THREE ESSAYS IN APPLIED MICROECONOMICS

AN ABSTRACT

SUBMITTED ON THE SIXTH DAY OF APRIL 2022

TO THE DEPARTMENT OF ECONOMICS

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

OF THE SCHOOL OF LIBERAL ARTS

OF TULANE UNIVERSITY

FOR THE DEGREE

OF

DOCTOR OF PHILOSOPHY

BY

Feng Chen
Feng Chen

APPROVED: Douglas N. Harris, Ph.D.
Douglas N. Harris, Ph.D.
Director

Wei Long
Wei Long, Ph.D.

Kevin Callison
Kevin Callison, Ph.D.

ABSTRACT

This dissertation consists of three independent chapters in applied microeconomics. In the first chapter, I study the causal effect of paid family leave on the infants mortality rate. Using linked U.S. birth and infant death data with a difference-in-differences framework, I find that a six-week PFL in California reduced the post-neonatal mortality rate by 0.135- that is, it saved approximately 339 infant lives. There were fewer deaths from health-related causes and larger effects for infants with married mothers and for infant boys. In the second chapter, my coauthor and I study the combined effects of charter schools on a national level and across multiple outcomes. Using difference-in-differences and fixed effects methods, we find that charter entry (above 10 percent market share) increases the high school graduation rate in geographic districts by about 2-4 percentage points and increases test scores by 0.06-0.16 standard deviations. Also, total effects are comprised not only of participant and competitive effects, but also the charter-induced closure of low-performing traditional public schools. Finally, the third chapter exploits the MeToo movement to study the role of sexism on sexual crime. Measuring the local sexist attitude toward women by the Google search index and a machine learning technique, my coauthor and I show that areas with low sexism experience a higher sexual crime rate than those with high sexism in the MeToo era. An analysis on the potential mechanism shows that the surge in documented sexual crimes in the low sexism areas is primarily sourced from reporting rather than an increase in actual incidents.

THREE ESSAYS IN APPLIED MICROECONOMICS

A DISSERTATION

SUBMITTED ON THE SIXTH DAY OF APRIL 2022

TO THE DEPARTMENT OF ECONOMICS

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

OF THE SCHOOL OF LIBERAL ARTS

OF TULANE UNIVERSITY

FOR THE DEGREE

OF

DOCTOR OF PHILOSOPHY

BY

Feng Chen

Feng Chen

APPROVED: D N H

Douglas N. Harris, Ph.D.
Director

Wei Long

Wei Long, Ph.D.

Kevin Callison, Ph.D.

ACKNOWLEDGMENT

Throughout the writing of this dissertation, I have received generous help and support from the countless people around me.

Foremost, I would like to express my sincerest gratitude to Professor Douglas N. Harris, who is my dissertation committee chair and is also the coauthor of chapter 2. His expertise was invaluable in formulating the research questions and methodology. His insightful feedback pushed me to sharpen my thinking and brought my work to a higher level. Without his meticulously clear roadmap and step-by-step guidance, I could not have finished my doctoral program timely. I am also grateful to Professor Wei Long, who is my dissertation committee member and coauthor of charter 3. His kindest words of encouragement were a deep source of strength and self-confidence when I felt discouraged. Also, his detailed suggestions over many earlier drafts challenged me to be more critical in thinking and clear in writing and allowed me to become a better researcher. I also thanks Professor Kevin Callison, my dissertation committee member, who has given me such strong boosts for the papers in their formative periods that, especially for chapter 1. Without his help, I just cannot imagine how the essay could have grown up beyond its initial stage. His willingness to listen to my ideas and to improve upon them has provided me with a sense of optimism and vigor that I was clearly lacking.

I would also like to thank Professors John Edwards (Graduate Program Director) for their advice and assistance during my graduate studies, and to all the faculty of the Economics Department at Tulane University for their contributions to my education, whether through formal instruction, seminar discussions, or informal conversation. I am also grateful to Erin Callhover for her assistance with administrative matters, and to the Ph.D. students in our program who have provided moral support, friendship, and many interesting discussions.

In addition, I could not have completed this dissertation without the support of my friends and neighbors, who provided stimulating discussions as well as happy distractions to rest my mind outside of my research.

Most importantly, I am grateful to my parents and my sister, for their love, trust, care, support, and many things I lack words for. I should have spent more time with all of them.

LIST OF FIGURES

LIST OF TABLES

Chapter 1

Chapter 2

Chapter 3

CONTENTS

CHAPTER 1 DOES PAID FAMILY LEAVE SAVE INFANT LIVES? EVIDENCE FROM CALIFORNIA'S PAID FAMILY LEAVE PROGRAM

1. Introduction

Paid family leave (PFL), or paid parental leave, is designed to provide compensated time off from work for parents to care for their infants, which is essential to child development (Baker and Milligan 2008, 2010, 2015, Liu and Skans 2010, Dustmann and Schönberg 2012, Carneiro, Løken, and Salvanes 2015, Dahl et al. 2016, Danzer and Lavy 2018, Bullinger 2019). Countries have taken different ways of creating maternity leave legislation to improve the welfare of families. For example, 25 of 34 OECD countries guarantee at least six months of paid leave for mothers to care for their infants (Raub et al. 2018), but women in the U.S. are only entitled to twelve weeks of unpaid leave. In 2004, California became the first state in the U.S. to offer six weeks of paid family leave (CA-PFL) for eligible workers. The paid time off allows for increased maternal-child interactions and better monitoring of children's health status, prolongs breastfeeding, and thereby benefits early childhood outcomes (Rossin-Slater, Ruhm, and Waldfogel 2013, Huang and Yang 2015, Baum and Ruhm 2016, Lichtman-Sadot and Bell 2017, Bartel et al. 2018, Pihl and Basso 2019).

Recent literature shows that children may benefit from parental leave if their mothers take prenatal leave. For example, Stearns (2015) found that in the U.S. paid prenatal leave through Temporary Disability Insurance (TDI) reduced the share of low-birth-weight births by 3.2 percent and decreased the likelihood of early-term births by 6.6 percent. In addition to the benefit to children before birth, parental leave may also influence infant health, and ultimately reduce infant deaths, through the following channels. First, paid parental leave could lead to more investment in parental care, lessening the need for non-parental care which is associated with increased risks of many infectious illnesses, e.g., diarrheal illness and respiratory infections. Second, more time off from work may allow parents to arrange preventative care for their children, such as immunizations and well-child visits, more easily (Berger, Hill, and Waldfogel 2005). Third, women with longer parental leaves can increase their breastfeeding duration (Huang and Yang 2015, Pac et al. 2019), and recent studies have found that longer breastfeeding duration is associated with a reduction in the risk of post-neonatal death. Fourth, parental leave could improve parents' mental health (Bullinger 2019), enabling them to be more attentive to an infant's needs.[1] Finally, in contrast to unpaid leave, paid family leave provides compensating benefits which could be used for better nutrition for children. Evidence from studies of transfer programs in the U.S. (e.g., earned income tax credit, food stamps, and WIC[2] program) has shown benefits for infant health outcomes (Almond, Hoynes, and Schanzenbach 2011).

Previous literature in economics has shown that parental leave can reduce the infant mortality rate (IMR), especially the post-neonatal mortality rate (PNMR) (Ruhm 2000, Tanaka 2005, Rossin-Slater 2011). However, most studies focused on European countries and the 20th century, where there was widespread adoption or expansion of parental leave. For example, Ruhm (2000) used aggregated data on 16 European countries from 1969 to 1994 and found that a 10-week extension of paid leave was predicted to reduce the PNMR by 3.7%-4.6%. Similarly, Tanaka (2005) extended Ruhm (2000) by adding U.S. and Japan from 1969 to 2000 and found similar results. Both studies found little or

[1] I thank one anonymous reviewer for reminding me of this mechanism.
[2] WIC is the special supplemental nutrition program for Women, Infants, and Children (WIC), which provides federal grants to states for supplemental foods, health care referrals, and nutrition education for low-income pregnant, breastfeeding, and non-breastfeeding postpartum women, and to infants and children up to age five who are found to be at nutritional risk.

no effect of unpaid leave. On the contrary, Rossin-Slater (2011) exclusively examined the twelve weeks unpaid leave of the 1993 Family and Medical Leave Act (FMLA) in the U.S. and found that it reduced PNMR by 10% for children with college-educated and married mothers as they were more likely eligible for the unpaid leave.

In this study, I examine the causal effect of CA-PFL on infant mortality using cohort-linked birth and infant death data from the National Vital Statistics System (NVSS) with a difference in differences (DD) framework. The outcome of interest is the PNMR, defined as infant deaths (between 28 and 365 days) per 1,000 live births – this generally overlaps with the periods that CA-PFL can be taken.[3] Using all states other than California as the comparison group, I find that the CA-PFL reduced PNMR by 0.135. There were fewer deaths from health-related causes and larger effects for infants with married mothers and for infant boys. My back-of-the-envelope calculation estimates that the reduction in infant deaths would save approximately $9.7 billion per year assuming that a 12-week national PFL policy had been in effect in 2020.

There is one obstacle in examining the causal effect of PFL on PNMR. Specifically, there might exist contemporaneous shocks that are correlated with the PFL policy and are also beneficial for infant health. If so, the result would be spurious rather than causal. To address the concern of contemporaneous shocks, I use the fetal mortality rate as the placebo outcome because it is less likely to be influenced by CA-PFL but should be impacted by contemporaneous shocks. Furthermore, I performed additional robustness checks using several alternative comparison groups and found that my results are not

[3] The PFL for new mothers runs from six weeks to 12 months after childbirth – new mothers with pregnancy have to start with State Disability Insurance (SDI) first, which provides them six weeks paid leave.

sensitive to comparison groups consisting of states with different backgrounds of family leave policies.

Another difficulty in conducting inference is that there is only one treated unit, which suffers the few clusters problem. Typically, studies that exploit policy variation across states conduct inference using standard errors clustered at the state level. However, this approach may be challenging in cases where the number of treated clusters is small and the conventional cluster-robust standard errors may be underestimated (Bertrand, Duflo, and Mullainathan 2004, Donald and Lang 2007, Conley and Taber 2011). In this study, I follow Ferman and Pinto (2019) to deal with the few clusters problem.

This paper is closely related to the literature on the impact of parental leave policies on infant health outcomes and extends it on several dimensions. First, this paper examines the overall effect of mothers' and fathers' leave on PNMR through the first PFL policy in the U.S. and finds that a six-week PFL program reduced PNMR by 0.135. Stearns (2015) focused on mothers' paid leave through TDI, and Rossin-Slater (2011) focused on unpaid leave which is typically taken only by mothers. Second, the findings in this study may enhance our interpretation of the effects of CA-PFL on early childhood health outcomes (e.g., Lichtman-Sadot and Bell 2017, Pihl and Basso 2019, Bullinger 2019). The earlier studies focused on surviving infants, and if CA-PFL reduced deaths of the most vulnerable infants in California, then their estimates would be higher bounds of these effects. Third, this study examines the heterogeneous effects of PFL for different sub-groups of mothers/infants. This is helpful to our understanding of how such policies would have a different impact on infant deaths and which groups of people are more likely to be influenced by them.

