman, 2019; Buckles et al., 2016).

A smaller but growing literature has shown that students often make these decisions with incomplete information, potentially leading to sub-optimal outcomes (Manski, 1993; Betts, 1996; Conlon, 2020; Wiswall and Zafar, 2015a). These studies also often find that students change their beliefs and behavior after receiving information about these decisions, providing scope for information interventions or "nudges" to correct for information deficiencies (Wiswall and Zafar, 2015b; Baker et al., 2018; Conlon, 2020; Rury and Carrell, 2020).

This information problem may be particularly acute during a student's major choice. This is likely true as most colleges in the U.S. offer students a great deal of freedom over when they select their major. Also, by virtue of the fact that education happens earlier in life, students rarely make this decision with much relevant life experience (eg. time spent in the labor force). Research has indeed demonstrated that students exhibit large errors in their beliefs about different characteristics of majors (Betts, 1996; Conlon, 2020; Wiswall and Zafar, 2018; Baker et al., 2018). Most work in this literature has focused on beliefs about the pecuniary returns to majors, although recent work has incorporated beliefs about non-pecuniary returns (Wiswall and Zafar, 2018; Carrell et al., 2020). Receiving much less attention, but certainly key to a student's college major choice, are beliefs about one's ability to complete a field of study (Zafar, 2011; Wiswall and Zafar, 2015a; Stinebrickner and Stinebrickner, 2013).

Despite their importance, we still do not know whether students' beliefs about their ability in different academic fields are accurate. Initial work on this question has focused on providing students with relative performance feedback based on classroom assessments and measuring the impact of this feedback on measures of persistence in that major, such as course-taking (Li, 2018; Owen, 2020). Beliefs about ability may be particularly important in science, engineering, technology and mathematics (STEM) fields, as descriptive evidence shows that STEM students are more likely to change majors than any other field (Chen and Soldner, 2014). This phenomenon is commonly referred to as the "leaky pipeline" out of STEM and has received a great deal of attention from the media, scholars and policymakers. Stinebrickner and Stinebrickner (2003) find that students enter college with incorrect beliefs about their ability to complete a science degree, causing more students to exit science, either for a different major or by dropping out entirely. This leaky pipeline out of STEM is especially important as the demand for STEM-capable workers increases (Olson and Riordan, 2012), with broad implications for the competitiveness of the domestic workforce. These facts raise important questions about whether students have correct beliefs about their ability to complete different fields of study, particularly STEM fields, and if not, whether providing students with information about their ability can increase persistence in those fields.

I study these questions using a field experiment in which I first elicit college students' top choices of major and then randomly provide them with information they are above average in those fields. In my descriptive analyses, I verify that students are indeed more likely to leave their preferred field of study if that field is a STEM field when compared to students whose top field is non-STEM. Specifically, STEM students are four times as likely to leave their top choice of major compared to non-STEM students. I also show that this is not due to STEM students placing less likelihood on graduating in that field compared to non-STEM students. STEM students believe they are just as likely to graduate in their field as non-STEM students. Focusing on beliefs about ability, I also find that over one in four STEM students who are actually above average in their top field believe they are "relatively bad" in that field. Focusing on those students who switch out of STEM, they are more likely to be female, believe that their peers are of higher ability relative to their own ability as well as overestimate the actual ability of their peers.

In my experimental results, I find that the information I provide to students increases persistence in STEM fields by over 20% of a standard deviation, although these results are imprecise. I do find, however, a significant differential treatment effect on STEM students when compared to non-STEM students. I find evidence that this may be due to a small negative effect on persistence for non-STEM students, who are then induced to switch *into* a STEM field as a result of the intervention. Combining these two effects, I find that my information intervention increased STEM major choice by a statistically significant 12% of a standard deviation across the entire sample.

These effects are all largest for first-generation students. I also find suggestive evidence that female STEM students were more likely to persist in their top major as a result of treatment, although these results are less precisely estimated. Lastly, I also study the existence of confirmation bias in major choice as students process new information about their ability. I do this by using survey responses to identify students as believing they are "above average" in their top fields of study. I find evidence that providing information to students that they are in fact above average in their top fields has a stronger relative effect on students who already believe they are above average. This effect is stronger for STEM students and is primarily driven by confident students, those for whom the information is least informative, increasing persistence in response to information. I argue that these results underscore the potential for behavioral biases to influence student decisions and outcomes and should motivate scholars of education to consider the role of these factors in their models of student decision making.

This paper contributes to several literatures. The first is the literature studying the determinants of college major choice. Most papers in this literature model a student's major choice as the result of a learning process that resolves uncertainty about various characteristics about different majors, including expected earnings, productivity and academic ability (Arcidiacano et al., 2012; Arcidiacano, 2004; Gong et al., 2019; Zafar, 2011). These papers employ structural modeling techniques to untangle the magnitudes of these different determinants of students' college major decisions. They find that while students often have large errors in their beliefs about the pecuniary returns to college majors, the elasticity of college major choice with respect to expected earnings is quite low (Wiswall and Zafar, 2015a). Most studies in this literature also highlight the role of perceptions of ability as an important determinant in the major choice decision. Another branch in this literature studies the major choice decision by providing information to students about characteristics of majors and measuring changes in students' major choice behavior (Conlon, 2020; Li, 2018; Carrell et al., 2020). Other work examines the effect of role models in major choice (Carrell et al., 2010; Porter and Serra, 2019) Looking at initial beliefs about ability to complete different fields, Stinebrickner and Stinebrickner (2013) find that students leave science fields because they enter college with incorrect beliefs about their ability. Similarly, Zafar (2011) finds that students hold biasesd beleifs about their ability in their major, but that they update their beliefs as the progress through college. Motivated by these findings, my paper tests the following thought experiment; how would students' major choices change if they were better informed about their relative ability? Through my experimental design, I am able to answer this question for students of different majors, demographic characteristics and beliefs, building on the work of previous structural work on the topic.

Secondly, this paper contributes to the study of beliefs in education (Bobba and Frasinco, 2019a,b; Zafar, 2011; Bleemer and Zafar, 2018; Dizon-Ross, 2019; Kaufman, 2014; Attanasio and Kaufman, 2014; Rury and Carrell, 2020; Ersoy, 2019a). In order to fully understand the casual factors behind academic decision making, researchers must collect and analyze agents' subjective expectations as observed decisions may be compatible with several different sets of preferences (Manski, 2004). As such, I collect and study several beliefs about students' major choice. I use these beliefs to study the the leaky pipeline out of STEM. I find that students who intend to study STEM believe they are just as likely to graduate in that field as non-STEM students. I also use students' beliefs to study *how* students incorporate new information about their ability into their decision making. I find suggestive evidence that ex ante beliefs may be inhibiting students who believe they are below average in a field of study from fully processing information that they are above average in that field.

Thirdly, this paper contributes to the literature on performance feedback in education. Papers in this literature typically exploit variation in the ability of students to learn their relative performance on an academic assessment, such as a grade in a course or on an exam, in order to study the effect of that information on subsequent academic achievement(Azmat and Iriberri, 2015; Azmat et al., 2019; Azmat and Iriberri, 2010; Bandiera et al., 2015; Murphy and Weinhardt, 2020; Goulas and Megalokonomou, 2018; Brade et al., 2018; Gonzalez, 2017). These studies argue that providing feedback to students changes students' perceptions of their ability, although some claim these effects operate through increases in self-confidence. The information I provide students is a unique form of performance feedback that allows students the rare opportunity to compare themselves to other students within a field of study while in college. The information I offer students is highly salient to students, as the measures of ability I provide in my information treatment are derived from statistics most students spend a great deal of time considering when applying to college; HS GPA, SAT and ACT scores. This paper is the first paper to assess how receiving performance feedback about academic performance derived completely before entering college influences college major choice. This is important for university administrators hoping to bolster academic success for those students about to enter college. It is also the first paper to provide feedback to college students across a set of majors.

My results also contribute to the study of labor supply in STEM occupations in the U.S. Along with a leaky pipeline found in STEM fields in higher education, there has also been a slowdown in the domestic supply of STEM students in the U.S. labor force (Carnevale et al., 2011). This slowdown has led policymakers to call for a drastic increase in the number of STEM-capable workers (Olson and Riordan, 2012). This paper studies a plausible explanation for this slowdown in STEM-capable students; that STEM majors switch fields because of misperceptions of their ability to succeed in that field. As I describe later, I also find evidence that by providing students with information that they are above average in their top field of study increases both STEM persistence and STEM major choice overall.

Lastly, I contribute to the literature studying the use of "nudges" and behavioral economics in education. While nudges are often deployed in educational settings, existing evidence is mixed on their efficacy in changing behavior (Oreopoulos et al., 2020; Oreopoulos and Ptronijevic, 2019; Damgaard and Nielsen, 2018). I show that light touch interventions can indeed induce meaningful behavioral change in educational decision making. I also find evidence of confirmation bias in how students update their beliefs about their ability and measure it's influence on major choice, extending the literature on potential behavioral biases in educational settings (Lavecchia et al., 2016; French and Oreopoulos, 2017; Reuben et al., 2015; Owen, 2020).

The rest of the paper is structured as follows; section 2 describes the experimental design and setting including a discussion of the information intervention and how it was constructed; section 3 describes the data, presents the descriptive results and discusses the leaky pipeline out of STEM; section 4 presents the experimental results; section 5 provides a discussion of my findings; and section 6 concludes.

2 Experimental Details

The experiment was conducted during the fall 2018 and winter 2019 academic quarters at University California at Davis (UC Davis). UC Davis is a large, selective public research university in Northern California with an undergraduate enrollment of over 30,000 students. The 75th percentile of high school GPA for the 2019 freshman class was 3.86, while the 25th percentile was 4.18. To study STEM major choice, the sampling frame was chosen to consist of large, introductory STEM courses, which are typically overly represented by underclassmen. Course instructors and department administrators were contacted for participation during the preceding academic terms, with over a dozen instructors agreeing to advertise the survey.

Several pieces of the experiment were arranged prior to students taking the survey. Firstly, to create my information treatment, the university's registrar provided me with average scores of recent graduates within the a set of 16 major groups. These averages were calculated for five different measures of ability; high school GPA, SAT combined score, SAT math score, SAT reading and writing score and ACT score. These average scores were derived using scores from the five most recent cohorts of UC Davis graduates and were calculated *within each major group*. This implies that I was given 16 x 5 different measures of ability that were calculated using the universe of recent UC Davis graduates. These averages were incorporated into the survey and used as part of the survey's internal display logic[2]. Secondly, for all students included in the experiment's sampling frame, scores from all five measures were also uploaded to be used within the survey.

Students were incentivized to take the survey by being entered into a raffle to win one of several amazon gift cards. Upon taking the survey, participants were asked several questions about their educational preferences and beliefs. First, students were asked to identify their top and second choices of majors. Students were given 16 options; economics, biological sciences, physics, chemistry, communications, psychology, engineering (any type), mathematics, statistics, foreign language, computer

[2]This was necessary due to a research design constraint from my campus partners that restricted me to only giving information to students whose own score was above the average score. I discuss this further later on in the paper.

science, English, history, philosophy, political science and sociology. As most of the literature on major choice has focused on the pecuniary returns to majors and it's roll in determining major choice, students were then asked about their expected earnings both five and 20 years after graduating in both their top and second choices of major.

To capture students' subjective expectations about graduating with different majors, student were also asked "what is the probability you will believe you graduate in your top choice of major, second choice of major, or some other major?", with the sum of the probabilities constrained by the survey to sum to one. In order to study their beliefs about their relative ability, students were then asked a series of questions concerning their top choices of major to elicit those beliefs. This series of questions was designed to avoid any direct elicitation of beliefs, as these responses may have been more subject to experimenter bias or motivated reasoning. In my elicitation procedure, students were first asked to select which academic measure best represents ability in their top and second choices of major. Students were given the same options for which I was provided average scores of recent graduates; HS GPA, SAT, SAT math, SAT reading and writing and ACT score. After students selected these measures for both their top and second choice of major, they were then asked "What do you believe is the average score of recent UC Davis graduates is in your top choice of major?", where "score" was replaced with the academic measure they selected and "top choice of major" with their selection for most preferred major. As we will see later, these beliefs are crucial to my analysis on confirmation bias, as they allow me to credibly characterize students as believing they are "above average" or "below average" in their top choices of major.

After students provide their beliefs of the average score of recent graduates in their top choices of major, the survey determines who is eligible to be randomized into either treatment or control groups. To do this, the survey compares the student's *own score* in the measure they selected for their top choice of major with the *average score* of recent graduates in that measure and major. If this criterion is satisfied, students are then randomized into either treatment or control groups. For students randomized into the treatment group, a message appears that reads "The average score of recent graduates in your top choice of major is (*actual average major-measure specific score*)", where again "score" is the measure students selected as most representative of ability and "top choice of major" is students selection for top choice of major. This message is then followed by a small nudge that states "Our records show that your score is above the average score of graduates in your top choice of major. We hope that this information helps you in your college major decision." The survey then follows the same procedure for the student's second choice of major, after which the survey concludes.

Due to design constraints imposed by my institutional collaborators, I was not permitted to provide information to students whose own score was below the average score of recent graduates. I discuss this restriction towards the end of the paper when I consider the policy implications of my findings. Lastly, students were also asked to sign a Family Educational Rights and Privacy Act (FERPA) release so that I could access their academic records, including their major choice history history, as well as their demographic and background characteristics.

3 Data and Descriptive Results

3.1 Data

The data for this project consist of 728 completed baseline surveys, as well as demographic and academic background information, and major choice history up to two years after the experiment for all students who signed a FERPA release. Due to the research design constraint described earlier, 483 of these students were eligible for treatment in their top choice of survey, as they had scores in majors/measures they selected that were in fact above the average score of recent graduates in that major/measure [3]. Slightly fewer, 456, were eligible to receive information in their second choice of major.

3.2 Experimental Balance

As mentioned above, students became eligible to receive treatment through their major choices and the measures they selected as most representative of ability in those majors. For those whose own score was above the average of recent graduates in that major, the survey software randomized them into treatment and control groups. This randomization was stratified by top choice of major.

In order to check that this randomization worked as designed, I conduct several tests. I received a rich set of variables on student demographics and pre-collegiate academic performance from the university's registrar, including race/ethnicity, gender, low-income and first-generation status as well as high school GPA and standardized test scores. Tables 1 and 2 show results from multiple models that regress survey responses and demographic variables on treatment status for both top and second choice of major. In each of these sets of analyses, only one coefficient is statistically different than zero, less than what would happen due to chance. I also fail to reject the null of joint significance in any of the models. Taken

[3] One clear question here is why is the number of students who are "above average" more than half of the sample. There are two answers, one slightly less obvious than the first. Firstly, this may represent sample selection, with those who are above average more likely to take the survey than those who are not. Another explanation come from the fact that standardized test and HS GPA scores at UC Davis have increased over the years. This would mechanically push up the number of students who are above average when compared to average scores of previous cohorts.

together, I see this as firm evidence that students were in fact randomly assigned between treatment and control groups for treatment in both a student's top and second choice of major.

3.3 Sample Descriptives

Table 3 provides demographic information for students in the sample. The sample is almost exactly two thirds female, slightly higher than the proportion of undergrads that are female at UC Davis (61%). Half of the sample is Asian or Asian-American, a quarter white, a fifth Hispanic and about 3% African American. While there are small differences, these numbers mirror the student population at UC Davis quite well[4]. Low-income and first-generation students, however, appear to be slightly under-represented in my sample when compared to the UC Davis population. They represent 26% and 37% of my sample respectively, while they constitute 37% and 42% of the student body. As my sampling frame consisted of large introductory STEM classes, 63% of students in my sample are freshman.

Table 3 presents results on top and second choice of major for the full set of survey respondents. For these results, I aggregate majors into four groups; social sciences (economics, sociology, political science, psychology); humanities (English, foreign language, history, philosophy); Biological Sciences; and STEM (physics, chemistry, mathematics, statistics, computer science and engineering (any type)). We see that biology is the most popular top choice of major (47.3% of students), followed by social sciences (24.9%), STEM (22.6%), and humanities (5.1%). The relative popularity of biology is striking, but somewhat consistent with major choices at UC Davis, with it's strong agricultural focus and popular biochemistry fields[5].

After selecting their top choices of major, students were then asked to select the academic measure they believe most represents ability in these majors[6]. This choice of measure selection by students is a novel element of my research design as most other work in the performance feedback literature restricts the type of feedback students receive, either through features of the institutional setting or by assumptions made by the researcher.

I next study whether students consider these measures to differentially represent ability, either within majors or across all majors If students did not believe scores in these measures carried any explanatory

[4]The Asian/Pacific Islander population at UC Davis represents about 32% of the student population. Another 17% of students are international, the vast majority of which are from Asian countries, predominantly China. The sum of these two terms is 49%, almost identical to the number of Asian students in my sample. Unfortunately, I do not have international status in my data.

[5]The sampling frame for this study was introductory STEM courses, which ended up including two large introductory biology courses. This fact may be skewing the distribution towards biology students

[6]Students were encouraged to only select a measure for which they themselves had a valid measure for. This implies that students should only select SAT combined score if they in fact had taken the SAT

power about ability , they would be indifferent between any measure and responses may be randomly selected. We see in table 3 that students are indeed selecting different measures and as we will see when we focus on STEM, they are choosing different measures for different majors. Also, if students did not believe these measures actually represented ability in majors differentially, students would be more likely to select the same measure for both their top and second choice of major. In fact, I find that more than a third of the sample select different measures for their top and second choice of major. Looking at the entire sample, the majority of student believe HS GPA is most representative of ability in their top choice of major, with more than five as many students selecting it than selecting the SAT.[7]. I discuss later in the paper why I believe these measures and the information they contain are in high demand from students and are likely to influence student behavior.

As part of survey, students were asked to provide their belief of the average scores in their selected measures of recent graduates in the top choices of major. These responses allow me to categorize students as believing their are "above average" or "below average" in these majors. To do so, I compare a student's belief of the average score with their own score (in that same measure). I then define a student as believing they are "above average" if a student's *own score* $>$ *belief of the average score*. Conversely, I categorize students as believing they are "below average" in a major if *own score* $<$ *belief of the average score*. Figure 2 shows the distribution of the difference between these two measures. We see that students appear to have on average unbiased estimates about the average score of recent graduates.

Table 4 presents results comparing students who select a STEM field as their top choice of major to those who select a non-STEM field across all survey outcomes[8]. Here we see that STEM are students slightly more likely to select a standardized test score as most representative of ability, and 8.8 percentage points more likely to select the math portion of the SAT. These beliefs align with stereotypes both of STEM students being math-focused and STEM curriculum being relatively math-heavy. In addition to the differences in academic measure selection, we see that students expect to earn significantly more than non-STEM students, both five and 20 years after graduation. STEM students are also 9 percentage points less likely to rate themselves as above average in their top field of study. Conditioning on being above average reduces this particular difference between STEM and non-STEM students, but as figure 3 shows, about 27% of STEM students who are actually above average believe they are relatively bad. All of these results are robust to controlling for majors within these groupings.

[7]Transfer students do not have valid HS GPA or standardized test scores in the administrative data as they are not required to report those scores to the university when applying. The only valid pre-Davis measure of academic ability I have for transfer students is community college GPA.

[8]As I discuss below, I exclude biology from STEM in all of my descriptive analyses.

3.4 The Leaky Pipeline Out of STEM

To explore the dynamics of major persistence and major switching over time, I compare the distribution of top choice of major with the empirical distribution of actual majors two years after the experiment. I exclude both those who where ineligible for treatment (those who had scores below the average) as well as students who received information in their top choice of major from these analyses. As we will see in the experimental results, there is reason to believe the information provided to them influenced their major choice behavior. These results on persistence for the control group can be found in Table 5. Here I compare the distribution of top choice of major to the empirical distribution of major choices I see two years later. Due to the large number of students intending to study biology, I have separated it from other STEM fields and presented the three major groups along with biology.

Looking at table 5, we see increases in the number of students majoring in each biology, social sciences and humanities two years after the experiment when compared to the number of students who selected those fields as their top choice of major at the time of the survey. While biology and the humanities see a small positive change (2.9% and 5% respectively), social sciences saw a 15.8% increase between the number of students who had a social science as their top choice of major and the number actually majoring in it two years later. In sharp contrast, however, STEM saw a decrease of 33.2% during this time.

Perhaps a more direct way to assess whether students of different majors are more or less likely to persist in those fields is to see which of these groups of majors has the highest persistence rate, conditional on selecting it a major in that group as your top choice . To do this, I create a variable that equals one if the student is majoring in their top choice of major two years after the experiment. I find that 91% of biology students and 83% of social science students are majoring in their top choice of major two years later. Looking at STEM students however, only 52% are still majoring in their top field [9]

Given that STEM fields are highly mathematical and are known for "weeding out" students with insufficient aptitude in mathematics, these transitions out of STEM may reflect differential mathematical preparation. Instead of observing a leaky pipeline that reflects incorrect beliefs about ability, we may simply be observing the least mathematically prepared students switching to other majors. To study this, I compare the SAT math scores of those STEM students who switched out of their top field to those who persisted as measured two years later. Table 9 presents descriptive comparisons between the "leakers" and non-leakers, namely those students who top choice of major was a STEM field, were eligible for

[9] Only 50% of students who chose a field in the humanities are still majoring in that field two years later, although there are only 14 of these students so it is hard to draw sharp conclusions about these fields.

treatment (meaning they were above average in their measure/major combination) but were randomized into control. While these samples are small (n = 25, n = 28 respectively), they provide a unique glimpse into the demographic makeup of students who switch out of STEM. Here we see that these students have an average SAT math score that is 29 points (800 point scale) higher than those students who continue on in their top field. While this result is not statistically significant, it calls into serious question the role of mathematical aptitude in explaining exit out of STEM.

Another explanation for why students are switching out of STEM is that once they learn they about the course content and curriculum involved in those fields, their preferences change and grow to like other fields more relative to STEM fields. While I would not doubt this is true for some students, there are two reasons why this is unlikely to explain the majority of this exit out of STEM. The first is that preferences for majors, particularly for STEM fields are likely explained by other factors such as expected study effort, difficulty of courses or expected grades/GPA. A true test of the role of preferences for STEM changing would be whether students' expectations of study effort or the efficacy of their study effort as well as any other characteristic other than the curriculum of the major remain fixed and we still observe major switching after students experiment with the major. I am unable to hold constant these expectations in my setting and therefore cannot explicitly rule out changing preferences as a potential mechanism for the leaky pipeline out of STEM. Secondly, however, if preferences were the primary driver of this phenomenon, my intervention would have little effect on persistence in STEM. As we will see, I find evidence that the information I provided students increased persistence in these fields, lending more support to a story about perceptions of ability rather than preferences.

Table 9 also shows that students who switch out of their top field are more likely to be female (significant at 10% level), first-generation students and Hispanic, although there is no statistically detectable difference for these last two characteristics. While these students do not appear to be more or less likely to rate themselves as "above average" in their top field, they place less distance on average between their own score and their belief of the average score (own score - belief of average) and is almost marginally significant (p-value = 0.154). They are also appear more likely to underestimate the actual average score of recent graduates in their top choice of major (p-value = 0.136).

These dynamics out of STEM may not truly reflect a leaky pipeline if student expectations are in line with the empirical distribution of major choices two years after the experiment. For example, students who begin with a STEM field as their top major may expect to switch out of that field more so than students whose top choice of major is another field. This might be true if students place a higher option value on starting college in a STEM field. As we saw earlier, students whose top choice of major was a

STEM field expected to earn significantly more in those fields conditional on graduating in those majors. This would imply that students experiment with STEM, but place a lower probability on the likelihood that they ultimately persist in that field. If students were indeed as earnest in their top choice of major in STEM as in other fields, the differences in likelihoods of graduating in your top majors would not be very different between STEM and non-STEM fields.

As part of the survey, students were also asked to provide the likelihood they believed they would graduate in their top choice, second choice or some other major. I incorporate these subjective probabilities into my analysis of the leaky pipeline to assess the above hypotheses. To do so, I condition on top choice of major group and compare the average subjective probability students report that they will graduate in that field to the average empirical probability they in fact do. Again here I condition on being in the control group for top choice of major. Table 6 presents results. Here we again see strong evidence of a leaky pipeline out of STEM fields. STEM students appear to report subjective probabilities of graduating in their top field that are only slightly different than students of other fields. Importantly, however, STEM students are much less likely to persist in that field two years after the survey than students of other fields.

These combined results provide evidence of a leaky pipeline out of STEM fields when excluding biology, but strikingly not for biology itself. If anything, students appear to underestimate the rate they will ultimately study biology. For this reason and those previously mentioned, I focus on STEM fields excluding biology when studying my experimental results.

4 Experimental Results

4.1 Main Effects on Persistence

To study whether the information provided to students in my intervention influence persistence in STEM, I estimate the following statistical model:

$$persist_i = \alpha + \beta TREAT_i + X_i \phi + S_i \gamma + \psi_m + \varepsilon_i$$

where *persist* is whether the student is majoring in their top choice of major two years after the intervention, $TREAT_i$ is an indicator if they receiving information in their top choice of major, X_i is vector of background characteristics such as race/gender and standardized test scores, S_i contains baseline survey responses and ε_i is random error term. Random assignment of students into treatment status assures me that $E[\varepsilon_i | TREAT_i] = 0$, allowing me to estimate the causal impacts of information on my outcomes of interest.

To conduct inference, I estimate standard errors and calculate statistical significance under standard asymptotic assumptions. I also follow Athey and Imbens (2017) and perform randomization inference. In this procedure, treatment assignment is permuted across the analysis sample and and treatment effects are calculated for each permutation[10]. Empirical p-values are then calculated by ranking these treatment effects and assessing where the true model lies within this distribution. This approach has many advantages for hypothesis testing. I present both standard errors and empirical p-values in my results.

I estimate four models for my primary outcomes, including a model with no controls (the pure experimental results), as well as models with major fixed effects and controls for demographic characteristics and survey responses. Table 10 presents results separately for both students who select a STEM field as their top choice of major as well as students who selected a non-STEM field. Focusing on STEM students, we see that the point estimates across all models are positive and economically significant, with treatment effects ranging from an eight and 12 percentage point increase in STEM persistence after two years. While these results are economically significant, I fail to reject the null in each model. Looking at non-STEM students, we see an opposite, negative effect of information on persistence in top choice of major. These effects range between -5.4 and -6.8 percentage point decrease in persistence. These estimates are also more precise, with one model achieving marginal significance (p value = .074). This is likely due to a larger sample size for this outcome, with over three in four students in the analysis sample selecting a non-STEM field as their top choice of major[11].

Table 12 shows results from a difference-in-differences style model that looks at the effect of information on STEM students compared to non-STEM students, effectively differencing out the effect on non-STEM students from the effect on STEM students. Given the two large and opposing effects found within each group discussed above, I find a large and statistically significant differential effect between the two groups. These results provide clear evidence that STEM students are responding differently to information that they are above average in their top choice of major than non-STEM students, although less can be said about the magnitudes for each group individually.

I next study why non-STEM students appear to be switching out of their top choice of major after receiving information that they are above average in that major. As we saw earlier, I find strong evidence of a leaky pipeline *out* of STEM fields. I now check to see if there is a similar phenomenon for students

[10]I perform 500 permutations for each outcome.

[11]This is driven by the large number of students who select biology as their top choice of major. The differential effect of information between STEM and non-STEM as I define them, however, lends credibility to excluding biology from STEM in these analyses.

switching *into* STEM. To study this, I perform similar analyses as before, comparing the distribution of second choice of major to the empirical distribution of majors two years after the intervention. For this second distribution, I restrict the sample to students who are not majoring in their top choice of major two years later. This comparison tells us whether students' second choice of major matches up with the distribution of majors, *for those who are not majoring in their top choice*. If there is no leaky pipeline into STEM, under certain assumptions, these distributions should be similar[12]. Table 8 shows that while biology, social sciences and humanities all have higher representation in the distribution of students who switch out of their top field compared to the distribution of second choice of major, STEM sees a large, sharp decline. Using student's subjective probabilities that they believe they will major in their second choice of major instead of the actual distribution of second choice of major yields a similar result. Students appear to over-estimate the rate they will switch into STEM, a result similar to those found in Stinebrickner and Stinebrickner (2013). Along with the problems of students switching out of STEM discussed earlier, I see this as evidence of a leaky pipeline *into* STEM.

In light of these facts, I estimate models that focus on students who chose a non-STEM major as their top choice of major, but a STEM field as their second choice of major. These models study whether a student's recent major is their second choice of major. Tables 13 presents results looking at the effect of receiving information in either your first or second choice of major. Somewhat surprisingly, receiving information in your second choice of major does not appear to significantly change major choice behavior. Treatment effects are modest (3 percentage points) and not significantly different than zero. Receiving information in your top choice of major offers larger treatment effects (6 percentage points) and are more precise, although not significant at conventional levels (p value = .128). Despite the lack of precision in these results, they are consistent with a story in which students switch out of their top choice of major and into STEM upon learning they are above in that non-STEM field, as the effects are of similar magnitudes and in opposing signs.

I next study the effect of information on the likelihood of major in a STEM field across all students. This combines the impact of information on persistence in STEM for STEM students as well as the impact on switching into STEM from non-STEM students. Within the group of STEM students, this treatment effect will be attenuated when compared to the results on persistence in top field of major. This is due to the fact some STEM students switch out of their top choice of major, but switch into a

[12]The primary assumption is that for any level of major switching, the individual transition probabilities from top choice of second choice of major do not favor any major over the other. Conditional on not majoring in your top choice of majo and being in the control group, the distribution of majors is much more uniform than top choice; social sciences and STEM are very similar with 34.8% and 32.6% respectively, while Biology and Humanities are smaller at 19.6% and 13%. The low representation of biology in the switcher distribution is likely driven by the fact that biology is by far the most popular top choice of major and only 1.6% of students who select biology as their top choice of major select a biology field as their second choice.

different STEM field. These students were seen as "not persisting" in the model studied earlier, while here they are categorized as having STEM as their most recent major. These estimates represent a unique policy parameter; how much STEM major choice would increase as a result of telling students they are above average in their top field[13]. Table 13 present these results. By combining these two separate effects into a single test, I am able to detect a significant positive treatment effect on STEM major choice of about 12% of a standard deviation in STEM major choice (significant at 10% level).

4.2 Heterogeneous Treatment Effects

I next study heterogeneous treatment effects across three different groups; female students, freshman and first-generation students. We saw earlier that female STEM students were more likely than males to switch out of their top choice of major, conditional on being above average in that field. Previous literature studying this has shown the female persistence in STEM is sensitive to role model effects (Carrell et al., 2010). The estimates studying heterogeneous treatment effect of treatment here may tell us if these effects found in the literature are partially driven by beliefs about ability. Table 14 presents results for female students. Looking at persistence in STEM fields, we see large heterogeneous treatment effects for female students, although these estimates are imprecisely estimated.

