The MIT Undergraduate Journal of Economics

Volume XXII

2022-2023

The MIT Undergraduate Journal of Economics Volume XXII

2022-2023

Mailing Address:

The MIT Undergraduate Journal of Economics
Massachusetts Institute of Technology, Building E52-301
Cambridge, MA 02139

Foreword

"Money… must always be scarce with those who have neither wherewithal to buy it, nor credit to borrow it."

- *Adam Smith*

As MIT undergraduate economics students progress through their coursework, they are continuously introduced to new economic topics, constantly learning the ideas and models of established economists, and relentlessly being challenged to think differently about the observable phenomena around them. It is this enthusiasm for learning that led undergraduates at MIT to proceed in their own research—to experience the excitement of asking a question and striving to answer it. We hope that this year's papers highlight the vigor with which our undergraduate students pursue economic research and the rigor with which they present their ideas.

The publication of this Journal is made possible by the support of many people. We especially thank Professor Dave Donaldson for selecting the articles for this year's publication.

These relevant student papers demonstrate the enduring importance of rigorous economic research in the days ahead.

The MIT Undergraduate Journal of Economics Volume XXII

2022-2023

Contents

The Effect of Low Income Housing Tax Credit Developments on Crime: A Study of New York City

Dev Patale
dpatale@mit.edu

January 9, 2023

Abstract

This paper analyzes the impact of new rental housing developments funded with the Low Income Housing Tax Credit (LIHTC), the largest federal project based housing program, on crime in the surrounding neighborhood. Employing a difference in differences event study with geo-coded microdata in New York City from 2010 to 2017, I do not find statistically significant evidence of an effect on crime by LIHTC developments. To provide more color, at 95% confidence, the minimum instantaneous treatment effect (i.e. the effect within the first quarter upon an LIHTC development's opening) implied is a decrease in the total quarterly crime rate by 51 quarterly crimes per square mile, while the maximum instantaneous treatment effect implied is an increase in the total quarterly crime rate by 41 quarterly crimes per square mile. Interestingly, I find that LIHTC developments tend to be systematically placed into high-crime neighborhoods, which is consistent with policy objectives and existing literature's findings. These results are interesting because they do not support the theory that LIHTC developments are causally responsible for higher crime rates and thus could have notable implications for the public's perception of such projects.

1 Introduction

Since the inception of federal public subsidized housing programs, there has been much controversy around the impacts of such projects, particularly on neighborhood property values, crime rates, and academic achievement for local school districts. For example, The Ethel Lawrence Homes project in Mount Laurel, New Jersey experienced decades of opposition, with the community expecting increasing crime rates and falling property values to result from its opening. The project became so contentious that the Mount Laurel Doctrine was conceived, in which the New Jersey Supreme Court ruled that municipalities had an obligation to provide affordable housing. Such concerns have extended to Low Income Housing Tax Credit (LIHTC) developments, given that the LIHTC represents a growing proportion of all federal subsidized housing programs of this nature. This leads to a natural question that I aim to explore: what is the effect of LIHTC developments on crime rates? To answer this question, I investigated the New York City setting from 2010 to 2017, employing a difference in differences event study using geo-coded microdata of LIHTC development openings and NYPD crime data. I also examine how the impacts of LIHTC developments vary across different types of crime, such as property and violent crime.

The LIHTC program was not only chosen for this study because it is the largest federal affordable housing program but also because all developments are a matter of public record and thus easier to examine. New York City has robust databases serving a variety of purposes; for this study, the New York City OpenData Project provides access to the NYPD Reported Crime Database, which I elaborate on in Section 4. Additionally, the NYU Furman Center provides a detailed record of LIHTC development openings over time, complete with geo-coded microdata. As Butts (2021) discusses, the use of geo-coded microdata for difference in differences studies has become increasingly common. Drawing on past literature - particularly Diamond and McQuade (2019) and Woo and Joh (2015) - I utilize the Ring Method in establishing my difference in differences event study. After determining an appropriate inner ring radius over which the treatment effect exists, as well as an outer ring radius in which the control neighborhood exists, I am able to examine the effect of an LIHTC development opening on crime rates. Figure 3 illustrates this method pictorially. In my event study, I examine 473 LIHTC development openings over the period 2010-2017, with 8 quarters of pre-opening crime data and 8 quarters of post-opening crime data for each LIHTC; I find no statistically significant evidence that LIHTC developments have an effect on crime, but I do find statistically significant evidence that LIHTC developments are systematically placed into neighborhoods that have high crime rates. This is an interesting result, as this study cannot support the causation theory, which many individuals believe, that LIHTC developments cause higher crime rates. It does, however, support the selection theory that LIHTC developments are systematically opened in higher-crime neighborhoods, which may have notable implications for the public's perception of such projects and thus help reduce opposition to new LIHTC developments. Providing more color on my findings regarding the causal impact of LIHTC developments on crime, I find that at 95% confidence, the minimum instantaneous treatment effect (i.e. the effect within the first quarter upon an LIHTC development's opening) is a decrease in the total quarterly crime rate by 51 quarterly crimes per square mile, while the maximum instantaneous treatment effect is an increase in the total quarterly crime rate by 41 quarterly crimes per square mile. This heterogeneity in possible treatment effects may result from differences in neighborhoods themselves, as I opine on in the next paragraph. More details on the 95% confidence intervals that I find are discussed in Section 6.

This paper's results add to a literature that have previously found varying results in different settings. For example, Woo and Joh (2015) investigate Austin, Texas from 2000 to 2009 using an Adjusted Interrupted Time Series Difference in Differences approach and find that LIHTC developments are opened in high-crime neighborhoods. However, they also found that LIHTC developments helped mitigate crime rates. Interestingly, they discuss how "additional research is needed to better understand the conditions under which subsidized housing developments may raise or decrease neighborhood crime, such as the income-level of neighborhoods, because subsidized housing developments in affluent neighborhoods may have different impacts on neighborhood crime than those in poorer neighborhoods." This directly leads to Diamond and McQuade (2019), who develop a new difference-in-differences style estimator and find that for low-income neighborhoods, LIHTC developments lower crime. They find that for high-income neighborhoods, LIHTC developments do not increase crime. These differences are further discussed in the Literature Review (Section 3), and for my paper, I leverage the 95% confidence intervals that accompany my estimates to provide more color around my results, with the understanding that heterogeneity may be found across different types of neighborhoods. My study still finds an interesting result in the systematic placement of LIHTC developments in higher-crime areas that aligns with literature. As discussed in Section 6, Freedman and Owens (2011) also find that the LIHTC pushes for new low-income housing development in poorer areas, which have higher crime rates (Krivo and Peterson (1996); Patterson (1991); Peterson, Krivo, and Harris (2000)).

The paper proceeds as follows. Section 2 provides institutional background details about the Low Income Housing Tax Credit program. Section 3 briefly reviews the associated literature on the topic. Section 4 discusses the data sources utilized as well as data cleaning methodologies and sample restrictions conducted for the study. In section 5, I discuss the empirical methodology behind my study and discuss relevant assumptions. Section 6 then analyzes the results of the study. Section 7 provides concluding remarks and final discussions.

2 Background: The Low Income Housing Tax Credit

Started in 1986, the Low Income Housing Tax Credit (LIHTC) has been a primary tool in the production of affordable rental housing across the country, with approximately 3 million units having been placed into service. Federal tax credits are distributed to states each year on the basis of population, and regulations seek to stimulate development in high poverty census tracts (e.g. Qualified Census Tracts) as well as other high-priority settings; states then competitively allocate the credits to developers, which they can sell to investors to raise equity capital for qualifying projects. In return, investors receive a dollar-for-dollar credit for their federal tax liabilities over a ten-year period.

Developers become eligible to receive credits to build low-income housing in any area as long as the project is for construction or rehabilitation of a residential rental property and meets one of two occupancy criteria: at least 20% of tenants must have household incomes below 50% of the area median gross income (AMGI), or at least 40% of tenants must have household incomes below 60% of the AMGI[1]. Also, rents must be restricted in low-income units to 30% of the relevant income limit, either 50% or 60% of AMGI for a minimum peiod of 30 years. Because the size of the credit partially depends on the share of units set aside for low-income units, over 90% of the units in LIHTC projects qualify as low-income in practice.

States tend to receive significantly more project proposals and tax credit allocation requests from developers than they have federal allotments, so each state must maintain a Qualified Application Plan to help govern the selection process. Typically, point scores are assigned on the basis of various project characteristics, after which tax credits are allocated based on these point totals. Characteristics include features such as tenant demographics, project location, alternative funding sources, and structural properties of the building.

As part of the Omnibus Reconciliation Act of 1989, Congress passed legislation that allowed specific LIHTC projects to be eligible for a 30% increase in their credit allocation; such projects must be built in very low-income areas, termed Qualified Census Tracts (QCTs), or in areas with relatively high construction costs, termed Difficult Development Areas (DDAs). A census tract qualifies as a QCT if at least 50% of its households have incomes below 60% of the AMGI or if the poverty rate of the tract is at least 25%. A DDA is a metropolitan area, county (or county equivalent), or census place with high construction, land, and utility costs relative to the AMGI. As found in literature, nearly all state QAPs indicate that developers locating in high poverty areas receive preference in the qualification process. Given the level of uncertainty that developers face about whether the state will approve their LIHTC application, locating in a QCT increases the

[1]The income limits are dependent on household size. For the 50% of AMGI limit, this is established for a family size of four. Limits are then are adjusted upward by 4% for each family member above four and adjusted downward by 5% for each family member below four. For the 60% of AMGI limit, the above figures are multiplied by 1.2.

probability of receiving a tax credit.

The value of the tax credits received by eligible developers is based on the project's qualified basis, which reflects the cost of constructing or rehabilitating the low-income units. Once the qualified basis has been determined, the annual tax credit is then calculated by applying the relevant tax credit rate: new construction or substantial rehabilitation projects receive a 9% credit rate, while all other projects receive a 4% credit rate[2]. These annual credits are then paid out over a 10-year period.

3 Literature Review

This section reviews the existing literature on LIHTC developments' neighborhood impacts. Before reviewing literature that studies crime as an outcome of interest, I briefly elaborate on literature that touches on other outcomes of interest. In New York City, Schwartz, Ellen, Voicu, and Schill (2006) find that low income housing developments have large positive effects on local housing values; according to their discussion, this may be a result of a positive amenity effect from new development construction. Green, Malpezzi, and Seah (2002) examine LIHTC projects in Milwaukee and find weak evidence that the developments decrease property values but find ambiguous evidence for other areas. Sinai and Waldfogel (2005) and Eriksen and Rosenthal (2010) both study the crowd-out effects of subsidized affordable housing construction on private rental development and find large crowd-out effects. Lee, Culhane, and Wachter (1999) examine the correlation between the location federally subsidized housing units and surrounding property values; they find that the relationship depends on the type of program. However, public housing developments, users of Section 8 vouchers, and LIHTC developments are associated with declines in housing values in the study. Baum-Snow and Marion (2009) study the effects of LIHTC-funded developments in low-income neighborhoods on new con-

[2]This is before any additional credit increase from being located in a QCT or DDA is applied.

struction, property values, and median incomes at the census block group level. They found that property values increase in low-income areas.

Regarding literature that examines crime as an outcome of interest, there have been several works worth noting. Roncek, Bell, and Francik (1981) and McNulty and Holloway (2000) have drawn links between subsidized housing and crime, discussing how the demographic groups most often involved in criminal activity are disproportionately tenants of low-income housing developments. Thus, they argue that building new affordable housing units could increase criminal activity by attracting individuals who might be more prone to such activity. According to Glaeser, Sacerdote, and Scheinkman (1996) and Bjerk (2010), the concentration of poverty itself could further exacerbate crime problems in neighborhoods. Multiple studies have examined crime at the individual level and have found that randomly moving people to more affluent communities does reduce individual criminal behavior levels (Kling, Ludwig, and Katz (2005); Harcourt and Ludwig (2006); Ludwig and Kling (2007)). Additionally, Freedman and Owens (2011) study the impact of LIHTC developments, specifically, on crime at the county level and do not find unambiguous results. Woo and Joh (2015) estimate the impact of the LIHTC program on neighborhood crime rates using an adjusted interrupted time series difference-in-difference specification. They find that LIHTC developments tended to be developed in neighborhoods where crime was already prevalent and that LIHTC developments had a mitigating impact on neighborhood crime.

This paper also builds on the seminal work by Diamond and McQuade (2019), which leverages microdata to study highly granular effects of LIHTC developments across the United States. They developed a novel difference-in-differences style estimator to analyze a wide array of neighborhoods and counties, demonstrating that affordable housing has significantly different effects on neighborhoods based on income and the minority share of the neighborhood population. They find that, in low-income neighborhoods, LIHTC development increases home prices 6.5%, lowers crime rates, and attracts racially and income diverse populations. LIHTC development in higher-income areas causes home prices to decline 2.5% and attracts lower-income households.

As discussed in the Introduction, Woo and Joh (2015) and Diamond and McQuade (2019) find meaningfully different results in terms of how LIHTC developments affect crime in the neighborhoods that they are placed. These differences in results from my paper are largely attributed to their ability to control for income levels across neighborhoods in the studies, which was not possible for my setting given how high-resolution the study was and the lack of institutionalized data for this parameter. Woo and Joh (2015) conclude their paper opining on the importance of understanding income levels of neighborhoods for the causal impact of LIHTC developments on crime, as they can be heterogeneous. Since my study does not control for neighborhood income levels, I discuss a few of the 95% confidence intervals implied by my estimates to provide a sense of possible maximum and minimum treatment effects.

4 Data

4.1 NYU Furman Center for Real Estate and Urban Policy

My first dataset is from the NYU Furman Center, which provides detailed information of each property funded with the Low Income Housing Tax Credit (LIHTC) in New York City, including address, the number of residential units, year built, subsidy start date, etc. The dataset also provides latitude and longitude points of the units for geospatial analysis. Before any sample restrictions, the database contained information on 21,762 low-income housing units placed into service as of the second quarter of 2022.

I restricted my analysis to LIHTC units whose subsidy began between 2010 and 2017, using this as a proxy for an LIHTC development opening, since this is the point at which subsidized units are available for low-income tenants. I also removed LIHTC units whose subsidy start date was unknown as well as whose number of residential units was unknown. Additionally, I did not consider LIHTC developments that were specific to elderly or disabled individuals, as this population is separate from that of interest for the study. To avoid treatment spillover effects for my analysis (described in the Empirical Methodology section), I removed instances in which LIHTC units overlapped within a specified distance of one another over the study period (as described in the Empirical Methodology section). Thus, my sample of interest is specific to LIHTC developments that were placed into service into a region where the direct neighborhood (treatment area) and immediate surrounding area (control area) were unaffected by other LIHTC development openings. This left me with 473 LIHTC development openings over the 7 year period from 2010-2017; Figure 1 provides a visual mapping of the units across New York City.

4.2 New York City OpenData Project

My second dataset is from the New York City OpenData Project, which publishes New York Police Department (NYPD) Reported Crime data; the dataset includes all valid felony, misdemeanor, and violation crimes reported to the NYPD from 2006 to the end of 2019, with latitude and longitude points for geospatial analysis. Before any sample restrictions, the database contained information on nearly 8 million reported crime instances.

In line with my Empirical Methodology described below, I restricted my analysis to reported crime instances that fall within the study period of 2008 to 2019; this helps me examine two years' worth of pre-trend crime data before any instance of an LIHTC opening as well as two years' worth of post-trend crime data. The database contained several redundant offense classifications, which I collapsed together based on their offense codes; note that this is not necessary and was only performed to simplify the process of defining crimes as Part I/II as well as Violent/Property/Neither as detailed shortly. I also collapsed theft-fraud into the broader classification of fraud crimes for lack of better granularity, as well as assuming that instances of felony assault were aggravated assault and that all instances of "Assault 3" were simple assault. This was a reasonable assumption stemming from the offense codes.

After cleaning the dataset and removing redundancies as described above, I utilized the FBI Uniform Crime Reporting (UCR)'s definition of Part I and Part II Crime to categorize each reported crime instance. Regarding granular assumptions, I considered all forms of negligent homicide as Part I crimes, harassment as Part I, all unclassified and miscellaneous crimes as Part II, criminal trespassing as Part II, and all vehicle and traffic violations as Part I. Note that under these definitions, all crimes are classified as either Part I or Part II. I then leveraged the FBI UCR's definition of Property and Violent Crimes to also classify crimes as Property, Violent, or Neither; note that these classifications are independent of whether the crime is Part I or Part II. Regarding granular assumptions, I considered "offenses against the person" as Violent and all sex crimes as Violent.

After cleaning the dataset and labeling crimes according to the methodology above, I was left with nearly 6 million instances of reported crime over the 11 year period from 2008 to 2019. Note that there exist a certain number of crime instances that are geotagged to the nearest NYPD precinct because of an invalid reported location; although the database's footnotes do not provide a number, it appears to be a relatively small number and thus should not be significant given the scale of the study.

4.3 Descriptive Statistics

I establish the inner ring radius to be 0.2 miles and the outer ring radius to be 0.3 miles, which remains consistent with Diamond and McQuade (2019) and Woo and Joh (2015) as well as leads to the reasonable descriptive statistics discussed below. I also elaborate on this choice in Section 5. Recall that the study runs from 2010 to 2017, with 473 LIHTC openings under study, and 8 quarters of pre-opening crime rates as well as 8 quarters of post-opening crime rates, each delineated by inner ring and outer ring (per LIHTC opening). This results in the summary statistics for quarterly crime rates across inner rings, outer rings, and both rings outlined in Tables 1, 2, and 3. Figure 2 also describes the distribution of LIHTC development openings across boroughs during the study period.

From tables 1, 2, and 3, we note that there are instances of quarters where the rates for specific types of crime in the corresponding ring are zero. With Property and Violent crime rates, this is sensible because even in instances where both types of crime rates are zero for a given quarter, there still exists crime that is not classified as one of the two (described earlier as "Neither"). For all types of crime, no LIHTC location or time period systematically accounts for a significant share of the zero-crime rate observations. More broadly, however, there may be heterogeneity given that the NYPD database naturally only contains reported crime; thus, there is a potential that certain neighborhoods have systematically lower reported crime rates because many go unreported. In addition, certain crime instances in the database are geo-tagged to the nearest NYPD location because invalid reported locations. However, these two sources of noise are mitigated by the large sample of LIHTC openings and the 17 quarter time-frame of the study design, which provides a large number of observations over which to smooth out noise.

5 Empirical Methodology

5.1 Empirical Test

Butts (2021) formalizes an approach for estimating the effects of treatment at a specific location using geocoded microdata. This "Ring" estimator compares units immediately next to treatment (an inner-ring) to units just slightly further away (an outer-ring); for this paper, the inner-ring represents the localized neighborhood in which a new LIHTC development opens, and the outer-ring represents the control group and comparison region.

An illustration of the Ring Method can be found in Figure 3. The triangle at the center of the figure represents an LIHTC development opening (i.e. the location of the treatment), and the inner circle represents the area defined by the localized neighborhood. Each dot represents an instance of reported crime within a defined time window - this paper analyzes quarterly crime figures - in the treated area, whereas each triangle in the outer ring represents an instance of reported crime in the control region. All other units outside of the rings are not considered in the analysis for that particular sample. The Ring estimate for the treatment effect compares average changes in outcomes between units in the inner 'treated' ring and the outer 'control' ring to form an estimate for the treatment effect, i.e. a difference-in-differences estimator.

For this study, each LIHTC development opening yields a set of rings with two years of quarterly reported crime data pre-dating the opening and two years of quarterly data post-dating the opening. This allows me to then compute quarterly crime rates for the inner ring as well as the outer ring, forming the basis of my estimate for the treatment effect. A regression formulation can be shown below:

$$Y_{igt} = \alpha + \sum_{k=T_0}^{-2}(\beta_k \cdot \text{treated}_{igk}) + \sum_{k=0}^{T_1}(\beta_k \cdot \text{treated}_{igk}) + \sum_{k \neq -1} \mu_{gk} + \gamma_t + \epsilon_{gt}$$

Where Y_{igt} is the crime rate outcome (e.g. total crime, Part I crime, Part II crime, etc) for ring type $i \in$ {inner, outer} and LIHTC development $g \in [1, 473]$ (given that there are 473 LIHTC developments in the sample) at observation time t. Note that the "group" specification g is used to account for differences in baseline levels of crime rates across LIHTC developments in the event study. T_0 and T_1 allow us to define the upper-bound and lower-bound on leads and lags for the event study window; with 8 quarters of pre-opening data and 8 quarters of post-opening data, $T_0 = -8$ and $T_1 = 8$. k refers to the event time: specifically, this is the time period relative to LIHTC development g's opening time period. For example, $k = -1$ refers to the quarter before LIHTC development g opened. As is standard practice with event studies, the "-1" time period is used as the reference period and thus dropped from the regression to avoid perfect multicollinearity; in other words, β_{-1} is normalized to be equal to 0. treated$_{igk}$ is a dummy variable that equals 1 when i = inner and the current observation's time t relative to LIHTC development g's opening is equal to k (to ensure you're analyzing the correct event times). For example, if the current observation is in the inner ring at a time t that is one quarter after LIHTC development g's opening, then treated$_{igk} = 1$ for $k = 1$. When i = outer, treated$_{igk}$ is always equal to 0. This is because only inner rings are treated, not outer rings. μ_{gk} represents event-time fixed effects for each LIHTC development g, and γ_t represents observation time fixed effects. This regression specification thus allows us to study the difference in differences in quarterly crime rates between inner and outer rings pre- and post-treatment, controlling for the heterogeneity across LIHTC developments and across observation time. Standard errors are clustered at the LIHTC development/"group" level.

5.2 Assumptions

The key insight of the Ring Method is that since the treated and control units are all very close in physical location, the counterfactual untreated outcomes will approximately be equal for units within each ring. The first assumption is clearly the parallel trends assumption for the treated and control units; specifically, the average change in (counterfactual) untreated outcomes in the treated ring should be equal to the average change in the control ring. If this is satisfied, then the control units will be able to estimate the counterfactual trend for the treated units. The second assumption for the Ring Method as described by Butts (2021) requires an understanding of how far treatment effects are experienced (i.e. the radius of the inner ring). If the treated ring is too narrow, then units in the control ring experience effects of treatment and the change among 'control' units would no longer identify the counterfactual trend. On the other hand, if the treated ring is too wide, then the zero treatment effect of some unaffected units are averaged into the change among 'treated' units. Therefore, results would be biased towards zero.

For the parallel trends assumption, although there is no statistical test, visual inspection is useful with observations over many time points. Additionally, showing that the inner and outer rings are statistically the same before treatment ($\beta_k = 0$ for $k < 0$) supports the parallel trends assumption because this indicates that the difference in differences in quarterly crime rates is not statistically significant pre-treatment, which is what we desire. To determine the radii of the inner and outer rings, this paper draws on the insights from previous works employing the Ring Method - e.g. Diamond and McQuade (2019) - and adjusts downward for the unique geographic idiosyncrasies of New York City (for example, the fact that there is significantly higher population density in New York City than in other urban or suburban studies, which justifies the examination of a smaller ring). Although it is difficult to account for variation in block sizes across boroughs, an estimate of approximately 20 blocks per mile is utilized; thus, with current ring sizes, the inner ring assumes that treatment exists over 4 blocks, with the region between 4 to 6 blocks serving as a valid control region. I then carry out a sensitivity analysis with varying ring radii to provide a sense check for my results.

Note that I seek to study LIHTC openings over a window where there are no treament spillovers from overlapping LIHTC openings (i.e. in each other's inner or outer rings). Thus, over the study period 2008 to 2019, which includes the two years' worth of pre-opening and post-opening crime data I analyze, I remove all instances of LIHTC openings that have overlapping inner or outer rings. As a result, I restrict my analysis to regions where only one LIHTC opening per region during the study window impacts an inner ring (while the corresponding outer ring remains unaffected, in theory).

There are challenges with establishing and testing the assumptions associated with the empirical study. First, with such a high-resolution analysis, it becomes difficult to accurately assess demographic data across inner and outer rings because of how granular they are. Most Census data is recorded at a higher level (i.e. larger regions), so this is an important limitation. Furthermore, although reported crime data has been cleaned, there are certain instances that are not tagged to the precise location in which the crime took place; instead, they may be tagged at the closest Police Department, which introduces a level of noise to the analysis. Thus, there is a trade-off between the size of the inner and outer rings, where smaller radii result in a higher resolution study, and the impact of the noise from the reported crime data, which decreases with larger radii. Finally, the largest challenge involves establishing the proper radii for the inner and outer rings to ensure that the Ring Method estimates are not biased upwards or downwards. The proposed methodology from earlier may not perfectly define these radii, so this is a key limitation to be aware of for the study. However, a critical mitigant is that we can sensitize our results across different inner and outer ring radii, allowing us to form a more complete picture for the study.

6 Results

As seen in Table 4, there appears to be no statistically significant effect of an LIHTC opening on any type of crime in the surrounding (inner ring) neighborhood, as also verified through Figures 4, 5, and 6 (which provide the 95% confidence intervals of relevant regression coefficients for varying types of crime). Specifically, after an LIHTC opens (where the event time is ≥ 0), the coefficients for the interaction terms between inner ring and quarter are not statistically significant. However, as discussed previously, it is still important to discuss the confidence intervals that accompany my estimates given the heterogeneity in impacts that may exist based on differences in neighborhoods, especially in terms of income levels. Although all of the relevant 95% confidence intervals can be derived from Table 4, I focus on the Total Crime regression estimates, which is the most encompassing. Recall that the units are quarterly crimes per square mile (i.e. crimes in the quarter normalized per square mile). Also note that event-time of 0 (i.e. $k = 0$) is the instantaneous treatment effect because it examines the quarter in which the LIHTC development opens.

For k ranging from 0 to 8, the 95% confidence intervals are [-50.9,40.6], [-50.0, 51.5], [-46.9, 32.8], [-42.1, 32.7], [-58.0, 33.5], [-50.3, 61.6], [-77.6, 3.4], [-64.3, 20.0], and [-56.9, 34.5], respectively. This implies that the minimum treatment effects implied by the regression estimates are a decrease in the total quarterly crime rate for all 8 quarters after an LIHTC development opens. On the other hand, the maximum treatment effect implied by the estimates is an increase in the total quarterly crime rate for all 8 quarters after an LIHTC development opens. We find similar patterns for specific types of crime as well. These results are particularly interesting given that this study does not control for neighborhood income levels or any other variation in neighborhoods, which can be especially important for identifying how treatment effects may differ across neighborhoods. Given that past studies have found that LIHTC developments decrease crime rates in low-income neighborhoods, it is important to note this study's confidence interval endpoints.

Importantly, however, the regression coefficient for the Inner Ring indicator is large, positive, and statistically significant across all types of quarterly crime rates (each column). That is, on average, simply being located within the inner ring of an LIHTC opening results in an increase of approximately 142 total quarterly crimes per square mile, 32 Part 1 quarterly crimes per square mile, 39 Part 2 quarterly crimes per square mile, 8 Property quarterly crimes per square mile, and 25 Violent quarterly crimes per square mile. In other words, LIHTC developments are specifically placed in areas with higher crime (in the inner ring) compared with the neighboring outer ring according to my findings. This is an interesting but not surprising finding. Freedman and Owens (2011) find that "the LIHTC steers new low-income housing development toward poorer areas" since it provides relatively larger tax incentives to developers who build low income housing developments in high-poverty areas, as discussed in the Institutional Background section. Baum-Snow and Marion (2009) also show that the program promotes more affordable rental housing construction in low-income neighborhoods. Fallon and Price (2020) then discuss how "a substantial body of literature has found that Low-Income Housing Tax

Credit (LIHTC) buildings are concentrated in high poverty regions (Cummings and DiPasquale (1999); Oakley (2008)) where crime rates are higher (Krivo and Peterson (1996); Patterson (1991); Peterson et al. (2000))." Thus, it is no surprise that LIHTC openings are found to be selectively opened in neighborhoods (inner rings) with higher crime rates in my study. A natural extension of my study would be to examine the income bands of the neighborhoods in which my LIHTC developments exist; however, my study's findings are consistent with the paper's discussions above. Within New York City itself, according to the NYC Independent Budget Office (IBO (2019)), the most common income band for affordable housing financing is the low-income band. They find that 37 percent of low-income units are located in low-income neighborhoods, 28 percent in very low income neighborhoods, and 19 percent in moderate-income neighborhoods. Additionally, they discuss how "a growing body of research demonstrates that higher-income neighborhoods in the city tend to have lower rates of violent crime." My study's findings appear to support the IBO's discussion.

Appendix A also contains sensitivity analysis on different values for the inner and outer ring radii to further validate the results.

7 Conclusion and Discussion

The Low Income Housing Tax Credit has become the largest federal affordable housing program, and with it has continued concerns about its effect on crime rates. In my study, I employed a difference in differences event study using the Ring Method with geo-coded micro-data in New York City from 2010 to 2017 to investigate the effect of LIHTC developments on crime. Although I did not find statistically significant evidence that LIHTC developments cause higher crime rates, I did find that LIHTC developments tended to be systematically placed into neighborhoods with higher baseline crime rates, as was clear from the large, statistically significant positive coefficient on the inner ring indicator in my regression. It is interesting to note that Woo and Joh (2015) and Diamond and McQuade (2019) discuss how differences in neighborhoods (i.e. income levels) can and do lead to differences in treatment effects from LIHTC development openings; since this study does not control for such differences, a lack of statistically significant evidence for a causal impact on crime is not necessarily surprising. For example, at 95% confidence, the minimum instantaneous treatment effect implied is a decrease in the total quarterly crime rate by 51 quarterly crimes per square mile, while the maximum instantaneous treatment effect implied is an increase in the total quarterly crime rate by 41 quarterly crimes per square mile. The statistically significant result that my paper does find is interesting, supporting those found in Woo and Joh (2015) and being consistent with the policy objectives of the LIHTC program, as established by the Qualified Census Tract.

My study comes with certain limitations that may be expanded upon in the future. First, a formal empirical test was not undertaken to determine the ring radii. Instead, I relied upon existing literature as well as sensitivity analysis as found in Appendix A to validate my results, but this is an area that could be expanded upon. Additionally, I restricted my sample to LIHTC openings that did not overlap with one another; that is, for each LIHTC opening, no other LIHTC development opened in the inner or outer ring over the 4 year period. There could be interesting results from multiple LIHTC development openings occurring in the same neighborhoods, allowing for an expansion of this current study. Finally, as previously discussed, interesting results may arise with additional analysis of how neighborhoods differ in terms of income; this is an area of expansion.

Figure 1: Visualization of LIHTC Developments in the Sample of Interest.

Note: This refers to the 473 LIHTC developments under study from 2010-2017 in New York City such that no developments overlap with one another in inner or outer rings to ensure a clean study.

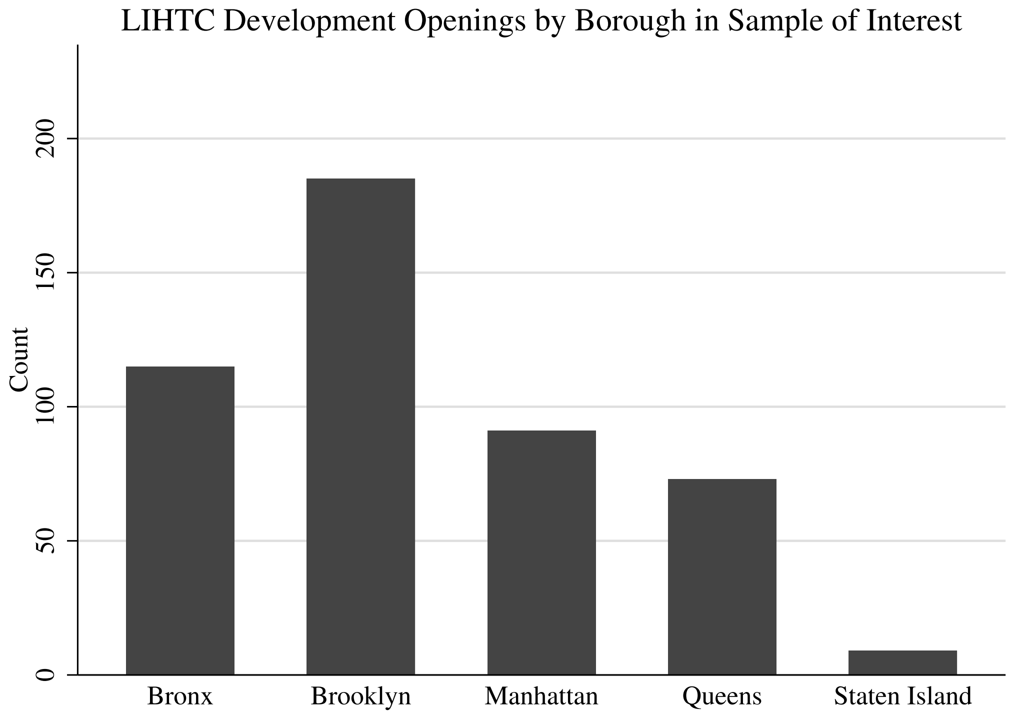


Figure 2: Note that this pertains to the 473 LIHTC openings under study.

Note: This refers to the 473 LIHTC developments under study from 2010-2017 in New York City such that no developments overlap with one another in inner or outer rings to ensure a clean study.

Table 1: Summary Statistics for Inner Ring Quarterly Crime Rates

	mean	sd	min	max	count
Total Crime Rate	1636.25	1434.53	15.92	20817.47	8025
Part 1 Crime Rate	335.03	416.70	0.00	8856.97	8025
Part 2 Crime Rate	483.09	398.59	0.00	3326.34	8025
Property Crime Rate	78.40	87.18	0.00	755.99	8025
Violent Crime Rate	254.09	381.36	0.00	8554.58	8025
Observations	8025				

Notes: Rates are defined as number of crimes in the quarter per square mile.
Recall that the inner ring is a radius of 0.2 miles from an LIHTC opening, and the outer ring is a radius of 0.3 miles (with the outer region being exclusive of the inner region).
The summary contains 8 quarters of pre-opening data and 8 quarters of post-opening data.

Table 2: Summary Statistics for Outer Ring Quarterly Crime Rates

	mean	sd	min	max	count
Total Crime Rate	1494.22	1209.52	12.73	15584.45	8041
Part 1 Crime Rate	305.13	343.40	0.00	6582.65	8041
Part 2 Crime Rate	441.98	343.87	0.00	2762.93	8041
Property Crime Rate	69.77	72.31	0.00	630.25	8041
Violent Crime Rate	231.85	314.71	0.00	6442.59	8041
Observations	8041				

Notes: Rates are defined as number of crimes in the quarter per square mile.
Recall that the inner ring is a radius of 0.2 miles from an LIHTC opening, and the outer ring is a radius of 0.3 miles (with the outer region being exclusive of the inner region).
The summary contains 8 quarters of pre-opening data and 8 quarters of post-opening data.

Table 3: Summary Statistics for Combined Ring Quarterly Crime Rates

	mean	sd	min	max	count
Total Crime Rate	1565.16	1328.55	12.73	20817.47	16066
Part 1 Crime Rate	320.07	382.05	0.00	8856.97	16066
Part 2 Crime Rate	462.51	372.76	0.00	3326.34	16066
Property Crime Rate	74.08	80.20	0.00	755.99	16066
Violent Crime Rate	242.96	349.76	0.00	8554.58	16066
Observations	16066				

Notes: Rates are defined as number of crimes in the quarter per square mile.
Recall that the inner ring is a radius of 0.2 miles from an LIHTC opening, and the outer ring is a radius of 0.3 miles (with the outer region being exclusive of the inner region).
The summary contains 8 quarters of pre-opening data and 8 quarters of post-opening data.

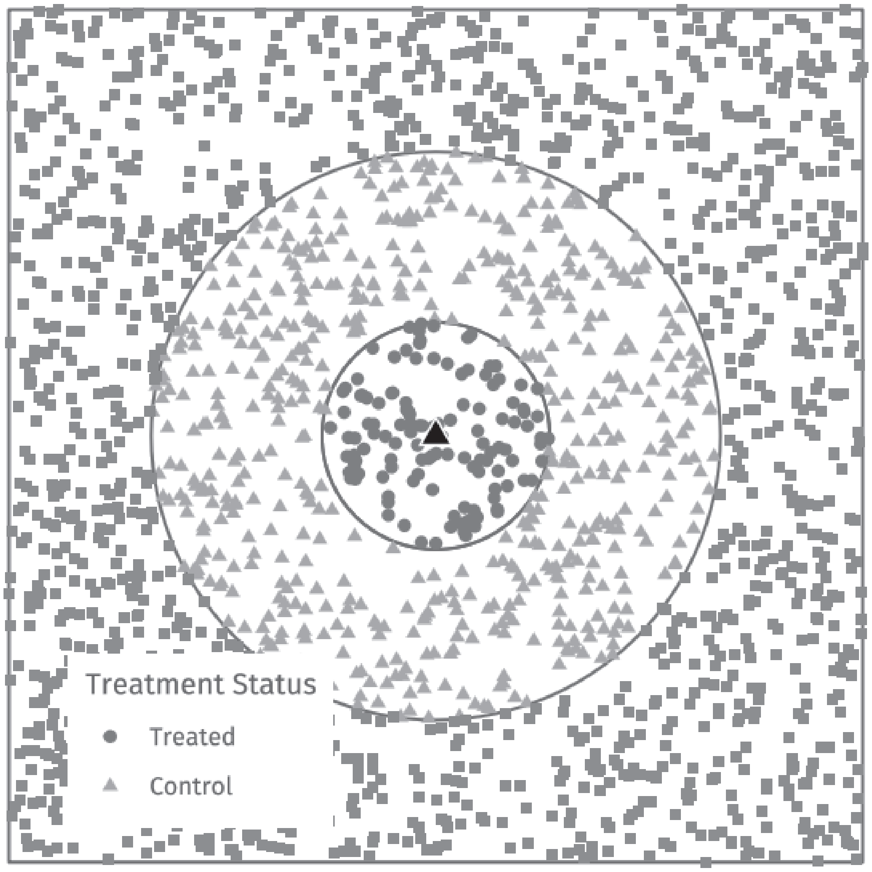


Figure 3: The Ring Method (Source: Butts (2021)).

Note: This is a method of conducting a difference-in-differences study with geo-coded micro-data.

Table 4: Regression Results for Quarterly Crime Rates

	(1)		(2)		(3)		(4)		(5)	
	Total Crime Rate		Part 1 Crime Rate		Part 2 Crime Rate		Property Crime Rate		Violent Crime Rate	
Inner Ring	141.76***	(46.73)	32.34**	(16.02)	38.54***	(11.86)	8.28**	(4.15)	25.16*	(15.24)
Inner Ring × Quarter: -8	11.89	(23.99)	2.34	(9.01)	3.60	(6.98)	-0.43	(2.35)	2.01	(8.38)
Inner Ring × Quarter: -7	22.35	(25.37)	4.16	(8.96)	7.02	(7.36)	-0.91	(2.50)	6.28	(8.02)
Inner Ring × Quarter: -6	-25.51	(19.45)	-9.55	(6.49)	-3.20	(6.54)	0.27	(2.28)	-8.73	(6.02)
Inner Ring × Quarter: -5	10.51	(20.52)	4.69	(6.53)	0.57	(7.19)	1.20	(2.21)	6.63	(5.90)
Inner Ring × Quarter: -4	3.66	(23.22)	5.36	(7.95)	-3.52	(7.11)	-0.90	(2.60)	4.26	(7.07)
Inner Ring × Quarter: -3	13.36	(25.47)	3.95	(8.96)	2.73	(7.26)	1.43	(2.44)	3.54	(7.90)
Inner Ring × Quarter: -2	11.78	(17.83)	-1.94	(5.81)	7.83	(6.26)	0.15	(2.45)	-4.19	(5.35)
Inner Ring × Quarter: 0	-5.16	(23.27)	-0.61	(7.54)	-1.97	(6.64)	1.63	(2.43)	-2.50	(6.59)
Inner Ring × Quarter: 1	0.73	(25.81)	1.34	(9.07)	-0.97	(6.78)	0.22	(2.50)	0.11	(7.98)
Inner Ring × Quarter: 2	-7.07	(20.28)	-2.58	(6.64)	-0.96	(6.82)	1.30	(2.46)	-4.62	(5.98)
Inner Ring × Quarter: 3	-4.67	(19.04)	-3.43	(5.86)	1.09	(7.00)	0.84	(2.37)	-3.86	(5.30)
Inner Ring × Quarter: 4	-12.29	(23.28)	-13.09*	(7.14)	6.94	(7.64)	-2.02	(2.44)	-12.11*	(6.33)
Inner Ring × Quarter: 5	5.67	(28.46)	0.64	(10.05)	2.19	(7.68)	0.75	(2.62)	0.02	(9.14)
Inner Ring × Quarter: 6	-37.12*	(20.61)	-14.50**	(6.46)	-4.06	(7.31)	-0.38	(2.35)	-14.61**	(6.11)
Inner Ring × Quarter: 7	-22.17	(21.46)	-13.14**	(6.44)	2.05	(7.89)	2.47	(2.34)	-15.72***	(5.85)
Inner Ring × Quarter: 8	-11.21	(23.28)	-15.31**	(7.52)	9.70	(7.98)	-2.05	(2.47)	-14.06**	(6.64)
Constant	1516.57***	(24.81)	311.41***	(8.01)	446.87***	(6.75)	71.74***	(2.28)	236.28***	(7.58)
Time Fixed Effects	Yes		Yes		Yes		Yes		Yes	
R^2	0.82		0.77		0.84		0.64		0.76	
Observations	16066		16066		16066		16066		16066	

Notes: Clustered standard errors in parenthesis. * p< 0.10, ** p< 0.05, *** p < 0.01.

The regressions establish the quarter preceding an LIHTC development opening (i.e. Quarter: -1) to be the baseline and thus omit the corresponding terms.

Recall that the inner ring is a radius of 0.2 miles from an LIHTC opening, and the outer ring is a radius of 0.3 miles (with the outer region being exclusive of the inner region).

The regression runs on 8 quarters of pre-opening data and 8 quarters of post-opening data.

One of the time periods must be dropped to avoid perfect multicollinearity. In this event study, the -1 time lag is used as the dropped reference.

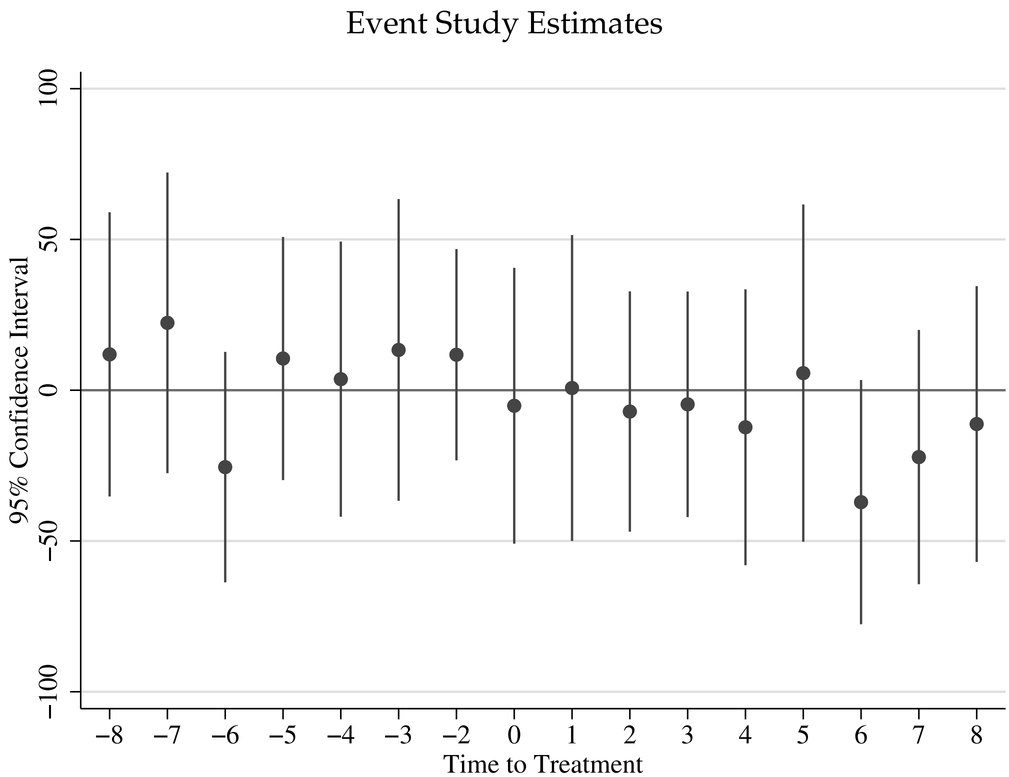


Figure 4: 95% CIs of Regression Coefficients for Total Crime Quarterly Rates

Note: The vertical axis refers to quarterly crimes per square mile. To avoid perfect multicollinearity, the -1 time lag is used as the dropped reference.

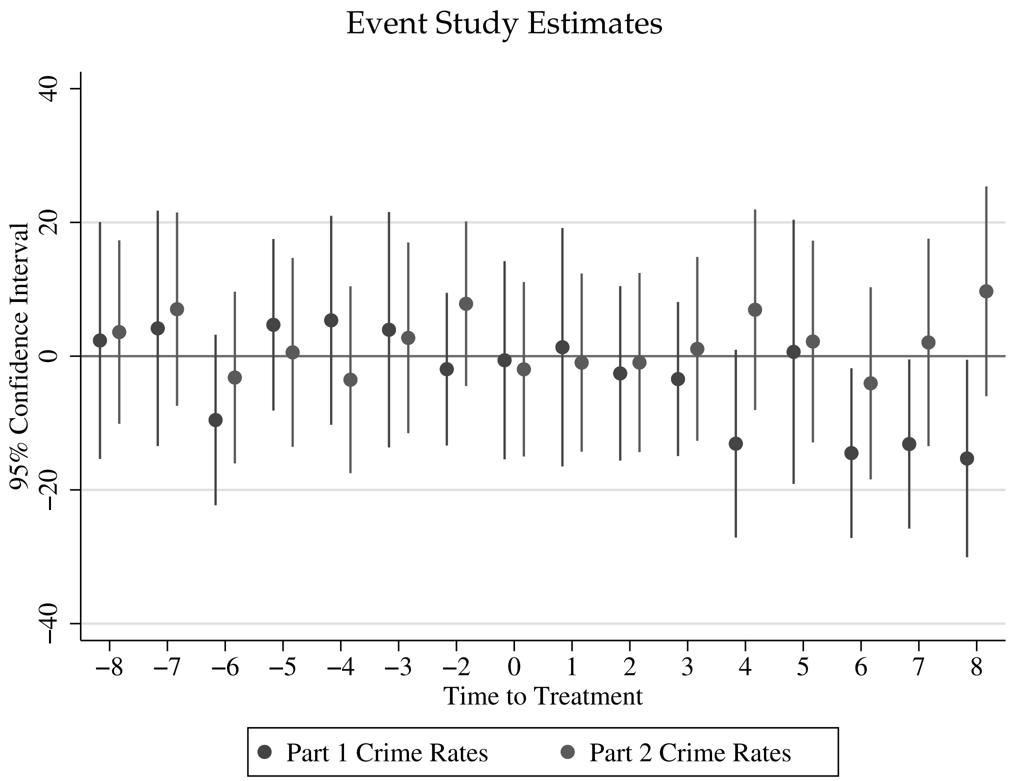


Figure 5: 95% CIs of Regression Coefficients for Part I/II Crime Quarterly Rates

Note: The vertical axis refers to quarterly crimes per square mile. To avoid perfect multicollinearity, the -1 time lag is used as the dropped reference.

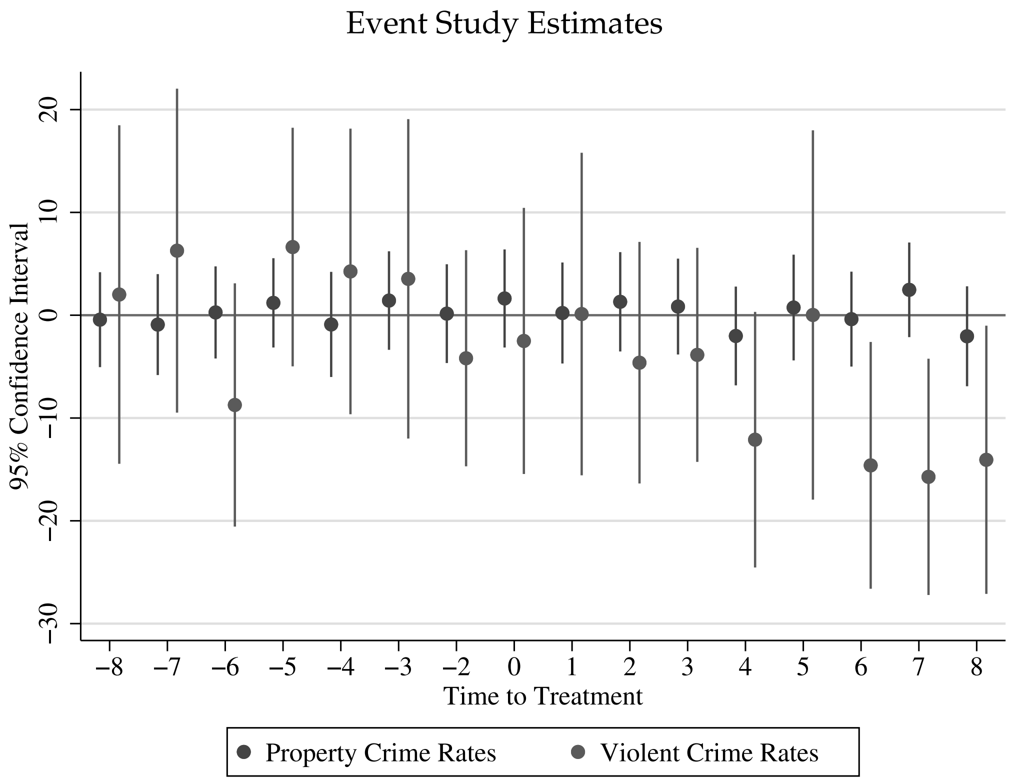


Figure 6: 95% CIs of Regression Coefficients for Property/Violent Crime Quarterly Rates

Note: The vertical axis refers to quarterly crimes per square mile. To avoid perfect multicollinearity, the -1 time lag is used as the dropped reference.