This study is also related to the literature on infant mortality which has shown that it is vulnerable to environmental and economic factors, such as air pollution (Currie, Neidell, and Schmieder 2009, Tanaka 2015), clean water (Mettetal 2019, Heft-Neal et al. 2019), and expenditures (Kiross et al. 2020). This study suggests that infant mortality also could be impacted by the public policy of parental leave.

There are also policy implications here for national PFL programs. Currently, two national PFL programs are under review. One is the Family and Medical Insurance Leave (FAMILY) Act, the other is the American Families Plan (AFP). Better understanding the benefits of CA-PFL may be helpful for policymakers to make their decisions for the two national PFL programs as they share many common elements with CA-PFL.

The paper proceeds as follows. Section 2 discusses family leave policies in the U.S. Section 3 describes the data and presents summary statistics. Section 4 discusses the identification strategy and inference methods. Section 5 presents the main results and the results of heterogeneous analyses. Section 6 discusses robustness checks, threats to identification, and external validity. Section 7 concludes and provides policy implications.

2 Family Leave Policies in the U.S.

The U.S. is the only developed country in the world that does not mandate paid parental leave. The only national policy, the 1993 FMLA, requires employers to provide twelve weeks of unpaid job-protected leave to qualified workers with a newborn or a sick child, or due to a personal or family illness. To be eligible for the FMLA, one must have worked at least 1,250 hours over twelve months for a firm that employs at least 50 workers within 75 miles of its physical establishment. Therefore, only 56 percent of U.S.

employees are eligible for FMLA. This is partly due to the stringent requirements of firm size and the length of time an employee must work for the same employer; further, many eligible workers cannot afford to take three months off without pay (Stearns 2015).

The U.S. 1978 Pregnancy Discrimination Act does require that employers treat pregnancy and childbirth like any other temporary disability. Consequently, five states (California, Hawaii, New Jersey, New York, and Rhode Island) have TDI programs that are required to provide partial wage replacement (50–66 percent) for medical leaves related to pregnancy and childbirth. Workers in California and New Jersey can claim benefits for up to four weeks before the expected delivery date and six weeks after birth (eight weeks for Cesarean sections). The other TDI states provide six to eight weeks of leave that can be used on either side of birth.

On September 23, 2002, the first PFL law in the U.S. was enacted in California; it became effective on July 1, 2004. The program provides six weeks of paid leave for eligible workers who take time off to care for an ill family member or to bond with a new child. Benefits are equal to 55 percent of weekly earnings, up to a weekly cap of $728, as of 2004. The PFL program is funded by the payroll tax on employees' wages; employers make no direct financial contribution. Unlike the FMLA, the CA-PFL is nearly universal in its coverage. Apart from some self-employed persons, all private-sector and nonprofit-sector workers are included, regardless of the size of their employer (Appelbaum and Milkman 2015). Workers need not have been with their current employer for any specific period to be eligible for the PFL; they need only to have earned at least $300 in a job that is covered by the State Disability Insurance (SDI), during any quarter in the 5 to 18 months prior to filing a CA-PFL claim. Most employed mothers in California already qualify for up to four weeks of paid pre-birth leave and six weeks of paid post-birth leave under SDI. Newly pregnant mothers have to start by filing an SDI claim; fathers can take leave through PFL immediately after their child's birth. The PFL does not include job protection unless individuals also qualify for FMLA or the California Family Rights Act (CFRA)[4]. PFL can be taken continuously or intermittently within the first twelve months after a child's birth or adoption.

As of 2021, six states and the District of Columbia (D.C.) had PFL programs in effect: California (2004), New Jersey (2009), Rhode Island (2013), New York (2018), D.C. (2020), Washington (2020), and Massachusetts (2021). PFL programs will soon take effect in Connecticut (2022), Oregon (2023), and Colorado (2024). In addition, two national PFL programs are proposed and are under review. The FAMILY Act, designed to provide twelve weeks of paid leave at a 66% wage replacement rate, was introduced in 2013 but has not been enacted yet. In 2021, President Joe Biden proposed an AFP that is similar to the FAMILY Act: it would guarantee twelve weeks of paid leave to new parents with benefits of 66 to 80 percent of their wages, capped at $4,000 a month. However, the full twelve weeks of paid leave is not expected to be available until the tenth year of the program.[5]

A number of studies have examined the effects of CA-PFL on various outcomes. For example, some found that CA-PFL increased parental leave-taking and improved early childhood outcomes. The leave-taking increased by about five weeks for the average

---

[4] CFRA generally requires employers with 50 or more employees to provide eligible workers unpaid time off to attend to the medical needs of themselves or certain family members.
[5] Fact Sheet: The American Families Plan, retrieved from https://www.whitehouse.gov/briefing-room/statements-releases/2021/04/28/fact-sheet-the-american-families-plan.

covered mother (Rossin-Slater, Ruhm, and Waldfogel 2013, Baum and Ruhm 2016) and by one week for fathers (Baum and Ruhm 2016). Or, fathers were 0.9 percentage points more likely to take leave (Bartel et al. 2018). Huang and Yang (2015) and Pac et al. (2019) concluded that the CA-PFL increased breastfeeding by about 5 percentage points. Lichtman-Sadot and Bell (2017) found improved health outcomes among elementary school children, while Bullinger (2019) found improvements in parent-reported overall child health. Pihl and Basso (2019) reported a decline in infant admissions to hospitals; they concluded that this may be due to more breastfeeding.

3. Data

This paper utilizes the cohort-linked birth and infant death data of the NVSS from the National Center for Health Statistics (NCHS 2020). The microdata contains cohort-linked births and infant deaths occurring in a given calendar year in the U.S., which includes information on birth characteristics (e.g., birth weight, gestational age, birth order, and sex) and maternal characteristics (e.g., age, race, ethnicity, and marital status). I use the data of all singleton births and infant deaths from 2000 to 2008 for analysis. Multiple births are excluded because of the increased risk of prematurity and low birth weight associated with multiple gestations. The sample period stops in 2008 because New Jersey implemented a PFL program in 2009 and I want to have a clear comparison group of states without PFL policy changes during my sample period. The sample period starts in 2000 because I want to have a balanced length of pre- and post-treatment periods (4.5 years pre and 4.5 years post).[6]

[6] Results using extended periods are similar to my main result and are available upon request.

Because there is no unique id linking individual deaths and births, I aggregate the death and birth data to state-month level and then link them by birth state and birth cohort. To reflect the underlying microdata, I use the birth counts in each cell at the aggregated level as the sample weight. The outcome of interest is the PNMR, defined as infant deaths between 28 and 365 days per 1,000 live births in each month at the state level. I include birth and maternal controls in my analyses. The birth controls are birth weight, gestational age, sex at birth, and birth order.[7] The maternal controls are age, race/ethnicity, marital status, educational attainment, employment status, and family income.[8] Maternal educational attainment, employment status, and family income come from the Current Population Survey (CPS) rather than from NVSS because the NVSS data are not comparable across states and years due to the 2003 revisions of the U.S. standard certificates (NCHS 2008).[9] To make the CPS data as representative as the NVSS data, I restrict the CPS sample to women whose youngest child is less than one year old. The final data are aggregated in 5,508 state-month cells for 36,039,789 total births. Table 1 presents the summary statistics of outcome and the controls variables. The PNMR is lower in California than in the comparison group in both pre- and post-PFL periods. However, the PNMR was reduced by 0.12 in California after PFL was effective, while it increased by 0.04 for the comparison group.

---

[7] Birth order is in three categories: first born, second born, and third born or later.
[8] Mother's age is in five categories: 20 years old or less, 21-25 years, 26-30 years, 30-35 years, 36 or more; Mother's race/ethnicity is in four categories: non-Hispanic white, non-Hispanic black, non-Hispanic other, and Hispanic. Education attainment is in four categories: less than high school diploma, high school diploma, some college, college degree and more.
[9] States implemented the 2003 revision of the birth certificate across different years that range from 2003 to 2016. Many data items are common to both the 1989 and 2003 standard birth certificates and are considered directly comparable between revisions. Several key items, however (i.e., educational attainment, tobacco use during pregnancy, month prenatal care began and type of vaginal or cesarean delivery), are not considered comparable between revisions (NCHS 2008).

4. Identification Strategy

To identify the effects of CA-PFL on PNMR, I use the DD method that compares the PNMR in California to that of the comparison group (49 non-CA states plus D.C.) before and after the implementation of PFL. I estimate the effects based on the following equation.

$$Y_{st}= \beta CA_s \times Post_t + \gamma X_{st} + \mu_s + \lambda_t + \varepsilon_{st} \qquad (1)$$

where Y_{st} is the measure of the PNMR in state s and time t (year-by-month); *CA* is an indicator of residence in California; *Post* is an indicator that the birth date was after July 1, 2004; X_{st} is a vector of the birth and maternal controls; μ_s is the state fixed effects; λ_t is the time (year-by-month) fixed effects; ε_{st} is the error term. The key coefficient of interest is β, which measures the DD estimate of the effect of the CA-PFL on PNMR. Standard errors are clustered at the state level.

A key assumption in the DD analysis is that the comparison group provides the appropriate counterfactuals for the trend that the treated state would have followed if it had not been treated – that is, the treated group and the comparison group would have had parallel trends. First, in Figure 1 I plot the raw trends of PNMR throughout the 2000 to 2008 period for California and the comparison group. The trends in PNMR are generally common for both groups before 2004; we can detect a downward trend in PNMR for California after 2004, while there is no similar pattern for the comparison group. More formally, I use an event study model to test for the parallel trends assumption by regressing the outcome on the interaction of the treatment variable (*CA*) with a series of event-time dummies based on the following equation:

$$Y_{st}= \Sigma\, \beta_r CA \times Event_r + \gamma_4 X_{st} + \mu_s + \lambda_t + \varepsilon_{st} \qquad (2)$$

In equation (2), $Event_r$ is a dummy of the r years of leads (+) or lags (-) since the implementation of PFL[10]. For example, $Event_{-1}$ is a dummy of the year from July 2003 to June 2004, $Event_0$ is a dummy of the year from July 2004 to June 2005, and $Event_{+1}$ is a dummy of the year from July 2005 to June 2006. The coefficients β_r are measures of cohort-specific effects compared with the comparison group. I plot the coefficients β_r and its 95% confidence interval in Figure 2. The coefficients of the interaction term are not statistically significant for the birth cohort prior to the implementation of PFL. This suggests that the pre-treatment trends in PNMR do not differ between California and the comparison group, and the states I have selected can be used as a valid comparison group for California.