When considering freshman, we might expect these students to experience the largest effects from this information ex ante, for multiple reasons. First, they have received the least amount of college-level information about their ability in their top choices of major. Because of this, when compared to students with more college experience who have received performance feedback containing information about their ability, freshman students represent a blank slate upon which this information can operate to influence behavior. Another reason we might expect students to respond more to information is that they are less exposed to "sunk costs" when deciding a major, as more experienced students may feel too much has been invested in other majors to switch[14]. Looking at table 14 again, we see large heterogeneous treatment effects for freshman students as well, although just as with female students, these estimates are statistically indistinguishable from zero. These estimates are, however, consistent with a story in which freshman students enter college with less information about their relative ability in STEM fields than more experienced students.

The last subgroup I study are first-generation students. Similar to female students, in my descriptive analyses I find modest evidence that, absent treatment, first-generation STEM students are more likely to

[13] Implementing this policy would be difficult as all students who did not receive information could easily infer that they were below average. I discuss this more in section 5.

[14] In order to test this, posterior beliefs must be collected to asses how the information changes the beliefs of both freshman and non-freshman. If they change beliefs in a similar way but non-freshman appear to be less likely to switch, this may be evidence of sunk-cost bias

switch out of their top field compared to other STEM students. Studying the effect of treatment, we see in table 14 large and statistically significant differential treatment effects for first generation students, ranging from 34 percentage points (p-value =.076) to 43 percentage points (p-value = .032). Similar to the results studying freshman students, these results are consistent with a story in which first-generation students have less information about their relative ability (as captured by the academic measures used here) and are poised to incorporate it into their behavior upon receiving it.

4.3 Confirmation Bias

Finally, I examine how students process information about their ability and whether students exhibit any behavioral biases in doing so, specifically confirmation bias. In my setting, I define confirmation bias as placing more weight on information that confirms your beliefs than on information that contradicts them when deciding on your college major. There is a large literature on confirmation bias in psychology and economics. Bénabou (2015) offers a nice review. Despite the amount of research on this topic more broadly, there is little to no work studying how confirmation bias might be impacting student decisions. This is surprising given the growing literature on the effects of performance feedback in education, which emphasizes how students respond to learning about their performance relative to their peers. To study how students respond to performance feedback, this literature exploits the fact that students are continuously receiving information about their ability via results of assignments, exams, grades and other assessments. This literature has yet to study, however, *how* students process this information.

I plan to study this question by devising a test for confirmation bias that studies how students behave in response to receiving information about their ability. To do so, I leverage student survey responses to characterize students as believing they are "above average" or "below average" in their top choices of major. As described in section 3, I compare a student's belief of the average score of graduates in a major to their own score to construct this characterization. I characterize students who believe the average score of recent graduates is below their *own* score as believing they are above average. Students who place their belief above their own score are therefore characterized as believing they are below average. In a standard Bayesian framework, information that you are above average should be more useful to students who have prior beliefs that they are below average, conditional on the variance of those beliefs. This implies that this information is less informative to students who hold prior beliefs that they are above average.

The simplest way to test for confirmation bias in this setting would be to collect posterior beliefs after students receive information. This is challenging in my setting as the information I give students is the average score of recent graduates in their top majors. Posterior beliefs assessing students beliefs about the average should be centered around this number, with no room for uncertainty about this quantity. It is therefore uninformative to ask students about their beliefs of the average score after they receive information that contains the average score. There are two other potential ways to collect posterior beliefs in this setting, although as they are not included in my experimental design, they must be left to future research. Firstly, alongside eliciting beliefs about average scores of recent graduates, students could be asked to provide a probability distribution over where they think they place within the *ability* distribution within a cohort. After information is provided, this question can again be asked to measure changes in beliefs. Also, posterior probabilities of the likelihoods of graduating in students' top choice, second choice or some other major could be collected after students receiving information. I leave these to future work.

To test for the existence of confirmation bias, I estimate the following statistical model:

$$persist = \alpha + \beta_1 TREAT_i * ABOVE_i + \beta_2 TREAT_i + \beta_3 ABOVE_i + X_i\phi + S_i\gamma + \psi_m + \varepsilon_i$$

where here $ABOVE_i$ is an indicator for whether the student is characterized as believing they are above average. In the above model, β_1 is the coefficient of interest, as it represents the differential effect of treatment on students who believe they are relatively good. If confirmation bias exists in this setting, students who believe they are above average should be relatively more likely to use this information in their decision making than students who do not. As such, the above model constitutes a test for confirmation bias, specifically if $\beta_1 > 0$. This would imply that the information I provide has a stronger effect on those students who already believed they were above average. Table 15 presents results. Estimates of β_1 are positive across all models, although only one model is marginally significant (p value = .058). These estimates suggest a differential treatment effect between 8 and 17 percentage points for those students who already believe they are above average. Focusing on STEM students, estimates of β_1 are twice as large as those for the entire sample, again with the model containing major fixed effects, survey responses and demographics significant at the 10% level. These effects for STEM students again suggest how different STEM fields are from other fields regarding the role of beliefs about ability.

In order for these estimates to be a test of confirmation bias, I must rule out the role of the changing variance in beliefs. If the effect of *reducing the variance* in beliefs as the result of treatment for those who believe they are above average is stronger than the effect of *shifting the mean* of beliefs for those who believe they are below average, the interpretation of these results is less clear regarding confirmation bias. There is some credibility to this mechanism, as Conlon (2020) finds that providing information about the pecuniary returns to majors to students influences behavior through a reduction in the variance of the beliefs about returns to fields of study. Unfortunately, as I do not collect posterior beliefs about ability of likelihood of graduating in top major, I cannot test this directly in my setting.

As I do not study the effect of information on posterior beliefs, but rather on the behavior itself, these results should be seen as a reduced form test for confirmation bias, measuring changes as the exogenous treatment (shocking student beliefs with information) influences behavior, while ignoring changes in the endogenous explanatory variable (beliefs). I leave to future work to measure more precisely how changes in students' beliefs about their ability in their top choice of major influence major choice behavior.

5 Discussion

The supply and demand of STEM-capable workers in the U.S. has received much attention from academics and policy-makers alike over the past decade (Olson and Riordan, 2012; Carnevale et al., 2011). Empirical work studying the supply of STEM-workers has focused on educational pipelines from the classroom to the labor force and has shown that STEM students are more likely to switch out of STEM more so than students of other fields (Chen and Soldner, 2014). This comes at a time when the growth in demand for STEM workers appears to be growing faster than other occupations[15] The results from this paper highlight both potential reasons why the supply of STEM-capable workers is failing to keep up with demand and ways to address it. I find evidence that students' beliefs about their ability to complete a STEM degree play a crucial role in STEM persistence in college. Moreover, these beliefs appear to be sensitive to information nudges that provide critical information to students about their relative ability in their top fields of study. Information nudges such as the one used in this paper may be powerful policy tools used in closing the gap between the supply and demand of STEM graduates.

While they may be tentative, my findings on confirmation bias highlight the potential role of behavioral biases in education. One implication that arises from these results is that student beliefs may be more difficult to influence if students believe they are not "good" at a certain subject or academically more broadly. This implications makes it incredibly important to understand how and when these beliefs are formed and how we can better design educational institutions to protect students from false negative beliefs. This is particularly true given the current design of educational institutions requires students of all ages receive performance feedback on a regular basis. This feedback can carry important content that is useful to the student, although this stream of relative feedback may exacerbate existing inequalities

[15]Bureau of Labor Statistics: https://www.bls.gov/emp/tables/stem-employment.htm

in beliefs about ability if confirmation bias is inhibiting students from positively updating their beliefs. Given how consequential students' decisions are, particularly those made in college, more research is needed to understand how biases in belief updating influence the distribution of beliefs we observe once students arrive at college.

One limitation of my research design is that I cannot test how students who are below average would respond to information. This is an important limitation as prior research has demonstrated that these students may respond negatively to such information (Franco, 2020; Azmat et al., 2019). Understanding how these students might respond to learning they are below average would help policy-makers balance the trade offs when considering making average scores of recent graduates in different majors public. Below average students may increase their study effort in response to learning they are in fact below average. Other work in the performance feedback literature shows that this may be the case (Azmat et al., 2019). Also, there may be cost savings from informing students they are below average and inducing them to change majors, as the net cost to both the student and the university of taking courses that do not contribute to a degree or skills that will be used later on may be high.

Lastly, there continues to be a dramatic under representation of female students in STEM fields. Previous work has (Carrell et al., 2010). I find large treatment effects on persistence for female students amounting to a significant increase in STEM persistence, although perhaps due to a small sample size, I am unable to rule out a null effect of information for female students. My results are promising, however, and highlight a need to pursue this mechanism further in future research.

6 Conclusion

One of many consequential decisions students must make is choosing their college major. As with most other academic decisions, students often make this decision with incomplete information, particularly how they compare academically to successful students in those majors. This may be especially true in STEM fields, as students are more likely to leave those fields compared to other subjects, evidence of a leaky pipeline in STEM. In this paper, I study the effect of providing students with information that they are above average in their top choice of major on major persistence, with a focus on STEM fields.

I find strong evidence of a leaky pipeline out of STEM that is consistent with a story where students leave STEM because they believe they are not academically suited for those fields. In my experimental results, I find evidence that providing STEM students with information that they are above average in that field increases STEM persistence by 22.7% compared to the control mean, although these estimates are imprecisely estimated. I also find that providing students with information they are above average in their top choice of major, even if that field is non-STEM, increases STEM major choice by about 30% of the control mean, or about 12.6% of a standard deviation in STEM major choice. The effects for STEM students are strongest for first-generation students, with mild evidence of a differential effect for female students. These results highlight an important potential mechanism as to why students are more likely to leave STEM fields.

I also leverage my novel survey design to offer the first test of confirmation bias in college major choice. I find larger treatment effects for students with ex ante beliefs that they are relatively good in their top choice of major, evidence that the information was relatively less useful for students with beliefs that they were below average. These results provide strong evidence that the leaky pipeline out of STEM is driven partly by beliefs about ability and that one way to tighten the leaky pipeline is to provide information to students that they are above average in those fields. The role of confirmation bias may dampen the effectiveness of this type of intervention. Future research should study how student beliefs arise and how we can protect students from incorrect beliefs that they are not academically well suited for STEM fields. Much research is also needed on how to support below average students in STEM. Finally, these findings raise broad questions about the role of behavioral biases in educational settings, a result that should motivate scholars to pursue a better understanding of how these biases operate on student decision making.

7 Figures and Tables

Table 1: Selection Regressions (Top Choice of Major)

	(1)	(2)	(3)	(4)	(5)
	second_treat	second_treat	second_treat	second_treat	second_treat
	b/se/pvalues	b/se/pvalues	b/se/pvalues	b/se/pvalues	b/se/pvalues
Biology 2nd Choice	-.0387612			-.0471451	-.0483135
	(.09463)			(.0990533)	(.1040674)
	[.708]			[.666]	[.642]
STEM 2nd Choice	.0010719			-.0103032	-.0187751
	(.059989)			(.0665136)	(.0709291)
	[.982]			[.874]	[.778]
Soc Sci 2nd Choice	-.0027364			-.0089652	-.0169423
	(.0628513)			(.0670663)	(.0727567)
	[.962]			[.872]	[.806]
Pr(Graduate 2nd Choice)		.0538891		.0547299	.0533238*
		(.1372488)		(.1388065)	(.1532983)
		[.668]		[.686]	[.084]
E[Salary 2nd Choice] 5y		-.0000584		-.0000391	-.0001286
		(.0008207)		(.0008499)	(.0009727)
		[.934]		[.954]	[.652]
Believe Good (2nd Choice)		.0121631		.023755	.002211
		(.0494838)		(.0557497)	(.0557634)
		[.81]		[.666]	[.652]
Gender			-.0908062		-.1004198
			(.0576425)		(.0621597)
			[.106]		[.548]
Black			-.0957025		-.0978263
			(.265269)		(.2685665)
			[.644]		[.652]
White			-.1149017		-.1127278
			(.1618172)		(.1639296)
			[.692]		[.596]
Asian/Asian American			-.1044739		-.1090483
			(.1590971)		(.1610263)
			[.53]		[.926]
Hispanic			-.0829779		-.0840887
			(.1772706)		(.1797265)
			[.67]		[.532]
Low-Income			-.0345989		-.0337812
			(.07533)		(.0767122)
			[.608]		[.724]
First Generation			.0441298		.0446684
			(.0684778)		(.0695066)
			[.846]		[.766]
SAT/ACT score			.0000972		.0000918
			(.0002499)		(.0002574)
			[.548]		[.86]
freshman			.0120216		.0058852
			(.0557821)		(.0567696)
			[.69]		[.954]
Observations	430	428	373	428	371

Standard errors in parentheses; empirical p vaues in brackets

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2: Selection Regressions (2nd Choice of Major)

Table 3: Top/2nd Choice of Major/Measures

	(1) Top Choice of Major	(2) Second Choice of Major
Majors		
Biology	47.3%	11.3%
STEM (No bio)	22.6%	41.1%
Social Sciences	24.9%	30.3%
Humanities	5.1%	50.0%
Academic Measures		
HS GPA	68.3%	55.6%
SAT Combined	10.9%	17.0%
SAT Math	8.6%	12.88%
SAT Reading/Writing	0.6%	6.99%
ACT	11.5	7.4
Observations	486	409

Notes: Sample consist of students in the control group

Table 4: Descriptive Statistics - All/top non-STEM/top STEM

	(1) All students	(2) Top non-STEM	(3) Top STEM	(4) Difference
Earnings 5y post grad (thousands)	78.3	75.8	86.9	11.0***
				(3.3)
Earnings 20y post grad (thousands)	113.0	110.5	121.4	10.9***
				(3.37)
Prob grad Top Major	72.0	72.0	72.2	0.2
				(0.09)
Believe good Top Major	68.9	69.4	67.3	-2.1
				(0.43)
HS GPA top measure	68.3	69.1	65.5	-3.7
				(0.73)
Standardized test top measure	31.7	30.9	34.5	3.6
				(0.73)
SAT math top measure	8.6	6.6	15.5	8.8***
				(2.91)
Female	66.6	73.7	42.7	-30.9***
				(6.28)
Observations	486	376	110	

Notes: Sample includes only those in control group. t-statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5: Leaky Pipeline out of STEM (Top Choice)

	(1)	(2)	(3)	(4)
	Top choice Prob.	Empirical Prob.	Δ	Percentage Δ
Biology	46.9%	48.3%	+1.4	+2.9%
	(50.0)	(49.8)		
STEM (No bio)	22.6%	15.1%	-7.5	-33.2%
	(41.9	(37.7)		
Social Sciences	24.7%	28.6%	+3.9	+15.8%
	(43.2)	(45.3)		
Humanities	5.6%	5.9%	+0.3	+ 5%
	(23.5)	(23.4)		
Observations	239	238		

Notes: First column summarizes indicator if student's top choice of major was in the corresponding major group. Standard deviation in parentheses. Sample consist of students in the control group for top choice of major (above average in top choice of major, no treatment)

Table 6: Leaky Pipeline out of STEM (Subjective Probabilities)

	(1)	(2)	(3)	(4)
	Subjective Prob.	Empirical Prob.	Δ	Percentage Δ
Biology (n = 112)	73.3%	90.9%	+17.6	+24.0%
	(23.5)	(28.8)		
STEM (No bio) (n = 54)	70.5%	51.9%	-18.6	-26.4%
	(25.4)	(50.4)		
Social Sciences (n = 59)	68.3%	83.1%	+14.8	+21.7%
	(26.1)	(37.8)		
Humanities (n = 14)	57.7%	50.0%	-7.7	-13.3%
	(20.1)	(51.9)		

Notes: First column provides probability student believed they would graduate in their top choice of major, and conditions on that major group. Standard deviation in parentheses. Sample consist of students in the control group for top choice of major (above average in top choice of major, no treatment)

Table 7: Leaky Pipeline out of STEM (Pr(persist))

	(1)	(2)
	Pr(Persist 2ys Later)	Percentage Drop
Biology (n = 111)	91.0%	-9.0%
	(28.8)	
STEM (No bio) (n = 54)	51.9%	-48.1%
	(50.4)	
Social Sciences (n = 59)	83.1%	-16.9%
	(37.8)	
Humanities (n = 14)	50.0%	-50.0%
	(51.9)	

Notes: First column summarizes indicator if student is majoring in top choice of major two years after survey. Standard deviation in parentheses. Sample consist of students in the control group for top choice of major (above average in top choice of major, no treatment)

Table 8: Leaky Pipeline into STEM (Pr(grad in 2nd choice))

	(1)	(2)	(3)	(4)
	Subjective Prob.	Empirical Prob.	Δ	% Δ
2nd Biology (n = 18)	18.6%	38.9%	+20.3	+109.1%
	(24.1)	(50.2)		
2nd STEM (No bio) (n = 65)	17.5%	7.7%	-9.8	-56.0%
	(16.1)	(26.9)		
2nd Social Sciences (n = 55)	18.8%	18.9%	+0.1	+0.5%
	(20.0)	(39.5)		
2nd Humanities (n = 31)	14.7%	0%	-14.7	-100%
	(12.6)	(0)		

Notes: First column provides probability student believed they would graduate in their 2nd choice of major, and conditions on that major group. Standard deviation in parentheses. Sample consist of students in the control group for top choice of major (above average in top choice of major, no treatment)

Table 9: Summary Stats - Leakers and non-Leakers

	(1) Persister	(2) Leaker	(3) Difference
SAT Math score	683	712	29
			(1.07)
Female	28.8	50.0	21.4*
			(1.62)
First-Generation	25.0	32.0	7.0
			(0.56)
Low-Income	22.7	20.8	-1.9
			(0.15)
Asian	69.4	67.3	-2.1
			(0.43)
Hispanic	14.3	28.0	13.7
			(0.73)
White	25.0	28.0	3.0
			(0.24)
Believe Good	64.3	61.5	-2.7
			(0.21)
(Own Score - Belief)/Belief	8.0	3.6	-4.4
			(1.50)
(Belief-Actual)/Actual	-1.3	2.9	-4.2
			(1.45)
Pr(Graduate in Top Choice)	76.1	64.5	-11.6*
			(1.71)
Observations	28	25	

Notes: Sample includes only those in control group. t-statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 10: Effects on Persistence in Top Choice (STEM and non-STEM)

	(1)	(2)	(3)	(4)
Panel A: Top STEM (no bio)				
	Recent = top	Recent = top	Recent = top	Recent = top
First Treat	.118	.110	.080	.113
	(.095)	(.097)	(.094)	(.100)
	[.212]	[.236]	[.392]	[.306]
Observations	109	109	109	88
Panel B: Top non-STEM (no bio)				
	(5)	(6)	(7)	(8)
	Recent = top	Recent = top	Recent = top	Recent = top
First Treat	-.054	-.057	-.064*	-.068
	(.039)	(.037)	(.037)	(.042)
	[.162]	[.122]	[.074]	[.106]
Observations	373	373	371	320
Top Major FE	N	Y	Y	Y
Survey Responses	N	N	Y	Y
Demographics	N	N	N	Y

Standard errors in parentheses; empirical p values in brackets. Statistical significance determined by empirical P-values

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 11: Diff-in-Diff effect of treatment Effects on Persistence in Top Choice (STEM and non-STEM)

	(1)	(2)	(3)	(4)
	Recent = top	Recent = top	Recent = top	Recent = top
First Treat x STEM (No bio)	.172*	.168*	.162*	.187*
	(.089)	(.087)	(.085)	(.096)
	[.086]	[.088]	[.084]	[.070]
Observations	482	482	480	408
Top Major FE	N	Y	Y	Y
Survey Responses	N	N	Y	Y
Demographics	N	N	N	Y

Standard errors in parentheses; empirical p values in brackets. Statistical significance determined by empirical P-values

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 12: Diff-in-Diff effect of treatment by subgroup (STEM vs. non-STEM)

	(1)	(2)	(3)	(4)
Panel A: Freshman				
	Recent = top	Recent = top	Recent = top	Recent = top
First Treat x STEM (No bio)	.269**	.242*	.205*	.247*
	(.109)	(.109)	(.107)	(.121)
	[.038]	[.062]	[.086]	[.086]
Observations	323	323	321	266
Panel B: First-gen				
	(5)	(6)	(7)	(8)
	Recent = top	Recent = top	Recent = top	Recent = top
First Treat x STEM (No bio)	.430**	.511***	.457**	.446**
	(.174)	(.175)	(.170)	(.213)
	[.022]	[.004]	[.010]	[.026]
Observations	152	152	152	137
Panel C: Female				
	(9)	(10)	(11)	(12)
	Recent = top	Recent = top	Recent = top	Recent = top
First Treat x STEM (No bio)	.239	.221	.228	.233
	(.123)	(.120)	(.121)	(.134)
	[.102]	[.158]	[.142]	[.192]
Observations	322	322	320	277
Panel D: Low-Income				
	(13)	(14)	(15)	(16)
	Recent = top	Recent = top	Recent = top	Recent = top
First Treat x STEM (No bio)	.358	.465**	.489**	.469*
	(.207)	(.241)	(.241)	(.271)
	[.110]	[.038]	[.038]	[.068]
Observations	100	100	100	91
Top Major FE	N	Y	Y	Y
Survey Responses	N	N	Y	Y
Demographics	N	N	N	Y

Standard errors in parentheses; empirical p values in brackets

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 13: Effect of info in top choice on 2nd choice of major (STEM 2nd choice, non-STEM top)

	(1)	(2)	(3)	(4)
	Recent = 2nd	Recent = 2nd	Recent = 2nd	Recent = 2nd
Treat (Top Choice) $\times$ STEM 2nd Choice	.086	.089	.102	.061
	(.074)	(.072)	(.073)	(.084)
	[.174]	[.18]	[.124]	[.416]
Observations	300	300	298	253
Top Major FE	N	Y	Y	Y
Survey Responses	N	N	Y	Y
Demographics	N	N	N	Y

Standard errors in parentheses; empirical p values in brackets. Statistical significance determined by empirical P-values

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 14: Effect of Information on STEM Major Choice

	(1)	(2)	(3)	(4)
	Recent = STEM	Recent = STEM	Recent = STEM	Recent = STEM
First Treat	.045	.045*	.043*	.055**
	(.035)	(.025)	(.026)	(.028)
	[.194]	[.062]	[.080]	[.046]
Observations	482	482	480	408
Top Major FE	N	Y	Y	Y
Survey Responses	N	N	Y	Y
Demographics	N	N	N	Y

Standard errors in parentheses; empirical p values in brackets. Statistical significance determined by empirical P-values

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 15: Effect of Information on STEM Major Choice

Panel A: Female				
	(1)	(2)	(3)	(4)
	Recent = top	Recent = top	Recent = top	Recent = top
Treat (Top Choice) × Gender	.112	.079	.133	.189
	(.191)	(.199)	(.191)	(.213)
	[.540]	[.666]	[.454]	[.388]
Observations	109	109	109	88
Panel B: Freshman				
	(5)	(6)	(7)	(8)
	Recent = top	Recent = top	Recent = top	Recent = top
Treat (Top Choice) × Freshman	.139	.142	-.011	-.177
	(.205)	(.210)	(.208)	(.230)
	[.472]	[.486]	[.966]	[.474]
Observations	109	109	109	88
Panel C: First-Gen				
	(9)	(10)	(11)	(12)
	Recent = top	Recent = top	Recent = top	Recent = top
Treat (Top Choice) × First Generation	.412**	.433**	.336*	.206
	(.211)	(.218)	(.210)	(.233)
	[.044]	[.032]	[.076]	[.402]
Observations	105	105	105	88
Top Major FE	N	Y	Y	Y
Survey Responses	N	N	Y	Y
Demographics	N	N	N	Y

Standard errors in parentheses; empirical p values in brackets

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 16: Confirmation Bias - Effect of Information x Believing You Are Good

Panel A: All students				
	(1)	(2)	(3)	(4)
	Recent = top	Recent = top	Recent = top	Recent = top
First Treat x Top Good (Belief)	.080	.091	.085	.170*
	(.083)	(.079)	(.077)	(.090)
	[.328]	[.236]	[.258]	[.058]
Observations	482	482	480	408
Panel B: Top-STEM (no Bio)				
	(5)	(6)	(7)	(8)
	Recent = top	Recent = top	Recent = top	Recent = top
First Treat x Top Good (Belief)	.204	.190	.295	.391*
	(.204)	(.214)	(.204)	(.267)
	[.298]	[.342]	[.112]	[.08]
Observations	109	109	109	88
Top Major FE	N	Y	Y	Y
Survey Responses	N	N	Y	Y
Demographics	N	N	N	Y

Standard errors in parentheses; empirical p values in brackets. Statistical significance determined by empirical P-values

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure 1: Subject Probabilities (all students)

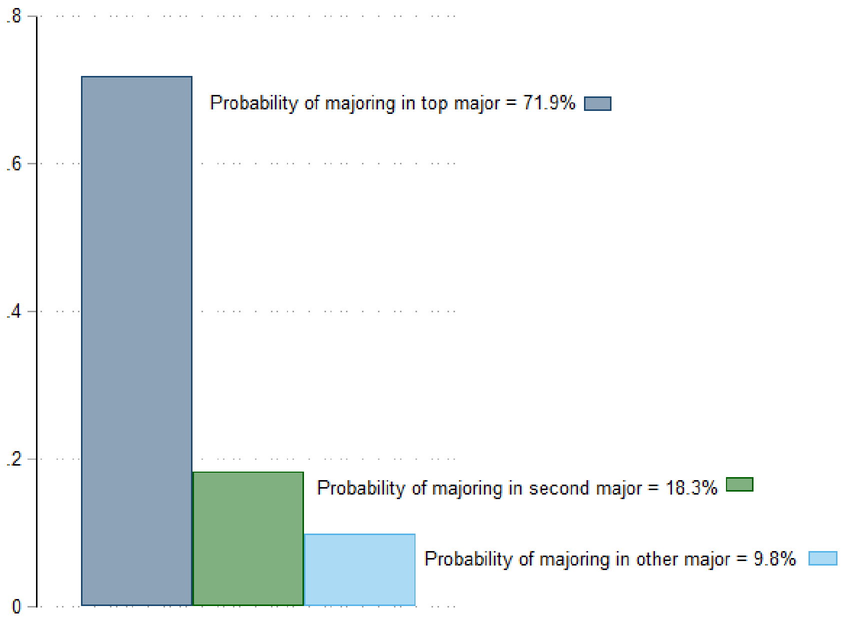


Figure 2: Beliefs about Ability

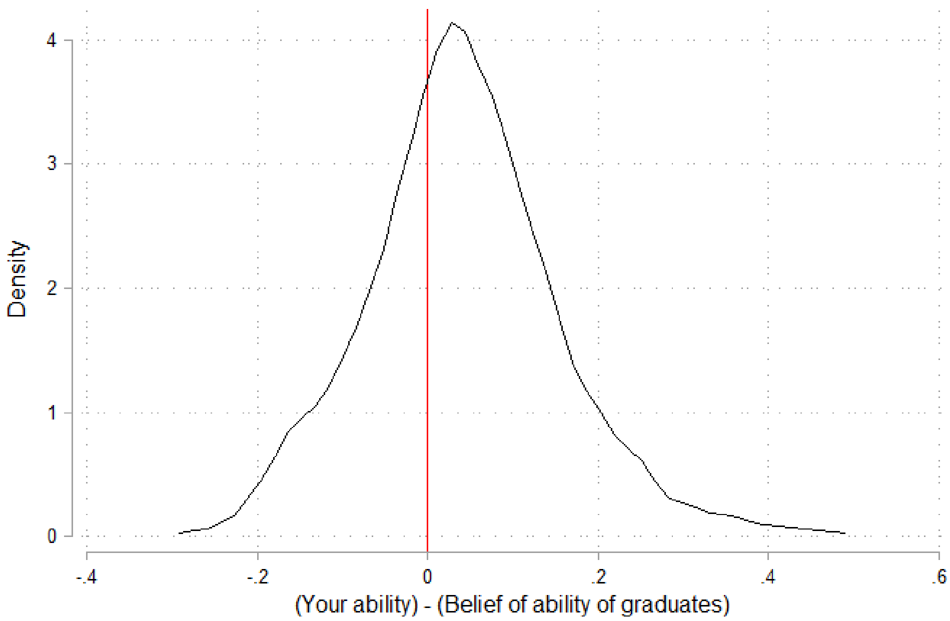
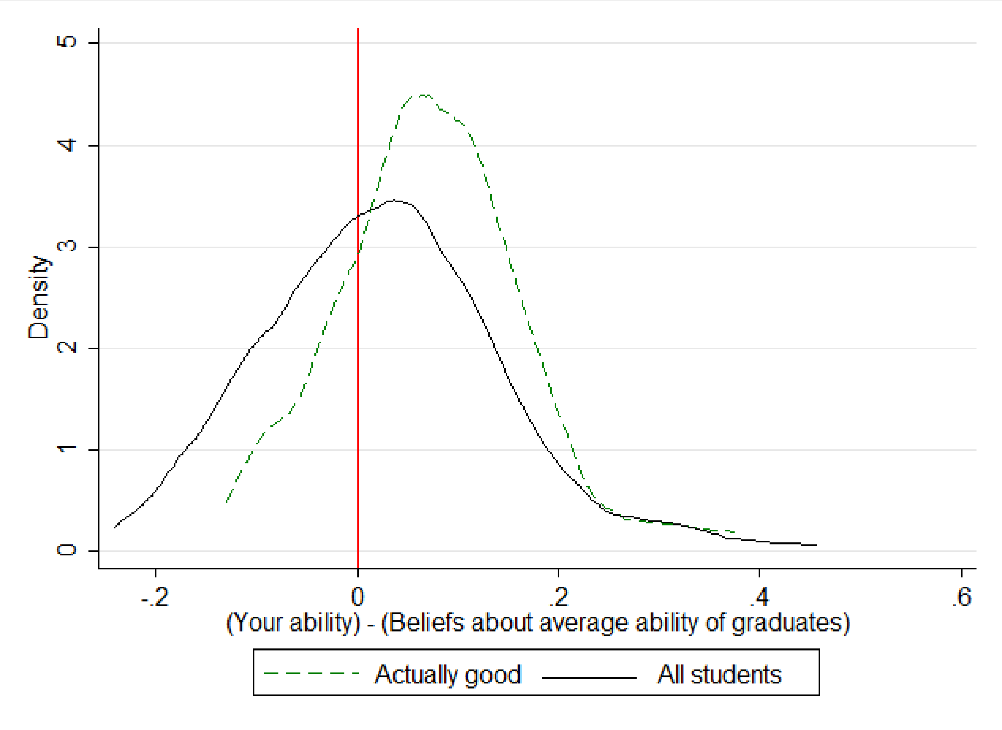


Figure 3: Beliefs about Ability (STEM Top Choice)



Knowing What it Takes: The Effect of Information About Returns to Studying on Study Effort and Achievement

Derek Rury and Scott E. Carrell

Abstract

We study the effect of providing students with information on returns to study effort in a large introductory microeconomics course. To do so, we use granular time-use data from the course's online homework module to estimate the association between study time and course performance. We use the same data as well as course outcome data to measure the impact of this information on several important outcomes, such as study effort as well as exam and overall course performance. We find that the treatment led to a 13% short-term increase in study effort, as well as a 9% increase in effort throughout the course. We find similar effects on homework scores. We also find that this long term effect on effort is mainly driven by students who originally under-estimated the returns to study effort. These students outperformed students who had over-estimated the returns to study effort both on measures of exam performance as well as overall course performance. We also see strong evidence that low-income students increased their study effort throughout the course, along with suggestive evidence of large effects on their exam and course performance.