References

Baum-Snow, N., & Marion, J. (2009). The effects of low income housing tax credit developments on neighborhoods. *Journal of Public Economics*, *93*. doi: 10.1016/j.jpubeco.2009.01.001

Bjerk, D. (2010). Thieves, thugs, and neighborhood poverty. *Journal of Urban Economics*, *68*. doi: 10.1016/j.jue.2010.06.002

Butts, K. (2021, 10). Difference-in-differences with geocoded microdata.

Cummings, J. L., & DiPasquale, D. (1999). The low-income housing tax credit: An analysis of the first ten years. *Housing Policy Debate*, *10*. doi: 10.1080/10511482.1999.9521332

Diamond, R., & McQuade, T. (2019). Who wants affordable housing in their backyard? an equilibrium analysis of low-income property development. *Journal of Political Economy*, *127*. doi: 10.1086/701354

Eriksen, M. D., & Rosenthal, S. S. (2010). Crowd out effects of place-based subsidized rental housing: New evidence from the lihtc program. *Journal of Public Economics*, *94*. doi: 10.1016/j.jpubeco.2010.07.002

Fallon, K. F., & Price, C. R. (2020). Evaluating exposure to crime among lihtc building types and characteristics in ohio. *Housing Policy Debate*. doi: 10.1080/10511482.2020.1839938

Freedman, M., & Owens, E. G. (2011). Low-income housing development and crime. *Journal of Urban Economics*, *70*. doi: 10.1016/j.jue.2011.04.001

Glaeser, E. L., Sacerdote, B., & Scheinkman, J. A. (1996). Crime and social interactions. *Quarterly Journal of Economics*, *111*. doi: 10.2307/2946686

Green, R., Malpezzi, S., & Seah, K.-Y. (2002). Low income housing tax credit housing developments and property values. *The Center for Urban Land Economics Research, The University of Wisconsin*.

Harcourt, B. E., & Ludwig, J. (2006). Broken windows: New evidence from new york city and a five-city social experiment. In (Vol. 73).

IBO, N. Y. C. (2019, 2). *Comparing affordability levels of the mayor's housing new york plan with neighborhood incomes.* Retrieved from https://ibo.nyc.ny.us/iboreports/affordable-for-whom-comparing-affordability-levels-of-the-mayors-housing-new-york-plan-with-neighborhood-incomes-february-2019.pdf

Kling, J. R., Ludwig, J., & Katz, L. F. (2005). Neighborhood effects on crime for female and male youth: Evidence from a randomized housing voucher experiment. *Quarterly Journal of Economics*, *120*. doi: 10.1162/0033553053327470

Krivo, L. J., & Peterson, R. D. (1996). Extremely disadvantaged neighborhoods and urban crime. *Social Forces*, *75*. doi: 10.2307/2580416

Lee, C. M., Culhane, D. P., & Wachter, S. M. (1999). The differential impacts of federally assisted housing programs on nearby property values: A philadelphia case study. *Housing Policy Debate*, *10*. doi: 10.1080/10511482.1999.9521328

Ludwig, J., & Kling, J. R. (2007). Is crime contagious? *Journal of Law and Economics*, *50*. doi: 10.1086/519807

McNulty, T. L., & Holloway, S. R. (2000). Race, crime, and public housing in atlanta: Testing a conditional effect hypothesis. *Social Forces*, *79*. doi: 10.1093/sf/79.2.707

Oakley, D. (2008). Locational patterns of low-income housing tax credit developments: A sociospatial analysis of four metropolitan areas. *Urban Affairs Review*, *43*. doi: 10.1177/1078087407309432

Patterson, E. B. (1991). Poverty, income inequality, and community crime rates. *Criminology*, *29*. doi: 10.1111/j.1745-9125.1991.tb01087.x

Peterson, R. D., Krivo, L. J., & Harris, M. A. (2000). Disadvantage and neighborhood violent crime: Do local institutions matter? *Journal of Research in Crime and Delinquency*, *37*. doi: 10.1177/0022427800037001002

Roncek, D. W., Bell, R., & Francik, J. M. A. (1981). Housing projects and crime: Testing a proximity hypothesis. *Social Problems*, *29*. doi: 10.1525/sp.1981.29.2.03a00060

Schwartz, A. E., Ellen, I. G., Voicu, I., & Schill, M. H. (2006). The external effects of place-based subsidized housing. *Regional Science and Urban Economics*, *36*. doi: 10.1016/j.regsciurbeco.2006.04.002

Sinai, T., & Waldfogel, J. (2005). Do low-income housing subsidies increase the occupied housing stock? *Journal of Public Economics*, *89*. doi: 10.1016/j.jpubeco.2004.06.015

Woo, A., & Joh, K. (2015). Beyond anecdotal evidence: Do subsidized housing developments increase neighborhood crime? *Applied Geography*, *64*. doi: 10.1016/j.apgeog.2015.09.004

Appendix A

A.1 Inner Ring: 0.1 Miles, Outer Ring: 0.2 Miles

Table 5: Summary Statistics for Inner Ring Quarterly Crime Rates

	mean	sd	min	max	count
Total Crime Rate	1904.04	1710.91	63.66	24573.52	13712
Part 1 Crime Rate	369.27	438.11	0.00	8976.34	13712
Part 2 Crime Rate	582.75	534.87	0.00	3787.89	13712
Property Crime Rate	89.10	124.40	0.00	2037.18	13712
Violent Crime Rate	270.87	386.16	0.00	8180.56	13712
Observations	13712				

Notes: Rates are defined as number of crimes in the quarter per square mile.
Recall that the inner ring is a radius of 0.1 miles from an LIHTC opening, and the outer ring is a radius of 0.2 miles (with the outer region being exclusive of the inner region).
The summary contains 8 quarters of pre-opening data and 8 quarters of post-opening data.

Table 6: Summary Statistics for Outer Ring Quarterly Crime Rates

	mean	sd	min	max	count
Total Crime Rate	1766.25	1464.87	21.22	19798.88	13862
Part 1 Crime Rate	353.27	392.08	0.00	9029.39	13862
Part 2 Crime Rate	529.85	443.50	0.00	3978.87	13862
Property Crime Rate	88.98	101.53	0.00	1018.59	13862
Violent Crime Rate	261.05	352.62	0.00	8849.01	13862
Observations	13862				

Notes: Rates are defined as number of crimes in the quarter per square mile.
Recall that the inner ring is a radius of 0.1 miles from an LIHTC opening, and the outer ring is a radius of 0.2 miles (with the outer region being exclusive of the inner region).
The summary contains 8 quarters of pre-opening data and 8 quarters of post-opening data.

Table 7: Summary Statistics for Combined Ring Quarterly Crime Rates

	mean	sd	min	max	count
Total Crime Rate	1834.77	1593.44	21.22	24573.52	27574
Part 1 Crime Rate	361.23	415.68	0.00	9029.39	27574
Part 2 Crime Rate	556.16	491.77	0.00	3978.87	27574
Property Crime Rate	89.04	113.48	0.00	2037.18	27574
Violent Crime Rate	265.94	369.70	0.00	8849.01	27574
Observations	27574				

Notes: Rates are defined as number of crimes in the quarter per square mile.
Recall that the inner ring is a radius of 0.1 miles from an LIHTC opening, and the outer ring is a radius of 0.2 miles (with the outer region being exclusive of the inner region).
The summary contains 8 quarters of pre-opening data and 8 quarters of post-opening data.

Table 8: Regression Results for Quarterly Crime Rates

	(1)		(2)		(3)		(4)		(5)	
	Total Crime Rate		Part 1 Crime Rate		Part 2 Crime Rate		Property Crime Rate		Violent Crime Rate	
Inner Ring	102.54**	(47.29)	18.58	(14.53)	32.68**	(13.46)	-1.20	(4.59)	13.73	(13.31)
Inner Ring × Quarter: -8	73.77**	(29.06)	7.18	(8.57)	29.70***	(10.33)	1.98	(3.29)	5.84	(7.37)
Inner Ring × Quarter: -7	26.19	(31.17)	-10.59	(9.44)	23.68**	(10.89)	1.27	(3.35)	-12.36	(8.16)
Inner Ring × Quarter: -6	76.50***	(26.41)	0.62	(7.92)	37.63***	(10.23)	3.05	(3.26)	-1.40	(7.12)
Inner Ring × Quarter: -5	22.21	(25.53)	-2.01	(7.25)	13.12	(10.16)	0.37	(3.26)	-1.34	(6.50)
Inner Ring × Quarter: -4	55.63**	(27.58)	5.39	(8.43)	22.43**	(9.93)	2.34	(3.34)	4.04	(7.40)
Inner Ring × Quarter: -3	2.41	(29.50)	-6.94	(9.24)	8.15	(9.78)	-3.67	(3.32)	-4.13	(7.83)
Inner Ring × Quarter: -2	64.18**	(25.47)	15.28**	(7.06)	16.81*	(9.43)	2.06	(3.44)	14.42**	(6.22)
Inner Ring × Quarter: 0	18.97	(26.67)	-6.06	(8.82)	15.54*	(9.29)	-2.71	(3.20)	-2.91	(7.72)
Inner Ring × Quarter: 1	0.78	(36.67)	-12.58	(13.92)	12.97	(9.39)	-0.94	(3.38)	-11.60	(12.83)
Inner Ring × Quarter: 2	10.84	(36.52)	-13.66	(13.77)	19.08**	(9.65)	1.00	(3.42)	-13.44	(13.04)
Inner Ring × Quarter: 3	49.57	(35.85)	-7.10	(13.15)	31.88***	(10.22)	3.55	(3.53)	-13.65	(12.51)
Inner Ring × Quarter: 4	0.02	(36.14)	-11.53	(12.97)	11.54	(10.38)	1.77	(3.45)	-13.58	(11.98)
Inner Ring × Quarter: 5	-30.12	(37.76)	-18.83	(13.97)	3.78	(10.08)	0.85	(3.40)	-21.04	(13.09)
Inner Ring × Quarter: 6	11.05	(36.07)	-2.51	(13.50)	8.04	(9.36)	2.09	(3.44)	-5.71	(12.62)
Inner Ring × Quarter: 7	-2.14	(35.98)	-15.90	(13.82)	14.83	(9.62)	-1.27	(3.29)	-15.76	(12.99)
Inner Ring × Quarter: 8	-6.11	(35.17)	-12.88	(13.55)	9.82	(9.32)	-1.15	(3.41)	-10.84	(12.79)
Constant	1816.91***	(24.05)	364.67***	(6.76)	543.78***	(7.33)	92.67***	(2.35)	269.10***	(6.02)
Time Fixed Effects	Yes		Yes		Yes		Yes		Yes	
R^2	0.79		0.73		0.79		0.57		0.72	
Observations	27574		27574		27574		27574		27574	

Notes: Clustered standard errors in parenthesis. * p< 0.10, ** p< 0.05, *** p < 0.01.

The regressions establish the quarter preceding an LIHTC development opening (i.e. Quarter: -1) to be the baseline and thus omit the corresponding terms.

Recall that the inner ring is a radius of 0.1 miles from an LIHTC opening, and the outer ring is a radius of 0.2 miles (with the outer region being exclusive of the inner region).

The regression runs on 8 quarters of pre-opening data and 8 quarters of post-opening data.

One of the time periods must be dropped to avoid perfect multicollinearity. In this event study, the -1 time lag is used as the dropped reference.

A.2 Inner Ring: 0.1 Miles, Outer Ring: 0.3 Miles

Table 9: Summary Statistics for Inner Ring Quarterly Crime Rates

	mean	sd	min	max	count
Total Crime Rate	1671.75	1435.78	63.66	11777.47	8166
Part 1 Crime Rate	329.02	325.92	0.00	3819.72	8166
Part 2 Crime Rate	506.86	483.28	0.00	4647.32	8166
Property Crime Rate	72.28	102.96	0.00	1687.04	8166
Violent Crime Rate	245.83	277.55	0.00	3787.89	8166
Observations	8166				

Notes: Rates are defined as number of crimes in the quarter per square mile.
Recall that the inner ring is a radius of 0.1 miles from an LIHTC opening, and the outer ring is a radius of 0.3 miles (with the outer region being exclusive of the inner region).
The summary contains 8 quarters of pre-opening data and 8 quarters of post-opening data.

Table 10: Summary Statistics for Outer Ring Quarterly Crime Rates

	mean	sd	min	max	count
Total Crime Rate	1571.81	1246.50	15.92	12445.92	8313
Part 1 Crime Rate	318.42	321.22	0.00	5188.45	8313
Part 2 Crime Rate	467.49	372.91	3.98	2606.16	8313
Property Crime Rate	74.34	69.42	0.00	473.49	8313
Violent Crime Rate	240.82	286.75	0.00	5049.19	8313
Observations	8313				

Notes: Rates are defined as number of crimes in the quarter per square mile.
Recall that the inner ring is a radius of 0.1 miles from an LIHTC opening, and the outer ring is a radius of 0.3 miles (with the outer region being exclusive of the inner region).
The summary contains 8 quarters of pre-opening data and 8 quarters of post-opening data.

Table 11: Summary Statistics for Combined Ring Quarterly Crime Rates

	mean	sd	min	max	count
Total Crime Rate	1621.34	1344.52	15.92	12445.92	16479
Part 1 Crime Rate	323.67	323.59	0.00	5188.45	16479
Part 2 Crime Rate	487.00	431.59	0.00	4647.32	16479
Property Crime Rate	73.32	87.67	0.00	1687.04	16479
Violent Crime Rate	243.30	282.23	0.00	5049.19	16479
Observations	16479				

Notes: Rates are defined as number of crimes in the quarter per square mile.
Recall that the inner ring is a radius of 0.1 miles from an LIHTC opening, and the outer ring is a radius of 0.3 miles (with the outer region being exclusive of the inner region).
The summary contains 8 quarters of pre-opening data and 8 quarters of post-opening data.

Table 12: Regression Results for Quarterly Crime Rates

	(1)		(2)		(3)		(4)		(5)	
	Total Crime Rate		Part 1 Crime Rate		Part 2 Crime Rate		Property Crime Rate		Violent Crime Rate	
Inner Ring	95.22**	(46.80)	9.71	(12.08)	37.90**	(16.07)	-2.62	(4.40)	5.07	(10.78)
Inner Ring × Quarter: -8	-2.51	(31.03)	-5.45	(9.30)	4.20	(11.59)	-1.64	(3.75)	-4.46	(8.01)
Inner Ring × Quarter: -7	3.10	(33.08)	-8.45	(9.44)	10.00	(12.38)	2.08	(3.45)	-10.27	(8.22)
Inner Ring × Quarter: -6	10.89	(29.92)	4.64	(8.42)	0.80	(11.99)	5.16	(3.60)	-0.65	(7.28)
Inner Ring × Quarter: -5	-42.66	(28.69)	-9.85	(7.93)	-11.48	(11.46)	1.00	(3.51)	-10.83	(7.08)
Inner Ring × Quarter: -4	-21.72	(26.99)	0.77	(8.64)	-11.63	(10.36)	0.95	(3.69)	-1.42	(7.33)
Inner Ring × Quarter: -3	-33.89	(28.87)	-7.27	(8.57)	-9.67	(10.98)	-1.58	(3.37)	-4.40	(7.25)
Inner Ring × Quarter: -2	28.27	(25.47)	11.14	(7.26)	3.00	(10.31)	2.59	(3.67)	8.94	(6.39)
Inner Ring × Quarter: 0	6.73	(26.86)	3.00	(8.09)	0.36	(10.34)	2.60	(3.40)	-1.22	(6.86)
Inner Ring × Quarter: 1	-32.67	(29.37)	-10.08	(8.59)	-6.26	(10.74)	-2.18	(3.65)	-8.37	(7.50)
Inner Ring × Quarter: 2	-23.29	(28.05)	-4.85	(8.68)	-6.79	(10.07)	-2.92	(3.78)	-0.11	(7.55)
Inner Ring × Quarter: 3	19.16	(27.75)	-3.52	(8.02)	13.11	(10.76)	-1.65	(3.53)	-3.38	(7.12)
Inner Ring × Quarter: 4	-50.90*	(27.20)	-4.99	(8.50)	-20.46*	(10.54)	-2.32	(3.63)	-4.91	(7.35)
Inner Ring × Quarter: 5	-27.09	(32.00)	-5.80	(9.66)	-7.74	(11.43)	-2.95	(3.48)	-3.76	(8.47)
Inner Ring × Quarter: 6	-3.53	(29.58)	-2.08	(9.07)	0.32	(10.84)	-0.71	(3.76)	-2.07	(8.17)
Inner Ring × Quarter: 7	-65.78**	(30.06)	-6.90	(9.68)	-25.99**	(11.06)	-0.73	(3.96)	-8.38	(8.58)
Inner Ring × Quarter: 8	-10.16	(30.72)	-3.07	(9.22)	-2.01	(11.17)	-3.56	(3.69)	1.73	(8.37)
Constant	1595.56***	(23.58)	324.76***	(6.06)	473.02***	(7.61)	75.35***	(2.15)	245.43***	(5.30)
Time Fixed Effects	Yes		Yes		Yes		Yes		Yes	
R^2	0.80		0.75		0.79		0.58		0.74	
Observations	16479		16479		16479		16479		16479	

Notes: Clustered standard errors in parenthesis. * $p< 0.10$, ** $p< 0.05$, *** $p < 0.01$.

The regressions establish the quarter preceding an LIHTC development opening (i.e. Quarter: -1) to be the baseline and thus omit the corresponding terms.

Recall that the inner ring is a radius of 0.1 miles from an LIHTC opening, and the outer ring is a radius of 0.3 miles (with the outer region being exclusive of the inner region).

The regression runs on 8 quarters of pre-opening data and 8 quarters of post-opening data.

One of the time periods must be dropped to avoid perfect multicollinearity. In this event study, the -1 time lag is used as the dropped reference.

A.3 Inner Ring: 0.2 Miles, Outer Ring: 0.275 Miles

Table 13: Summary Statistics for Inner Ring Quarterly Crime Rates

	mean	sd	min	max	count
Total Crime Rate	1706.53	1449.78	15.92	20817.47	9209
Part 1 Crime Rate	345.39	410.09	0.00	8856.97	9209
Part 2 Crime Rate	507.88	416.81	0.00	3151.27	9209
Property Crime Rate	84.42	94.69	0.00	748.03	9209
Violent Crime Rate	258.00	366.48	0.00	8554.58	9209
Observations	9209				

Notes: Rates are defined as number of crimes in the quarter per square mile.
Recall that the inner ring is a radius of 0.2 miles from an LIHTC opening, and the outer ring is a radius of 0.275 miles (with the outer region being exclusive of the inner region).
The summary contains 8 quarters of pre-opening data and 8 quarters of post-opening data.

Table 14: Summary Statistics for Outer Ring Quarterly Crime Rates

	mean	sd	min	max	count
Total Crime Rate	1612.71	1337.17	17.87	18924.36	9206
Part 1 Crime Rate	328.07	390.46	0.00	8381.04	9206
Part 2 Crime Rate	478.29	382.71	0.00	2877.07	9206
Property Crime Rate	74.10	79.79	0.00	893.50	9206
Violent Crime Rate	249.02	360.93	0.00	8247.02	9206
Observations	9206				

Notes: Rates are defined as number of crimes in the quarter per square mile.
Recall that the inner ring is a radius of 0.2 miles from an LIHTC opening, and the outer ring is a radius of 0.275 miles (with the outer region being exclusive of the inner region).
The summary contains 8 quarters of pre-opening data and 8 quarters of post-opening data.

Table 15: Summary Statistics for Combined Ring Quarterly Crime Rates

	mean	sd	min	max	count
Total Crime Rate	1659.63	1395.37	15.92	20817.47	18415
Part 1 Crime Rate	336.73	400.48	0.00	8856.97	18415
Part 2 Crime Rate	493.09	400.39	0.00	3151.27	18415
Property Crime Rate	79.26	87.71	0.00	893.50	18415
Violent Crime Rate	253.51	363.73	0.00	8554.58	18415
Observations	18415				

Notes: Rates are defined as number of crimes in the quarter per square mile.
Recall that the inner ring is a radius of 0.2 miles from an LIHTC opening, and the outer ring is a radius of 0.275 miles (with the outer region being exclusive of the inner region).
The summary contains 8 quarters of pre-opening data and 8 quarters of post-opening data.

Table 16: Regression Results for Quarterly Crime Rates

	(1)		(2)		(3)		(4)		(5)	
	Total Crime Rate		Part 1 Crime Rate		Part 2 Crime Rate		Property Crime Rate		Violent Crime Rate	
Inner Ring	90.68*	(48.18)	18.93	(16.27)	26.41**	(12.20)	10.13**	(4.29)	10.24	(14.92)
Inner Ring × Quarter: -8	18.98	(23.80)	1.29	(8.71)	8.20	(7.70)	-0.16	(2.59)	2.79	(7.99)
Inner Ring × Quarter: -7	17.55	(22.86)	7.06	(8.14)	1.72	(7.54)	0.97	(2.59)	6.35	(7.37)
Inner Ring × Quarter: -6	3.15	(18.88)	-5.21	(6.36)	6.79	(7.02)	-1.09	(2.44)	-3.07	(5.99)
Inner Ring × Quarter: -5	20.03	(20.12)	5.58	(6.79)	4.44	(6.96)	2.94	(2.48)	3.66	(6.31)
Inner Ring × Quarter: -4	10.23	(22.06)	6.29	(7.52)	-1.17	(7.17)	-0.28	(2.49)	7.51	(6.59)
Inner Ring × Quarter: -3	26.14	(24.10)	6.73	(8.22)	6.34	(7.52)	0.68	(2.47)	5.78	(7.51)
Inner Ring × Quarter: -2	9.91	(17.60)	-2.94	(6.03)	7.89	(6.36)	-3.50	(2.41)	-0.79	(5.56)
Inner Ring × Quarter: 0	-1.32	(21.06)	1.18	(6.95)	-1.84	(6.66)	3.04	(2.48)	-1.39	(6.11)
Inner Ring × Quarter: 1	8.22	(23.13)	4.41	(8.34)	-0.30	(6.85)	1.62	(2.52)	3.83	(7.39)
Inner Ring × Quarter: 2	-0.84	(19.90)	-3.15	(6.21)	2.73	(7.38)	-1.29	(2.43)	-4.40	(5.70)
Inner Ring × Quarter: 3	3.61	(18.53)	-3.96	(5.54)	5.77	(6.63)	1.97	(2.46)	-3.72	(4.87)
Inner Ring × Quarter: 4	-11.20	(21.53)	-9.66	(6.92)	4.06	(7.18)	-1.79	(2.45)	-5.09	(5.91)
Inner Ring × Quarter: 5	2.13	(26.07)	2.78	(9.19)	-1.72	(7.29)	1.68	(2.60)	3.07	(8.22)
Inner Ring × Quarter: 6	-1.88	(21.29)	-6.33	(6.76)	5.40	(7.28)	1.08	(2.41)	-6.94	(6.03)
Inner Ring × Quarter: 7	-10.35	(21.63)	-11.02	(6.88)	5.85	(7.55)	-1.87	(2.49)	-9.63	(6.40)
Inner Ring × Quarter: 8	-31.91	(22.29)	-18.59**	(7.78)	2.64	(7.35)	-0.24	(2.49)	-17.96***	(6.93)
Constant	1640.38***	(25.49)	333.73***	(8.37)	486.46***	(6.78)	76.04***	(2.42)	253.32***	(7.56)
Time Fixed Effects	Yes		Yes		Yes		Yes		Yes	
R^2	0.82		0.76		0.83		0.60		0.76	
Observations	18415		18415		18415		18415		18415	

Notes: Clustered standard errors in parenthesis. * $p< 0.10$, ** $p< 0.05$, *** $p < 0.01$.

The regressions establish the quarter preceding an LIHTC development opening (i.e. Quarter: -1) to be the baseline and thus omit the corresponding terms.

Recall that the inner ring is a radius of 0.2 miles from an LIHTC opening, and the outer ring is a radius of 0.275 miles (with the outer region being exclusive of the inner region).

The regression runs on 8 quarters of pre-opening data and 8 quarters of post-opening data.

One of the time periods must be dropped to avoid perfect multicollinearity. In this event study, the -1 time lag is used as the dropped reference.

A.4 Inner Ring: 0.2 Miles, Outer Ring: 0.4 Miles

Table 17: Summary Statistics for Inner Ring Quarterly Crime Rates

	mean	sd	min	max	count
Total Crime Rate	1430.85	1224.02	15.92	8291.97	5879
Part 1 Crime Rate	281.55	262.64	0.00	2076.97	5879
Part 2 Crime Rate	433.87	387.29	0.00	2761.34	5879
Property Crime Rate	70.30	85.26	0.00	1034.51	5879
Violent Crime Rate	208.55	209.31	0.00	1726.83	5879
Observations	5879				

Notes: Rates are defined as number of crimes in the quarter per square mile.
Recall that the inner ring is a radius of 0.2 miles from an LIHTC opening, and the outer ring is a radius of 0.4 miles (with the outer region being exclusive of the inner region).
The summary contains 8 quarters of pre-opening data and 8 quarters of post-opening data.

Table 18: Summary Statistics for Outer Ring Quarterly Crime Rates

	mean	sd	min	max	count
Total Crime Rate	1315.56	1099.29	42.44	9124.88	5882
Part 1 Crime Rate	265.82	254.07	0.00	2904.58	5882
Part 2 Crime Rate	391.96	331.80	5.31	2058.40	5882
Property Crime Rate	62.97	63.66	0.00	429.72	5882
Violent Crime Rate	199.52	213.28	0.00	2774.60	5882
Observations	5882				

Notes: Rates are defined as number of crimes in the quarter per square mile.
Recall that the inner ring is a radius of 0.2 miles from an LIHTC opening, and the outer ring is a radius of 0.4 miles (with the outer region being exclusive of the inner region).
The summary contains 8 quarters of pre-opening data and 8 quarters of post-opening data.

Table 19: Summary Statistics for Combined Ring Quarterly Crime Rates

	mean	sd	min	max	count
Total Crime Rate	1373.19	1164.69	15.92	9124.88	11761
Part 1 Crime Rate	273.69	258.50	0.00	2904.58	11761
Part 2 Crime Rate	412.91	361.20	0.00	2761.34	11761
Property Crime Rate	66.64	75.32	0.00	1034.51	11761
Violent Crime Rate	204.04	211.35	0.00	2774.60	11761
Observations	11761				

Notes: Rates are defined as number of crimes in the quarter per square mile.
Recall that the inner ring is a radius of 0.2 miles from an LIHTC opening, and the outer ring is a radius of 0.4 miles (with the outer region being exclusive of the inner region).
The summary contains 8 quarters of pre-opening data and 8 quarters of post-opening data.

Table 20: Regression Results for Quarterly Crime Rates

	(1)		(2)		(3)		(4)		(5)	
	Total Crime Rate		Part 1 Crime Rate		Part 2 Crime Rate		Property Crime Rate		Violent Crime Rate	
Inner Ring	125.71***	(38.55)	20.34*	(10.41)	42.52***	(11.38)	6.46*	(3.58)	14.42	(9.02)
Inner Ring × Quarter: -8	-21.25	(23.95)	-7.54	(8.09)	-3.09	(6.46)	-0.18	(2.29)	-7.60	(7.26)
Inner Ring × Quarter: -7	20.85	(27.67)	1.28	(10.03)	9.14	(6.87)	-0.34	(2.57)	1.29	(8.83)
Inner Ring × Quarter: -6	-27.86	(18.94)	-8.73	(6.26)	-5.20	(6.35)	0.71	(2.49)	-9.74*	(5.81)
Inner Ring × Quarter: -5	-12.49	(19.09)	-4.11	(5.63)	-2.13	(6.77)	0.91	(2.06)	-3.67	(5.19)
Inner Ring × Quarter: -4	-6.75	(26.01)	-2.25	(8.27)	-1.13	(7.53)	1.94	(2.49)	-4.51	(7.24)
Inner Ring × Quarter: -3	-23.90	(30.15)	-0.28	(10.40)	-11.67	(7.57)	-1.50	(2.64)	1.27	(8.87)
Inner Ring × Quarter: -2	-24.39	(17.81)	-10.69*	(5.63)	-1.51	(5.81)	1.12	(2.25)	-11.31**	(5.00)
Inner Ring × Quarter: 0	-7.59	(25.44)	0.50	(8.28)	-4.29	(7.14)	4.06*	(2.18)	-3.32	(7.30)
Inner Ring × Quarter: 1	9.69	(32.29)	7.77	(11.23)	-2.93	(7.05)	0.90	(2.60)	6.61	(9.73)
Inner Ring × Quarter: 2	-27.72	(19.69)	-12.74**	(6.26)	-1.12	(6.52)	0.15	(2.33)	-11.97**	(5.71)
Inner Ring × Quarter: 3	-4.28	(17.26)	-5.03	(5.15)	2.89	(6.77)	2.64	(2.30)	-7.52	(4.61)
Inner Ring × Quarter: 4	-24.87	(24.55)	-7.60	(7.67)	-4.84	(7.94)	0.80	(2.30)	-9.09	(6.66)
Inner Ring × Quarter: 5	11.32	(31.88)	3.14	(10.66)	2.52	(8.28)	0.85	(2.80)	0.53	(9.38)
Inner Ring × Quarter: 6	-20.02	(21.75)	-10.40	(6.56)	0.39	(7.85)	0.13	(2.33)	-10.23*	(6.00)
Inner Ring × Quarter: 7	-7.91	(18.94)	-10.10*	(5.75)	6.14	(7.11)	0.79	(2.55)	-9.94*	(5.37)
Inner Ring × Quarter: 8	-19.78	(25.28)	-13.65*	(7.91)	3.76	(7.85)	1.27	(2.44)	-13.89**	(6.91)
Constant	1323.57***	(19.49)	269.05***	(5.18)	392.74***	(6.01)	63.24***	(1.93)	202.81***	(4.39)
Time Fixed Effects	Yes		Yes		Yes		Yes		Yes	
R^2	0.88		0.82		0.88		0.71		0.80	
Observations	11761		11761		11761		11761		11761	

Notes: Clustered standard errors in parenthesis. * p< 0.10, ** p< 0.05, *** p $<$ 0.01.

The regressions establish the quarter preceding an LIHTC development opening (i.e. Quarter: -1) to be the baseline and thus omit the corresponding terms.

Recall that the inner ring is a radius of 0.2 miles from an LIHTC opening, and the outer ring is a radius of 0.4 miles (with the outer region being exclusive of the inner region).

The regression runs on 8 quarters of pre-opening data and 8 quarters of post-opening data.

One of the time periods must be dropped to avoid perfect multicollinearity. In this event study, the -1 time lag is used as the dropped reference.

Analyzing the “Penalty” on Asian American Applicants in College Admissions

Nina Wang

Massachusetts Institute of Technology

Abstract

The *Students for Fair Admissions vs. Harvard* case highlights the possibility that affirmative action for under-represented minorities reduces the likelihood of selective college enrollment for Asian Americans, who are not seen as disadvantaged, resulting in significantly higher test score averages among accepted Asian applicants when compared to the rest of the accepted applicants. This paper uses data from the National Longitudinal Survey of Youth 1997 (NLSY97) to estimate the relationship between Asian ethnicity and college selectivity, as proxied by observing characteristics of the college that individuals eventually enroll in. The analysis shows that Asian applicants are less likely to go to a private institution, but other than that, the results suggest that there is no strong evidence of an Asian American penalty in college admissions.

1. Introduction

Asian Americans are in a unique position from a college admissions standpoint. While Asian Americans are considered a minority in the United States, accounting for around 3% of American 18-22 year olds in 1990 and 6% in 2020, they have been *over*-represented at top institutions since the late 1900s (Harvard, as an example, reported that 27.9% of their most recent freshmen class was Asian American). This is partly because Asian Americans notoriously outperform their non-Asian counterparts in the classroom and on standardized tests, resulting in a large number of highly qualified Asian American applicants (VerBruggen, 2022).

However, the college admissions process is not as simple as admitting the top students based on measures of academic achievement. One of the goals of the college admissions process, among other considerations, is to maintain a diverse student body and prevent any extreme skew in the campus's racial balance. This practice is more generally known as affirmative action. Parties such as Students for Fair Admissions allege that these conflicting objectives put the individual Asian applicant at a disadvantage. If two students with comparable qualifications are competing for a spot in the admitted class, but one student is Asian and one is not, then many would argue that the non-Asian student would be favored. This happens 1) to prevent further overrepresentation of Asian American students and "balance out" the demographics, and 2) because there are most likely a higher number of Asian American applicants who are just as, if not more, qualified for the spot. These factors create an implicit "penalty" that is put on the Asian American's college application.

This paper uses data from the National Longitudinal Survey of Youth 1997 (NLSY97) to examine the validity of such a penalty at universities nationwide by comparing measures of college selectivity between Asian Americans and non-Asian American cohorts. If all Asian applicants receive such a penalty in the process of college admissions, then Asian applicants would go to less selective schools than non-Asian applicants with the same qualifications on average. I include a vector of covariates that help control for ability, income, geographic region, and parents' education, in order to reduce omitted variable bias. The analysis shows that Asian Americans are less likely to attend a private school than their non-Asian peers. However, other measures of selectivity do not differ significantly between Asian and non-Asian cohorts. These results suggest that there is no observable Asian American penalty when we generalize to schools of varying selectivity.

This study adds to a wealth of literature that focuses on racial preferences in college admissions dating back to the mid 20th century. Many studies that address this topic focus on whites vs. under-represented minorities (usually referring to black and Hispanic students). For example, Bowen and Bok (1998) find that the probabilities of gaining admittance into five selective institutions vary distinctly between black and white cohorts of students conditional on SAT score. Similarly, Long (2002a) finds that for the average student, being black instead of white would raise the probability of acceptance to the student's best college by 8.4 percentage points, while being Hispanic would raise the probability of acceptance by 5.8 percentage points. He also notes that the coefficient on Asian American is negative but not significant, although this is not his main finding. Arcidiacono, Kinsler and Ransom (2022) find that at Harvard, the admit rates for African American applicants are 4 times larger than if they had been treated as white, and at UNC Chapel Hill, out of state applicants are admitted at a rate more than 10 times that of if they had been treated as white.

There are relatively few studies that have focused on the disparity between Asian and non-Asian applicant cohorts. This is partly because the proportion of applicants that identify as

Asian is still small (not to be confused with the proportion of admitted students that identify as Asian which is much larger), and partly because Asian-Americans have historically downplayed the negative impacts of racial preferences, and thus they have only recently become a topic of discourse (Izumi, 1997).

However, studies focusing on the impact of a removal of race-conscious admission practices shed some light on how racial preferences affect Asian applicants. Espenshade and Chung (2005) find that Asian Americans would be the largest beneficiaries of a ban on affirmative action, with average acceptance rate rising from 18 percent to over 23 percent. Data from Lynch (1997) revealed that in California's Master Plan, where the top 12.5 percent of the high school graduating class gains admission into the UC system, Asian American students qualify at a rate over double that of non-Asian students. In more recent literature connected to the *Students for Fair Admission vs. Harvard* case, Arcidiacono (2020) finds that Asian Americans face a substantial penalty in the admissions process at Harvard due to having higher average observable measures of ability than their non-Asian counterparts. The data further yields a model that suggests that Asian Americans would be admitted at a rate of 19 percent higher without the penalty.

While my study is partly inspired by Arcidiacono's paper, it is not able to replicate the results. Part of the reason for this is because measurable Asian American outperformance (SAT, GPA, etc.) is naturally more concentrated at the tail ends of the distribution. While this *will* have a pronounced effect at an extremely selective school such as Harvard, it is less likely to have a significant impact on schools that are less selective and will admit the average applicant.

The rest of this paper is structured as follows. Section 2 describes the data used in this study, as well as its strengths and weaknesses, Section 3 describes the empirical approach, Section 4 looks at descriptive statistics, Section 5 covers the results of the empirical analysis, and Section 6 concludes.

2. Data and Variables

The dataset used in this paper is a subset of the data collected as part of the National Longitudinal Survey of Youth 1997 surveys. This family of surveys was conducted by the Bureau of Labor Statistics and consists of data on 8,984 respondents, 51% male and 49% female, all of whom were born in the years 1980-1984. After the initial survey in 1997, researchers conducted annual interviews with the same cohort up until 2020, with a separate COVID follow up. Over 77% of initial respondents remained in the program in round 19 (fielded in 2019-2020). The respondents of the survey were all between the ages of 12-18 when the survey was first conducted in 1997, meaning all respondents were between the ages of 34-40 at the time of the last follow up. The surveys capture complete college information on all respondents who decided to go to college. For each respondent, the data contains details on general demographics, pre-college academic history, standardized test scores, higher education history, income (if applicable), and a host of other variables.

This study obtains information on the first college each respondent attended using a combination of the initial survey and the annual follow ups, as only a select few respondents were in college at the time of the initial survey. Due to restrictions with the dataset, I was not able to obtain the names and locations of the colleges each respondent attended. Rather, I will be relying on more general measures associated with college selectivity, specifically, private college vs. public college, 2-year program vs. 4-year program, and college tuition. The two subsets within each of the first two categories are mutually exclusive, while tuition is a continuous variable. These metrics provide an alternative way to gauge college selectivity without needing explicit information on which schools each respondent attended. Private schools and 4-year programs are generally more selective than public schools and 2-year programs, respectively.

There are two other outcomes I investigate in my analysis. The first is the proportion of students who graduated from the first college they attended. The first reason this is of interest is because it acts as a measure of "fit" within the college. If Asian American applicants face a higher barrier of entry at selective schools, resulting in higher rates of rejection of qualified applicants, perhaps they are more likely to succeed at the school that the *do* end up going to. Second, this metric serves as an additional measure of ability. If Asian students have higher average measures of ability than their non-Asian peers at any given institution, then that ability might be reflected in graduation rates. Lastly, I look at the amount of financial aid awarded via scholarships. Long (2007) notes that targeted financial aid is one method of institutions' broader efforts to increase the diversity of its class, and so another potential way colleges could enact racial penalties is through the amount of aid they offer each admitted applicant.

High school GPA and other measures of ability are collected from a combination of the original survey and annual follow ups. However, we obtain reports of household, demographic, and geographic characteristics from round 1 of the survey.

The dataset covers respondents of all major racial groups, but oversamples black and Hispanic respondents, while slightly under sampling Asian respondents. To adjust for this, I plan on utilizing weight variables in my analysis. I obtain these weights from the dataset directly. By suggestion of the BLS, to account for inconsistent participation throughout survey years (some respondents may miss a year of surveys due to scheduling problems), I will limit the sample to respondents sampled in the terminal year of my analysis and use the weights from that year.

3. Empirical Approach

To conduct my analysis, this study utilizes a simple regression model. The explanatory variable is a dummy that takes on the value of 1 if the respondent identifies as Asian, and 0 if not. The ability control is a variable that represents the log of the student's highest recorded combined SAT score out of 1600. Other covariates include parents' education, geographic region, household income, etc. To examine household income, I segment the range of household incomes into four segments and assign a dummy variable to each one.

The outcome variables of interest are indirect measures of college selectivity and college performance. These measures are taken from the first college each respondent attended. There are three separate outcome variables: the first is a dummy that takes on a value of 1 if the college is private and 0 if the college is public. The second is a dummy that takes on a value of 1 if the respondent entered a four-year program and 0 if the respondent entered a 2-year program. The last metric is log tuition, which is a logarithm of the full "sticker price" of the college each respondent attended. The "sticker price" of a college reflects full tuition and additional fees that a respondent would have had to pay to attend the school, *before* any scholarships, tuition remission, or financial aid.

There are two other outcome variables of interest. The first is dummy that takes on a value of 1 if the respondent graduated from college, and 0 if the respondent did not graduate *and* did not transfer to a different college. The second is a variable that represents the percentage of tuition that is paid using tuition remission, scholarships, and/or financial aid. However, not all the aid reflected in this variable is guaranteed to come directly from the school.

The proposed empirical test I'll be performing is the following regression:

$$Y_i = \alpha + \beta_1 S_i + \ \beta_2 D_i + \beta_3 X_i + \epsilon_i$$

Where the dependent variable describes college characteristics, S_i is the log of the student's highest recorded combined SAT score, D_i is a dummy for Asian, and X_i is a vector of controls.

The main assumption required for the validity of the empirical test is that the treatment effect is as good as randomly assigned. To test this, I create a table of descriptive statistics for Asian respondents and for non-Asian respondents and compare the averages. For this analysis, the focus is on characteristics that would affect the colleges any given respondent applies to, as well as characteristics that would affect their chance of acceptance. This includes information about the individual, information about their pre-college education, and information about their household and geographic characteristics.

Even if the descriptive statistics are not similar between Asian and non-Asian cohorts, there are a few strategies I can employ that will simulate race being a random assignment. Specifically, I plan on using the descriptive statistics as controls in my analysis. The controls used in this analysis can be divided into 4 groups: Personal Characteristics, Family Characteristics, High School Characteristics, and Geographic Characteristics. Personal characteristics covers variables such as gender and measures of ability, family characteristics covers household income, parent's education and if both biological parents are present, high school characteristics cover private vs. public, Catholic vs. non-Catholic, and whether the high school offered AP or IB curriculum, and geographic characteristics covers geographical variables such as region (Midwest, south, etc.) and urban vs. rural environment. Adding these covariates will help regulate any omitted variable bias and neutralize any differences between the two groups of respondents.

4. Descriptive Statistics

Table 1 displays descriptive statistics for the NLSY97 cohort separated by Asian vs. non-Asian respondents. High school attributes tend to be similar, while measures of ability, household attributes, and geographic attributes differ between the groups. For measures of ability, the Asian cohort of students scored an average of around one hundred points higher on the SAT, while their Grade Point Average (GPA) in high school is 0.16 points higher (out of 4.0). For household attributes, there are some differences as well. One major difference is that households in which the primary respondent identified as Asian (we will refer to these as "Asian households" to simplify) tend to have higher incomes. While 34% of non-Asian households reported having an income of $70,000 or higher, 45% of Asian households reported an income of above $70,000. Another major difference is that 74% of Asian households report having both biological parents at home, while for non-Asian households that proportion is only 57%. Lastly, there are some geographic differences between the two groups. More Asian households live in urban areas, and less households live in Southern and Midwestern areas, when compared to non-Asian households. To accommodate for these differences in my analysis, I plan on adding these variables as covariates.

5. Results

Column 1, Column 2, and Column 3 of Table 2 display the effect of identifying as Asian on the probability of attending a private institution, probability of attending a four-year program, and the tuition of the institution attended, respectively. Column 4 and Column 5 display the effect of identifying as Asian on the secondary outcome variables, which are the probability of graduating from first college attended and percentage of tuition covered by scholarships and other types of aid. The most notable finding from this table is that Asian respondents were around 12 percentage points less likely to go to a private school as their first college, and this result was statistically significant at a 95% confidence level.

Asian respondents were also 4.7 percentage points less likely to go to a four-year program and 0.18 percentage points more likely to graduate from their first college, although these results were not significant. Additionally, on average, Asians went to institutions that had around 3% higher tuition than their non-Asian counterparts, and the percentage of their tuition covered by scholarships and other types of aid was 2 percentage points lower. However, neither of these findings were significant.

SAT score is highly correlated with the three measures of selectivity. A 1% increase in SAT score raises probability of going to a private college and going to a four-year program by 24 percentage points and 21 percentage points, respectively. A 1% increase in SAT score also increases the tuition of the school attended by 1.3% (this is not a measure of cause and effect, rather, the higher SAT score results in attendance at a more selective institution, which in turn has a higher sticker price). All three of these measures are statically significant at the 1% level. GPA is also correlated with measures of selectivity, with a 1 point increase in grade point average raising the probability of going to a four-year program by 21 percentage points. A 1 point increase in GPA is also correlated with a 37% increase in college tuition, a significant finding. The coefficient on private is positive but insignificant.

These results deviate from the uncontrolled differences in means that are shown in Table 1 (these results are visually represented in Figure 1 and Figure 2). A higher proportion of Asians attended a four-year college, but after adding controls, the probability of going to a four-year college drops between non-Asian and Asian cohorts. A similar discrepancy occurs with the private school outcome and the tuition outcome. The positive coefficients on log SAT score and GPA provide some insight into this result. Asian respondents had higher average SAT scores, which are highly correlated with going to a private school, or to a four-year program. Similarly, students with higher average SAT scores go to schools with higher average tuition (an indirect measure of selectivity). Once we control for SAT score, the probability of going to a more selective school for Asian respondents drops. The private school outcome drops, and the difference is statistically significant.

Since private schools are not perfect indicators of selectivity, it is difficult to interpret this finding in more general contexts. However, private schools have several characteristics that are indicative of a more selective school, including access to more resources and higher graduation rates (IPEDS). Thus, this result seems to align with results of previous literature that finds a penalty placed on Asian applicants during the college admissions process. Since Asians tend to do well on measures of ability such as the SAT, their scores are seen as less competitive than an applicant of a different race with the same score, and thus are less likely to be admitted to private colleges.

However, our other measures of selectivity do not support this theory. Our coefficients on the four year and log tuition outcome are insignificant. Additionally, the analysis shows that even conditional on SAT score and other covariates, Asian applicants tend to go to schools with a higher sticker price. Altogether, the results do not fully support the existence of an Asian American penalty when it comes to college admissions.

One reason for this discrepancy is that this analysis is unique in that it is conducted using data from a wide range of colleges in terms of selectivity. This is important because the distribution of applicants applying to top schools such as Harvard are not representative of the broader distribution of all individuals applying to college. They are more clustered at the upper tail end of the distribution of applicant "quality", and so differences between Asian and non-Asian cohorts are more likely to be very pronounced. This leads to a vast overrepresentation of Asian Americans in upper levels of selectivity that is not observable at more average levels.

The results from my analysis indicate that all else being equal, Asian American applicants are 12 percentage points less likely to go to a private college than their non-Asian counterparts. However, aside from this finding, there is no other evidence supporting the claim that there is a penalty placed on Asian American applicants when it comes college admissions. Research using more recent data could potentially shed more light on this issue and would be more comparable to recent literature that argues otherwise, but these results suggest an interesting finding: that racial discrepancies between Asian and non-Asian applicants at top schools may be due to the unique selection of applicants that the schools attract coupled with the goal of maintaining a diverse group of students, and that this discrepancy is not generalizable to other schools of lower selectivity.

6. Conclusion

Racial discrepancies in college admissions have been a contentious issue since the late 20^{th} century, but more recent discourse is unique in that it focuses on Asian American cohorts, a small but growing share of all college applicants. Specifically, the *Students for Fair Admissions vs. Harvard* case and corresponding literature has raised the question of whether Asian American applicants face a disadvantage, or “penalty”, when it comes to college admissions at Harvard. My study uses data from the NLSY97 cohort to investigate whether there is any difference between Asian and non-Asian college selectivity conditional on measures of ability, and sheds light on the existence of an Asian penalty outside of Harvard. While my analysis found that Asian American applicants were 12 percentage points less likely to go to a private institution, there was no other evidence of disparities in college selectivity between Asian and non-Asian cohorts.

There are several limitations to this study. Firstly, the data was not representative of the overall college applicant pool, as there was a slight underrepresentation of Asian American individuals. Second, most respondents were college aged (18-22) from 1998 to 2006. More recent data might provide more accurate and generalizable results. Lastly, I was unable to observe true college selectivity. Future studies may use more exact measures of selectivity that can further shed light on the landscape of college admissions for Asian American applicants.

Table 1: *Descriptive statistics, by Asian vs. non-Asian*

	Non-Asian	Asian	Total
Primary Outcome variables			
Attended a private college	0.22	0.20	0.22
Attended a four-year college	0.63	0.74	0.63
Total Tuition	$18,915.67	$32,740.94	$19,488.17
Add'l Outcome Variables			
Graduated first college attended	0.88	0.90	0.88
Percent tuition paid through scholarships	25.95%	31.35%	26.17%
Covariates			
Female	0.53	0.45	0.52
Combined SAT score	1064.41	1141.82	1068.26
High School GPA	3.00	3.16	3.01
Household attributes			
Income < $25,000	0.30	0.23	0.30
Income ≥ $25,000 and < $45,000	0.22	0.21	0.22
Income ≥ $45,000 and < $70,000	0.15	0.12	0.14
Income ≥ $70,000	0.34	0.45	0.34
Both biological parents at home	0.57	0.74	0.58
At least one parent went to college	0.44	0.59	0.44
High school attributes			
Attended a private high school	0.04	0.04	0.04
Attended a Catholic high school	0.05	0.02	0.05
High school offered AP or IB curriculum	0.67	0.68	0.67
Geographic attributes			
Urban	0.73	0.92	0.73
South	0.33	0.17	0.32
Midwest	0.28	0.19	0.28

Notes: Sample is weighted using cumulative case weights from the NLSY97 data. Weights are intended to make the sample more representative of the US population. Outcome variables are taken from the year respondents first entered and/or graduated from college. High school attributes are taken from respondents' last year of high school. Household and geographic attributes are taken from the initial survey when possible, and from the earliest year data was available if not possible.