In studies that leverage policy changes across states, the inference is usually conducted using standard errors that are clustered at the state level. However, the cluster-robust standard errors are underestimated when the number of treated clusters is small. (Bertrand, Duflo, and Mullainathan 2004, Donald and Lang 2007, Conley and Taber 2011). Further, since the number of births varies greatly across states, the residuals in the regression equation tend to exhibit substantial heteroscedasticity. Accordingly, I use a method of inference, developed by Ferman and Pinto (2019), that provides an improvement in the hypothesis testing for situations where there are few or even one treated unit(s) and many control units in the presence of heteroskedasticity. It is important to note that Ferman and Pinto's method is not robust to any form of unknown heteroskedasticity, but it provides an improvement relative to existing methods that rely

[10] The PNMR is noisy at shorter periods (e.g., monthly). Conducting event study analysis at the yearly level would use 12 months before the implementation of the policy as the reference period, which is more reliable. Event study results at monthly, quarterly, and semiyearly levels are available upon request.

on homoskedasticity. Specifically, they model the heteroskedasticity of the pre-post difference in average errors. Under this assumption, they rescale the pre-post difference in average residuals of the control groups using the pre-post difference in average errors. In this way, a cluster residual bootstrap with this heteroskedasticity correction provides asymptotically valid hypothesis testing when the number of control groups goes to infinity, even when there is only one treated group. This method produces a bootstrapped distribution of the pseudo-treatment effects for determining the significance of the estimate of the treatment effects, rather than being a test statistic (Ferman and Pinto 2019).[11] I report Ferman Pinto p-values (F-P p-values) and conventional p-values for all specifications. F-P p-values are my preferred inference results, and conventional p-values are listed only for reference purposes.

5. Results

5.1 Effects of CA-PFL on PNMR

Based on a sample of all singleton births in the U.S. from 2000 to 2008, Table 2 shows the estimates of the effects of CA-PFL on the PNMR.[12] Table 2 presents estimates of equation (1) with three model specifications. In column (1), I consider a baseline model with state and time fixed effects only. The point estimate suggests that there was a significant decrease in PNMR in California after the implementation of CA-PFL. This result hinges on the assumption that there are no omitted time-varying and state-specific

---

[11] The purpose of this bootstrap is to provide asymptotic refinement in the presence of clustered errors. It essentially performs a permutation test by imposing the null hypothesis and resampling the residuals—rather than re-estimating the key coefficient of interest—and then asking how likely it would be to observe the key coefficient by chance. As such, only p-values (not standard errors and R-squared) are generated.

[12] I present a version of estimates using all births in the U.S. during 2000 to 2008 in Table A1 in the Appendix.

factors that correlated with the PNMR. In columns (2) and (3), I relax this assumption by adding a set of time-varying state-level birth controls and maternal controls. The model in Column 3 is my preferred specification as it includes both fixed effects and time-varying controls. The magnitude of the coefficient in Column 3 indicates that the CA-PFL reduced the PNMR by 0.135 at the one percent level of significance, or about an 8 percent reduction of its pre-treatment sample mean (1.65).

Because the data do not contain information on who is eligible to benefit from the CA-PFL, the estimated effect will represent the intention-to-treat (ITT) effect. The treatment-on-the-treated (TOT) effects could be estimated using the ITT effect scaled by the inverse of parents' take-up rate of CA-PFL. One way to estimate the take-up rate is to consider the number of claims divided by the number of likely eligible parent. Table 3 presents the estimates of mothers' and fathers' take-up rates of CA-PFL. I estimate that 44 percent (5 percent) of employed new mothers (employed new fathers) made a bonding claim in 2005, which is similar to estimates of Bana, Bedard, and Rossin-Slater (2018) – 40 percent for mothers and 4 percent for fathers. According to Table 3, the estimated take-up rate for California families is about 27.12%, and the ITT effects could be scaled to 0.5 (0.135*(1/0.27)) to get the TOT effects for both mothers and fathers taking the full length of leave.

5.2 Interpretation

It is worthwhile to compare my estimate to studies that focus on other parental leave studies on PNMR. Table 4 presents the comparison of the effect size (ITT effects) in this study with those in related studies. Ruhm (2000) examined 16 European countries and found that the equivalent effect of a one-week extension on PNMR is 0.020; Tanaka

(2005) extended Ruhm's (2000) study by adding the U.S. and Japan and found a similar result (0.015). Rossin-Slater (2011) reported that PNMR is reduced by 0.017 (one-week equivalent effect) for infants with highly educated and married mothers after the FMLA becomes effective. In this study, I find that the one-week equivalent effect on PNMR is 0.023, which is similar in magnitude but slightly larger than that of previous studies.

5.3 Heterogeneous Effects

Thus far, I have considered PNMR for all births. However, the heterogeneous effects on subgroups of infants are also important to understanding how PFL affects PNMR. In this section, I conduct several analyses according to the cause of death, maternal race, maternal marital status, and sex of birth.

First, I examine the heterogeneous effects of PFL on PNMR by causes of infant deaths which can be categorized as health-related and non-health-related causes.[13] The health-related causes include infectious diseases, nutritional diseases, respiratory diseases, and influenza. The non-health-related causes include accidents and assault, for example. The results in Panel A of Table 5 suggest that the reduction in PNMR is driven by health-related causes rather than non-health-related causes. This result is supportive of two mechanisms mentioned earlier: parental care and preventive care. With more time bonding with children, parents can better monitor children's health status and children are more likely to receive timely medical treatment. Also, PFL allows parents to arrange appropriate preventative care more easily for their children. Thus, PFL reduces PNMR for health-related causes rather than for non-health-related causes. Death from external causes may occur stochastically, and it is nearly impossible for parents to anticipate it and

[13] Deaths with NCHS's ICD-10 130 Groups for selected causes of infant mortality codes from 001 to 137 are classified as health related, while those codes from 138 to 158 are in the non-health related category.

to take leave in advance to avoid it. This is also consistent with Rossin-Slater's (2011) finding that FMLA has no impact on infant deaths with non-health-related causes.

I next examine the heterogeneous effects of CA-PFL by mothers' race/ethnicity and marital status. Panel B and C in Table 5 present these results. Only the results for births to married mothers are statistically significant using the Ferman-Pinto inference method. The results for births to non-Hispanic black and non-Hispanic white mothers and for unmarried mothers are only statistically significant when using the cluster robust inference method, which is less reliable. Why are there larger effects on infants with married mothers? First, the likelihood of having more total leave per birth is higher among married couples because both parents may be eligible for leave. Second, given that the PFL only has partial wage replacement, married mothers are more likely to be able to afford the partial paid leave than single mothers because married couples may have dual incomes.

Finally, I investigate the effect on PNMR by sex of births. The results in Panel D suggest that there was a larger, and significant reduction in PNMR for infant boys than for infant girls. Infant mortality is often higher in boys than girls, which has been explained by sex differences in genetic and biological characteristics: boys are biologically weaker and more susceptible to infectious diseases and adverse risks (Pongou 2013). This result is supportive of the mechanism that the PFL might lessen the need for non-parental care, and non-parental care is often associated with an increased risk of many infectious illnesses. It is possible that the reduction in non-parental care reduced the risk of infant deaths caused by infectious diseases, and therefore that the effect was disproportionately larger on infant boys who are more vulnerable to infectious diseases.

6. Robustness

6.1 Alternative Comparison Groups

So far, I used states other than California as the comparison group for the main analysis. However, some may have concerns that this is not a good comparison group because they are different from California in terms of family leave policies. Ness, Shabo, and Fink (2016) ranked all states based on policies that support expecting and new parents, and California was ranked first.[14] To further explore the sensitivity of my estimates, I, therefore, conduct robustness checks using comparison groups based on family leave related policies. Specifically, I use: (1) states with (past) TDI programs (Hawaii, New Jersey, New York, and Rhode Island); (2) states without TDI programs; (3) states and D.C. with (future) PFL programs (New Jersey, Rhode Island, New York, D.C., Washington, Massachusetts, Connecticut, Oregon, and Colorado); (4) states without PFL programs[15]; (5) the top 25 family-friendly states (other than California); and (6) the bottom 25 family-friendly states as the alternative comparison groups.[16] Figure 3 presents the estimates using all non-CA states plus D.C. as the comparison group and

---

[14] They assess state laws and policies that guarantee access to family or medical leave to expecting and new parents, paid sick days, reasonable accommodations for pregnant workers and support for breastfeeding mothers.

[15] However, one drawback of using TDI states and PFL states as the comparison groups is that Ferman and Pinto's inference method is not applicable as there are only very limited control units (N<10). Ferman-Pinto inference method works well for cases of few treated units and many control units, however, when the number of control units is small (e.g., N<10), it is difficult to estimate the conditional variance accurately. According to Ferman (personal communication, 2021), it is hard to tell how many control units are sufficient to use this method, but from the simulations in their paper, an N>=20 is good to use.

[16] Table A2 in the Appendix lists ranks of all states based on the assessment of family-friendly policies.

then using the above-mentioned comparison groups, respectively. Figure 3 shows that the estimates using all comparison groups are statistically significant and very similar in magnitude; this indicates that my estimates are not sensitive to comparison groups having different family leave policies.

To further examine the sensitivity of the main result, I use 25 randomly chosen states as the comparison group and repeated 1,000 times. The resulting distribution of estimates is displayed in Figure 4. As it shows, all coefficients are negative and range from -0.059 to -0.125 with a mean of -0.135, and the 95% confidence interval is [-0.085, -0.185]. Overall, this suggests that the main result is robust and not sensitive to using different comparison states.

6.2 Placebo Check

One threat to identification is that the reduction in PNMR is due to contemporaneous shocks. For example, if there were unknown factors or contemporaneous shocks (e.g., less air pollution or more clear water) in California that also benefit infant health, then the result would be a correlation rather than a causation. To address this concern, I use the fetal mortality rate as the placebo outcome and redo the analyses. Fetal mortality is infant death during pregnancy, so it should not be influenced by CA-PFL but could be influenced by other factors or contemporaneous shocks that benefit infant health. The results in Table 6 indicate that CA-PFL has no significant impact on the fetal mortality rate, which suggests the reduction in PNMR in California is less likely due to other unknown factors or contemporaneous shocks.

One may also be interested in the effect of CA-PFL on the neonatal mortality rate. Neonatal mortality is defined as infant deaths within 28 days of birth. Recall that mothers with new births start with a six-week SDI first and then take the PFL, so the neonatal period may less likely be influenced by the mother's leave through PFL. However, fathers can take leave immediately after birth, even though their take-up rate is only about 5%. Also, PFL may affect the mother's use of neonatal leave if it raises awareness of SDI pregnancy-related benefits. For these reasons, access to PFL may affect parents' neonatal leave; therefore, neonatal mortality may not be a clean placebo outcome. I also conducted an analysis using the neonatal mortality rate as the outcome, and the results presented in Table A3 in the Appendix suggest that CA-PFL has no significant impact on the neonatal mortality rate.

Another placebo examination assumes that the treatment state is a different state and to see if there is a similar effect. Specifically, I replicate the estimation of equation (1) but assume that the treatment state is a different state. I repeat this procedure for all states other than California plus D.C. and then plot the coefficients and F-P p-values in Figure 5. As it shows, the F-P p-values are generally randomly distributed,[17] and there are six estimates with F-P p-values less than 0.1.[18] I then conduct event study analyses for the six states, respectively, and none of them show a similar pattern to that of California or display any meaningful patterns.[19] I also conduct another, robustness check that excludes these six states from the analysis and the result is consistent with that of including them.[20] Overall, there is little evidence that the effects of CA-PFL on PNMR are driven by inappropriate identification assumptions, and the effect is indeed not spurious but causal.

[17] Specifically, 6 of them in range of 0 to 0.1, 6 of them in range of 0.1 to 0.2, 4 of them in range of 0.2 to 0.3, 5 of them in range of 0.3 to 0.4, 4 of them in range of 0.4 to 0.5, 3 of them in range of 0.5 to 0.6, 8 of them in range of 0.6 to 0.7, 7 of them in range of 0.7 to 0.8, 5 of them in range of 0.8 to 0.9, and 2 of them in range of 0.9 to 1.