1 Introduction

A student's study effort is a critical part of their education production function (Stinebrickner and Stinebrickner, 2004, 2008; Fraja et al., 2010; Bonesrønning and Opstad, 2015; Gneezy et al., 2019). As such, understanding how students make decisions on how much time to allocate towards studying is of great importance for education policy-makers. Study effort decisions contain important trade-offs for students, as increased time towards studying implies less time for other activities such as leisure and work (Stinebrickner and Stinebrickner, 2003; Bound et al., 2010, 2012; Metcalf et al., 2019). However, in order to make these trade-offs efficiently, students must know the actual returns to effort. That is, how study effort maps onto academic outcomes such as performance on exams and course grades.

Despite the central importance of these study effort decisions, the research is decidedly thin regarding the returns to study effort and how students make these choices. This is likely due to the fact that valid measures of study effort are both difficult to obtain and are endogenous to other factors affecting achievement. As a consequence, it is unclear how and whether students incorporate information about the returns to study effort into their beliefs and behavior, and whether changing those beliefs can lead to increases in academic achievement.

Previous work has also shown that students often have incorrect beliefs about their own education production function (Fryer, 2016; Ersoy, 2019a). Absent accurate information on their returns to study effort, students may over or under-invest in studying, leading to an inefficient allocation of time. While there is a growing literature examining information interventions in college classes, to our knowledge, no study has attempted to update students beliefs about the returns to study effort in a similar setting.

To fill this gap in the literature, we derive and administer an information intervention that both elicits and shocks students' beliefs about their returns to effort in an introductory microeconomics course. To obtain a valid measure of study effort to create the information treatment, we leverage historical, granular time-use data derived from the course's online homework application. After eliciting student's own beliefs about returns to study effort in a baseline survey, we randomly provide one-half of the students in the class information regarding the "actual" returns to effort. We then track subsequent study effort and course performance for students in the treatment and control.

Results show that the information contained in our intervention increased study effort in the short run (2-3 weeks after intervention) by 14% for all students and increased median homework score by 3 percentage points, or about 14% of a standard deviation. We also see positive, but diminishing, effects on time spent on homework assignments later in the course as well as large, but somewhat imprecise, effects on class attendance.

We also explore the role of beliefs about returns to study effort when estimating heterogeneous treatment effects. Results from our baseline survey indicate that about 80% of students underestimate their returns to study effort and we find meaningful differences in our treatment effects based on this broad characterization. For students who originally overestimated returns to study effort, we see large increases in study effort in the weeks directly after the intervention. However, the effects dissipate entirely when examining time spent on homework throughout the entire course and these students are *less* likely to attend class. In contrast, we see consistent positive effects on study effort and class attendance for those students who originally underestimated their returns to study effort, resulting in a near 10% increase in study time throughout the entire course.

We also find differential response to the treatment when examining heterogeneity by family income. Specifically, low-income students–those students most likely to face the trade-off between study and work time–in the treatment group significantly increase study effort throughout the entire course by 20%. In contrast, the treatment effect on study effort for higher-income students is smaller and short-lived, though higher income students in the treatment group respond through increased class attendance.

When translating these effects on study effort to course performance, we find small positive, but insignificant effects for the entire sample on homework scores and exam performance. However, for low-income students, the experimentally induced increase in study effort led to large gains in course performance, particularly on homework and exam scores in the weeks immediately following the treatment.

We cast these changes in beliefs and subsequent study effort decisions as products of opposing income and substitution effects under a binding student time constraint. Do students feel "richer" in their ability to achieve academically, or do they substitute more into studying upon learning they are more able? The resulting effect on study effort decisions is ambiguous and depends both on the initial beliefs as well as on the effect of performance feedback on beliefs. We outline this framework more explicitly in Section 2 and discuss its policy importance in our conclusion.

This paper makes contributions to several related literatures. First, we contribute to the research examining student effort decisions and the effect of study effort on academic achievement (Metcalf et al., 2019; Fraja et al., 2010; Stinebrickner and Stinebrickner, 2008; Ahn et al., 2019). This literature has primarily focused on how changing incentives for achievement affect study effort and subsequent performance (Hishleifer, 2016; Azmat and Iriberri, 2015; Golightly, 2020; Stinebrickner and Stinebrickner, 2008), while our paper manipulates beliefs about effort while keeping incentives for achievement fixed. Additionally, in a pair of papers most related to ours, Ersoy (2019b,a) uses data from an online language platform to demonstrate that students have incorrect beliefs about returns to study effort and shows that those beliefs become more accurate upon receiving information. In the spirit of Ersoy (2019a), our study focuses on updating students' beliefs in a classroom setting about the returns to effort. As such, we contribute to this literature by positing a framework that incorporates beliefs about returns to study effort into a study effort decision process in which students make trade-offs with their time.

Second, this paper contributes to the literature on performance feedback by examining how changes in beliefs translate into changes in achievement. Prior studies have found strong effects on achievement as a result of performance feedback (Azmat and Iriberri, 2010; Bandiera et al., 2015; Bobba and Frasinco, 2019b,a; Goulas and Megalokonomou, 2018; Brade et al., 2018; Gonzalez, 2017; Li, 2018), although these effects are not always positive (Azmat et al., 2019). These papers rightly interpret these effects as a result of changing beliefs about students' own abilities. What is less clear, however, is the mechanisms by which changes in beliefs translate into changes in achievement. As students learn about their ability, some inputs into the education production function must also change. The inputs most under the student's control, as well as those we believe most likely to be related to beliefs about ability, are those related to study effort. We aim to study this potential link and assess whether changes in beliefs about returns to study effort mimic findings from the research on beliefs about ability.

Third, we contribute to the literature examining beliefs, specifically in an education setting. A large literature has emerged over the past decade which demonstrates the importance of students' beliefs in decision making (Bobba and Frasinco, 2019a,b; Conlon, 2020; Wiswall and Zafar, 2015a,b; Zafar, 2011, 2013). To our knowledge, we are the first to document heterogeneous beliefs about returns to study effort in a common educational setting; a large introductory course at a selective public four-year university. We also demonstrate that our experimental results hinge importantly on ex ante beliefs about returns to study effort, depending on whether or not students over or underestimate returns to study effort.

Finally, we contribute to the literature studying "nudges" in education that attempt to alter student behavior via light-touch interventions, although with varying success (Damgaard and Nielsen, 2018; Li, 2018; Carrell et al., 2020; Oreopoulos et al., 2020). We show that our light-touch intervention changes behavior in a way that is consistent with a common decision framework used in economics, further demonstrating that nudges may yet play an important role in the classroom and economic research.

The rest of the paper is organized as follows; section 2 works through a simple framework that connects beliefs about returns to effort, achievement and behavior; section 3 provides details on our field experiment; section 4 describes our data; section 5 presents our results; section 6 concludes.

2 Beliefs, Effort and Achievement

To motivate our discussion of the role of beliefs in study effort decisions, consider the following utility function for a representative student. Following Ersoy (2019a), the student receives utility from both academic achievement, A, and leisure, L, and allocates their time, $\overline{T}$, over both study effort, e, and leisure, l

$$\max_{A,L} U(A,L) \tag{1}$$

$$s.t.\ e+l=\overline{T} \tag{2}$$

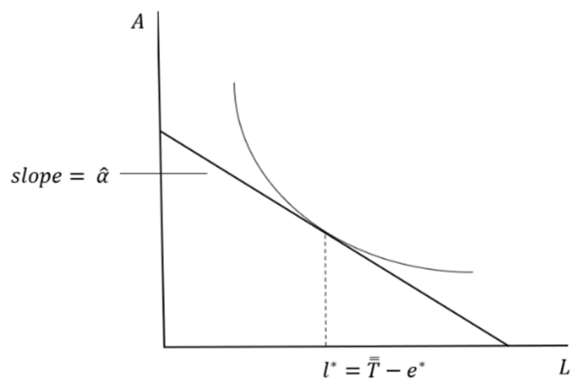
We also assume that study effort maps onto academic achievement linearly. Specifically, we assume that $A = f(e) = \alpha e$. Throughout this paper we will refer to this rate, α, as the "returns to study effort". For simplicity we also assume that $L = l$. After substituting into equation (1), the student's optimization problem becomes:

$$\max_{e,l} U(\alpha e, l) \tag{3}$$

$$s.t.\ e+l=\overline{T} \tag{4}$$

where e^* and l^* are the solution to the above problem. Lastly, we assume the utility function is strictly concave so that a unique solution exists and that marginal utility is decreasing for both arguments. Under this framework, the student faces a linear budget constrain in *time* with which they are endowed $\overline{T}$ and over which they allocate leisure and study effort. We represent the student's problem in Figure 1. using the familiar graph used in studying consumption decisions between two goods.

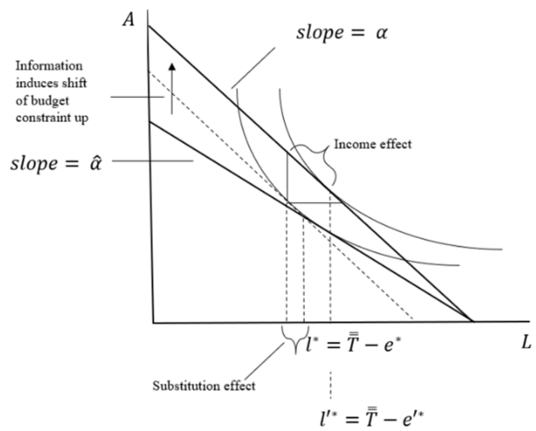
Figure 4: Utility maximization over study effort and leisure



Because effort maps onto achievement at a rate of α, the slope of the budget line is $-\alpha$. The linear axis represents both time spent on leisure, l, as well as time devoted to study effort, e, as $e = \overline{T} - l$. As in consumer theory, optimal effort and leisure are found where the student's indifference curve is tangent to the budget line, or more formally where $MU_A = MU_L$. Assume also that students do not know the value of α, but have beliefs, $\hat{\alpha}$, about its value. In this paper, we explore how e^* changes when students update their beliefs about $\hat{\alpha}$.

One way to frame how student behavior changes as beliefs about returns effort change is in terms of the "income" and "substitution" effects. Assume that a student holds beliefs about returns to study effort such that $\hat{\alpha} < \alpha$.[16] When the student is provided information on the true value of α, we assume the student fully updates their beliefs about their returns to study effort. This leads to a rotation of the budget constraint up the vertical axis, due to this relative "price" decrease. Similar to a relative price change in consumer theory, this rotation leads to a new equilibrium e^* and l^* resulting from a combination of the income and substitution effects as depicted in Figure 2.

Figure 5: Substitution and income effect from a shock to beliefs about the returns to effort



The substitution effect in this case will lead the student to study more, as academic achievement is now "cheaper". Likewise, the income effect makes the student study less, as they are now "richer" in time available and leisure is a normal good. As a result of these two opposing effects, the students overall change in study effort, e^*, is ambiguous. The same is true in cases where $\hat{\alpha} > \alpha$, although the income and substitution effects moves e^* in the opposite direction.

Connecting this framework to our study, our experiment elicits students' beliefs about $\hat{\alpha}$ and provides a randomly selected subset of students (e.g. the treatment group) with information about α. We

[16] We assume here that α is the same for all students but understand that, in reality, returns to study effort are likely to be heterogeneous. This would imply that $\alpha = \alpha_i$ for each student i. In our information intervention, we provide students with the average returns across a large sample of students from a previous course. In doing so, we make a trade off between offering specific information to students with offering feasible information in the form of average returns. In the end, our aim is less to provide individualized information to students but rather shock their beliefs about returns to study effort.

then examine the effect this information treatment on study effort and subsequent academic achievement. Importantly, we focus on how a student's initial beliefs of $\hat{\alpha}$ play in mediating these effects. In doing so, we are able to quantitatively estimate the relative importance of the income and substitution effects on students' study effort decisions as they learn about the returns to study effort.

Specifically, we estimate heterogeneous treatment effects based on ex-ante beliefs about returns to study effort. We do this by interacting whether students *under* or *over* estimated returns to study effort with treatment status. This allows us to address whether students who originally under (over) estimated their returns to study effort respond differently as a result of receiving the information treatment and whether the income or substitution effect dominates in this decision.

These results have import important implications for education policy as previous research has shown the potential for adverse effects from information treatments (Azmat et al., 2019; Beshears et al., 2015; Ringold, 2002). As mentioned above, it is ambiguous ex-ante whether students will study more or less upon learning the true valie of α.

3 Experimental Details

We administered our study effort experiment during the spring quarter of 2019 in two large introductory microeconomics courses with a total enrollment of over 700 students at a large selective public university on the west coast. We administered a *baseline survey* during the first week of class and short surveys at the beginning of each exam asking about time spent studying during the previous week. All surveys were completed by hand.

The baseline survey asked students questions regarding beliefs about their academic ability, preferences for majors, expected grade in the course, as well as beliefs about returns to study effort. Specifically, students were asked "how many hours do you think you would have to study per week to increase your grade by one letter?". The baselines survey also asked students to sign a Family Education Rights Protection Act (FERPA) to release their academic and demographic information from the university registrar.

The second survey was administered to all students in class prior to start of the first exam during the third week of the course. Two different surveys were randomly distributed to students: a *treatment* survey and a *control* survey. Both surveys asked questions regarding the amount of time spent studying for this class as well as other classes in the past week, followed by a paragraph of text and a short yes or no question, which was included to measure whether students actually read the paragraph. The paragraph in the *treatment* survey contained information on the returns to study effort. Specifically, students were told, "Using data from Prof. X's course last year, we found a significant relationship between the time students spent on homework and their course grade. Specifically, we found that for the average student, an additional three and a half hours of study time per week was associated with an improvement of a full letter grade in the course." The *control* survey contained a paragraph describing the benefits of participating in research on campus. The font and amount of text used in both the *treatment* and *control* surveys were designed so that the two surveys would appear identical at a quick glance.[17]

The second survey containing the information treatment appeared on the back of the first page of the exam. Exams were ordered such that treatment and control surveys alternated in their placement in order to provide an exogenous distribution of surveys. Teaching assistants handed out exams to students as they entered the classroom and took their seats. In section 5.2, we verify that assignment to treatment and control groups appears to be as good as random across student characteristics. Once the class began, students were given five minutes to turn over the first page of the exam and complete the survey. Students then turned the survey into the teaching assistants prior to the exam starting.

4 Data and Results

4.1 Data

A total of 456 students completed the baseline survey and 566 students completed the second survey. Our primary analytic sample consists of the 313 students who completed both surveys and signed the FERPA release, which represents just over two thirds of the baseline survey sample.[18]

We then matched these survey data to course administrative data containing our primary outcomes of interest, including, time spent studying, class attendance, homework scores and exam scores. Table 1 presents summary statistics for our outcome variables, survey results and background characteristics for the entire sample and separately for students by treatment status. Importantly, our data contains time-use data on each of the nine on-line homework assignments throughout the course. The time use data are measured in seconds, providing a granular measure of study effort. Specifically, these data measure the time spent between the moment when a student begins the homework assignment and either completes it or exits the homework module[19]. Homework scores are measured as the percentage of questions

[17]The *treatment* and *control* surveys can be found in the appendix.

[18]Though we cannot rule out selection into the sample, we note that students in the analytic sample performed similar to those in the entire class.

[19]We used the same time use data from the previous time the instructor taught the course, spring 2018, to create our information treatment. To do so, we regress time spent on homework on overall percentage points in the course. We find that three and a half hours of study time is associated with a statistically significant ten percentage point increase in the course, which corresponds to an increase in one letter grade. Not surprisingly, we also find a strong positive relationship in our study sample between measures of time spent on homework throughout the course and overall course performance. Specifically, we find that an increase of one unit in log time spent on homework assignments is associated with a statistically significant increase in

answered correctly. Since the homework module allowed for multiple attempts, we view homework scores as a measure of both effort and learning. We also measure the effect of our treatment on a course measure of classroom attendance, offering us the ability to study multiple dimensions of study effort.

We believe the information treatment to be most salient for study effort on the four homework assignments (homeworks two through five) that were due immediately following the information treatment and prior to students taking the second exam. That is, we believe time spent on these homework assignments to be our most valid measure of behavioral changes induced by the treatment. First, save for exams, there were no other activities other than homework for which students received course credit. Second, students are most likely to remember the information and incorporate it into their studying decisions immediately after receiving the treatment. Third, students will likely (endogenously) update their own estimates of $\hat{\alpha}$ after receiving new signals (e.g., follow-on exam performance) about the returns to study effort. Lastly, time spent studying is the variable over which students have the most control, while other outcomes such as exam and homework scores are the result of a mapping of effort onto achievement.

4.2 Balancing Tests and Descriptive Results

First, we perform checks to examine whether our treatment was, in fact, assigned exogenously. As mentioned above, teaching assistants distributed exams by handing out the top exam from their pile to students as they entered the room. While we acknowledge that this mechanism is not truly "random", we see no reason, ex ante, that assignment to treatment would be significantly correlated with observable or unobservable student characteristics.

Table 2 shows results when regressing treatment status on pre-treatment time spent studying and academic achievement, demographic characteristics, and responses to questions in the baseline survey. For statistical inference, and to address for multiple hypothesis testing, in these regressions and our main results, we follow Athey and Imbens (2017) and List et al. (2019) and use a bootstrap-based procedure for testing the null hypotheses using random sampling to assign treatment status. Hence, in addition to reporting traditional standard errors in parentheses, square brackets contain empirical p-values from randomization-based inference using a counterfactual of randomly assigning treatment status 500 times.[20]

Specifications 1-3 show results from separate regression for each of our three groups of pre-treatment

course performance of 2.9 percentage points (p value = 0.000). Converting these results into time units, the amount of study time per week associated with a ten percentage point increase found in our study sample ranges from 3.8 to 4.5 hours

[20] Athey and Imbens (2017) recommend the use of randomization-based inference in lieu of sampling-based inference for experiments. Additionally, as discussed by List et al. (2019), "by incorporating information about dependence ignored in classical multiple testing procedures, such as the Bonferroni (1935) and Holm (1979) corrections randomization-based inference has much greater ability to detect truly false null hypotheses."

variables (academic, demographics, and baseline survey), Specification 4 contains all of these variables in a single regression, and Specification 5 additionally adds teaching assistant fixed effects. Results indicate that treatment assignment is largely uncorrelated with all of our pre-treatment measures, with no coefficients statistically significant at conventional levels. As a further test, in Specifications 6-8 we present results when regressing treatment status on predicted measures[21] of time spent studying, homework score, and exam performance. Results show treatment status is not significantly correlated with the predicted homework measures, though we do find a negative and significant (p=0.068) relationship between treatment and predicted exam 1 performance. Specifically, students in the treatment group are predicted to score 1.5 percentage points lower on the first exam. Given this, in our outcome specifications we control for (pre-treatment) time spent on homework 1, homework 1 score, and exam 1 score.[22]

Next, Figure 4 shows the distribution of students' estimates of $\hat{\alpha}$ by examining responses to the question asking "How many hours a week do you think you need to study to increase your grade by one letter?". We see a wide dispersion across our sample, with a dramatic right skewness containing numerous high-value outliers. The median study hours students believe are required to increase one's grade by a full letter is six, almost twice as large as the estimate provided in the information treatment. This implies that most students in our sample (80%) overestimated the number of hours required to increase their grade by one letter. We categorize these students as underestimating the returns to study effort, as they believe it takes more hours than necessary to increase their grade. Conversely, we have far fewer students (20%) who overestimate their returns to study effort.[23].

Finally, as a fidelity check on whether students read the information paragraph in our intervention, we examine the responses to the yes or no question asked to students below the information paragraph in the second survey. Specifically, the treatment survey asked students if they found the information on returns to study effort useful, while the control survey asked if students wanted to learn more about research on campus. For those who were given the treatment text, we find that 91.3% of students answered "yes" when asked if found the information useful. In contrast, for those students given the control text, when asked if they would like to learn about participating in research, only 49.6% of students answered "yes". We see this as evidence that students not only read the treatment information provided to them carefully, but broadly speaking, they found it beneficial and were poised to incorporate

[21] We predicted these measures by regressing time spent on homework 1, homework 1 score, and exam 1 score on all of the pre-treatment demographic and survey response variables.

[22] Appendix Table A1 shows our main results with no controls.

[23] Following Ersoy (2019a), we also asked students about their beliefs regarding how much "control" they have over their ability. This question was meant to elicit beliefs about whether the students had what is called a "growth mindset" or whether they believed their ability was "fixed". Unfortunately, while we find this line of inquiry interesting, we found little variation in responses to this question, with only one student believing they had a below average amount of control over their ability.

the information into their beliefs.

4.3 Experimental Results

To study the effects of the information treatment on our outcomes of interest, we estimate the following statistical model:

$$y_i = \alpha + \beta TREAT_i + PT_i\gamma + X_i\phi + B_i\lambda + \psi + \varepsilon_i$$

where y_i is an outcome of interest (eg. time spent on homework), PT_i (pre-treatment) is a vector containing exam one and homework one scores as well as time spent studying on the first homework. X_i is a vector of background characteristics, including gender, race/ethnicity indicators, socio-economically disadvantaged, first-generation status, and SAT score. We also include responses from the baseline survey, found in B_i, which capture important beliefs about choice of major, ability in economics, expected grade in the course, as well as beliefs about typical study habits and beliefs about the returns to study effort. We control for teaching assistant fixed effects with ψ, to control for potential differences in the quality of teaching assistants. Lastly, ε_i represents a random error term. $TREAT_i$ represents an indicator for assignment to the treatment group. Random assignment to the treatment group ensures that $corr(TREAT_i, \varepsilon_i) = 0$, allowing us to estimate the causal effect of the information treatment on our outcomes of interest. For all results, traditional standard errors are in parentheses, while square brackets contain empirical p-values from randomization-based inference using a counterfactual of randomly assigning treatment status 500 times.

4.3.1 Average Treatment Effects

Results showing the average treatment effects for the entire sample are shown in Table 3. Odd numbered specifications include controls for (pre-treatment) time spent studying on homework 1, homework 1 performance, and exam 1 performance. Even numbered specifications add baseline survey responses, demographic controls, and TA fixed effects. Panel A presents results from time spent studying on the graded homework assignments and an indicator measure of class attendance.[24] For the time spent studying outcomes, we take the log of total time spent to approximate a percentage change in study time as a result of treatment. Panel B presents results for the percentage of points earned on the homework

[24] We measure class attendance using an app called Pocket Points. Specifically, the instructor in the course incentivized students to come to class by providing extra credit. Students earned 5 points (1 percentage point of the overall course grade) if they attended at least 18 hours (75%) of course lectures. To track attendance, Pocket Points requires students to be in class, open the app, and put their phone in sleep mode. The outcome is an indicator variable for whether the student met the 16 hour threshold.

assignments. Panel C presents results for exam performance.

Specifications 1 and 2 of Panel A show the treatment effect on time spent on homework two (the first assignment due after administering the information treatment). The coefficient of 0.129 in Specification 1 indicates that students in the treatment exhibited a 13% (45 minute) increase in study effort, relative to students in the control, with the effect significant at the 5% level. When adding baseline survey and demographic controls in Specification 2, the effect size increases only slightly (0.143) and remains significant at 5% level. We interpret these results as strong evidence that the students who received information about returns to study effort significantly increased their study effort in the period immediately following the intervention. In Specifications 3 and 4 of Panel A, we also see positive, but diminishing, effects on time spent when extending our outcomes to measure time spent on homework assignments later in the course. We explore and discuss what may be driving these diminishing effects when we present results based on beliefs about returns to study effort below. Finally, in Specifications 5 and 6 we see large, but somewhat imprecise, effects on class attendance. Specifically, results in Specification 5 indicate that students in the treatment group were nearly nine percentage points (16.4%) more likely to meet the class attendance incentive (p-values=0.104).

Next, in Panel B, examining homework scores, we also find positive and marginally significant treatment effects on the four homework assignments due in the weeks following the treatment.[25] Results in Specification 7 indicate that students in the treatment group showed a 3 percentage point (14 percent of a standard deviation) increase in homework score performance, relative to students in the control (p=0.088). Similar to the results on time spent, the effects become (slightly) smaller and less precise when examining performance on homework assignments later in the course.

In Panel C, we examine measures of exam performance. Though positive, the treatment effects are economically small and indistinguishable from zero.[26] These results are somewhat surprising, especially for exam 2, given the large treatment effects found in both study time and scores for homeworks two through five. Again, we consider why this may be the case when we study potential heterogeneous treatment effects by ex-ante beliefs and low-income status.

[25]When considering multiple homeworks, as is the case when estimating the treatment effect on scores on homeworks two through five or two through nine, our preferred measure of homework scores is median homework score. We take this approach for a specific reason. As part of the course design, along with their lowest exam score, each student's lowest homework scores is automatically dropped from their final course point total. This creates an incentive for students to skip one homework. Taking the median homework score helps us avoid measurement issues that arise from this incentive.

[26]For reasons similar to those regarding homework scores, our preferred measures of exam performance are percentage score on exam two (0 - 1 scale), as well as median percentage score on exams two through four (i.e. those after exam one, all which took place after the treatment was administered). Along with dropping their lowest homework score, students' lowest exam score was also dropped from their total course points total.

4.3.2 Heterogeneous Treatment Effects by Beliefs

As mentioned earlier, our baseline survey asked students, "How many hours would you need to study each week to increase your grade by one letter in this course?". This provides us with a measure of students' beliefs about the returns to their study effort, $\hat{\alpha}$, in this course. To explore heterogeneous treatment effects across beliefs about returns to study effort, we estimate separate treatment effects depending on whether student's over or underestimated the returns to study effort. Specifically, we categorize students as overestimating (underestimating) if their estimate of α was greater (less) than the three and a half hours contained in our information treatment. Connecting back to our framework and hypothesis from Section 2, interpreting the signs of the coefficients on homework time allow us to determine whether the income or substitution effect dominates, on average, as students in the treatment group update their beliefs about study effort. More specifically, if the coefficient for those who underestimated is positive (negative), and the coefficient for those who overestimated is negative (positive), then the substitution (income) effect dominates over the other.

Table 4 presents these results in the same format as Table 3. Panel A presents results from time spent studying on the graded homework assignments, Panel B presents results for the percentage of points earned on the homework assignments, and Panel C presents results for exam and course performance. Focusing on time spent on homeworks 2-5, we find large effects, though not statistically significant at conventional levels, for those who originally overestimated their returns to study effort. The treatment increased study effort by approximately 15% for these students. Notably, when examining time spent on homework throughout the entire course, these effects dissipate entirely and these students are *less* likely to attend class.

The pattern of results are noticeably different for those students who originally underestimated the returns to study effort. For our outcome measuring time spent on homeworks 2-5, we find a positive and significant effect of 12 to 14% and these effects largely persist when we examine study effort throughout the entire course. Additionally, we find a large and significant treatment effect of 15.5 percentage points (28%) on our measure of class attendance. Results in Panel B and Panel C, examining homework and exam scores are less pronounced, but show some evidence of increased effect size for student who underestimate returns to study effort.

These results are relatively consistent with a story in which students who originally underestimated returns to study effort substitute *into* studying after learning that academic achievement is *cheaper*. For these students, this implies that the substitution effect dominates the income effect, as the income effect would lead them to study less. Interpreting the results for those who originally overestimated returns is a bit less clear as the initial surge in study effort would signify a strong income effect, as students study more upon learning they are *richer* in achievement. We see that these effects do not persist throughout the course implying a potential substitution *out of* studying.

4.3.3 Heterogeneous Treatment Effects by Family Income

In Table 5, we estimate heterogeneous treatment effects for low and high income students. Our motivation stems from the fact that low-income students, nationwide, have been shown to have systematically lower levels of achievement, persistence, and graduation success (Bailey and Dynarski, 2011; Bound and Turner, 2002). Additionally, low-income student are most likely to face the potential trade-off between work ans study time. As such, we are interested in knowing whether providing information on the returns to study effort deferentially impacts these two groups.

Results for this analysis are mixed. While we see a large and significant treatment effect for low-income students on time spent studying, with a roughly 20% increase in study effort throughout the course (Specification 4), we find a negative, but insignificant, effect on attendance and negligible effects on homework scores. Importantly, for low-income students, we find evidence that the treatment led to increased exam performance, particularly on exam 2 as shown by the 5.2 percentage point treatment effect in Specification 12 (p=0.058).

In contrast, for high-income students, we find significant effects on short term time spent studying (Specifications 1 and 2), which dissipate throughout the course (Specifications 3 and 4). In addition, high income students in the treatment respond through increased class attendance, but show no overall improvements in homework or exam scores.

We take these results as suggestive evidence that updating beliefs about study effort has potentially larger benefits to low-income students.

4.3.4 Effect of Study Effort on Achievement

As previously discussed, our information treatment provided students with the correlation between time spent studying and academic achievement. As such, an important remaining question is whether or not increased study effort leads to a *causal* effect on academic achievement? This question is difficult to answer due to the likely engogeneity between study effort and academic performance. For example, students with higher ability may study less, on average, while still performing well on exams. Conversely, low-ability student may study more, on average, but perform worse on exams. In these scenarios, a simple regression of study effort on academic achievement will be biased downward, though we note that other scenarios may lead to bias in the opposite direction.

Fortunately, our experiment lends us the rare opportunity to answer this question. That is, because our experiment exogenously increased study effort, we can causally estimate the returns to study effort by instrumenting for time spent studying with treatment status.

Table 6 presents results from this analysis where we estimate two-staged least square models of time spent studying on homework and exam scores. Panel A presents results for the full sample, while the remaining panels present results for those subgroups where we found the largest response to treatment (students who underestimate returns and low-income students). Odd numbered specifications show OLS estimates, while even number Specifications show IV estimates. Across all outcomes and subgroups, we find a consistent pattern of substantially larger estimates (2-3 times) for the IV compared to OLS, suggesting a large negative bias in the OLS estimates. For example, the 0.099 OLS coefficient in Specifications 1 suggests that a 10 percent increase in time spent studying is associated with a nearly 1-percentage point increase in median homework score, whereas the IV estimate in Specification 2 shows this effect more than doubles at 2.4 percentage points.

Turning to the effects on exam performance, we note the differences between the OLS and IV are even larger, though the IV estimates are not precise for the full sample. Notably, for low-income students, we find large significant effects in both the OLS and IV estimates for exam 2 performance. The estimates suggest a 10 percent increase in study effort is associated with a 3.1 percentage point increase in exam 2 performance, which is significant at the 5-percent level.