Table 2: *Effect of identifying as Asian on college characteristics*

	(1) Private	(2) 4-year	(4) Log Tuition	(3) Graduated	(5) Scholarship*
Asian	-0.118*	-0.0466	0.0336	0.00182	-2.363
	(0.0493)	(0.0491)	(0.348)	(0.0485)	(8.929)
Female	0.0584*	0.00200	0.0708	-0.00622	-9.069*
	(0.0259)	(0.0217)	(0.138)	(0.0222)	(4.071)
Log SAT Score	0.236***	0.214**	1.338**	0.0259	22.78
	(0.0678)	(0.0725)	(0.450)	(0.0734)	(13.77)
High School GPA	0.0236	0.210***	0.374*	0.0256	15.38**
	(0.0290)	(0.0268)	(0.161)	(0.0288)	(5.263)
Income < \$25,000	-0.0344	-0.0733*	-0.178	0.00181	7.336
	(0.0309)	(0.0296)	(0.286)	(0.0265)	(9.011)
Income ≥ \$25,000	0.0263	0.0158	0.109	-0.0118	0.366
and < \$45,000	(0.0340)	(0.0276)	(0.279)	(0.0281)	(10.87)
Income ≥ \$45,000	0.0225	0.0235	-1.110	-0.0541	-18.14***
and < \$70,000	(0.0397)	(0.0319)	(0.842)	(0.0346)	(5.077)
Both biological	0.0482	0.0382	0.122	-0.0129	-2.832
parents' home	(0.0254)	(0.0248)	(0.136)	(0.0236)	(4.365)
At least one parent	0.00176	0.0887***	-0.246	0.0177	-0.941
went to college	(0.0259)	(0.0245)	(0.200)	(0.0224)	(5.879)
Went to private	0.0501	0.0355	-0.802	-0.0424	2.265
high school	(0.0598)	(0.0532)	(0.482)	(0.0623)	(9.395)
Went to Catholic	0.0936	0.109***	0.335*	0.0246	-0.714
high school	(0.0497)	(0.0294)	(0.162)	(0.0361)	(6.381)
Went to high school	0.0164	0.0441	-0.294	-0.0501	1.508
that offered AP or IB	(0.0385)	(0.0397)	(0.233)	(0.0266)	(6.172)
Urban	-0.0595	0.0681*	0.110	-0.0149	-3.666
	(0.0318)	(0.0286)	(0.160)	(0.0236)	(4.784)
South	-0.0937***	-0.0345	-0.348*	0.0192	0.378
	(0.0266)	(0.0244)	(0.149)	(0.0227)	(4.210)
Midwest	-0.0274	0.0481	0.0734	0.0278	9.858
	(0.0412)	(0.0297)	(0.198)	(0.0285)	(6.254)
N	1349	1352	459	846	480

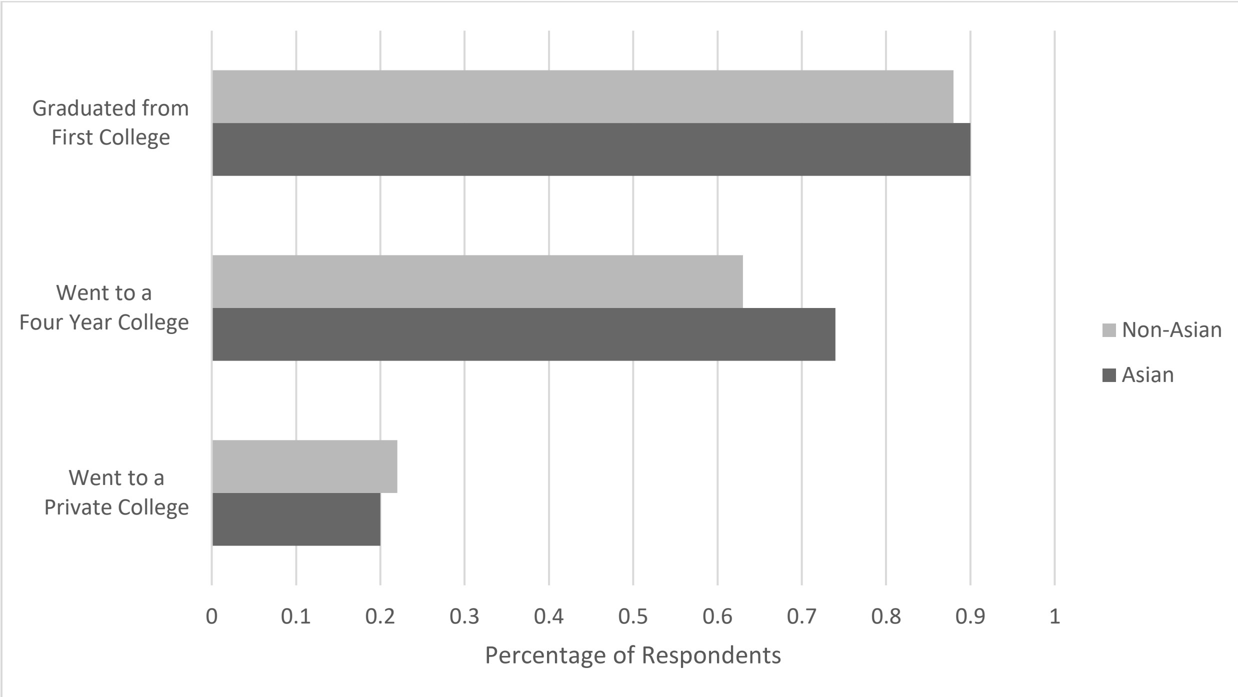
Standard errors in parentheses.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Regression is weighted using cumulative case weights from the NLSY97 data. Weights are intended to make the sample closer to representative of the US population.

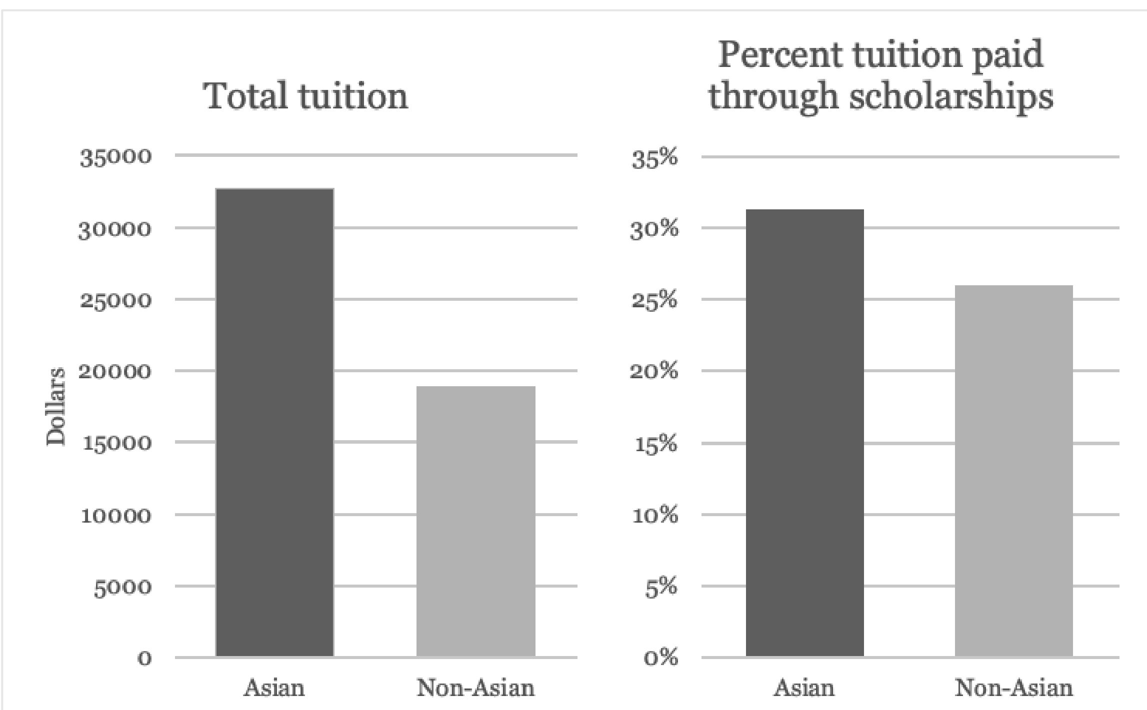
*Scholarship refers to percent of tuition paid through scholarships, financial aid, and/or tuition remission

Figure 1: *Differences in Measures of College Selectivity and Performance, by whether respondent is Asian*



Note: This figure is not indicative of regression results (i.e., no controls added)

Figure 2: *Differences in Measures of College Tuition and Financial Aid, by whether respondent is Asian*



Note: This figure is not indicative of regression results (i.e., no controls added)

References

Harvard Admissions, Harvard (2022). A Brief Profile of the Admitted Class of 2026 n.d. Harvard College. Accessed November 29, 2022.

Bowen, William G., and Derek Bok (1998). *The Shape of the River: Long Term Consequences of Considering Race in College and University Admissions*. Princeton University Press.

Long, Mark (2002a). Race and College Admissions: An Alternative to Affirmative Action? University of Michigan.

Arcidiacono, Peter, Kinsler, Josh and Tyler Ransom (2020). Asian American Discrimination in Harvard Admissions. Duke University.

Arcidiacono, Peter, Kinsler, Josh and Tyler Ransom (2022). What the Students for Fair Admissions cases reveal about racial preferences. Duke University.

Lynch, M. W. (1997). Affirmative Action at the University of California. *Notre Dame Journal of Law, Ethics & Public Policy, 11(1),* 139-158.

Izumi, L. T. (1997). Confounding the Paradigm: Asian Americans and Race Preferences. *Notre Dame Journal of Law, Ethics & Public Policy, 11(1),* 121-138.

Griffith, Amanda L. and Rothstein, Donna S. (2009). Can't get there from here: The decision to apply to a selective college. Economics of Education Review, Volume 28, Issue 5.

VerBruggen, Robert (2002). Racial Preferences on Campus: Trends in Asian Enrollment at U.S. Colleges. Manhattan Institute.

Weisman, Dennis L. and Robinson, Glen O. (2012). Eliminating Racial Preferences in College Admissions *The Economists' Voice*, vol. 9, no. 1, 2012.

U.S. Department of Education, National Center for Education Statistics. (2009). Table 326.27. Number of degree/certificate-seeking undergraduate students entering a postsecondary institution and percentage of students 4, 6, and 8 years after entry, by completion and enrollment status at the same institution, institution level and control, attendance level and status, Pell Grant recipient status, and acceptance rate: Cohort entry year 2009. In U.S. Department of Education, National Center for Education Statistics (Ed.), Digest of Education Statistics (2009 ed.).

Students for Fair Admissions v. Harvard, Pending (2022).

References

Harvard Admissions, Harvard (2022). A Brief Profile of the Admitted Class of 2026 n.d. Harvard College. Accessed November 29, 2022.

Bowen, William G., and Derek Bok (1998). *The Shape of the River: Long Term Consequences of Considering Race in College and University Admissions*. Princeton University Press.

Long, Mark (2002a). Race and College Admissions: An Alternative to Affirmative Action? University of Michigan.

Arcidiacono, Peter, Kinsler, Josh and Tyler Ransom (2020). Asian American Discrimination in Harvard Admissions. Duke University.

Arcidiacono, Peter, Kinsler, Josh and Tyler Ransom (2022). What the Students for Fair Admissions cases reveal about racial preferences. Duke University.

Lynch, M. W. (1997). Affirmative Action at the University of California. *Notre Dame Journal of Law, Ethics & Public Policy, 11(1),* 139-158.

Izumi, L. T. (1997). Confounding the Paradigm: Asian Americans and Race Preferences. *Notre Dame Journal of Law, Ethics & Public Policy, 11(1),* 121-138.

Griffith, Amanda L. and Rothstein, Donna S. (2009). Can’t get there from here: The decision to apply to a selective college. Economics of Education Review, Volume 28, Issue 5.

VerBruggen, Robert (2002). Racial Preferences on Campus: Trends in Asian Enrollment at U.S. Colleges. Manhattan Institute.

Weisman, Dennis L. and Robinson, Glen O. (2012). Eliminating Racial Preferences in College Admissions *The Economists' Voice*, vol. 9, no. 1, 2012.

U.S. Department of Education, National Center for Education Statistics. (2009). Table 326.27. Number of degree/certificate-seeking undergraduate students entering a postsecondary institution and percentage of students 4, 6, and 8 years after entry, by completion and enrollment status at the same institution, institution level and control, attendance level and status, Pell Grant recipient status, and acceptance rate: Cohort entry year 2009. In U.S. Department of Education, National Center for Education Statistics (Ed.), Digest of Education Statistics (2009 ed.).

Students for Fair Admissions v. Harvard, Pending (2022).

The Impact of State Happy Hour Bans on Drunk-Driving-Related Fatal Motor Vehicle Accidents in the US

Carson Collard

November 29th, 2022

Abstract

Happy hour bans – or laws which prohibit the sale of discounted alcoholic beverages – are relevant to current political discourse, yet limited evidence exists to evaluate their effectiveness. To contribute to this discourse, this paper aims to answer, "What is the impact of state-level happy hour bans on drunk driving related fatal motor vehicle accidents in the United States?" This paper exploits the staggered implementation of 10 state-level bans between 1980 and 2019 to estimate the change in fatal drunk driving accidents due to a treated state's happy hour ban using a difference in differences regression. I find no statistically significant difference in rates of drunk driving related fatal motor vehicle accidents between states with and without a ban. Further research is needed to better separate the impact of happy hour ban from other state-level legislation aimed at preventing drunk driving.

Keywords: Drunk driving, Happy hour, DUI

1. Introduction

United States public outrage surrounding drunk driving peaked in the early 1980s after 13-year-old Cari Lightner was killed by a drunk motorist with three previous DUI convictions. Cari's death prompted the foundation of Mothers Against Drunk Driving (MADD), a nationwide non-profit organization which advocates for stricter legislation to prevent drunk driving. In response to these events, both US federal and state governments passed laws aimed at reducing incidences of driving under the influence (DUI) (*Our History*, MADD). Many of these laws – such as raising the minimum legal drinking age to 21, lowering the BAC threshold for drunk driving from 0.15 to 0.08, improving police enforcement of drunk driving laws, and increasing penalties for drunk driving offenders – have well-founded empirical evidence supporting their effectiveness (Gerstein, 1985; Carpenter & Dobkin, 2011). In part due to these laws, alcohol-related driving fatalities have fallen significantly since 1980 (**Figure 2**).

In line with this effort, ten states passed a ban on happy hour drink specials (**Table 1**). The bans originated from the hypothesis that a happy hour's inexpensive drinks lead to excessive drinking within a shortened period and, in turn, heightened incidents of drunk driving and drunk driving accidents (Kim, 2022). Unlike the more prominent legislative changes listed above, there is surprisingly limited evidence on the efficacy of happy hour bans in preventing drunk driving (Puac-Polanco et al., 2020).

Happy hour bans are a current topic of political discourse, as several states have begun to question whether the bans are an effective part of their legislative strategy to limit drunk driving. Three states (Illinois, Kansas, and Oklahoma) have already repealed their happy hour bans and at least one other (Massachusetts) has recently (July 2022) introduced legislation to overturn its ban (Kim, 2022). The purpose of this research is to provide evidence to inform discussion surrounding the efficacy of happy hour bans. More specifically, this research seeks to answer, "what is the impact of state-level happy hour bans on drunk driving related fatal motor vehicle accidents in the United States?"

To answer this question, I exploit the staggered implementation of 10 state happy hour bans between 1980 and 2019 to set up a state-level difference in differences regression. This regression aims to estimate the change in drunk driving related motor vehicle fatalities per annual vehicle miles traveled (VMT) due to a state's happy hour ban. I then use an event study specification to further estimate the impact of a happy hour ban each year after the ban is implemented. Data on drunk driving related fatal motor vehicle accidents was retrieved from the NHTSA FARS database. This data was then translated to crash rates per vehicle miles traveled using data from the FHA.

Both my regressions yielded estimates not statistically significant from zero for the difference in drunk driving related fatal motor vehicle accidents between states treated with a happy hour ban and states not treated with a happy hour ban. However, the results are still useful in estimating a range of possible impacts from happy hour bans. The 95% confidence interval of the estimate, which has bounds of -0.390 and +0.607 fatal drunk driving accidents per billion vehicle miles traveled, offers policymakers a window within the true effect of a happy hour ban is most likely to lie.

This research builds upon previous literature investigating the impact of drink pricing specials on alcohol consumption and drunk-driving outcomes. *Babor et al 1978*, the earliest paper on the impact of happy hours I could find, aimed to answer whether happy hours led to increased alcohol consumption. To test this, the researchers conducted an RCT where 34 male volunteers who self-identified as either a casual or heavy drinker were given the option to purchase alcoholic drinks during a 20-day period. Half the subjects in each category were treated with the option to purchase 50% discounted drinks for three hours each day during the afternoon. The researchers observed 2 times higher consumption in self-identified casual drinkers and 2.4 times higher consumption in self-identified heavy drinkers among the treated group.

A second study, *Smart & Adlaf 1986*, aimed to answer whether happy hour bans in Ontario impacted alcohol consumption or impaired-driving charges. To test this, the researchers collected observational data from five taverns in Ontario before and after a province-wide ban on happy hour discounts on alcohol. The researchers found no difference in alcohol purchasing or consumption before and after the ban. The researchers did observe a decrease in impaired driving charges but could not causally attribute the decline to the independent effect of happy hour ban because there was no observed change in alcohol consumption at the tavern or individual level.

A 2020 meta-analysis from *Puac-Polanco et al.* analyzing the effects of drink specials concluded "studies examining drink specials showed consistency in reporting negative individual-level consequences related to higher alcohol use" but determined "further research is needed to determine whether regulations of drink specials… can help to discourage high-risk groups from engaging in problematic drinking behavior". This study's goal is to estimate whether happy hour bans can discourage the problematic drinking behavior of drunk driving.

The remainder of the paper proceeds as follows: Section 2 outlines the three sources used to construct my dataset, as well as potential risks of measurement error. Section 3 outlines the empirical methods used and the assumptions necessary for those methods to hold validity. Section 4 details the results from the difference in differences regression and offers an interpretation for the statistical insignificance of the result. Section 6 concludes the research and explains potential implications for public policy.

2. Data

This analysis uses three separate categories of data: data on motor vehicle accidents from the NHTSA FARS database, data on vehicle miles traveled per state from the FHA's Table VM-2, and information on the effective date of happy hour bans from the ProQuest Historical Newspapers database.

2.1 NHTSA FARS

The principal source of data for this analysis is the National Highway Traffic Safety Administration's (NHTSA) Fatality Analysis Reporting System (FARS). The NHTSA FARS is a nationwide census of fatal injuries resulting from motor vehicle accidents and has been published annually between 1975 and 2020. Each year the FARS database reports on the individual characteristics of between 30,000 and 40,000 vehicle crashes which resulted in at least one fatality.

The data in the FARS is broken down into three primary datasets: ACCIDENT, VEHICLE, and PERSON. This research uses the ACCIDENT data file, which lists data for each recorded accident over the course of the calendar year. The ACCIDENT file contains a binary variable codifying whether the police reported at least one driver involved in the accident was driving under the influence of alcohol, which is used to construct the dependent variable. The dataset also includes variables for YEAR and STATE, which prove useful in constructing a diff-in-diff empirical test. The FARS contains data for over 1.6 million crashes where the police reported alcohol involvement between 1980 and 2019.

Data for the FARS is compiled primarily using police crash reports, but is supplemented with state highway department data, medical records, toxicology reports, and vehicle registration data. Due to the resourcing, auditing, visibility, and comprehensiveness of the FARS, it is unlikely to contain significant measurement error.

The FARS is limited in its usefulness for creating regression controls. Variables such as urban vs. rural and pavement type were not added to the FARS till after most of the treated states had implemented their happy hour ban. As such, any fixed-effect estimates for regressions using FARS data must come from external datasets (such as Table VM-2). Fortunately, it is unlikely these factors describing road conditions changed significantly between the pre- and post-treatment periods.

2.2 FHA Table VM-2

The most pronounced difference between states for the purposes of this analysis is the amount of driving done within a state each year. Regardless of happy hour ban status, California will have more fatal vehicle accidents than Alaska simply because there are more people driving in California. To control for this difference, results in this paper are reported in crashes per million vehicle miles traveled (VMT). Data on VMT per state was retrieved from the Federal Highway Administration's Highway Statistics Series Table VM-2. Table VM-2 includes state-level estimates for the total VMT each year. VMT is calculated by each state's department of transportation using a combination data from travel surveys, fuel sales, and odometer samples before being reported to the FHA. Because each state uses its own procedure for estimating VMT, there is likely to be some measurement error in Table VM-2. However, there is no reason to believe this error would be different for states with happy hour bans.

Table VM-2 only includes estimates for the VMT in each state for the year 2019. While this provides relative levels of travel volume for each state, it is less precise than coding a separate VMT estimates for each state and each year in the sample. If states have significantly different VMT in treated states after treatment compared to before treatment, it would limit the efficacy of Table 2's VMT estimates as controls.

2.3 ProQuest Historical Newspapers

The exact dates for the implementation of each happy hour ban were determined using historical newspaper articles from ProQuest's Historical Newspaper database. Links to each newspaper article are included in **Table 1**. Because each news source is from the time and local where each ban was implemented, it is unlikely there are significant reporting errors. For this study, a happy hour ban is defined as the complete prohibition of discounted alcoholic beverages for limited hours during the day. As shown in **Figure 1**, there are several states with laws restricting happy hour to specific times of day or limiting the amount that alcohol can be discounted during happy hour. In the specifications described in **Section 3**, these states are included in the untreated group. Further analysis outside the scope of this research is necessary to separately estimate the impacts of these happy hour restrictions.

3. Empirical Methods

To test the impact of happy hour bans on drunk driving accidents, I begin with a staggered difference in differences estimate of the change in drunk driving related motor vehicle fatalities per annual VMT due to a state's happy hour ban. Then, to confirm the difference in differences result and better visualize the estimation, I use an event study to estimate the change in drunk driving fatalities in treated states for each year before and after a ban is implemented.

3.1 Staggered Difference in Differences Estimation

The specification for my staggered difference in differences estimate is as follows:

$$Y_{st} = \alpha + \delta_{rDD} HHBAN_{st} + \sum_{i=AL}^{WY} \beta_i STATE_{is} + \sum_{k=1980}^{2019} \gamma_k YEAR_{ks} + \varepsilon_{st}$$

Where Y_{st} is the number of drunk driving-related fatal vehicle accidents per million VMT in state *s* at year *t*, $HHBAN_{st}$ is a dummy coding for whether state *s* has a happy hour ban at year *t*, β_i is the state-specific effects, γ_k is the year-specific effects, and δ_{rDD} is the difference in differences estimate. This specification was used to run three separate regressions. The first regression uses a sample of only accidents where the police reported an intoxicated driver. The second uses only accidents where the police *didn't* report an intoxicated driver, for use as a "placebo". The third uses the combined samples from the first two regressions. The results of these regressions are listed in **Table 4**. Standard errors in this specification are clustered by state.

While running this empirical test, the biggest assumption I make is happy hour bans are the only major causal factor which could create an increase or decrease in rates of fatal drunk driving accidents in treated states after the treatment relative to the untreated states. Unfortunately for this analysis, there were several other pieces of legislation in the mid-to-late 1980s also targeted at reducing instances of drunk driving. For example, the Drunk Driving Prevention Act of 1988 increased penalties for drivers caught driving under the influence, created open container laws, mandated alcohol education for new drivers, and created a zero-tolerance policy for drivers under the age of 21 (Gerstein 1985). The National Minimum Drinking Age Act of 1984 also effectively forced all states to raise their drinking age to 21 by October 1986 (Gerstein 1985). However, most pieces of drunk driving legislation were passed at the national level, meaning they should have a roughly similar effect on both the treatment and control groups.

Another classic difference in differences assumption is the existence of parallel trends between the treatment and control groups in each state. Based on the estimates displayed in **Figure 3**, I argue this assumption is reasonably upheld. Prior to the treatment at t = 0, the difference between treated and untreated states is not statistically significant from zero for any of the estimates.

3.2 Event Study Estimation

The specification for my event study is as follows:

$$Y_{st} = \alpha + \sum_{j=-\infty}^{\infty} \delta_j HHBAN_{i,t-j} + \sum_{i=AL}^{WY} \beta_i STATE_{is} + \sum_{k=1980}^{2019} \gamma_k YEAR_{ks} + \varepsilon_{st}$$

This event study specification is very similar to the difference in differences specification, with identical methods for estimating state-specific effects (β_i) and year-specific effects (γ_k). This specification diverges from the difference in differences only in its estimation of the treatment effect, δ_j. The treatment effect in this specification is allowed to vary in each time period before and after the treatment. In other words, the treatment effect is estimated separately for each year leading up to and each year after the implementation of a happy hour ban for each of the 10 treated states. The results of this specification are shown in **Figure 3**.

The estimation window begins four years prior to the implementation of a ban, since the first happy hour ban (Massachusetts) occurred in 1984 four years after the beginning of the available data. The estimation window ends ten years after the implementation of a ban since any long-term effects from a happy hour ban would be detectable by ten years after implementation.

4. Results

Results for the difference in differences estimate for the impact of state happy hour bans on rates of fatal drunk driving related vehicle accidents are summarized in **Table 4.** The estimated treatment effect is 0.109 additional fatal car crashes each year for every billion VMT. For the state of Massachusetts, which Table VM-2 estimates as accumulating 64.9 billion VMT each year, this estimate translates to an increase of seven crashes each year due to its happy hour ban. Though every drunk driving crash is tragic, this number is relatively small compared to the hundreds of fatal car crashes in Massachusetts each year. The placebo regression, which uses a sample of fatal vehicle accidents where alcohol wasn't involved, estimates a similarly statistically insignificant result.

There are two important things to note about this estimate. First, the estimate is slightly positive. This is opposite of the hypothesis supporting happy hour bans as a tool to decrease drunk driving. Second, the estimate is not statistically significant from zero. Because of this, it is not conclusive based on this estimate whether happy hour bans have had any impact on rates of drunk driving related vehicle fatalities.

Results for the event study estimates tell a very similar story. Surprisingly, the rates of drunk driving fatalities appear to increase in the two years following a happy hour ban for treated states when compared to untreated states. However, neither of these increases are statistically significant from zero. In fact, in no year throughout the 14-year treatment window were the impact of happy hour bans statistically significant from zero.

Despite the statistical insignificance of these outcomes, the estimates from both the difference in differences and event study specifications can be useful tools in discussions surrounding the efficacy of happy hour bans. For those most optimistic about the impact of happy hour bans in decreasing drunk driving, who take the lower bound of the 95% confidence interval as fact, happy hour bans could eliminate 0.390 fatal drunk driving car crashes each year per billion VMT. Back to the Massachusetts example, this would result in 25 fewer fatal crashes each year. For those most pessimistic about the impact of happy hour bans in decreasing drunk driving, who take the upper bound of the 95% confidence interval as fact, happy hour bans could result in 0.607 extra fatal drunk driving car crashes each year per billion VMT. For Massachusetts, this would result in 39 additional fatal crashes each year. As a policymaker, it is useful to understand the range of feasible outcomes which a happy hour ban could cause.

It's difficult to imagine a scenario where a happy hour ban would cause a legitimate increase in drunk driving fatalities. In my opinion, its most likely happy hour bans have little to no effect on drunk driving fatalities and the marginally positive treatment effect is a product of estimation error.

5. Conclusion

Reducing deaths from drunk driving crashes is a valiant goal. One policy lawmakers have implemented in an attempt to reduce drunk driving is bans on happy hour drink specials. In recent years, the effectiveness of these bans has come into question, yet little empirical evidence exists as to how happy hour bans impact drunk driving outcomes.

In this analysis, the impact of happy hour bans on drunk driving related motor vehicle fatalities was estimated using difference in differences and event study specifications. These specifications utilize crash data from the NHTSA's FARS census and VMT data from the FHA's Table VM-2. Both specifications estimate treatment effects of happy hour bans that are not statistically significant from zero. Despite statistical insignificance, these estimates still are informative though estimating a window of possible impacts from happy hour bans, which ranges from -0.390 to +0.607 fatal drunk driving accidents per billion vehicle miles traveled.

Several extensions of this research are necessary to fully inform policy decisions. First, a similar analysis with yearly VMT data could eliminate a major potential source of estimation error. Second, an analysis which separates treatment into the three buckets of happy hour ban, happy hour restriction, and unrestricted happy hour could more precisely estimate the outcomes of the spectrum of policy options. Third, a localized analysis estimating the treatment effect of one state with a happy hour ban against one state without a happy hour ban could better control for idiosyncrasies in state laws.

Despite the limitations of this study, it creates some empirical basis to the claim that happy hour bans are an ineffective tool in preventing drunk driving. If further analysis were to corroborate this claim, it could help policymakers to focus drunk driving legislation on restrictions with proven positive outcomes.

6. References

Alcohol Policy Information System (APIS). (2021). *Alcohol Beverages Pricing: Drink specials*. National Institute on Alcohol Abuse and Alcoholism. Retrieved November 10, 2022, from https://alcoholpolicy.niaaa.nih.gov/apis-policy-topics/drink-specials/2

Babor, T. F., Mendelson, J. H., Greenberg, I., & Kuehnle, J. (1978, January). *Experimental Analysis of the 'Happy Hour": Effects of purchase price on alcohol consumption*. PubMed. Retrieved November 3, 2022, from https://pubmed.ncbi.nlm.nih.gov/97717/

Carpenter, C., & Dobkin, C. (2011). The minimum legal drinking age and public health. U.S. National Library of Medicine. Retrieved November 3, 2022, from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3182479/

Eddie Kim, "How Happy Hour Became Illegal in a Number of Different States," MEL Magazine, April 26, 2022, https://melmagazine.com/en-us/story/why-is-happy-hour-illegal-in-massachusetts/

Enomoto, K. (2015). *2015 Report to Congress on the Prevention and Reduction of Underage Drinking*. Stop Alcohol Abuse. Retrieved November 28, 2022, from https://www.stopalcoholabuse.gov/about-iccpud/data/national-reports/report-to-congress/default.aspx

Fatality Analysis Reporting System (FARS). (2021). *NHTSA file downloads*. National Highway Traffic Safety Administration. Retrieved November 10, 2022, from https://www.nhtsa.gov/file-downloads?p=nhtsa%2Fdownloads%2FFARS%2F

Gerstein, O. S. (1985). *Preventing drunk driving - Alcohol in America - NCBI Bookshelf*. National Library of Medicine. Retrieved November 3, 2022, from https://www.ncbi.nlm.nih.gov/books/NBK217455/

Our History. MADD. (2022, April 20). Retrieved November 28, 2022, from https://madd.org/our-history/

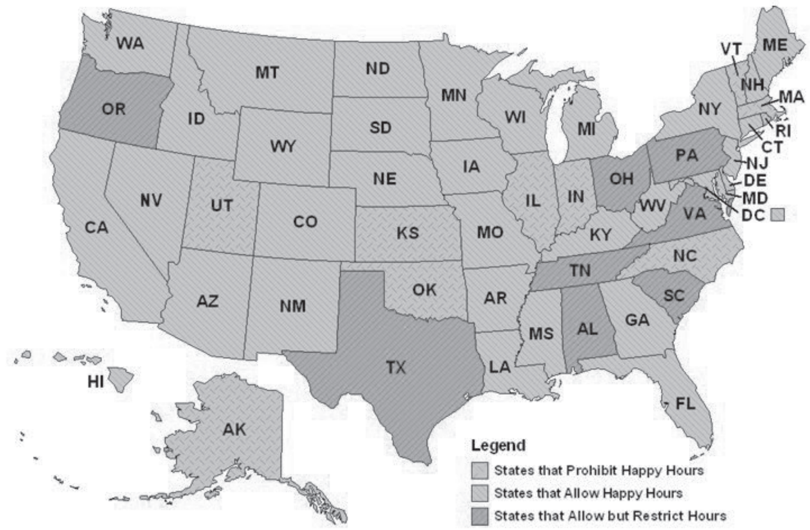
Puac-Polanco, V., Keyes, K. M., Mauro, P. M., & Branas, C. C. (2020). A Systematic Review of Drink Specials, Drink Special Laws, and Alcohol-Related Outcomes. *Current epidemiology reports*, *7*(4), 300–314. https://doi.org/10.1007/s40471-020-00247-0

Smart, R. G., & Adlaf, E. M. (1986, May). *Banning happy hours: The impact on drinking and impaired-driving charges in Ontario, Canada*. US National Library of Medicine. Retrieved November 3, 2022, from https://pubmed.ncbi.nlm.nih.gov/3724165/

Table VM-2. (2019). *Highway Statistics Series*. Office of Highway Policy Information. Retrieved November 10, 2022, from https://www.fhwa.dot.gov/policyinformation/statistics/2019/vm2.cfm

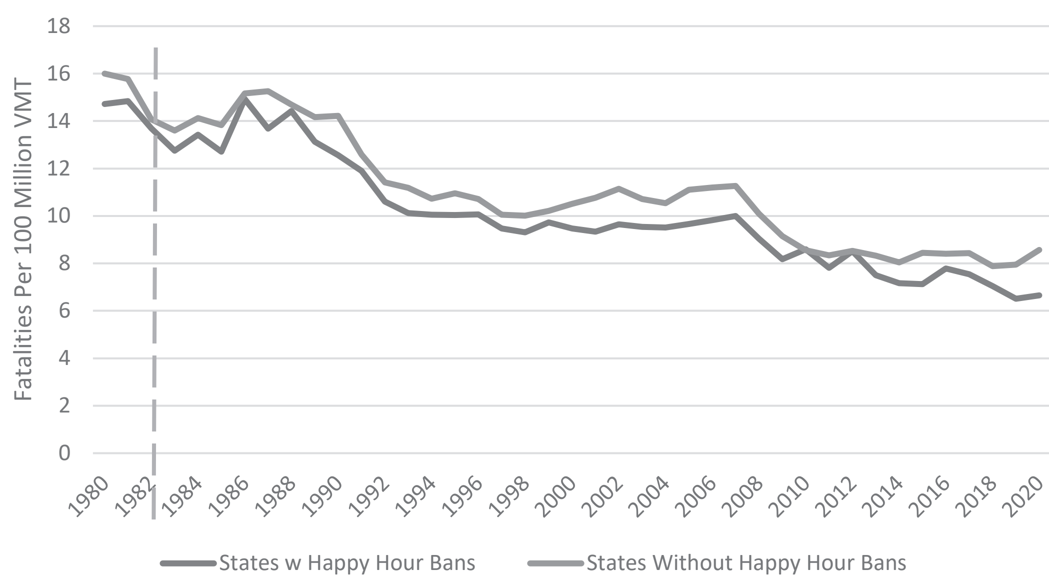
7. Appendix A: Tables and Figures

Figure 1: *Status of State Happy Hour Legislation in 2012*



Note: Map of state-level happy hour laws in the United States. Only states in green, which completely prohibit(ed) happy hours, were considered treated for this research. Retrieved from Enomoto et al., 2015.

Figure 2: *Alcohol-Related Fatal Driving Accidents Per 100 Million Vehicle Miles Traveled*



Note: Line graph comparing treated and untreated states' rates of drunk driving crashes between 1980 and 2019. A vertical grey line is added in 1984, when MA became the first state to explicitly ban happy hour. Analysis using annual data for all 50 states retrieved from the FARS. VMT numbers retrieved from 2019 table VM-2.

Table 1: *Summary of State-Level Happy Hour Bans*

Jurisdiction	Postal Code	Ban Start Date	Ban End Date	Source
Alaska	AK	Sep-86	Ongoing	Link
Illinois	IL	Sep-89	7/14/2015	Link
Indiana	IN	Oct-85	Ongoing	Link
Kansas	KS	Jun-85	6/30/2012	Link
Massachusetts	MA	Dec-84	Ongoing	Link
North Carolina	NC	Aug-85	Ongoing	Link
Oklahoma	OK	Jan-80*	9/30/2018	Link
Rhode Island	RI	Jun-85	Ongoing	Link
Utah	UT	Jul-11	Ongoing	Link
Vermont	VT	Aug-86	Ongoing	Link

*Note: Overview of state legislation banning happy hours. Dates retrieved from ProQuest Historical Newspapers. *In Oklahoma, it was illegal for licensed liquor vendors to sell alcohol by the drink until 1984. After 1984, Oklahoma legalized the practice of selling "Bar Drinks", but at the same time implemented an explicit happy hour ban. As such, Oklahoma had a de facto happy hour ban from the beginning of the dataset until October 2018.*

Table 2: *Rates of Vehicular Accidents Resulting in at Least One Fatality Per Million Vehicle Miles Traveled (1980-2019)*

	State	Without Alcohol Involvement (1)	With Alcohol Involvement (2)	Total Fatal Vehicular Accidents (3)
States With a Happy Hour Ban	Alaska	1.069	0.751	1.821
	Illinois	1.042	0.567	1.609
	Indiana	1.004	0.392	1.396
	Kansas	1.191	0.525	1.716
	Massachusetts	0.700	0.329	1.028
	North Carolina	0.989	0.464	1.453
	Oklahoma	1.317	0.644	1.961
	Rhode Island	0.841	0.579	1.420
	Utah	0.823	0.262	1.085
	Vermont	0.864	0.632	1.496
States Without a Happy Hour Ban	Alabama	1.205	0.505	1.710
	Arizona	1.093	0.490	1.583
	Arkansas	1.218	0.646	1.864
	California	0.946	0.544	1.490
	Colorado	0.785	0.513	1.297
	Connecticut	0.848	0.544	1.392
	Delaware	0.877	0.572	1.449
	Florida	1.059	0.456	1.515

Georgia	0.970	0.380	1.350
Hawaii	0.836	0.623	1.459
Idaho	1.107	0.573	1.680
Iowa	1.147	0.500	1.647
Kentucky	1.368	0.662	2.030
Louisiana	1.450	0.676	2.126
Maine	0.978	0.603	1.582
Maryland	0.878	0.413	1.291
Michigan	1.022	0.573	1.595
Minnesota	0.710	0.398	1.108
Mississippi	1.737	0.443	2.181
Missouri	1.077	0.482	1.559
Montana	1.118	0.971	2.089
Nebraska	0.975	0.532	1.507
Nevada	0.804	0.479	1.283
New Hampshire	0.749	0.511	1.260
New Jersey	0.889	0.387	1.276
New Mexico	1.249	0.719	1.969
New York	1.307	0.407	1.714
North Dakota	0.844	0.539	1.382
Ohio	1.029	0.513	1.541
Oregon	1.070	0.631	1.701
Pennsylvania	1.247	0.645	1.892
South Carolina	1.305	0.678	1.983
South Dakota	1.095	0.740	1.835
Tennessee	1.119	0.568	1.686
Texas	1.056	0.435	1.492
Virginia	0.861	0.457	1.317
Washington	0.709	0.561	1.270
West Virginia	1.671	0.810	2.480
Wisconsin	0.766	0.544	1.311
Wyoming	1.118	0.708	1.826

Note: State-by-state summary comparing treated and untreated states' rates of >5 million individual fatal crashes reported between 1980 and 2019. Analysis using annual data for all 50 states retrieved from the FARS. VMT numbers retrieved from 2019 table VM-2.

Table 3: *Rates of Vehicular Accidents Resulting in at Least One Fatality Per Million Vehicle Miles Traveled (1980-2019)*

	Without Alcohol Involvement	With Alcohol Involvement	**Total Fatal Vehicular Accidents**

	(1)	(2)	**(3)**
States with a Happy Hour Ban	0.993	0.471	1.464
States without a Happy Hour Ban	1.048	0.513	1.561

Note: Summary comparing treated and untreated states' rates of >5 million individual fatal crashes reported between 1980 and 2019. Analysis using annual data for all 50 states retrieved from the FARS. VMT numbers retrieved from 2019 table VM-2.

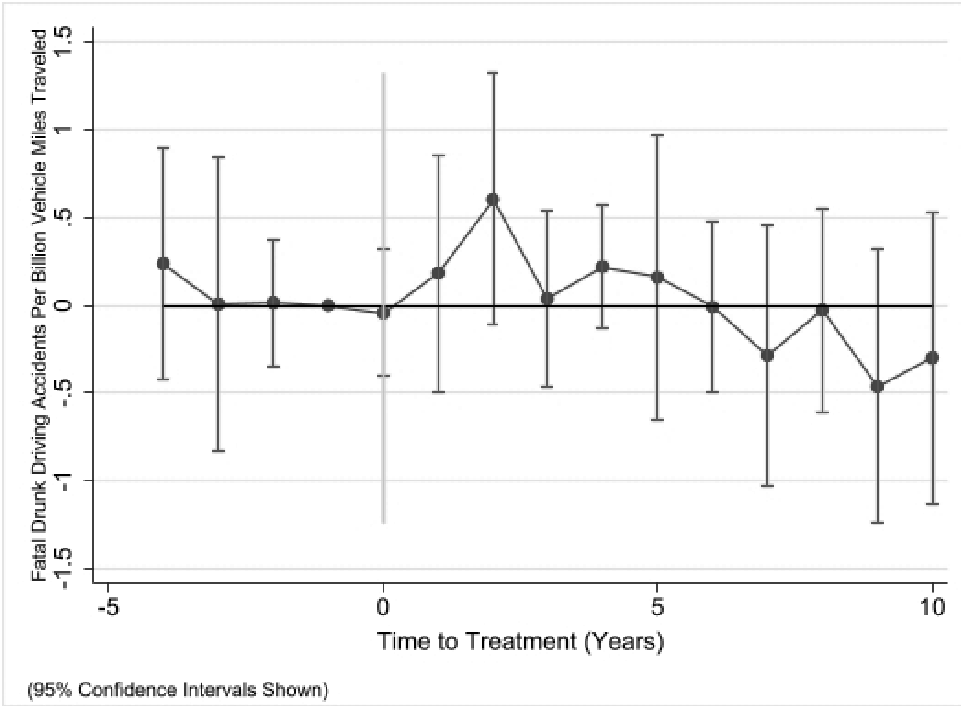
Table 4: *Staggered Difference-in-Differences Regression*

$$Y_{st} = \alpha + \delta_{rDD}HHBAN_{st} + \beta_i STATE_{is} + \gamma_k YEAR_{ks} + \varepsilon_{st}$$

	(1)	(2)	(3)
Treatment Effect	0.109 (0.248)	-0.419 (0.411)	-0.311 (0.506)
Constant	5.862 (0.273)	9.470 (0.324)	15.331 (0.386)
Drunk Driving Crashes Included	Yes	No	Yes
Sober Driving Crashes Included	No	Yes	Yes
N	1799	1799	1799

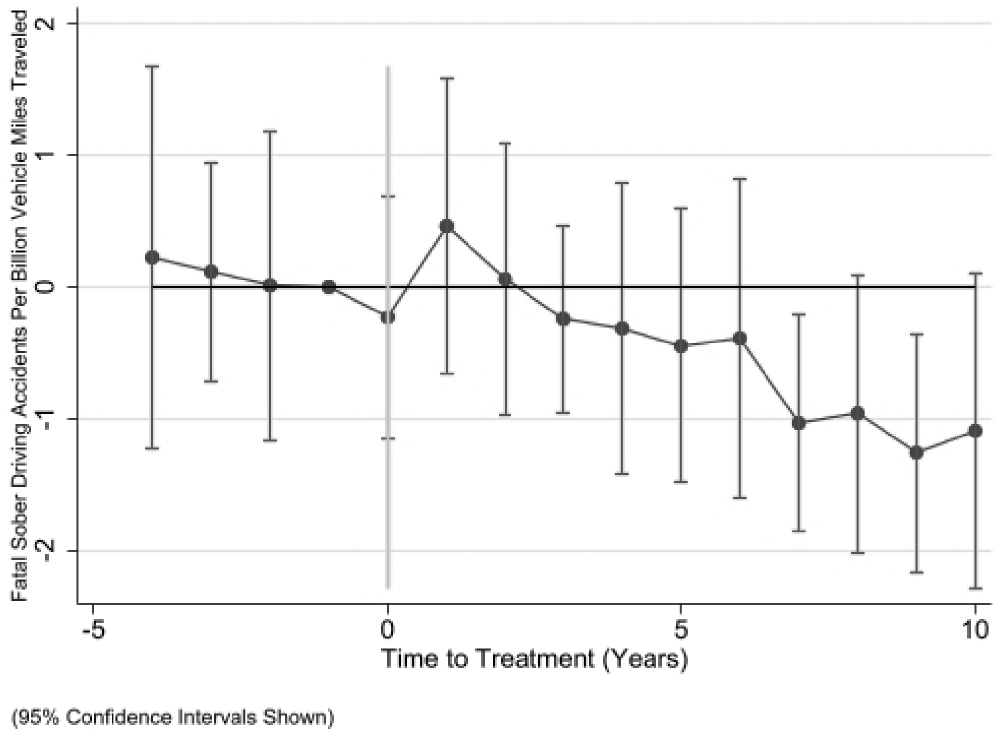
Note: Analysis using annual data for all 50 states retrieved from the FARS. VMT numbers retrieved from 2019 table VM-2. Treatment effect represents the estimated change in motor vehicle fatalities per billion annual VMT due to a state's happy hour ban. Difference-in-differences models. Sample in Column (1) includes only drunk driving crashes, sample in Column (2) includes only sober driving crashes, sample in Column (3) includes all crashes. Robust standard errors clustered at the state level shown in parentheses.

Figure 3: *Staggered Difference-in-Differences Regression Event Study Plot (Drunk Driving Accidents)*



Note: The blue line plots the difference in drunk driving crash rates between treated and untreated states in each year before and after the implementation of a happy hour ban. Happy hour bans implemented at time t=0, represented by the vertical cyan line. Analysis using annual data for all 50 states retrieved from the FARS. VMT numbers retrieved from 2019 table VM-2. Treatment effect represents the estimated change in motor vehicle fatalities per billion annual VMT due to a state's happy hour ban. Difference-in-differences models. Robust standard errors clustered at the state level represented as vertical red bars. See section 2 for more detailed data and variable descriptions.

Figure 4: *Staggered Difference-in-Differences Regression Event Study "Placebo" Plot (Sober Driving Accidents)*



Note: The blue line plots the difference in sober driving crash rates between treated and untreated states in each year before and after the implementation of a happy hour ban. Happy hour bans implemented at time t=0, represented by the vertical cyan line. Analysis using annual data for all 50 states retrieved from the FARS. VMT numbers retrieved from 2019 table VM-2. Treatment effect represents the estimated change in motor vehicle fatalities per billion annual VMT due to a state's happy hour ban. Difference-in-differences models. Robust standard errors clustered at the state level represented as vertical red bars. See section 2 for more detailed data and variable descriptions.

The Effect of Non-Medical Vaccine Exemption Policy Changes on School Enrollment and Use of Exemptions

Sarah Aaronson

April 25, 2022

Abstract

Throughout the 21st century, misinformation has led to concerns about vaccinations within the United States. In response to increasing use of exemptions for required school vaccines, many states have made it more difficult to obtain them including fully banning non-medical vaccine exemptions (NMEs). This paper uses comprehensive state-level data on vaccination rates, exemption rates, and school enrollment rates to explore the effect of these different policies. Overall, these policies lead to around a 2 percentage point increase in many types of vaccination in the year the policy comes into effect. Further, they lead to a sustained decrease in the use of exemptions up to seven years after the policy is enacted. These effects appear to be driven by stricter policy changes in California, Colorado, and New York. In certain states, I also find effects on 5-year-old child school enrollment..

In the United States, each state requires some level of vaccination in order for children to attend school. While these requirements help create a high level of vaccination across the country, each state has varying policies on gaining exemption from the vaccine requirements. All states allow some sort of medical exemption (NCSL 2022), but there are significant differences in non-medical exemptions (NMEs). Many states allow for religious and/or philosophical exemptions under the justification of protecting parents' autonomy. However, opposition to vaccination has been growing in the US along with the use of these vaccine exemptions in many states (Olive et. al. 2018). Because of this and the instance of notable outbreaks of vaccine-preventable diseases (VPDs), state legislatures have moved to make it more difficult to obtain these exemptions through steps such as education modules, obtaining a healthcare provider's signature, or outright banning non-medical exemptions as a whole. This context took on even more relevance with the onset of the COVID-19 pandemic and controversies regarding the vaccines used to combat it. Thus, in this study, I exploit the variation in policies on NMEs across time and space to understand the effect of making obtaining NMEs more difficult on the use of these exemptions and the level of kindergarten vaccination for VPDs. Results show that the policies on average increased the level of vaccination for each type of vaccine by around 1 to 2 percentage points each in the school year the policy took effect. The next year also lead to further increases in vaccination of around 1 to 2 percentage points again. This effect appears to be mainly driven by the policies in California, Colorado and New York, which when viewed separately, had positive and statistically significant effects on all types of required vaccinations within the state. Considering California's policy was a ban on NMEs, Colorado implemented an annual renewal for exemption and New York banned religious exemptions, this is potentially unsurprising.