[18] These states are Illinois, Louisiana, Maine, Montana, North Carolina, and South Carolina.

[19] Figure A1 in the Appendix presents the figures of these event studies.

[20] Table A4 in the Appendix presents the results.

6.3 Other Threats to Identification: Migration, Fertility, and Birth Outcomes

The CA-PFL was announced on September 23, 2002 and became effective on July 1, 2004. Some may have concerns that the 21-months-prior announcement may have made it possible for pregnant women in other states to migrate to California to take advantage of this policy. However, the maximum weekly benefit of the CA-PFL program was $728 before 2012, or $4,368 for six weeks, which is less than the average cost of an interstate move ($5,630).[21] The relatively small financial incentive is not sufficient enough to encourage mass migration of pregnant women in other states to California.

Another threat to the identification is that the CA-PFL may induce a change in fertility and, thereby affect the PNMR by changing the number of new births. This could happen if some women find that motherhood would be more appealing when they have access to PFL. Previous studies have examined the impacts of the CA-PFL on fertility and found no evidence of a response or in changes in the composition of births after the policy (Rossin-Slater, Ruhm, and Waldfogel 2013, Pihl and Basso 2019). However, Lichtman-Sadot (2014) found some shifts in the number of births from the earlier part of 2004 to the latter part. To address the concern over fertility changes in 2004, I exclude that year and redo the analysis; the results are consistent with those that include 2004.[22] To formally examine whether CA-PFL impacts fertility during the sample period, I perform additional analyses using the general fertility rate and the log of the number of births as the outcomes in the DD regressions, and the results are reported in Table 7.

[21] According to the American Moving & Storage Association, the average cost of an interstate move is $5,630 in 2016.
[22] Table A5 in the Appendix presents the results.

There is little evidence indicating that there is a fertility change in new births due to the CA-PFL.

I also analyze the effects of CA-PFL on two measures of birth outcomes – low birth weight and preterm birth, two of the leading causes of infant death (Ely and Driscoll 2020).To examine if the CA-PFL impacts birth outcomes and then ultimately affects post-neonatal deaths, Table 7 reports the estimates of the effects of CA-PFL on birth outcomes. There are no significant effects on either birth outcome using the Ferman-Pinto inference method, which is consistent with Bullinger (2019). Overall, the reduction in PNMR is less likely to be explained by better health outcomes at birth.

6.4 External Validity

So far, I have evaluated the effect of PFL on PNMR in CA, but it is also important to know if a similar effect would be expected for other states with similar programs. Given that New Jersey (in 2009) and Rhode Island (in 2014) also implemented a similar PFL program, and that the data is also available to evaluate them, I followed Callaway and Sant'Anna (2021)'s method of staggered DD to explore the effect of PFL in CA, NJ, and RI on PNMR. I did two different versions of the staggered DD. One uses the balanced panel data from 1999 to 2017; the other uses the balanced pre-and post-period data, four years before and after implementation. Table 8 presents the estimates of the overall effect. The estimates are only statistically significant in specifications without controls or with birth controls. However, the parallel trend assumption does not hold for NJ and RI; thus, these estimates would be more like suggestive evidence.

Alternatively, the synthetic control method is useful for constructing a comparison group with parallel trends when there is only one treated unit and many control units.

Thus, I estimate the effect of PFL on PNMR in CA, NJ, and RI using that method. The result, as Figure 6 presents, is that the synthetic state shares parallel trends in PNMR with the treated state before the effects of the PFL program. The PNMR in all three states was reduced after the implementation of the PFL program, and this was especially true for CA and NJ. Galiani and Quistorff (2017) proposed a method for inference for the synthetic control method by comparing the estimated main effect with the distribution of placebo effects (estimations for the same treatment period but on all the control units). It could also provide inference for several units that received treatment at different times. I followed their method for inference and the results are in Table 9. As shown, the overall estimates and the estimates for all states are negative. However, the estimates are only statistically significant for CA and the overall effect of the three programs but are not statistically significant for NJ and RI individually.

Overall, the results of the staggered DD and synthetic control methods provide suggestive evidence that PFL also reduced PNMR in NJ and RI; along with the solid evidence in CA, it is supportive that we should expect a similar effect for a national PFL program.

7. Conclusion and Policy Implication

The PFL aims to help working parents balance their careers and family responsibilities, which is essential to child development. The benefits of PFL on infant mortality previously have been documented in large cohort studies using data from European countries, where there has been widespread adoption of paid family leave at a national level. This study examines the first PFL program in the U.S. and finds that a six-week PFL reduced PNMR by 0.135, saving approximately 339 infant lives in California from

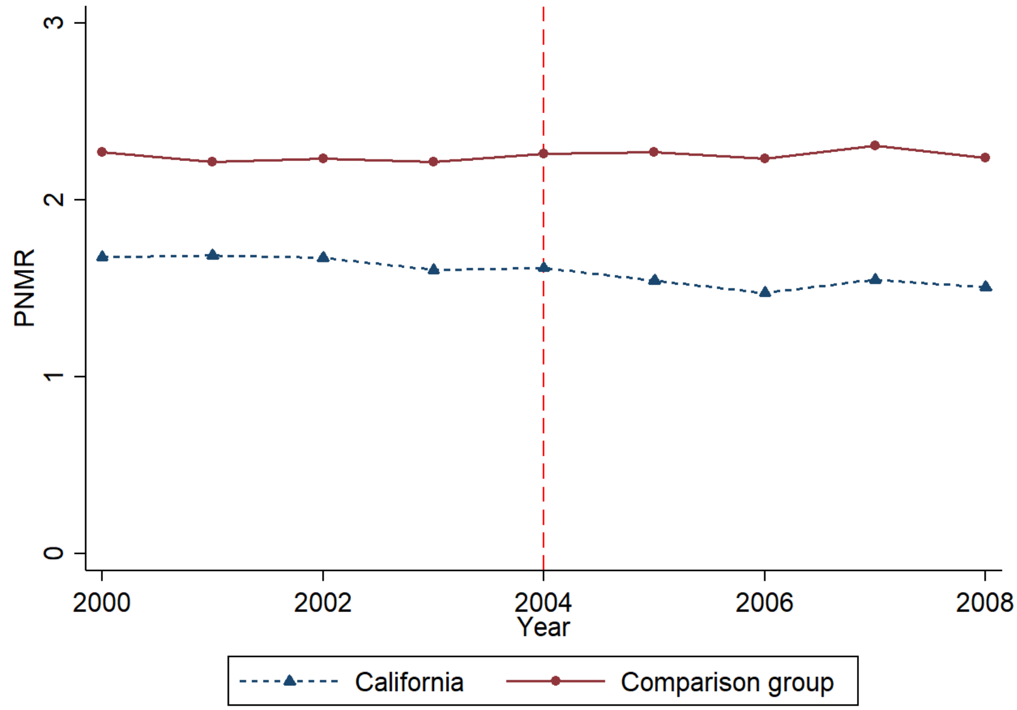
2004 to 2008.[23] Heterogeneous analyses suggest that there were fewer deaths from health-related causes and larger effects for infants with married mothers and for infant boys.

One policy implication of this study is that the benefits of the PFL program may be understated if the effects on infant mortality are not taken into account. This is especially significant as more states in the U.S. are developing their own PFL program and the national PFL plans, FAMILY Act and AFP, are currently under review. The FAMILY Act and AFP were estimated to cost approximately $228 billion and $225 billion across 10 years, respectively.[24] Therefore, the average cost of a 12-week PFL plan would cost around 23 billion per year. The value of a statistical life (VSL) is commonly used by policy analysts and researchers to estimate life values, even though human life is priceless – estimates of the VSL for the U.S. are around $10 million (Kniesner and Viscusi 2019). In this study, I find that a six-week PFL reduces PNMR by 0.135. Assuming a similar effect for a national PFL policy like FAMILY Act or AFP, the 12-week PFL plan could reduce the infant mortality rate by 0.27. In 2020, there were about 3.6 million new births, which implies that approximately 972 additional infants could survive to one year of age if the FAMILY Act or AFP were effective. When a statistical life is valued at $10 million, the reduction in infant mortality is worth approximately $9.7 billion per year. The estimated dollar benefit is substantial and is nearly half of the estimated cost. The back-of-the-envelope calculations of benefits associated with the

[23] 339=0.135*(the number of total births in this period)/1000.
[24] The cost of the FAMILY Act was estimated by the Congressional Budget Office. https://www.cbo.gov/publication/56129 The cost of the AFP is from the *Fact Sheet: The American Families Plan*, retrieved from https://www.whitehouse.gov/briefing-room/statements-releases/2021/04/28/fact-sheet-the-american-families-plan.

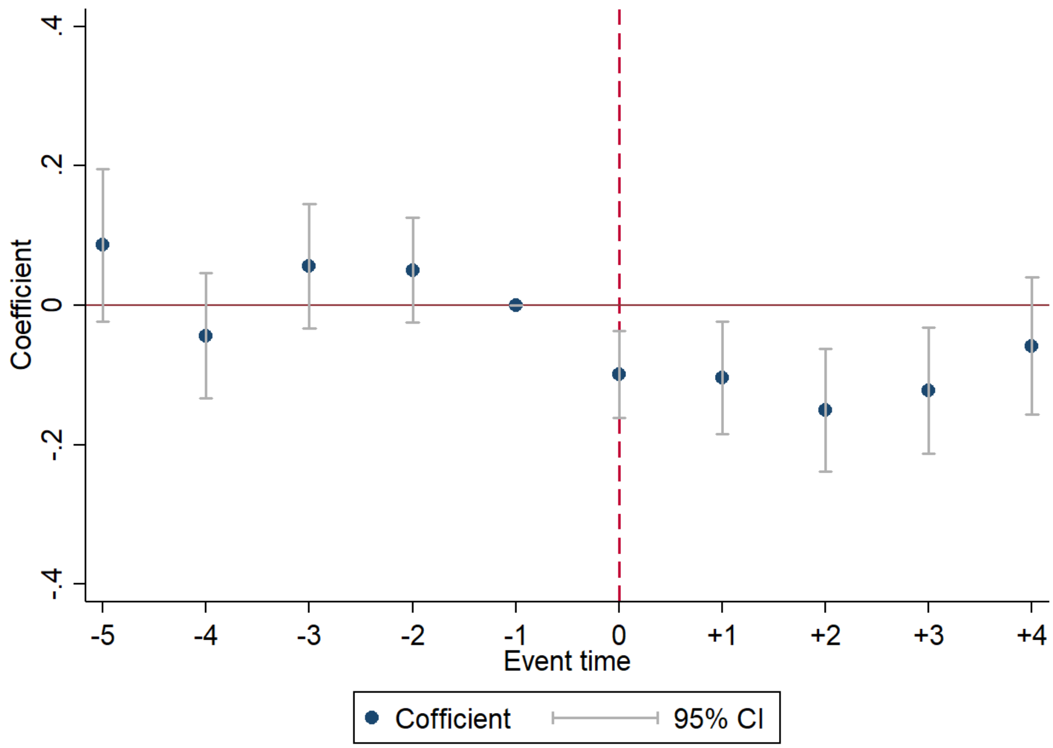
reduction in infant deaths might be helpful for policymakers to make decisions on the national PFL programs.