5 Conclusion and Discussion

In this paper, we study the effect of providing students with information about returns to study effort. We measure the impact of receiving this information on several important outcomes such as study effort, class attendance, homework scores and exam performance. We are able to measure study effort using a granular measurement of effort based on time-use data from the course's online homework software.

We also administer a survey that captures students beliefs about returns to study effort. In doing so, we find that around 80% of students underestimate their returns to study effort. We also formulate a framework to test how students might respond to receiving information found in our intervention based on those ex-ante beliefs.

We find that our treatment led to a significant increase in study effort as well as homework performance, but the results are heterogeneous depending on initial beliefs about the returns to study effort and family income.

We find that those students in the treatment who had originally underestimated the returns to study effort studied more throughout the course and were more likely to attend class. On the contrary, those students who had originally overestimated returns to study effort increased their study effort immediately following the intervention, but this effect dissipated. Finally, we find large and persistent treatment effects on study effort for low-income students, with evidence of positive treatment effects on exam performance for this subgroup.

We view these results as being consistent with a story where students in the treatment group who originally underestimated the returns to study effort substituted into studying more for the class upon receiving information. On the contrary, those student in the treatment who originally overestimated the returns to study effort did not substantively change their effort during the duration of the course.

We believe these results speak to findings from the performance feedback literature, which finds that student achievement can be manipulated by providing students with information about how their performance compares to a relevant standard or peer group (Azmat and Iriberri, 2010; Bandiera et al., 2015; Bobba and Frasinco, 2019b,a; Goulas and Megalokonomou, 2018; Brade et al., 2018; Gonzalez, 2017; Li, 2018). As such, we demonstrate how similar results can be achieved through influencing beliefs about returns to study effort. Though, further research should attempt to measure beliefs about returns to study effort as a result of performance feedback to capture the relationship more formally.

Our results are also important for policy makers who wish to increase achievement and persistence in college. We document that the vast majority of students in our study underestimate the returns to study effort. Importantly, our results show that these students, on average, increase their study effort upon learning about the "true" returns to effort. However, depending on the distribution of beliefs within classrooms, our results show there may be unintended consequences as students with high beliefs of the returns to study effort could react by substituting out of study effort (in our case by attending class less).

Finally, our results suggest that providing students with information about previous cohorts studying patterns and their return to effort may provide gains in achievement at a low cost. Our findings compare favorably to other "light touch" interventions aimed at increasing academic achievement, particularly for low-income students.

6 Tables and Figures

Table 17: Selection Regressions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	TREAT	TREAT	TREAT	TREAT	TREAT	TREAT	TREAT	TREAT
	b/se/pvalues	b/se/pvalues	b/se/pvalues	b/se/pvalues	b/se/pvalues	b/se/pvalues	b/se/pvalues	b/se/pvalues
HW 1 time	.033			.005	.011			
	(.049)			(.056)	(.057)			
	[.520]			[.898]	[.828]			
HW 1 score	-.096			-.063	-.102			
	(.192)			(.215)	(.220)			
	[.642]			[.750]	[.614]			
Exam 1	-.336			-.187	-.168			
	(.262)			(.308)	(.311)			
	[.168]			[.532]	[.598]			
Female		-.010		-.041	-.028			
		(.061)		(.064)	(.065)			
		[.848]		[.532]	[.682]			
Low income		-.019		-.031	-.030			
		(.084)		(.086)	(.088)			
		[.812]		[.724]	[.742]			
African american		.073		.075	.091			
		(.185)		(.193)	(.194)			
		[.718]		[.716]	[.638]			
Hispanic		.096		.082	.130			
		(.103)		(.104)	(.106)			
		[.356]		[.420]	[.228]			
Asian		.070		.073	.087			
		(.077)		(.079)	(.079)			
		[.350]		[.356]	[.278]			
other		-.202		-.281	-.233			
		(.264)		(.270)	(.273)			
		[.526]		[.316]	[.440]			
SAT/ACT		-.000		-.000	-.000			
		(.000)		(.000)	(.000)			
		[.372]		[.784]	[.794]			
First generation		.026		.029	-.006			
		(.077)		(.078)	(.081)			
		[.730]		[.70]	[.918]			
Study habits			-.011	-.013	-.015			
			(.010)	(.010)	(.010)			
			[.254]	[.192]	[.158]			
Returns to study effort			.012	.013	.014			
			(.008)	(.009)	(.009)			
			[.156]	[.136]	[.110]			
Econ top major choice			.044	.015	.022			
			(.065)	(.069)	(.069)			
			[.516]	[.834]	[.740]			
High control over ability			-.482	-.513	-.580			
			(.501)	(.542)	(.545)			
			[.668]	[.508]	[.350]			
Medium control over ability			-.402	-.424	-.520			
			(.503)	(.542)	(.546)			
			[.954]	[.684]	[.468]			
Expected grade A			-.079	-.074	-.073			
			(.066)	(.071)	(.072)			
			[.234]	[.284]	[.290]			
High ability in econ (belief)			.013	.048	.034			
			(.083)	(.089)	(.090)			
			[.894]	[.600]	[.710]			
Predicted HW 1 Study Time						.148		
						(.093)		
						[.104]		
Predicted HW 1 Score							-.224	
							(.356)	
							[.558]	
Predicted Exam 1 Score								-.772*
								(.403)
								[.068]
Observations	313	313	313	313	313	313	313	313
Pre-scores	Y	N	N	N	Y	Y	N	N
Demos	N	N	Y	N	Y	Y	Y	Y
Beliefs	N	N	N	N	Y	Y	Y	Y
TA FE	N	N	N	N	N	Y	Y	Y

Standard errors in parentheses; Empirical P values in brackets

Note: Statistical significance determined by empirical P values

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 18: Summary statistics

	(1) Mean	(2) Mean (TREAT=1)	(3) Mean (TREAT=0)
HW 1 time (log)	5.49	5.51	5.46
	(.624)	(.633)	(.616)
HW 1 score	.864	.857	.872
	(.166)	(.171)	(.160)
Exam 1	.774	.763	.785
	(.118)	(.126)	(.108)
Female	.588	.588	.588
	(.493)	(.494)	(.494)
Low income	.256	.275	.235
	(.437)	(.448)	(.426)
African American	.029	.031	.026
	(.167)	(.175)	(.160)
Hispanic	.220	.250	.190
	(.415)	(.434)	(.393)
Asian	.521	.525	.516
	(.500)	(.501)	(.501)
Other	.013	.001	.020
	(.113)	(.063)	(.139)
SAT/ACT	1239	1228	1250
	(302)	(300)	(305)
First generation	.377	.413	.340
	(.485)	(.494)	(.475)
Study habits	5.30	5.27	5.33
	(3.73)	(3.54)	(3.93)
Returns to study effort	6.91	7.23	6.57
	(4.49)	(4.59)	(.4.37)
Econ top major choice	.284	.306	.261
	(.451)	(.462)	(.441)
High control over ability	.776	.744	.810
	(.417)	(.438)	(.393)
Medium control over ability	.220	.250	.190
	(.415)	(.434)	(.393)
Expected grade A	.281	.237	.327
	(.450)	(.427)	(.471)
High ability in econ (belief)	.150	.150	.150
	(.358)	(.358)	(.359)
HW 2-5 time	6.20	6.27	6.12
	(.564)	(.588)	(.536)
HW 2-9 time	6.66	6.71	6.61
	(.564)	(.588)	(.536)
HW 2-5 median score	.816	.822	.809
	(.211)	(.212)	(.211)
HW 2-9 median score	.839	.843	.834
	(.199)	(.200)	(.198)
Exam 2 score	.746	.743	.750
	(.145)	(.153)	(.137)
Exam 2-4 median score	.712	.709	.716
	(.115)	(.118)	(.112)
Observations	313	160	153

Standard deviation in parentheses

Table 19: The effect of treatment on homework time and scores, class attendance as well as exam and course performance

Panel A: Effects on Time Spent on Homeworks						
	(1) HW 2-5 time	(2) HW 2-5 time	(3) HW 2-9 time	(4) HW 2-9 time	(5) Attendance	(6) Attendance
TREAT	.129**	.143**	.073	.081	.089	.078
	(.055)	(.057)	(.052)	(.054)	(.056)	(.056)
	[.032]	[.018]	[.156]	[.128]	[.104]	[.174]
Panel B: Effects on Scores on Homeworks						
	(7) Median HW score (2-5)	(8) Median HW score (2-5)	(9) Median HW score (2-9)	(10) Median HW score (2-9)		
TREAT	.030*	.034*	.023	.027		
	(.018)	(.018)	(.017)	(.018)		
	[.088]	[.074]	[.202]	[.150]		
Panel C: Effects on Exams And Course Performance						
	(11) Exam 2 Score	(12) Exam 2 Score	(13) Median exam Score (2-4)	(14) Median exam Score (2-4)		
TREAT	.005	.004	.007	.010		
	(.015)	(.014)	(.012)	(.011)		
	[.706]	[.798]	[.516]	[.392]		
Observations	313	313	313	313	313	313
Pre-scores	Y	Y	Y	Y	Y	Y
Demos + Beliefs	N	Y	N	Y	N	Y

Standard errors in parentheses; empirical p values in brackets

Note: Statistical significance is based on empirical p values. Pre-scores, demographics and beliefs controls are based on the relevant columns in table 1 (eg. pre-scores are column one)

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 20: Heterogeneous effects by beliefs of treatment on homework, exam and course outcomes

Panel A: Effects on Time Spent on Homeworks						
	(1)	(2)	(3)	(4)	(5)	(6)
	HW 2-5 time	HW 2-5 time	HW 2-9 time	HW 2-9 time	Attendance	Attendance
Over Estimate Returns	.158	.147	.009	.005	-.168	-.160
	(.122)	(.124)	(.117)	(.119)	(.124)	(.122)
	[.130]	[.204]	[.944]	[.956]	[.176]	[.190]
Under Estimate Returns	.123*	.138**	.093	.098	.155**	.144**
	(.062)	(.064)	(.060)	(.062)	(.063)	(.063)
	[.058]	[.026]	[.120]	[.118]	[.020]	[.016]
Panel B: Effects on Scores on Homeworks						
	(7)	(8)	(9)	(10)		
	Median HW score (2-5)	Median HW score (2-5)	Median HW score (2-9)	Median HW score (2-9)		
Over Estimate Returns	.035	.033	.032	.031		
	(.040)	(.040)	(.038)	(.039)		
	[.406]	[.430]	[.446]	[.488]		
Under Estimate Returns	.032	.0370*	.020	.027		
	(.020)	(.021)	(.019)	(.020)		
	[.114]	[.084]	[.280]	[.180]		
Panel C: Effects on Exams And Course Performance						
	(11)	(12)	(13)	(24)		
	Exam 2 Score	Exam 2 Score	Median exam Score	Median exam Score		
Over Estimate Returns	-.016	-.010	.011	.014		
	(.033)	(.031)	(.026)	(.025)		
	[.682]	[.808]	[.692]	[.570]		
Under Estimate Returns	.008	.006	.005	.008		
	(.017)	(.016)	(.013)	(.013)		
	[.626]	[.686]	[.710]	[.536]		
Observations	313	313	313	313	313	313
Pre-Treatment	Y	Y	Y	Y	Y	Y
Demos + Beliefs	N	Y	N	Y	N	Y

Standard errors in parentheses; empirical p values in brackets

Note: Statistical significance is based on empirical p values. Pre-scores, demographics and beliefs controls are based on the relevant columns in table 1 (eg. pre-scores are column one)

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 21: Heterogeneous effects by beliefs of treatment on homework, exam and course outcomes

Panel A: Effects on Time Spent on Homeworks						
	(1) HW 2-5 time	(2) HW 2-5 time	(3) HW 2-9 time	(4) HW 2-9 time	(5) Attendance	(6) Attendance
Low Income (Treat)	.212*	.258**	.154	.201*	-.051	-.091
	(.108)	(.111)	(.104)	(.106)	(.110)	(.109)
	[.088]	[.036]	[.168]	[.082]	[.642]	[.408]
High Income (Treat)	.097*	.102*	.043	.039	.136**	.137**
	(.063)	(.066)	(.061)	(.063)	(.065)	(.065)
	[.096]	[.088]	[.444]	[.516]	[.032]	[.044]
Panel B: Effects on Scores on Homeworks						
	(7) Median HW score (2-5)	(8) Median HW score (2-5)	(9) Median HW score (2-9)	(10) Median HW score (2-9)		
Low Income (Treat)	.016	.042	.022	.0463		
	(.035)	(.036)	(.034)	(.035)		
	[.638]	[.286]	[.530]	[.200]		
High Income (Treat)	.034	.031	.021	.020		
	(.021)	(.021)	(.020)	(.021)		
	[.118]	[.156]	[.312]	[.352]		
Panel C: Effects on Exams And Course Performance						
	(11) Exam 2 Score	(12) Exam 2 Score	(13) Median exam Score	(24) Median exam Score		
Low Income (Treat)	.052	.052*	.032	.039		
	(.029)	(.027)	(.023)	(.022)		
	[.104]	[.058]	[.188]	[.122]		
High Income (Treat)	-.010	-.013	-.001	.000		
	(.017)	(.016)	(.014)	(.013)		
	[.492]	[.394]	[.950]	[.994]		
Observations	313	313	313	313	313	313
Pre-Treatment	Y	Y	Y	Y	Y	Y
Demos + Beliefs	N	Y	N	Y	N	Y

Standard errors in parentheses; empirical p values in brackets

Note: Statistical significance is based on empirical p values. Pre-scores, demographics and beliefs controls are based on the relevant columns in table 1 (eg. pre-scores are column one)

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 22: OLS and IV results of study effort on achievement

	(1) (OLS) HW 2-5 score	(2) (IV) HW 2-5 score	(3) (OLS) HW 2-9 score	(4) (IV) HW 2-9 score	(5) (OLS) Exam 2 score	(6) (IV) Exam 2 score	(7) (OLS) Exam 2-4 median	(8) (IV) Exam 2-4 median
Panel A: All Students								
HW time	0.099***	0.239*	0.104***	0.333	0.018	0.025	0.027**	0.117
	(0.018)	(0.127)	(0.019)	(0.243)	(0.015)	(0.093)	(0.012)	(0.135)
Observations	313	313	313	313	313	313	313	313
Panel B: Under-estimators								
HW time	0.096***	0.308*	0.086***	0.249	0.015	0.056	0.016	0.071
	(0.021)	(0.180)	(0.020)	(0.219)	(0.016)	(0.116)	(0.013)	(0.128)
Observations	247	247	247	247	247	247	247	247
Panel D: Low-income Students								
HW time	0.122***	0.097	0.174***	0.183	0.060**	0.313**	0.085***	0.234
	(0.042)	(0.144)	(0.040)	(0.173)	(0.029)	(0.155)	(0.026)	(0.145)
Observations	80	80	80	80	80	80	80	80

Standard errors in parentheses

Note:All models include demographic and beliefs controls as well as TA fixed effects. Models 1-2 and 5-6 use time spent on HW 2-5 2-5 as main explanatory variable while models 3-4 and 7-8 use time spent on homeworks 2-9

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The Economic Impacts of Migrants from Hurrican Maria[27]

Giovanni Peri, Derek Rury and Justin C. Wiltshire

Abstract

We examine the economic impact of the large migration of Puerto Ricans to Orlando after Hurricane Maria in 2017. Using a synthetic control approach, we find that employment in Orlando increased, especially in construction and retail, and find positive aggregate labor market effects for non-Hispanic and less-educated workers. While we find that earnings for these workers decreased slightly in construction, this was balanced by earnings growth in retail and hospitality. These results are consistent with small negative impacts on earnings in sectors exposed to a labor supply shock, offset by positive effects in sectors impacted by an associated positive demand shock.

[27]Derek Rury and Justin Wiltshire gratefully acknowledge financial support for this project from the Center for Growth and Opportunity. We thank seminar participants at Simon Fraser University, UC Davis, and the 2019 EEA Annual Conference for useful comments on early versions of this paper, and particularly thank Marianne Bitler, Brendan Price, and Geoffrey Schnorr for helpful comments. Any errors or omissions are the sole responsibility of the authors. Addresses: Giovanni Peri, Department of Economics, UC Davis, One Shields Avenue, email:gperi@ucdavis.edu; Derek Rury, Department of Economics, UC Davis, One Shields Avenue, email:drury@ucdavis.edu; and corresponding author Justin C. Wiltshire, UC Davis, One Shields Avenue, email:jcwiltshire@ucdavis.edu. The online appendix can be found at https://justinwiltshire.com/research-1

1 Introduction

In September 2017 Hurricanes Irma and Maria struck Puerto Rico in rapid succession, bringing catastrophic loss in property and lives to hundreds of thousands. In the months following the hurricane, an estimated 44,000 individuals rapidly fled Puerto Rico for the U.S. mainland. This paper examines the short-run economic impact of this sudden migration event on Orlando, the city which, because of pre-existing network connections, received a plurality of these refugees. We study the impact of this event on the aggregate Orlando labor market and we examine whether there is evidence of effects on employment and earnings for incumbent workers, as well as which sectors were affected and whether natives were crowded out of the labor market. To do so, we employ a synthetic control approach using a data source that provides virtually-complete coverage of county-level employment outcomes at a high frequency throughout our study period. Our results are consistent with a story in which migrants place modest downward pressure on earnings for natives in sectors most exposed to the new migrant labor, while having positive effects on employment and earnings for workers in sectors which meet the consumer demand of migrants.In aggregate the effects on wages and employment were very small.

Several papers have analyzed the local labor market effects of sudden waves of immigration. Examples include the analysis of the inflow of Soviet Jews to Israel during the 1990s (Cohen-Goldner and Paserman, 2011); the inflow of Syrians to Turkey in 2013-15 (Ceritoglu et al., 2017); and the inflow of Algerians to France in the 1950's (Hunt, 1992)). The only case studied in the U.S.—potentially *over*-studied, given the limited amount of data available and the fact that the event now happened forty years ago—is the Mariel Boatlift.[28] This was an episode in which approximately 100,000 people fled from Cuba, and many of them relocated to Miami. While early studies established that this large inflow had no impact on local wages even in the short run Card (1990), recent re-analyses of the event have generated some disagreement on the effects, especially with respect to the impact on the small subgroup of male, less-educated, native workers (Borjas, 2017; Peri and Yasenov, 2019; Clemens and Hunt, 2019). The details of the controversy are centered on measurement, sample choice, and methods used.

For several reasons, the event we study here while somewhat smaller in scale, seems more relevant than the Mariel Boatlift for informing our current understanding of the local labor market impact of immigration in the U.S. First, the episode is much more recent, and the economy of Orlando in 2018 is more comparable to current U.S. metropolitan areas than was the economy of Miami in 1980. Second,

[28] Another weather event similar to Maria, Hurricane Katrina, has also been studied, but most of the papers focus on the impact on the *out*-migrants who fled New Orleans and the surrounding area (rather than on people in the receiving communities). De Silva et al. (2010) look at the impact on wages of a large relocation of evacuees on Houston, and find that there was an aggregate, small wage depression of 0.7 percent in low-skill industries. Compared to De Silva et al. (2010), our paper uses a more recently developed and, we argue, more appropriate estimation strategy to test a similar set of outcomes.

the impact of this episode is a consequence of migration in response to an extreme weather event—an occurrence likely to become more frequent in the future. Third, the case of immigration from Puerto Rico involves people who have similar levels of education to native mainland U.S. workers, while the Cubans in the Mariel Boatlift (Marielitos) had much lower levels of education. This means that the Hurricane Maria migration closely resembled the sort of immigration the U.S. has received in the last 20 years (i.e. 2000-2018)—which included a large share of college-educated immigrants—while the low-skilled immigration of Marielitos was more similar to the immigration of the 1990s. We also note that, similarly to the Cuban Marielitos, immigrants from Puerto Rico had immediate access to the U.S. labor market and other benefits available to U.S. citizens. This implies that their potential crowding-out impact on native workers would be at a maximum. Hence our findings of no crowding-out and of small overall wage effects, positive in certain sectors and negative in others, are particularly compelling.

We employ the most current 'synthetic control' estimating strategy (similarly to Peri and Yasenov (2019)) first proposed by Abadie and Gardeazabal (2003), developed in Abadie et al. (2010, 2015), and subject to recent important improvements.[29] An additional advantage of our study, relative to the Mariel Boatlift literature—which relies on small, weighted samples of survey data from the Current Population Survey (CPS)—is that we use data from the Quarterly Census of Employment and Wages (QCEW) and the Quarterly Workforce Indicators (QWI). These are county-by-industry-level administrative data covering 95+% of all workers, observed on a quarterly basis.[30] The detail and precision of our analysis is therefore significantly greater than in those studies. Moreover, we have higher frequency (quarterly) data for this episode and we focus much more specifically on the short-run effects. Additionally we analyze the impact on number of firms—an important and under-researched margin of adjustment to immigration (recently, Beerli et al. (2018) show the importance of firms' response to increased availability of immigrants for understanding labor market impacts). Finally, through the analysis of different sectors we separate likely supply and demand effects of the immigrants' wave (the previous standard for this type of analysis was Bodvarsson et al. (2008)). There are limitations, in our case, relative to the Mariel Boatlift. First, the inflow of refugees in Orlando was smaller relative to the local labor force, second it likely had a larger temporary component as some of them likely returned in Puerto Rico one year later. However, as our much larger sample reduces the measurement error and we only analyze short-run effects, in spite of these two limitations, we are able to find significant and precisely estimated effects.

We consider data from many different sources to gauge an idea of the size and destinations of the flow of Puerto Ricans refugees. Examining data from FEMA applications for assistance related to the

[29] Readers are pointed to Abadie (Forthcoming) for an excellent current review of the synthetic control literature, which is continually expanding.

[30] Employment is observed on a monthly basis in the QCEW. The other variables are observed quarterly.

hurricanes and other data on social media usage, school enrollments, and air travel of Puerto Ricans in the months after Hurricane Maria hit Puerto Rico, we establish that the Orlando Metropolitan Area received significantly more people from the Island than did any other city. Orlando was also the area with largest pre-existing community of Puerto-Ricans, relative to its population. A handful of other cities also receiving non-trivial numbers.[31] In the months from September 2017 to March 2018 we estimate that about 24,000 Puerto-Ricans ended up in Orlando. Some of them may have left subsequently, during 2018. This represents around 1 percent of Orlando's pre-hurricane population. Both in levels and as a percentage of the population, Orlando received many more refugees than the next most-impacted metropolitan areas (Lakeland-Winter Haven, FL and Springfield, MA). We therefore consider it as the treated unit in our analysis, when performing a synthetic control analysis. In several robustness checks we find similar results when considering the average effects on the three- and five- cities receiving highest immigrant numbers as share of population (most treated cities). Those cities were also among the 10 with largest pre-existing network of Puerto-Ricans before the hurricane. Ultimately, the sudden out-migration of Puerto Ricans from the Island and into Orlando, as a result of Hurricane Maria, constitutes a good approximation to a *natural experiment*—namely a sudden and unexpected event, uncorrelated with conditions in the Orlando local economy.

The period of analysis post-event is between September 2017 and September 2018. We look first at *aggregate* effects on employment and earnings per person, then shift our focus to specific sectors—namely the *construction*, *retail* trade, and accommodations & food services sectors (hereafter, *hospitality*). We analyze first all workers, then specifically non-Hispanic and less-educated subgroups (those with a high-school degree or less education) as approximation, respectively, of the incumbent native-worker population (Orlando non-Hispanic residents are likely to be U.S.-born) and of those who may be in competition with the new Puerto Rican arrivals who, due to initial downgrading, took largely manual and less-skilled jobs.

Going into the the sector-specific dimension in the short-run helps us to identify potential effects of the immigrant wave on labor demand and labor supply. We make the case that the construction sector was the most likely to experience a sudden increase in labor supply in the short run. A large share of Hispanics were employed in the construction sector before the hurricane, and most construction jobs require limited use of the English language—which less than 40% of Puerto Rican residents spoke "well". Moreover, while increased demand for new buildings may eventually follow the arrival of immigrants, we see no evidence that the housing market was affected in the short run, hence the construction sector

[31] We take Orlando to be the five counties which comprise the Orlando Commuting Zone: Lake, Orange, Osceola, and Seminole—which comprise the Orlando Metropolitan Area—plus Sumter County. We also perform analysis using the other cities that received significant inflow

likely received a larger labor supply than labor demand shock. On the other hand, the retail and hospitality sectors were likely to experience positive demand shocks in the short run, as new additions to a local population require accommodations, retail goods, and food services. While some immigrants were employed in these sectors, the short-run impact of the new arrivals should mainly reflect an increase in demand for retail and hospitality workers, resulting from the increased consumer demand in those sectors. By quantifying the effect on employment and wages in these two sectors, we learn about the potential labor supply and labor demand shifts produced by immigration.

Additionally, we consider the response of firm establishments in the area. The speed and magnitude of the response by firms to this expansion is relevant for understanding the mechanisms through which the labor supply shock was absorbed by the local economy in the short run. While our estimated effects are all positive, their pre-treatment trends are noisy enough that none of them reach a conventional level of significance according to our test statistics.[32] The overall evidence is suggestive of a positive firm-creation response, but not large enough to be significant in the short-run.

Broadly, our results are consistent with a story in which Puerto Rican refugees in Orlando constituted a labor supply shock concentrated in the construction sector, as well as a labor demand shock driven by demand of goods and services, especially visible in the retail and hospitality sectors. While we do find evidence of a negative impact on construction sector wages of natives and less-educated workers 12 months after the migration event (-2.5% and -1.5%, respectively), we find symmetric evidence that retail earnings increased for natives and less-educated workers (+2.1% and +1.2%, respectively) over that same period.[33] Looking across all sectors and all workers, despite the sudden influx of tens of thousands of new workers into the area, we find no evidence of any negative impact on wages (the point estimates are, in fact, positive and not significant), suggesting that the sector-specific positive wage effects more than offset the negative ones in the aggregate. This aggregate balancing effect holds for native and less-educated workers as well. In fact, we also estimate a 0.8% increase in aggregate employment for each of those groups 12 months after the hurricane. These results are broadly robust to several controls and specifications, and suggest that this large inflow of workers was quickly absorbed into and grew the local economy.

The rest of this paper proceeds as follows: in Section 2 we describe the demographic and employment trends of immigrants from Puerto Rico in the mainland United States, then we provide a measure of the immigrant wave and justification for focusing on Orlando as the (most intensely) treated unit.

[32] Our 12-month estimate of the impact on construction establishments is significant according to our moving block p-value, but not according to our more conservative test statistics.

[33] The construction sector employed between 5 and 6.5% of Orlando workers over the period, while the retail sector typically employed between 12 and 13%.

Section 3 describes our data and empirical methodology. Section 4 describes the results from our analyses, and presents a series of robustness checks—including a test for whether our results were impacted by Hurricane Irma's effects on Orlando, estimates using three or five metropolitan areas, rather than Orlando only and an analysis of housing market conditions before and after the migrants' arrival. Section 5 concludes.

2 Characteristics and Trends of Puerto-Rican Immigration

2.1 Historical Migration and Characteristics of Migrants

During the first half of the 20th century and for several decades thereafter, rates of migration from Puerto Rico to the mainland remained lowThe Puerto Rican population on the mainland was an estimated 1,513 in 1910 and 226,110 in 1950, respectively constituting 0.002 percent and 0.15 percent of the total population in those years.. During the later half of the 20th century and continuing into the 21st, however, Puerto Rican migration to the mainland United States increased significantly. As of 2015 an estimated 5.4 million Puerto Ricans resided in the mainland U.S. (constituting 1.6 percent of the U.S population). This number is particularly striking when compared to the total population of 3.5 million people residing in Puerto Rico as of 2018 (Whalen and Vazquez-Hernandez, 2005). During these years, large Puerto Rican communities formed in New York City, in Philadelphia and Chicago. In terms of share of the local population, however, the largest communities were in Florida, and Orlando in particular emerged as the biggest pole of attraction. Migration followed the usual pattern, with new immigrants settling in areas with relatively large Puerto Rican communities.

In recent decades, the Orlando area became even more important as destination for Puerto-Ricans (see online Appendix Figure A1[34]). Puerto Ricans' full access to labor opportunities in the mainland United States suggests that the availability of jobs in the U.S. may have been a key driver of mobility from the island to the mainland.

Table 23 shows differences in average characteristics of Puerto Rican migrants who arrived in the year preceding Hurricane Maria, compared to reference groups.[35] This gives an idea of the selection prevailing among Puerto Rican migrants right before the hurricane. The first column shows the difference with US natives, the second with residents of Puerto Rico, the third with US natives in Florida.[36].

[34] The figures are constructed using American Community Survey from 2005 to 2017, downloaded from IPUMS USA (Ruggles et al., 2021).

[34] Migrants coming to the United States are often positively "selected" from their country in terms of education, mainly because of the large skill premium paid in the U.S. relative to countries of origin (Grogger and Hanson, 2011). Borjas (2008), however, documented negative selection of pre-2000 Puerto-rican immigrants, in terms of education.

[35] Individuals are restricted to those who are in the labor force.

[36] A negative value implies that the Puerto Rico migrants have a smaller value for that variable relative to the comparison

The largest difference between recent Puerto Rican immigrants and US natives is in their average age, as Puerto Rico immigrants are significantly younger and less likely to be married than natives. They are also more likely to be male. On the other hand, education levels are similar. While the difference in average years of schooling compared to U.S. natives is statistically significant, compared to the average Florida native it is *not* significant and it is small—equal to 0.38 and 0.13 years of schooling, respectively. There is no significant difference in the share of college-educated immigrants from Puerto Rico and natives in the U.S. or in Florida. The inflow can therefore be characterized as mostly young, with an education composition similar to that of natives.

2.2 Size of the migration wave After Hurricane Maria

Early estimates put the number of Puerto Ricans who fled to Florida in the months after September 2017 at around 200,000 people. Later estimates, based on Facebook data, substantially revised this number down, to 44,000 people who moved from Puerto Rico to Florida *and stayed* there until at least March 2018 (Alexander et al., 2019). Estimates based on flight passenger lists confirm between 30,000 and 50,000 people moved out of Puerto Rico to Florida in the months after the hurricane (Rayer, 2018). Additionally, Ozek (Forthcoming) documents that 12,000 children, who were Maria refugees registered for school in Florida after the hurricane[37] We complement these estimates with additional evidence to get an idea of the distribution of those evacuees across the US. Data from applications for disaster relief from the Federal Emergency Management Agency (FEMA) are useful to gather a geographic distribution of evacuees. These application data, obtained through a Freedom of Information Act (FOIA) request, represent claims made to FEMA to obtain disaster relief funds, filed by people who had a home in Puerto Rico which was damaged by Hurricanes Irma and/or Maria. By looking at the ZIP code of residence before the hurricane and the ZIP code at the time of filing (after Hurricane Maria occurred), we can identify the likely residence of these Puerto Ricans on the mainland in the months after the hurricane. The distribution of these applications (aggregated to commuting zones and shown in Figure A2 in the online Appendix, in both level and per capita terms) reveals that these FEMA applications were heavily concentrated in relatively few areas. Especially in per capita terms, the Orlando area exhibits by far the largest concentrations: with nearly 4,000 applications, Orlando had more than two-and-a-half times as many applications as the next most heavily affected commuting zones (Fort Lauderdale-Miami and New York-Nassau-Suffolk).

group.