There are four main recent articles in this area that I have referenced in writing this paper. First, in Olive et al. (2018), the authors find that non-medical (both religious and philosophical) school vaccine exemptions are negatively correlated with measles, mumps, and rubella (MMR) vaccine coverage of kindergarteners. While this article acknowledges the various changes in exemption policies during their study period, they do not harness the variation to examine the impact of these policies. The results are only descriptive in nature. In Hair, Gruber, and Carly (2020), the authors find that 2003 legislation in Texas and Arkansas that permitted personal belief exemptions led to decreased vaccination coverage among black and low-income preschoolers by 16.1% and 8.3%, respectively and that those cohorts performed worse on standardized tests in middle school. While the authors obtain data on legislative changes across the country up until the present day, they restrict their results to only examining the two changes in Texas and Arkansas that changed the landscape from no personal belief exemptions to allowing them. Finally, Blank, Caplan, and Constable (2013) classify the difficulty of obtaining a NME and assess the correlation between this measure and the number of NMEs in a given state. Similarly to Olive et al. (2018), they do not examine the effects of changes in legislation on the use of the NMEs, although they do discuss the various policy changes. My work will likely most closely mirror that of Richwine, Dor, and Moghtaderi (2019). The authors find, using a DD design, that California's elimination of all NMEs in 2016 led to an increase in vaccination coverage for all required vaccines, ranging from 2.5% for MMR to 5% for Polio. They also observe that while NMEs decreased, they were potentially substituted with medical exemptions due to a lack of monitoring of the process. To the best of my knowledge, there is no literature connecting US vaccination exemption policy to school enrollment in general or public versus private school enrollment. I hypothesize a relationship due to the seriousness of some parents' beliefs regarding vaccination. However, these results may only be visible on a more granular level such as within a country or school district.

This paper is organized as follows. Section 1 provides background on the school vaccine policy landscape in the United States. Section 2 outlines the data sources used. Section 3 describes the empirical strategy and evidence for satisfying assumptions. Section 4 presents main results of vaccination policy on usage rates and school enrollment. Finally, section 5 concludes.

1 Background

School vaccination policy within the US begins in the 1850s when Massachusetts required inoculation against smallpox. Since then, all 50 states have enacted laws requiring schoolchildren to be vaccinated against certain diseases, many in line with the Center for Disease Control and Prevention's (CDC) Advisory Committee on Immunization Practices recommendations. Vaccination requirements have been deemed constitutional in Supreme Court cases such as Zucht v. King in 1922 and Maricopa County Health Department vs. Harmon in 1987. However, this has not stopped states from allowing for exemptions to vaccination requirements for both medical and non-medical reasons. In fact, 44 out of 50 US states allow for religious and/or philosophical exemptions to protect parent autonomy. Misplaced concerns over vaccine safety and side effects have made these allowances a liability to states in recent years. Figure 1 shows the increasing proportion of kindergartners on vaccine exemptions over the past ten years. Outbreaks of measles in California in 2015, Minnesota in 2017, New York in 2018 and other examples have led states to crack down on exemption policies with more tedious requirements and outright bans of the exemptions. Figure 2 summarizes policy changes in six states designed to make obtaining non-medical exemptions (NMEs) more difficult.

2 Data

My analysis draws upon two principal data sources to parse out the effect of these policies on exemption, vaccination, and school enrollment rates. First, data on the percent of kindergartners vaccinated in each state or using exemptions comes from the Centers for Disease Control and Prevention, in particular the National Center for Immunization and Respiratory Diseases (NCIRD). Data is available from the 2009-10 school year through the 2019-20 school year. It includes data on the Varicella (chickenpox); Measles, Mumps, and Rubella (MMR); Hepatitis B; Diphtheria, tetanus, & acellular pertussis (DTaP); and polio vaccines; as well as whether an exemption was medical or non-medical. It does not break down religious versus philosophical vaccine exemptions. Thus, I will be looking at the combined effect of all non-medical exemptions (NMEs). The data set also includes US medians for each year.

Table 1 shows summary statistics for various required vaccinations and types of exemptions. Observations are at the state-year level. The CDC's goal for vaccination coverage is 95% for reference.

Secondly, school enrollment data comes from the American Community Survey (ACS) 2009 through 2019. I specifically look at enrollment rates for 5-year-old children since that is the most common kindergarten entry age in the US. Table 1 also provides summary statistics for this data.

Next in order to classify a state-year observation as treated or in the control group, I combined a list of state policy changes on NMEs between 2009 and 2016 from Olive et al. (2018) with tables from Hair, Gruber, and Carly (2020) on the year various exemptions policies came into effect. This was supplemented by information from Immunize.org (formerly the Immunization Action Coalition) and the National Conference of State Legislatures. The CDC data also notes whether religious or philosophical exemptions were permitted in a given state during a given school year.

Furthermore, since parents' decisions to vaccinate their children might be correlated with observable characteristics, I include data from the ACS on race, gender, household income, employment and parental education.

3 Empirical Strategy

In order to explore differential impacts of these policies across time and space, I employ three different empirical strategies. First, I use a pooled difference-in-differences (DD) strategy. Second, I perform an event study with a common time variable based on the year a state's policy is implemented. Finally, I examine each state individually with a DD design.

A DD strategy is useful here to examine the effect of policy changes on school enrollment and use of the exemptions due to the high variation in NME policies across time and space in the US. The equation is as follows:

$$Y_{st} = \alpha + \delta_{rDD} EXEMPT_{st} + \sum_{k=Alaska}^{Wyoming} \beta_k STATE_{ks} + \sum_{j=2009}^{2019} \gamma_j YEAR_{jt} + \theta_{st} + e_{st}$$

Here δ_{rDD} is our coefficient of interest on the interaction term between being treated and in the post-period, $EXEMPT_{st}$. β_k captures state fixed effects and γ_j captures year fixed effects. Y_{st} is our outcome of interest, either percentage of students using NMEs, vaccinated against certain diseases, or enrolled in school. θ_{st} is a vector of control variables including the percentage of the state population that is white, that is male, that has graduated from college, and that is employed. I also control for average income from wages in a state.

The model for the event study is as follows:

$$Y_{ts} = \gamma_s + \lambda_t + \sum_{\tau=-10}^{-1} \gamma_\tau D_{s\tau} + \sum_{\tau=-0}^{8} \delta\tau D_{s\tau} + \theta_{st} + e_{st}$$

Here treatment occurs in year $\tau = 0$. The coefficients of interest are the δ_τ's which show the effect of each additional year on the outcome of interest. γ_s and λ_t are again state and year fixed effects. This model allows us to test that the coefficients in the pre-period are indistinguishable from zero and see if policies have an effect beyond the first year of enactment.

Finally, for each individual state, we use the same model as in the pooled DD except that now the definition of treatment is specific to each state.

In order for the DD strategy to yield a causal interpretation, we must assume that unobserved variables affecting use of NMEs and school enrollment are captured by the state and year fixed effects. The goal is to isolate solely the effect of a change in exemption policy rather than the effect of changing beliefs regarding vaccination, for example. If Oregon generally has a greater number of anti-vaccination parents than California and California makes obtaining NMEs more difficult, then the DD model should control for this difference in anti-vaccination parents through the state fixed effects. Secondly, if in one year there is a significant measles outbreak that captures nationwide attention and causes multiple states to make obtaining NMEs more difficult, then the model controls for this effect through year fixed effects. The standard Ordinary Least Squares (OLS) assumptions must also hold, including that the errors in the regression have conditional mean zero.

To test this assumption, I compare state trends in enrollment and use of vaccination exemptions (where applicable) before policy changes come into effect. By graphing the control and treated states' school enrollment and use of NMEs over time, I will be able to see if there are parallel trends in the pre-period among treatment and control groups. Secondly, I can perform event studies and test placebo effects around moments that may have led to a push in the policy change seen within a state. For example, if a measles outbreak in California puts the NME landscape on lawmakers' radars, then I can see if the instance of the outbreak affects use of exemptions and, if so, control for the presence of an outbreak. Figure 3 showcases some preliminary evidence that this assumption holds. Note that Oregon and Colorado are treated in the same year (2014), so the forest green dashed line also applies for Colorado.

The main problem I encountered was quantifying and encoding the different levels of policy change enacted in each state during the study period. The examined policies are not simply allowing versus prohibiting NMEs but instead a variety of educational modules, discussions with health providers, and signatures required to obtain an NME along with the broader level allowance or prohibition of the NMEs. I made simplifying assumptions on the relationship between different types of policy changes by treating all of the policies that were designed to make obtaining exemptions more difficult as the same. In order to check the robustness of this assumption I also examined the policy changes in each state individually, which showed that the different policies did lead to different outcomes.

4 Results

4.1 Pooled DD

Table 2 shows the results of the pooled DD specification on different outcome variables including each type of required vaccination, the types of exemptions, and enrollment of 5 year old children in school. Here treatment is defined as whether an observation is for a given state in the years of and after the policy change (i.e. all observations in Oregon from 2014 on since that is the year the start requiring educational videos or certificates). The proportion of kindergarteners on any type of exemption decreases by 1.5 percentage points and on non-medical exemptions decreases by 1.1 percentage points. There is no effect on vaccination levels however except for two doses of varicella, which is potentially just a statistical fluke since six different types of vaccines are used as dependent variables. There is also no effect on enrollment. Since the policies like signatures and educational supplements are not much more work for parents, the states with these changes likely weaken the changes in outcomes in states that entirely banned NMEs.

4.2 Event Study

Figure 4 shows the results of the event study on each outcome variable of interest. This also provides further evidence for parallel trends between treatment and control states in the pre-period. For all the variables except any exemption and non-medical exemption, all of the coefficients in the pre-period are indistinguishable from zero. There is no obvious monotonic trend for those two variables though. The most striking results of this specification are for any exemption and NMEs where results last up to seven years after the policy was enacted. For any exemption, the year of the policy changes results in a 0.9 percentage point decrease with further years seeing coefficients ranging from 0.89 to 1.6 percentage points, all statistically significant at the 5% level. Non-medical exemptions are not quite as consistent, but in the first year after treatment, rates decrease 0.98 percentage points. All of the vaccinations also see a 1-2 percentage point increase in take-up in the year of the policy change. However, additional years do not appear to cause further changes. There is no impact on medical exemptions, which is to be expected since the policies targeted NMEs. There is also no impact on general enrollment or public vs. private enrollment.

4.3 Individual States

Tables 3-8 showcase results from individual state DD specifications. They show that the pooled results are driven mostly by California, Colorado, and New York.

4.3.1 California

California has significant increases in all types of vaccination (except 2 doses of varicella since there is only one observation) and significant decreases in any exemptions and NMEs. Notably, though, there is a 0.5 percentage increase in medical exemptions. This aligns with results from Richwine, Dor, and Moghtaderi (2019), which found that Californians substituted non-medical exemptions with medical ones. These results also show that school enrollment for 5-year-olds decreases by 1.2 percentage points. This is one of only two states with this results, so the evidence is not as robust.

4.3.2 Colorado

For Colorado, there are also statistically significant increases in vaccination rates for each vaccine with sufficient data, ranging from 1.3 percentage points for polio to 4.9 for DTaP. Colorado also sees decreases in any exemption and NMEs, but no increase in medical exemptions and no decrease in enrollment.

4.3.3 New York

New York banned religious exemptions in 2018, so I only have two years of treated data. However, there are still significant increases in MMR, Hep. B, and 2 doses Varicella vaccination. These range from 0.7 to 1.2 percentage points. It also sees a 1.4 percentage point decrease in any vaccine exemptions.

4.3.4 Washington

Washington's results are likely confounded by the fact that there are two policy changes within the time period of interest. For simplicity, the earliest one is used to define treatment. However, a stricter policy change only relevant to the MMR vaccine comes later. This may explain why the regression produces simultaneously a decrease in MMR vaccination rates and decreases in any exemption and NMEs. Washington also sees a decrease in enrollment.

4.3.5 Oregon

Oregon's results are statistically insignificant across the board except for a 0.7 percentage point decrease in Hepatitis B. Since the policy change was simply the addition of an educational module, these results align with my priors.

4.3.6 Michigan

Michigan only sees significant changes in exemption rates with all types decreasing. However, none of the vaccination rates change.

5 Conclusion

Overall, there is strong evidence that strict bans of non-medical exemptions make a significant change in vaccination levels within states and have lasting effects. With vaccination rates above 90% on average, changes of 1 or 2 percentage points are evidence that these policy changes are useful at changing the outcomes of people at the margin. One explanation for the usefulness of these policies at the margin is explained in Hair, Gruber, and Carly (2020), who write that when states introduced non-medical exemptions in the early 2000s, the costs of not vaccinating one's children likely dipped below the costs of obtaining vaccinations. Parents facing structural barriers toward obtaining primary healthcare were more likely to not vaccinate their children when exemptions were easier to obtain. Eliminating exemptions in these cases takes away a potentially convenient option for parents who otherwise do not take the time and resources to obtain vaccinations. Similarly, with exemption rates averaging only 2 or 3%, changes around 1 percentage point are high in relative magnitude. However, small barriers like education modules and health practitioner signatures do not appear to be as meaningful. It also appears that there is not a significant portion of the population that holds anti-vaccine beliefs intense enough to pull their children out of school based on these policy changes. However the results in California may call this into question. Additional research on the topic could include using county-level data in California similar to Richwine, Dor, and Moghtaderi (2019) to more robustly explore the question of substituting vaccination with de-enrollment. Overall, as policymakers look to increase vaccination rates and avoid outbreaks, they will need to turn toward these larger legislative moves. Small hurdles do not have enough evidence to support their impact.

References

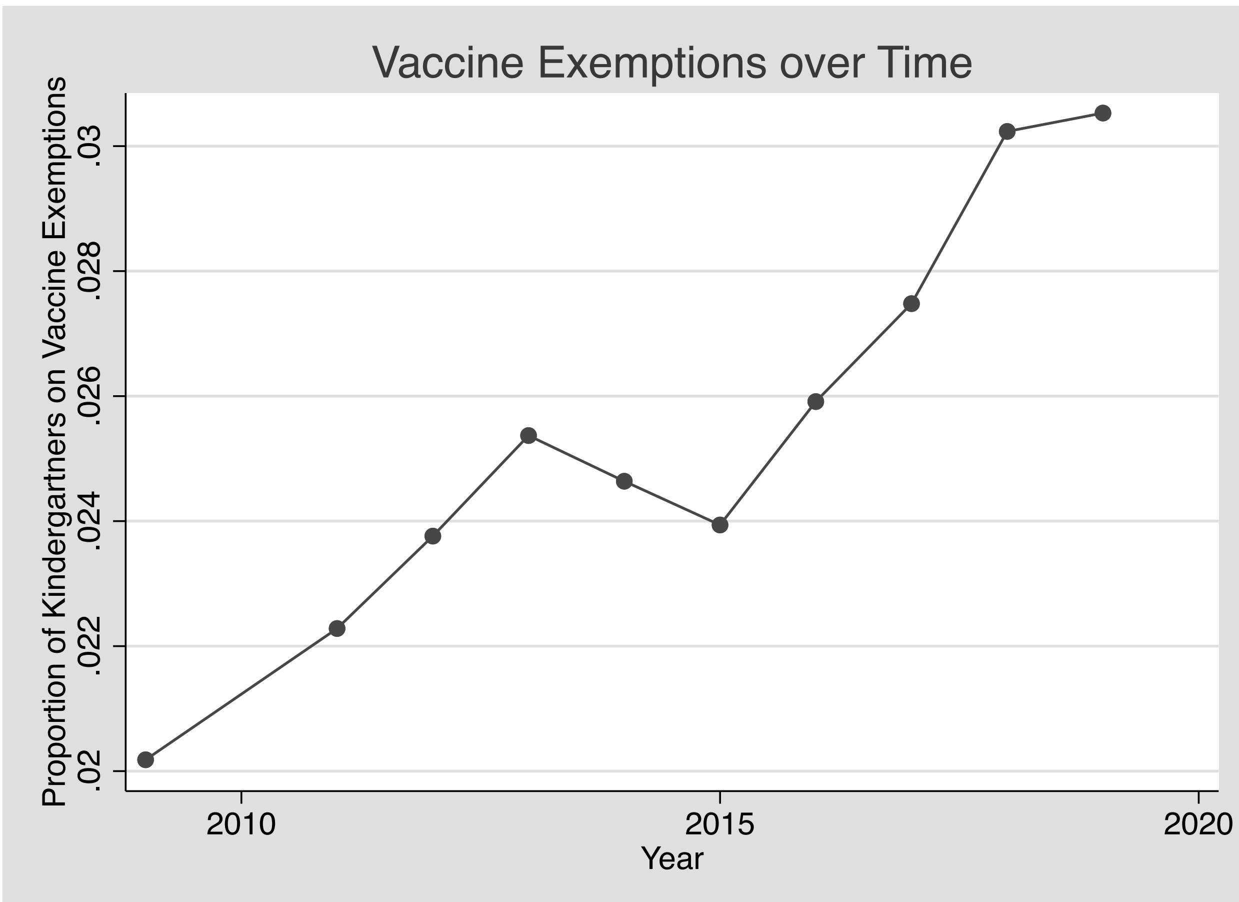
- Bednarczyk, Robert A., et al. "Current Landscape of Nonmedical Vaccination Exemptions in the United States: Impact of Policy Changes." Expert Review of Vaccines, U.S. National Library of Medicine, 2019, https://pubmed.ncbi.nlm.nih.gov/30572729/.
- Blank, Nina R., et al. "Exempting Schoolchildren from Immunizations: States with Few Barriers Had Highest Rates of Nonmedical Exemptions: Health Affairs Journal." Health Affairs, 1 July 2013, https://www.healthaffairs.org/doi/10.1377/hlthaff.2013.0239?url_ver=Z39.88-2003amp;rfr_id=ori
- Erik Skinner, Alise Garcia. "States With Religious and Philosophical Exemptions From School Immunization Requirements." States with Religious and Philosophical Exemptions from School Immunization Requirements, 10 Jan. 2022, https://www.ncsl.org/research/health/school-immunization-exemption-state-laws.aspx.
- Hair, Nicole L., et al. "Personal Belief Exemptions for School-Entry Vaccinations, Vaccination Rates, and Academic Achievement." IZA, Feb. 2020, https://www.iza.org/publications/dp/12978/personal-belief-exemptions-for-school-entry-vaccinations-vaccination-rates-and-academic-achievement.
- Olive, Jacqueline K, et al. "The State of the Antivaccine Movement in the United States: A Focused Examination of Nonmedical Exemptions in States and Counties." PLoS Medicine, Public Library of Science, 12 June 2018, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5997312/.
- Richwine, Chelsea J., et al. "Do Stricter Immunization Laws Improve Coverage? Evidence from the Repeal of Non-Medical Exemptions for School Mandated Vaccines." NBER, 20 May 2019, https://www.nber.org/papers/w25847.

Table 1: Summary Statistics

	Mean	SD	Min	Max	N
Any Exempt.	0.03	0.02	0.00	0.08	476.00
Med. Exempt.	0.00	0.00	0.00	0.02	472.00
NME	0.02	0.02	0.00	0.07	455.00
DTaP	0.94	0.03	0.81	1.00	478.00
Hep. B	0.95	0.03	0.79	1.00	448.00
MMR	0.94	0.03	0.82	1.00	485.00
Polio	0.95	0.03	0.78	1.00	485.00
VAR 1	0.96	0.03	0.76	1.00	119.00
VAR 2	0.93	0.04	0.70	1.00	378.00
Wage and Salary Income	26,421.06	6,171.20	16,564.80	62,390.46	561.00
White Pop.	0.79	0.14	0.24	0.97	561.00
Male Pop.	0.49	0.01	0.46	0.53	561.00
College Educ. Pop.	0.14	0.03	0.08	0.24	561.00
Employed Pop.	0.46	0.04	0.38	0.59	561.00
5YOs in School	0.86	0.05	0.71	1.00	561.00
5 YOs in Public School	0.78	0.06	0.57	0.94	561.00

The first 9 variables are measured as a percentage of kindergarteners with either the relevant exemption or vaccine. Wage and Salary Income is in whole dollar units. The next four variables are dummies with 1 = white, 1 = man, 1 = graduated college, 1 = employed, respectively. 5YOs in school is the percentage of 5 year olds in a state enrolled in any level of school. 5 YOS in Public School is the percentage of the previous group enrolled in a public school as opposed to a private school.

Figure 1: Vaccine Exemptions



Notes: Data were obtained from the CDC.

Figure 2: NME Policies

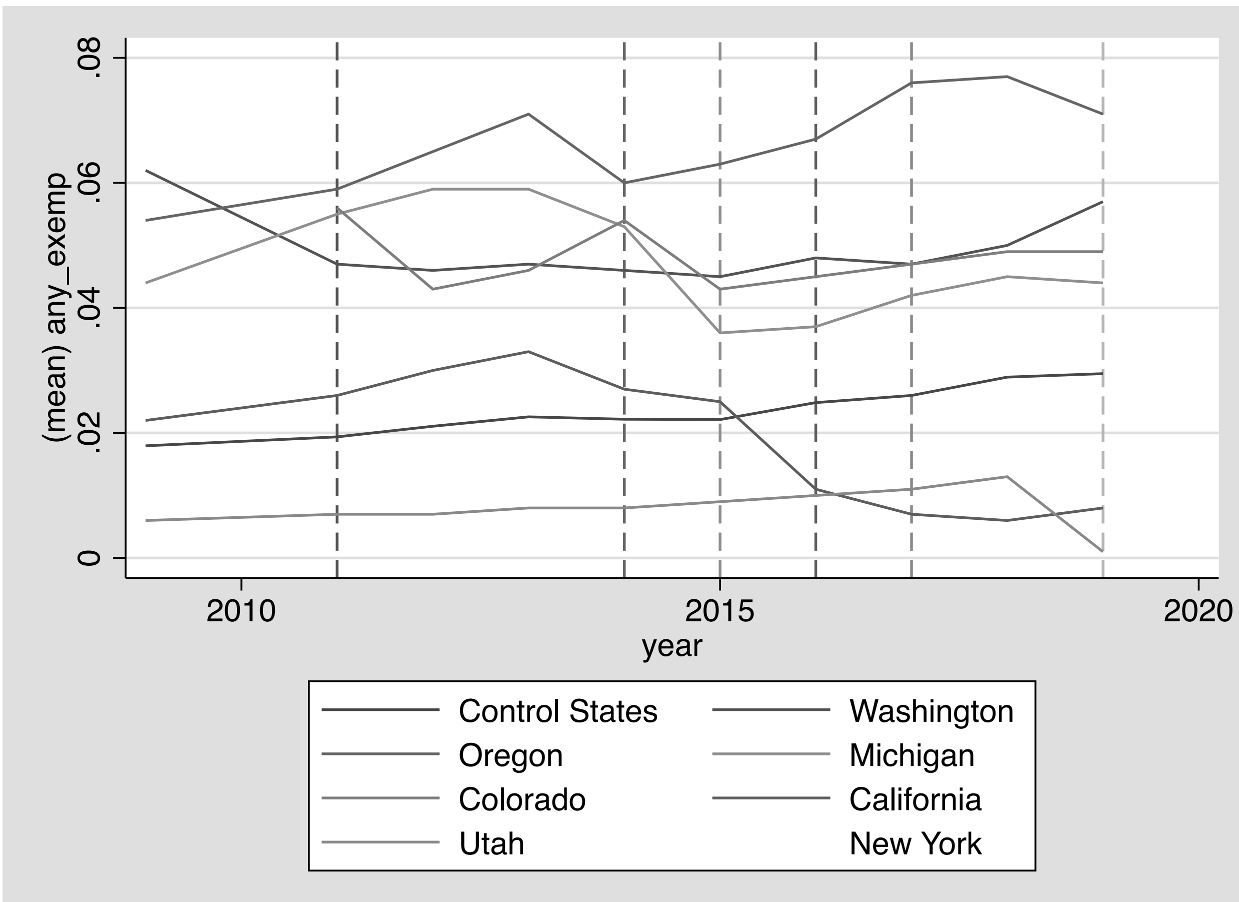
Figure 2 - State policies influencing NME rates

State	Policy Change(s), 2009-2019	Year
New York	Senate Bill 2994: removes the religious exemption for public school immunization requirements.	2019
Washington	House Bill 1638: removes the personal belief exemption for the measles, mumps and rubella vaccine requirement for public schools, private schools and day care centers	2019
California	SB 277: prohibits NMEs from vaccination requirements of public and private schools	2016
Michigan	R 325.176 (update to MDHHS Communicable and Related Diseases Administrative Rules): parents must obtain exemption waivers at county health department, sit through education session, and sign a form before obtaining NMEs	2015
Colorado	HB-1288: schools/childcares must make vaccination and exemption rates publicly available upon request and online education module created; 25-4-902 C.R.S.: requires annual application for vaccine exemption	2014; 2016
Oregon	ORS 433.267: parents must watch an educational video online or get an education certificate at a doctor's office before obtaining NMEs	2014
Washington	Wash. Rev. Code Ann. § 28A.210.080, 90: requires health care practitioner signature before obtaining NMEs	2011

*Sources:

- Originally based on Table S1 from Olive et al. (2018)
- Further years of policy changes added from https://www.ncsl.org/research/health/school-immunization-exemption-state-laws.aspx

Figure 3: Vaccine Exemption Pre-Trends



Notes: Data were obtained from the CDC. Dashed lines indicate year that policy came into effect for state in same color. Colorado and Oregon's policy changes were both in 2014.

Table 2: Pooled DD Estimates

	(1) MMR	(2) DTaP	(3) Hep. B	(4) Polio	(5) VAR 1	(6) VAR 2	(7) Any Ex.	(8) Med. Ex.	(9) NME	(10) Enrollment
Treatment	0.012	0.017	0.009	0.010	0.017	0.012	-0.015	0.001	-0.011	-0.003
	(0.009)	(0.010)	(0.006)	(0.008)	(0.010)	(0.006)	(0.005)	(0.001)	(0.004)	(0.004)
White	0.066	0.196	0.199	0.182	-0.088	-0.209	-0.057	0.011	-0.074	-0.124
	(0.111)	(0.150)	(0.100)	(0.148)	(0.332)	(0.139)	(0.033)	(0.012)	(0.035)	(0.131)
Income	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Male	0.213	0.199	0.258	0.009	-0.523	0.889	0.144	-0.011	0.116	-1.392
	(0.337)	(0.388)	(0.414)	(0.474)	(1.030)	(0.479)	(0.150)	(0.032)	(0.153)	(0.490)
College Edu.	0.326	0.359	0.263	0.090	0.079	0.837	0.199	-0.048	0.237	-0.023
	(0.341)	(0.356)	(0.397)	(0.382)	(0.739)	(0.364)	(0.097)	(0.029)	(0.085)	(0.417)
Employed	-0.148	-0.128	-0.301	-0.051	-0.222	-0.141	-0.068	0.018	-0.082	-0.362
	(0.240)	(0.275)	(0.239)	(0.276)	(0.337)	(0.315)	(0.077)	(0.021)	(0.074)	(0.206)
Observations	465	458	428	465	99	358	456	452	435	561

Figure 4: Event Study

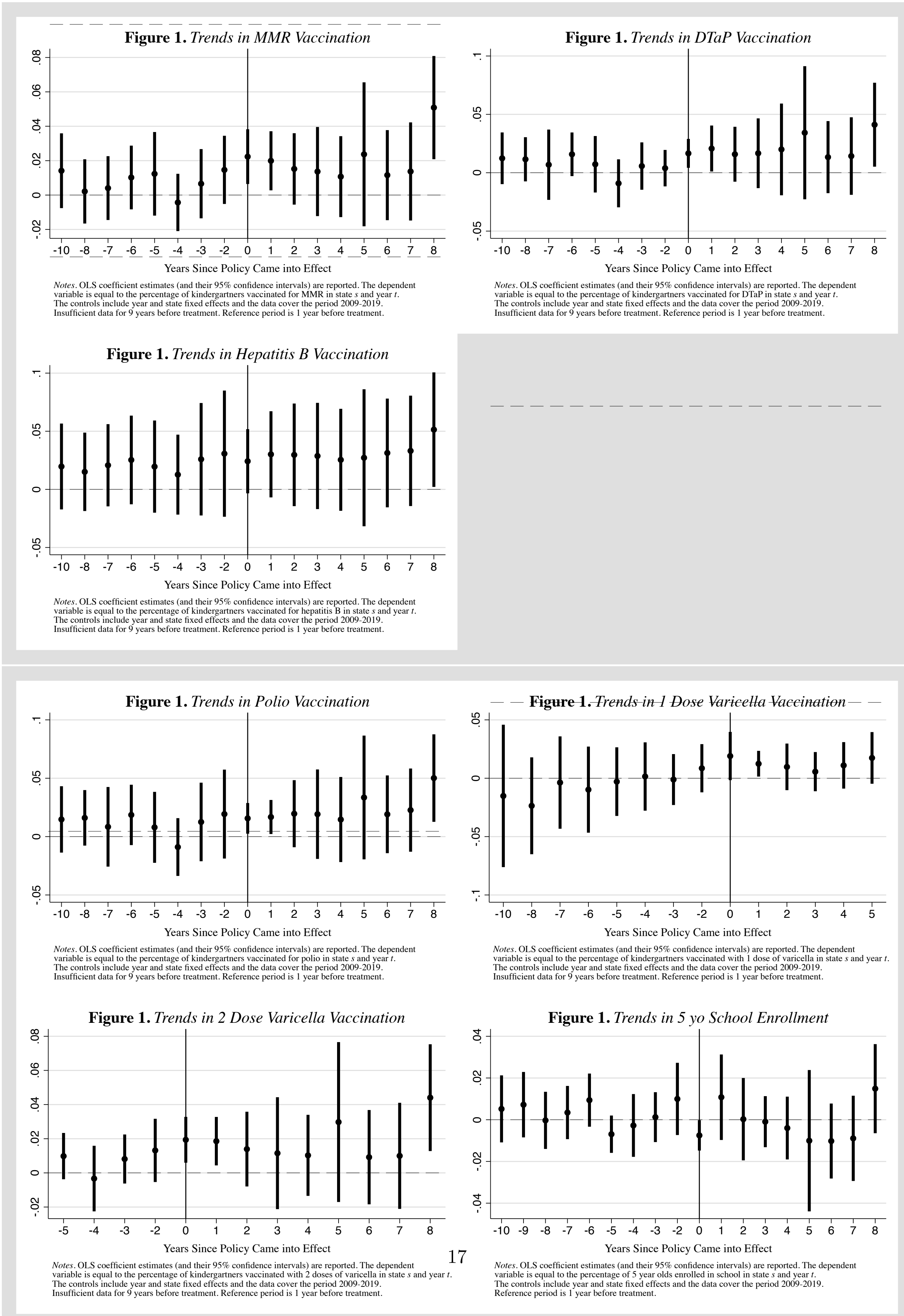
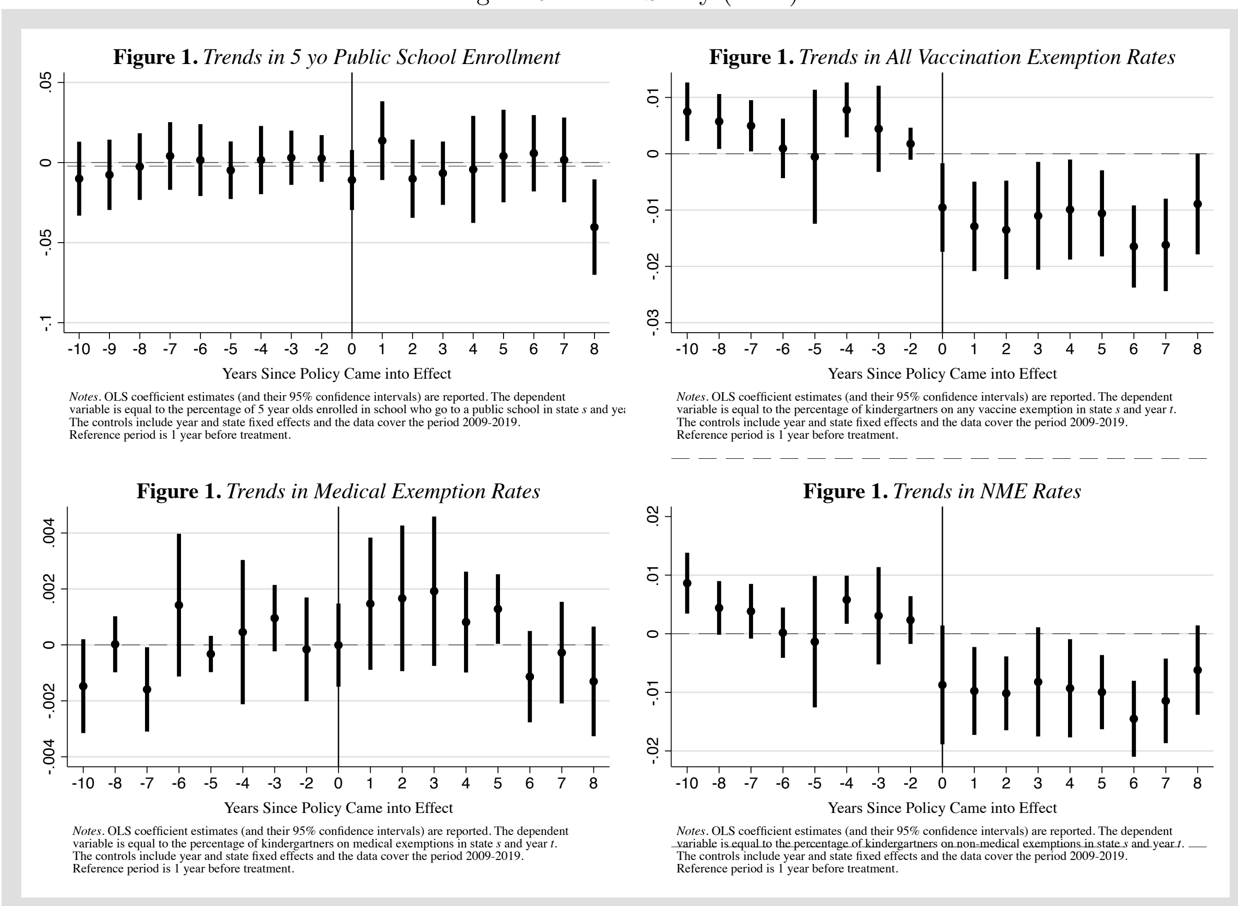


Figure 5: Event Study (cont.)



Notes: Data were obtained from the CDC. 0 indicates the control observations and 1 indicates the treated observations.

Table 3: California

	(1) MMR	(2) DTaP	(3) Hep. B	(4) Polio	(5) VAR 1	(6) VAR 2	(7) Any Ex.	(8) Med. Ex.	(9) NME	(10) Enrollment
Treatment	0.037	0.041	0.030	0.044	0.030	0.000	-0.027	0.005	-0.025	-0.012
	(0.005)	(0.006)	(0.004)	(0.006)	(0.007)	(.)	(0.002)	(0.000)	(0.001)	(0.005)
White	0.093	0.215	0.225	0.225	-0.050	-0.214	-0.060	0.017	-0.069	-0.133
	(0.106)	(0.145)	(0.098)	(0.139)	(0.345)	(0.136)	(0.032)	(0.011)	(0.038)	(0.132)
Income	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000	0.000	-0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Male	0.281	0.266	0.317	0.094	-0.564	0.909	0.106	0.000	0.104	-1.413
	(0.335)	(0.385)	(0.413)	(0.471)	(1.013)	(0.477)	(0.149)	(0.029)	(0.153)	(0.493)
College Edu.	0.384	0.424	0.315	0.156	0.261	0.844	0.161	-0.041	0.214	-0.036
	(0.334)	(0.349)	(0.391)	(0.374)	(0.691)	(0.357)	(0.093)	(0.028)	(0.086)	(0.418)
Employed	-0.153	-0.113	-0.310	-0.074	-0.236	-0.115	-0.090	0.014	-0.107	-0.354
	(0.232)	(0.265)	(0.234)	(0.265)	(0.333)	(0.314)	(0.089)	(0.022)	(0.083)	(0.208)
Observations	465	458	428	465	99	358	456	452	435	561

Table 4: Colorado

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	MMR	DTaP	Hep. B	Polio	VAR 1	VAR 2	Any Ex.	Med. Ex.	NME	Enrollment
Treatment	0.029	0.049	0.016	0.013	0.000	0.023	-0.006	-0.000	-0.002	0.003
	(0.004)	(0.004)	(0.003)	(0.004)	(.)	(0.005)	(0.001)	(0.000)	(0.001)	(0.004)
White	0.023	0.133	0.167	0.150	-0.258	-0.221	-0.020	0.009	-0.055	-0.118
	(0.118)	(0.160)	(0.104)	(0.159)	(0.320)	(0.136)	(0.051)	(0.014)	(0.042)	(0.131)
Income	-0.000	-0.000	0.000	-0.000	-0.000	-0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Male	0.169	0.130	0.228	-0.020	-0.901	0.886	0.182	-0.014	0.119	-1.391
	(0.342)	(0.395)	(0.415)	(0.477)	(0.982)	(0.478)	(0.163)	(0.032)	(0.152)	(0.490)
College Edu.	0.309	0.331	0.253	0.084	0.175	0.819	0.199	-0.047	0.218	-0.030
	(0.339)	(0.352)	(0.395)	(0.380)	(0.735)	(0.361)	(0.103)	(0.030)	(0.087)	(0.419)
Employed	-0.103	-0.066	-0.265	-0.007	0.002	-0.134	-0.130	0.022	-0.123	-0.373
	(0.240)	(0.276)	(0.239)	(0.278)	(0.372)	(0.315)	(0.098)	(0.023)	(0.086)	(0.205)
Observations	465	458	428	465	99	358	456	452	435	561

Table 5: Michigan

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	MMR	DTaP	Hep. B	Polio	VAR 1	VAR 2	Any Ex.	Med. Ex.	NME	Enrollment
Treatment	-0.004	-0.002	0.002	-0.000	0.000	0.009	-0.017	-0.004	-0.015	-0.001
	(0.005)	(0.005)	(0.004)	(0.005)	(.)	(0.005)	(0.002)	(0.000)	(0.001)	(0.003)
White	0.031	0.147	0.173	0.153	-0.258	-0.212	-0.027	0.008	-0.060	-0.118
	(0.119)	(0.163)	(0.105)	(0.160)	(0.320)	(0.137)	(0.051)	(0.014)	(0.041)	(0.131)
Income	-0.000	-0.000	0.000	-0.000	-0.000	-0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Male	0.201	0.174	0.239	-0.009	-0.901	0.894	0.215	-0.006	0.150	-1.386
	(0.343)	(0.387)	(0.414)	(0.475)	(0.982)	(0.479)	(0.163)	(0.030)	(0.153)	(0.490)
College Edu.	0.329	0.367	0.267	0.093	0.175	0.853	0.193	-0.048	0.213	-0.027
	(0.337)	(0.355)	(0.394)	(0.380)	(0.735)	(0.357)	(0.102)	(0.030)	(0.085)	(0.417)
Employed	-0.079	-0.039	-0.262	-0.001	0.002	-0.142	-0.089	0.032	-0.084	-0.371
	(0.250)	(0.286)	(0.247)	(0.288)	(0.372)	(0.325)	(0.093)	(0.021)	(0.080)	(0.210)
Observations	465	458	428	465	99	358	456	452	435	561

Table 6: New York

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	MMR	DTaP	Hep. B	Polio	VAR 1	VAR 2	Any Ex.	Med. Ex.	NME	Enrollment
Treatment	0.011	0.005	0.007	0.000	0.000	0.012	-0.014	0.000	0.000	0.003
	(0.004)	(0.005)	(0.003)	(0.004)	(.)	(0.003)	(0.002)	(0.000)	(.)	(0.004)
White	0.032	0.147	0.172	0.153	-0.258	-0.218	-0.021	0.009	-0.055	-0.117
	(0.118)	(0.163)	(0.105)	(0.159)	(0.320)	(0.136)	(0.051)	(0.014)	(0.042)	(0.131)
Income	-0.000	-0.000	0.000	-0.000	-0.000	-0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Male	0.197	0.172	0.245	-0.009	-0.901	0.916	0.171	-0.014	0.117	-1.387
	(0.340)	(0.388)	(0.413)	(0.475)	(0.982)	(0.478)	(0.162)	(0.032)	(0.152)	(0.490)
College Edu.	0.336	0.369	0.271	0.093	0.175	0.854	0.187	-0.047	0.216	-0.026
	(0.338)	(0.355)	(0.394)	(0.380)	(0.735)	(0.357)	(0.103)	(0.030)	(0.086)	(0.417)
Employed	-0.088	-0.043	-0.257	-0.001	0.002	-0.111	-0.134	0.022	-0.123	-0.373
	(0.240)	(0.276)	(0.238)	(0.278)	(0.372)	(0.315)	(0.098)	(0.023)	(0.086)	(0.206)
Observations	465	458	428	465	99	358	456	452	435	561

Table 7: Oregon

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	MMR	DTaP	Hep. B	Polio	VAR 1	VAR 2	Any Ex.	Med. Ex.	NME	Enrollment
Treatment	-0.005	-0.007	-0.007	-0.004	-0.000	0.000	0.003	-0.000	0.000	0.005
	(0.004)	(0.005)	(0.003)	(0.005)	(0.005)	(.)	(0.002)	(0.000)	(0.001)	(0.003)
White	0.028	0.142	0.168	0.150	-0.258	-0.214	-0.019	0.009	-0.055	-0.114
	(0.120)	(0.166)	(0.106)	(0.162)	(0.324)	(0.136)	(0.052)	(0.014)	(0.043)	(0.131)
Income	-0.000	-0.000	0.000	-0.000	-0.000	-0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Male	0.185	0.161	0.233	-0.015	-0.905	0.909	0.182	-0.014	0.118	-1.384
	(0.340)	(0.389)	(0.413)	(0.476)	(1.013)	(0.477)	(0.163)	(0.032)	(0.153)	(0.490)
College Edu.	0.335	0.374	0.275	0.097	0.178	0.844	0.192	-0.047	0.216	-0.032
	(0.339)	(0.357)	(0.395)	(0.381)	(0.754)	(0.357)	(0.103)	(0.030)	(0.087)	(0.418)
Employed	-0.084	-0.036	-0.250	0.004	0.003	-0.115	-0.135	0.022	-0.124	-0.378
	(0.243)	(0.280)	(0.240)	(0.281)	(0.385)	(0.314)	(0.099)	(0.023)	(0.087)	(0.206)
Observations	465	458	428	465	99	358	456	452	435	561

Table 8: Washington

	(1) MMR	(2) DTaP	(3) Hep. B	(4) Polio	(5) VAR 1	(6) VAR 2	(7) Any Ex.	(8) Med. Ex.	(9) NME	(10) Enrollment
Treatment	-0.011	-0.004	-0.006	-0.012	0.000	-0.008	-0.022	0.006	-0.027	-0.015
	(0.005)	(0.006)	(0.006)	(0.006)	(.)	(0.008)	(0.002)	(0.001)	(0.002)	(0.004)
White	0.028	0.146	0.170	0.149	-0.258	-0.221	-0.029	0.011	-0.065	-0.123
	(0.119)	(0.164)	(0.106)	(0.161)	(0.320)	(0.139)	(0.049)	(0.014)	(0.038)	(0.131)
Income	-0.000	-0.000	0.000	-0.000	-0.000	-0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Male	0.186	0.168	0.240	-0.016	-0.901	0.906	0.164	-0.010	0.099	-1.407
	(0.339)	(0.387)	(0.412)	(0.474)	(0.982)	(0.477)	(0.161)	(0.032)	(0.149)	(0.490)
College Edu.	0.337	0.370	0.272	0.101	0.175	0.854	0.210	-0.052	0.235	-0.025
	(0.337)	(0.356)	(0.395)	(0.379)	(0.735)	(0.360)	(0.102)	(0.028)	(0.085)	(0.416)
Employed	-0.101	-0.048	-0.264	-0.013	0.002	-0.125	-0.153	0.029	-0.151	-0.385
	(0.244)	(0.279)	(0.243)	(0.281)	(0.372)	(0.322)	(0.097)	(0.023)	(0.084)	(0.207)
Observations	465	458	428	465	99	358	456	452	435	561

The Effect of School Voucher Programs on Enrollment and Private School Diversity

Diana Degnan

Massachusetts Institute of Technology

Department of Economics

ABSTRACT

School vouchers are meant to make private schools more accessible to low-income and minority students. Do they? This paper looks at the effect of school vouchers on the demographic mix of private school enrollment. I use an event study model with a matching specification. This allows me to analyze the effect of school vouchers on private school demographics. School voucher programs appear to change the race and gender mix of private school enrollment little. On average, school voucher programs are associated with a small but significant *decrease* in hispanic private school enrollment post-implementation. However, school voucher programs with specific low-income or poor-performing public school requirements are associated with a small but significant *increase* in hispanic private school enrollment. The results indicate that programs targeting low-income students or students from poor-performing public schools may be more effective at increasing minority representation in private schools.

1 Introduction

Private schools are widely regarded to provide higher quality education than public schools (Pierce, 2021). For better educational opportunities, many families choose to move their children to private school. However, due to their cost, private schools are particularly inaccessible for low-income students. School voucher programs provide an opportunity to enable school choice for low-income and minority students. But do they actually work? Voucher programs have generated significant political debate due to concerns that they increase private school enrollment at the expense of public schools and primarily benefit the white middle class (Viteritti, 1996). Very little is known about the effect of school vouchers on enrollment and demographics. Existing literature has focused on the educational outcomes of school voucher recipients instead. I evaluate the enrollment and demographic effects of school vouchers by answering two questions: (1) How do school voucher programs affect enrollment? (2) Do school voucher programs encourage demographic diversification among private school students?

School vouchers are educational programs administered at the state-level that aim to encourage school choice. School vouchers transfer federal funds to eligible recipients who may use these funds to pay for private school tuition, online classes, or private tutoring. Since the mid 2000s school voucher programs have taken shape in numerous states to a varying degree and scale. To determine the effect of school vouchers on enrollment and demographics, I estimate an event study model using census data from the American Community Survey 2000-2019.

My research design compares urban districts in school voucher states with urban districts in nearby non-voucher states. I analyzed the effect of school voucher programs on public school enrollment, private school enrollment, and the demographic mix of private school students. I conducted my analysis over 5 years post-voucher program. I evaluate effects relative to a base year of one year prior to school voucher program implementation. I find that on average, school voucher programs have no significant effect on school enrollment or black private school diversity. However, school voucher programs are associated with a small but significant *decrease* in hispanic private school students by approximately 2-3% nearly every year post-voucher program. These results change when comparing voucher programs with low-income or poor-performing public school requirements to programs without these requirements. I find that school voucher programs with low income requirements are associated with a small but significant *increase* in hispanic private school students by 1-3% every year post-voucher program.

My results imply that school voucher programs change private school enrollment little. School voucher programs do not appear to be increasing private school enrollment at the expense of public schools, nor do they appear to be encouraging more minority students to attend private school. School voucher programs that limit eligibility to low-income students or students from poor-performing public schools are more effective at increasing private school diversity, particularly for hispanic students. This suggests that school voucher programs without these requirements may disproportionately benefit white students, possibly at the expense of other racial groups.

The rest of the paper will proceed as follows. Section 1 provides a description of data sources and manipulations. Section 2 outlines my event study and matching specification. Section 4 interprets the results of my regression model and discusses the implications in both statistical and real-world significance. Section 5 offers a conclusion and speaks to the limitations of my study. Section 6 consists of all regression tables and figures, as well as an appendix of additional analysis.

2 Data

My data source is American Community Survey (ACS) Census data from 2000 to 2019. The survey for the year 2000 provides a 1-in-750 national random sample of the United States population. For 2001, 2002, 2003, and 2004, data consists of a 1-in-232, 1-in-261, 1-in-236, and 1-in-239 weighted national random sample, respectively. For the years 2005 to 2019, the data consists of a 1-in-100 national random sample. Data in the sample is at the individual level. The ACS data provides necessary information on the geographic, demographic, and educational composition of primary and secondary school-aged children in the US.

In addition to US Census data, I compiled a qualitative dataset on United States school voucher programs from state-level departments of education and the nonprofit EdChoice. I used both their annual reports and their "School Choice in American Dashboard" to create a reference set of school voucher programs in the United States by state, year implemented, size, and eligibility requirements. I only included school voucher programs implemented between 2000 and 2019. I excluded from study any school voucher programs too small to have an effect on overall private school enrollment and demographics. I defined programs of such small size as those with less than 600 participants after the fifth year of implementation.

My analysis focused on school voucher programs, ignoring tax credit scholarships and education savings accounts (ESAs). In narrowing the focus of my study to school vouchers exclusively, I do not assess the impacts of two large school choice programs that have garnered significant academic and political attention in recent years: Florida and Arizona's ESA programs. These specifications left me with 8 school voucher programs in the states of Indiana, Wisconsin, Ohio, North Carolina, Maryland, Louisiana, Georgia, and the District of Columbia.