Figure 1 Raw trends in PNMR in California and the comparison group, 2000-2008



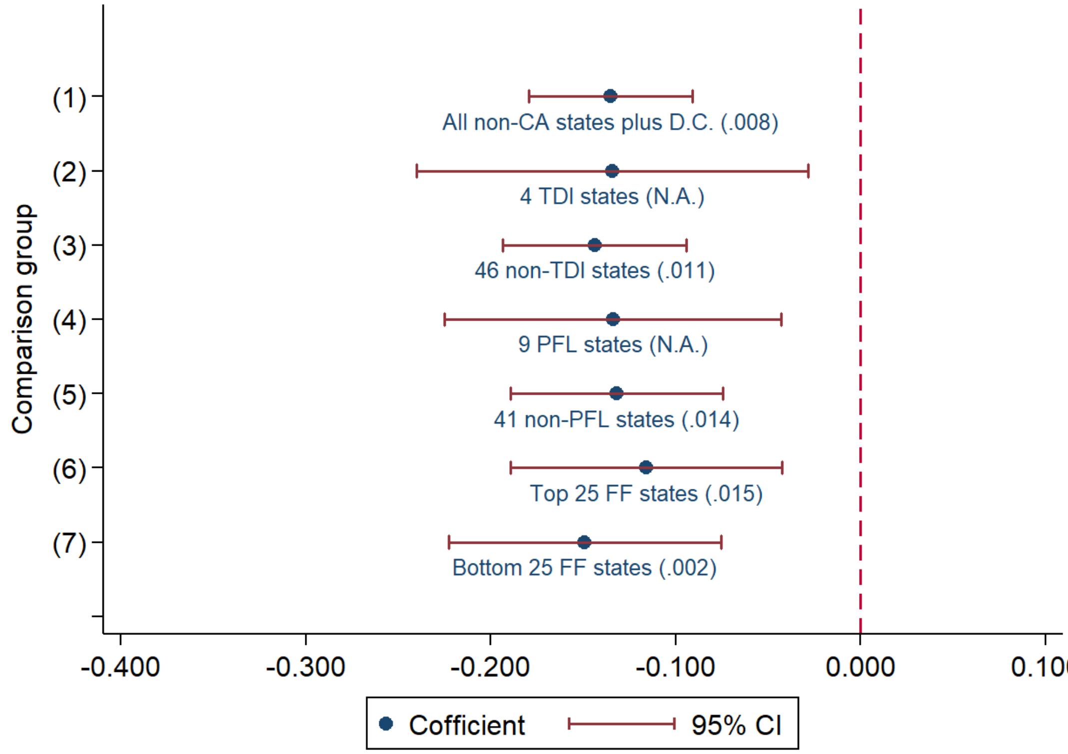
Notes: This figure plots the raw trends in PNMR in California and the comparison group.

Figure 2 Event study estimates of effects of CA-PFL on PNMR



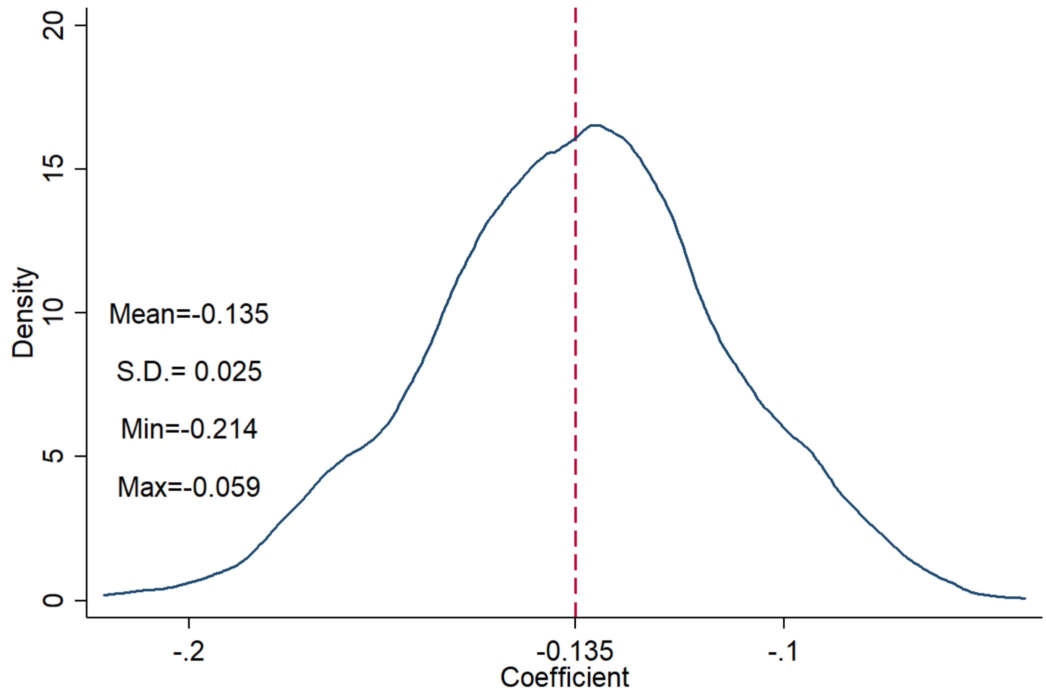
Notes: This figure displays coefficients and 95% confidence intervals of event study estimates. Event time is a dummy of the year(s) of leads or lags since CA-PFL is effective, for example, the event time 0 is a dummy of the year PFL effective (July 2004 to June 2005).

Figure 3 Robustness checks: alternative comparison groups



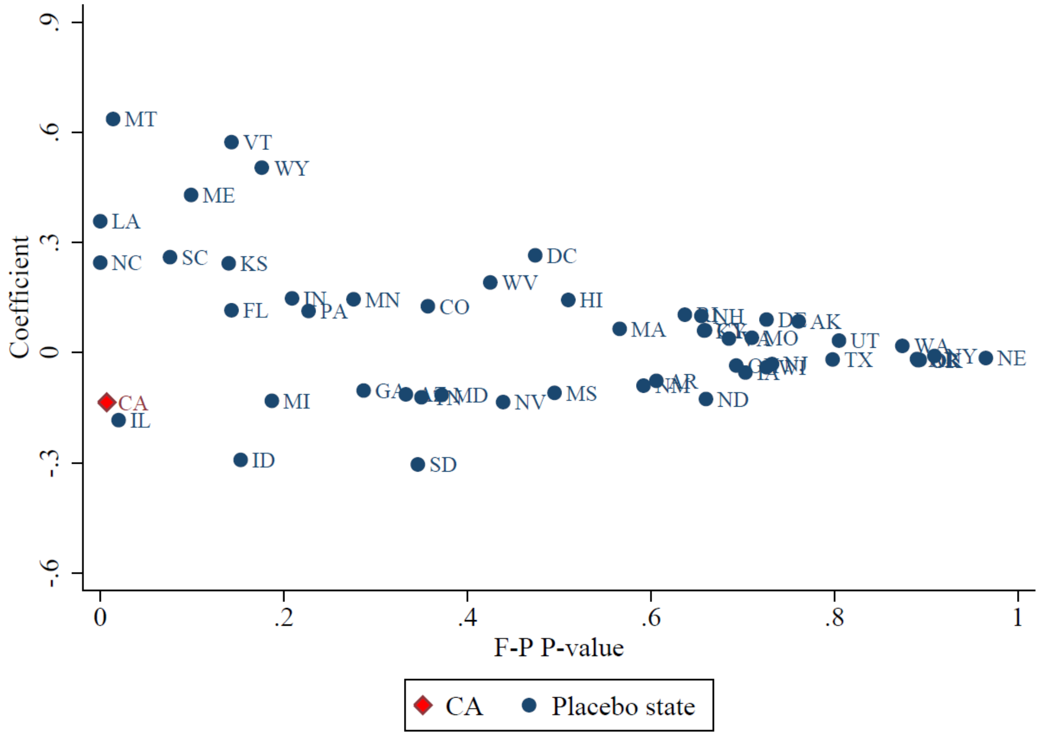
Notes: This figure plots coefficients and 95% confidence intervals of estimates using alternative comparison groups. The 95% confidence intervals are based on the conventional cluster-robust standard errors, and F-P p-values are in parentheses (if applicable). Estimate (1) is the same as estimate in column (3) of Table 3. TDI comparison states are New York, New Jersey, Rhode Island, and Hawaii. PFL comparison states are New Jersey, Rhode Island, New York, D.C., Washington, Massachusetts, Connecticut, Oregon, and Colorado. See Table A2 in Appendix for the list of the top 25 family friendly (FF) states and the bottom 25 FF states.

Figure 4 Permutation results using 25 randomly chosen states as the comparison group



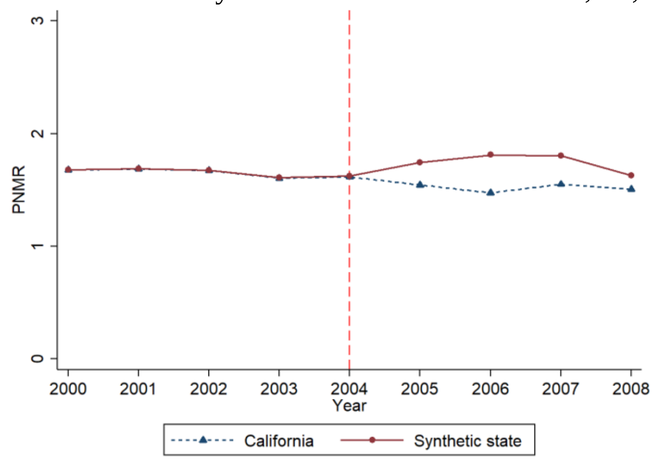
Notes. This figure plots the density distribution of estimates of using 25 randomly chosen states as the comparison group and permuted 1,000 times. The vertical dashed line corresponds to -0.135, the estimate of our preferred specification in column (3) of Table 3.

Figure 5 Results of placebo tests using every other state as the treated state

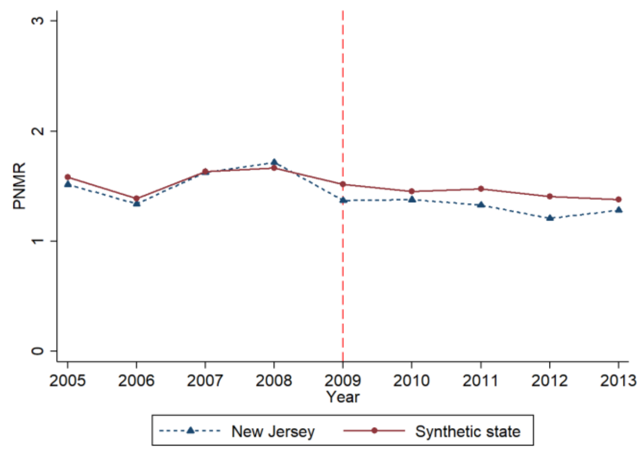


Notes: This figure plots coefficients and F-P p-values of placebo tests using every other state as the treated state. The solid diamond dot is the main result that using California as the treated state, and the solid circle dots are results of placebo tests using every other state as the treated state. The F-P p-values are generally randomly distributed: 6 of them in range of 0 to 0.1, 6 of them in range of 0.1 to 0.2, 4 of them in range of 0.2 to 0.3, 5 of them in range of 0.3 to 0.4, 4 of them in range of 0.4 to 0.5, 3 of them in range of 0.5 to 0.6, 8 of them in range of 0.6 to 0.7, 7 of them in range of 0.7 to 0.8, 5 of them in range of 0.8 to 0.9, and 2 of them in range of 0.9 to 1.