[37] Analysis of ACS/PRCS data from the U.S. Census Bureau puts the number of net out-migrants from Puerto Rico in the year between July 2017 and July 2018 at 123,399 (U.S. Census Bureau, 2018), and the Center for Puerto Rican Studies estimates this number to be more than 135,000 (Centro, 2018). These number may include people who migrated before and other that returned, hence may not be too accurate

Another approach to measuring the size of the inflow to Orlando from Puerto Rico would be to compare data from the 2017 and 2018 American Community Survey (ACS) to track the number of people who moved between Puerto Rico and the mainland during this time. The Census Bureau itself notes that there are significant issues with this approach and the ACS figures should be treated with caution (Schachter and Bruce, 2020).[38] For example, obtaining survey responses from Puerto Rican migrants right after the hurricane was likely very difficult for several reasons, including the fact that they were often not technically eligible for participation in the ACS due to the survey's two-month residency requirement. Moreover, the hurricane didn't hit Puerto Rico until September of 2017, and those who migrated and stayed in Florida for several months may have returned at any point in 2018.[39] These factors make the ACS microdata quite imprecise in measuring the size of the migration event. Despite these issues, we show in Figure A3 in the online appendix using the ACS that there is an observable jump in migrants from Puerto Rico to the Orlando area in 2018 after the hurricane, as proportion of the population. Particularly when compared to a city for which we expect the migrants have a much smaller relative impact on population such as New York City that graph shows an increase of Puerto Rican immigrants by 0.2-0.3 percent of the population. While likely an underestimate of the inflow, due to inability to reach all the new arrived, the chart confirms the sudden increase in Puerto-ricans following Hurricane Maria.

To get a more accurate sense of the size of the total wave in Orlando we combine the U.S. Census Bureau estimate of 123,399 net out-migrants from Puerto Rico between July 2017 and July 2018 with the FEMA application data to approximate the share of these migrants in Orlando. We conservatively estimate that around 24,000 Puerto Ricans relocated to Orlando in the months following the hurricanes. This represents about 1 percent of Orlando's population, 1.5 percent of its working-age population (aged 16-64), and around 2 percent of its pre-hurricane total employment. We conservatively estimate that at least half of these migrations are directly attributable to the hurricanes (based on out-migration trends from Puerto Rico in the years prior). This inflow was smaller than that of Cubans arriving in Miami after the Boatlift (which equaled 5-6% of the Miami labor force). Nevertheless, the post-Maria Puerto Rican refugee flows provided a significant shock to Orlando's population and labor force, especially as the large majority of this inflow took place within a relatively short period of 1-3 months.[40] To put this

[38] The note can be read at https://www.census.gov/library/stories/2020/08/estimating-puerto-rico-population-after-hurricane-maria.html

[39] In fact, the other estimates of Puerto Rican migration patterns following Maria suggest substantial numbers of refugees returned to Puerto Rico by the end of 2018. As the ACS provides only the year and not the month the participant was surveyed, it cannot capture temporary, mid-year migration patterns of the sort prompted by Hurricane Maria. The Schachter and Bruce (2020) note from the Census Bureau argues the ACS data need to be adjusted to accurately capture net migration from Puerto Rico to the mainland U.S., allowing them to arrive at the 123,399 number.

[40] As we demonstrate in robustness checks using the three- and five-most treated CZs, this shock was more than large enough to generate the effects we observe.

in context, an inflow of immigrants equal to one percent of the labor force in one year is larger than the annual inflow of Mexicans during the 1990s, when Mexican immigration into the US was at its peak. This inflow as a percent of Orlando's labor force is thus comparable to that peak and it was concentrated in the months between September 2017 and March 2018. To compare this inflow to another well-known international event, the inflow of Syrian refugees in Turkey was comparable to our setting, at 2.5 percent of the local population over the course of a couple of years (see Tumen (2015)).

Finally, to mention other studies that used the unusually high inflow of Puerto-Ricans in 2017-18, Ozek (Forthcoming) uses this migration event to study an anonymous school district in Florida which received a plurality of refugees from Maria. Using administrative data that captures whether a student is a Maria refugee, the author documents that 4,000 students enrolled in this district as a result of the hurricane. This school district is likely to be Orlando Unified.

2.3 Sector Distribution of Puerto Rican Migrants

Before the considered event the distribution of Puerto Rican migrants across sectors of employment in Orlando was not very different from that of natives (online Appendix Table A2). As of 2016, Puerto Rican natives were slightly more concentrated in local services such as *transportation* and retail, and slightly less concentrated in *education* and *management*. However, when we consider the sector distribution of all Hispanics and of Hispanics born abroad, (columns 3 and 4 of online Appendix Table A2) we see a significant over-representation in the construction sector. Because of language and cultural commonalities, this extended group (and not just Puerto Ricans) can be a very important network to find jobs in the short run after arrival. The construction sector may have been a particularly attractive industry initially as, in Orlando, it employs a larger share of workers who do not speak English well relative to any other sector (11.8% do not speak English or do not speak it well). The majority of Puerto Rican residents are not fluent in English (only about a third speak English "well" according to the Puerto Rico ACS), hence a job in construction would have been more attainable in the short run. Additionally Bureau of Labor Statistics (2015) shows that in 2014 Hispanic workers in the U.S. were more likely to work in the construction sector than any other sector in the economy. Finally, we use the QWI to look at employment in Orlando and demonstrate that construction saw substantially larger growth in the Hispanic share of employment between Q2 2017 (before the hurricane hit), and Q2 2018 compared to any other industry (online Appendix Figure A4).Combining all these pieces of information, it is reasonable to think that the evacuees from Maria found in Orlando's construction sector a quickly accessible and attractive source of jobs, at least in the short run, relative to other sectors.

3 Data and Methodology

3.1 Overview

The migration push out of Puerto Rico was very sudden and, triggered by the hurricanes. The presumption, therefore, is that it is not correlated with economic fundamentals in Orlando. The presence of the largest community and network of Puerto Ricans already in Orlando, as share of population, likely attracted the migrants making this the main destination in the sudden post-Maria flow. Taken together, these features constitute a natural experiment and make this event a good candidate to estimate the causal effects on Orlando's economic outcomes, by comparing them to those of similar and unaffected cities. We focus on the Orlando-area labor market captured by its commuting zone (CZ)—an aggregation of counties constituting the Orlando labor market—to internalize most of the economic effects.[41]

We primarily call the Orlando CZ our "treated unit" and we adopt a synthetic control approach to estimate the effects on local (log) employment, compensation per worker, and establishment counts. We analyze the local economy in aggregate, and also focus on three sectors: the construction sector (NAICS 23) which, as argued, received potentially the largest labor supply increase from the migrants' arrival; the retail trade sector (NAICS 44-45) and the accommodation and food services, or hospitality, sector (NAICS 72). Retail and hospitality are non-tradable sectors, most likely to have experienced a labor demand shock associated with the increased demand for local accommodations, hospitality services and goods for purchase generated by the migrants. While some immigrants may have worked in these sectors, Puerto Rican employment does not appear to have disproportionately concentrated here. Therefore, while any effects on construction employment and wages are primarily the result of a labor supply shock in that sector, the effects on retail and hospitality should primarily be the result of a labor demand shock (with a possible small labor supply increase). The aggregate effects, then, are a combination of both the supply of new workers and the expenditures of new consumers in the Orlando economy.

After analyzing overall employment and average wage effects, we turn our attention to a second group of outcomes which approximate the effects on native workers (as done in Card (1990); Borjas (2017); Peri and Yasenov (2019)). To do this we analyze the employment and wages of non-Hispanic workers, who are very likely to be U.S. born, both in the overall economy and in the specific sectors where separate labor demand and supply effects should be more visible.

We also try to identify effects on the group of workers that is more vulnerable and in the short-run

[41] These have become the standard units of analysis when considering labor market impacts (see, for instance, Autor et al. (2013); Autor and Dorn (2013)). On a county basis (i.e. prior to cross-walking our QCEW and QWI data into 1990 CZs), we combine the five boroughs of New York City, and individually combine Virginia's independent cities with their surrounding counties.

may be more affected by labor market competition from immigrants—who may take jobs that do not require sophisticated language skills and are disproportionately concentrated in the construction sector. These are typically workers with low levels of education (high school or less) and who often receive relatively low pay (on average). While immigrants from Puerto Rico did not have levels of schooling significantly lower than natives, in the short run they may have been willing to "downgrade" their job expectations in order to find work (e.g. if their English language skills were not particularly strong). In such a case, even with a similar education distribution among immigrants and natives, there may still have been stronger competition for less-skilled jobs in the short run.

Finally, we look at changes in the number of local establishments as a proxy of the response of local investment, at least in the short-term. Firm-creation is an important mechanism of adjustment to immigration in the long run, but there has been very little work done studying the rapidity of its response in a local economy.

3.2 Data

Our primary analysis is conducted using data from the Quarterly Census of Employment and Wages (QCEW), published by the U.S. Census Bureau. The QCEW is derived from the Unemployment Insurance (UI) accounting system in each state, and effectively covers 95%+ of all employed individuals from UI-reporting establishments. The data are available at the industry-by-county level down to the 6-digit NAICS level definition, and are observed monthly for employment and quarterly for earnings and establishments. For cells that are particularly small there may be "suppression" of data due to privacy reasons, but there are very few suppressions at the 2-digit NAICS level. This provides near-complete coverage of employment and earnings, and the availability of these data each month or quarter (depending on the variable) allow us to conduct a reliable short-run analysis along several dimensions. The complete coverage and high reliability are also substantial advantages relative to studies of the Mariel Boatlift, which are all based on small, weighted samples from the Current Population Survey. Of our variables of interest, employment is observed monthly while compensation, establishments, and the derived compensation-per-worker variables are observed on a quarterly basis. We focus on the period 2014 Q1 - 2018 Q3 for the quarterly-observed variables, and January 2014 - August 2018 for employment. This allows us to include three and a half year of pre-Hurricane data and one year of post-Hurricane data, hence the short-run effect of the shock.

Additional analysis is conducted using data from the Quarterly Workforce Indicators (QWI). A product of the U.S. Census Bureau, the QWI data are the result of the Longitudinal Employer-Household

Dynamics (LEHD) program, which (as with the QCEW) uses data from state UI accounting systems, as well as other sources. The QWI data report employment and earnings of employees at the industry-by-county level, and they also break the data down by education level and by ethnicity.[42] This, as mentioned above, allows us to measure labor market outcomes for a subset of workers who are likely to be natives in the Orlando labor market (non-Hispanic workers), as well as for workers with a high school degree or less education, who could face more precarious employment. There are some issues with the QWI data relative to the QCEW: all variables including employment are only observed on a quarterly basis; for less-educated workers, only non-youth individuals (aged 25+) are observed; a few states are missing observations for some quarters of interest; and the available data series do not cover all counties in all states for a full year after Hurricane Maria.[43] This final issue somewhat restricts the number of commuting zones that can be included in the donor pools for the QWI-based estimates.

3.3 Empirical Approach

3.3.1 Synthetic Control Estimator

Our main econometric analysis follows the synthetic control estimator approach introduced by Abadie and Gardeazabal (2003) and further refined by Abadie et al. (2010, 2015). This method is most often utilized with a single observed unit which experienced a treatment beginning at a single point in time, while a 'donor pool' of potential 'control' units did *not* receive such a treatment.[44] In our case, treatment is the inflow of Puerto Ricans to the city of Orlando, which began immediately after the sudden and unanticipated Hurricanes Irma and Maria devastated Puerto Rico. The advantage of this estimating strategy, relative to an ad-hoc choice of controls, is that the appropriate control unit—the *synthetic control*—as a weighted average of a subset of units in the "donor pool" (an appropriate set of untreated units) that best match the pre-treatment values of a set of predictors of the outcome, including linear combinations of the period-specific pre-treatment outcomes. The path of synthetic control outcomes after treatment represents the path of counterfactual outcomes for the Orlando commuting zone, had it not been treated. The difference between the observed value in Orlando and the value for the corresponding synthetic control is the causal estimate of the treatment effect on the outcome of interest .

Formally, we observe $N = J + 1$ units (commuting zones), indexed by j, where $j = 1$ is the single treated unit (Orlando, in our case) and the remaining J units are the untreated units in the donor pool.

[42] Cells are more likely to be suppressed as they become more focused on a particular group.

[43] States missing QWI observations for some or all counties include Arkansas, Maine, Minnesota, Mississippi, Missouri, New Jersey, Pennsylvania, South Dakota, Virginia, and Washington.

[44] Recently the method has been extended to deal with multiple treated units e.g. Cavallo et al. (2013); Abadie and L'Hour (2019); Ben-Michael et al. (2019) and we apply this version in section 44.4

Each unit is observed T total periods, indexed by t, with T_0 total pre-treatment periods and $T-T_0>0$ treated periods. $Y^N_{j,t}$ is the potential outcome observed in $\{j,t\}$ if j is *not* treated at t, and $Y^I_{j,t}$ is the potential outcome observed in $\{j,t\}$ if j *is* treated at t. The treatment effect (also often referred to as the "gap") can be defined as:

$$\alpha_{j,t} = Y^I_{j,t} - Y^N_{j,t} \tag{5}$$

The observed outcome in j,t is: $Y_{j,t} = Y^N_{j,t} + \alpha_{j,t} D_{j,t}$. As only $j=1$ is treated, we have:

$$D_{jt} = \begin{cases} 1 & \text{if } j=1 \text{ and } t > T_0 \\ 0 & \text{otherwise} \end{cases}$$

The goal is to estimate the dynamic path of treatment effects, $\alpha_1 = (\alpha_{1,T_0+1},\ldots,\alpha_{1,T})$. As $Y^I_{1,t}$ is observable $\forall\ t > T_0$, we need only estimate $Y^N_{1,t}$. The synthetic control estimator for $Y^N_{1,t}$ is a weighted sum of the same outcomes for the non-treated units:

$$\hat{Y}^N_{1,t} = \sum_{j=2}^{J+1} \hat{w}_j Y_{j,t} \ \ \forall\ t \tag{6}$$

The weights are obtained by minimizing the distance, between the treated unit and the J non treated ones in the donor pool, of a set of k predictor variables plus M linear combinations of the outcomes before the treatment year. Specifically the synthetic control method selects the vector of weights $\hat{\mathbf{W}} = (\hat{w}_2 \ldots \hat{w}_{J+1})' = \hat{\mathbf{W}}(\hat{\mathbf{V}})$ on the J units in the donor pool, given the matrix $\hat{\mathbf{V}}$ of weights on the K predictor variables[45] that solve the following problem:

$$\hat{\mathbf{W}} = \arg\min_{\mathbf{W}} \sqrt{(\mathbf{X_1}-\mathbf{X_0}\mathbf{W})'\mathbf{V}(\mathbf{X_1}-\mathbf{X_0}\mathbf{W})} \qquad \text{s.t.} \qquad \sum_{j=2}^{J+1} w_j = 1,\ w_j \geq 0\ \forall\ j \in \{2,\ldots,J+1\} \tag{7}$$

where $\mathbf{X_1}$ of dimension $K \times 1$ includes the values of the r selected covariates in the pre-treatment period plus the M linear combinations of the pre-treatment period observations of the outcome variable, while matrix $\mathbf{X_0}$ includes in each row k the vector of values for the same variables and time periods as in $\mathbf{X_1}$ but for untreated units $j \in \{2,\ldots,J+1\}$ in the donor pool.[46]

[45] Given the relatively small number of pre-treatment periods, we follow the advice of Abadie et al. (2015) and estimate $\mathbf{V}$ using the regression-based method adopted as the default in the `synth` command for Stata Abadie et al. (2011) described in detail in Kaul et al. (2015)

[46] Our estimator imposes the No-intercept, Adding-up, Non-negativity and Exact-balance constraints considered by Doudchenko and Imbens (2016). These constraints are commonly imposed on synthetic control estimators. Abadie et al. (2015) emphasize the value of the non-negativity condition, in particular, as it ensures estimates are not subject to potential extrapolation bias and helps preserve interpretability (see, also, Abadie and L'Hour (2019)).

These estimated weights are used to calculate $\hat{Y}^N_{1,t}$, and finally our estimated elements of α_1:

$$\hat{\alpha}_{1,t} = Y_{1,t} - \hat{Y}^N_{1,t} \;\; \forall \;\; t \in \{T_0+1, ..., T\}$$

As it is important that units in the donor pool are not affected by the treatment (see e.g. Cao and Dowd (2019)) we restrict the donor pool by excluding any CZs that received any significant number of Puerto Rican evacuees (10 or more Puerto Rican Irma- or Maria-associated FEMA applications per 100,000 population). When analyzing sector-specific outcomes, to reduce the possibility of interpolation bias (i.e. including in the synthetic control some units which may match well on certain important variables but poorly on others), we further restrict the donor pool to include only those CZs at or above the 75^{th} percentile of sector-specific employment levels. This ensures we are only allowing the synthetic Orlando (for each sector of focus) be comprised of other CZs with large pools of workers in the same sector.[47] For each outcome variable we also drop any commuting zone for which the outcome is not observed in every period.

This yields the group of 148-170 commuting zones[48] (depending on the sector of focus, and including Orlando) in the United States for which data is consistently available at a monthly or quarterly frequency for the time period before and after Hurricane Maria (September 2017) and which meets our qualification thresholds for inclusion (see Figure 11). These criteria exclude every CZ which borders Orlando, and nearly every CZ in Florida, which also minimizes the risk of any treatment spillover on the control group. Finally in order to increase homogeneity and comparability across commuting zones, for each outcome we cleaned each series of seasonal variation (details in the Appendix).

September 2017 is the first *treated* period, $T_0 + 1$, for employment (observed monthly); quarter Q3 in 2017 is the first *treated* period for variables observed on a quarterly basis. The pre-treatment period starts at the beginning of 2014, when the recovery from the Great Recession had firmly taken root. This avoids heterogeneous, short-run local labor market dynamics that occurred after the Great Recession.[49] In the analysis using the QCEW data, we have $T_0 = 44$ pre-treatment periods (months) and consider $T - T_0 \in \{6, 12\}$ treated periods corresponding to half a year and one year after the hurricanes hit Puerto Rico, for our outcomes observed on a monthly basis (namely employment). For the quarterly-observed outcomes (earnings and establishments), and when using the QWI data, we have $T_0 = 14$ pre-treatment periods and consider $T - T_0 \in \{2, 4\}$ treated periods.

[47] As Abadie (Forthcoming) notes, "Including in the donor pool units that are regarded by the analyst to be unsuitable controls because of large discrepancies in the values of their observed attributes... or because of suspected large differences in the values of the unobserved attributes... relative to the treated unit is a recipe for bias."

[48] 128-137 CZs for the 12-month QWI estimates

[49] One of our robustness checks extends the pre-treatment period back to the beginning of 2013.

The predictor variables included in matrix X_0 are: average quarterly local construction and hospitality employment, each as proportions of aggregate local employment;[50] and the pre-treatment values of the outcomes at 6-month intervals from 2014 to 2015, and at quarterly intervals from 2016 through Q2 2017.[51] In total we have $K = 11$ predictor variables in our primary specifications.[52] Figure 11 shows the commuting zones included in the donor pool when analyzing employment in aggregate and in the construction, retail and hospitality sectors. Figure 12 shows those CZs that receive positive weight in the synthetic control for the employment and earnings outcomes, in aggregate and in the construction sector. Note that the Las Vegas, NV commuting zone, the Reno, NV commuting zone, and some commuting zones in the Los Angeles area often receive large positive weights. This is reasonable as their economies, which strongly rely on tourism and construction, and have large numbers of Hispanic workers, each resemble the economy of Orlando.[53]

Additionally we present estimates of the average treatment effects using the three- and five-most treated CZs, rather than only the most treated one (Orlando). We apply the more recent SCM improvements, developed in the case of multiple treated units, where the treatment effect is constructed by weighting the paths of synthetic control gaps for each treated CZ by per-capita FEMA applications.[54]

3.3.2 Inference

Once we have estimated the treatment effects $Y_{1,t} - \hat{Y}^N_{1,t} \;\; \forall \;\; t \in \{1, ...T\}$ a key question is whether they are significantly different from 0. Hypothesis testing using a synthetic control approach comes with challenges, as this method does not produce standard errors, and large-sample inferential approaches are not appropriate. Most of the proposed approaches involve the construction of a test statistic based on some form of falsification test.[55] The one proposed in Abadie et al. (2015) and endorsed in Firpo and Possebom (2018) as the uniformly most powerful is the ratio of the treated-period mean squared prediction error (MSPE) to the pre-treatment-period MSPE, referred to as the RMSPE. A substantial benefit of a test statistic based on the RMSPE is that, by construction, post-treatment deviations from the null are normalized by the pre-treatment fit, such that large post-treatment deviations are not attributed

[50] To match the economy of the synthetic Orlando with that of the actual Orlando

[51] Kaul et al. (2015) argue that including the full set of pre-treatment outcomes as predictors will result in certain zero $\hat{v}_k$ weights for some covariates which may actually be important for predicting future values of the outcome variables. This can bias estimated treatment effects; thus we ensure a number are excluded.

[52] For each specification of interest and given $\hat{\mathbf{V}}$, the existence of sparse solutions to (7) with no more than $K+1$ strictly positive weights $\hat{w}_j$ follows from Carathéodory's theorem, while uniqueness obtains with a maximum of K strictly positive weights provided Orlando does not fall within the convex hull of the donor pool units, and provided the columns of the predictor matrix $\mathbf{X_0}$ are in general position (see Abadie and L'Hour (2019)).

[53] Online Appendix Table A3 lists the positively-weighted CZs for these outcomes.

[54] Wiltshire (2021) details how to extend the synthetic control estimator to cases with many treated units.

[55] The literature on conducting inference with synthetic control estimators is relatively young and rapidly evolving. See for instance Abadie et al. (2015); Doudchenko and Imbens (2016); Hahn and Shi (2017); Ferman and Pinto (2017); Chernozhukov et al. (2017); Firpo and Possebom (2018); Abadie and L'Hour (2019)

undue significance if the pre-treatment fit is poor.

Alternatives to the RMPSE are proposed by Hahn and Shi (2017), who argue that permutation tests of this sort may suffer from incorrect statistical size and suggest the Andrews (2003) end-of-sample instability test, and by Chernozhukov et al. (2017), who propose an alternative *moving block* permutation test which may be ideal but is not the current standard. We thus apply all three tests to our estimates and present the corresponding p-values in Table 24. We also present the Andrews and moving block p-values for our secondary estimates in Table A4 in the online Appendix.[56] As the RMPSE test is the most conservative of these three (as can generally be seen in Table 24) and is the current standard for synthetic control inference, we generally base our claims of statistically significant treatment effects on RMSPE tests and report the results of that test for our secondary estimates in Table 25 and for our robustness checks in online Appendix Tables A4-A6.

The RMSPE p-value is constructed by repeating the synthetic control estimation procedure for each commuting zone in our donor pool, effectively conducting falsification or 'placebo' tests by reassigning treatment to each of the $j \in \{2,...,J+1\}$ *untreated* CZs in our donor pool to estimate $\hat{Y}^N_{j,t}\ \forall\, t$. For each $j \in \{1,...,J+1\}$ we then calculate the summary statistic:

$$RMSPE_j = \frac{\sum_{t=T_0+1}^{T}(Y_{j,t} - \hat{Y}^N_{j,t})^2/(T-T_0)}{\sum_{t=1}^{T_0}(Y_{j,t} - \hat{Y}^N_{j,t})^2/T_0} \tag{8}$$

Our test statistic is then constructed as a p-value based on the empirical distribution of these $RMSPE_j$:

$$p = \frac{\sum_{j=1}^{J+1} \mathbb{1}[RMSPE_j \geq RMSPE_1]}{N} \tag{9}$$

$N = J+1$ is the total number of units and our actual treated unit (Orlando) is $j = 1$. Thus if the deviations between observed post-treatment outcomes and the synthetic control relative to the pre-treatment fit are large enough in Orlando relative to the distribution of differences from our placebo tests, our p-values will be small and we will reject our null hypothesis of no effect.[57]

[56] Andrews p-values and moving block p-values for our other estimates are available upon request.

[57] We calculate the RMSPE p-values for our estimates of the ATEs using the three- and five-most treated CZs by constructing a distribution of 1,000 randomly sampled (weighted) averages of the placebo gaps, as described in Wiltshire (2021).

4 Results

4.1 Effects on All Workers

Results for employment are illustrated in the four panels of Figure 6. Each panel of the figure plots the observed (de-seasonalized, logarithm of) employment in Orlando against that of its synthetic control, first in aggregate (Panel A) then for the construction (Panel B), retail (Panel C), and hospitality (Panel D) sectors. We set the value to zero at the beginning of our pre-treatment period, which is January 2014. The month in which the hurricane hit (the initial period of treatment), September 2017, is the shaded area. The six months following the shaded area are those in which Orlando experienced the largest inflow of people from Puerto Rico fleeing Hurricane Maria. If they caused any significant effect on Orlando's labor market relative to the synthetic control, it should be visible in the portion to the right of the shaded area. The distance—or "gap"—between the line representing Orlando and that of its synthetic control, after September 2017 represents our estimates of the causal effect of the inflow of Puerto Ricans as a result of the Hurricane. These gaps are shown as percentages in Figure 3, as the solid black line in each panel.

Three facts emerge from inspection of Figure 6. First, for each of our outcomes, the pre-Hurricane match between Orlando and its synthetic control is remarkably good. This implies that the commuting zones constituting our synthetic controls mimic well the short run fluctuations and long run trends in Orlando employment before September 2017. Figure 12 shows the commuting zones included in each particular donor pool. In particular, nine donor pool CZs were assigned positive weights for Orlando's aggregate deseasonalized log employment synthetic control. These donor CZs are, in ascending order of weights: Fort Walton Beach-Pensacola, FL (2.7%); Fresno-Visalia-Tulare-Parterville, CA (5.3%); Las Vegas, NV-AZ (6.5%); Boise City, ID (6.7%); El Paso, TX-Las Cruces, NM (7%); Nashville, TN (11%); Provo-Orem, UT (15.3%); Fayetteville-Springdale-Rogers, AR (18.6%); and Gainesville, GA (26.9%)[58] Second, Figure 6 shows that, for aggregate as well as sector-specific employment, Orlando realized positive employment effects as a result of the migration event.[59] Third, this positive employment effect is largest and most clearly seen in the construction sector and begins a few months after the hurricane hit Puerto Rico. One year after the hurricane hit, Orlando's construction employment was about four log points (4 percent) larger than its synthetic control.

Treatment effects (the post-treatment gaps) from Figure 6 are quantified in the first two columns of Table 24. The entries are the treatment effects 6 months (column 1) and 12 months (column 2) after

[58] online Appendix Table A3 provides these lists for a broader set of synthetic controls. Comparing this map with Figure A2 in the online Appendix, we can see that none of the donor CZs received any meaningful number of evacuees

[59] This is more obviously shown by the black lines in Figure 8.

September 2017. The table also shows the RMSPE statistics and their p-values, as well as the Andrews and the moving block p-values described in Section 3.3.2. These statistics indicate the significance of the treatment effect in Orlando by conducting numerous "placebo" tests in which each CZ in our donor pool is assigned treatment status (none of which were exposed to the inflow of Puerto-Ricans). High positive estimated treatment effects and low p-values imply that the treatment effect is positive and significant relative to the placebos.

The treatment effect on aggregate employment 12 months after the hurricane is positive and significant at the 5% level using any of our three p-value calculation procedures. Despite the seemingly small effect size, this result is consistent with most evacuees finding jobs in Orlando, and corroborates estimates of the number of evacuees that, at least in the year after the hurricane, remained on the mainland. Specifically, this effect is about 0.4 percent of aggregate employment in Orlando. Approximately 24,000 Puerto Ricans arrived in Orlando over the 6 months following the hurricanes, at least 11,000 of whom can be attributed to the hurricanes, constituting 0.65 percent of the pre-treatment working-age population. In the years preceding the hurricanes, around 75 percent of Puerto Ricans in Orlando (and more widely) were aged 16 and over, around 60 percent of whom were employed (U.S. Census Bureau, 2017). If these patterns held among the 11,000 refugees, this would amount to about a 0.3 percent increase in the local labor force. That is, three-fourths of Orlando's post-treatment growth in aggregate employment (relative to the synthetic control) was due to the labor supply shock, with no evidence that local workers were displaced.

The construction sector experienced the largest and most significant increase in employment, perfectly in line with the idea that this sector received a supply "shock" from the inflow. We estimate a 1.5 percent increase in construction employment 6 months into the treated period, although this estimate is not quite significant at the 10 percent level using our most conservative p-value calculation. Our estimated treatment effect 12 months into the treated period, however, corresponds to 4 percent increase in construction employment and is highly significant (RMSPE p-value of 0.01). This growth in construction employment would account for a full 80 percent of the employed refugees if their age and employment patterns were similar to those of the Puerto Ricans who preceded them. retail and hospitality employment show a more modest treatment effect equal to about one percent of their employment. We have argued above that the construction sector was most likely to absorb the immigrants, many of whom were unlikely to be fluent in English.

The retail and hospitality sectors, on the other hand, were much more likely to be affected by this migration event through increased demand for consumer goods and services, which could translate through to increased labor demand. Looking at the retail sector, employment did significantly increase—slowly at first, with a 0.3 percent increase seen after 6 months (RMSPE p-value of 0.04) and then a 0.9 percent increase seen after 12 months (RMSPE p-value of 0.06). We also estimate a 1.2 percent increase in hospitality sector employment 12 months after treatment, though here significance is only seen using the Andrews and moving block p-values (0.02 and 0.05, respectively).

To visualize the significance of the treatment effects on aggregate employment in Orlando and on each sector—especially construction—Figure 8 shows the estimated (treatment-synthetic control) gap for Orlando together with the same gap calculated for all of the other (non-treated) commuting zones in the donor pool. As in Figure 6, Panel A is relative to (residualized logarithm of) aggregate employment, Panel B represents employment in the construction Sector, Panel C employment in retail, and Panel D employment in hospitality. The gaps for the other commuting zones are shown as light grey lines. Additionally, the dark dashed line shows the gap for Los Angeles—a tourism-based, Latino-dense large metropolitan area that provides a useful reference.[60] Note that for aggregate and retail employment and, much more clearly, for construction employment, the Orlando-synthetic control gap in the post-Hurricane period is in the top part of the grey trajectories, while it is not far from the average in the pre-Hurricane period. This means that Orlando had an unusual increase in aggregate and especially construction employment in the post-Hurricane period. We also observe a one-month dip in employment in every sector, but especially construction, exactly in September 2017, before the increases occurred. We will show that this was likely the impact of Hurricane Irma, causing a temporary slow-down in economic activity when it hit Orlando. In section 4.4, below, we consider the possible confounding effects of this event and show that our estimates are likely free of associated bias.