Table 1 displays summary statistics for my key outcome variables clustered by state. The sample is restricted to individuals ages 5-18 within the United States. The Not Enrolled, Enrolled, Attending Public School, and Attending Private School variables give the portion of the sample not enrolled in school, enrolled in school, attending public school, and attending private school, respectively. The Black, Female, and Hispanic variables give the portion of the sample identifying as black, female, or hispanic, respectively. Private School Students, Black (indicator) gives the portion of black private school students. Private School Students, Female (indicator) gives the portion of female private school students. Private School Students, Hispanic (indicator) gives the portion of hispanic private school students. Mixed race students are not included. An exception is for black and hispanic private school students. These students are included in both Private School Students, Black (indicator) and Private School Students, Hispanic (indicator).

Private School Students, Black (indicator), Private School Students, Female (indicator), and Private School Students, Hispanic (indicator) give the demographic mix of private school students. I am interested in whether school vouchers increase racial diversity among private school students, so these are my outcome variables of interest. My regression interpretations therefore speak to the effect of school vouchers on private school diversity. Most school voucher programs target low-income students or students from poor-performing public school districts. Empirical findings have established that minority populations, particularly blacks and hispanics, are more likely to be from high-poverty or poor-performing public school districts than white students. If school vouchers enable school choice for minority populations, I expect higher rates of minority representation in private schools after school vouchers are implemented. Relatively larger portions of blacks and hispanics should be receiving school vouchers, thereby becoming eligible for private school. As a result, private schools should be admitting more black and hispanic students compared to pre-voucher levels. I expect to see this change manifested in the demographic mix of private school students.

This data source has two key limitations. First, ACS does not contain a variable indicating school voucher recipiency. Therefore, this study cannot directly track the enrollment and demographic mix of actual school voucher recipients. Second, geographic specification is limited for early timeframes in certain states. My study uses a matching specification to link the metropolitan area of a school voucher state to a comparable metropolitan area of a nearby non-voucher state. The details of this matching strategy are explained in depth in the Empirical Methods section. In five of the eight voucher states, the metropolitan area of a major city in a school voucher state is matched with the metropolitan area of a major city in a neighboring non-voucher state. However, there are three exceptions: DC, Wisconsin, and Ohio. The ACS has no smaller geographic units within DC. DC is therefore matched with Delaware, which is of similar population size. Wisconsin faces a similar problem. Wisconsin's school voucher program extends to all students outside of the Milwaukee school system. Milwaukee has its own school voucher program that began in 1990. The ACS only contains metro-area geographic data for the city of Milwaukee within Wisconsin. To circumvent this problem, I matched Dane County, Wisconsin, which contains the city of Madison, with Kent County, Michigan, which contains the city of Grand Rapids. This is the only match for which county level data is used. The last exception is Ohio. Ohio's school voucher program was initiated in 2006. ACS data only has geographic units smaller than a state available for the years since 2005. In order to analyze prior trends in Ohio, I consider Ohio on the statewide level. Ohio is paired with Pennsylvania due to their proximity and similar population size. I anticipated that the inclusion of state-level data for Ohio and DC may have a significant effect on my results. To address this, I run my analysis on school voucher programs a second time with Ohio and DC excluded from study. I do not find dramatically different results between my analyses. This suggests that the inclusion of Ohio and DC does not bias my results significantly. I explore this fully in the Appendix.

3 Empirical Method

To test the effects of school vouchers on school enrollment and private school diversity, I employ an event-study empirical strategy with a matching specification. To construct my matches I paired a metropolitan area within a school voucher state with a metropolitan area of similar population size in a nearby non-school voucher state. For example, Raleigh, North Carolina (a voucher state) is matched with Charleston, South Carolina (a non-voucher state). I do this by creating a region variable with a value for the 16 metropolitan areas, states, and counties comprising the matched pairs. I create a separate variable to indicate the 8 areas with school voucher programs that comprise the treated group. I cluster at the region level and control for region-fixed effects. My dependent variables are public school enrollment, private school enrollment, and the mix of black, female, and hispanic private school students. I also group the matched pairs into two categories defined by the eligibility requirements of the voucher programs. I create one variable containing only the areas with voucher programs that target low income students or students from poor-performing public school districts. I create a second variable containing only the areas with voucher programs that require students to have been enrolled in public school the year prior. Using the matches, I conduct an event-study regression model with a matching specification. I conduct my analysis over the 3 years prior and 5 years after a school voucher program has been implemented. This type of analysis is seen in Jager's 2016 paper. Jager employs a dynamic difference-in-differences model and matched sampling procedure (see Section 4.2 and Section 4.4, particularly Equation 12). A very similar method is seen in Sachs' 2019 paper (see Section 5.1, particularly Equations 2 and 4). My regression equation for the event-study model is:

$$\mathrm{Y}_{\mathrm{i,t}} = \beta_0 + \alpha_{\mathrm{t}} + \gamma_{\mathrm{i}} + \sum_{j=-3}^{j=5} \beta_{1,\mathrm{j}}\, \mathrm{D}_{\mathrm{j}} + \sum_{j=-3}^{j=5} \beta_{2,\mathrm{j}}\, \mathrm{D}_{\mathrm{j}}^{\mathrm{Treated}} + \epsilon_{i,t} \qquad (1)$$

Y is the outcome of interest. α_{t} represents time fixed-effects. γ_{i} represents region-fixed effects. D_{j} is a dummy variable for each year of study. $\mathrm{D}_{\mathrm{j}}^{\mathrm{Treated}}$ represents the group of areas with school voucher programs in each year. $\beta_{2,\mathrm{j}}$ is the coefficient of interest. It indicates the average effect of a school voucher program on Y relative to the base year. The base year is set as one year prior to a school voucher program's start. The regression residual is represented by $\epsilon_{i,t}$.

The event study model and matching specification I employ requires the satisfaction of two key assumptions for validity. The first assumption I must satisfy is the parallel trends assumption. I must verify that there are no relevant time-varying differences between the metropolitan areas of the school voucher states and their matched counterparts. To test for the parallel trends assumption, I analyze enrollment and the demographic mix of private school students 3 years prior to the implementation of a school voucher program. I find no significant prior trends across any outcome variable. Therefore, I assume the parallel trends assumption is satisfied.

The second assumption required for validity is the good as random assumption. This requires that a state's decision to implement a school voucher program is essentially random. However, school voucher programs are the result of significant economic, political, and social considerations by state officials. This is a key limitation of my event study model. I am unable to fully test for or satisfy the good as random assumption. However, I have taken measures to qualitatively account for possible violations of the good as random assumption. The school voucher programs under study exhibit significant variation. They were not all initiated under a common political party, at either the state or national level. The school voucher states are also not clustered in a specific geographic region. In addition, I conducted a review into news articles and state government publications on the creation of each state's voucher program. I did not identify any common non-random justifications for implementing a school voucher program that could serve as a confounding variable (Hsu, 2006). I also tested a hypothesis that funding for public schools may have increased within school voucher states around the same time as school voucher program implementation. An increase in public school funding would be evidenced by an increase in property taxes. I reviewed the property tax rates of each school voucher state. I saw no sudden increase in property taxes around the time of school voucher program implementation (Urban Institute, 2020 & Hanson, 2022). Yet, my empirical strategy remains limited by its inability to fully satisfy the good as random assumption.

4 Estimation

4.1 Grouped Event Study Model:

Table 2 displays regression coefficients from the event-study model and matching specification. Each coefficient gives the average effect of a school voucher program on the outcome variable in a given year since the voucher program began. Effects are relative to the base year. The base year is one year prior to a school voucher program's start. One year pre-voucher program is omitted from the regressions, as it serves as the base year. Each column indicates a separate regression with a separate outcome of interest. Pub. School represents total public school enrollment. Priv. School represents total private school enrollment. Black Priv. School represents black private school students. Female Priv. School represents female private school students. Hispanic Priv. School represents hispanic private school students. Table 2 indicates that on average school voucher programs had no significant effect on enrollment or the black and female mix of private school students. This finding is consistent across each of the 5 years post-voucher program. School voucher programs are associated with a small but significant decrease in hispanic private school enrollment. The percentage of hispanic private school students decreases by 2-3% in nearly every year post-voucher program.

Figure 1 shows event study plots for the regressions in Table 2. Panel (a) of Figure 1 refers to the public school enrollment outcome variable. Panel (b) of Figure 1 refers to the private school enrollment outcome variable. Panel (c) of Figure 1 refers to the black private school students outcome variable. Panel (d) of Figure 1 refers to the female private school students outcome variable. Panel (e) of Figure 1 refers to the hispanic private school students outcome variable. The coefficients are indexed from -3 to 5. Negative values represent years pre-voucher program. Positive values represent years post-voucher program. 0 represents the voucher program's start year. The coefficients measure the average change in each outcome variable for all school voucher regions compared to all matched non-school voucher regions. Coefficients are relative to a base year of 1 year prior to treatment. The coefficient for t=-1 is omitted to serve as the base year. 95% confidence intervals are illustrated by red error bars. No outcome variable shows a consistent and statistically significant prior trend. Therefore, I am reasonably certain prior trends do not bias the results. Figure 1 conveys the same findings as Table 2.

Despite enabling school choice and providing a pathway for students to attend private school, my results find that school vouchers have no significant effect on overall enrollment. School voucher programs do not appear to increase private school enrollment at the expense of public schools. My results suggest that school voucher programs are not large enough to significantly affect overall enrollment. I also find school vouchers to have a negligible effect on female private school enrollment. This result is expected. It is unlikely that school vouchers are given to dramatically more female than male students. However, the lack of increase in black private school enrollment is more surprising. Most school voucher programs target low-income students or students from poor performing public school districts. Empirical research shows that black students are far more likely to attend high-poverty and low-performing public schools than other racial groups (Jordan 2014). Therefore, I expect that higher relative portions of black students are eligible for school vouchers. Despite this, my results find that school voucher programs do not significantly affect black private school enrollment. This may indicate that black students do not receive a proportional share of school vouchers. It could also indicate that black populations do not have equal access to school voucher information or applications.

School voucher programs are associated with a small but significant decrease in the level of private school hispanic students nearly every year post-voucher program. The coefficients on private school hispanic students are significant at the 5% level. On average, private school hispanic enrollment decreased 2-3% each of the five years after a school voucher program was implemented. Similar to black students, empirical evidence has found that hispanic students are much more likely to be from high-poverty or poor-performing public school districts than their white counterparts (Carnoy & Garcia, 2017). Therefore, I expect that higher relative portions of hispanic students are eligible for school vouchers. As a result, I would expect to see higher percentages of hispanic students enrolled in private school post-voucher program. My empirical findings contradict this hypothesis. I find an association between school voucher programs and lower levels of hispanic private school students. These results suggest that hispanic populations may not receive a proportional share of school vouchers or have equal access to voucher applications. Furthermore, it could indicate that privately enrolled hispanic students are losing their share of private school enrollment to an influx of new non-hispanic school voucher recipients.

4.2 Event Study Model By Voucher Requirements:

Table 3 and Table 4 display regression coefficients from my event study model with school voucher states grouped by program requirements. The regression coefficients in Table 3 give the average effect of school vouchers with eligibility limited to low-income students or students from poor-performing public school districts on each outcome variable. These effects are compared to school voucher programs without such requirements. Coefficients are estimated relative to a base year of one year pre-voucher program. School voucher programs define low-income as households within 300% of the federal poverty level. Poor-performing public schools are defined as school districts with a D or F rating from the US Department of Education. The regressions from Table 3 find that school vouchers do not have a significant effect on enrollment or the demographic mix of black and female private school students. I regard the significant coefficients on public school enrollment as spurious due to their inconsistency and extremely small magnitude. All significant coefficients on public school enrollment indicate a <1% change in the enrollment level. School voucher programs with low-income or poor-performing school requirements are associated with a small but significant increase in private school hispanic enrollment. Private school hispanic enrollment increased by 1-3% each year post-voucher program. This result is the opposite of the trend seen in the event study model for all school voucher programs, as shown in Table 2 and Figure 1. These findings imply that school voucher programs with low-income or poor-performing public school requirements are more effective at increasing racial diversity among private school students, specifically for hispanics. These results are consistent with the demographic characteristics of hispanic students. Hispanic students are much more likely than white or Asian students to be from high-poverty or poor-performing public school districts (Carnoy & Garcia, 2017). School voucher programs without low-income or poor-performing public school requirements may disproportionately benefit white students, possibly at the expense of other minority groups.

Figure 2 shows event study plots for the regressions in Table 3. Panel (f) of Figure 2 refers to the public school enrollment outcome variable. Panel (g) of Figure 2 refers to the private school enrollment outcome variable. Panel (h) of Figure 2 refers to the black private school students outcome variable. Panel (i) of Figure 2 refers to the female private school students outcome variable. Panel (j) of Figure 2 refers to the hispanic private school students outcome variable. Coefficients are indexed -3 to 5, as in Figure 1. It conveys the same results as Table 3.

The regression coefficients in Table 4 give the average effect of school vouchers with eligibility limited to students enrolled in public school the year prior on each outcome variable. These effects are compared to school voucher programs without such requirements. Coefficients are estimated relative to a base year of one year pre-voucher program. Table 4 shows that school vouchers with prior public school enrollment requirements have no significant effect on enrollment or the demographic composition of private school students. I regard the significant coefficients on female private school enrollment as spurious due to their small magnitude and inconsistency. Although not statistically significant, Table 4 shows that the coefficients on black and hispanic private school enrollment are positive when controlling for programs with prior public school enrollment requirements. This suggests that such voucher programs may have more racially diverse recipients. These results also imply that prior public school enrollment requirements may be somewhat effective at encouraging minority enrollment in private school. However, I cannot make these conclusions with statistical certainty.

School vouchers with prior public school enrollment requirements are not associated with a significant change in private school enrollment. I expected school voucher programs with prior public school requirements to decrease public enrollment and increase private enrollment. However, the results of Table 4 do not support this hypothesis. In fact, although not statistically significant, the coefficients on public school enrollment are positive and the coefficients on private school enrollment are negative post-voucher program. These results suggest that prior public school enrollment requirements do not significantly affect private or public enrollment overall.

Figure 3 shows event study plots for the regressions in Table 4. Panel (k) of Figure 3 refers to the public school enrollment outcome variable. Panel (l) of Figure 3 refers to the private school enrollment outcome variable. Panel (m) of Figure 3 refers to the black private school students outcome variable. Panel (n) of Figure 3 refers to the female private school students outcome variable. Panel (o) of Figure 3 refers to the hispanic private school students outcome variable. Coefficients are indexed -3 to 5, as in Figure 1. It conveys the same results as Table 4.

5 Conclusion

I use an event study model with a matching specification to analyze the effect of school vouchers on enrollment and private school diversity. I employ ACS census data from 2000-2019 for my analysis. I match a metropolitan area in a voucher state with a major metropolitan area in a non-school voucher state. I analyze the effects of school vouchers 3 years prior and 5 years post-voucher implementation. Coefficients are estimated relative to a base year of one year prior to school voucher program implementation. My results find that school vouchers have little effect on enrollment or the demographic mix of black and female private school students. I find that school vouchers are associated with a small but significant *decrease* in private school hispanic enrollment by approximately 2-3% nearly every year post-school voucher program. However, school voucher programs with low-income or poor-performing public school requirements are associated with a small but significant *increase* in private school hispanic enrollment by 1-3% each year post-school voucher program. As a result, I do not find school vouchers effective at increasing enrollment or diversifying private school students. School vouchers do not appear to increase private school enrollment at the expense of public schools. Voucher programs absent low-income requirements may disproportionately benefit white students, possibly at the expense of other minority groups.

However, the results of my study are limited by several key factors. First, metropolitan area-level data is unavailable prior to 2005. Therefore, it is not possible to see prior trends greater than three years pre-voucher program in several states. As a result, I cannot be sure the parallel trends assumption is fully satisfied. Similarly, I do not study several recent school voucher programs due to lack of data. In addition, census data may be too broad to fully capture the effects of school voucher programs. Even the largest school voucher programs affect relatively few students compared to the US population. Finally, unknown omitted variables may exist. Further research should be devoted to strengthening these results and analyzing the effect of school voucher programs on minority private school enrollment in the long-term.

Bibliography:

Abdulkadiroglu, Atila, Pathak, Parag, Walters, Christopher, "School Vouchers and Student Achievement: Evidence from the Louisiana Scholarship Program" *NBER Working Paper Series*, December 2015.

EdChoice, "School Choice In America Dashboard," *EdChoice,* March 2021. https://www.edchoice.org/school-choice-in-america-dashboard-scia/.

EdChoice, "Who We Are" *EdChoice*, March 2021.

Carnoy, Martin & Emma Garcia, "5 Key Trends in US Student Performance," *Economic Policy Institute*, January 2017.

Epple, Dennis, Romano, Richard, Urquiola, Miguel, "School Vouchers: A Survey of the Economics Literature," *NBER Working Paper Series*, September 2015.

Hanson, Melanie, "US Public Education Spending Statistics," *Education Data Initiative*, June 15, 2022.

Hsu, Spencer, "How Vouchers Came to D.C." Education Next, June 30, 2006. Jager, Simon, "How Substitutable Are Workers? Evidence from Worker Deaths," *Department of Economics, Harvard University*, January 2016.

Jordan, Reed, "Millions of black students attend schools that are highly segregated by race and by income," *Urban Institute*, October 2014.

Krueger, Alan & Pei Zhu, "Another Look at the New York City School Voucher Experiment," *American Behavioral Scientist*, January 2004.

Mayer, Daniel, Peterson, Paul, Myers, David, Clark Tuttle, Christina and Howell, William, "School Choice in New York City after Three Years: An Evaluation of the School Choice Scholarships Program," *Mathematica Policy Research*, February 2002.

Peterson, Paul, Howell, William, Wolf, Patrick and Cambpell, David, "School Vouchers: Results from Randomized Experiments," *The University of Chicago Press*, April 2003.

Pew Research Center, "U.S. Born Hispanics Increasingly Drive Population Developments," *Pew Research Center*, January 2002.
Sachs, Rebecca, "Safety Net Cutbacks and Hospital Service Provision: Evidence from Psychiatric Care," *Congressional Budget Office*, June 2019.

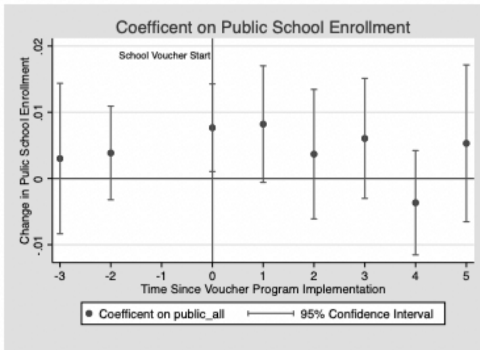
Pierce, Emily, "Private School vs. Public School," *US News*, September 2021.

Urban Institute, “State and Local Backgrounders: Property Taxes,” *The Urban Institute*, 2020.
Viteritti, Joseph, “Stacking the Deck for the Poor: The New Politics of School Choice,” *The Brookings Institute*, December 1996.

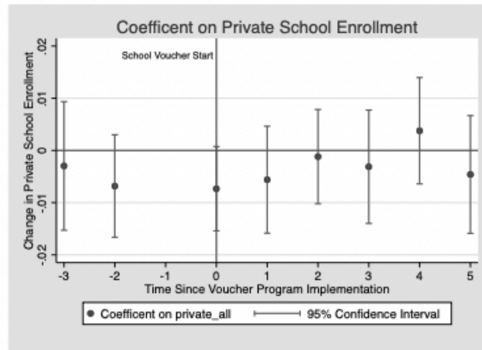
Witte, John., “Achievement Effects of the Milwaukee Voucher Program,” *University of Wisconsin at Madison*, January 1997.

Wolf, Patrick, Gutmann, Babette, Puma, Michael, Kisida, Brian, Rizzo, Lou, Eissa, Nada, Carr, Matthew, “Evaluation of the DC Opportunity Scholarship Program Final Report,” *National Center for Education Evaluation - US Department of Education*, March 2010.

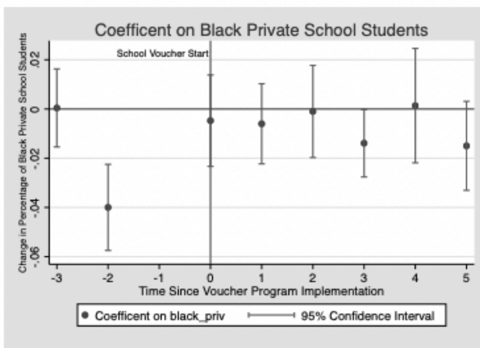
Figure 1: Enrollment and Demographic Mix for All Voucher Programs



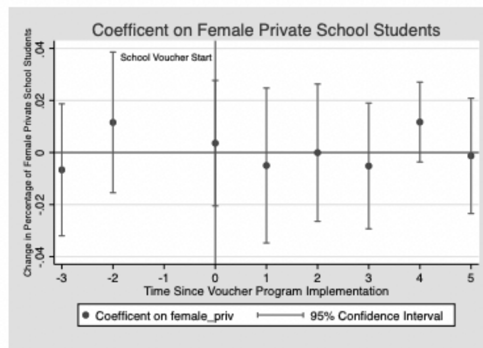
(a) Public School Enrollment



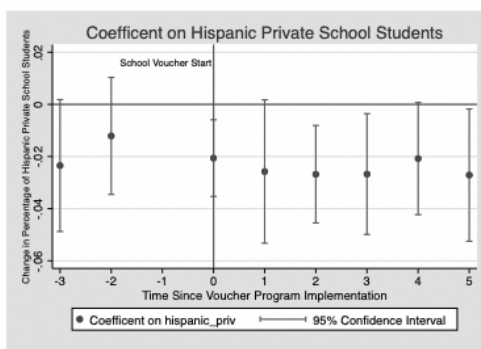
(b) Private School Enrollment



(c) Black Private School Enrollment



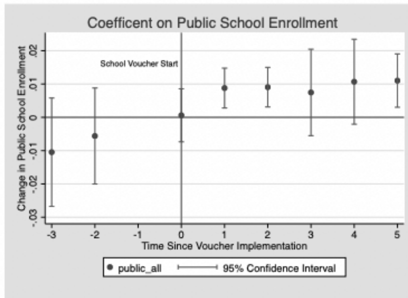
(d) Female Private School Enrollment



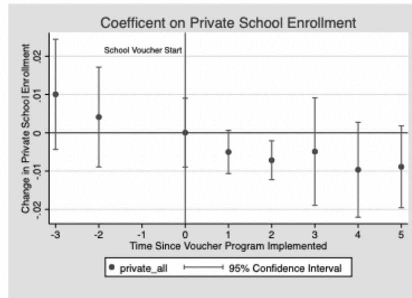
(e) Hispanic Private School Enrollment

Figure 1 shows the regression coefficients from the event study model across all voucher programs for each outcome variable. Coefficients are displayed 3 years prior and 5 years after a school voucher program has been implemented. Estimates are relative to one year pre-voucher program. The coefficients are indexed from -3 to 5. 95% confidence intervals are illustrated by red error bars. Figure 1 conveys the same findings as Table 2.

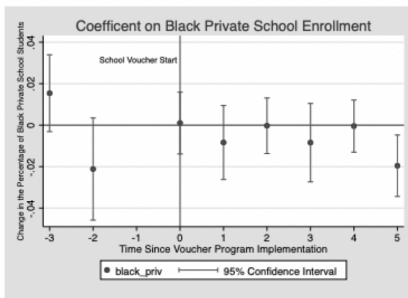
Figure 2: Enrollment and Demographic Mix for Low-Income/Poor-Performing Public Schools Voucher Programs



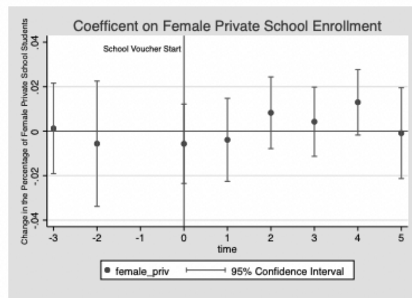
(a) Public School Enrollment



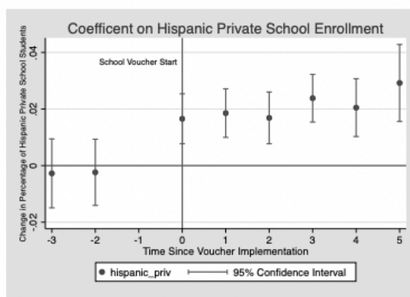
(b) Private School Enrollment



(c) Black Private School Enrollment



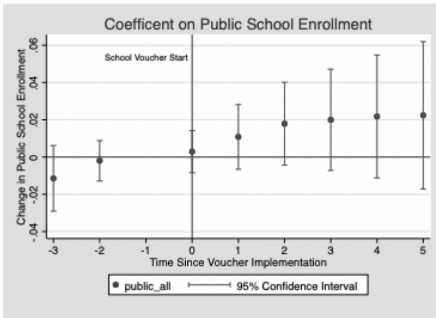
(d) Female Private School Enrollment



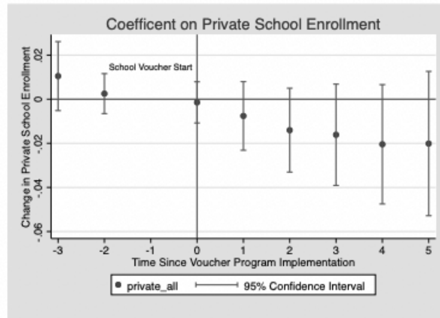
(e) Hispanic Private School Enrollment

Figure 2 shows the regression coefficients from the event study model for voucher programs with low-income/poor-performing public schools requirements for each of the outcome variables. Coefficients are estimated by comparing school voucher programs with these requirements to programs without such requirements. Estimates are relative to one year pre-voucher program. Coefficients are displayed across 3 years prior and 5 years after a school voucher program has been implemented, indexed -3 to 5. 95% confidence intervals are illustrated by red error bars. Figure 2 conveys the same findings as Table 3.

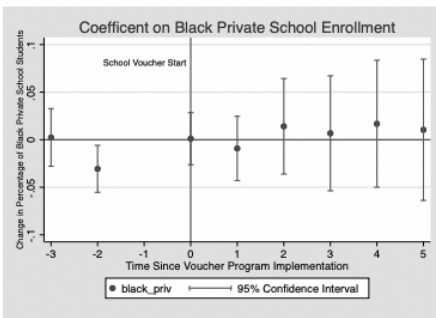
Figure 3: Enrollment and Demographic Mix for Prior Public School Enrollment Voucher Programs



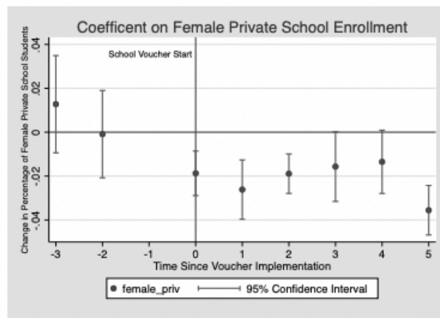
(a) Public School Enrollment



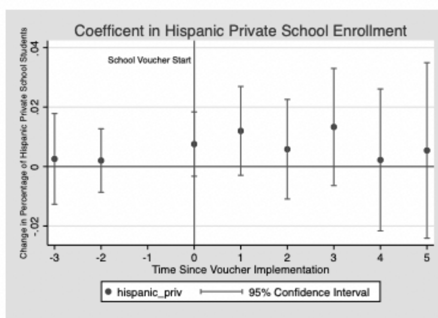
(b) Private School Enrollment



(c) Black Private School Enrollment



(d) Female Private School Enrollment



(e) Hispanic Private School Enrollment

Figure 3 shows the regression coefficients from the event study model for voucher programs with prior public school enrollment requirements for each of the outcome variables. Coefficients are estimated by comparing school voucher programs with these requirements to programs without such requirements. Estimates are relative to one year pre-voucher program. Coefficients are displayed across 3 years prior and 5 years after a school voucher program has been implemented, indexed -3 to 5. 95% confidence intervals are illustrated by red error bars. Figure 3 conveys the same findings as Table 4.

Table 1: Summary Statistics Averaged by State

	Mean	Min	Max	N
Not Enrolled	.0451843	.0306323	.0630124	51
	(.0067326)			
Enrolled	.9548157	.9369876	.9693677	51
	(.0067326)			
Attending Public School	.835257	.7057803	.8910091	51
	(.0356286)			
Attending Private School	.1195588	.0599536	.2527296	51
	(.0373697)			
Black	.1082477	.0043804	.6312138	51
	(.1218116)			
Female	.4872414	.4806582	.5016699	51
	(.0033375)			
Hispanic	.1214481	.0174539	.5020753	51
	(.1145425)			
Private School Students, Black (indicator)	.0561587	.0055444	.3123253	51
	(.0523728)			
Private School Students, Female (indicator)	.4971704	.4768145	.5177627	51
	(.008191)			
Private School Students, Hispanic (indicator)	.0765574	.0181452	.3673111	51
	(.0677475)			

Note: Table 1 shows summary statistics for key enrollment and demographic variables, averaged by state across the entire time period of study, 2005-2019. Standard error is given in parentheses.

Table 2: Coefficients for Event Study Model Across All Voucher Programs

	(1) Pub. School	(2) Priv. School	(3) Black Priv. School	(4) Female Priv. School	(5) Hispanic Priv. School
3 Yrs Pre-Voucher	0.00301	-0.00297	0.00043	-0.00663	-0.02346
	(0.00580)	(0.00628)	(0.00808)	(0.01293)	(0.01291)
2 Yrs Pre-Voucher	0.00386	-0.00682	-0.03998***	0.01157	-0.01209
	(0.00361)	(0.00501)	(0.00892)	(0.01378)	(0.01145)
Voucher Start Yr	0.00767**	-0.00733	-0.00473	0.00360	-0.02061***
	(0.00338)	(0.00412)	(0.00948)	(0.01229)	(0.00750)
1 Yr Post-Voucher	0.00821	-0.00561	-0.00601	-0.00499	-0.02575
	(0.00449)	(0.00522)	(0.00830)	(0.01518)	(0.01403)
2 Yrs Post-Voucher	0.00368	-0.00118	-0.00101	-0.00005	-0.02679***
	(0.00499)	(0.00460)	(0.00956)	(0.01346)	(0.00954)
3 Yrs Post-Voucher	0.00605	-0.00312	-0.01390**	-0.00516	-0.02674**
	(0.00462)	(0.00554)	(0.00699)	(0.01232)	(0.01182)
4 Yrs Post-Voucher	-0.00366	0.00378	0.00138	0.01169	-0.02079
	(0.00402)	(0.00519)	(0.01186)	(0.00782)	(0.01097)
5 Yrs Post-Voucher	0.00531	-0.00460	-0.01498	-0.00129	-0.02713**
	(0.00604)	(0.00575)	(0.00922)	(0.01129)	(0.01295)
Constant	0.79601***	0.17454***	0.09646***	0.46654***	0.64634***
	(0.00784)	(0.00851)	(0.00929)	(0.02036)	(0.02180)
Number of Observations	778081	778081	117153	117153	117153

Note: Standard errors are given in parentheses. ***p<0.01, **p<0.05. The coefficients displayed in Table 2 give the average effect of a school voucher program on each outcome variable. Coefficients are relative to a base year of one year pre-voucher program. Pub. School is an indicator for whether an individual attends public school. Priv. School is an indicator for whether an individual attends private school. Black Priv. School is an indicator for whether an individual is black and attends private school. Female Priv. School is an indicator for whether an individual is female and attends private school. Hispanic Priv. School is an indicator for whether an individual is hispanic and attends private school. Changes in the number of observations is due to moving from the subset of total enrolled students to the smaller subset of only privately enrolled students.

Table 3: Coefficients for Voucher Programs with Low-Income or Poor-Preforming Schools Requirements

	(1) Pub. School	(2) Priv. School	(3) Black Priv. School	(4) Female Priv. School	(5) Hispanic Priv. School
3 Yrs Pre-Voucher	-0.01046	0.01004	0.01540	0.00126	-0.00275
	(0.00830)	(0.00733)	(0.00945)	(0.01038)	(0.00622)
2 Yrs Pre-Voucher	-0.00560	0.00410	-0.02115	-0.00562	-0.00237
	(0.00736)	(0.00665)	(0.01258)	(0.01437)	(0.00597)
Voucher Start Yr	0.00061	0.00003	0.00103	-0.00567	0.01656***
	(0.00406)	(0.00459)	(0.00760)	(0.00910)	(0.00450)
1 Yr Post-Voucher	0.00879***	-0.00502	-0.00834	-0.00390	0.01856***
	(0.00305)	(0.00288)	(0.00909)	(0.00953)	(0.00438)
2 Yrs Post-Voucher	0.00905***	-0.00716***	-0.00028	0.00827	0.01687***
	(0.00302)	(0.00258)	(0.00684)	(0.00822)	(0.00466)
3 Yrs Post-Voucher	0.00746	-0.00491	-0.00842	0.00425	0.02383***
	(0.00663)	(0.00716)	(0.00965)	(0.00794)	(0.00430)
4 Yrs Post-Voucher	0.01068	-0.00965	-0.00044	0.01298	0.02049***
	(0.00650)	(0.00633)	(0.00642)	(0.00751)	(0.00521)
5 Yrs Post-Voucher	0.01102***	-0.00887	-0.01954***	-0.00087	0.02921***
	(0.00408)	(0.00545)	(0.00756)	(0.01042)	(0.00694)
Constant	0.81412***	0.14147***	0.03474***	0.45486***	0.09731***
	(0.01399)	(0.01175)	(0.01176)	(0.01747)	(0.00981)
Number of Observations	373835	373835	56277	56277	56277

Table 4: Coefficents for Voucher Programs with Prior Public School Enrollment Requirements

	(1) Pub. School	(2) Priv. School	(3) Black Priv. School	(4) Female Priv. School	(5) Hispanic Priv. School
3 Yrs Pre-Voucher	-0.01147	0.01049	0.00233	0.01276	0.00256
	(0.00899)	(0.00799)	(0.01544)	(0.01130)	(0.00779)
2 Yrs Pre-Voucher	-0.00200	0.00253	-0.03065**	-0.00090	0.00201
	(0.00553)	(0.00463)	(0.01260)	(0.01015)	(0.00545)
Voucher Start Yr	0.00291	-0.00142	0.00110	-0.01873***	0.00755
	(0.00576)	(0.00478)	(0.01399)	(0.00518)	(0.00551)
1 Yr Post-Voucher	0.01082	-0.00757	-0.00913	-0.02615***	0.01198
	(0.00886)	(0.00794)	(0.01725)	(0.00689)	(0.00762)
2 Yrs Post-Voucher	0.01786	-0.01402	0.01409	-0.01891***	0.00583
	(0.01134)	(0.00971)	(0.02561)	(0.00458)	(0.00854)
3 Yrs Post-Voucher	0.01996	-0.01612	0.00674	-0.01567	0.01330
	(0.01386)	(0.01172)	(0.03083)	(0.00810)	(0.01005)
4 Yrs Post-Voucher	0.02174	-0.02043	0.01679	-0.01350	0.00222
	(0.01682)	(0.01380)	(0.03405)	(0.00735)	(0.01218)
5 Yrs Post-Voucher	0.02239	-0.02009	0.01041	-0.03554***	0.00541
	(0.02019)	(0.01669)	(0.03790)	(0.00575)	(0.01505)
Constant	0.80442***	0.14897***	0.07859***	0.43697***	0.06668***
	(0.01509)	(0.01287)	(0.02065)	(0.00874)	(0.01070)
Number of Observations	373835	373835	56277	56277	56277

Note: Standard errors are given in parentheses. ***p<0.01, **p<0.05. Table 3 gives the average effect of a school voucher program for programs with low-income or poor-performing public school requirements. Table 4 gives the average effect of a school voucher program for programs with prior public school enrollment requirements. Pub. School is an indicator for whether an individual attends public school. Priv. School is an indicator for whether an individual attends private school. Black Priv. School is an indicator for whether an individual is black and attends private school. Female Priv. School is an indicator for whether an individual is female and attends private school. Hispanic Priv. School is an indicator for whether an individual is hispanic and attends private school. Changes in the number of observations is due to moving from the subset of total enrolled students to the smaller subset of only privately enrolled students.

Appendix

Coding Details:

My outcome variables for enrollment and the demographic mix of students were created using the following ACS variables: race, sex, year, age, predhisp (a variable indicating hispanic ethnicity), and schltype, a variable indicating if a student is enrolled in public school, private school, is not enrolled at all (including homeschooled). In addition to these variables, my dataset includes three geographic specifications: statefip (indicates each state within the US), countyfip (indicates each county within the US), and met2013 (indicates each metropolitan area within the US, as defined by US Office of Management and Budget's 2013 metropolitan statistical areas).

Private School Students, Black (indicator) is coded as 1 for black students attending private school, 0 for non-black students attending private school, and missing for all students not attending private school. Private School Students, Female (indicator) is coded as 1 for female students attending private school, 0 for male students attending private school, and missing for all students not attending private school. Private School Students, Hispanic (indicator) is coded as 1 for hispanic students attending private school, 0 for non-hispanic students attending private school, and missing for all students not attending private school. Because all three of these variables are drawn from the subset of students attending private school, they all have the same lower number of observations (1103953).

Difference-in-Differences Model:

Table 5: Grouped Difference-in-Differences Model Coefficients

	(1) Pub. School	(2) Priv. School	(3) Black Priv. School	(4) Female Priv. School	(5) Hispanic Priv. School
Voucher_Treat_Post	0.00289	-0.00043	0.00548	-0.00201	-0.01569**
	(0.00302)	(0.00311)	(0.00536)	(0.00650)	(0.00631)
Constant	0.78504***	0.16112***	0.05029***	0.46432***	0.07117***
	(0.00393)	(0.00398)	(0.01017)	(0.01321)	(0.00769)
Number of Observations	778081	778081	117153	117153	117153

Note: Standard error is given in parentheses. ***p<0.01, **p<0.05.

I conducted a separate difference-in-differences model with the same matching specification. The regression equation for this model is:

$$Y_{i,t} = \beta_0 + \alpha_t + \gamma_i + \beta_1 D_j + \beta_2 D_j^{Treated} + \epsilon_{i,t} \qquad (2)$$

Y is the outcome of interest. α_t represents time-fixed effects. γ_i represents region-fixed effects. D_j is a dummy variable indicating if the year is before or after the voucher program is implemented. $D_j^{Treated}$ represents the areas with school voucher programs over the 5 years after the program had been implemented. The regression residual is represented by ϵ. β_2 is the coefficient of interest. It indicates the average effect of a school voucher program on Y across all 5 years post-implementation.

Table 5 displays regression coefficients from the difference-in-differences regression model. Table 5 reaches the same conclusions as Table 2. The regressions in Table 5 find that school vouchers had no significant effect on enrollment or the demographic mix of black and female private school students across all 5 years post-voucher program. Table 5 confirms my finding that school vouchers are associated with a significant decrease in private school hispanic enrollment. Due to its enhanced specificity, the coefficients from the matched event-study regression model in Table 2 are preferred.

Analysis Excluding State-Level Data (Ohio & DC):

Table 6: Grouped Event Study Model Coefficents, Excluding OH & DC

	(1) Pub. School	(2) Priv. School	(3) Black Priv. School	(4) Female Priv. School	(5) Hispanic Priv. School
3 Yrs Pre-Voucher	0.00584	-0.00761	0.01332	-0.02185	-0.04613***
	(0.00847)	(0.00841)	(0.01183)	(0.01698)	(0.01234)
2 Yrs Pre-Voucher	0.00756	-0.01315**	-0.04336***	-0.00576	-0.02549
	(0.00439)	(0.00614)	(0.01264)	(0.01694)	(0.01410)
Voucher Start Yr	0.00755	-0.01110	-0.00458	-0.01975	-0.03373***
	(0.00577)	(0.00683)	(0.01497)	(0.01202)	(0.00882)
1 Yr Post-Voucher	0.01510***	-0.01247	-0.01571	-0.03857***	-0.05278***
	(0.00467)	(0.00754)	(0.01166)	(0.01098)	(0.01057)
2 Yrs Post-Voucher	-0.00144	0.00044	0.00694	-0.02469	-0.04311***
	(0.00506)	(0.00752)	(0.00728)	(0.01487)	(0.01168)
3 Yrs Post-Voucher	0.01255	-0.01017	-0.01002	-0.03069***	-0.05057***
	(0.00754)	(0.00906)	(0.01010)	(0.01067)	(0.01069)
4 Yrs Post-Voucher	0.00369	-0.00636	-0.01301	0.00677	-0.03581**
	(0.00698)	(0.00857)	(0.01174)	(0.01484)	(0.01697)
5 Yrs Post-Voucher	0.00108	-0.00100	-0.01458	-0.02288**	-0.05379***
	(0.00696)	(0.00753)	(0.01263)	(0.01088)	(0.01122)
Constant	0.80625***	0.15123***	0.10321***	0.47879***	0.36489***
	(0.00337)	(0.00459)	(0.00989)	(0.01476)	(0.00936)
Number of Observations	414702	414702	61770	61770	61770

Note: Standard error is given in parentheses. ***p<0.01, **p<0.05.

Table 6 displays coefficients from the event study model in Table 2 adjusted to exclude state-level data from Ohio and DC. The results indicate that school voucher programs have no significant effect on enrollment or the demographic mix of black and female private school students.

Table 6 strengthens the finding that school vouchers are associated with a decrease in hispanic private school enrollment. School voucher programs are associated with a small but significant decline in hispanic private school enrollment by 3-5% each year post-voucher program. Removing DC and Ohio rendered the decrease in hispanic private school enrollment significant across all years post-voucher program. This specification also increased the magnitude of the decline by roughly 2% each year. These results are consistent with the demographic characteristics of DC and Ohio. Across 2000-2019, DC's population is about 9.4% hispanic on average. DC's hispanic private school enrollment is approximately 7%. Across 2000-2019, Ohio's population is about 4% hispanic on average. Ohio's hispanic private school enrollment is approximately 3.6%. Compared to Table 1, Ohio and DC are below the mean hispanic population and hispanic private school enrollment. In addition, hispanic populations are extremely concentrated in urban areas and city centers. In 2000, almost 90% of hispanics lived in urban areas. Almost half of hispanics resided in a central city (Pew Research Center, 2002). Therefore, the stronger association between school vouchers and lower hispanic private school enrollment is consistent with DC and Ohio's relatively small and concentrated hispanic populations.

Excluding DC and Ohio from study introduced a possible association between school voucher programs and lower female private school enrollment. Table 6 shows that female private school enrollment declined by 2-3% nearly every year post-voucher program. I conclude that the association between school voucher programs and female private school enrollment is spurious due to inconsistency of this trend. It is unlikely that school vouchers would be given to substantially more female than male students.

Illustration of Prior Trends:

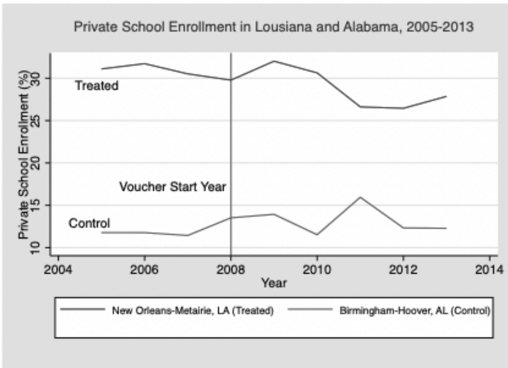


Figure 4: Private School Enrollment

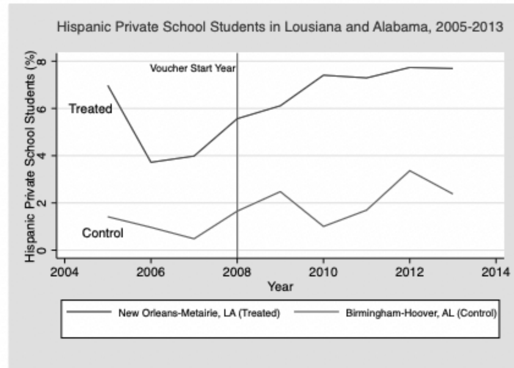


Figure 5: Hispanic Private School Students

Figure 4 and Figure 5 display graphical representations of the summary statistics for two key outcome variables. Figure 4 shows trends in private school enrollment within the Louisiana and Alabama match. Figure 5 shows trends in hispanic private school enrollment within the same match. These figures provide a match-level supplement to the summary statistics in Table 1.

Figure 4 shows private school enrollment in the New Orleans-Metairie, Louisiana metro area and the Birmingham-Hoover, Alabama metro area from 2005 to 2013. This is a matched pair in my study. Louisiana implemented a school voucher program in 2008. Louisiana is the treated unit. Alabama did not have a school voucher program within this time period. Alabama is the control unit. The time frame 2005-2013 was selected to show trends in both states 3 years prior and 5 years post-voucher program. Figure 4 shows that private school enrollment increased more sharply in Louisiana than in Alabama in 2008. Private school enrollment dropped steadily 2009-2011 in the Louisiana metro area. Private school enrollment spiked sharply in the Alabama metro area 2010-2011. The implementation of a voucher program in Louisiana in 2008 appeared to cause a short-term increase in private school enrollment relative to the control region.

However, this effect is small in magnitude and begins decreasing 2 years post-voucher program. The trends seen in Figure 4 appear consistent with the results in Table 2 and Figure 1.

Figure 5 shows the percentage of hispanic private school students in the New Orleans-Metairie, Louisiana metro area and the Birmingham-Hoover, Alabama metro area from 2005 to 2013. The time frame 2005-2013 was selected to show trends in both states 3 years prior and 5 years post-voucher program. Figure 5 shows that hispanic private school enrollment increased in both regions beginning in 2007. This increase was sustained in Louisiana until 2013. In Alabama private school hispanic enrollment increased 2008-2009 and fluctuated until 2013. Relative to Alabama, Lousiana's voucher program appeared to cause a sustained increase in private school hispanic enrollment. However, these effects are small in magnitude. While Louisiana displayed a more sustained increase in private school hispanic enrollment, Alabama showed a similar, though inconsistent, increase in private school hispanic enrollment over the same time period. The trends in Figure 5 appear consistent with the regression coefficients in Table 2 and Figure 1.

Matched Voucher Programs Reference Table:

Match	School Voucher State	School Voucher Metro Area	Non-School Voucher City	School Voucher Start Year	Low Income/Poor-Preforming School Req?	Prior Public School Req?
1	Indiana	Indianapolis-Carmel-Anderson	San Antonio-New Braunfels, TX	2011	Yes	Yes
2	Wisconsin	Dane County	Kent County, MI	2013	Yes	Yes
3	DC	N/A	Delaware	2004	Yes	No
4	Ohio	N/A	Pennsylvania	2006	Yes	Yes
5	North Carolina	Raleigh	Charleston-North Charleston, SC	2014	Yes	Yes
6	Maryland	Baltimore-Columbia-Towson	Nashville-Davidson-Murfreesboro-Franklin, TN	2016	Yes	No
7	Lousiana	New Orleans-Metairie	Birmingham-Hoover, AL	2008	Yes	Yes
8	Georgia	Atlanta-Sandy Springs-Rothwell	Miami-Fort Lauderdale-West Palm Beach, FL	2007	No	Yes

Note: The reference table above lists the matches between voucher and non-voucher metropolitan areas used in my matching specification. Each voucher program's start year and eligibility requirements are included as well.