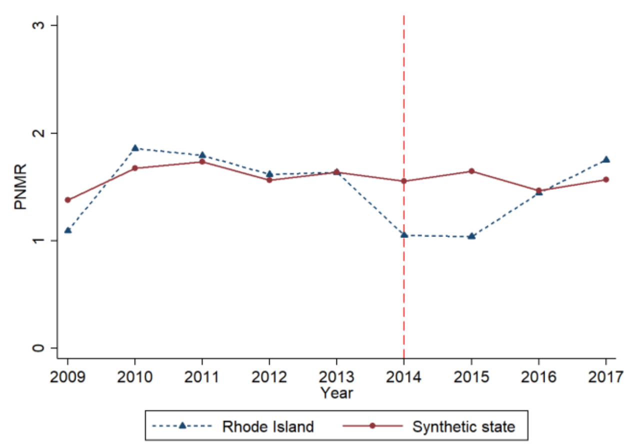
Figure 6 Results of the synthetic control method for CA, NJ, and RI



(a) California



(b) New Jersey



(c) Rhode Island

Note: This figure presents the synthetic control results for CA, NJ, and RI. The red vertical line is the year that the PFL policy was effective.

Table 1 Summary statistics

Variable	All	Pre-CA-PFL		Post-CA-PFL	
		CA	Comparison	CA	Comparison
Outcome of interest					
Post-neonatal mortality rate	2.16	1.65	2.23	1.53	2.27
Placebo outcome					
Neonatal mortality rate	3.81	3.07	4.01	2.97	3.85
Fetal mortality rate	5.97	5.25	6.25	4.92	5.98
Fertility outcome					
General fertility rate	65.20	67.52	63.51	69.06	65.88
Number of births	15,798	43,114	11,202	45,193	11,751
Birth control					
Birth weight	3,317	3,372	3,328	3,338	3,294
Gestational age	38.75	38.98	38.79	38.83	38.66
Male	0.51	0.51	0.51	0.51	0.51
First born	0.40	0.39	0.41	0.40	0.41
Second born	0.32	0.32	0.32	0.31	0.32
Third or later born	0.28	0.29	0.27	0.29	0.28
Maternal control					
Age<=20	0.16	0.14	0.16	0.14	0.15
20<Age<=25	0.26	0.24	0.26	0.24	0.27
25<Age<=30	0.27	0.27	0.27	0.27	0.28
30<Age<=35	0.21	0.23	0.21	0.23	0.20
Age>35	0.11	0.13	0.10	0.13	0.10
Non-Hispanic black	0.14	0.06	0.16	0.06	0.16
Non-Hispanic white	0.56	0.31	0.61	0.28	0.58
Non-Hispanic other	0.06	0.11	0.05	0.11	0.05
Hispanic	0.24	0.51	0.18	0.54	0.21
Married	0.63	0.67	0.66	0.62	0.61
Less than high school completion	0.17	0.23	0.16	0.23	0.15
High school diploma	0.28	0.26	0.29	0.24	0.28
Some college	0.26	0.26	0.26	0.26	0.27
Bachelor's degree or higher	0.29	0.25	0.29	0.28	0.30
Share of employed	0.51	0.45	0.51	0.44	0.52
Family income	47,950	51,274	47,797	47,956	47,613
N	5,508	54	2,700	54	2,700

Notes: The table presents the summary statistics (means) of the outcome and control variables obtained from the NVSS for the whole sample, California (pre- and post-CA-PFL samples), and the comparison group (pre- and post- CA-PFL samples), from 2000 to 2008. The comparison group is the all non-CA states plus D.C.

Table 2 Effects of CA-PFL on the PNMR

	(1)	(2)	(3)
CA*Post	-0.155	-0.161	-0.135
P-value	(0.000)	(0.000)	(0.000)
F-P p-value	[0.098]	[0.050]	[0.008]
R-squared	0.456	0.458	0.460
Observations	5,508	5,508	5,508
State FE, Time FE	Y	Y	Y
Birth control	N	Y	Y
Maternal control	N	N	Y

Notes: The table presents the DD estimates of the effects of the CA-PFL on PNMR. The birth controls include birth weight, gestational age, sex of birth, and birth order; and the maternal controls include maternal age, race/ethnicity, marital status, educational attainment, employment status, and family income. All regressions are clustered at the state level and weighted by the number of births in each state-month cell. The cluster-robust p-values are in parentheses, and the Ferman-Pinto p-values are in brackets.

Table 3 Estimates of the CA-PFL take-up rate

Year	Number of bonding claims		Children for bonding		Eligible parent%		Take-up rate	
	Mother	Father	New birth	Adoption	Mother	Father	Mother	Father
2004	56,279	10,178	282,643	3,778	45.09%	91.12%	43.57%	3.90%
2005	112,155	24,810	548,882	7,556	44.21%	88.29%	45.59%	5.05%
2006	118,112	28,223	562,440	7,393	42.13%	90.71%	49.19%	5.46%
2007	127,754	33,804	566,414	7,622	43.67%	90.54%	50.97%	6.50%
2008	137,566	39,833	551,804	7,777	45.91%	90.34%	53.55%	7.88%
Total	551,866	136,848	2,512,183	34,126	44.20%	90.20%	48.58%	5.76%

Notes: The table presents the estimates of the CA-PFL take-up rate. The take-up rate = (number of bonding claims) ÷ (sum of new births and adoptions × eligible parent %). For example, take-up rate in 2005 for mothers: 45.59%=112,155 ÷ ((548,882 + 7,556) × 44.21%); and for fathers: 5.05%=24,810 ÷ ((548,882 + 7,556) × 88.29%). The number of bounding claims is from the California Employment Development Department (2020). The number of total births is from NCHS (2020), and the number of births in 2004 is the total births from July to December of 2004. The number of adopted children (2005-2008) is from the U.S. Department of Health and Human Services (2015), and the number of adopted children in 2004 is estimated as half of the number in 2005. The percent of eligibility is estimated as the share of the employed parent with the youngest child less than one-year old using data from CPS.

Table 4 Comparison of effect size in this study with that of previous studies

Study	Sample period	Country	Effect	Mean	Percent change	Length (week)	1-week effect
Ruhm (2000)	1969–1994	16 European countries	0.20	4.30	5%	10	0.020
Tanaka (2005)	1969–2000	16 European countries, U.S., and Japan	0.15	3.60	4%	10	0.015
Rossin-Slater (2011)	1989–1997	U.S. (FMLA)	0.20	2.00	10%	12	0.017
This study	2000–2008	U.S. (CA-PFL)	0.14	1.65	8%	6	0.023

Note: This table presents the comparison of effect size in this study with that of previous studies. The "Effect" column is the (ITT) effect on PNMR. The "1-week effect" column is the estimate divided by leave length assuming linear effect.

Table 5 Heterogeneous effects of CA-PFL on the PNMR

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A Cause of death: health-related vs non-health-related						
Group (mean)	Health-related Cause (1.51)			Non-health-related Cause (0.14)		
CA*Post	-0.151	-0.160	-0.147	-0.004	-0.002	0.012
P-value	(0.000)	(0.000)	(0.000)	(0.754)	(0.906)	(0.452)
F-P p-value	[0.136]	[0.063]	[0.031]	[0.922]	[0.972]	[0.789]
R-squared	0.394	0.397	0.398	0.281	0.283	0.286
Panel B Race: Non-Hispanic black vs non-Hispanic white						
Group (mean)	Non-Hispanic Black (3.80)			Non-Hispanic White (1.50)		
CA*Post	-0.280	-0.242	-0.305	-0.137	-0.119	-0.067
P-value	(0.002)	(0.006)	(0.001)	(0.000)	(0.000)	(0.004)
F-P p-value	[0.676]	[0.710]	[0.602]	[0.163]	[0.179]	[0.333]
R-squared	0.097	0.104	0.107	0.331	0.333	0.336
Panel C Mother's marital status: married vs unmarried						
Group (mean)	Married (1.28)			Unmarried (2.39)		
CA*Post	-0.152	-0.176	-0.167	-0.050	-0.076	-0.064
P-value	(0.000)	(0.000)	(0.000)	(0.363)	(0.195)	(0.310)
F-P p-value	[0.004]	[0.004]	[0.001]	[0.895]	[0.802]	[0.818]
R-squared	0.273	0.275	0.277	0.280	0.286	0.288
Panel D Child sex: female vs male						
Group (mean)	Female (1.48)			Male (1.82)		
CA*Post	-0.100	-0.101	-0.078	-0.207	-0.224	-0.201
P-value	(0.001)	(0.001)	(0.017)	(0.000)	(0.000)	(0.000)
F-P p-value	[0.379]	[0.362]	[0.404]	[0.031]	[0.015]	[0.004]
R-squared	0.271	0.273	0.274	0.333	0.336	0.338
Observations	5,508	5,508	5,508	5,508	5,508	5,508
State FE, Time FE	Y	Y	Y	Y	Y	Y
Birth control	N	Y	Y	N	Y	Y
Maternal control	N	N	Y	N	N	Y

Notes: The table presents the heterogeneous effects of the CA-PFL on PNMR. The pre-treatment mean of PNMR is in parentheses. The observation for the heterogeneous analysis applies to all subgroup analyses as all analyses were conducted at the state by month level. At the state by month level, each state in each month has one observation of the outcome variable for each subgroup, thus, the number of 5,508 observations is applied for to subgroup analyses. See notes to table 2 for other details.

Table 6 Placebo outcome: Fetal mortality

	(1)	(2)	(3)
CA*Post	-0.049	-0.059	-0.027
P-value	(0.291)	(0.231)	(0.582)
F-P p-value	[0.701]	[0.641]	[0.832]
R-squared	0.602	0.604	0.606
Observations	5,508	5,508	5,508
State FE, Time FE	Y	Y	Y
Birth control	N	Y	Y
Maternal control	N	N	Y

Notes: The table presents the DD estimates of the effects of the CA-PFL on fetal mortality rate. The fetal mortality rate is the number of deaths during pregnancy per 1,000 live births. See notes to table 2 for other details.