Focusing on earnings, the panels of Figure7 show the plots relative to (de-seasonalized, natural logarithm of) earnings per worker for Orlando relative to synthetic control, and follows a similar structure to Figure 6. The impact on earnings per worker was a combination of the increased labor supply and increased local demand produced by the new arrivals. Our estimates of treatment effects on earnings are reported, 6 and 12 months after the hurricane, in columns 3 and 4 of Table 24. The table also shows the significance of the estimates from each of our three p-values.

Interestingly, in spite of the significant increase in aggregate employment driven by the evacuees inflow, earnings per worker was stable in aggregate over the first 6 and 12 months of treatment. Provided labor demand is not perfectly elastic, this implies that labor demand grew along with labor supply. When looking at the sector-level, we do not find any negative treatment effect on earnings per worker in the

[60] L.A. and some surrounding CZs, along with Reno, NV and Las Vegas, NV typically receive substantial weighting in the synthetic controls. See Figure 12 as well as Table A3 in the online Appendix.

sectors considered above. In the construction sector, compensation per worker actually grew by 3.3 percent (RMSPE p-value of 0.05). Thus the large increase in the construction-sector workforce did not have a negative effect *overall* on construction sector earnings per worker. This may have been a mixture of composition effects, as more people became employed with different effects on incumbent workers of different skills, depending on their complementarity and substitutability with evacuees. We investigate this further in the next section.

The hospitality sector, on the other hand, experienced the most significant effect on earnings per worker (using the RMPSE p-value, and due especially to the particularly good fit in the pre-treatment period). Wages were a significant 1.4 percent higher 12 months after the migration event (RMSPE p-value of 0.03).[61] The estimated effects on retail earnings per worker are positive but small, and not significant. Together with the employment effects on these sectors, these patterns may have emerged because the arrival of the refugees constituted a shock to local demand for consumer goods and services. Hospitality sector employers would then have had to raise wages to compete with other sectors to maintain staffing levels as demand increased. While this story is plausible and consistent with the patterns of our estimated treatment effects, our data do not allow us to test it directly. Overall, the results suggest a significant increase in labor demand, particularly in hospitality and retail, and a labor supply increase in construction that was matched by increased aggregate demand, such that aggregate earnings per worker did not decline.

Finally, columns 5 and 6 of Table 24 show the estimated treatment effects on numbers of establishments. In aggregate, the treatment effect on the number of establishments amounted to a positive, though not statistically significant, increase of 0.5 to 0.6 percent. Similarly, for each of our sectors of interest we have positive effects around 1 percent, but no statistical significance. This is suggestive (but not conclusive) evidence that even as quickly as a few months after the arrival of the evacuees, firms may have started to open new local establishments to take advantage of higher labor supply and consumer demand. The noise in the data—there is a notable dip in Orlando establishments in mid-2015 (see, for example, the Orlando trend in Panels B and D of Figure 13 which eliminates our RMPSE p-value significance. If we instead consider the moving block p-value for construction, the treatment effect of around 1% is significant (p-value of 0.06)—and the short period of time, however, do not allow us to produce robust and conservative statistically significant results on this outcome.

Overall, Table 24 paints a picture of a significant wave of Puerto Rican evacuees being absorbed into the Orlando labor market, with positive effects on employment and per-worker earnings. The estimates

[61] We estimate earnings per worker grew by a statistically significant 0.1 percent 6 months after treatment began, but view this estimate as too small to be considered economically significant.

suggests significant positive effects even on employment and earnings per worker in sectors which likely did not absorb these new arrivals as workers, but may have been affected by their demand for goods and services. The employment effects of the labor supply shock were particularly large in the construction sector which, as argued earlier, was the most likely place for the refugees to find work. The effects of the consumer demand shock were clear increases in retail employment and in hospitality earnings per worker, and possibly in hospitality employment as well. We find no conclusive evidence of significant changes in the numbers of local establishments—though the point estimates are all positive—and thus cannot draw any conclusions about the impact of this migration event on local investment. We find no evidence of displacement or decline in aggregate wages or in sector wages overall.

4.2 Effects on Non-Hispanic and Less-Educated Workers

How did immigrants impact the wages of natives and, among them, of specific groups they may compete with? To study these questions in our setting, we require data which measure employment and wages of subgroups of workers in the local economy. We thus turn to the Quarterly Workforce Indicators (QWI), which allow us to analyze the same labor market outcomes as the QCEW but for various subgroups of workers stratified by ethnicity and educational attainment. Similar to the QCEW data used in our previous analyses, the QWI collects information on employment and wages for each industry and county and offers the same geographical coverage as the QCEW, although all variables are observed at quarterly frequencies. One cost of using the QWI for these analyses is that we lose about 20 percent of our sample when looking one full year after the hurricane, primarily due to issues around data-reporting from particular states.

While we are not able to directly observe worker nativity, we can approximate native workers quite well by focusing on ethnicity. Using ACS data we calculate that, five years prior to the hurricane, 88 percent of "non-Hispanic" workers were "native", meaning they were born in one of the 50 states in the U.S. We also observe that the overwhelming majority of people from Puerto Rico (99%) identify as Hispanic. Hence the subgroup of non-Hispanic workers serves as a good proxy for US natives, and we can be certain that group will not include the evacuees when observed in the period after the Hurricane.

We thus repeat the analysis performed in the previous section, using the QWI data and focusing on non-Hispanic workers. The estimated treatment effects on log employment (columns 1 and 2) and log compensation per worker (columns 3 and 4), 6 and 12 months after the Hurricane, are presented in Table 25. We do not find any evidence of a negative effect on non-Hispanic employment either in aggregate or in any of the sectors considered. Rather, the point estimates of the employment effects on non-Hispanic workers are positive and, though not statistically significant in our sectoral analyses, are significant for the aggregate analysis (the construction employment effect of +0.7 percent is significant according to the moving block p-value only). We estimate an increase of 0.8 percent in non-Hispanic employment 12 months after the Hurricane. This may represent some "crowding-in" driven by complementarities between immigrants and natives. As explained in Ottaviano and Peri (2012) if immigrants perform jobs/tasks that are different from Natives' and if production requires combining those tasks, the increase in supply of one of workers increases demand for the other group. However, as the estimates of this coefficeant are rather large and imprecise we remain cautious when interpreting this result. We also fail to find an effect, positive or negative, on employment of non-Hispanic workers in the construction, retail, or hospitality sectors. Small positive and non-significant point estimates suggest, again, no crowding-out of non-Hispanic workers.

Focusing on earnings per worker of non-Hispanics, we estimate a very small and non-significant increase in the aggregate, both six and 12 months after the hurricane. This null aggregate effect seems to average a significant negative 2.5 percent impact on construction and a positive 2 percent impact on retail sector earnings. The retail sector result is not significant according to the RMSPE p-value, but *is* significant according to the Andrews and moving block p-values (0.07 and 0.00, respectively, identical to those for the construction estimate) reported in Table A4 in the online Appendix. A short-run decrease in non-Hispanic construction wages by 2.5 percent in response to a 4 percent increase in construction employment would imply a partial demand elasticity of -0.6, which is plausible. The fact that we do not find any effect on aggregate wages of construction workers in Table 24 (which include Hispanic sand non-Hispanic) and a positive employment effect in this sector, suggests that inflow of workers and some change in composition may be responsible for this estimate. In general, however, for non-Hispanics there may have been small negative effects on construction wages and small positive effects on retail wages which balanced out in the non-Hispanic aggregate. The positive wage effect in retail would be consistent with a positive labor demand shock.

A second group that is often the focus within the immigration-labor literature are those individuals with a high school degree or less, as their wages have declined in recent decades (Autor et al., 2008) and as competition with immigrants in relatively low skilled jobs may hurt their opportunities Borjas (2003). This group of workers constituted about 40 percent of the US labor force in 2016. In the retail and hospitality sectors in Orlando, 49 percent of workers had a high school degree or less as of 2016 (the year before Hurricane Maria hit Puerto Rico). In the construction sector this share was 60 percent. We repeat our primary analysis using the QWI data, limited to workers with high school degree or less.

As noted earlier, the education levels of workers under age 25 are not observed, meaning this part of our analysis is limited to non-youth workers. Columns 5-8 in Table 25 present our results. We find no evidence of a significant impact on overall employment or compensation per worker in this group 6 months after the hurricane. However, 12 months after the Hurricane we see a significant 0.8 percent increase in aggregate employment for this subgroup (p-value of 0.04) and a 1.9 percent, significant decrease in their per-worker earnings. We view this latter result with some skepticism, as that estimate is not particularly precise and the increase in employment may have somewhat changed some of the workers' composition. The +1.7 percent effect on construction employment is only significant according to the Andrews and moving block p-values. Interestingly, the wage effects for less-educated workers in specific sectors range from a positive 1.2 percent effect for retail workers (p-value of 0.02. Positive and significant effects in this sector are already evident 6 months after the Hurricane hit) to a 1.5 percent decline in the construction sector (p-value of 0.06). The estimate on hospitality earnings is positive 1.3 percent after 12 months, but this is not even close to significant according to the RMSPE or the Andrews p-values. These results are again consistent with the idea that retail and hospitality received a positive labor demand shock that put upward pressure on their workers' employment and compensation, while the construction sector received a supply shock which increased employment but possibly reduced wages for incumbent workers.

Putting together the results for all workers (Table 24), and for non-Hispanics and less-educated workers (Table 25), as well as the sector-specific analyses, we are left with three significant findings. First, during the post-Hurricane period the large inflow of evacuees from Puerto Rico led to an increase in total employment and also employment of both non-Hispanic and less-educated workers in Orlando increased somewhat. Second, the retail and hospitality sectors experienced not only employment growth, but also positive growth of compensation per worker, both overall and within the non-Hispanic and less educated subgroups (although this last effect was not significant). We see this as clear evidence of a labor demand increase in those sectors, likely as a consequence of new demand for local goods and services from the evacuees. Finally, the construction sector saw the most significant increase in employment due to the evacuees and potentially also due to increases in non-Hispanic and less-educated employment. While this increase in labor supply did not seem to produce a depressive wage effect on the average worker in that sector, it may have hurt the earnings of non-Hispanic and less educated construction workers. The very large short-run labor increase may have depressed wages of some groups in this sector. We interpret these results as evidence that, even in the very short run, this significant inflow of migrants likely had an overall positive impact on native and less-educated workers' employment and wages, though native and less-educated workers in the sector most heavily exposed to the labor supply shock did face some downward pressure on earnings, while those working in sectors producing goods and services that immigrants consumed had a positive upward effect on earnings.

4.3 Regression-based Estimates

An alternative, but much more simplistic, method to conduct inference, adopted in Peri and Yasenov (2019), is to use the synthetic control method to identify the control units (i.e. those getting positive weight in the distance-minimization problem) and then perform difference-in-differences regression-based estimates considering Orlando as the only treated unit and the other units as control(s). Abadie et al. (2015) note how regression-based estimates implicitly weight the control units in a way similar to how a synthetic control estimator does.[62] While this method allows a very simple and straightforward check of our estimates and is intuitive, the standard errors may not be quite correct, especially when we consider the "synthetic Orlando" as the only control unit in the regression. Moreover, any simple regression model with a single post-treatment period fails to capture any trend-changes post-treatment. For these reasons, we view our regression-based DD estimates with a deal of caution and only as additional evidence on the qualitative findings and robustness of our synthetic control approach.

For each regression-based DD model we consider two ways of constructing the control groups. In Table 26, columns (1) and (2), we include in the control group all commuting zones that are positively weighted in the synthetic control unit. In column (3), we instead only include the constructed synthetic control (the weighted composite) as the 'control group'.[63] To implement the difference-in-differences estimate, we regress the outcome variable (alternatively [deseasonalized] log employment, log compensation per worker and log establishments) on a treatment dummy which equals to one for Orlando after September 2017 and zero otherwise, and on time fixed effects (column (1) of Table 26) or on time and unit fixed effects (column (2)). For column (3), where we have only a single treatment and a single control unit, we can include only time effects. In Table 27, instead of one single 'post-Hurricane' Orlando treatment dummy, we include an Orlando dummy interacted with each half-year period, both before and after the hurricanes, and estimate against the composite synthetic control (showing results for employment and earnings). This serves to test whether there are deviations from zero in the pre-hurricane period, which would imply non-similar trends between Orlando and the control group. The latter half of

[62]The regression approach, however, does not constrain the weights to be non-negative; thus a regression-based approach allows extrapolation outside the support of the data. Additionally, failing to restrict the sample of donor pool units to match the pre-treatment characteristics of the treated unit (conditional on included covariates) increases the probability that the "parallel trends" condition will be violated.

[63]Thus the estimates in columns (1) and (2) of Table 26 more likely suffer from bias realized through interpolation and extrapolation, while those in column (3) are free from extrapolation bias and less-likely to suffer from interpolation bias.

2017 is the period of the "treatment" (when the hurricanes hit Puerto Rico and the migration occurred) and is therefore treated as the reference period in the analysis.

The estimates in Table 26 mostly confirm the synthetic control results. Because the estimates in column (3) are most likely to be free of extrapolation and interpolation bias (see above), we focus primarily on these. First, aggregate employment exhibits a positive and significant treatment effect of 0.3 percent, and the construction sector shows the largest positive employment effect, although now it is only 1.5 percent rather than 4 percent (as this is an average over the entire post-hurricane period and there is evidence of temporary inflow only). Second, log earnings per worker show small effects, usually non-significant, that vary between positive and negative depending on which control group (column) one chooses. Third, these estimates show significantly positive effects on establishments, and especially large in the construction sector. In Table 27 we show the estimates of the Orlando dummy interacted with each half-year period, showing employment (columns 1-4) and compensation per worker (columns 5-8). The four columns for each outcome variable correspond, left to right, to the aggregate economy, and to the construction, retail, and hospitality sectors. In general, the deviations of Orlando from control, both in employment and per-worker earnings, are small in the pre-Hurricane period. However, there are small but significant deviations which may suggest modest *negative* pre-trends for construction and hospitality employment and for aggregate and hospitality earnings per worker, and possibly a modest positive pre-trend for retail employment. It is also clear that the post-Hurricane employment and earnings effects are positive and significant, especially by Q3 of 2018 (one year after the hurricanes) and the largest positive employment effect is in construction. In general, the regression results confirm the findings of the synthetic control method and, while they reveal a certain variability of the pre-Hurricane differences between Orlando and its control, they do not show systematic and significant departures.

4.4 Robustness Checks

In addition to our difference-in-differences estimates, we conduct five additional robustness checks to address a variety of concerns.

First, we adopt an alternative hypothesis testing procedure implying that in our RSMPE and relative p-value calculations we correct for possible deviations between treatment and synthetic control at the time right before treatment. Specifically, we define our treatment effects as the difference between observed outcomes and synthetic control outcomes *minus* that difference in the period immediately prior to treatment. This has essentially the flavor of a difference-in-differences estimator on the treated and synthetic control units. That is, we define $f_0 = Y_{j,T_0} - \hat{Y}^N_{j,T_0}\ \forall\ j$ and our alternative estimated treatment effects are then $\hat{\alpha}'_{1,t} = Y_{1,t} - \hat{Y}^N_{1,t} - f_0 \; \forall \; t$. This subtracts from any estimated treatment the gap in T_0, defining the relevant statistic as the change in the gap between T_0 and T. Using this statistic and evaluating the corresponding p-value (Table A5 in the online Appendix) shows very similar significance levels as with the usual method. This indicates that there was no significant deviation between Orlando and control at the time of the shock.

Our second robustness check extends the length of the pre-treatment period. In our primary specification, the pre-treatment period begins in the first period of 2014 (quarter or month, as appropriate) and continues through to the period immediately prior to the hurricane. The reason we begin in 2014 was due to concerns that the impact and duration of the recovery from the Great Recession was heterogeneous across commuting zones and could affect pre-2014 trends. Still, extending the pre-treatment period to the beginning of 2013 helps test the stability of our results. We conduct this robustness check by extending the pre-treatment period to the beginning of 2013, de-seasonalizing the raw data over this period, and then implementing our estimating strategy as before. The results are presented in online Appendix Table A6 and confirm our primary findings from Table 24, showing very similar treatment effects for all variables considered.

For our third robustness check, we consider the role of Hurricane Irma, which hit Florida in September 2017, downing trees and power lines and causing other damage (although the damage was much less severe than either Irma or Maria did in Puerto Rico). The immediate local economic impact was almost certainly negative as a result of Irma and is likely captured in the significant dip in Orlando's employment in September 2017, evident in Figures 6, 7, and 8. It is theoretically possible that the aftermath would yield movement in economic activity in either direction (e.g. Groen et al. (2020)). Damage to infrastructure, for example, could hamper economic activity for months, or there could be a boost to the economy as the result of rebuilding efforts. This motivates our attempt to isolate potential consequences of exposure to Irma. To do this, we consider a large commuting zone which received very few Puerto Rican refugees but which lay approximately the same distance from the path taken by Irma's eye as did Orlando: Jacksonville, FL.

Compared to the 3,972 relevant FEMA applications filed in Orlando, Jacksonville saw fewer than one-seventeenth (221) relevant FEMA applications. On a per-capita basis (around 17 applications per 100,000 population), Jacksonville saw less than a tenth of those seen in Orlando, which saw 178 applications per 100,000 population.[64] Any robust estimated labor market treatment effect in Jacksonville—particularly in the construction sector, which we view as the most likely to demonstrate any positive post-hurricane impact as a result of reconstruction efforts—is thus much more likely to measure the

[64]We note this number of applications still disqualified Jacksonville from inclusion in our donor pools for the analyses above.

lingering impact of exposure to Irma rather than the impact of an inflow of Puerto Rican evacuees.

The gaps for aggregate employment and construction employment between treatment and synthetic control for Orlando and Jacksonville, respectively, are shown in Figure 13, Panels A and C (Table A7 of the online Appendix quantifies the results). The panels include the other placebo gaps for all donor pool units. Two things are clear from these figures. First, the gaps (vs synthetic control) in Jacksonville experienced very similar dips to those seen in Orlando in September 2017, when both labor markets were directly exposed to Hurricane Irma, for both aggregate employment and construction employment. This confirms the similarity of potential effects of Irma. Second, the employment gaps—both in aggregate and in the construction sector—6 and 12 months after the Hurricane hit are significantly larger for Orlando than for Jacksonville. While we observe some ups and down in the gap for construction employment in Jacksonville before and after Irma, we do not see anything comparable to the significant increase in Orlando (which also has more stable paths of gaps in the pre-Hurricane period). In Jacksonville the estimated treatment effect on construction employment 12 months after Irma is 2.5 percent, but the RMSPE is quite small (1.01), reflecting the substantial amount of noise in the period before the Hurricane relative to the period afterward. The associated RMSPE p-value, at 0.81, makes clear this point estimate is not in any way significant when compared to the associated placebo tests (nearly 81 percent of the placebo runs yielded a larger RMSPE).

Overall, Jacksonville seems a valid counterfactual of the impact of Irma on Orlando in the absence of the Puerto Rican migrant inflow. These results show that the one-month dip in employment can be attributed to that event, while the one-year surge in employment in Orlando, especially in construction, should be attributed to the impact of the migrants [65] We also plot the results for establishments in Orlando and Jacksonville in Panels B and D. While none of our synthetic control estimates on Orlando establishment numbers were significant, they were all positive (around 1 percent) and these Panels demonstrate how the DD estimates (averaged over the whole post-treatment period) were significant. They also show that Jacksonville did not have any apparent establishment response post-Irma, further supporting our claim of no positive bias from Irma in our Orlando estimates.

Our fourth robustness check extends our analysis to analyze the effect on the top-three and top-five most treated CZs as defined by per-capita FEMA requests. This addresses concerns that when considering only one treated unit some other events in the Orlando labor market could have produced the result. These top-five most treated CZs include (in descending order of treatment exposure) Orlando,

[65] We note that Jacksonville may be less of a good counterfactual for the retail or hospitality sectors because several large retailers started operations here in the treatment period. Amazon opened a distribution center in September 2017 and another in October 2017, both of which were around 1,000,000 square feet, and opened a sorting center in September 2018; Ikea opened a store in November 2017; and Walmart opened a new Supercenter in June 2018 and began hiring in August 2018 for another new store which opened in November 2018.

FL; Lakeland-WinterHaven, FL; Springfield, MA; Daytona Beach, FL; and Tampa-St. Petersburg-Clearwater, FL. Table A1 in the online Appendix shows that each of these CZs were among the top ten destinations for Puerto Rican migration per-capita in 2015, and all but one were among the top ten destinations in the year directly preceding the hurricane. Hence the sudden outflow of Puertoricans, combined with pre-existing network density would indicate these five (and three) metro areas as most likely to receive migrants. Figures 9 and 10, as well as Table A8 in the online Appendix, present treatment effects averaged across these multiple locations. Again, we see virtually the same patterns as when focusing on Orlando; positive effects on aggregate employment, as well as employment and wage effects in construction and positive employment effects in both retail and hospitality.[66] We note the one big difference from Orlando alone is in the hospitality sector: we now see a significant positive impact on hospitality employment, while the growth in hospitality earnings is only significant according to the Moving Block and the Andrews *p*-values. These results demonstrate that the shock to Orlando (by far the most-heavily treated CZ) was consistent with effects on top-treated areas, affirming our primary focus on Orlando. Moreover, these results suggest more broadly that we can consistently estimate the wage and employment effects of immigration shocks even when they are not too large, i.e. in the order of a fraction of percent of the labor force.

Finally, to address the possibility that Orlando experienced a housing boom before September 2017, continuing into the treatment period, and whether such a boom might be driving our construction employment results, we conduct two checks. First, in Figure 14 we plot the (log) Zillow Home Value Index of all homes in Orlando and any of the MSAs approximately corresponding to commuting zones which receive positive weights in our construction employment synthetic control unit. We normalize this value to be equal to one in September 2017 for all units and plot the values through May 2020. We see that during the pre- and post- Hurricane periods, Orlando's trend appears to be rather average and is in no way indicative of a housing price boom relative to the control units. We also estimate a difference-in-difference model similar to those mentioned in Section 4.3, using as the outcome the same (log) Zillow Home Value Index but restricting the treatment period through September 2018 for consistency with our 12-month estimates. The regression results confirm the visual inspection and do not show any evidence that Orlando experienced a housing price boom in the periods before or after September 2017.

These results give us confidence that there were no secular trends or trend breaks in housing prices in the Orlando area, relative to cities in the construction employment synthetic control, before or around the period of Hurricane Maria. More broadly, the combined results of our robustness checks strengthen

[66]Conversely, the distributions of average placebo treatment effects tighten as the number of placebo donors in each placebo average increases.

[66]Results can be found in Table A9 of the online Appendix.

our confidence that our estimates reflect the un-confounded causal impacts of the inflow of Puerto Rican migrants.

5 Conclusions

Using various sources, including FEMA disaster relief applications, we show that a large wave of evacuees relocated, at least temporarily, from Puerto Rico to Orlando after Hurricane Maria struck in September 2017. This episode provides an opportunity to measure the short-run impact of a sudden, sizeable increase in predominantly Spanish-speaking labor supply on the local labor market outcomes of Orlando. We look specifically at aggregate and sector-specific employment and wages for all workers, and also for natives and less-educated workers.

We use the QCEW data, a comprehensive dataset including all workers, aggregated by commuting zone, and implement a synthetic control estimation strategy to measure the impact of these migrant inflows on outcomes in Orlando relative to a synthetic control. We find that overall employment in Orlando increased by 0.4 percent one year after the inflow began, and increased by 4 percent in the construction sector. The quantitative increase in employment compared to the estimated arrival of evacuees is consistent with absorption of this wave with no crowding-out effects, and with a large share finding employment in the construction sector, easier to access for Hispanic workers. We also find that, one year after the inflow began, retail-sector employment rose by nearly 1 percent and earnings per worker in the hospitality sector grew by 1.4 percent. We view these as the results of growth in local labor demand for production of goods and services resulted from the consumer demand of these new arrivals. These results are broadly robust to a series of checks. We perform additional analysis with the QWI data and find that, when focusing on subgroups like non-Hispanic or less-educated workers, the migration event increased their employment by 0.8 percent with little effect on aggregate earnings per worker for those groups, and with increased earnings for them in the retail sector accompanied by a decrease in earnings for likely-native and less-educated workers in the construction sector which absorbed the majority of the labor supply shock.

Our results support a story in which immigration constituted a positive shock to local labor supply and to local consumer demand, which spurred greater labor demand in the local economy. Orlando's construction sector, in particular, was exposed to the labor supply shock, while the retail and hospitality portions of the Orlando economy were primarily exposed to the consumer demand shock. The new workers were absorbed by the local economy without displacing native workers and without any significant overall negative effect on earnings, though there was some downward pressure on construction earnings for natives and less-educated workers, in construction, specifically. However, retail employment and hospitality earnings were both positively impacted, such that the overall effect on natives was to increase their employment without any clear impact on their earnings.

Figures and Tables

Figure 6:

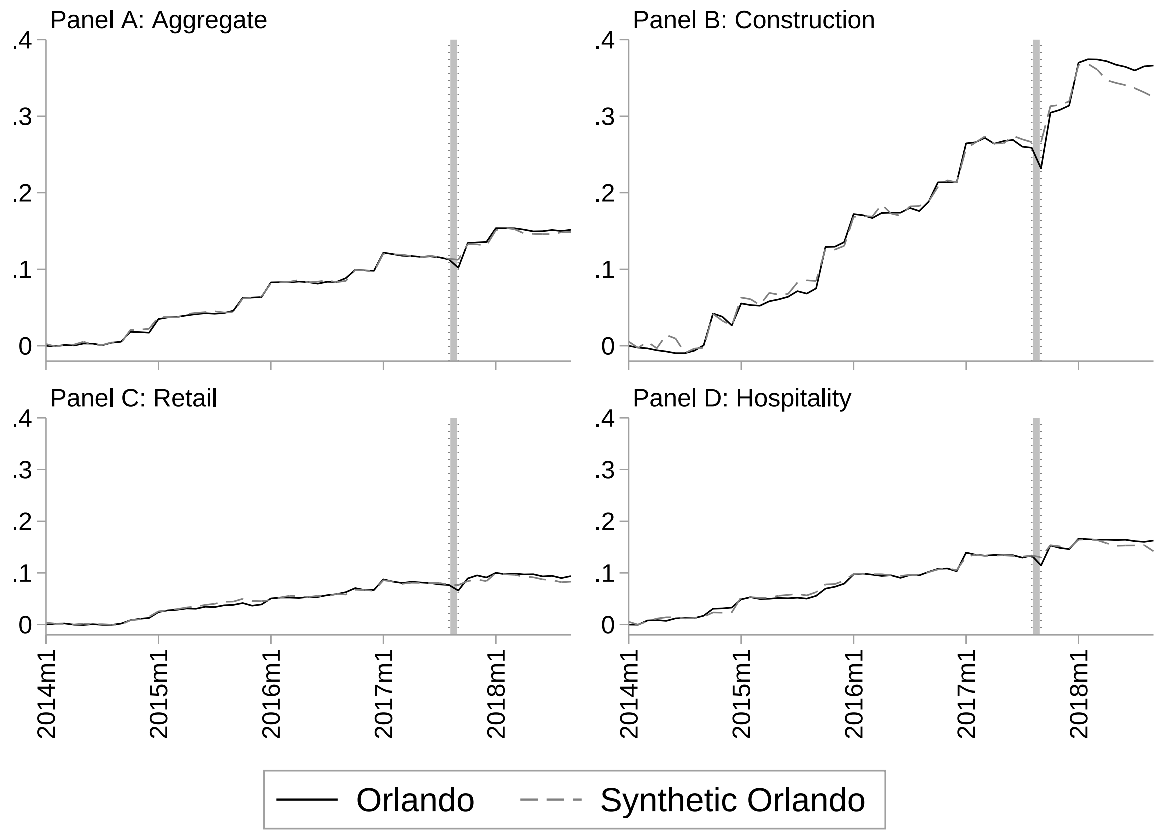


FIG. 6.—Log Employment, Orlando vs Synthetic Orlando. Each panel represents the residualized (after accounting for seasonal component and intercept) logarithm of employment in aggregate (Panel A), in construction (Panel B), retail (Panel C) and hospitality (Panel D) for Orlando and its synthetic control. The data, from the QCEW, are at a monthly frequency. The shaded period is September 2017, when Hurricane Maria hit Puerto Rico.

Figure 7:

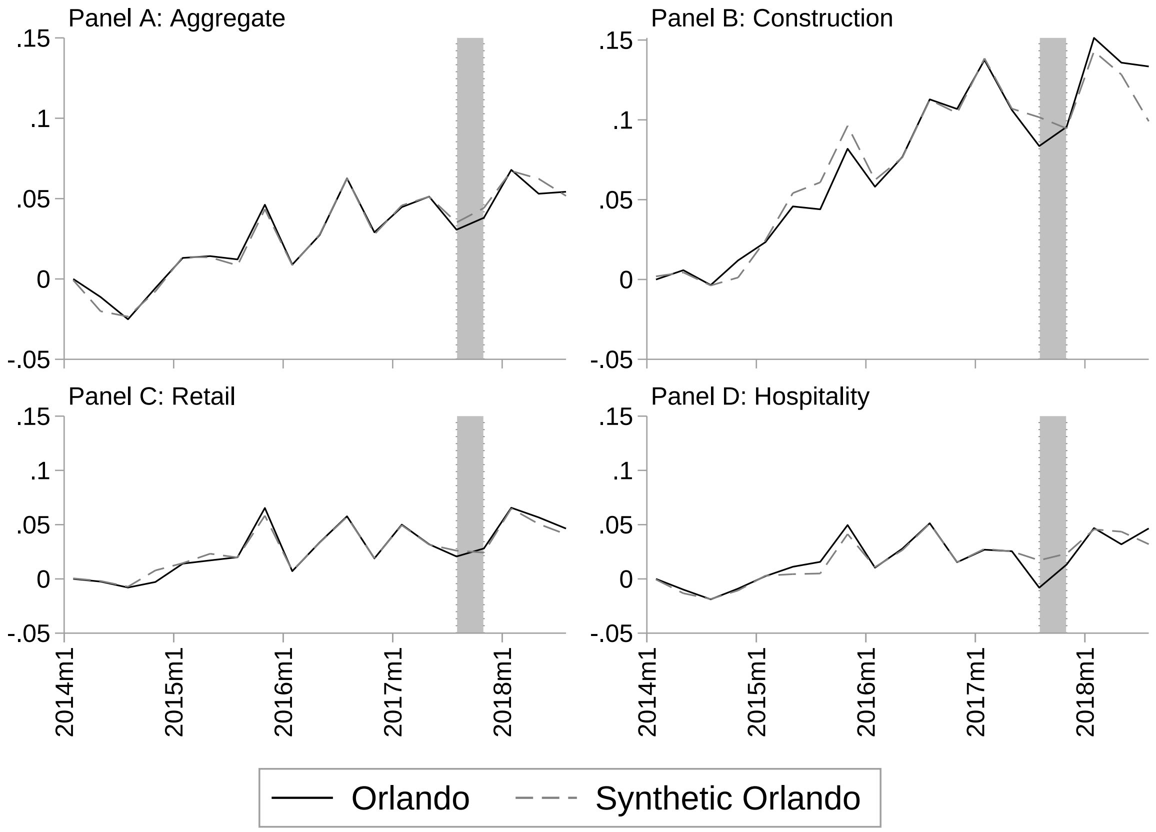


FIG. 7.—Log Earnings per Worker, Orlando vs Synthetic Control. Each panel represents the residualized (after accounting for seasonal component and intercept) logarithm of earnings per worker in aggregate (Panel A), in construction (Panel B), retail (Panel C) and hospitality (Panel D) for Orlando and its synthetic control. The data, from the QCEW, are at quarterly frequency. The shaded period is Q3 2017, when Hurricane Maria hit Puerto Rico.