Individual Match Level Analysis:

Table 7: Effect of School Vouchers on Private School Black Enrollment at the Match Level

	Match 1	Match 2	Match 3	Match 4	Match 5	Match 6	Match 7	Match 8
3 Yrs Pre-Voucher	-0.01345	-0.00320	-0.03448	-0.02345*	0.03180	0.00088	0.08634*	0.00000
	(0.02342)	(0.01805)	(0.05088)	(0.01139)	(0.04244)	(0.02229)	(0.04337)	(.)
2 Yrs Pre-Voucher	-0.00733	-0.04520*	-0.07186	-0.02623*	-0.05788	-0.03625	-0.06997	-0.03534
	(0.02307)	(0.01894)	(0.05188)	(0.01077)	(0.03820)	(0.02299)	(0.04540)	(0.01931)
Voucher Start Yr	0.00015	-0.00780	-0.06269	0.00418	-0.07418	-0.02029	0.05370	0.01084
	(0.02304)	(0.01848)	(0.05392)	(0.00874)	(0.04195)	(0.02361)	(0.04213)	(0.01940)
1 Yr Post-Voucher	0.01230	-0.05047**	0.00379	0.00610	-0.02678	-0.01155	-0.01177	0.00049
	(0.02395)	(0.01852)	(0.05773)	(0.00867)	(0.03914)	(0.02241)	(0.04376)	(0.01830)
2 Yrs Post-Voucher	0.01162	-0.00956	-0.04414	-0.00278	-0.03961	0.00263	0.03480	0.01911
	(0.02230)	(0.01901)	(0.05702)	(0.00883)	(0.03926)	(0.02082)	(0.04480)	(0.01965)
3 Yrs Post-Voucher	0.00122	-0.03426	-0.04005	-0.01468	-0.06539	0.00576	0.01778	-0.00014
	(0.02570)	(0.01910)	(0.05397)	(0.00881)	(0.03731)	(0.02143)	(0.04530)	(0.01946)
4 Yrs Post-Voucher	0.02708	-0.05968***	-0.06758	0.02070*	-0.00426	0.00000	-0.00018	-0.00244
	(0.02709)	(0.01664)	(0.05486)	(0.00913)	(0.03361)	(.)	(0.04596)	(0.02091)
5 Yrs Post-Voucher	-0.03443	-0.02913	-0.11551*	-0.00653	-0.06110	0.00000	0.07131	0.00177
	(0.02288)	(0.01903)	(0.05448)	(0.00943)	(0.03797)	(.)	(0.04217)	(0.01990)
Constant	0.08427***	0.09829***	0.40311***	0.08535***	0.11914***	0.13565***	0.08799*	0.17993***
	(0.02018)	(0.01437)	(0.04089)	(0.00912)	(0.03455)	(0.01885)	(0.04065)	(0.01604)
Number of Observations	6983.00000	14117.00000	4762.00000	50652.00000	3914.00000	9570.00000	7055.00000	20100.00000

Table 8: Effect of School Vouchers on Private School Hispanic Enrollment at the Match Level

	Match 1	Match 2	Match 3	Match 4	Match 5	Match 6	Match 7	Match 8
3 Yrs Pre-Voucher	-0.07944*	-0.08658***	0.01202	0.00126	0.02368	-0.03226*	0.02057	0.00000
	(0.03846)	(0.02033)	(0.02607)	(0.00788)	(0.02506)	(0.01500)	(0.01639)	(.)
2 Yrs Pre-Voucher	-0.04602	-0.05421**	-0.00376	-0.00371	-0.00464	-0.01332	-0.00748	-0.00209
	(0.03844)	(0.02094)	(0.02663)	(0.00698)	(0.02161)	(0.01395)	(0.01484)	(0.01913)
Voucher Start Yr	-0.08051*	-0.05882**	-0.01040	-0.00440	-0.09017*	-0.00686	0.00412	-0.02222
	(0.04002)	(0.01967)	(0.02847)	(0.00542)	(0.03526)	(0.01664)	(0.01627)	(0.01875)
1 Yr Post-Voucher	-0.02971	-0.08145***	-0.03668	0.00677	0.00884	-0.02812	0.00133	-0.07053***
	(0.04067)	(0.02167)	(0.02990)	(0.00586)	(0.02481)	(0.01620)	(0.01703)	(0.01914)
2 Yrs Post-Voucher	-0.16025***	-0.04324*	-0.01919	-0.00830	-0.00906	-0.02283	0.02901	-0.03951*
	(0.03922)	(0.02201)	(0.02334)	(0.00596)	(0.02103)	(0.01550)	(0.01630)	(0.01967)
3 Yrs Post-Voucher	-0.11048**	-0.05084*	-0.01481	0.00026	-0.02035	-0.04048**	0.02096	-0.04748*
	(0.03966)	(0.02272)	(0.02589)	(0.00615)	(0.02792)	(0.01524)	(0.01697)	(0.01988)
4 Yrs Post-Voucher	-0.10439*	-0.04916*	-0.03783	-0.00150	-0.07778**	0.00000	0.00868	-0.01937
	(0.04086)	(0.02196)	(0.02650)	(0.00648)	(0.02878)	(.)	(0.01970)	(0.01939)
5 Yrs Post-Voucher	-0.07935*	-0.10036***	-0.02197	0.00193	0.00640	0.00000	0.01810	-0.06084**
	(0.04006)	(0.02045)	(0.02611)	(0.00676)	(0.02479)	(.)	(0.01823)	(0.01968)
Constant	0.12235**	0.19239***	0.05856**	0.02848***	0.04236*	0.06580***	0.04916***	0.05246**
	(0.03729)	(0.01679)	(0.02053)	(0.00624)	(0.01832)	(0.01364)	(0.01287)	(0.01792)
Number of Observations	6983.00000	14117.00000	4762.00000	50652.00000	3914.00000	9570.00000	7055.00000	20100.00000

Note: Standard errors are in parentheses. * p<0.1, ** p<0.05, *** p<0.01.

Table 7 and 8 analyze the effect of school vouchers on enrollment and the demographic mix of private school students at the match-level. Table 7 displays the coefficients on private school black enrollment from an event study model conducted within each match. Each column gives the effect of school vouchers on the demographic mix of black private school students within each match. Coefficients are estimated relative to a base year of one year prior to school voucher program implementation. The matched voucher program reference table (see page above) lists each match number. Match 8 (Georgia) does not have data for 3 years pre-voucher program due to the unavailability of metropolitan area data prior to 2005. Georgia's school voucher program began in 2007. Match 6 (Maryland) does not have data 4 and 5 years post-voucher program due to the unavailability of data past 2019. Maryland's school voucher program began in 2016. Table 7 indicates that school voucher programs do not have a significant effect on private school black enrollment at the match level individual matches. I find Match 2 (Wisconsin) as one possible exception. School vouchers may be associated with a slight decrease in private school black enrollment within Wisconsin's match. However, statistically significant coefficients are displayed only 1 and 4 years post-voucher implementation. Therefore, I cannot conclude with certainty that this association exists. These results are consistent with those found in Table 2. Just as in the aggregated event study model, the match-level analysis suggests that school vouchers did not encourage increased black racial diversity among private school students. It may also suggest that black students did not receive a proportional number of school vouchers.

Table 8 displays the coefficients on private school hispanic enrollment from an event study model conducted within each match. Each column gives the effect of school vouchers on the demographic mix of hispanic private school students within each match. Coefficients are estimated relative to a base year of one year prior to school voucher program implementation. Table 8 indicates that school voucher programs are associated with a small but significant decrease in hispanic private school enrollment in Indiana (Match 1), Wisconsin (Match 2), Georgia (Match 8). I regard the association seen in Wisconsin as spurious due to the presence of negative and prior to school voucher program implementation. The prior parallel trends assumption may not be satisfied for Wisconsin at the match level. School vouchers are associated with the most significant decrease in private school hispanic enrollment in Indiana. School vouchers are associated with a decrease in private school hispanic enrollment by 8-16% nearly every year post-voucher program. This result is surprising. Indiana's school voucher program specifically targets low-income students or students from poor-performing public school districts. It also requires students to have been previously enrolled in public school. These results indicate that school vouchers may have been given primarily to non-hispanic students in Indiana.

Graduate Student Unions on Enrollment, Doctorates, and Papers

Sarah Gao

Abstract

In 2016 the National Labor Relations Board (NLRB) ruled that graduate students are employees allowing graduate students at private universities to begin unionizing. Since then over a dozen universities have held votes and many have unionized. This paper examines the impacts of unionization on the number of earned doctorates, graduate student enrollment, and number of papers published. Using a comparative case studies design we find there are no significant effects of unionization on earned doctorates, graduate student enrollment, or number of papers published. However schools that have a union vote that ends up failing publish more papers after the vote.

1 Introduction

For some time union membership has been declining in the US with only 10.3% of wage and salary workers belonging to a union compared to 54% at the peak of union membership in 1954 according the Bureau of Labor Statistics (BLS). However, recent unionization movements at companies like Amazon, Starbucks, and Delta, seem to suggest a resurgence in the labor movement across the country. Another group that has recently been unionizing are graduate students. In 2016 the National Labor Relations Board (NLRB), ruled that graduate students at private universities are employees, opening the doors to graduate student unions across the country. Since their 2016 decision, dozens of universities have had union votes including 12 who have established unions on their campuses.

In this paper we aim to understand the impact of graduate student unions on universities both to contribute to the literature's broader discussion on the effect of unions but also to provide insight to graduate students across the country who are deciding how to vote in their union elections. Here we focus on three main outcomes: graduate student enrollment, number of earned doctorates, and number of published papers. While not all enrolled graduate students work as teacher's assistants, research assistants, or other student jobs, groups that are officially represented by the union, we use enrollment as a proxy for employment, because many of the issues that graduate student unions advocate for do impact all graduate students including providing dental and vision coverage and 3rd party arbitration in discrimination cases. Unions affect a university's enrollment because as unions engage in bargaining and advocate for increased wages and other benefits it could make the cost of enrolling graduate students more expensive causing universities to potentially admit fewer graduate students. The next measure we look at is earned doctorates. This also serves to measure how many students a university can fund but additionally it is a measure for the satisfaction of the employees. A PhD is long and arduous degree often taking 4-6 years to complete. In fact, according a Wiley report only 50-60% of PhD candidates go on to earn their doctorates. If we were to see an increase in earned doctorates while enrollment remained the same, we could conclude that unions improve student satisfaction causing more people to finish their degrees. The final outcome that we look at is the number of papers a university publishes. One of the main products that research universities produce are research papers. In order to examine the effect of unionization on a university's productivity, we see how unions impact the number of papers a university produces. A union could effect the number of papers that a firm produces by making employees feel like that have more agency, empowering them to do better and more efficient work. Research assistants are not the only job that graduate students have, in fact many graduate students work as teachers assistants. However, the productivity of graduate students in their role of a teachers assistant is much more difficult to examine.

In order to see if unions have an effect on these outcomes we use a comparative case studies design to understand the effect of unions on each of our desired outcomes. Namely, we build a synthetic control for each university that establishes a graduate student union, and aggregate the results across all treated universities to see if there is any effect of unionization.

Additionally, we compare our treated universities to a set of placebos, universities that either did not pass their union vote or passed their union vote but failed to unionize. From this we find that there are no significant effects on enrollment, earned doctorates, or number of papers published. However, we find that in our placebo universities there are more papers published after they vote and do not establish a union. One possible mechanism for this is that students at these universities were impassioned enough about their working conditions to organize a union vote; however because they don't actually unionize are able to maintain a better relationship with their school's administration and become more productive.

Economists have long tried to understand the effect of unions on a firm's productivity and employment. Here our finding of a null result on graduate student enrollment and earned doctorates complements the results of Wang and Young (2021) that find heterogeneity in the effect of unionization on overall employment and firm survival by sector. Namely, that there are little to no effects on the overall employment and survival rates for firms in the service sector, whereas overall employment and survival rates decline in manufacturing firms. Similarly, Dinardo and Lee (2004) and Freeman and Kleiner (1999) find no relationship between unions and business survival. As a consequence our findings contradict Sojourner et al. (2015) and Lalonde et al. (1996) that find employment declines following successful unionization. Some of these differences are likely due to the fundamentally distinct nature of universities compared to most firms: universities are non-profit organizations with a commitment to education. Although some differences may also be due to a time lag effect, since enrollment and earned doctorates may take many years to update and our data set only has data for a couple of years after unionization.

There has also been a lot of work on the effect of unionization on a firms productivity. Notably, Dube et al. (2016) find that unionization has positive effects on patient care after nurses unionize and Mas (2006) find police officers are more effective after successful arbitration. However, Lalonde et al. (1996) find productivity declines following unionization. Our findings, that productivity increases only in firms that having a union election but do not actually unionize diverge somewhat from both of these findings. Similarly to employment, some of these differences are likely due to the nature of universities. Universities may be more willing to hear graduate students demands than other firms after a unionization attempt because one aspect of a universities success is their ability to educate graduate students, different than most firms where their goal is not to train workers.

The remaining paper is structured as follows. Section 2 discusses the data sources, Section 3 describes our empirical test, Section 4 describes are results, first discussing are main results, followed by robustness checks, and finally our placebo tests.

2 Data

We use five data sources to examine the effects of labor relations: a report published by Hunter College enumerating the universities that have unionized, data released by the US Department of Education about US universities, the Web of Science database, surveys conducted by the National Science Foundation on number of earned doctorates, and finally university rankings published by US News.

1. **2020 Supplementary Directory of New Bargaining Agents and Contracts in Institutions of Higher Education, 2013-2019** Herbert et al. (2020). This report published in 2020 by the National Center is a compilation of data collected on new bargaining units, bargaining agents, and contracts in higher education from 2013 to 2019. This data set contains a list of every union that exists at a higher education institution including those representing graduate students, professors, librarians, postdoctoral researchers, and other employees hired by a university. Additionally, it gives explicit definitions of what groups each union covers as well as the date of the creation. In particular, we use this to see which private universities formed graduate student unions from 2013 to 2019 and when their union was formed.

2. **College Scorecard** The College Scorecard is an API maintained by the US Depart-

ment of Education that has time series data on a variety of aspects of higher education institutions including institutional characteristics, enrollment, student aid, costs, and student outcomes. The College Scorecard aggregates data from the Integrated Postsecondary Education Data System (IPEDS) which is a series of surveys that the Department of Education collects annually from every higher education institution that participates in federal student financial aid programs. In particular, we will be leveraging this data set to see what graduate student enrollment is per year at each institution.

3. **Web of Science** Another axis that we are interested in looking at is the productivity of universities after a graduate student union is formed. We do this by examining the number of papers a university publishes. We use the Web of Science Core Collection database to see how many papers each university publishes a year. This database aggregates 21,100 peer-reviewed, high-quality scholarly journals across hundreds of disciplines and provides a variety of descriptors for each paper including author affiliation and year published. From this database we are able to collect data on the number of papers each university publishes each year.

4. **National Survey of Earned Doctorates (NSED)** The National Science Foundation (NSF) conducts a variety of surveys across universities, in particular this data set collects the number of earned doctorates that each university has each year.

5. **US News Ranking** In order to create a synthetic control of similar universities we want our synthetic control to be similar in ranking to the treated group. The US News Ranking provides a ranking of higher education institutions across the US. While no university ranking list perfectly captures how good a university is, the US News ranking is widely regarded as a reputable list and uses a combination of reputation surveys, quantitative measures (eg student to teacher ratio), measures of student quality, and graduation and retention rates to aggregate their rankings. With this data set, we are

able to gauge a rough ranking of the universities in question.

Once we have all this data we merge it into a long panel data set such that we have ranking, enrollment, number of earned doctorates, and number of papers for each year and university. Additionally we have a dummy variable, treated, that is coded for 0 if there is no union at that university during that year and 1 if there is one during that year. We have data for the top 188 schools spanning the years 2001 to 2019. We began with looking at the top 202 schools in 2017; however, we drop 14 universities because of missing data.[1]

Table 1 presents the summary statistics of our data. The data is averaged over each year and university. We note that there are very few private universities that unionize from 2013-2019. There are only 11 schools that unionize: American University, Georgetown University, University of Chicago, Loyola University Chicago, Brandeis University, Harvard University, Tufts University, Columbia University, New School, New York University, and Brown University.

Table 2 presents the summary statistics of our outcome variables by treatment, notably the universities that unionize tend to have a higher rank and publish more papers than universities that do not unionize.

3 Empirical Analysis

Finding the causal effect of the adoption of a graduate student union can be difficult since the universities that unionize tend to be schools that are ranked relatively high and the sample size of universities that have unionized is still small, making a standard difference in difference approach difficult. In order to overcome these concerns we employ the use of

[1]We drop Lesley University, Maryville University of St. Louis, University of St Thomas, St. John Fisher College, Immaculata University, Robert Morris University, Lipscomb University, Union University, and Edgewood College because there is no data for number of earned doctorates for those universities. We drop Azusa Pacific University because there is no data for number of papers published. We drop La Verne, Pepperdine, Indiana University, and UC Merced because of missing graduate student enrollment data. Some schools were missing only one year of enrollment data. If this was the case we take the average of the preceding and succeeding year's enrollment data.

comparative case studies. We follow Cavallo et al. (2013) by constructing a synthetic control for each university that unionizes and aggregating the results over the universities that have unionized by taking the average for each year after treatment.

There have been only 11 private universities that have implemented a graduate student union before 2020. To find a control for each of these universities we use the synthetic control method as described by Abadie et al. (2010). For each university that has implemented a union we create a synthetic control that minimizes the mean squared prediction error. To construct each of these synthetic controls we choose weights for untreated universities such that the resulting synthetic control best resembles the pre intervention values of the treated university. We use a series of covariates for each outcome variable to create the synthetic control. For example, in creating a synthetic control for a school's enrollment we create a synthetic control such that it matches the treated university in school ranking, number of papers published, and graduate student enrollment pre intervention. Once we construct a synthetic control for each university, we compare the results between the post intervention outcomes of the treated university and the synthetic control.

Additionally, we are interested in doing inference on the average effect of unionization across the 11 universities. After creating synthetic controls for each of the treated universities, we follow Cavallo et al. (2013) and take a simple average over the effect of unionization for each outcome variable for each year after the treatment period. Specifically, we use the notation as follows. For example, lets look at the outcome variable of papers published. We observe J universities, without loss of generality, let the first G be the universities that unionize and the remaining $J - G$ universities be those that did not unionize. Additionally, we observe over the time period $\{1, ...T\}$. For each treated university i let T_0^i be the year that a union was established at university i, where $T_0^i < T$ for all universities $i, i \in \{1, ...G\}$ (a treated university).

Let our estimates of the effect of unionization on papers published for some treated university i be $(\hat{\alpha}_{i,T_0^i+1}, ..\hat{\alpha}_{i,T})$ Where $\hat{\alpha}_{i,j}$ is the difference between the number of papers published at our synthetic university i and our actual university i in year j. Now to find the aggregated effect of unionization across 11 treated universities we take the average at each year after treatment, so the estimated average effect is given by

$$\bar{\alpha} = (\bar{\alpha}_{T_0+1}, ..\bar{\alpha}_T) = \frac{1}{G}\sum_{g=1}^{G}(\hat{\alpha}_{g,T_0^g+1}, ..\hat{\alpha}_{g,T}).$$

Next we follow Cavallo et al. (2013)'s structure to find the statistical significance of each of our estimates. First, we discuss how to find the p-value for a single university and then how to find the p-value for the aggregated universities. Since we have a relatively small number of treatment groups we use a permutation test. We are trying to examine whether the difference of the outcome variable between a treated university and its synthetic control is large relative to the universities that did not unionize. Without loss of generality, we describe how to find the p-value of the effect of unionization on papers published for University 1, as we recall a treated unit. We go through each of our control universities and assign a placebo unionization during T_0^1, the year University 1 unionized, create a synthetic control for this unionization, find the difference between the the placebo unionization and the synthetic control, and see if the magnitude of the placebo effect is larger than the effect of the unionization at University 1. Specifically, we compute a p-value for each year, l, after T_0^1, as follows:

$$\text{p-value}_l = \Pr(\hat{\alpha}^{PL}_{1,l+T_0^1} < \hat{\alpha}_{1,l+T_0^1}) = \frac{\sum_{c=G+1}^{J} \mathbb{1}_{\hat{\alpha}^{PL(c)}_{1,l+T_0^1} < \hat{\alpha}_{1,l+T_0^1}}}{\text{num of control universities}} = \frac{\sum_{c=G+1}^{J} \mathbb{1}_{\hat{\alpha}^{PL(c)}_{1,l+T_0^1} < \hat{\alpha}_{1,l+T_0^1}}}{J-G},$$

where $\hat{\alpha}^{PL(c)}_{1,l+T_0^1}$ is the placebo effect of a union l years after T_0^1 on control university c.

Next, we discuss how we find the p-value of our aggregated synthetic controls. Similar to before, we want to iterate through all possible placebos and see what percentage of placebos result in an effect that is greater than than the effect that we find in our treated university; however, here we want to compute over all possible placebo averages instead of single universities. To do this we do the following steps:

1. For each treated university we compute all the placebo effects using the available donors. So we construct synthetic controls for all non treated universities and find

$$\hat{\alpha}^{PL(j)}_{g,l+T_0^g}, \text{ for all } j \in \{G+1, ..J\}, g \in \{1, ..G\} \text{ and } l \in \{1, ..T - \max_{i\in\{1,..G\}} T_0^i\}$$

2. At each lead, l we compute every possible placebo average effect by picking G placebos then taking the average over the G placebos. The number of possible placebo averages is

$$N_{\bar{PL}} = (J-G)^G$$

3. With out loss of generality we index all the placebo averages from $1, ..N_{\bar{PL}}$ such that $\bar{\alpha}_l^{PL(p)}$ where $p \in \{1, ..N_{\bar{PL}}\}$ is the average placebo effect for the pth placebo average.

4. We compute the p-value l years after the treatment as:

$$\begin{aligned} \text{p-value}_l &= \Pr(\bar{\alpha}_l^{PL} < \bar{\alpha}_l) && (1)\\ &= \frac{\sum_{p=1}^{N_{\bar{PL}}} \mathbb{1}_{\bar{\alpha}_l^{PL(p)} < \bar{\alpha}_l}}{\#\text{ of possible placebo averages}} && (2)\\ &= \frac{\sum_{p=1}^{N_{\bar{PL}}} \mathbb{1}_{\bar{\alpha}_l^{PL(p)} < \bar{\alpha}_l}}{N_{\bar{PL}}} && (3) \end{aligned}$$

4 Results

We consider three outcomes: earned doctorates, enrollment, and number of papers. In this section we present the results of our synthetic control for each outcome, followed by robustness checks, and placebo tests.

4.1 Main Results

We now show the results of our empirical test for our three different outcome variables: earned doctorates, enrollment, and number of papers. Here we find that there are no significant effects on any of these outcomes after a graduate student union is formed.

4.1.1 Earned Doctorates Results

First, we discuss number of earned doctorates. Especially prominent in manufacturing sectors, after unionizing the employment and survival rates of the firm may go down. While universities don't employ graduates students in the same way as most firms do, here we look at the number of earned doctorates both to measure the number of people that the university is able to fund and to measure how many students are satisfied and are willing to finish their course of study. To construct the synthetic control for earned doctorates, we try to create a control that is similar to the pre intervention values of the treated universities in the following covariates: graduate student enrollment, papers, and rank across all years leading up to the treatment and earned doctorates in 2001, 2005, and 2010.

In the first column of Table 3 and Figure 1 we see the results of our aggregated synthetic control for number of earned doctorates. Note that, with a slight abuse of notation, in Table 3-6 we have p-values in the parenthesis as opposed to standard deviation. Additionally, the first graph in Figures 8 - 18 show the results of the earned doctorates synthetic control for each of the treated universities. Here we see that the number of earned doctorates for each university is both small in magnitude and noisy in quality. The additional drawback is that it takes four to six years to get a PhD meaning that this metric takes a lot of time to reflect any changes. Accordingly, we find that unionization has little to no effect on the number of earned doctorates, noting large p-values and noisy graphs.

4.1.2 Enrollment Results

Another outcome that we are interested in looking at related to employment is the number of graduate students enrolled at an university. Distinct from earned doctorates, enrollment also includes masters students, medical students, law students, and MBA students, as well as a variety of other graduate students. This larger sample base gives us less noisy data that becomes easier for our synthetic control method to create an accurate control. To construct the synthetic control for enrollment, we try to create a control that is similar to the pre intervention values of the treated universities in the following covariates: rank, papers, and earned doctorates across all years leading up to the treatment and enrollment in 2001, 2005, and 2010.

The second column of Table 3 and Figure 2 show the results of our aggregated synthetic control for enrollment, again noting the p-values instead of standard deviations in the parenthesis. Additionally, the second graph in Figures 8 - 18 show the results of enrollment synthetic controls for each of the treated universities. Here, we see that enrollment for each university is much less noisy. However, we also find no significant effect on enrollment.

4.1.3 Number of Papers

Finally, we show the results of our empirical test for number of papers. We look at the number of papers published to measure the productivity of the university after unionization. To construct the synthetic control for number of papers published, we try to create a control that is similar to the pre intervention values of the treated universities in the following covariates: number of earned doctorates, enrollment numbers, and rank across all pre intervention years and number of papers in 2001, 2005, and 2010.

The third column of Table 3 and Figure 3 we see the results of our synthetic control for number of papers, again noting the p-values instead of standard deviations in the parenthesis. Additionally the third graph in Figures 8 - 18 show the results of papers for each of the treated universities. Here we see that the number of papers matches quite closely to the constructed synthetic control, especially once smoothed out in the aggregated value. We find that unionization has little to no effect on the number of papers published.

4.2 Robustness Checks

When we conduct our synthetic control matching for papers as one of our covariates, we run into an issue: Harvard is an outlier in the number of papers that it publishes, publishing far more papers than any other university. In the time period we are looking at, Harvard published on average 27,703 papers per year, significantly more than the average number of papers published by the controls (2,727 with a standard deviation of 2,773.89 see Table 2) and much more than the control university that publishes the most papers: Washington University in St. Louis that publishes on average 6,078 papers per year. The third graph in Figure 16 depicts Harvard's synthetic control for number of papers, we see that when we try to create a synthetic control that tries to match Harvard in the number of papers that it creates, it is still much less than the number that Harvard actually publishes.

In order to ensure that this outlier doesn't have too large of an effect on our results, since we only have 11 treated groups, we try two approaches. First, we create a synthetic control that doesn't match on papers, next we create a synthetic control that matches on papers but we drop the Harvard observation. We do this for both earned doctorates and enrollment since both of those synthetic controls are trying to control for papers. In Table 4 we see that the number of earned doctorates has a similar magnitude and p-value both to each other and our original results. Similarly, in Table 5 we see that enrollment has a similar magnitude and p-value both to each other and our original results. Thus, we conclude that this outlier should not have a large effect on our results.

4.3 Placebo Test

Additionally, we compare our results to a couple of placebo universities. Five other schools held union votes but failed to unionize for a variety of reasons. Here we compare our results to these schools, using the year of their vote as the beginning of their treatment. These schools provide a good comparison to our treated units because they were all campuses that were able to organize and willing to put on a union election; however, because of various factors they did not end up being able to form a union.

In Cornell and Duke's union vote more than half of students voted against a union resulting in no union formed. However, at Yale, Boston College, and the University of Pennsylvania graduate student unions won their respective votes but the union organizers decided to withdraw their petitions from the NLRB resulting in no union being formally recognized at these universities. This is because in 2018, the year these universities were holding their vote, Trump appointed a 3rd republican to NLRB making the board majority Republican. University unions worried that if they were to submit their petitions the board could make a decision that overturned the 2016 precedent allowing graduate students to unionize. This is in line with Frandsen (2017)'s findings that when Republicans control the NLRB their is evidence of manipulation in favor of employers.

By comparing these universities to our actual treated units we can parse out any effects that are had by merely having an election or going through the process of trying to unionize rather than having an actual union.

4.3.1 Earned Doctorates (Placebo)

In the first column in Table 6 and Figure 8 we see the results of our aggregated synthetic control for number of earned doctorates in our placebo universities, and the first column of Figures 22-26 show the synthetic control for number of earned doctorates at each individual placebo university. Similar to our main results, we find the number of earned doctorates to be relatively noisy and the effects insignificant.

4.3.2 Enrollment (Placebo)

In second column of Table 6 and Figure 9 we see the results of our aggregated synthetic control for graduate student enrollment in our placebo universities, and the second column of Figures 22-26 show the synthetic control for enrollment at each individual placebo university. Similarly, we find the effect of unionization on graduate student enrollment to be insignificant.

4.3.3 Number of Papers (Placebo)

In the third column of Table 6 and Figure 10 we see the results of our aggregated synthetic control for number of published papers in our placebo university. Interestingly, we see that there is significant effect on number of papers published in our placebo universities. Namely, our placebos, the universities that have had union votes but do not end up actually forming a union, publish significantly more papers than their controls. One year after the union vote the placebos publish on average 10,278 papers, 683 more papers than the control, and two years after the vote publish 10,920 papers, 881 more papers than the control. One possible mechanism is that having a union election signals to the university that graduate students' are very upset by their working conditions. However, the lack of actual unionization allows graduate students to have a more harmonious relationship with administration allowing for positive change to occur and thus increasing productivity. We hesitate to draw too strong of a long term conclusion from this because there is time lag in how quickly an effect would be reflected in our data as papers can take years to publish.

5 Conclusion

In this article we examine the effect of unionization at graduate schools. We look at three main outcome variables: enrollment, number of earned doctorates, and number of published papers. Using a synthetic control set up, we find that unions don't appear to have any effect on number of earned doctorates, graduate student enrollment, or number papers published.

However, our results do indicate that schools that have union election but fail to actually unionize do publish more papers after their union election than universities that do not have a union vote. We note that one of the main drawbacks of our empirical results is the time horizon. Since graduate student unions at private schools were only allowed to unionize starting in 2016 of the 11 universities that we look at six unionize in 2017 and four unionize in 2018.[2] We worry that some of the effects of unionizing will not yet be captured by current data because it takes time after the creation of a union for changes from bargaining to occur and our outcome variables, especially number of papers published and doctorates earned take time to update, since both endeavors can take years to complete.

Additionally, one of the biggest appeals of a graduate student union to graduate students is the promise of higher wages. Conducting this analysis would be meaningful in continuing to understand the role that unions have in universities across the United States and inform graduate students across the country on the impact of their vote for or against a graduate student union

[2] New York university voluntary agreed to recognize a graduate student union in 2013

6 Tables and Figures

6.1 Tables

Table 1: Summary Statistics

	Sum	Mean	SD	Min	Max	N
Year	7,103,340	2,010.00	5.48	2,001	2,019	3,534
Rank	336,224	95.14	56.34	1	197	3,534
Union	209	0.06	0.24	0	1	3,534
Num Papers	10,219,956	2,891.89	3,538.83	0	43,195	3,534
Num of Earned Doctorates	768,351	217.42	196.25	0	917	3,534
Enrollment	19,716,382	5,579.06	4,012.05	5	27,970	3,534

Note: Year indicates a calendar year. Rank is the ranking as given by the US News Ranking. Union is a 1 if that schools unionizes in the years 2013-2019 and 0 if it does not. Num Papers is the number of papers that a university published as noted by the Web of Science Core Collections. Num of Earned Doctorates is the number of earned doctorates at each university at a given year as indicated by NSED. Enrollment is the number of graduate students enrolled at a given university at a given year as indicated by the College Scorecard

Table 2: Summary Statistics by Treatment

	All	Unionized University	Non-Unionized University
Rank	95.14	40.27	98.59
	(56.48)	(42.30)	(55.56)
Num of Earned Doctorates	217.42	239.30	216.04
	(192.97)	(200.30)	(193.01)
Enrollment	5,579.06	8,225.36	5,412.72
	(3,668.33)	(6,318.30)	(3,397.43)
Papers	2,891.89	5,522.78	2,726.52
	(3,326.22)	(7,918.51)	(2,773.89)

Note: Rank is the ranking as given by the US News Ranking. Papers is the number of papers that a university published as noted by the Web of Science Core Collections. Num of Earned Doctorates is the number of earned doctorates at each university at a given year as indicated by NSED. Enrollment is the number of graduate students enrolled at a given university at a given year as indicated by the College Scorecard

Table 3: Aggregated Results

	Earned Doctorates	Enrollment	Papers
	Treatment - Control	Treatment - Control	Treatment - Control
After 1 Yr	-9.26	723.09	-25.4
	(0.51)	(0.17)	(0.88)
After 2 Yr	5.08	896.29	-126.8
	(0.72)	(0.15)	(0.51)

Note: With a slight abuse of notation, we have p-values in the parenthesis as opposed to standard deviation.

Table 4: Earned Doctorates Results (Robustness)

	Drop Harvard, Look at Papers	Don't Look at Papers
	Treatment - Control	Treatment - Control
After 1 Yr	-19.26	-11.48
	(0.19)	(0.43)
After 2 Yr	2.52	12.24
	(0.86)	(0.39)

Note: With a slight abuse of notation, we have p-values in the parenthesis as opposed to standard deviation.

Table 5: Enrollment Results (Robustness)

	Drop Harvard, Look at Papers	Don't Look at Papers
	Treatment - Control	Treatment - Control
After 1 Yr	826.5	774.7
	(0.14)	(0.14)
After 2 Yr	1070.7	938.7
	(0.10)	(0.12)

Note: With a slight abuse of notation, we have p-values in the parenthesis as opposed to standard deviation.

Table 6: Aggregated Results (Placebo)

	Earned Doctorates	Enrollment	Papers
	Treatment - Control	Treatment - Control	Treatment - Control
After 1 Yr	-14.4	-24.5	682.3
	(0.43)	(0.96)	(0.04)
After 2 Yr	5.1	262.5	881.05
	(0.79)	(0.67)	(0.02)

Note: With a slight abuse of notation, we have p-values in the parenthesis as opposed to standard deviation.

6.2 Figures

Figure 1: Aggregated Synthetic Control for Earned Doctorates

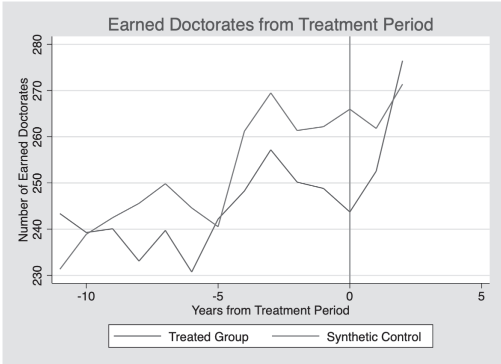


Figure 2: Aggregated Synthetic Control for Enrollment

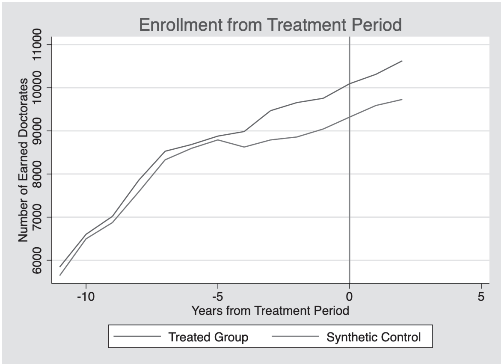


Figure 3: Aggregated Synthetic Control for Paper

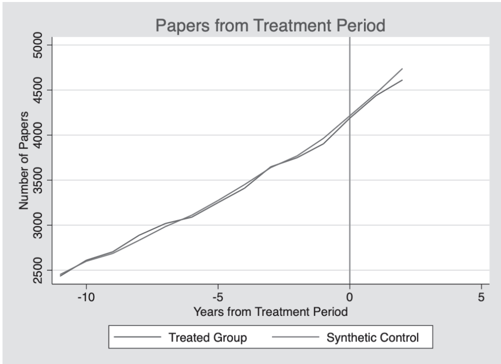


Figure 4: Aggregated Synthetic Control for Enrollment (drop Harvard, match on papers)

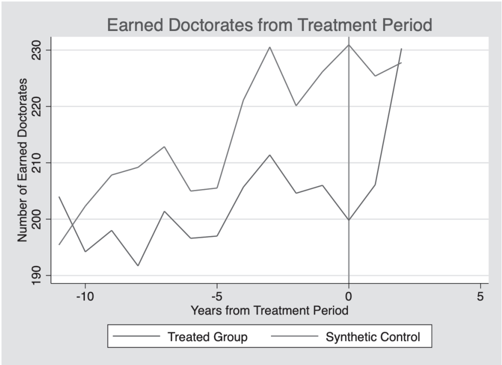


Figure 5: Aggregated Synthetic Control for Enrollment (don't match on papers)

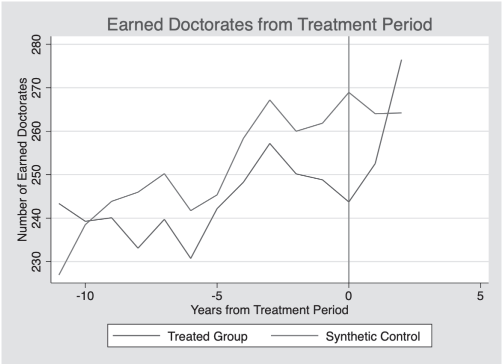


Figure 6: Aggregated Synthetic Control for Enrollment (drop Harvard, match on papers)

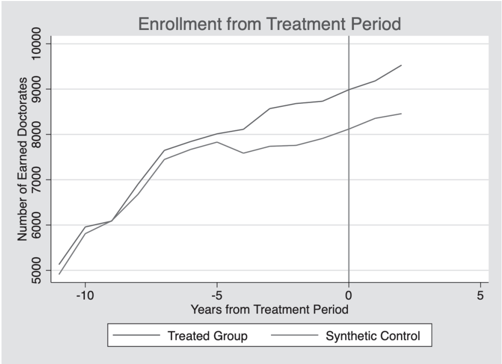


Figure 7: Aggregated Synthetic Control for Enrollment (don't match on papers)

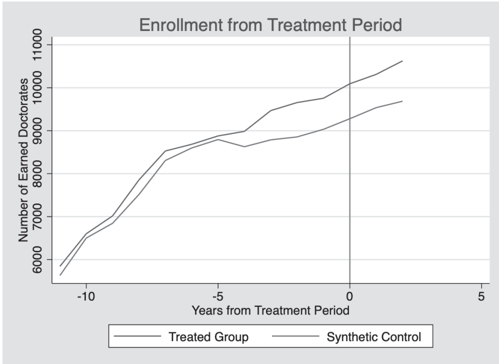


Figure 8: Placebo Aggregated Synthetic Control for Earned Doctorates

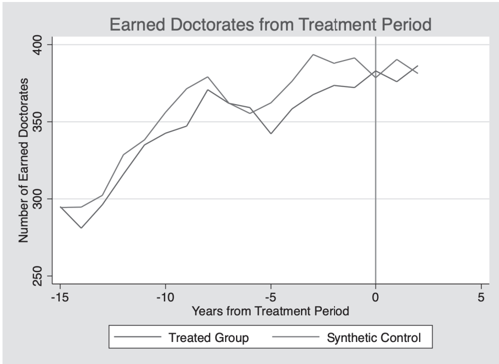


Figure 9: Placebo Aggregated Synthetic Control for Enrollment

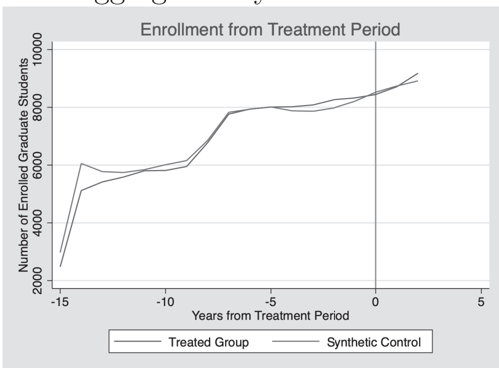


Figure 10: Placebo Aggregated Synthetic Control for Papers

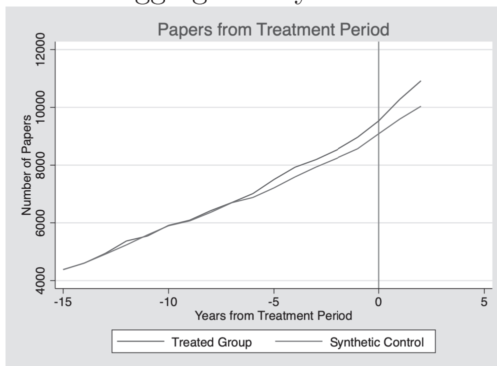


Figure 11: Synthetic Controls for American University

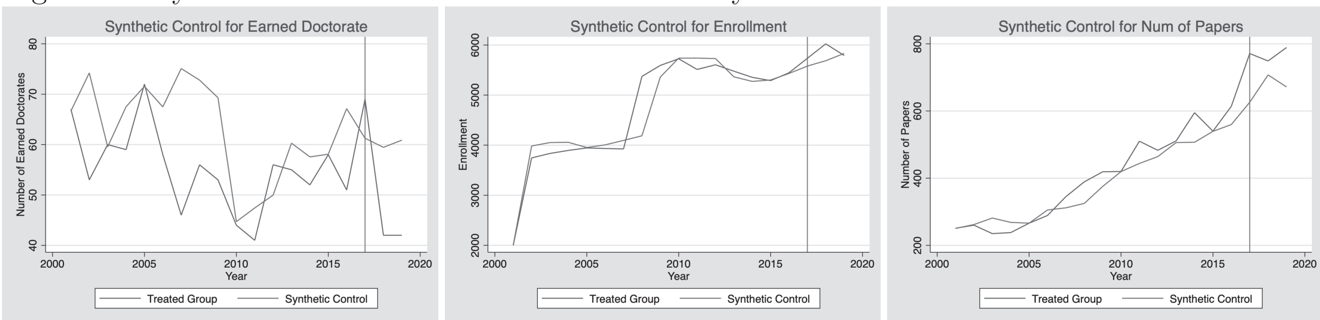


Figure 12: Synthetic Controls for Georgetown University

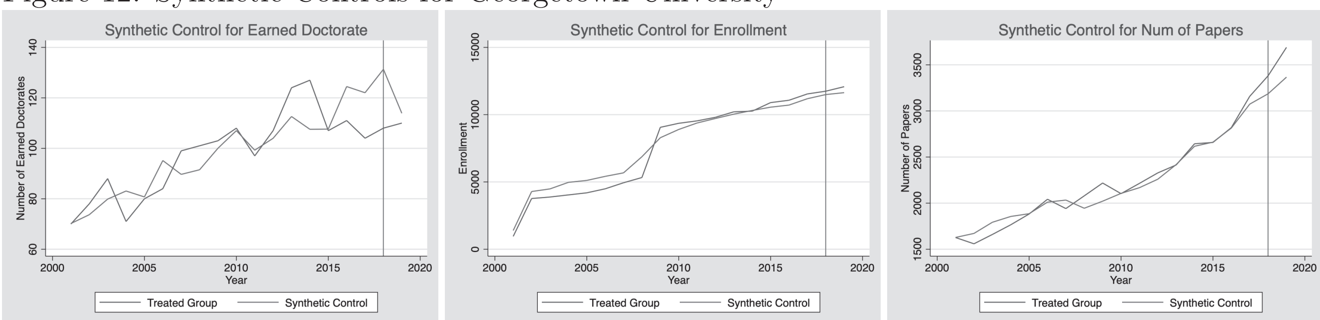


Figure 13: Synthetic Controls for University of Chicago

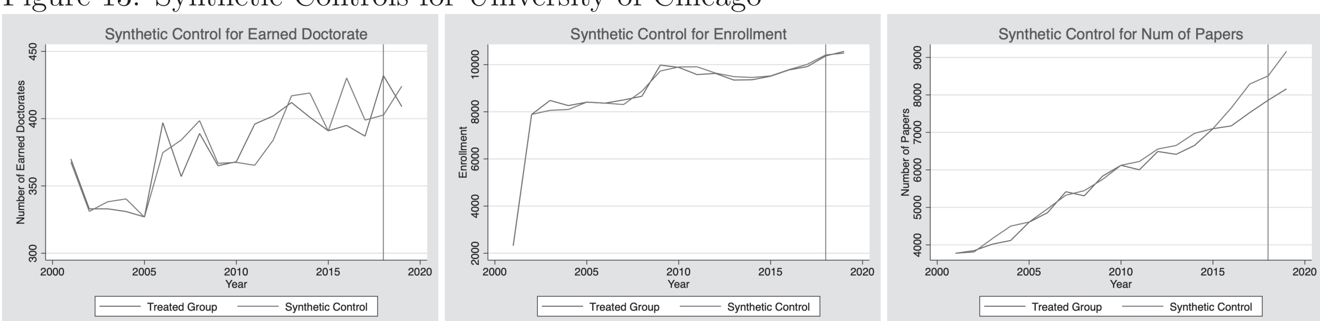


Figure 14: Synthetic Controls for Loyola University Chicago

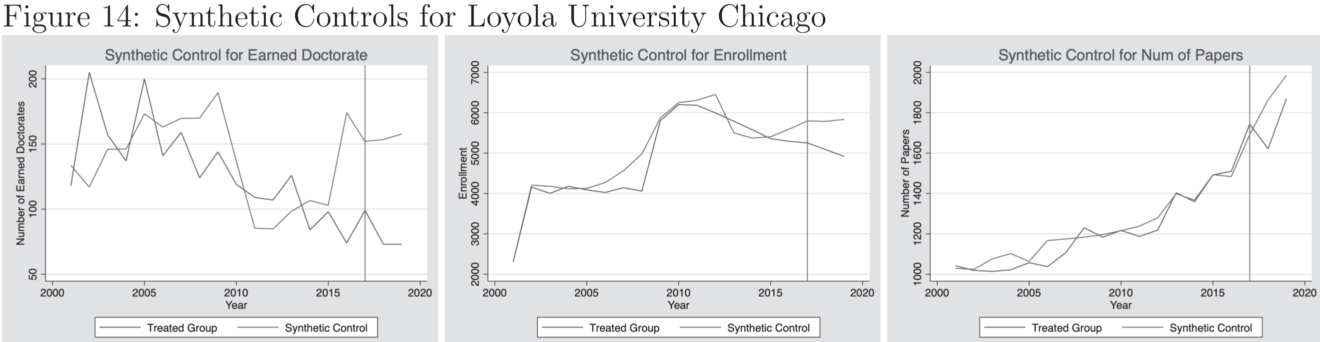


Figure 15: Synthetic Controls for Brandeis University

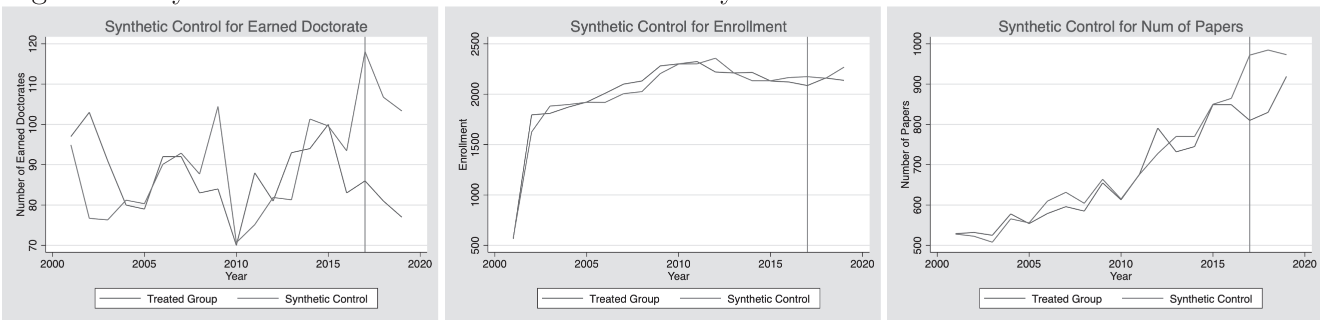


Figure 16: Synthetic Controls for Harvard University

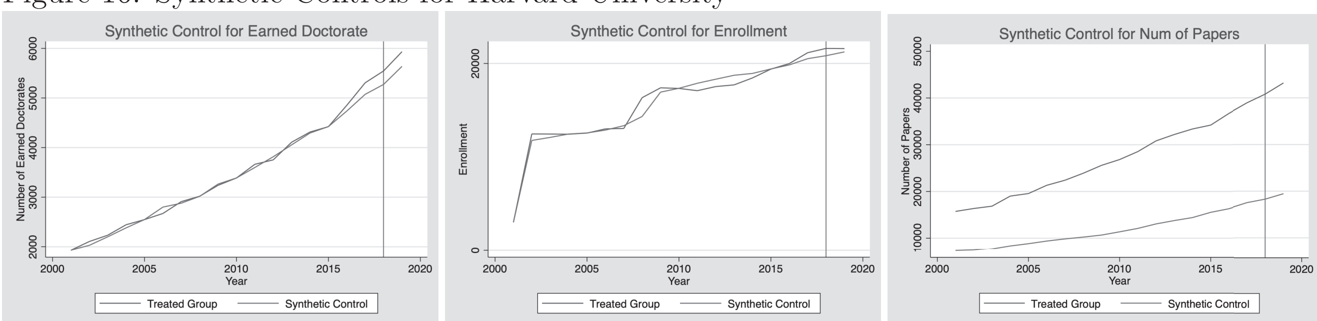


Figure 17: Synthetic Controls for Tufts University

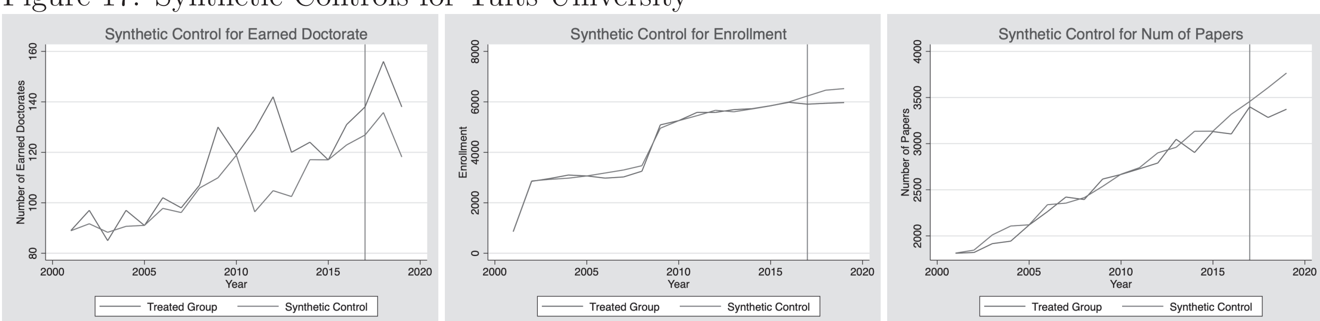


Figure 18: Synthetic Controls for Columbia University

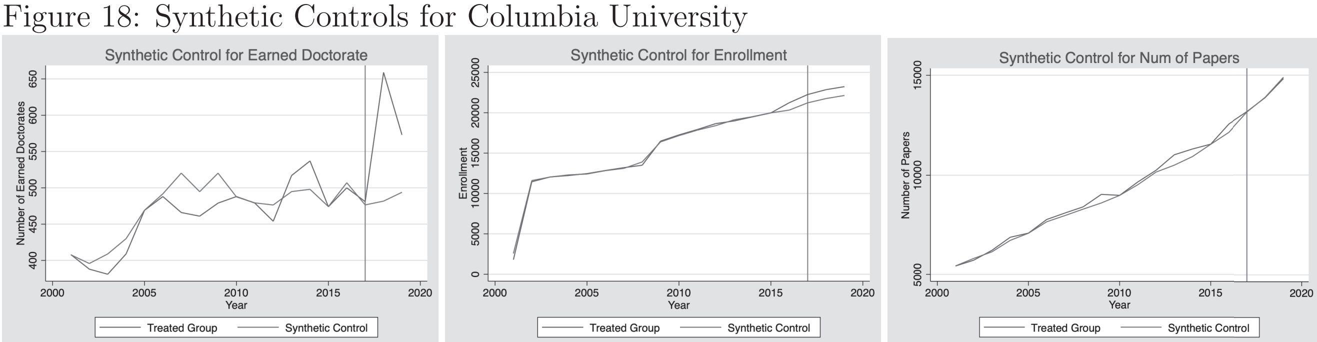


Figure 19: Synthetic Controls for New School

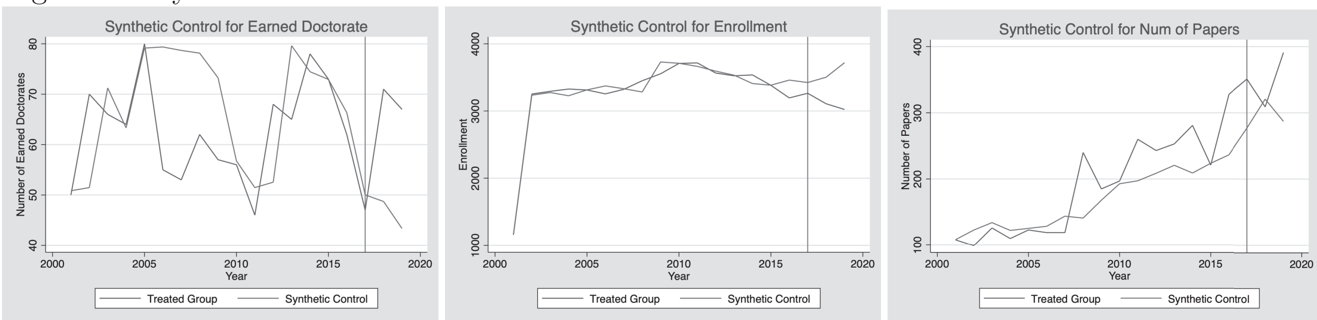


Figure 20: Synthetic Controls for New York University

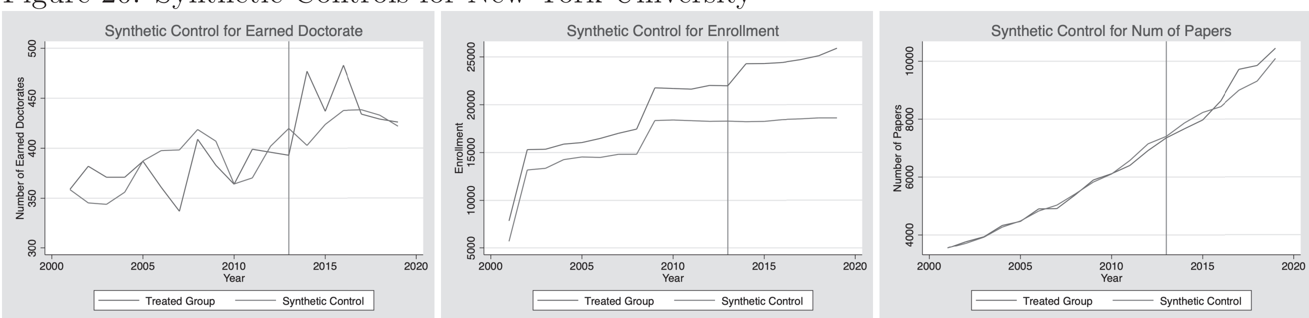


Figure 21: Synthetic Controls for Brown University

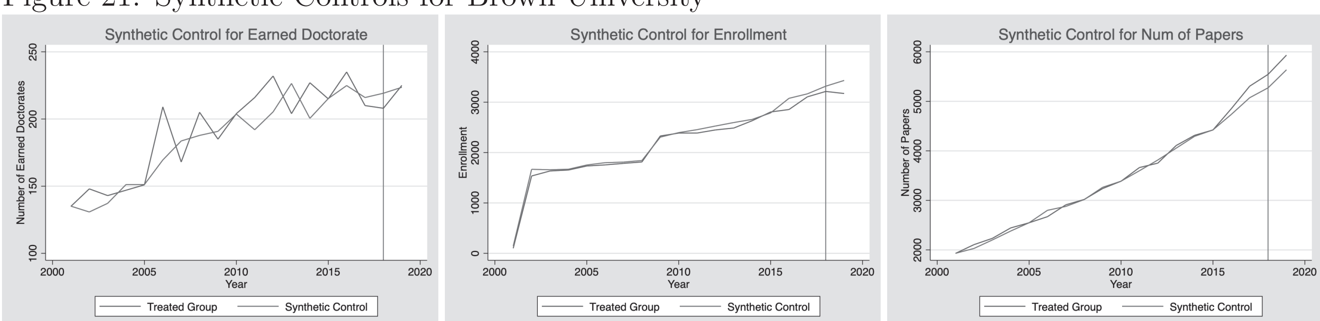


Figure 22: Synthetic Controls for Yale (Placebo)