Table 7 Effects of CA-PFL on fertility and birth outcomes

	(1)	(2)	(3)
Panel A General fertility rate			
CA*Post	-0.632	-0.553	0.047
P-value	(0.017)	(0.035)	(0.831)
F-P p-value	[0.310]	[0.308]	[0.923]
R-squared	0.977	0.977	0.982
Panel B Log of N(birth)			
CA*Post	-0.002	0.002	0.002
P-value	(0.844)	(0.875)	(0.785)
F-P p-value	[0.959]	[0.968]	[0.891]
R-squared	0.999	0.999	0.999
Panel C Low birth weight			
CA*Post	-0.001	-0.001	-0.001
P-value	(0.004)	(0.008)	(0.001)
F-P p-value	[0.371]	[0.403]	[0.403]
R-squared	0.910	0.911	0.913
Panel D Preterm birth			
CA*Post	0.000	0.001	0.001
P-value	(0.608)	(0.399)	(0.309)
F-P p-value	[0.910]	[0.805]	[0.735]
R-squared	0.909	0.909	0.911
Observations	5,508	5,508	5,508
State FE, Time FE	Y	Y	Y
Birth control	N	Y	Y
Maternal control	N	N	Y

Notes: Notes: The table presents the DD estimates of the effects of the CA-PFL on fertility and birth outcomes. The general fertility rate is the number of live births per 1,000 females of childbearing age between the ages of 15-44 years. Low birth weight is defined as a weight of fewer than 2,500 grams, and preterm is defined as babies born alive before 37 weeks of pregnancy are completed. See notes to table 2 for other details, except for the birth controls exclude birth weight and gestational age.

Table 8 Results of staggered DD

	(1)	(2)	(3)
Panel A Balanced panel (1999-2017)			
Estimate	-0.098	-0.193	-0.090
P-value	(0.000)	(0.000)	(0.429)
Panel B Balanced pre- and post-period			
Estimate	-0.120	-0.165	-0.101
P-value	(0.000)	(0.000)	(0.522)
State, time FE	Y	Y	Y
Birth control	N	Y	Y
Maternal control	N	N	Y

Note: This table presents the results of staggered DD on the effect of PFL of CA, NJ, and RI on PNMR. Panel A shows the results using the balanced panel from 1999 to 2017, and Panel B shows the results using the balanced pre- and post-period, 4 years before and after.

Table 9 Synthetic Control Results

State	CA		NJ		RI		Overall	
Post-Period	Coef.	P-value	Coef.	P-value	Coef.	P-value	Coef.	P-value
1	-0.008	0.340	-0.149	0.720	-0.502	0.700	-0.145	0.298
2	-0.201	0.020	-0.075	0.920	-0.610	0.700	-0.203	0.005
3	-0.336	0.040	-0.145	0.760	-0.020	0.960	-0.106	0.023
4	-0.256	0.000	-0.199	0.800	0.184	0.840	-0.074	0.017
5	-0.121	0.040	-0.096	0.820	NA	NA	NA	NA
Average	-0.184		-0.133		-0.237		-0.132	

Note: This table presents the inference for the synthetic control results. The inference was conducted using Galiani and Quistorff (2017) proposed a synth runner package by comparing the estimated main effect with the distribution of placebo effects (estimations for the same treatment period but on all the control units).

CHAPTER 2 THE COMBINED EFFECTS OF CHARTER SCHOOLS ON STUDENT OUTCOMES: A NATIONAL ANALYSIS OF SCHOOL DISTRICTS

1. Introduction

Charter schools have become one of the most hotly contested topics in American education since they first started three decades ago. These publicly funded schools are exempt from many of the state laws and regulations that govern traditional public schools (TPSs). For example, they operate with more autonomy in hiring teachers and in choosing curricula and instructional methods, allowing charter schools to differentiate themselves from other schools. Families can also choose charter schools regardless of whether they live within the neighborhood attendance zones that apply to traditional public schools.

Proponents argue that this more market-driven approach could increase innovation, create a better fit between schooling options and student needs, reduce the inefficiencies of bureaucracy, and increase competition among schools. Theoretically, this improves outcomes for all students, including families who do not actively choose (Goldhaber and Eide 2003), so that a “rising tide lifts all boats” (Hoxby 2003). Others, however, argue that charter schools engage in strategic behavior by selecting motivated, high-performing students (Bergman and McFarlin 2020) and focusing on superficial improvements, such as marketing, rather than improving actual school efficiency (Lubienski 2007, Loeb, Valant, and Kasman 2011, Jabbar 2015, Harris 2020). Charter schools might also divert funds in ways that make it more difficult for TPSs to succeed, due to economies of scale (Ni 2009, Ladd and Singleton 2020).

Empirical research has examined parts of these theories, but not their combined effects on whole school systems. We provide arguably the first evidence on this topic, especially on a national level, using two decades of data from the National Longitudinal

School Database (NLSD), which includes nearly all districts in the U.S. from school years 1995 to 2016. Using a matched difference-in-differences (DD) and fixed-effects (FE) identification strategies, we study the effects of charter entry on geographic-district-level high school graduation rates and test scores.

A second contribution of our study is understanding the roles played by the three main mechanisms of these systemwide influences: competitive effects, participant effects, and closure effects. Studies of the competitive effects of charter schools on student outcomes in TPSs generally find small positive effects on student achievement, but these effects vary across contexts, research methods, and measures of competition (Hoxby 2003, Bettinger 2005, Bifulco and Ladd 2006, Sass 2006, Ni 2009, Zimmer and Buddin 2009, Linick 2014, Cordes 2018, Ridley and Terrier 2018, Griffith 2019), and there are a few examples where charter competition reduced student outcomes (Ni 2009, Imberman 2011, Han and Keefe 2020).[1] Overall, the existing literature suggests that charter competition effects on TPSs are small.[2]

Participant effects represent the second key branch of literature. Even aside from their competitive effects on TPS, charter schools might serve their own students more or less effectively than other schools. The empirical strategies in this growing literature include matching (Furgeson et al. 2012, CREDO 2013), fixed effects (Brewer et al. 2003,

---

[1] Results from Arizona (Hoxby 2003), Florida (Sass 2006), and Texas (Booker et al. 2008) suggest that there are positive competitive effects of charters on TPSs, and results from California (Zimmer and Buddin 2009) and North Carolina (Bifulco and Ladd 2006) suggest no effects. However, results from Michigan are mixed, with positive effects (Hoxby 2003), zero effects (Bettinger 2005), and negative effects (Ni 2009). Han and Keefe (2020) also use nationwide SEDA district-level data and district-fixed-effects models to study the effects of charter competition, but their study is more descriptive in nature. Their analysis does not weight by enrollment, which has the effect of counting outcomes in small districts too heavily and missing important effect heterogeneity. Also, their study does not test for or address endogeneity. Finally, they frame their analysis as a study of competitive effects only.

[2] A related literature exists on the competitive effects on TPS from private school vouchers (e.g. Figlio and Hart (2014).

Bifulco and Ladd 2006, Sass 2006, Booker et al. 2007, Hanushek et al. 2007), and lotteries (Hoxby and Rockoff 2004, Abdulkadiroğlu et al. 2011, Curto and Fryer 2014, Angrist et al. 2016). As with the competitive effect literature, the results vary, but the average is small, positive, and improving over time. The one other national study of participant effects, in particular, suggests that the average charter participant effect is small and positive (CREDO 2013).

A third mechanism through which charter schools might affect student outcomes is through the closure of TPS. While this can be viewed as an extension of competitive effects, the competition literature focuses only on the effects on TPS that remain open. If charter schools induce low-performing schools to exit the market, then this might improve student outcomes by forcing students into higher-performing schools.[3] Previous studies have examined the effects of school closure on student performance (Bross, Harris, and Liu, 2016) and the effects of charter schools on private school enrollment (Chakrabarti and Roy 2016). However, to our knowledge, no prior research has assessed the causal effect of charter entry on TPS closure.

We start by analyzing the combined effects of all these mechanisms on whole school systems. To accomplish this, our dependent variables are a weighted average of TPS and the charter schools located in the same geographic district. Using DD and FE analysis, combined with propensity score matching (PSM) or propensity score weighting (PSW), we find that charter entry increases district-level student outcomes. For example, districts with more than 10 percent charter market share increased high school graduation rate by 2-4 percentage points and increase test scores by 0.06-0.16 standard deviations. Applying

[3] We also note that, if charter schools induce low-performing TPS to close, then this might bias "participant effects" downward as charter schools get compared with increasingly effective TPS.

the approach proposed by Goodman-Bacon (2021), we find that the effects are dominated by the treated versus never-treated groups. We also employ Sun and Abraham's (2021) method in our event study model to address potential heterogeneity in effects across cohorts and time (De Chaisemartin and d'Haultfoeuille 2020, Callaway and Sant'Anna 2021, Sun and Abraham 2021).

While this average effect is of considerable policy interest, there are also good reasons to explore heterogeneity by market share. Some states, for example, have charter caps that limit charter entry to a small share of statewide enrollment, but there is little research on whether there might be an optimal charter school market share that would justify these state-level or district-level caps. When we vary the charter threshold from 1-20 percent, we find that the effects become evident at about five percent market share and plateau at 15 percent market share. While this implies that some sort of charter cap might be efficient, these results suggest different kinds of caps than those employed by some states.

We also examine heterogeneous effects by location, grade levels, and baseline achievement levels. Prior literature suggests that charter schools have more positive effects in urban areas (CREDO 2013, Chabrier, Cohodes, and Oreopoulos 2016), a finding that we confirm. Harris (2020) hypothesizes that this might be because urban districts have lower baseline achievement, which makes improvement somewhat easier to achieve. We find some suggestive evidence of this, though the estimates are imprecise. Additionally, we find that the effects are concentrated in middle schools and high schools, not elementary schools.

Our analysis of the mechanisms behind these overall system effects suggests that participant, competitive, and closure effects likely all play a role. In particular, we find that when charter market share reaches 10 percent, the TPS school closure rate increases by 40 percent. Finally, we find little evidence that charter entry reduces private school enrollment.[4] This finding for private schools is useful both for ruling out selection bias from charter schools induced students to move in or out of private schools, and for understanding possible effects of charter schools on private schools, which could also affect educational outcomes and social welfare.

The main threat to identification is that charter school location and timing are endogenous. There is evidence about where charter schools tend to locate based on time-invariant, observable factors (Bettinger 2005, Glomm, Harris, and Lo 2005, Bifulco and Buerger 2015), which the DD and FE methods are effective in addressing. However, these methods, by themselves, cannot rule out time-varying unobserved factors. We address this by leveraging the timing of state laws in two ways: (a) we restrict the comparison group in the DD to districts in states that had no charter laws in place during the entire panel; and (b) we compare districts in the early-adopting state to those in late adopting states. The assumption in these analyses is that the existence and/or timing of state laws is independent of the time-varying changes arising specifically in the districts that are likely to be treated. While we conclude that this assumption is reasonably plausible, we also use placebo analyses as a third method for addressing endogenous charter entry and location. In one case, we test for effects of charter market share in

[4] One might argue that this is inconsistent with the evidence that voucher programs create competition that improves TPS (Figlio and Hart 2014); however, those situations are different because voucher programs make private schools accessible to students who typically attend TPS. Here, we are focusing on the effects of charter schools on private school that are financially inaccessible.

particular grades on the outcomes in grades where effects are implausible and, in the other version, we test for effects of subsequent charter entry on pre-treatment outcomes. Fourth, and finally, we follow Oster (2019) and use selection on observables to evaluate the extent of potential of selection on unobservables. These analyses generally reinforce the conclusion that charter schools improve test scores and/or high school graduation.

The paper proceeds as follows. Section 2 introduces the data. Sections 3 discuss the methods. Average treatment effects are discussed in section 4 and heterogeneous analyses in Section 5. Section 6 provides some discussion and Section 7 concludes.