Figure 8:

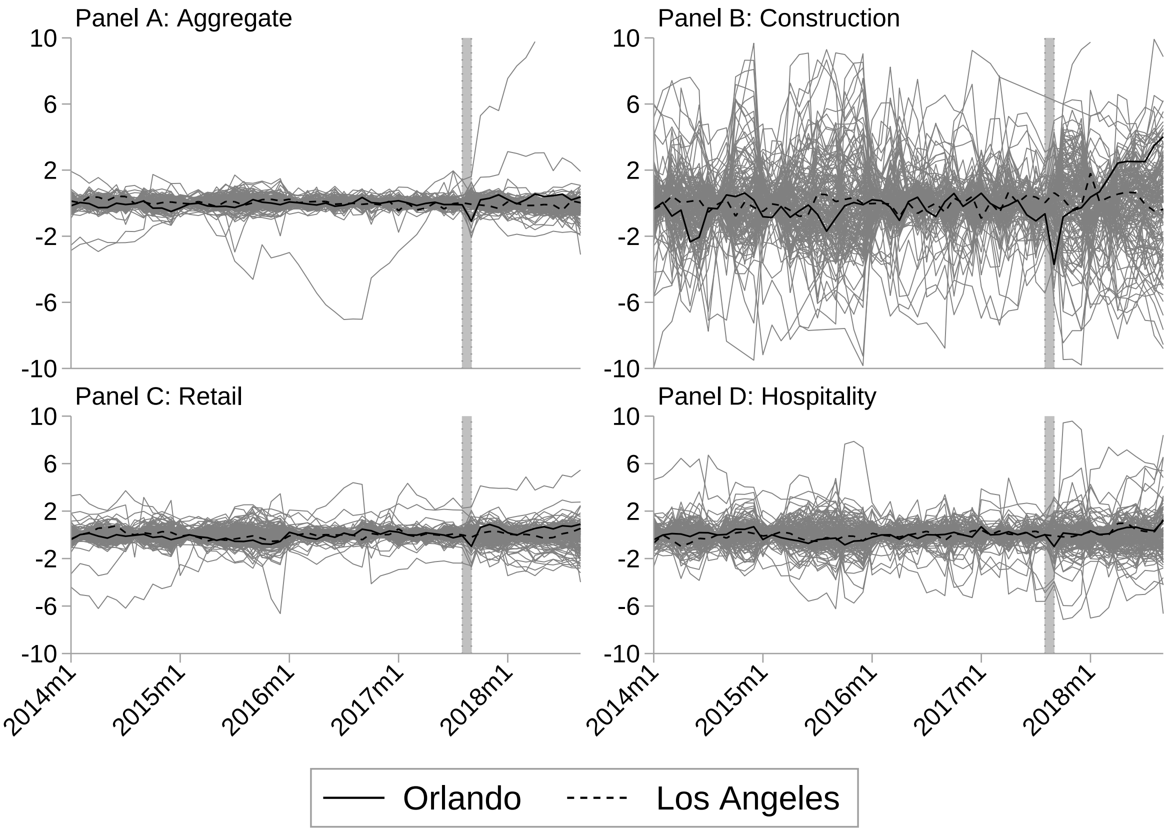


FIG. 8.—Estimated Treatment Effects and Placebo Treatment Effects for Employment in Orlando. Each panel represents the % gap between the 'treated' unit and its control, for Orlando and for each commuting zone in the donor pool (placebo treatments), for each outcome's synthetic control. The variable plotted is 100× the gap in residualized (after accounting for seasonal component and intercept) logarithm of employment in aggregate (Panel A, $N = 165$), in construction (Panel B, $N = 149$), retail (Panel C, $N = 170$) and hospitality (Panel D, $N = 148$). The data, from the QCEW, are at a monthly frequency. The shaded period is September 2017, when Hurricane Maria hit Puerto Rico. The dark line represents Orlando, the dashed dark line represents Los Angeles

Figure 9:

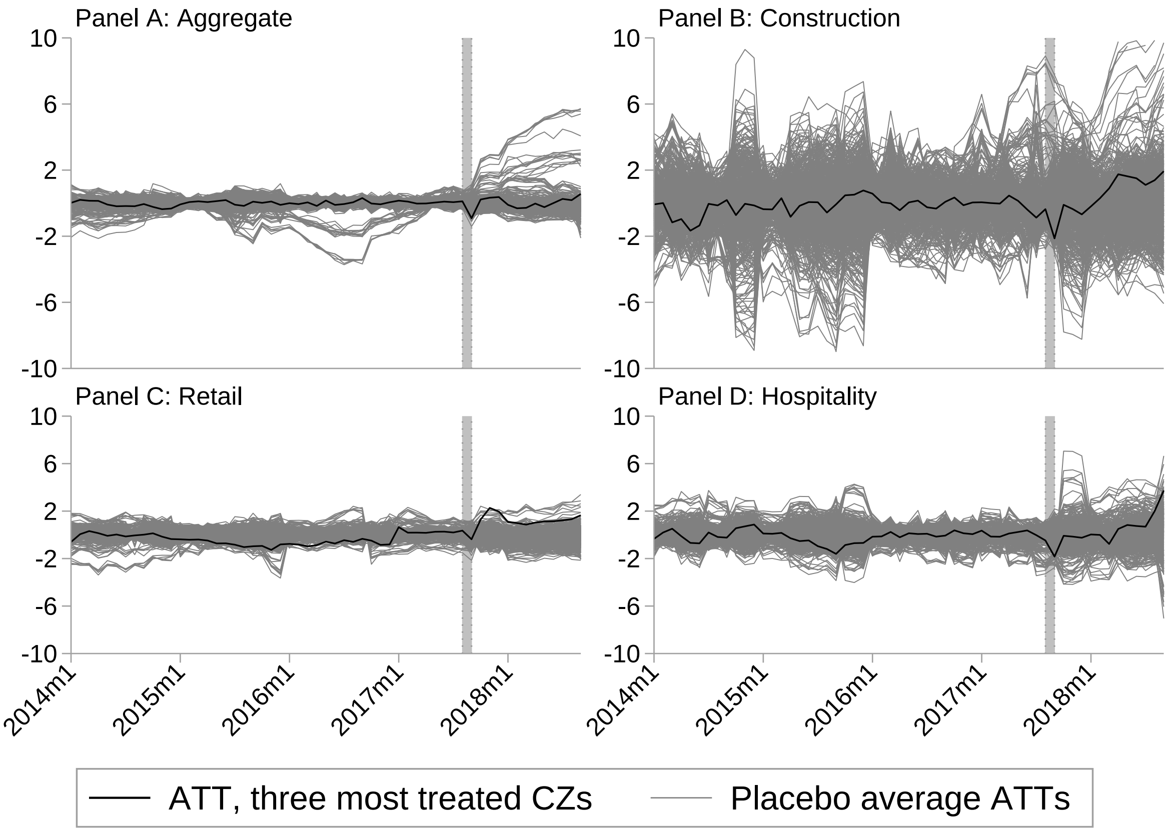


FIG. 9.—Estimated Average Treatment Effects and Placebo Treatment Effects for Employment in the three CZs which received the most FEMA applications per capita. Each panel represents the applications-per-capita-weighted average % gap between the 'treated' unit and its control, for Orlando and for 1000 random samples of three placebo-treated commuting zones in the donor pool, for each outcome's synthetic control. The variable plotted is 100× the average gap in residualized (after accounting for seasonal component and intercept) logarithm of employment in aggregate (Panel A), in construction (Panel B), retail (Panel C) and hospitality (Panel D). The data, from the QCEW, are at a monthly frequency. The shaded period is September 2017, when Hurricane Maria hit Puerto Rico. The dark line represents the average treatment effect for the top three treated CZs (Orlando, FL; Lakeland-Winter Haven, FL; and Springfield, MA)

Figure 10:

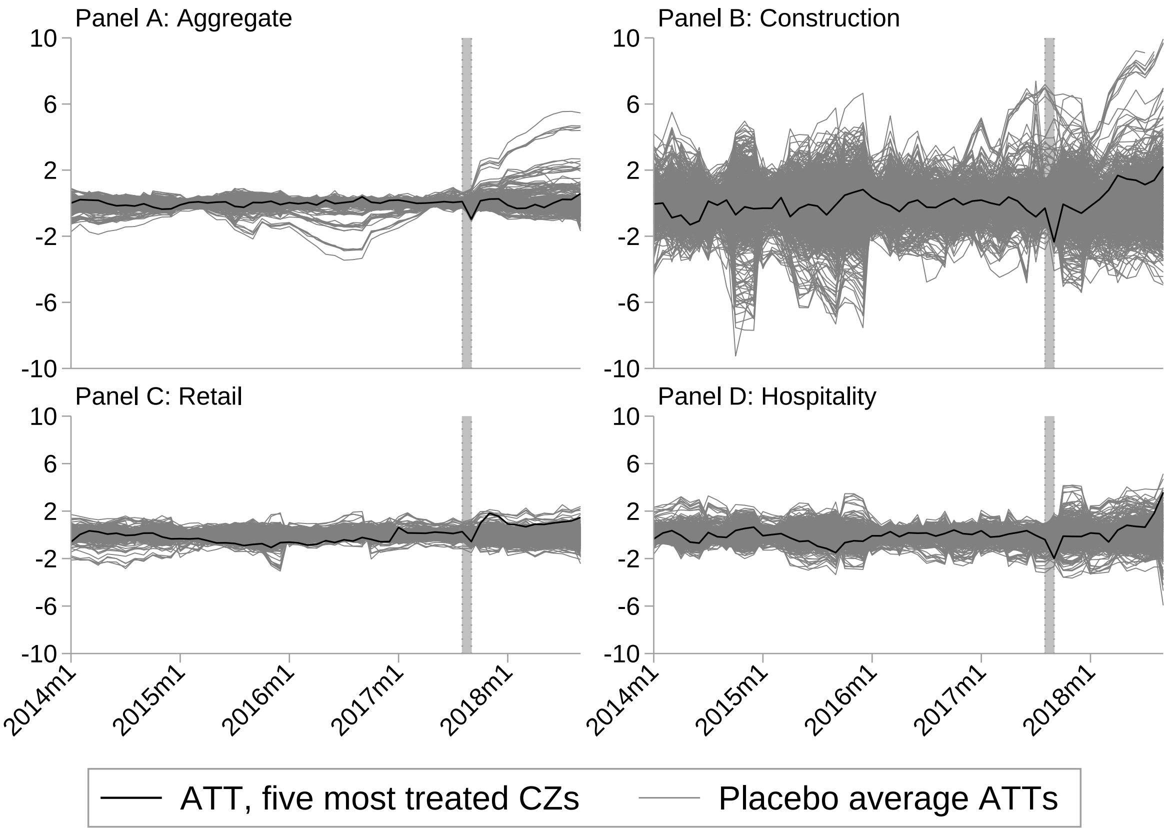


FIG. 10.—Estimated Average Treatment Effects and Placebo Treatment Effects for Employment in the five CZs which received the most FEMA applications per capita. Each panel represents the applications-per-capita-weighted average % gap between the 'treated' unit and its control, for Orlando and for 1000 random samples of five placebo-treated commuting zones in the donor pool, for each outcome's synthetic control. The variable plotted is 100× the average gap in residualized (after accounting for seasonal component and intercept) logarithm of employment in aggregate (Panel A), in construction (Panel B), retail (Panel C) and hospitality (Panel D). The data, from the QCEW, are at a monthly frequency. The shaded period is September 2017, when Hurricane Maria hit Puerto Rico. The dark line represents the average treatment effect for the top five treated CZs (Orlando, FL; Lakeland-Winter Haven, FL; Springfield, MA; Daytona Beach, FL; and Tampa-St. Petersburg-Clearwater, FL)

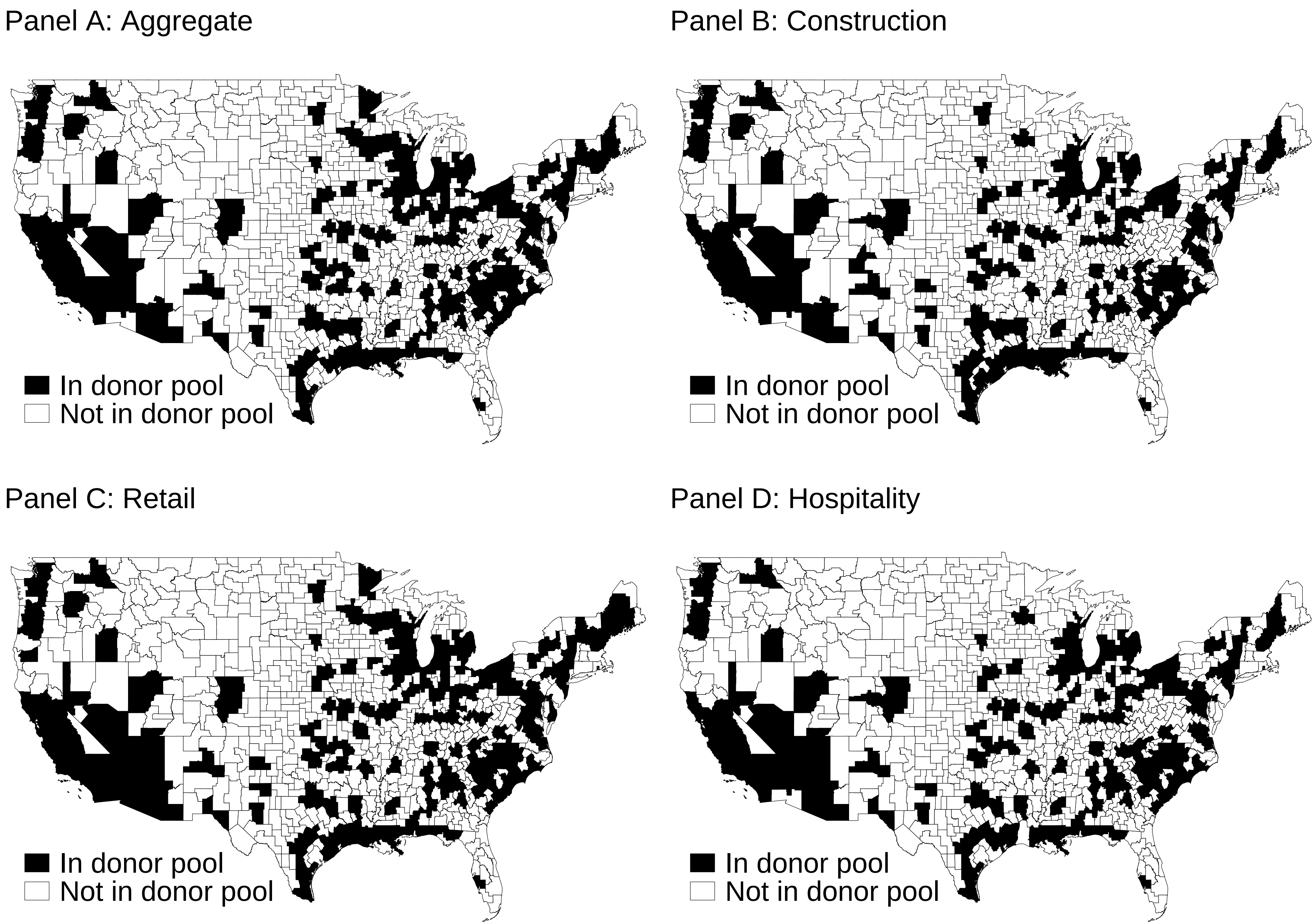


FIG. 11.—Synthetic Control Donor Pools, by Sector. Each panel represents the entire donor pool of commuting zones for the synthetic Orlando, for the indicated outcome. The donor pools are restricted to exclude CZs which received more than .0001 FEMA applications per capita or were below the 75^{th} percentile of industry-specific employment levels.

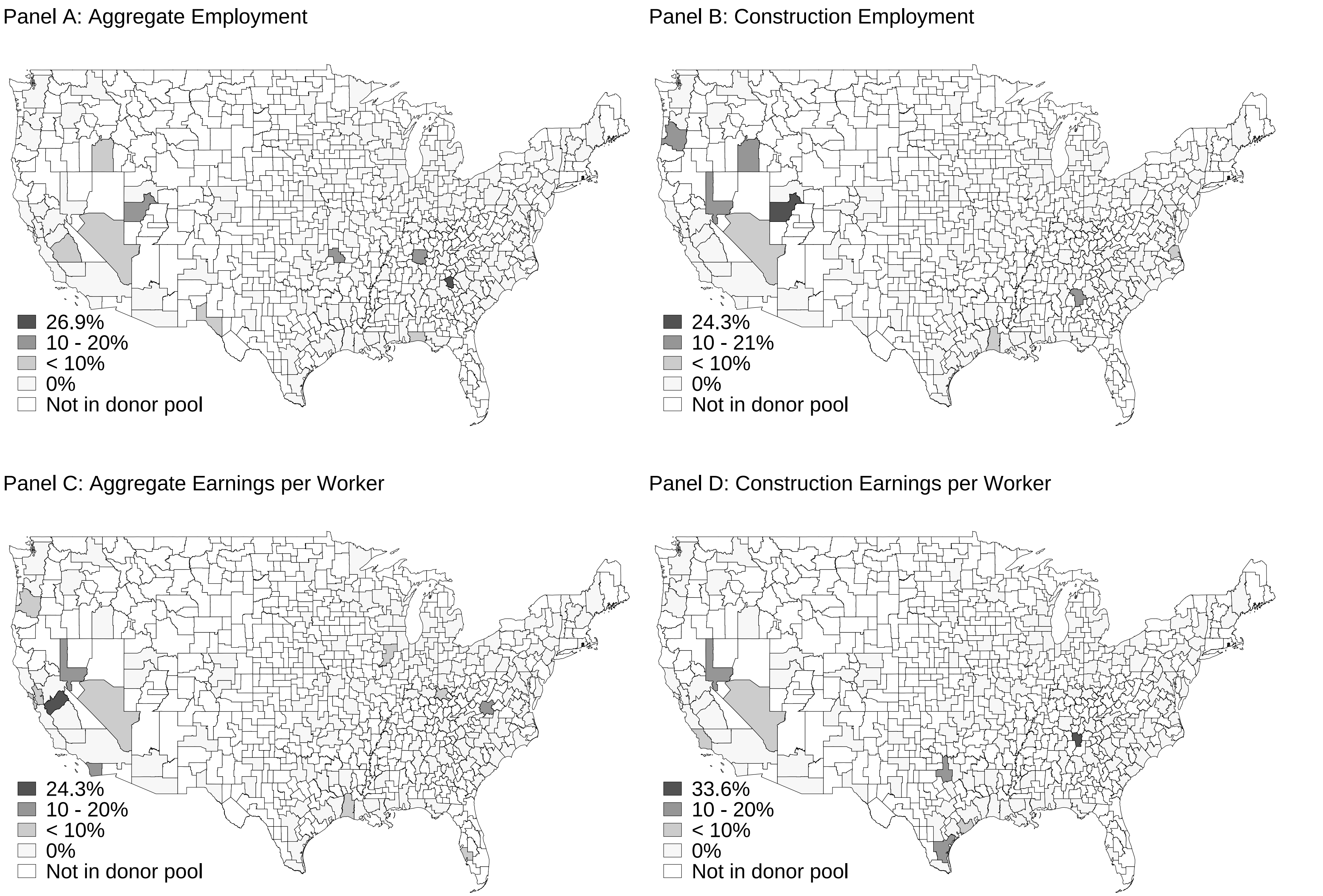


FIG. 12.—Synthetic Control Donor Pools, by Weight. Each panel represents the donor pool of the synthetic Orlando for the indicated outcome by its assigned weight. Shown for aggregate and construction-sector employment and earnings. The data, from the QCEW, are at a monthly frequency for employment and a quarterly frequency for earnings.

Figure 13:

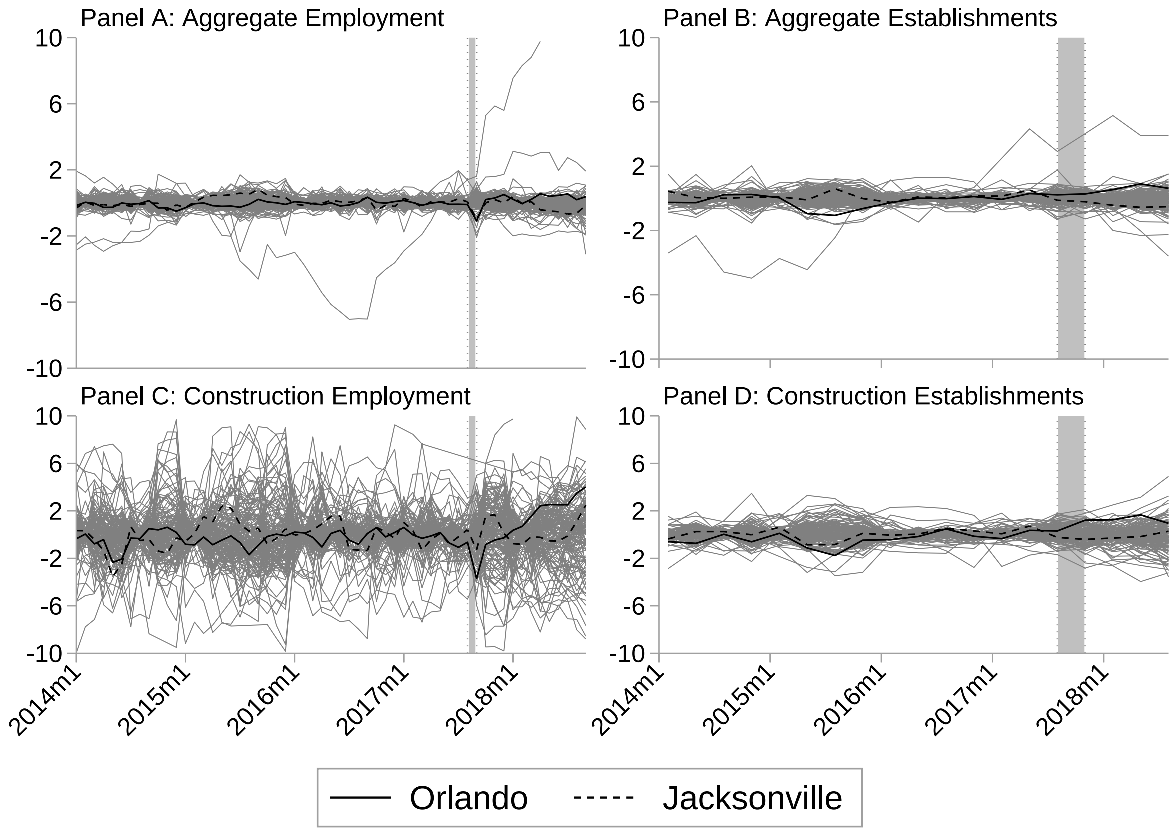


FIG. 13.—Test for Impact of Hurricane Irma on Orlando Treating Jacksonville as Counterfactual Orlando. Each panel represents the gap between treated unit and control for Jacksonville and for each commuting zone in the Jacksonville donor pool, for each outcome's synthetic Jacksonville. The gap for Orlando vs the outcome's synthetic Orlando is also plotted for comparison. The variable plotted is the gap in residualized (after accounting for seasonal component and intercept) logarithm of aggregate employment (Panel A), aggregate establishments (Panel B), construction employment (Panel C) and construction establishments (Panel D). The data, from the QCEW, are at a monthly frequency for employment and a quarterly frequency for establishments. The shaded period is September 2017 for employment and Q3 2017 for establishments, when Hurricane Maria hit Puerto Rico and Hurricane Irma hit both Orlando and Jacksonville at the same strength and approximately the same distance from its eye. The dark line represents Orlando, the dashed dark line represents Jacksonville. Because Jacksonville's exposure to Irma was comparable to Orlando's, and because Jacksonville did not receive a notable number of Puerto Rican FEMA applications related to Irma or Maria, any significant and robust post-Irma effects seen in Jacksonville may represent the impact of Irma on Orlando, which would confound our estimates of the impact of the Puerto Rican migrants.

Figure 14:

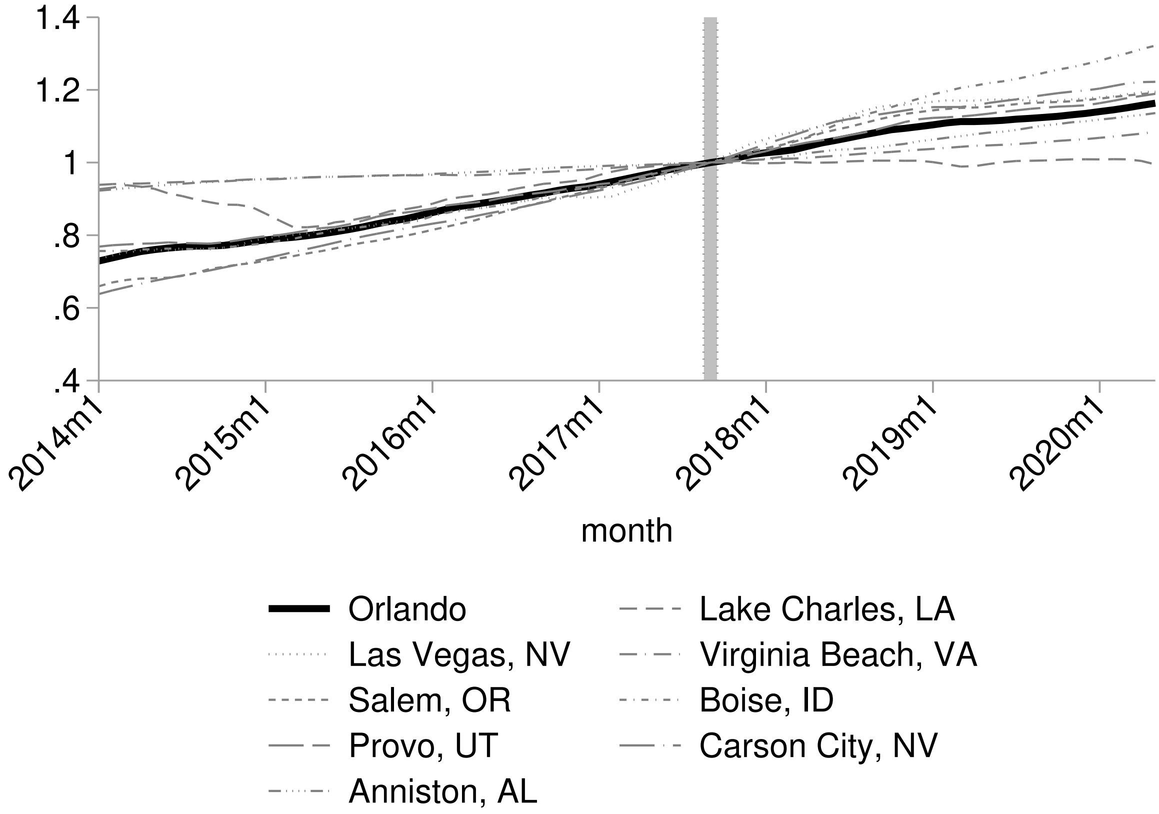


FIG. 14.—House Prices, Orlando and Synthetic Control Donors. The lines plot (the natural logarithm of) the Zillow Home Value Index for Orlando and each of the positively weighted MSAs corresponding to the CZs for the construction sector employment synthetic control (normed to September 2017). The synthetic control weights by commuting zone for construction employment are listed in online Appendix Table A3.

Table 23: Difference in Means, Recent Puerto Rican Migrants vs Comparators

	Natives (all mainland)	Puerto Rican Islanders	Natives (Florida only)
Age (years)	-9.383***	-7.445***	-8.827***
	(-18.29)	(-15.83)	(-9.59)
Male	0.0698***	0.0675***	0.0352
	(4.07)	(3.90)	(1.16)
Married	-0.157***	-0.0279	-0.117***
	(-9.17)	(-1.64)	(-3.84)
Yrs. education completed	-0.386***	-0.253***	-0.132
	(-5.27)	(-3.04)	(-1.02)
High school graduate	-0.0648***	-0.0272***	-0.0469***
	(-8.17)	(-2.68)	(-3.33)
4+ years of college	-0.0154	-0.0207	0.0292
	(-0.95)	(-1.27)	(1.03)
Comparison observations	6,431,276	48,855	333,182

NOTE.—Analysis using data from the 2016 ACS five year sample. All results are for recipients that are in the labor force. *t* statistics in parentheses. * Significance at the 10% level; ** significance at the 5% level; *** significance at the 1% level.