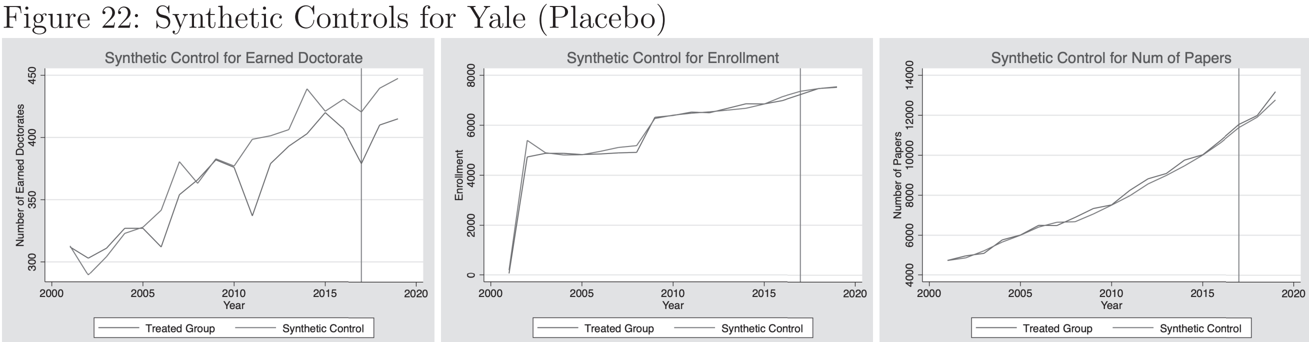


Figure 23: Synthetic Controls for Boston College (Placebo)

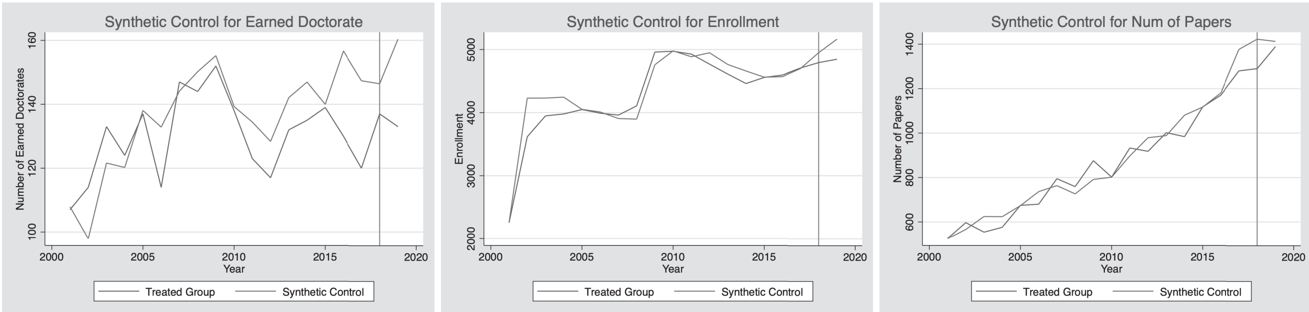


Figure 24: Synthetic Controls for Cornell (Placebo)

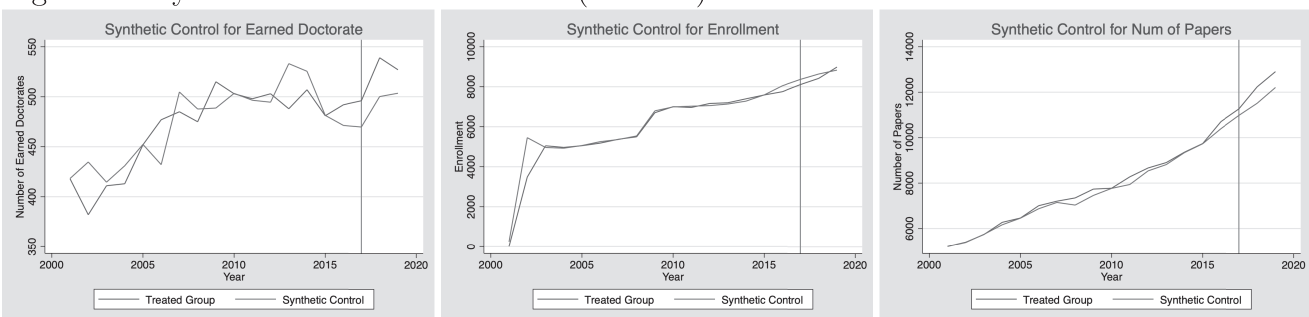


Figure 25: Synthetic Controls for University of Pennsylvania (Placebo)

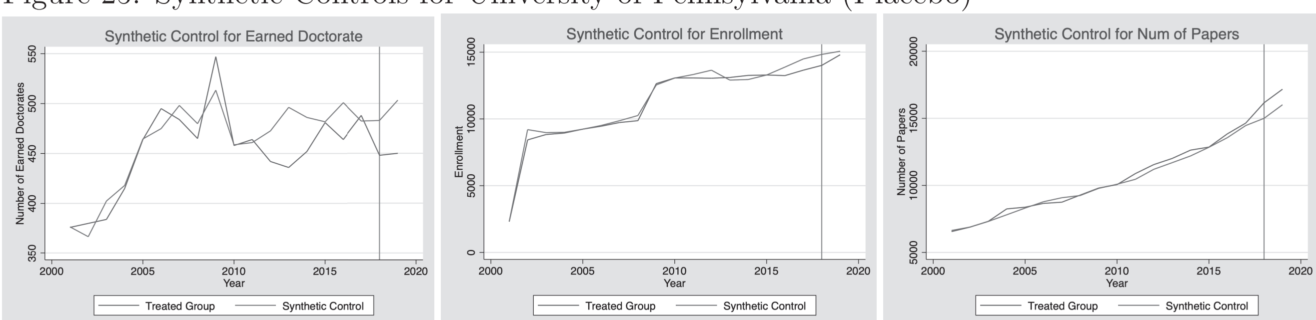
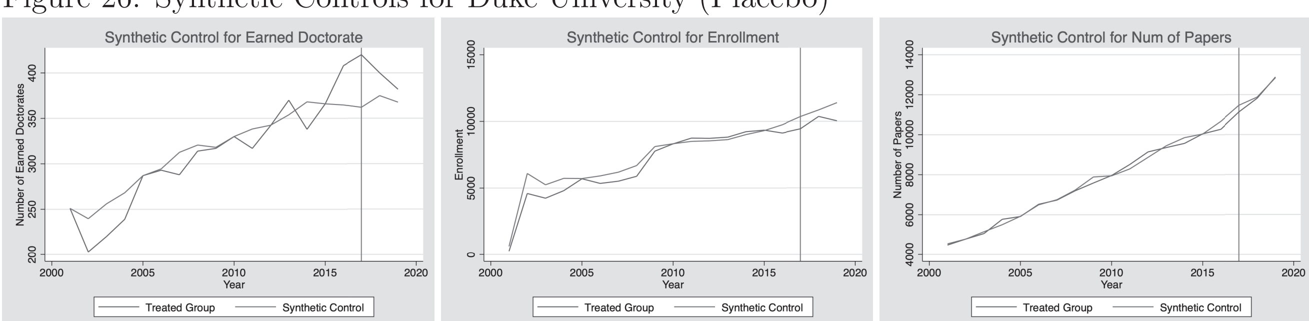


Figure 26: Synthetic Controls for Duke University (Placebo)



References

Sean Wang and Samuel Young. Unionization, employer opposition, and establishment closure. 2021.

John Dinardo and David S. Lee. Economic impacts of new unionization on private sector employers: 1984-2001. *Quarterly Journal of Economics*, page 1383–1441, 2004.

Richard B. Freeman and Morris M. Kleiner. Do unions make enterprises insolvent? *ILR Review*, 52(4):510–527, 1999. doi: 10.1177/001979399905200401. URL https://doi.org/10.1177/001979399905200401.

Aaron J. Sojourner, Brigham R. Frandsen, Robert J. Town, David C. Grabowski, and Min M. Chen. Impacts of unionization on quality and productivity: Regression discontinuity evidence from nursing homes. *ILR Review*, 68(4):771–806, 2015. doi: 10.1177/0019793915586380. URL https://doi.org/10.1177/0019793915586380.

Robert J. Lalonde, Gérard Marschke, and Kenneth Troske. Using Longitudinal Data on Establishments to Analyze the Effects of Union Organizing Campaigns in the United States. *Annals of Economics and Statistics*, (41-42):155–185, 1996. URL https://ideas.repec.org/a/adr/anecst/y1996i41-42p155-185.html.

Arindrajit Dube, Ethan Kaplan, and Owen Thompson. Nurse unions and patient outcomes. *ILR Review*, page 803–833, 2016.

Alexandre Mas. Pay, reference points, and police performance. *The Quarterly Journal of Economics*, 236:783–821, 2006.

William A. Herbert, Jacob Apkarian, and Joseph van der Naald. Supplementary directory of new bargaining agents and contracts in institutions of higher education, 2013-2019. 2020.

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

Alberto Abadie, Dimitri P. Bertsekas, and David A. Castanon. Synthetic control methods for comparative case studies: Estimating the effect of california's tobacco control program. *Journal of the American Statistical Association*, pages 493–505, 2010.

Brigham R. Frandsen. Party bias in union representation elections: Testing for manipulation in the regression discontinuity design when the running variable is discrete. *Advances in Econometrics*, 38:281–315, 2017.

Salary Considerations in Matching Markets within the Medical Residency System

Sofie Kupiec[1]

[1]**skupiec@mit.edu**

ABSTRACT

Within this paper, I propose an algorithm that adds a salary consideration to the preexisting doctor-proposing-deferred-acceptance algorithm. This algorithm allows for applicants to "propose" to hospitals with a certain salary-ask and adds budget to a constraint for hospitals. The algorithm/model I propose is stable and changes assignments when budgets are the constraining factor. However, I believe it will benefit less desirable hospitals more when we allow hospitals to deem certain salaries unacceptable.

INTRODUCTION

Each year, the Medical Residency System matches graduating medical students with hospital programs for internships. At the inception of medical internships in 1900, there was high competition for applicants and the system in place was wildly inefficient. Thus, in 1950, an algorithm was proposed that matched applicants with hospital programs through a central clearing house. The goal was to create a stable matching between applicants and programs so these previous inefficiencies would be eliminated. This trial run algorithm was implemented in 1951, was improved upon the following year, and has since been improved upon to better represent some nuances of the market.

Within my survey article, I explored some of the literature that has been published since that initial trial run algorithm, each detailing algorithms to produce a stable matching. Through that exploration, it stood out to me that within stable matching algorithms the same set of hospital programs and applicants who are unassigned by the algorithm are left unassigned within every possible stable matching; this has become known as the rural hospital theorem because less-desirable hospitals remain unmatched. Thus, hospitals in preferable areas (typically urban areas) get the better applicants and hospitals in less-desired areas (typically rural and inner city areas) sometimes even get no matches. In my opinion, this perpetuates an unbalanced health care system. It seems as though this would be a cycle of certain hospitals attracting the best talent, becoming an even better program due to having the top talent, and then continuing to attract the best talent. While on the other hand, other hospitals get little to no talent which could lead to not being an attractive program the next year and perpetuating that negative cycle.

It is my understanding that hospitals and applicants negotiate salaries outside of the central clearing house (or the salaries are nonnegotiable). However, I would like to explore if we could include this within the matching algorithm. My hope is that it would permit typically less desirable hospitals to attract more talent.

THE MODEL

Players and Constraints

Let us consider a game, much like that used within Roth's "The Evolution of the Labor Market for Medical Interns and Residents: A Case Study in Game Theory"(Roth, 1984), whose players are a set $H = \{h_1, \cdots, h_m\}$ hospital programs offering positions to first-year graduates and a set $S = \{s_1, \cdots, s_n\}$ students. Each hospital program h_i has some capacity q_i. Also, let each hospital have some budget b_i that can be spent on salaries for interns. Henceforth we will refer to hospital programs simply as hospitals. Let us also introduce the set $A = \{a_1, \cdots, a_l\}$ (where $a_i < a_{i+1} \forall i \in l$) of allowed salary amounts that a hospital can pay an intern.

Preferences

We will use the preference notation used in stable matching literature where $cP(s_j)b$ means that student s_j prefers match c over match b. Also, note that $cP(s_j)u$ means that s_j finds c to be an acceptable match and $uP(s_j)c$ means that s_j finds c to be an unacceptable match. This notation extends to hospitals as well, where $s_jP(h_i)s_k$ means that hospital h_i prefers student s_j to student s_k.

Each hospital h_i has strict preferences over the set of all students it finds acceptable: $\{s_j \in S \mid s_jP(h_i)u\}$. Each student s_j has strict preferences over the $O(l \cdot n)$ set of hospital salary pairs the student finds acceptable: $\{(h_i, a_v) \in H, A \mid (h_i, a_v)P(s_j)u\}$.

Assumptions

We assume all students to be rational in that a student s_j will always prefer a hospital h_i at a higher salary than a lower salary: $(h_i, a_v)P(s_j)(h_i, a_w) \quad \forall a_v > a_w \in A$. Since each hospital is simply making preferences over applicants (rather than applicant-salary pairs), for the purposes of the algorithm (which I will soon explain) the strong assumption is made that if $s_jP(h_i)s_k$ and $s_jP(h_i)s_r$ then $s_jP(h_i)\{s_k, s_r\}$. This is a pretty strict assumption that I will touch upon more later in the paper.

Outcome

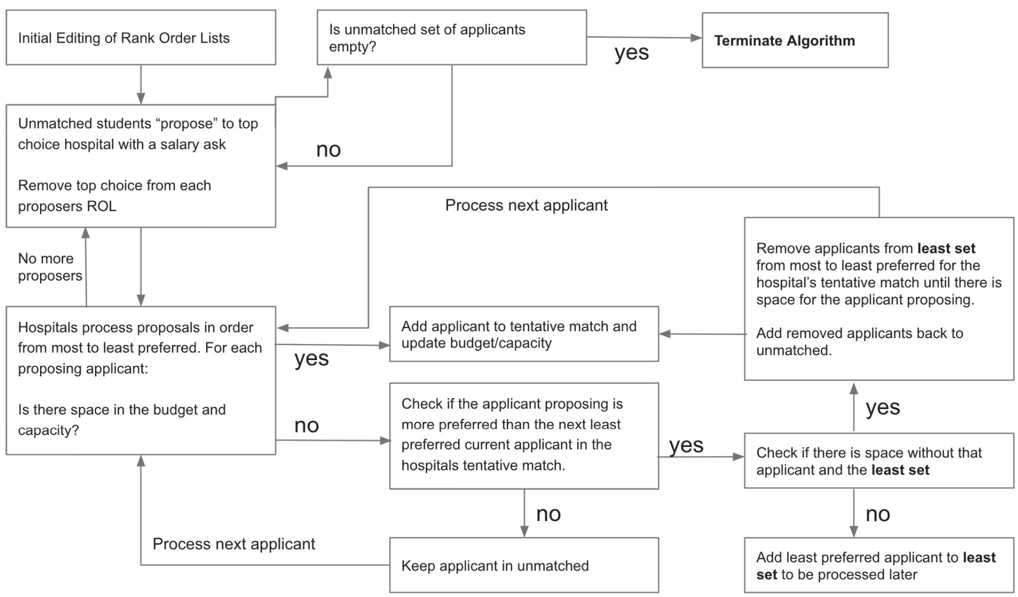
An outcome is represented by the function x that maps the set of students to the set of hospitals (and unassigned) where each student is mapped with a corresponding salary they will be paid. For any student s_j, $x(s_j) = (h_i, a_v)$ denotes that s_j is assigned to a position in hospital program h_i for a salary a_v. Let $x(s_j) = u$ denote that student s_j is unassigned. For any hospital program h_i in H, let $x(h_i) = \{(s_j, a_v) \in S \mid x(s_j) = (h_i, a_v)\}$ be the set of student-salary pairs assigned to h_i with the constraint that $|x(h_i)| \leq q_i$ and $\sum_{a_v \forall (s_j, a_v) \in x(h_i)} a_v \leq b_i$. Let $|x(h_i)|_q$ be defined as the leftover capacity of h_i in outcome x and $|x(h_i)|_b$ be defined as the leftover budget in outcome x.

Stability

An outcome x is unstable if for some student s_j, $uP(s_j)x(s_j)$; for some hospital h_i, $uP(h_i)(s_j, a_v)$, for some (s_j, a_v) in $x(h_i)$; or if there exists some hospital h_i and student s_j such that for some salary a_v (i)$(h_i, a_v)P(s_j)x(s_j)$ and (ii) $(s_j, a_v)P(h_i)\sigma$ where $\sigma \subseteq x(h_i)$ such that $|x(h_i)|_b + \sum_{(s_j, a_w) \in \sigma} a_w \geq a_v$ the leftover budget plus the total salary of the set of applicants in sigma is greater than or equal to a_v. So, no hospital or student can prefer being unassigned than to an assignment given in an outcome. And, no hospital-student pair can have a student prefer the hospital at certain salary over his current assignment and have a hospital prefer a student at said certain salary over a subset of current students assigned to the hospital and the sum of their allocated salaries plus the leftover budget.

THE ALGORITHM

At a high level, the algorithm works very similar to the Doctor-Proposing Deferred Acceptance Algorithm (DAA). However, rather than just being constrained by a capacity q_i, a hospital is also constrained by a budget b_i. And, rather than just simply ”proposing” to a hospital, at each step, each applicant proposes with a specific salary-ask.



Pre-processing

Much like the doctor-proposing DAA, each student and hospital submit their preferences. Again, the hospital will submit preferences over the subset of students it finds acceptable and marks each student it finds unacceptable. Students will submit preferences over the set of hospital-salary pairs the student finds acceptable and marks each hospital-salary pair it finds unacceptable. The lists will then be edited to remove each hospital-salary from a student's rank order list if said hospital finds the student unacceptable. In a similar fashion, each student will removed from a hospital's rank order list if the student finds the hospital unacceptable at ALL salaries.

These lists are then processed in what is similar to an applicant-proposing-deferred-acceptance problem but, rather than just having a capacity constraint, the hospital program has an additional budget constraint. And, applicants are proposing with a salary ask. We begin with a tentative matching where every hospital has no applicants matched to it, each applicant is unmatched, each hospital h_i has an initial available budget of b_i, each hospital h_i has an initial available capacity of q_i, and the algorithm updates these as follows:

Proposal Step

At any step, if the applicants rank order list is empty (they have no hospital to propose to), assign the applicant to the unassigned set and removed them from the set of unmatched students. All applicants within the unmatched set ”propose” to their top choice hospital-salary pair from their rank order list. Update the applicants’ rank order lists by removing the top choice pair from their lists.

For each hospital, the hospital processes its proposals in order of the hospitals rank order list (most preferred first) and checks if the proposer-salary pair fits within its available budget and capacity constraints. If it does, it adds the proposer-salary pair to its tentative assignment, updates its budget and capacity constraints, removes the proposer from the set of unmatched applicants, and moves on to the next proposer-salary pair. If it does not, it moves on to the ”least set” step.

Once there are no more proposals to process, repeat the proposal step until there are no more students in set of unmatched students. The algorithm terminates when the set of unmatched students is empty and returns the final matching.

Least Set Step

We introduce this empty set called the *least set* as well as a *tentative budget* and *tentative capacity* that we will use to track changes to the hospitals tentative matching, budget, and capacity without actually changing the current matching. Since the proposer does not fit in the hospital, the algorithm will look at each applicant in the hospital’s tentative matching from least to most preferred.

Backwards Pass Through Tentative Matching Step

At each applicant, it will check if the applicant is less preferred than the proposing student. If it is not, leave the proposer in the unmatched set of students and continue on to processing the next proposer in the proposal step. However, if so, add the applicant to the *least set*. It will then add back the amount allocated to his salary to the *tentative budget* and the space he took up (1) back to the *tentative capacity*. Check if the proposer fits in the hospital given the new *tentative budget* and *tentative capacity*. If it does, go to the ”Forward Pass Through Least Set Step”. If it does not, continue the backwards pass.

Forward Pass Through Least Set Step

Here, the algorithm knows that the proposer is more preferred than all applicants within this *least set*. Thus, we know we want to remove members from this set to create space (both budget and capacity) for the proposing applicant. We know that the salaries a hospital takes more preferred applicants in a matching is greater than or equal to salaries for a lesser preferred applicant (See **Lemma 1** in Further Explanation of Least Set below) and we know that we have just exceeded the threshold for necessary space with the addition of the most preferred applicant to the *least set*. So, we will pass through from most to least preferred, each time removing the applicant from the actual tentative matching for the hospital (adding said applicant to the unmatched set) and updating the available budget/capacity. After removing each applicant, the algorithm checks if there is now space for the proposer in the hospital, and if there is the proposer is added to the tentative match (and update the budget/capacity accordingly) and if not it will go on to remove the next applicant in the *least set*. We know that the proposer will eventually be added by removing some subset or the entirety of the *least set* due to the nature of creating the *least set*. Once the proposer has been added, return back to the proposal step to process the next proposer.

Further Explanation of Least Set

Lemma 1: If $s_j P(h_i) s_k$ and $\{(s_j, a_{s_j}), (s_k, a_{s_k})\} \subseteq x(h_i)$, then $a_{s_j} \geq a_{s_k}$.
Proof of Lemma 1: Proof by Contradiction.
Let us consider an outcome in which $s_j P(h_i) s_k$,$\{(s_j, a_{s_j}), (s_k, a_{s_k})\} \subseteq x(h_i)$ and $a_{s_j} \leq a_{s_k}$. There are two orders in which this can occur: (1) s_j proposed before s_k or (2) the other way around. In order (1), give our assumption that $(h_i, a_{s_k}) P(s_i) (h_i, a_{s_j})$, we know that h_i needed to reject s_j at salary a_{s_k}. Thus, if s_k proposes after s_j at salary a_{s_k}, h_i must also reject s_k at salary a_{s_k} because $s_j P(h_i) s_k$. And so, this outcome is not possible in this proposal ordering. In order (2), if s_k proposes to h_i and is accepted at salary a_{s_k}, since $s_j P(h_i) s_k$, if s_j proposes at a_{s_k} after s_k (which s_j must do before proposing at a_{s_j}), the algorithm will replace s_k if there is inadequate space for them both or add s_j at a_{s_k} if there is space. Thus, it is impossible for $a_{s_k} \geq a_{s_j}$ and we have a contradiction.

Importance of Processing Least Set in this Manner

Consider a scenario with the following preferences:

s_1	s_2	s_3	s_4
$(h_2, \$125)$	$(h_1, \$125)$	$(h_1, \$125)$	$(h_1, \$125)$
$(h_1, \$125)$	$(h_2, \$125)$	$(h_1, \$100)$	$(h_1, \$100)$
$(h_2, \$100)$	$(h_2, \$100)$	$(h_2, \$125)$	$(h_2, \$125)$
$(h_1, \$100)$	$(h_1, \$100)$	$(h_2, \$100)$	$(h_2, \$100)$

	h_1	h_2
	s_1	s_4
	s_2	s_2
	s_3	s_3
	s_4	s_1
b	\$225	\$225
q	2	2

At a certain point you will have the scenario:

	Tentative Match
h_1:	$(s_2, \$125)$, $(s_3, \$100)$
h_2:	$(s_4, \$125)$

Budgets/Capacity

	h_1	h_2
b:	\$0	\$100
q:	0	1

Without the processing the *least set* in the way we do, and you kicked out everyone required to create space, the algorithm will progress as follows:

	New Pointers
h_1:	$(s_1, \$125)$
h_2:	

	Tentative Match
h_1:	$(s_1, \$125)$
h_2:	$(s_4, \$125)$

Budgets/Capacity

	h_1	h_2
b:	\$100	\$100
q:	1	1

New Pointers

h_1:	
h_2:	$(s_2, \$125)$, $(s_3, \$125)$

Reject both

New Pointers

h_1:	
h_2:	$(s_2, \$100)$, $(s_3, \$100)$

Tentative Match

h_1:	$(s_1, \$125)$
h_2:	$(s_4, \$125)$, $(s_2, \$100)$

Budgets/Capacity

	h_1	h_2
b:	\$100	\$0
q:	1	0

Final Matching

h_1:	$(s_1, \$125)$
h_2:	$(s_4, \$125)$, $(s_2, \$100)$
u:	s_3

Leftover Budgets/Capacity

	h_1	h_2
b:	\$100	\$0
q:	1	0

This is *unstable* as $s_3 P(h_1) u$ and $(h_1, 100) P(s_3) u$. Thus, it is imperative we process the least set from most to least preferred to avoid this inefficiency.

FORMAL RESULTS

Theorem 1: Given the preferences submitted by students and hospitals, the Algorithm produces a stable matching.

Proof of Theorem 1: Recall there are 3 scenarios that would make an outcome unstable: (1) no hospital can prefer having a vacancy over a student assigned it it in the outcome; (2) no student can prefer being unassigned over his assignment given in the outcome; (3) no hospital-student pair can have a student prefer the hospital at certain salary over his current assignment and have a hospital prefer a student at said certain salary over a subset of current students assigned to the hospital and the sum of their allocated salaries. Given the pre-processing of the rank order lists of the hospitals and students, where hospitals are removed from a student's preferences that deemed the student unacceptable and vise versa, the algorithm will not ever reach an outcome that is scenario (1) or (2). Applicants will never propose to hospitals that deem them unacceptable and applicants will never propose to hospitals they deem unacceptable.

Let us examine the third scenario more closely. Define outcome x to be an outcome produced by the algorithm. Let $x(s_j)$ be the assignment of s_j within outcome x. And let $x(h_i)$ be the set of applicants assigned to h_i within outcome x. Let outcome x be unstable as $(h_i,a_v)P(s_j)x(s_j)$ and $(s_j,a_v)P(h_i)\sigma$ where $\sigma \subseteq x(h_i)$ such that $\sum_{(s_j,a_w)\in\sigma} a_w \geq a_v$. We know that s_j finds (h_i,a_v) acceptable and h_i finds s_j acceptable and thus will be in each other's rank order lists.

Within the algorithm, each applicant proposes down his rank order list. So, if $(h_i,a_v)P(s_j)x(s_j)$, we know that s_j will propose to (h_i,a_v) prior to $x(s_j)$. In order to subsequently propose to and be accepted by $x(s_j)$, (s_j,a_v) must be rejected by h_i. We defined σ such that the cost of salaries for all applicants in σ is greater than or equal to a_v. If $(s_j,a_v)P(h_i)\sigma$, by definition of the algorithm, h_i would remove/reject all applicants within σ to make space/hold on to s_j since the cost of $\sigma \geq a_v$. And so s_j's proposal is accepted and held on to by h_i. Thus, we have a contradiction in that if $(h_i,a_v)P(s_j)x(s_j)$ and $(s_j,a_v)P(h_i)\sigma$, the algorithm will not assign s_j to $x(s_j)$ and σ to h_i because s_j will be held on to by h_i.

Theorem 2: When, for all h_i, $b_i \geq a_l * q_i$, the assignment produced by the algorithm is the same assignment produced by the stable matching. (Assuming that a student's preferences at the highest salary tier is his true overall preferences.)
Proof of Theorem 2: We know that if $b_i \geq a_l * q_i$, a matching $x(h_i)$ will not be constrained by budget, as capacity will be filled before the budget constraint can even be reached. For each h_i, we know that s_j will first propose to h_i with salary a_l, given our assumption that $(h_i,a_l)P(s_i)(h_i,a_w)$ where $a_w < a_l$. Given that budget is not a constraining factor, h_i will tentatively hold onto its up to top q_i students it can fit in its capacity. And if h_i rejects a s_j, that means that h_i is already tentatively holding on to q_i better candidates than s_j and will reject s_j at any lower salary tier.

Theorem 3: Hospitals have the possibility of achieving a better outcome with strategic truncation thus making the algorithm not strategy proof.
Proof of Theorem 3: Consider a simple scenario with 2 salary tiers, 3 hospitals, and 3 applicants with the following preferences:

s_1	s_2	s_3
(h_3,a_2)	(h_2,a_2)	(h_1,a_2)
(h_3,a_1)	(h_2,a_1)	(h_1,a_1)
(h_1,a_2)	(h_1,a_2)	(h_2,a_2)
(h_1,a_1)	(h_1,a_1)	(h_2,a_1)
(h_2,a_2)	(h_3,a_2)	(h_3,a_2)
(h_2,a_1)	(h_3,a_1)	(h_3,a_1)

	h_1	h_2	h_3
	s_1	s_1	s_1
	s_2	s_3	s_2
	s_3	s_2	s_3
b	a_2	a_2	a_2
q	1	1	1

With these preferences, we will have the following matching:

$$\begin{aligned} h_1&: \ (s_3,a_2) \\ h_2&: \ (s_2,a_2) \\ h_3&: \ (s_1,a_2) \end{aligned}$$

However, consider if h_1 were to truncate its preferences:

s_1	s_2	s_3
	(h_2,a_2)	(h_1,a_2)
	(h_2,a_1)	(h_1,a_1)
(h_1,a_2)	(h_1,a_2)	(h_2,a_2)
(h_1,a_1)	(h_1,a_1)	(h_2,a_1)
(h_2,a_2)	(h_3,a_2)	(h_3,a_2)
(h_2,a_1)	(h_3,a_1)	(h_3,a_1)

	h_1	h_2	h_3
	s_1	s_1	s_1
	s_2	s_3	s_2
		s_2	s_3
b	a_2	a_2	a_2
q	1	1	1

With these preferences, we will have the following matching:

$$\begin{array}{rc} h_1\colon & (s_2,a_2) \\ h_2\colon & (s_3,a_2) \\ h_3\colon & (s_1,a_2) \end{array}$$

Thus, h_1 was able to improve its match with strategic truncation. However, there is also the possibility of over truncation. Consider the following preferences:

s_1	s_2	s_3
	(h_2,a_2)	(h_1,a_2)
	(h_2,a_1)	(h_1,a_1)
(h_1,a_2)		(h_2,a_2)
(h_1,a_1)		(h_2,a_1)
(h_2,a_2)	(h_3,a_2)	(h_3,a_2)
(h_2,a_1)	(h_3,a_1)	(h_3,a_1)

	h_1	h_2	h_3
	s_1	s_1	s_1
		s_3	s_2
		s_2	s_3
b	a_2	a_2	a_2
q	1	1	1

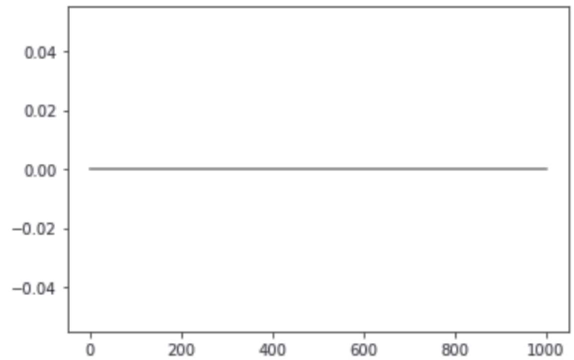
With these preferences, we will have the following matching:

$$\begin{array}{rc} h_1\colon & \emptyset \\ h_2\colon & (s_3,a_2) \\ h_3\colon & (s_1,a_2) \\ u\colon & s_2 \end{array}$$

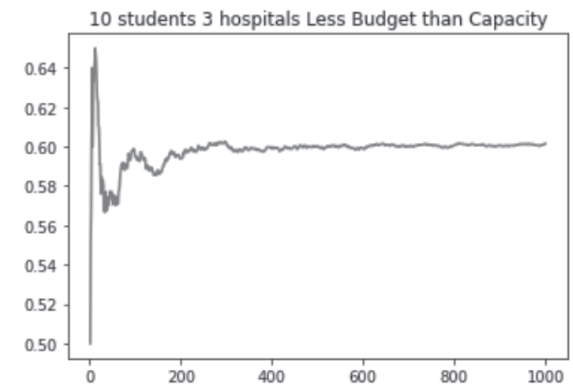
In reality, $s_2 P(h_1) u$ but, due to over-truncation of preferences, h_1 is left without a match and s_2 is left unassigned. Thus, though strategic truncation is possible, hospitals run the risk of having a worse-off outcome so it is not advised.

SIMULATIONS

Demonstrating **Theorem 2**, I ran simulations with various hospital-student ratios and budgets of hospitals all exceeding the capacity threshold. Each time, regardless of the randomly generated preferences, number of hospitals/students, etc, this graph was produced:



Additionally, I ran simulations to see the difference in students assignments when budget was less than capacity and a high number of students received a different assignment.



I ran more simulations, however, I believe that more interesting, and realistic, results will be produced when we allow for hospitals to deem salaries unacceptable. I will discuss this further in my future work section.

CONCLUSION AND FUTURE WORK

Through this exploration, I have come up with a way to handle salary in the medical residency problem. However, I think that there is more that can be done to help undesirable hospitals through the way hospitals can report their preferences. The next step I would like to explore would be to consider salary bounds that a hospital is willing to pay, as the current preference format does not consider salary from the hospital side. My hypothesis is that if hospitals were able to deem salary tiers unacceptable, rather than just students, there would be more opportunity for less desirable hospitals to get matches at higher salary tiers. I believe this would be the case as right now more desirable hospitals are still preferred at the higher salary tiers. But, if some high-desirable hospitals found the higher salary tiers unacceptable, less-desirable hospitals would find themselves in lower-ranked positions by students. For example, if there are 3 salary tiers and 2 hospitals we could have the following preferences:

$$s_1: \quad (h_1,a_3),\, (h_2,a_3),\, (h_1,a_2),\, (h_2,a_2),\, (h_1,a_1),\, (h_2,a_1)$$

However, perhaps h_1 only finds salaries a_2 and a_1 acceptable:

$$s_1: \quad (h_2,a_3), (h_1,a_2), (h_2,a_2), (h_1,a_1), (h_2,a_1)$$

We can see how this would help h_2 who is willing to pay more for applicants. Additionally, this would help mitigate the affect of the strict assumption that if $s_jP(h_i)s_k$ and $s_jP(h_i)s_r$ then $s_jP(h_i)\{s_k,s_r\}$. For example, if a hospital h_i has budget $b_i = a_l$, this could lead to h_i only being matched to one applicant who will get a salary a_l. However, if h_i could deem certain budgets unacceptable, this could ensure that a scenario like this would not occur. I believe this would be an easy addition to the problem and would allow for the algorithm to stay the same. My intuition is that the only thing this would affect is Theorem 2 as the algorithm would not necessarily produce the same outcome as the stable matching.

APPENDIX

Algorithm Psedocode

Algorithm 1 Deferred Acceptance with Additional Salary Constraint

$b[h_i] \leftarrow$ available budget of $h_i \quad \forall h_i \in H$
$q[h_i] \leftarrow$ available capacity of $h_i \quad \forall h_i \in H$
$T[h_i] \leftarrow \{\} \quad \forall h_i \in H$ ▷ tentative assignment starts empty for all h_i
unmatched $\leftarrow \{s_j \in S\}$ ▷ all students start unmatched
while $|\text{unmatched}| > 0$ **do**
  for $s_j \in$ unmatched **do**
    if s_j his ROL is empty ($\forall s_j \in$ unmatched) **then**
      Assign s_j to the unassigned set
      Remove s_j from the unmatched set
    else
      Have all $s_j \in$ unmatched point to their top-choice (hospital,salary) pair
      Remove top choice pair from preferences of s_j
    end if
  end for
  for $h_i \in H$ **do**
    for each (s_j,a_v) pointing to h_i **do** ▷ Process in order of h_i's preferences over s_j's
      if $b[h_i] - a_v \geq 0$ AND $q[h_i] - 1 \geq 0$ **then**
        Add (s_j,a_v) to $T[h_i]$
        Subtract a_v from $b[h_i]$and subtract 1 from $q[h_i]$
        Remove s_j from unmatched
      else
        least_set $\leftarrow \{\}$
        tentative_$b_i \leftarrow b[h_i]$ ▷ to be tentatively updated in the next steps
        tentative_$q_i \leftarrow q[h_i]$
        for $(s_k,a_w) \in T[h_i]$ **do** ▷ starting from least preferred to most
          if $s_jP(h_i)s_k$ **then**
            Add a_w to tentative_b_i
            Add 1 to tentative_q_i

```
                    Add (s_k, a_w) to least_set
                    if tentative_b_i − a_v ≥ 0 and tentative_q_i + 1 ≥ 0 then
                        for (s_k, a_w) ∈ least_set do   ▷ starting from most preferred to least
                            Remove (s_k, a_w) from T[h_i]
                            Add a_w to b[h_i] and add 1 to q[h_i]
                            Add s_k to unmatched
                            if b[h_i] − a_v ≥ 0 AND q[h_i] − 1 ≥ 0 then
                                Add (s_j, a_v) to T[h_i]
                                Subtract a_v from b[h_i] and subtract 1 from q[h_i]
                                Remove s_j from unmatched
                            end if
                        end for
                    else
                        Continue to next pair                ▷ Keep s_j in unmatched
                    end if
                else
                    Break out of for loop
                end if
            end for
        end if
        Continue to next pair
    end for
    Continue to next hospital
  end for
end while                          ▷ Terminates when no students are unmatched
return T
```

Worked out Example

Consider a scenario with the following preferences:

s_1	s_2, s_3, s_4
$(h_2, \$125)$	$(h_1, \$125)$
$(h_1, \$125)$	$(h_1, \$100)$
$(h_2, \$100)$	$(h_2, \$125)$
$(h_1, \$100)$	$(h_2, \$100)$

	h_1	h_2
	s_1	s_3
	s_2	s_2
	s_3	s_4
	s_4	s_1
b	\$225	\$225
q	2	2

The algorithm will work as follows:

Step 1

	Pointers
h_1:	$(s_2, \$125)$, $(s_3, \$125)$, $(s_4, \$125)$
h_2:	$(s_1, \$125)$

Tentative Match

h_1:	$(s_2, \$125)$
h_2:	$(s_1, \$125)$

Budgets/Capacity

	h_1	h_2
b:	\$100	\$100
q:	1	1

Step 2

Pointers

h_1:	$(s_3, \$100)$, $(s_4, \$100)$
h_2:	

Tentative Match

h_1:	$(s_2, \$125)$, $(s_3, \$100)$
h_2:	$(s_1, \$125)$

Budgets/Capacity

	h_1	h_2
b:	\$0	\$100
q:	0	1

Step 3

Pointers

h_1:	
h_2:	$(s_4, \$125)$

Note: Not enough space to simply add s_4 to h_2 so the algorithm add applicants currently tentatively assigned to h_2 from least preferred to most to the ”least set” until there is enough space to add s_4 if s_4 is more preferred to all the applicants needed to be removed

Least Set

Least Set:	$(s_1, \$125)$
b without least set:	\$225
q without least set:	2

Note: Remove applicants from least set from most preferred to least until there is enough space to add s_4

Tentative Match

h_1:	$(s_2, \$125)$, $(s_3, \$100)$
h_2:	$(s_4, \$125)$

Budgets/Capacity

	h_1	h_2
b:	\$0	\$100
q:	0	1

Step 4

Pointers

h_1:	$(s_1, \$125)$
h_2:	

Note: Not enough space to simply add s_1 to h_1 so the algorithm add applicants currently tentatively assigned to h_1 from least preferred to most to the "least set" until there is enough space to add s_1 if s_1 is more preferred to all the applicants needed to be removed

Least Set

Least Set:	$(s_3, \$100)$
b without least set:	\$100
q without least set:	1

Least Set

Least Set:	$(s_2, \$125)$, $(s_3, \$100)$
b without least set:	\$225
q without least set:	2

Note: Remove applicants from least set from most preferred to least until there is enough space to add s_1. Only need to remove $(s_2, \$125)$.

Tentative Match

h_1:	$(s_1, \$125)$, $(s_3, \$100)$
h_2:	$(s_4, \$125)$

Budgets/Capacity

	h_1	h_2
b:	\$0	\$100
q:	0	1

Step 5

Pointers

h_1:	$(s_2, \$100)$
h_2:	

Least Set

Least Set:	$(s_3, \$100)$
b without least set:	\$100
q without least set:	1

Tentative Match

h_1:	$(s_1, \$125)$, $(s_2, \$100)$
h_2:	$(s_4, \$125)$

Budgets/Capacity

	h_1	h_2
b:	\$0	\$100
q:	0	1

Step 6

Pointers

h_1:	
h_2:	$(s_3, \$125)$

Least Set

Least Set:	$(s_4, \$125)$
b without least set:	\$225
q without least set:	2

Tentative Match

h_1:	$(s_1, \$125)$, $(s_2, \$100)$
h_2:	$(s_3, \$125)$

Budgets/Capacity

	h_1	h_2
b:	\$0	\$100
q:	0	1

Step 7

Pointers

h_1:	
h_2:	$(s_4, \$100)$

Tentative Match

h_1:	$(s_1, \$125)$, $(s_2, \$100)$
h_2:	$(s_3, \$125)$, $(s_4, \$100)$

Budgets/Capacity

	h_1	h_2
b:	\$0	\$0
q:	0	0

Algorithm Terminates as all applicants are assigned.

Final Match

h_1:	$(s_1, \$125)$, $(s_2, \$100)$
h_2:	$(s_3, \$125)$, $(s_4, \$100)$

Code to Make Matching

```
def make_assignment(students, student_preferences, hospitals, hospital_preferences, budgets, capacity):
    '''
    students - list of students ints [1,2,3,4]
    student_prefernces - Dict mapping students to list of preferences [(h (str), salary (int))]
    hospitals - list of hospitals (str)
    hospital_preferences - list of perferences over students (int)
    budgets - Dict mapping hospitals to budget (int)
    capacity - Dict mapping hospitals to capacity (int)
    '''
    unassigned = set(students)
    tenative_assignment = {
    'u': []
    }
    for j in hospitals:
        tenative_assignment[j] = []


    ##############
    while unassigned:
        #INITIALIZE STUDENT POINTER w prefs
        student_pointer = {}
        empty = set()
        for x in unassigned:
            if len(student_preferences[x]) == 0:
                tenative_assignment['u'].append(x)
                empty.add(x)
            else:
                student_pointer[x] = student_preferences[x].pop(0)
        unassigned = unassigned - empty
        ##############
        ##############
        #Proposals to each hospital are in list and asking price is in the student
        proposals = {}
        for j in hospitals:
            proposals[j] = []

        for x in unassigned:
            tup = student_pointer[x]
            h = tup[0] #hosptial
            proposals[h].append(x)
```

```
            proposals[x] = tup[1] #asking price
        #############

        for h in hospitals:
            order = {}
            options = proposals[h]
            order = {k:i for i,k in enumerate(hospital_preferences[h])}
            options.sort(key=order.get)
            for x in options:
                ask = proposals[x] #asking salary
                #if there is space for x
                if budgets[h] - ask >= 0 and capacity[h] - 1 >= 0:
                    flag=True
                    for i, y in enumerate(tenative_assignment[h]):
                        if order[y[0]] > order[x]:
                            tenative_assignment[h].insert(i,(x,ask))
                            flag=False
                            break
                    if flag:
                        tenative_assignment[h].append((x,ask))
                    budgets[h] -= ask
                    capacity[h] -= 1
                    unassigned.remove(x)
                else:
                    least_set = []
                    ten_bud = budgets[h]
                    ten_cap = capacity[h]
                    # iterate backwards through tenative assignemnet
                    space = False
                    for i in range(len(tenative_assignment[h])-1,-1,-1):
                        #look at each applicant
                        app = tenative_assignment[h][i]
                        #if x is more preferred to applicant
                        if order[app[0]] > order[x]:
                            ten_bud += app[1]
                            ten_cap += 1
                            least_set.append(app)
                            #if removing that applicant creates enough space
                            if ten_bud - ask >= 0 and ten_cap - 1 >= 0:
                                least_set.reverse()
                                for mat in least_set:
                                    tenative_assignment[h].remove(mat)
                                    budgets[h] += mat[1]
                                    capacity[h] += 1
                                    unassigned.add(mat[0])
                                    if budgets[h] - ask >= 0 and capacity[h] - 1 >= 0:
                                        flag=True
                                        #placing x in tenative assignment
                                        for i, y in enumerate(tenative_assignment[h]):
                                            if order[y[0]] > order[x]:
                                                tenative_assignment[h].insert(i,(x,ask))
                                                flag=False
                                                break
                                        if flag:
                                            tenative_assignment[h].append((x,ask))
                                        #update
                                        budgets[h] -= ask
                                        capacity[h] -= 1
                                        unassigned.remove(x)
                                        space = True
                                        break
                        else:
                            break
                        if space:
                            break
    return tenative_assignment, budgets, capacity
```

REFERENCES

Roth, A. E. (1984). The evolution of the labor market for medical interns and residents: A case study in game theory. *Journal of Political Economy*, (6):991–1016.

Impact of Implementing Weighted School Funding on High School Educational Attainment

Alice Martynova

February 2023

Abstract

Since 1993, many large school districts across the United States have shifted away from deploying federal funds to schools based on uniform staffing formulas and towards weighted school funding (WSF), which deploys a fixed-dollar amount to schools for each student type with larger increments going to students from low-income backgrounds, with special needs, and/or who are English-language learners. Using publicly available NCES data, I study the impact of WSF on high school graduation rates, dropout rates, and pupil-per-teacher ratio. The difference-in-difference empirical strategy finds that WSF has a limited statistical significant impact on any of these educational attainments suggesting that WSF's effects still need to be further studied to fully understand the power and drawbacks of this new and growing funding schematic.

1 Introduction

Over the last two decades, many large school districts across the United States have shifted away from deploying federal funds to districts based on uniform staffing formulas to allocating funds to schools based on the particular mix of students within a school. This new allocation strategy, known as weighted student funding (WSF), deploys a fixed-dollar amount to schools for each student type with larger increments going to students from low-income backgrounds, with special needs, and/or who are English-language learners. It is important to note that WSF does not change how much money a district receives, but rather how the money is allocated amongst the schools within the district.

New funding methods, like WSF, have the potential to fight the poverty cycle, reduce inequality, and have significant effects on student educational outcomes (Johnson and Jackson

2019) [9]. In this paper, I will study how WSF affects high school educational attainments specifically graduation rates, dropout rates, and pupil-per-teacher ratio. Although WSF was first implemented in 1995, this hypothesis has yet to be tested due to the limited number of schools that have adopted WSF and more so the lack of public data, but available data creates robust empirical tests that allow us to begin to understand WSF's potential.

To analyze the effect of WSF on high school district graduation rates, dropout rates, and pupil-per-teacher ratios, this paper implements a staggered difference-in-difference for each educational attainment to compare outcomes at control schools that never implemented WSF to treated districts that implemented WSF between 1995-2018. The treated group consists of the 27 WSF school districts documented in the U.S. Department of Education report while the control districts are chosen from the NCES yearly table of "Selected statistics on enrollment, teachers, dropouts, and graduates in public school districts enrolling more than 15,000 students." By only selecting districts from this category, I guarantee districts have similar sizes and are nationally representative which leads to more robust results.

Ultimately this paper finds minute, positive but mostly insignificant effects of WSF on graduation rates and pupil-per teacher ratio and inconclusive negative effects on dropout rates. Following treatment, the pupil-per-teacher ratio and graduation rates remain unchanged relative to the pre-treatment mean and both effects are statistically insignificant. In addition to NCES documented graduation rates, this paper creates and studies its own statistic, "pseudo-graduation" rate, which is equivalent to the number of graduates divided by total district enrollment. The purpose of this statistic is to further the understanding of the effect on graduation rates and be able to consider the effect on total district enrollment. WSF increases pseudo-graduation rates by about 0.3 percentage points which is an overall 6% increase, but this is only at the 10% significance level and must be interpreted with caution. Finally, WSF appears to decrease dropout rates by 1 percentage point which is an overall 11% decrease, but this result must also be interpreted with caution as the dropout rate regression does not satisfy the parallel trend assumption (further discussed in Section 4 and 5). Although there is no apparent sizable and significant effect, WSF does not negatively affect any educational attainment at the district level which questions theories that WSF has an overall harmful impact.

The current WSF literature only focus on describing the WSF model and studying its impact on standardized test scores; furthermore, many WSF studies use datasets that limit the robustness and interpretation of results. A National Study by the U.S. Department of Education, focuses on describing WSF policy, its intended changes and benefits, and details school districts that adopted WSF before 2018 (Levin, Manship, Hurlburt, and Atchison 2019) [6]. However, this report does not study any quantified effect on educational attainments. Another paper from the Edunomics Lab in Georgetown focuses on understanding the financial details of WSF at the district level, specifically the unique weight formulation of each district and whether the formulations are aligned with WSF's goal of increasing equity (Roza et.al 2019) [4]. Roza begins to scrape at the surface of understanding the academic outcomes of WSF and finds that at the state-level, districts that implement WSF see higher average student outcomes on standardized Math and ELA tests and outcomes are even larger in higher-poverty schools. However, the Edunomics report notes that the state-level results should be interpreted with caution since WSF districts tend to be different than others in their state, in both enrollment size and student composition, and the effects of WSF cannot be isolated from the effects of other policies implemented around the same time. As previously mentioned, my paper produces more robust results by only selecting treatment and control districts with over 15,000 students from a nationally representative sample.

The rest of the paper is organized as follows. Section 2 provides an in-depth WSF policy debrief. Section 3 presents the data. Section 4 describes the empirical strategy. Section 5 presents the results. Section 6 presents a summary discussion. Section 7 details future work, and Section 8 concludes the paper.

2 Policy Background

Historically U.S. school districts distribute federal funds to schools through tangible resources, rather than allocating specific dollar amounts to individual schools. These traditional uniform staffing allocation systems typically determine the number of teachers, school administrators, and other types of staff for each school based on its total student enrollment. Many educators and researchers have noted that traditional resource allocation systems can contribute to and increase inequity amongst schools as schools with higher concentrations of at-risk students may not receive additional resources to meet the complex needs of such students (Rubenstein, Schwartz, and Stiefel 2006) [8].

The WSF Federal Government Program attempts to mitigate these inequities by allowing districts to deploy a fixed-dollar amount to schools for each student type with larger increments going to students from low-income backgrounds, with special needs, and/or who are English-language learners. Under the WSF approach, districts may allocate resources more effectively to meet the specific needs of each of their school's students.

Policymakers from the federal government to the district level are always researching and creating new programs and funding methods to improve public education. Districts choose to adopt WSF to increase equity, transparency, flexibility, and school-level autonomy to focus on improving student outcomes (Roza et.al 2019) [4]. WSF has been around since 1995 and over the past 2 decades, 27 school districts have implemented WSF with these goals in mind. This paper sets out to understand whether WSF indeed improved student educational outcomes. This is relevant to today as Biden plans to double funding for K-12 education through the "Build Back Better" plan as schools struggle to successfully emerge out of the pandemic and help students meet standards following the past year of virtual learning (Camera 2021) [3]. Understanding the effects of WSF can help schools and the federal government use their budget effectively.

3 Data

To study the impact of implementing WSF on high school educational attainment through a difference-in-difference model, I need funding data at the district level to understand which districts implemented WSF as well as school district performance data. Both datasets are further detailed below.

I will rely on the findings of existing WSF literature to identify control and treated districts. The U.S. Department of Education made a detailed 2019 WSF report (Levin, Manship, Hurlburt, and Atchison 2019) [6]. This report includes a table of 27 well-documented districts that have implemented WSF, have continued to implement it, and the year in which they implemented it. The Roza et al. (2019-Present) Georgetown report provides a similar table of 18 districts that have implemented WSF, and these 18 districts align with the 27 districts provided by the U.S. Department of Education. I use both lists to develop my treatment group. However, both WSF papers, anonymize schools that did not implement WSF increasing the difficulty of creating a control group. Through direct discussion with Hannah Jarmalowski, a Research Fellow at Georgetown Edunomics Lab, she explained that there are very few districts that have implemented WSF and districts that have are documented in the literature. With no other existing papers, I use the two papers described above to create a thorough table of WSF implementing school districts (Table 1) and have confidence that unlisted districts have never implemented WSF.

To explore multiple levels of educational attainment, the main data resource will be the NCES, the National Center for Education Statistics. The NCES provides yearly tables of "Selected statistics on enrollment, teachers, dropouts, and graduates in public school districts enrolling more than 15,000 students" from 1995-2018. The pupil per teacher ratio is one of the only variables available every year from 1995-2018. It is important to note that the ratio itself is not a measurement of educational attainment, but in the literature, lower pupil per teacher ratio is correlated with higher educational achievement (Jackson, Rucker and Persico 2015) [2]. The NCES tables also contain high school dropout rates by district from

1996-2009. Although this does not cover up to 2018, there are 9 schools that adopted WSF around 2002 and dropout rates can be observed for those sub-selected districts. The NCES also documents high school graduation rates by district from 2007 to 2018, and this data can be used for the 10 schools that adopted WSF between 2007 and 2018. Note intuitively it should be possible to get graduation rates from 1996-2009 by using 1-dropout rate, but for 2007 and 2008 in which both graduation and dropout rates are available, graduation rates are not equivalent to 1-dropout rates.

Due to changes in data collection methods over the years, it is difficult to find consistent data measurements over the past 25 years. The NCES does provide the number of high school graduates at the district level from 1995 to 2009, but this raw number is unusable because it does not separate number of graduates from national migration changes and general population growth. In addition to the number of graduates, the NCES provides the total enrollment count at every district. As a rough estimation, I divide the number of graduates by total enrollment to get a "pseudo-graduation" rate from 1995 to 2009.

To perform a robust staggered difference-in-difference, the data must be divided into treated and control groups using information from the U.S. Department of Education report and NCES. The treated group for each educational attainment will be selected from the 27 WSF school districts documented in the U.S. Department of Education report. I do not use all 27 school districts currently implementing WSF as the Minneapolis School District implemented WSF in 1993 and the Prince William County Public Schools implemented WSF in 1994, but there is insufficient NCES data prior to 1995. Atlanta Public Schools and Shelby County Schools districts implemented WSF in 2018, but NCES has yet to upload the needed data beyond 2018. Following these adjustments, the treated group is selected from a pool of 23 districts. The control districts will be chosen from the NCES yearly table of "Selected statistics on enrollment, teachers, dropouts, and graduates in public school districts enrolling more than 15,000 students." Only districts with consistent data for the respective time period for each attainment will be chosen.[1] All the districts will be from this table since most WSF implementing districts are large urban school districts, and the NCES only provides district-level statistics on districts with more than 15,000 students.

Summarizing the NCES data reveals the number of treated and control observations for every educational attainment. Table 2 summarizes the pupil-per-teacher ratio, dropout rate, graduation rate, and pseudo-graduation rate by control and treatment group. Notice that dropout rate has the most number of observations (total and by control/treated) even though it does not cover the full time period from 1995-2018 because it does not omit treated or control districts that are missing data for any year in between 1996-2009. On the other hand, the other 3 measurements only include data for districts with measurements for every year. This is necessary because there are no treated districts that had dropout data for every year between 1996 and 2009. This effects the interpretation of dropout regression results which will be further discussed in Section 6, the discussion section. Figure 1 visually summarizes the data by graphing arbitrarily chosen districts treated in the same year versus control schools for each educational attainment. Even before running the empirical tests, this figure hints at two findings: parallel trends are likely to be unsatisfied for dropout rates and results for all educational attainments are likely to be small in size and effect.

4 Methods

To understand the effect of WSF on various high school educational attainments at the district-level, this study will rely on the difference-in-difference method to compare outcomes at control districts that never implemented WSF and treated districts that implemented WSF between 1995-2018.

Difference-in-difference is the best method, given the available data and nature of WSF

[1] For pupil-per-teacher ratio, actual graduation rate, and pseudo-graduation rate, only districts with data for every year will be chosen. The exception is the dropout rate controls because dropout data is not available for every year for any treated district. This will limit the interpretation of the dropout rate results which is further discussed in Section 6.

implementation across districts, to estimate the educational effects of WSF. However, among the 23 treated school districts, many districts were treated at different times. This makes it difficult to perform a simple regression and a traditional difference-in-difference. Thus, I propose the following regression which should find coefficients on the pre-treated periods are statistically insignificant and hence demonstrate parallel trends leading into the treatment while the coefficients on post-treated periods will show the effect of WSF on the specific measurement of educational attainment:

$$E_{d,t} = \alpha_d + \delta_t + \sum_{y=0}^{T_1} \gamma_y D_{d,y} + \sum_{y=T_0}^{-2} \gamma_y D_{d,y} + \epsilon_{d,t} \quad \text{(Equation 1)}$$

Where $E_{d,t}$ is the educational outcome for a district d at time t. α_d and δ_t are the district and year fixed effects respectively. ϵ_{dt} is the error term. T_0 and T_1 in the summation are, respectively, the lowest lag year and highest lead year to consider surrounding the treatment period. $D_{d,y}$ is a dummy variable that is equal to 1 if the observation's period relative to district d's first treated period is the same value as y; otherwise the dummy is equal to 0 and is 0 for all never-treated observations. The regression coefficients are the γs which are for each year leading and lagging the treatment. Note the -1 is omitted from the summation to avoid multicolinearity and serves as the point of reference.

Equation 1 describes a dynamic regression which will give detailed insight into the effect of WSF on the educational attainment every year after treatment. However, for simplicity of understanding the overall effects of WSF, I will also run a static regression (Equation 2).

$$E_{d,t} = \alpha_d + \delta_t + \beta * (POST_t * TREAT_d) \quad \text{(Equation 2)}$$

In Equation 2, I regress the outcome for district d in year t on a dummy variable that is the interaction between $POST_T$ (year t is after WSF has been implemented in that district) and $TREAT_d$ (district d is a district in which WSF has been or will be implemented). Like in Equation 1, α_d and δ_t are the district and year fixed effects respectively.

Graphing the γ coefficients from Equation 1 will show the sign and size of the treatment, but to be able to effectively interpret these results, several assumptions need to be satisfied. First, the allocation of intervention must not be determined by the outcome; meaning if an increase in educational attainments is found following the implementation of WSF, it is due to the new funding scheme rather than prior characteristics or other novel changes of the school district. This assumption is satisfied because the 23 WSF implementing school districts and the control group are nationally representative. Potential educational attainment changes can be attributed to WSF because it is unlikely multiple schools passed similar policies other than WSF at the same time and achieved similar educational results.

Additionally there must be no spillover effects from treated to untreated school districts. Historically, school districts are very isolated, and students within one district are within the same city and their education is unaffected by the policies of nearby districts. Furthermore, there have been numerous peer-reviewed, economic studies that have compared various school districts in the same area using a difference-in-difference model (Harris and Larsen 2018) [5].

The most important assumption to satisfy is the parallel trend assumption. As in most economic studies, it is impossible to observe the treatment group in the absence of treatment. Thus, I will show the γ coefficients leading into treatment in Equation 1 are zero indicating parallel trends into treatment. Graphing these coefficients in Figures 2-5 for all districts across all years reveals the coefficients on the pre-period dummies are statistically indistinguishable from 0. These findings are further discussed in Section 5.

5 Results

This study considers the impact of WSF on high school district pupil-per-teacher ratio, actual and pseudo-graduation rates, and dropout rates. I run the dynamic regression described in Equation 1 and the static regression described in Equation 2 for the selected control and treated districts while accounting for district and year fixed effects. The rest of this section will present the results from the static regression followed by the dynamic regression.

The static regression reveals positive but only marginally significant effects on pupil-per-teacher ratios, actual and pseudo-graduation rates, and harder-to-interpret negative effects on dropout rates. Table 3 shows these raw results of the static regression. It is clear that pupil per teacher ratio and graduation rate coefficients are slightly positive but statistically insignificant at the 5% and even 10% level. The effect of WSF on the dropout rate is negative and statistically significant at the 10% level. However, Figure 5, a graph of coefficients on dropout rates from the dynamic regression, clearly shows that parallel trends are unsatisfied for dropout rates, thus these results are not robust. Most notably, the pseudo-graduation rate appears to be slightly positive and to be statistically significant at the 10% level. However, pseudo-graduation is a measurement created for this study and is difficult to interpret. It will be further discussed in Section 6.

Dynamically regressing on pupil-per-teacher ratio leads to coefficients of negligible size leading into and lagging out of treatment. Referencing Figure 2, the confidence intervals on the regression coefficients for every lag year cover 0. However, the lagging coefficients also cover 0 and do not seem to have a constant trend which signals that WSF does not have a significant effect on pupil-per-teacher ratio. Figure 2 was created using Appendix Table A1 which includes raw coefficients and standard errors.

While the pupil-per-teacher regression satisfies parallel trends, the dynamic graduation rate regression shows not all of leading coefficients cover zero in their 95% confidence interval (Figure 3 created using Appendix Table A2). This is likely due to the limited number of treatment schools and smaller time period compared to the pupil-per-teacher data. Due to the noise of these results, the actual graduation rate results are unusable in identifying the effect of WSF.

WSF appears to have a noticeable effect on pseudo-graduation rates at the 10% significance level. Pseudo-Graduation was calculated from 1995-2009 for 4 treated districts and 137 control districts. Starting with satisfying the parallel trends assumption, all leading treatment coefficients in Figure 4 have confidence intervals that cover 0. This helps support the parallel trend assumption leading into treatment. In this case, the lagging coefficients appear to have an upward trend that becomes slightly significant around 6 years after treatment. Figure 4 was created using Table A3 attached in the appendix which includes raw coefficients and standard errors.

As discussed previously, the dropout rate coefficients are difficult to interpret as they do not satisfy the parallel trend assumption (Figure 5). It is important to note that after treatment, the regression coefficient confidence intervals do follow a negative trend, however, the coefficients continue to cover 0 indicating an absence of a statistical significant effect of WSF on dropout rates.

6 Discussion

Overall, WSF has limited impact in size and significance on high school pupil-per-teacher ratio, actual and pseudo-graduation rates, and dropout rates. As described in the results section, the coefficient on pupil-per-teacher ratio is close to 0 and statistically insignificant. This is not immensely surprising because as noted in the Section 1 and 2, WSF does not increase the total sum of money a district receives. Even though some higher-risk schools within a district may receive additional funding through WSF to invest in more teachers, at the district level and nationally WSF has limited impact on the pupil-per-teacher ratio.

The effect on actual graduation rates is close to null which is unsurprising given the literature on the challenges of improving high school graduation rates. Following treatment the mean graduation rates rise for treated districts from 62.89 to 63.056 (Table 3) which is a close to 0 effect and again this result is statistically insignificant. Again this is not immensely surprising, as high school graduation rates are historically difficult to improve even through programs targeted at improving graduation rates (Abele and Iver 2011) [1]. Furthermore, graduation rates do not fully satisfy parallel trends making the interpretation less robust (Figure 3). This is likely due to the fact that there are only 5 treated districts which increases the noise.

Dropout rates slightly decrease following WSF, but it is imminent to remember that dropout rates fail to satisfy parallel trends. Districts that implement WSF appear to decrease dropout rate by about 1% compared to districts that do not implement WSF which is a relative 11% decrease of the pre-treatment mean, at the 10% statistical significant level. However, this result is not robust as the dropout rate regression does not satisfy the parallel trend assumption (Figure 5). Without this vital assumption, there is no definitive conclusion. The data for dropout rates was not panel data which likely increased the noise of the data leading into treatment. If more data is acquired, parallel trends can be satisfied, and a definitive effect of WSF on dropout rates can be identified.

Finally, I find after implementing WSF, district pseudo-graduation rates increase from 4.32 to 4.616 which is a 6% increase, but there are many limitations to this result. First of all, it is at the 10% significance level and should be approached with caution. Furthermore, this educational attainment measurement was made for this paper due to limited public available district level data. The original goal of the pseudo-graduation rate measurement was to support the results of the effect of WSF on standard graduation rates. However, the pseudo-graduation result should not be fully discarded and rather further studied. Remember pseudo-graduation is equal to the number of graduates divided by total enrollment within a district. Since I found no effect of WSF on graduation rates, WSF increasing pseudo-graduation rate could indicate WSF leads to a decrease in total enrollment within a district. This could signal a decrease in high school enrollment which is not necessarily an adverse effect. For example, decreasing high school enrollment within the studied schools could imply migration of families to less urban and crowded schools.

Ultimately WSF has no sizable and significant effects on pupil-per-teacher ratio, dropout rates, graduation rates, and even pseudo-graduation rates. However, even this finding should not be discarded. One major critique of WSF is that it reallocates money from higher-income students to those who are qualified for WSF funding which could negatively impact more privileged students. However, the 95% confidence interval of every coefficient covers 0 which indicates that WSF does not harm the general student population.

7 Future Work

The inconclusive results of this study indicate a need to continue understanding WSF's effect at the level of students directly targeted by WSF. Due to time constraints, I was unable to also explore the effects of WSF at the level of students who are English-Language learners, have disabilities, or come from low-income backgrounds. After exploring literature and data sets from the Equality of Opportunity project, I identified two promising data sets: the EDFacts data set and Neighborhood Characteristics by County.

The EDFacts data set details the percentage of students in every district who score above proficient on their state's ELA and Math standardized test from 2009-2018 and breaks the statistic down at the low-income, disability, and English language learner level. The EDFacts data set is extremely large and encoded and needs to be thoroughly processed and separated by control and treatment districts, about 10 treated districts in the given time period. The empirical method for standardized testing will follow the same dynamic regression described in Equation 1 in Section 4. Although it is disappointing that the effect of WSF on standardized testing must be left as a future study, there are many drawbacks in current literature that hindered this study from focusing on standardized testing. First of all, the Georgetown WSF study already explores the effect of WSF on standardized testing. As the purpose of this study was to expand upon WSF's overall effects, I chose to put full focus into exploring other educational attainments. Another reason this study did not focus on testing is over 40 states changed their standardized tests in 2010 with the adoption of national common core increasing the difficulty of isolating the effect of WSF from drastic changes in standardized testing (Polleck and Jeffery 2017) [7]. However, I can try to mitigate this effect by adding a fixed state effect.

It is also important to study the effect of WSF on pupil-per-teacher ratio, actual and pseudo graduation rates, and dropout rates at the low-income level using the Equal Opportunity data source Neighborhood Characteristics by County. This data set details the percentage of low-income county residents. However, this data set is by county, so I would only use control and treated districts that cover full counties and assume that the county-level and district level percentage of low-income backgrounds are similar. The regression will follow a similar format to Equation 1 from Section 4, but with an additional variable $I_{d,y}$, representing the low-income population percentage in district d and year y (Equation 3). The γ coefficient will find the isolated effect on districts implementing WSF, ρ coefficient will represent the isolated effect on continuous income levels, and σ, our main coefficient of interest, finds the interaction for every year for varying income levels.

$$E_{d,t} = \alpha_d + \delta_t + \sum_{y=0}^{T_1} \gamma_y D_{d,y} + \sum_{y=0}^{T_1} \rho_y I_{d,y} + \sum_{y=T_0}^{-2} \gamma_y D_{d,y} + \sum_{y=T_0}^{-2} \rho_y I_{d,y}$$

$$+ \sum_{y=T_0}^{-2} \sigma_y D_{d,y} * I_{d,y} + \sum_{y=0}^{T_1} \sigma_y D_{d,y} * I_{d,y} + \epsilon_{d,t} \quad \text{(Equation 3)}$$

8 Conclusion

Weighted School Funding has been around for over 2 decades and over 20 districts have implemented the funding policy in an attempt to solve inequities between students by allocating additional funds to students from low-income backgrounds, who are english-language learners, or who have a disability. However, WSF is largely unstudied and little is known about its effects on educational attainments. Using available public data, I studied the effect of WSF on pupil-per-teacher ratio, graduation rates, pseudo-graduation rates, and dropout rates.

Although overall WSF has limited impact on these educational attainments or produces inconclusive results, I discover WSF has no apparent negative effect and must be further studied. First, the mostly null effects of WSF indicate that WSF appears to not harm students from privileged backgrounds which was one of the only policy concerns. This study also emphasizes that researchers have barely scraped the surface of thoroughly understanding WSF. As described in Section 7, there are already 2 potential analyses; however, there are even more undiscovered empirical tests that can further the understanding of WSF such as the effect of WSF on college enrollment, primary school attainments, etc. WSF is implemented by some of the largest and most innovative school districts like New York City and Boston, and as WSF continues to spread nationally, it is necessary to stop blindly adopting WSF and instead begin to fully comprehend its drawbacks and benefits.

References

[1] Martha Abele and Mac Iver. The challenge of improving urban high school graduation outcomes: Findings from a randomized study of dropout prevention efforts. *Journal of Education for Students Placed at Risk*, pages 167–184, 2011.

[2] Rucker C. Johnson C. Kirabo Jackson and Claudia Persico. The effects of school spending on educational and economic outcomes: Evidence from school finance reforms. *The Quarterly Journal of Economics*, pages 157–218, 2015.

[3] Lauren Camera. Biden's budget significantly boosts k-12 education spending. *U.S. News*, 2021.

[4] Marguerite Roza et. al. Lessons learned: Weighted student funding. 2020.

[5] Douglas N. Harris and Matthew F. Larsen. The effects of the new orleans post-katrina market-based school reforms on medium-term student outcomes. 2019.

[6] Steve Hurlburt Drew Atchison Ryoko Yamaguchi Adam Hall Stephanie Stullich Jesse Levin, Karen Manship. Districts' use of weighted student funding systems to increase school autonomy and equity: Findings from a national study. 2019.

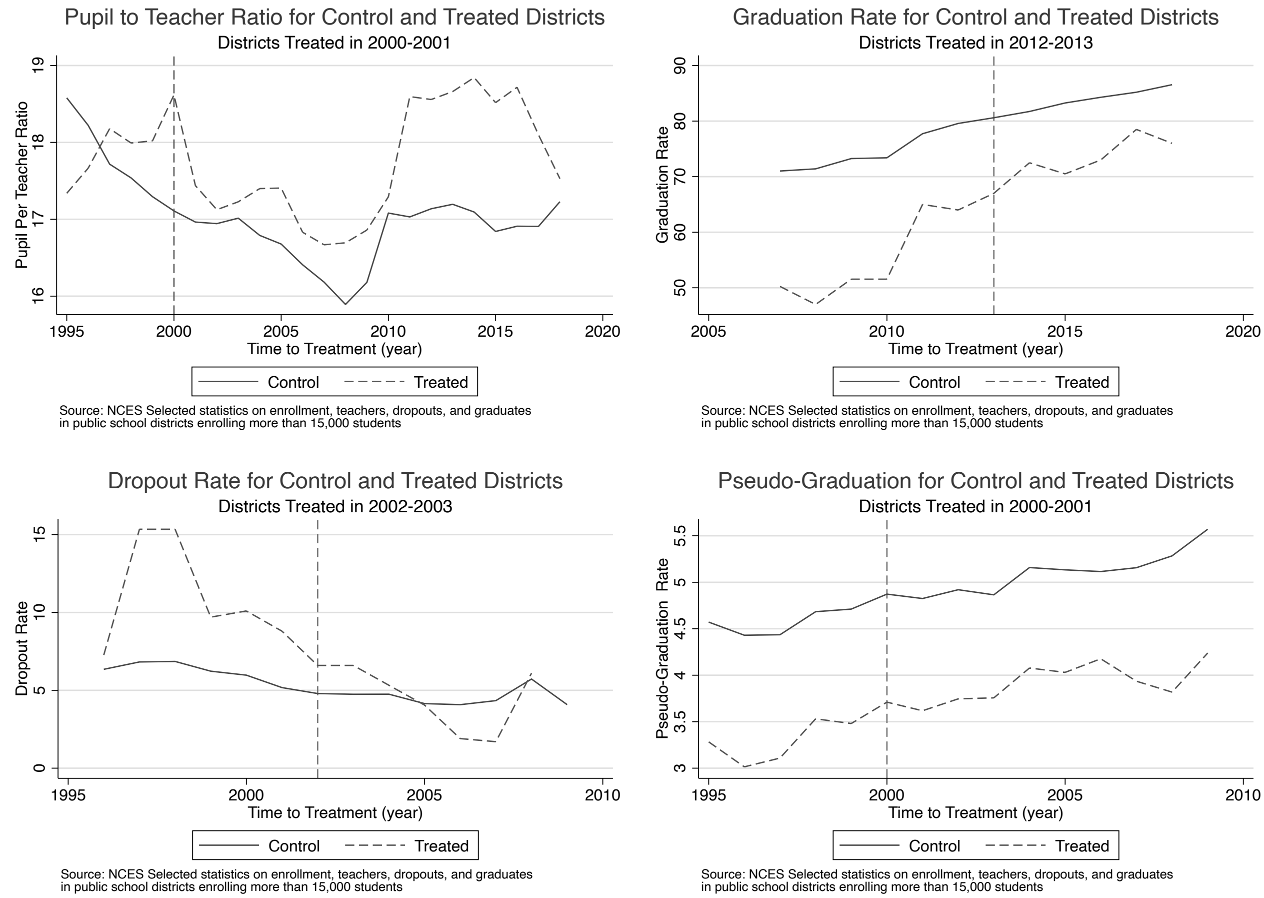
[7] Jill V. Jeffery Jody N. Polleck. Common core standards and their impact on standardized test design: A new york case study. *The High School Journal*, pages 1–26, Fall 2017.

[8] Amy Ellen Schwartz Ross Rubenstein and Leanna Stiefel. Rethinking the intradistrict distribution of school inputs to disadvantaged students. 2006.

[9] C. Kirabo Jackson Rucker C. Johnson. Reducing inequality through dynamic complementarity: Evidence from head start and public school spending. *American Economic Journal: Economic Policy*, pages 310–349, 2019.

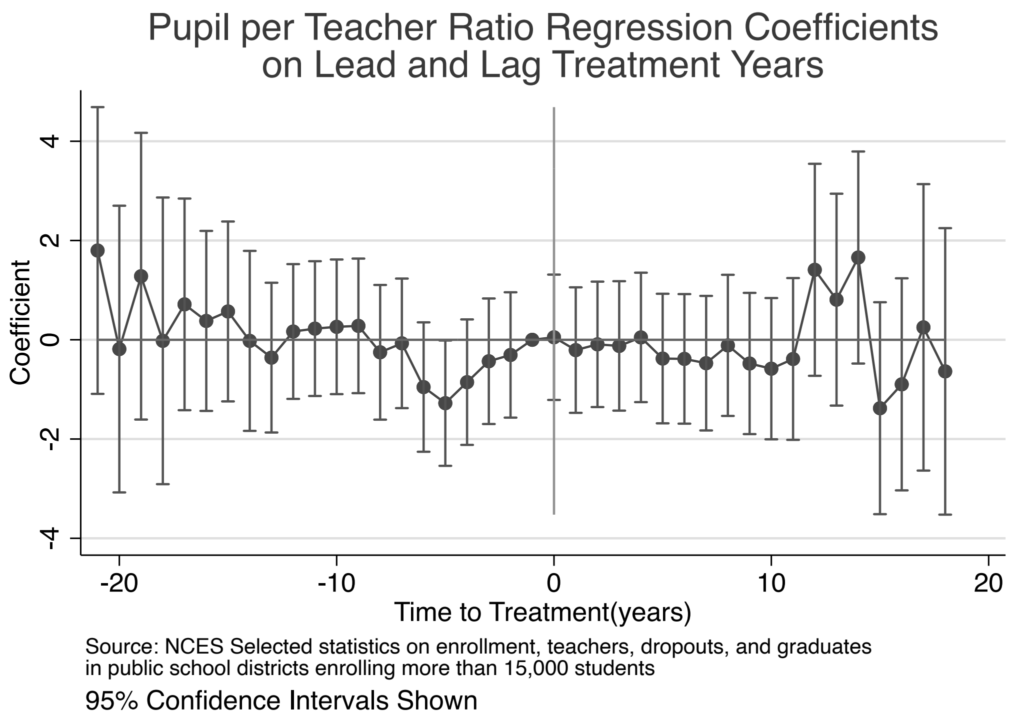
Figures and Tables

Figure 1: Educational Attainment Trends Over Time



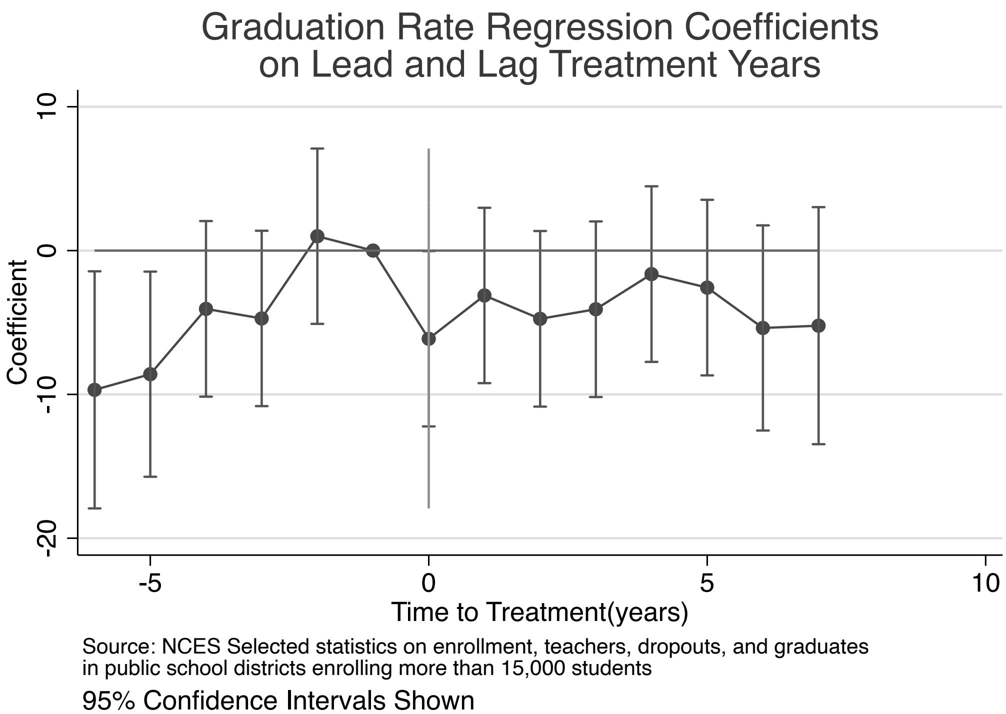
Note: Pupil Per Teacher Ratio from 1995-2018 for 144 control districts versus 2 districts treated in 2000-2001 school year. Graduation Rate from 2007-2018 for 138 control districts versus 1 district treated in 2012-2013 school year. Dropout rates were calculated from 1996-2009 for 343 control districts versus 3 districts treated in 2002-2003 school year. Pseudo-Graduation was calculated from 1995-2009 for 272 control districts versus 2 district treated in 2000-2001 school year using $\frac{\text{\# of high school graduates within district}}{\text{total enrollment within the district}} * 100$. Treated districts were identified using Georgetown Edunomics WSF Report (Roza et. al 2019-Present).

Figure 2



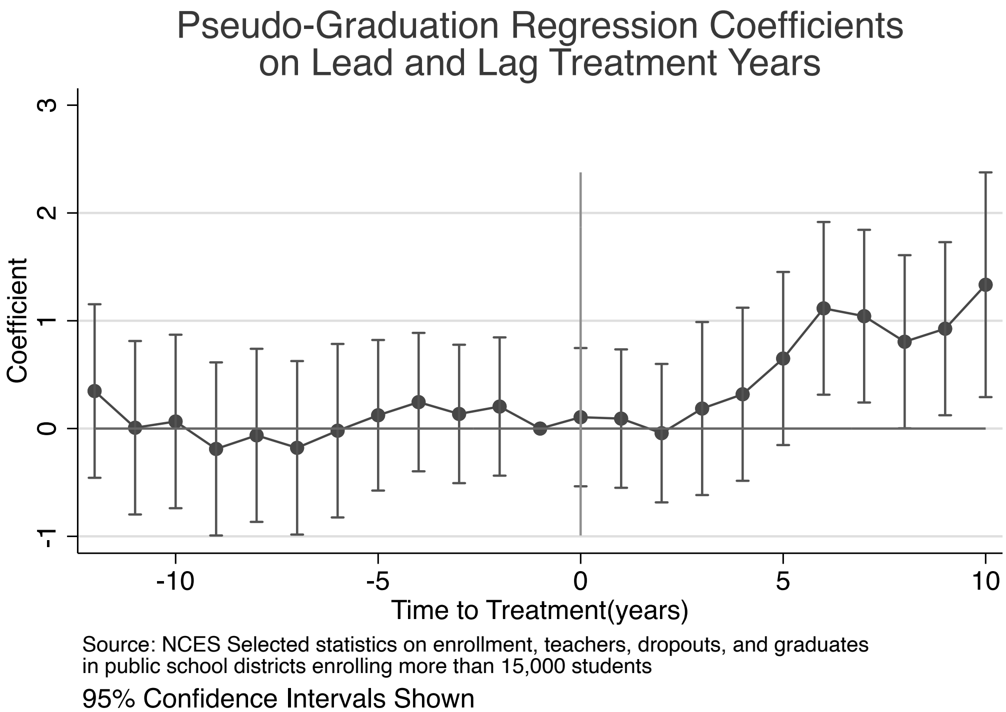
Note: Regression coefficients with confidence intervals on lead and lag years for Pupil per teacher ratio from 1995-2018 for 8 treated districts and 144 control districts. Note the year before treatment has been omitted to avoid multicolinearity and have a relative time reference. The graph was created using Appendix Table A1.

Figure 3



Note: Regression coefficients with confidence intervals on lead and lag years for Graduation Rates from 2007-2018 for 5 treated districts and 272 control districts. Note the year before treatment has been omitted to avoid multicolinearity and have a relative time reference.The graph was created using Appendix Table A2.

Figure 4



Note: Regression coefficients with confidence intervals on lead and lag years for Pseudo-Graduation rate:

$$\frac{\#\text{ of high school graduates within district}}{\text{total enrollment within the district}} * 100$$

Pseudo-Graduation was calculated from 1995-2009 for 4 treated districts and 137 control districts. Note the year before treatment has been omitted to avoid multicolinearity and have a relative time reference. The graph was created using Appendix Table A3.

Table 1: Districts Implementing WSF as of 2019

District Names	State	Year Adopted	Enrollment	Number of Schools	Poverty Rate	Urbanicity
Minneapolis Public Schools	MN	1993–94	36,793	86	24%	City
Prince William County Public Schools	VA	1994–95	87,793	92	9%	Suburb
Cincinnati Public Schools	OH	1999–2000	34,227	54	33%	City
Houston Independent School District	TX	2000–01	215,627	283	31%	City
Milwaukee School District	WI	2000–01	75,749	158	34%	City
San Francisco Unified School District	CA	2002–03	58,865	116	12%	City
St. Paul Public School District	MN	2002–03	37,698	103	27%	City
Hawaii Department of Education	HI	2006–07	181,995	289	10%	Suburb
Denver Public Schools	CO	2007–08	90,235	189	20%	City
New York City Public Schools	NY	2007–08	981,667	1,579	26%	City
Poudre School District	CO	2007–08	29,527	53	9%	City
Baltimore City Public Schools	MD	2008–09	83,666	182	31%	City
Douglas County School District	CO	2008–09	66,896	89	2%	Suburb
Falcon School District 49	CO	2010–11	20,561	22	8%	City
Boston Public Schools	MA	2011–12	53,885	120	28%	City
Charlotte-Mecklenburg Schools	NC	2011–12	146,211	164	17%	City
Newark Public School District	NJ	2011–12	40,889	65	33%	City
Prince George's County Public Schools	MD	2012–13	128,936	207	12%	Suburb
Adams 12 Five Star Schools	CO	2013–14	39,287	53	10%	Suburb
City of Chicago School District 299	IL	2013–14	387,311	591	27%	City
Cleveland Municipal School District	OH	2013–14	39,410	101	43%	City
Metro Nashville Public Schools	TN	2015–16	85,598	154	23%	City
Jeffco Public Schools	CO	2015–16	86,731	165	7%	Suburb
Santa Fe Public Schools	NM	2015–16	13,265	33	20%	City
Indianapolis Public Schools	IN	2016–17	31,371	67	41%	City
Atlanta Public Schools	GA	2018–19	51,500	89	33%	City
Shelby County Schools	TN	2018–19	114,487	208	34%	City

Note: Table reads: Minneapolis Public Schools adopted a WSF system in the 1993–94 school year, enrolls 36,793 students, has 86 schools, a poverty rate of 24 percent, and is located in a city.

Sources: Information gathered from U.S. Department of Education report, *Districts' Use of Weighted Student Funding Systems to Increase School Autonomy and Equity:Findings From a National Study*.

Note: Poverty rates are based on the 2016 Census Small Area Income Poverty Estimate (SAIPE) data for school districts.

Table 2: Summary Statistics

	Sum	Mean	SD	Min	Max	N
Pupil per Teacher 1995-2018						
Control	58,880	17.05	3.12	9	57	3,456
Treated	3,242	16.88	1.87	12	22	192
Total	62,121	17.04	3.07	9	57	3,648
Pseudo-Graduation Rate 1995-2009						
Control	10,172	4.93	0.95	0	9	2,070
Treated	258	4.29	1.23	2	7	60
Total	10,429	4.91	0.97	0	9	2,130
Graduation Rate 2007-2018						
Control	258,454	79.09	11.76	35	100	3,264
Treated	4,110	68.51	9.72	37	83	60
Total	262,564	78.90	11.81	35	100	3,324
Dropout Rate 1996-2009						
Control	23,768	4.95	3.49	0	33	4,802
Treated	689	8.20	3.96	1	21	84
Total	24,457	5.01	3.53	0	33	4,886

Note: Description: Pupil per teacher ratio from 1995-2018 for 8 treated districts and 144 control districts. Pseudo-Graduation Rate which is $\frac{\#\text{ of high school graduates within district}}{\text{total enrollment within the district}}$. Pseudo-Graduation was calculated from 1995-2009 for 4 treated districts and 138 control districts. Graduation Rates in percentage form from 2007-2018 for 5 treated districts and 272 control districts. Dropout rates from 1996-2009 for 9 treated districts and 343 control districts. Treated districts were identified using Georgetown Edunomics WSF Report (Roza et. al 2019-Present) and U.S. Department of Education WSF Report (Levin, Manship, Hurlburt, and Atchison 2019).

Table 3: Static Regression Coefficients

	Pupil Per Teacher	Graduation	Pseudo-Graduation	Dropout
Coefficient	0.0414	0.166	0.296	-1.004
	(0.18)	(1.52)	(0.17)	(0.57)
	[0.26]	[0.11]	[1.76]	[-1.78]
Mean Pre-Treatment	16.65	62.89	4.32	9.14
Observations	3,648	3,324	2,130	4,886

Note: Standard errors in parentheses and t-statistics in brackets. Static regression coefficients for every educational attainment. Note that pseudo-graduation is $\frac{\text{\# of high school graduates within district}}{\text{total enrollment within the district}} * 100$.Table of β coefficients Based on Equation 2. Each regression covers unique set of years and has its own set of control and treated schools based on which districts have data: Pupil per teacher ratio from 1995-2018 for 8 treated districts, Graduation Rates from 2007-2018 for 5 treated districts, Pseudo-Graduation from 1995-2009 for 4 treated districts, and Dropout Rates from 1996-2009 for 9 treated districts.

Appendix

Appendix Tables

Table A1: Leading and Lagging Coefficients for Pupil Per Teacher Ratio Dynamic Regression

	(1) Coefficient
lead/lag year=-21	1.768
	(1.47)
lead/lag year=-20	-0.212
	(1.47)
lead/lag year=-19	1.253
	(1.47)
lead/lag year=-18	-0.0495
	(1.47)
lead/lag year=-17	0.685
	(1.09)
lead/lag year=-16	0.354
	(0.93)
lead/lag year=-15	0.545
	(0.93)
lead/lag year=-14	-0.0487
	(0.93)
lead/lag year=-13	-0.648
	(0.77)
lead/lag year=-12	-0.0144
	(0.69)
lead/lag year=-11	0.220
	(0.69)
lead/lag year=-10	0.0362
	(0.69)
lead/lag year=-9	0.275
	(0.69)
lead/lag year=-8	-0.136
	(0.69)
lead/lag year=-7	-0.0745
	(0.67)
lead/lag year=-6	-0.885
	(0.67)
lead/lag year=-5	-1.264*
	(0.64)
lead/lag year=-4	-0.642
	(0.65)
lead/lag year=-3	-0.432
	(0.65)
lead/lag year=-2	-0.307
	(0.64)
lead/lag year=-1	0
	(.)
lead/lag year=0	0.0511
	(0.64)
lead/lag year=1	-0.281
	(0.65)
lead/lag year=2	-0.0913
	(0.65)
lead/lag year=3	-0.120
	(0.67)
lead/lag year=4	-0.0774
	(0.67)
lead/lag year=5	-0.374
	(0.67)
lead/lag year=6	-0.381
	(0.67)
lead/lag year=7	-0.464
	(0.69)
lead/lag year=8	-0.0997
	(0.73)
lead/lag year=9	-0.323
	(0.73)
lead/lag year=10	-0.569
	(0.73)
lead/lag year=11	-0.378
	(0.83)
lead/lag year=12	1.426
	(1.09)
lead/lag year=13	0.824
	(1.09)
lead/lag year=14	1.675
	(1.09)
lead/lag year=15	-1.362
	(1.09)
lead/lag year=16	-0.881
	(1.09)
lead/lag year=17	0.267
	(1.47)
lead/lag year=18	-0.621
	(1.47)
Constant	17.01***
	(0.63)
Observations	3648

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Raw regression coefficients with standard errors on lead and lag years for Pupil per teacher ratio from 1995-2018 for 8 treated districts and 144 control districts. Note the year before treatment has been omitted to avoid multicolinearity and have a relative time reference. This table was used to create Figure 2.

Table A2: Leading and Lagging Coefficients for Graduation Rate Dynamic Regression

	(1) Coefficient
lead/lag year=-6	-9.674*
	(4.19)
lead/lag year=-5	-8.585*
	(3.63)
lead/lag year=-4	-4.043
	(3.10)
lead/lag year=-3	-4.713
	(3.10)
lead/lag year=-2	1.002
	(3.10)
lead/lag year=-1	0
	(.)
lead/lag year=0	-6.132*
	(3.10)
lead/lag year=1	-3.117
	(3.10)
lead/lag year=2	-4.739
	(3.10)
lead/lag year=3	-4.068
	(3.10)
lead/lag year=4	-1.625
	(3.10)
lead/lag year=5	-2.564
	(3.10)
lead/lag year=6	-5.371
	(3.63)
lead/lag year=7	-5.211
	(4.19)
Constant	84.98***
	(3.07)
Observations	3324

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Raw regression coefficients and standard errors in parenthesis on lead and lag years for Graduation Rates from 2007-2018 for 5 treated districts and 272 control districts. Note the year before treatment has been omitted to avoid multicolinearity and have a relative time reference. These values were used to create Figure 3.

Table A3: Leading and Lagging Coefficients for Pseudo-Graduation Rate Dynamic Regression

	(1) Coefficient
lead/lag year=-12	0.348
	(0.41)
lead/lag year=-11	0.00444
	(0.41)
lead/lag year=-10	0.0655
	(0.41)
lead/lag year=-9	-0.189
	(0.41)
lead/lag year=-8	-0.0631
	(0.41)
lead/lag year=-7	-0.179
	(0.41)
lead/lag year=-6	-0.0200
	(0.41)
lead/lag year=-5	0.123
	(0.36)
lead/lag year=-4	0.245
	(0.33)
lead/lag year=-3	0.135
	(0.33)
lead/lag year=-2	0.204
	(0.33)
lead/lag year=-1	0
	(.)
lead/lag year=0	0.106
	(0.33)
lead/lag year=1	0.0923
	(0.33)
lead/lag year=2	-0.0428
	(0.33)
lead/lag year=3	0.186
	(0.41)
lead/lag year=4	0.318
	(0.41)
lead/lag year=5	0.650
	(0.41)
lead/lag year=6	1.115**
	(0.41)
lead/lag year=7	1.043*
	(0.41)
lead/lag year=8	0.806*
	(0.41)
lead/lag year=9	0.926*
	(0.41)
lead/lag year=10	1.334*
	(0.53)
Constant	4.798***
	(0.32)
Observations	2130

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Raw regression coefficients with standard errors on lead and lag years for Pseudo-Graduation rate:

$$\frac{\text{\# of high school graduates within district}}{\text{total enrollment within the district}} * 100$$

Pseudo-Graduation was calculated from 1995-2009 for 4 treated districts and 137 control districts. Note the year before treatment has been omitted to avoid multicolinearity and have a relative time reference. This table was used to create Figure 4.

Table A4: Leading and Lagging Coefficients for Dropout Rate Dynamic Regression

	(1) Coefficient
lead/lag year=-9	-5.515*
	(2.22)
lead/lag year=-8	-4.321***
	(1.28)
lead/lag year=-7	-5.227***
	(1.42)
lead/lag year=-6	-1.684
	(1.45)
lead/lag year=-5	2.922*
	(1.41)
lead/lag year=-4	2.824*
	(1.41)
lead/lag year=-3	1.163
	(1.27)
lead/lag year=-2	-1.068
	(1.06)
lead/lag year=-1	0
	(.)
lead/lag year=0	0.240
	(1.07)
lead/lag year=1	-0.792
	(1.03)
lead/lag year=2	-1.810
	(1.08)
lead/lag year=3	-1.488
	(1.16)
lead/lag year=4	-1.667
	(1.34)
lead/lag year=5	-1.642
	(1.49)
lead/lag year=6	-1.472
	(1.34)
lead/lag year=7	-2.260
	(1.45)
lead/lag year=8	-1.839
	(1.45)
lead/lag year=9	-3.231*
	(1.45)
lead/lag year=10	-2.431
	(2.22)
Constant	4.806***
	(1.07)
Observations	4886

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Raw regression coefficients with standard errors on lead and lag years for Dropout Rates from 1996-2009 for 9 treated districts and 343 control districts. Note the year before treatment has been omitted to avoid multicolinearity and have a relative time reference. This table was used to create Figure 5.