Table 24: Estimated Treatment Effects, All Workers

Sector	Log Employment		Log Compensation per Worker		Log Establishments	
Aggregate	6 months	12 months	6 months	12 months	6 months	12 months
Treatment effect	0.0021**	0.0038**	0.0005	0.0028	0.0054	0.0062
RMSPE	9.9813	8.4823	2.5299	3.8088	1.0399	1.9979
RMSPE *p*-value	0.0121	0.0121	0.5273	0.4000	0.6424	0.6061
Andrews *p*-value	0.2222	0.0444	0.8000	0.2667	0.2667	0.2667
Moving block *p*-value	0.0000	0.0000	0.1875	0.1111	0.3750	0.2778
N	165	165	165	165	165	165
Construction	6 months	12 months	6 months	12 months	6 months	12 months
Treatment effect	0.0153	0.0402**	0.0092	0.0331*	0.0125	0.0095
RMSPE	5.5120	11.0658	5.2998	9.2167	3.3304	3.6134
RMSPE *p*-value	0.1141	0.0134	0.2617	0.0537	0.3289	0.4430
Andrews *p*-value	0.0889	0.0222	0.2000	0.0667	0.1333	0.2000
Moving block *p*-value	0.0000	0.0000	0.1250	0.0000	0.1875	0.0556
N	149	149	149	149	149	149
Retail	6 months	12 months	6 months	12 months	6 months	12 months
Treatment effect	0.0030**	0.0090*	0.0005	0.0033	0.0037	0.0159
RMSPE	4.6217	4.8695	0.3605	0.7814	0.1335	2.2044
RMSPE *p*-value	0.0412	0.0588	0.9000	0.9588	0.9765	0.6882
Andrews *p*-value	0.3111	0.0222	0.4000	0.3333	0.3333	0.1333
Moving block *p*-value	0.0600	0.0000	0.5000	0.3889	0.6875	0.3889
N	170	170	170	170	170	170
Hospitality	6 months	12 months	6 months	12 months	6 months	12 months
Treatment effect	0.0009	0.0118	0.0014**	0.0141**	0.0025	0.0076
RMSPE	1.6708	2.8762	17.1481	13.6413	0.1277	0.6853
RMSPE *p*-value	0.3986	0.2635	0.0270	0.0338	0.9595	0.9054
Andrews *p*-value	0.6444	0.0222	0.4000	0.0667	0.6667	0.2000
Moving block *p*-value	0.4000	0.0536	0.0000	0.0000	0.8125	0.3889
N	148	148	148	148	148	148

NOTE.—Synthetic control estimates of the impact of the immigration inflow on the residualized (after accounting for seasonal component and intercept) logarithms of the indicated outcomes in Orlando (6 and 12 months after Hurricane Maria hit Puerto Rico), using data from the QCEW, unrestricted donor pool. Significance based on RMSPE *p*-values: * significance at the 10% level; ** significance at the 5% level; *** significance at the 1% level.

Table 25: Estimated Treatment Effects, Non-Hispanic and Less-Educated Workers

	Non-Hispanic Workers				Less-Educated Workers			
Sector	Log Employment		Log Earnings per Worker		Log Employment		Log Earnings per Worker	
Aggregate	6 month TE	12 month TE	6 month TE	12 month TE	6 month TE	12 month TE	6 month TE	12 month TE
Treatment effect	-0.0003**	0.0082**	0.0001	0.0009	-0.0006	0.0080**	0.0008	-0.0199*
RMSPE	10.9700	16.2198	0.8780	1.9797	4.1896	9.7029	1.7400	7.5027
RMSPE *p*-value	0.0448	0.0448	0.6791	0.6119	0.1765	0.0441	0.3750	0.0588
N	134	134	134	134	136	136	136	136
Construction	6 month TE	12 month TE	6 month TE	12 month TE	6 month TE	12 month TE	6 month TE	12 month TE
Treatment effect	0.0004	0.0068	0.0004**	-0.0245**	0.0082	0.0166	0.0007**	-0.0147*
RMSPE	2.0665	3.7870	6.7462	7.4052	3.6389	3.8433	8.8591	8.4125
RMSPE *p*-value	0.4141	0.3203	0.0234	0.0469	0.2857	0.3308	0.0226	0.0602
N	128	128	128	128	133	133	133	133
Retail	6 month TE	12 month TE	6 month TE	12 month TE	6 month TE	12 month TE	6 month TE	12 month TE
Treatment effect	-0.0001	0.0009	0.0036	0.0205	-0.0004	0.0089	0.0060**	0.0119**
RMSPE	1.4148	3.4864	1.0713	3.0396	0.1467	0.8946	6.6239	7.1356
RMSPE *p*-value	0.3869	0.2993	0.5839	0.3577	0.9265	0.8015	0.0221	0.0221
N	137	137	137	137	136	136	136	136
Hospitality	6 month TE	12 month TE	6 month TE	12 month TE	6 month TE	12 month TE	6 month TE	12 month TE
Treatment effect	-0.0016	0.0120	-0.0002	0.0028	-0.0041	0.0042	0.0003	0.0134
RMSPE	0.5137	0.9949	1.2531	3.2994	1.3909	1.1233	0.9401	1.1408
RMSPE *p*-value	0.7883	0.8540	0.4818	0.2993	0.4962	0.8647	0.6165	0.8045
N	137	137	137	137	133	133	133	133

NOTE.—Synthetic control estimates of the impact of the immigration inflow on the residualized (after accounting for seasonal component and intercept) logarithms of the indicated outcomes in Orlando (6 and 12 months after Hurricane Maria hit Puerto Rico), using data from the QWI. Columns 1-4 show estimates for non-Hispanic workers. Columns 5-8 show estimates for less-educated workers. The donor pool is restricted to include only those commuting zones which are observed for four quarters after the Hurricane, allowing 12-month estimates of the treatment effects. Significance based on RMSPE p-values: * significance at the 10% level; ** significance at the 5% level; *** significance at the 1% level.

Table 26: DD Estimated TEs of Hurricane Maria on Orlando

Outcome by Sector	(1)	(2)	(3)
Aggregate			
Log employment	0.0108***	-0.0009	0.0028**
	(0.0026)	(0.0080)	(0.0012)
Log earnings/worker	-0.0034	-0.0047	-0.0039*
	(0.0039)	(0.0083)	(0.0021)
Log establishments	0.0001	0.0137	0.0070***
	(0.0096)	(0.0188)	(0.0016)
Construction			
Log employment	0.0097	0.0303*	0.0148**
	(0.0081)	(0.0158)	(0.0058)
Log earnings/worker	-0.0009	-0.0043	0.0089
	(0.0092)	(0.0216)	(0.0080)
Log establishments	0.0045	0.0010	0.0146***
	(0.0071)	(0.0177)	(0.0027)
Retail			
Log employment	0.0064	-0.0197	0.0057***
	(0.0041)	(0.0156)	(0.0014)
Log earnings/worker	-0.0028	-0.0052	0.0031**
	(0.0031)	(0.0076)	(0.0013)
Log establishments	0.0053	-0.0143	0.0089**
	(0.0082)	(0.0162)	(0.0038)
Hospitality			
Log employment	0.0097***	0.0098	0.0034**
	(0.0034)	(0.0096)	(0.0014)
Log earnings/worker	-0.0017	-0.0043	-0.0087
	(0.0093)	(0.0138)	(0.0063)
Log establishments	-0.0059	0.0175	0.0059*
	(0.0065)	(0.0179)	(0.0028)

NOTE.—Difference-in-differences estimates of the impact of the immigration inflow on the residualized (after accounting for seasonal component and intercept) logarithms of the indicated outcomes in Orlando, in the post-treatment period, relative to the pre-treatment period. Estimated using the QCEW data. All positively-weighted CZs in the synthetic Orlando are individually included as controls in (1) and (2), and the composite synthetic Orlando for each outcome is the only control unit in (3). Authors own analysis using the QCEW. Robust standard errors in parentheses. * Significance at the 10% level; ** significance at the 5% level; *** significance at the 1% level.

Table 27: Difference-in-Differences pre-Trend Tests

		Employment				Earnings per Worker			
Year	Period	All sectors	Construction	Retail	Hospitality	All sectors	Construction	Retail	Hospitality
2014	Q1 - Q2	-0.0006	-0.0100**	-0.0019**	-0.0045**	0.0053	0.0006	-0.0010***	0.0023
		(0.0007)	(0.0040)	(0.0008)	(0.0018)	(0.0042)	(0.0017)	(0.0002)	(0.0015)
	Q3 - Q4	-0.0015	0.0006	-0.0011	0.0027	0.0007	0.0064	-0.0062	0.0010
		(0.0011)	(0.0020)	(0.0007)	(0.0021)	(0.0020)	(0.0054)	(0.0051)	(0.0011)
2015	Q1 - Q2	-0.0012*	-0.0066***	-0.0030***	-0.0050***	0.0006	-0.0039	-0.0038	0.0035
		(0.0006)	(0.0022)	(0.0007)	(0.0016)	(0.0009)	(0.0037)	(0.0029)	(0.0037)
	Q3 - Q4	-0.0001	-0.0050	-0.0082***	-0.0085***	0.0038***	-0.0147***	0.0032	0.0098***
		(0.0008)	(0.0042)	(0.0006)	(0.0013)	(0.0007)	(0.0013)	(0.0035)	(0.0013)
2016	Q1 - Q2	-0.0006	-0.0010	-0.0020*	-0.0033**	0.0005	-0.0010	-0.0005	0.0007
		(0.0007)	(0.0029)	(0.0011)	(0.0014)	(0.0007)	(0.0023)	(0.0004)	(0.0011)
	Q3 - Q4	0.0006	-0.0014	0.0009	-0.0020	0.0011	0.0024*	-0.0003	0.0004
		(0.0008)	(0.0024)	(0.0008)	(0.0014)	(0.0009)	(0.0012)	(0.0004)	(0.0005)
2017	Q1 - Q2	-0.0004	-0.0125**	0.0001	-0.0056*	-0.0048***	-0.0076	-0.0013	-0.0173*
		(0.0022)	(0.0049)	(0.0030)	(0.0028)	(0.0010)	(0.0098)	(0.0047)	(0.0078)
2018	Q1 - Q2	0.0026***	0.0153***	0.0022*	0.0035	-0.0038	0.0089***	0.0029	-0.0049
		(0.0009)	(0.0044)	(0.0011)	(0.0025)	(0.0051)	(0.0005)	(0.0027)	(0.0066)
	Q3	0.0033***	0.0323***	0.0080***	0.0099**	0.0031***	0.0353***	0.0048***	0.0149***
		(0.0011)	(0.0050)	(0.0009)	(0.0043)	(0.0006)	(0.0001)	(0.0002)	(0.0005)

NOTE.—Difference-in-differences estimates of the impact of the immigration inflow on the residualized (after accounting for seasonal component and intercept) logarithms of the indicated outcomes in Orlando, on two-quarter periods, relative to the last half of 2017 (when the hurricanes hit and the migration from Puerto Rico to Orlando occurred). Columns 1-4 show estimates for the employment. Columns 5-8 show estimates for per-worker earnings. Estimated using the QCEW data with the synthetic Orlando for each outcome as the only control unit. Robust standard errors in parentheses. * Significance at the 10% level; ** significance at the 5% level; *** significance at the 1% level.

References

ABADIE, A. (Forthcoming): "Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects," *Journal of Economic Literature*.

ABADIE, A., A. DIAMOND, AND J. HAINMUELLER (2010): "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program," *Journal of the American Statistical Association*, 105, 493–505.

——— (2011): "SYNTH: Stata module to implement Synthetic Control Methods for Comparative Case Studies," *Statistical Software Components S457334*.

——— (2015): "Comparative Politics and the Synthetic Control Method," *American Journal of Political Science*, 59, 495–510.

ABADIE, A. AND J. GARDEAZABAL (2003): "The Economic Costs of Conflict: A Case Study of the Basque Country," *American Economic Review*, 93, 113–132.

ABADIE, A. AND J. L'HOUR (2019): "A Penalized Synthetic Control Estimator for Disaggregated Data," Working Paper.

AHN, T., P. ARCIDIACONO, A. HOPSON, AND J. THOMAS (2019): "Equilibrium Grade Inflation with Implications for Female Interest in STEM Majors," *Working Paper*.

ALEXANDER, M., E. ZAGHENI, AND K. POLIMIS (2019): "The impact of Hurricane Maria on out-migration from Puerto Rico: Evidence from Facebook data," Working Paper, arXiv.

ANDREWS, D. W. (2003): "End-of-Sample Instability Tests," *Econometrica*, 71, 1661–1694.

ARCIDIACONO, P. (2004): "Ability Sorting and the Returns to College Major," *Journal of Econometrics*, 121, 343–375.

ARCIDIACONO, P., J. V. HOTZ, AND S. KANG (2012): "Modeling College Major Choices Using Elicited Measures of Expectations and Counterfactuals," *Journal of Econometrics*, 166, 3–16.

ATHEY, S. AND G. IMBENS (2017): "The Econometrics of Randomized Experiments," *Handbook of Economic Field Experiments*, 1, 73–140.

ATTANASIO, O. AND K. KAUFMAN (2014): "Educational Choices and Returns to Schooling: Intrahousehold Decision Making, Gender and Subjective Expectations," *Journal of Development Economics*, 109, 203–2016.

AUTOR, D. H. AND D. DORN (2013): "The Growth of Low-skill Service Jobs and the Polarization of the US Labor Market," *American Economic Review*, 103, 1553–97.

AUTOR, D. H., D. DORN, AND G. H. HANSON (2013): "The China Syndrome: Local Labor Market Effects of Import Competition in the United States," *American Economic Review*, 103, 2121–2168.

AUTOR, D. H., L. F. KATZ, AND M. S. KEARNEY (2008): "Trends in US Wage Inequality: Revising the Revisionists," *The Review of Economics and Statistics*, 90, 300–323.

AZMAT, G., M. BAGUES, A. CABRALES, AND N. IRIBERRI (2019): "What You Don't Know... Can't Hurt You? A Natural Field Experiment on Relative Performance Feedback in Higher Education," *Management Science*, 65, 3714–3736.

AZMAT, G. AND N. IRIBERRI (2010): "The Importance of Relative Performane Feedback Information: Evidence from a Natural Experiment Using High School Students," *Journal of Public Economics*, 94, 435–452.

——— (2015): "The Provision of Relative Perofmance Feedback: An Analysis of Performance and Satisfaction," *Journal of Economics and Management Strategy*, 25, 77–110.

BAILEY, M. AND S. DYNARSKI (2011): "Gains and Gaps: Changing Inequality in US College Entry and Completion," *Working Paper*.

BAKER, R., E. BETTINGER, B. JACOB, AND I. MARINESCU (2018): "The Effect of Labor Market Information on Community College Students' Major Choice," *Economics of Education Review*, 65, 18–30.

BANDIERA, O., V. LARCINESE, AND I. RASUL (2015): "Blissful Ignorance? A Natural Experiment on the Effect of Feedback on Students' Performance," *Journal of Labour Economics*, 34.

BEERLI, A., J. RUFFNER, M. SIEGENTHALER, AND G. PERI (2018): "The Abolition of Immigration Restrictions and the Performance of Firms and Workers: Evidence from Switzerland," Working Paper, National Bureau of Economic Research.

BEN-MICHAEL, E., A. FELLER, AND J. ROTHSTEIN (2019): "Synthetic Controls and Weighted Event Studies with Staggered Adoption," Working Paper, arXiv:1912.03290.

BÉNABOU, R. (2015): "The economics of motivated beliefs," *Revue d'économie politique*, 125, 665–685.

BESHEARS, J., J. CHOI, D. LAIBSON, B. MADRIAN, AND K. MILKMAN (2015): "The Effect of Providing Peer Information on Retirement Savings Decisions," *The Journal of Finance*, LXX, 1161–1201.

BETTS, J. (1996): "What Do Students Know About Wages? Evidence From a Survey of Undergraduates," *Journal of Human Resources*, 31, 741–765.

BLEEMER, Z. AND B. ZAFAR (2018): "Intended College Attendance: Evidence from an Experiment on College Returns and Cost," *Journal of Public Economics*, 157, 184–211.

BOBBA, M. AND V. FRASINCO (2019a): "Learning About Oneself: The Effect of Signaling Academic Ability on School Choice," *Working Paper*.

——— (2019b): "Self-Perceptions About Academic Achievement: Evidence from Mexico City," *Working Paper*.

BODVARSSON, Ö. B., H. F. VAN DEN BERG, AND J. J. LEWER (2008): "Measuring Immigration's Effects on Labor Demand: A Reexamination of the Mariel Boatlift," *Labour Economics*, 15, 560–574.

BONESRØNNING, H. AND L. OPSTAD (2015): "Can Student Effort Be Manipulated?" *Applied Economics*.

BONFERRONI, C. (1935): "Il calcolo delle assicurazioni su gruppi di teste," *Rome: Tipografia del Senato*.

BORJAS, G. J. (2003): "The Labor Demand Curve is Downward Sloping: Reexamining the Impact of Immigration on the Labor Market," *The Quarterly Journal of Economics*, 118, 1335–1374.

——— (2008): "Labor Outflows and Labor Inflows in Puerto Rico," *Journal of Human Capital*, 2, 32–68.

——— (2017): "The Wage Impact of the Marielitos: A Reappraisal," *ILR Review*, 70, 1077–1110.

BOUND, J., M. LOVENHEIM, AND S. E. TURNER (2010): "Why Have College Completion Rates Declined?: An Analysis of Changing Student Preparation and College Resources," *American Economic Journal: Applied Economics*, 7, 1–31.

——— (2012): "Increasing Time to Baccalaureatte Degree in the United States," *Education Finance and Policy*, 7, 365–425.

Bound, J. and S. Turner (2002): "Going to War and Going to College: Did the G.I. Bill Increase Educational Attainment," *Journal of Labor Economics*, 20, 784–815.

Brade, R., O. Himmler, and R. Jäckle (2018): "Normatively Framed Relative Performance Feedback - Field Experiment and Replication," *MRPA Working Paper*.

Buckles, K., A. Hagemann, O. Malamud, and M. Morrill (2016): "The Effect of College Education on Mortality," *Journal of Health Economics*, 50, 99–114.

Bureau of Labor Statistics (2015): "Hispanics and Latinos in industries and occupations," https://www.bls.gov/opub/ted/2015/hispanics-and-latinos-in-industries-and-occupations.htm.

Cao, J. and C. Dowd (2019): "Estimation and Inference for Synthetic Control Methods with Spillover Effects," Working Paper, arXiv:1902.07343.

Card, D. (1990): "The Impact of the Mariel Boatlift on the Miami Labor Market," *ILR Review*, 43, 245–257.

Carnevale, A. P., N. Smith, and M. Melton (2011): "STEM," *Georgetown University Center on Education and the Workforce*.

Carrell, S., L. Lusher, and D. Rury (2020): "Major Disappointment: A Large Scale Experiment on (Non-)Pecuniary Information and College Major," *Working Paper*.

Carrell, S., M. Page, and J. West (2010): "Sex and Science: How Professor Gender Perpetuates the Gender Gap," *Quarterly Journal of Economics*, 125.

Cavallo, E., S. Galiani, I. Noy, and J. Pantano (2013): "Catastrophic Natural Disasters and Economic Growth," *Review of Economics and Statistics*, 95, 1549–1561.

Centro (2018): "New Estimates: 135,000+ Post-Maria Puerto Ricans Relocated to Stateside," Tech. rep.

Ceritoglu, E., H. B. G. Yunculer, H. Torun, and S. Tumen (2017): "The Impact of Syrian Refugees on Natives' Labor Market Outcomes in Turkey: Evidence from a Quasi-Experimental Design," *IZA Journal of Labor Policy*, 6, 5.

Chen, X. and M. Soldner (2014): "STEM Attrition: COllege Students' Paths Into and Out of STEM Fields," *IES Statistical Analysis Report*.

CHERNOZHUKOV, V., K. WUTHRICH, AND Y. ZHU (2017): “An Exact and Robust Conformal Inference Method for Counterfactual and Synthetic Controls,” Working Paper, arXiv:1712.09089.

CHETTY, R., J. N. FRIEDMAN, E. SAEZ, N. TURNER, AND D. YAGAN (2020): “Income Segregation and Intergenerational Mobility Across College in the United States,” *Quarterly Journal of Economics*, 1567–1633.

CHIAPPORI, P.-A., B. SALANIE, AND Y. WEISS (2017): “Partner Choice, Investment in Children, and the Marital College Premium,” *American Economic Review*, 107, 2109–2167.

CLEMENS, M. A. AND J. HUNT (2019): “The Labor Market Effects of Refugee Waves: Reconciling Conflicting Results,” *ILR Review*, 72, 818–857.

COHEN-GOLDNER, S. AND M. D. PASERMAN (2011): “The Dynamic Impact of Immigration on Natives’ Labor Market Outcomes: Evidence from Israel,” *European Economic Review*, 55, 1027–1045.

CONLON, J. (2020): “Major Malfunction: A Field Experiment Correcting Undergraduates’ Beliefs about Salaries,” *Journal of Human Resources*, Forthcoming.

DAMGAARD, M. T. AND H. S. NIELSEN (2018): “Nudging in Economics,” *Economics of Education Review*, 64, 313–342.

DE SILVA, D. G., R. P. MCCOMB, Y.-K. MOH, A. R. SCHILLER, AND A. J. VARGAS (2010): “The Effect of Migration on Wages: Evidence from a Natural Experiment,” *American Economic Review*, 100, 321–26.

DIZON-ROSS, R. (2019): “Parents’ Beliefs about Their Children’s Academic Ability: Implications for Educational Investments,” *American Economic Review*, 109, 2728–2765.

DOUDCHENKO, N. AND G. W. IMBENS (2016): “Balancing, Regression, Difference-in-Differences and Synthetic Control Methods: A Synthesis,” Working Paper, National Bureau of Economic Research.

ERSOY, F. (2019a): “Effects of Perceived Productivity on Study Effort: Evidence from a Field Experiment,” *Working Paper*.

——— (2019b): “Returns to Effort: Experimental Evidence from an Online Language Platform,” *Working Paper*.

FERMAN, B. AND C. PINTO (2017): “Placebo Tests for Synthetic Controls,” Working Paper.

FIRPO, S. AND V. POSSEBOM (2018): "Synthetic Control Method: Inference, Sensitivity Analysis and Confidence Sets," *Journal of Causal Inference*, 6.

FRAJA, G. D., T. OLIVEIRA, AND L. ZANCHI (2010): "Must Try Harder: Evaluating the Role of Effort in Educational Attainment," *Review of Economics and Statistics*, 92, 577–597.

FRANCO, C. (2020): "How Does Relative Performance Feedback Affect Beliefs and Academic Decisions," *Working Paper*.

FRENCH, R. AND P. OREOPOULOS (2017): "Behavioral Barriers Transitioning to College," *Labour Economics*, 47, 48–63.

FRYER, R. (2016): "Information, non-Financial Incentves, and Student Achievement: Evidence from a Text Messaging Experiment," *Journal of Public Economics*, 144, 109–121.

GNEEZY, U., J. LIST, X. JIN, AND S. SADOFF (2019): "Measuring Student Success: The Role of Effort on the Test Itself," *American Economic Review: Insights*, Forthcoming.

GOLIGHTLY, E. K. (2020): "Does College Access Increase High School Effort? Evaluating the Impact of the Texas Top 10% Rule," *Working Paper*.

GONG, Y., T. R. STINEBRICKNER, AND R. STINEBRICKNER (2019): "Uncertainty About Future Income: Initial Beliefs and Resolution During College," *Quantitative Economics*, 10, 607–641.

GONZALEZ, N. (2017): "How Learning About One's Abilty Affects Educational Investments: Evidence From the Advanced Placement Program," *Mathematica Policy Research Working Paper*.

GOULAS, S. AND R. MEGALOKONOMOU (2018): "Knowing Who You Are; The Effect of Feedback on Short and Long-Term Outcomes," *Working Paper*.

GROEN, J. A., M. J. KUTZBACH, AND A. E. POLIVKA (2020): "Storms and Jobs: The Effect of Hurricanes on Individuals' Employment and Earnings over the Long Term," *Journal of Labor Economics*, 38, 653–685.

GROGGER, J. AND G. H. HANSON (2011): "Income Maximization and the Selection and Sorting of International Migrants," *Journal of Development Economics*, 95, 42–57.

HAHN, J. AND R. SHI (2017): "Synthetic Control and Inference," *Econometrics*, 5, 52.

HISHLEIFER, S. R. (2016): "Incetives for Inputs or Outputs? A Field Experiment to Improve Student Performance," *Working Paper*.

HOLM, S. (1979): “A Simple Sequentially Rejective Multiple test Procedure,” *Scandinavian Journal of Statistics*, 6, 65–70.

HUNT, J. (1992): “The Impact of the 1962 Repatriates from Algeria on the French Labor Market,” *ILR Review*, 45, 556–572.

KAUFMAN, K. M. (2014): “Understanding the Income Gradient In College Attendance In Mexico: The Role of Heterogeneity In Expected Returns,” .

KAUL, A., S. KLÖSSNER, G. PFEIFER, AND M. SCHIELER (2015): “Synthetic control methods: Never Use All Pre-Intervention Outcomes Together With Covariates,” Working Paper, MPRA:83790.

KIRKEBOEN, L. J., E. LEUVEN, AND M. MOGSTAD (2016): “Field of Study, Earnings, and Self Selection,” *Quarterly Journal of Economics*, 131, 1057–1112.

LAVECCHIA, A., H. LIU, AND P. OREOPOULOS (2016): “Behavioral Economics of Education: Progress and Possibilities,” *Handbook of Economics of Education (Eric Hanushek, Stephen Machin, Ludger Woessmann, editors*, 5, 1–74.

LI, H.-H. (2018): “Do Mentoring, Information and Nudge Reduce the Gender Gap in Economics,” *Economics of Education Review*, 64, 165–183.

LIST, J., A. M. SAIK, AND Y. XU (2019): “Multiple Hypothesis Testing in Experimental Economics,” *Experimental Economics*, 22, 773–793.

MANSKI, C. F. (1993): “Adolescent Econometricians: How Do Youth Infer the Returns to Schooling,” *Studies of Supply and Demand in Higher Education*.

——— (2004): “Measuring Expectations,” *Econometrica*, 72, 1329–1376.

METCALF, R., S. BURGESS, AND S. PROUD (2019): “Students’ Effort and Educational Achievement: Using the timing of the World Cup to Vary the Value of Leisure,” *Journal of Public Economics*, 172, 111–126.

MURPHY, R. AND F. WEINHARDT (2020): “Top of the Class: The Importance of Ordinal Rank,” *Review of Economic Studies*, Forthcoming.

OLSON, S. AND D. G. RIORDAN (2012): “Engage to Excel: Producing One Million Additional College Grduates with Degrees in Science, Technology, Engineering, and Mathematics,” *Report to the President*.

OREOPOULOS, P., R. W. PATTERSON, U. PETRONIJEVIC, AND N. G. POPE (2020): "Lack of Study Time is the Problem, but What is the Solution? Unsuccessful Attempts to Help Traditional and Online College Students," *Journal of Human Resources*, Forthcoming.

OREOPOULOS, P. AND U. PTRONIJEVIC (2019): "The Remarkable Unresponsiveness of College Students to Nudging and What We Can Learn from It," *NBER Working Paper 26059*.

OTTAVIANO, G. I. AND G. PERI (2012): "Rethinking the Effect of Immigration on Wages," *Journal of the European Economic Association*, 10, 152–197.

OWEN, S. (2020): "College Field Specialization and Beliefs about Relative Performance: An Experimental Intervention to Understand Gender Gaps in STEM," *Working Paper*.

OZEK, U. (Forthcoming): "Examining the Educational Spillover Effects of Severe Natural Disasters: The Case of Hurricane Maria," *Journal of Human Resources*.

PERI, G. AND V. YASENOV (2019): "The Labor Market Effects of a Refugee Wave: Synthetic Control Method Meets the Mariel Boatlift," *Journal of Human Resources*, 54, 267–309.

PORTER, C. AND D. SERRA (2019): "Gender Differences in the Choice of Major," *American Economic Journal: Applied Economics*.

RAYER, S. (2018): "Estimating the Migration of Puerto Ricans to Florida Using Flight Passenger Data," Tech. rep., Bureau of Economics and Business Research.

REUBEN, E., M. WISWALL, AND B. ZAFAR (2015): "Preferences and Biases in Educational Choices and Labour Market Expectations: Shrinking the Black Box of Gender," *Economic Journal*, 127, 2153–2186.

RINGOLD, D. J. (2002): "Boomerang Effects in Response to Public Health Interventions: Some Unintended Consequences in the Alcoholic Beverage Market," *Journal of COnsumer Policy*, 25, 27—63.

RUGGLES, S., S. FLOOD, S. FOSTER, R. GOEKEN, J. PACAS, M. SCHOUWEILER, AND M. SOBEK (2021): "Changes to Counties and County Equivalent Entities: 1970-Present," [dataset], IPUMS USA: Version 11.0, Minneapolis, MN: IPUMS, https://doi.org/10.18128/D010.V11.0.

RURY, D. AND S. CARRELL (2020): "Knowing What it Takes: The Effect of Information About Returns to Studying on Study Effort and Achievement," *Working Paper*.

SCHACHTER, J. AND A. BRUCE (2020): "Revising Methods to Better Reflect the Impact of Disaster," Tech. rep., U.S. Census Bureau, https://www.census.gov/library/stories/2020/08/estimating-puerto-rico-population-after-hurricane-maria.html.

STINEBRICKNER, R. AND T. R. STINEBRICKNER (2003): "Working During School and Academic Performance," *Journal of Labor Economics*, 93, 473–491.

——— (2004): "Time Use and College Outcomes," *Journal of Econometrics*, 21, 243–269.

——— (2008): "The Causal Effect of Studying on Academic Performance," *The B.E. Journal of Economic Analysis and Policy*, 14, 1–53.

——— (2013): "A Major in Science? Initial Beliefs and Final Outcomes for College Major and Dropout," *Review of Economic Studies*, 81, 426–472.

TUMEN, S. (2015): "The Use of Natural Experiments in Migration Research," *IZA World of Labor*.

U.S. CENSUS BUREAU (2017): "EMPLOYMENT STATUS, 2017 American Community Survey 1-Year Estimates," https://data.census.gov/cedsci/deeplinks?url=https%3A%2F%2Ffactfinder.census.gov%2Ffaces%2Ftableservices%2Fjsf%2Fpages%2Fproductview.xhtml%3Fpid%3DACS_17_1YR_S2301&prodType=table.

——— (2018): "Estimates of the Components of Resident Population Change: April 1, 2010 to July 1, 2018," https://data.census.gov/cedsci/deeplinks?url=https%3A%2F%2Ffactfinder.census.gov%2Ffaces%2Ftableservices%2Fjsf%2Fpages%2Fproductview.xhtml%3Fpid%3DPEP_2018_PEPTCOMP&prodType=table.

WHALEN, C. AND V. VAZQUEZ-HERNANDEZ (2005): *Puerto Rican Diaspora: Historical Perspectives*, Temple University Press.

WILTSHIRE, J. C. (2021): "Walmart Supercenters and Monopsony Power: How a Large, Low-Wage Employer Impacts Local Labor Markets," Working Paper.

WISWALL, M. AND B. ZAFAR (2015a): "Determinants of College Major Choices: Idenfitication from an Information Experiment," *Review of Econoic Studies*, 82, 791–824.

——— (2015b): "How Do College Students Respond to Public Information about Earnings," *Journal of Human Captial*, 9, 117–169.

——— (2018): "Preference for the Workplace, Investment in Human Capital, and Gender," *Quarterly Journal of Economics*, 457–507.

ZAFAR, B. (2011): "How Do Students Form Expectations?" *Journal of Labor Economics*, 29, 301–348.

——— (2013): "College Major Choice and the Gender Gap," *Journal of Human Resources*, 48, 545–595.

ZIMMERMAN, S. D. (2014): "The Returns to College Admission for Academically Marginal Students," *Journal of Labor Economics*, 32, 711–754.

——— (2019): "Elite College and Upward Mobility to Top Jobs and Top Incomes," *American Economic Review*, 109, 1–47.