2. Data and Descriptive Statistics

This paper uses data from the National Longitudinal School Database (NLSD), which contains some data for a near-census of all TPSs, charter schools, and private schools in the U.S. from 1991 to the most recent academic year.[5] The NLSD merges school and district level data from Common Core of Data (CCD), Stanford Education Data Archive (SEDA), Census Small Area Income and Poverty Estimates (SAIPE), and other sources. We are most interested in the student outcomes—test scores and graduation rates, and we include a series of districts covariates into analysis, such as enrollment, school type, district finance, school-age population and poverty rate. The addresses of charter school also allow us to place the location of each charter school within its geographic school district.[6] These districts define the scope of the market, i.e., which schools are assumed to compete with one another.

[5] All the school years mentioned in this paper are spring school years unless specifically stated otherwise.
[6] We note that some charter schools located within a given geographic district are authorized or regulated by a government agency or delegate located outside the district, such as a state government or a university. Also, some states have elementary and high school districts that overlap one another; in these cases, we assign the charter school to the district whose schools serve the same grade levels as the charter school.

The specific dependent variables are the average freshmen graduation rate (AFGR)[7] (hereafter, graduation rate) and student Math and English Language Arts (ELA) test scores. The test scores are available in 3rd through 8th grade in Math and ELA over the 2009-2016 school years from the Stanford Education Data Archive (SEDA). The SEDA provides nationally comparable, publicly available test score data for U.S. public school districts[8] (Ho 2020). For interpretation purposes, we normalize the test scores to have means of zero and standard deviations of unity within the grade, year, and subject.

The graduation rate sample includes schools covering grades 9-12 annually from 1995 to 2010. These data include both TPS and charter schools, which constitute the policy treatment in this study. We use the charter enrollment share, or the percentage of public-school students enrolled in charter schools to define district treatment status. These market share measures are created separately by grade level under the theory that TPSs compete with charters when there is a threat their students will leave for another school.

For both outcomes, we then calculate the geographic district outcome by averaging TPS and charter school outcomes (weighted by enrollment share). This is crucial to the analysis as it allows us to capture charter effects operating through all of the various market mechanisms. In contrast, prior studies of charter schools have focused on the

---

[7] The AFGR uses aggregate student enrollment data to estimate the size of an incoming freshman class and aggregate counts of the number of diplomas awarded four years later. For example, the AFGR for a school year in 2006 is the total number of diploma recipients in 2006 is divided by the average enrollment of the 8th-grade class in 2002, the 9th-grade class in 2003, and the 10th-grade class in 2004.

[8] To make estimates are comparable across states, grades, and years. The SEDA research team took the following steps: (1) estimate the location of each state's proficiency "thresholds"; (2) place the proficiency thresholds on the same scale using the National Assessment of Educational Progress (NAEP); (3) estimate the mean test scores in each school, district, county, metropolitan statistical area, commuting zone, and state from the raw data and the threshold estimates, and (4) create estimates of average scores, learning rates, and trends in average scale scores. See details in SEDA website https://edopportunity.org/methods/.

outcomes of individual schools in the examination of school participant effects and the competitive effects of individual charter schools on nearby. Here, we are interested in the net effect of all the mechanisms, which means we need to account for the outcomes of all traditional and charter schools. Aggregating these to the geographic district level also allows us to sidestep the main problem with prior studies, i.e., selection bias from students moving across schools (see further discussion later).

One potential threat to identification in our analysis is that charter entry coincides with changes in the population of students attending publicly funded schools (charter or TPS). To test this, we also sometimes include the following (time-varying) covariates: total enrollment (log form); the share of students who are Hispanic, black, white; the share of students who are in special education programs; the share of students eligible for free or reduced-price lunches (FRL); student-teacher ratio; average teacher salary; number of magnet schools; number of schools; total revenue per student; total expenditure per student; and the poverty rate of the school-age population. Since these measures are potentially endogenous, we view results with these controls as an endogeneity test, not as the preferred effect estimates.

Table 1 presents the summary statistics for the outcome and control variables.[9] Compared with TPSs, charter schools nationwide tend to enroll a larger proportion of African American and Hispanic students. Charter schools also more likely to be in urban districts and where achievement is relatively low. These observable differences are minimized after matching (see later discussion).

---

[9] Table 1 is weighted by enrollment size. Table A1 in the Appendix presents the unweighted summary statistics for the outcome and control variables of our samples.

The test score (graduation) data are available for 10,439 (9,278) districts, including 607 (416) districts that have at least one charter school. The final samples for the test scores (graduation) analyses include 83 (70) percent of the nation's publicly funded schools and 85 (78) percent of those schools' enrollment.[10] We also have complete data for 30 (31) percent for charter schools in the analysis. Appendix B tests for observable differences between included and missing districts. The student demographics between the two are very similar; however, unsurprisingly, the enrollment sizes of the missing districts are much smaller than those in the included districts. Also, the missing charter schools are more likely missing if they were authorized by an entity other than the local district (e.g., a state agency or university).[11] Prior research suggests that the performance of charter schools is not related to authorizer type (Carlson, Lavery, and Witte 2012), so we do not expect this to influence our main conclusions.[12] We discuss potential implications in the discussion of results.

The timing of the passage of charter laws also figures prominently in our analysis. Charter schools have seen dramatic growth over the last three decades. The first law allowing the establishment of public charter schools was passed in Minnesota in 1991. As of fall 2020, charter school legislation had been passed in 45 states and the District of

[10] Some districts have data missing for all years in the panel, while others are missing for so many years that imputation is a questionable strategy. For test score (graduation) data, our main estimates remove 16 (9) percent of districts with charters where we have some data, but too few time periods. We also ran a version of results that includes all available observation-years and these estimates are similar to those reported when these districts are dropped (results are available upon request).

[11] The missing data is most likely because charter schools authorized by non-local-district entities do not report data through the local districts; they function as their own Local Education Agencies (LEA) for governing and reporting purposes.

[12] Zimmer et al. (2014) find that Ohio charter schools authorized by non-profit organizations are typically more effective, but it is extremely rare, nationally, for non-profit organizations to be charter authorizers.

Columbia.[13] Table 2 presents the years of charter school legislation as of 2020. As a result of the growing number of states allowing charter schools, the percentage of publicly funded schools that are charter schools increased from 0 to 7.3 percent, and their percentage of enrollment increased from 0 to 6.2 percent.[14]

The analysis that follows focuses on the effects of charter entry on whole districts. Table A2 in the Appendix presents the Top 20 districts with the largest charter enrollment share among districts with at least 10,000 total students in the 2018 spring year. New Orleans tops this list, followed by Gary (Indiana), Queen Creek (Arizona), Washington DC, and Detroit. The question of interest in this study is whether charter school entry in these (and other) districts has improved or reduced student outcomes years after they have opened and, if so, through what mechanisms.

We begin by simply plotting the descriptive trends in graduation rate and Math, and ELA scores throughout the sample period for treated districts and comparison districts in Figure 1. For purposes here, we define a treated district as one that is ever at or above 10 percent market share and the comparison group is districts that never have charter schools. The comparison group outcomes are consistently below the treatment group, reflecting that charter schools tend to locate in low-performing districts. Figure 1 also plots the gradual increase in the overall charter market share to show the intensity of treatment over time. For all three outcomes, the comparison and treatment groups display parallel trends in the early years and generally show convergence by the last period. This

[13] The states in which public charter school legislation had not been passed by that time were Montana, Nebraska, North Dakota, South Dakota, and Vermont.
[14] Figure A1 in the Appendix presents the trends in charter school share and charter enrollment share from spring 1991 to spring 2018.

provides some visual evidence of a system effect of charter schools, at least at a descriptive level. In what follows, we describe our identification strategy.

3. Identification Strategy

3.1 Difference-in-Differences

We rely mainly on a DD strategy to evaluate the effect of charter enrollment share on student achievement. The treatment group includes those districts whose charter market share is ever at or above the threshold τ. In our baseline model, we use $\tau = 10$ percent of charter enrollment share.[15] The first comparison group includes districts (in all states) that did not have charter schools during the panel (i.e., never treated). Districts with charter shares that are above zero but below the threshold are omitted from the analysis to create a clear treatment contrast.

We use "ever above" the threshold because TPSs and other local education-related organizations (e.g., university schools of education) are likely to be aware in advance that their areas are going to have large charter market shares and they may therefore start reacting before districts reach the τ threshold. Since τ is inherently arbitrary, we estimate the models assuming a wide variety of threshold levels. In addition to providing a robustness check on the threshold choice, this provides initial evidence regarding non-linear market share effects.

We estimate the DD effects using equations (1) and (2):

$$GR_{it} = \alpha + \beta(T_i \cdot Post_{it}) + X_{it}\gamma + \mu_i + \lambda_t + \varepsilon_{it} \quad (1)$$

$$Test_{ijt} = \alpha + \beta(T_i \cdot Post_{it}) + X_{it}\gamma + \mu_i + \lambda_t + \omega_j + \varepsilon_{ijt} \quad (2)$$

[15] The charter market share here is for the sector of public schools and does not include private schools.

In equation (1), the dependent variable is the high school graduation rate (GR_{it}) for the high schools in district *i* in year t. T_i is an indicator variable equal to unity if the district *i* charter enrollment share is ever above the threshold and equal to 0 if district i has no charter schools during the sample period; $Post_{it}$ is an indicator set to unity starting in the period that district i had its first charter entrant and continuing thereafter. Finally, μ_i is district fixed effects; λ_t is year fixed effects; X_{it} is a vector of district characteristics, and $\varepsilon_{\mathrm{it}}$ is the error term. The coefficient of interest is β, which measures the charter effects on graduation rate.

In equation (2), Math and English Language Arts (ELA) test scores are the dependent variables, where $Test_{ijt}$ is the test score in district *i* and grade *j* (specifically, grades 3-8) during the year *t*, ω_j is a vector of grade fixed effects, and other terms are defined same as equation (1). In other words, (2) is the same as equation (1) except for the dependent variable and the fact that the outcome is grade specific. Standard errors are clustered at the district level, and the estimates are weighted by high school enrollment for graduation rate and grade-level enrollment for Math and ELA. (We report results clustered at the state level in Table A3 in the appendix, but this has minimal influence on inference.)

In some specifications, we combine our DD design with matching methods to minimize the differences between treated districts and comparison districts (i.e., to achieve common support). For example, states may intervene in districts that are persistently low-performing and/or otherwise similar. In this case, matching on observables reduces comparison-treatment differences in unobserved policy interventions that may also shape student outcomes. Matching therefore may increase the probability that the identifying assumptions hold for the DD analysis.

We use district covariates to match treated to comparison districts (nearest neighbor).[16] This yields the Propensity Score Matching (DD-PSM) sample. We also use the propensity score as the inverse probability weight for a Propensity Score Weighting (DD-PSW) analysis. These approaches minimize the difference between the control and treated groups, as reflected in density plots before and after matching (Appendix Figure A2).

In the above two-way fixed effects DD specification, the indicator for charter entry switches from 0 to 1 at varying times for different treated districts. Goodman-Bacon (2021) notes that the overall DD estimate is a weighted combination of each possible pairwise DD estimate comparing treated districts with: (a) never-treated districts, (b) districts that are not yet treated but are treated later, and (c) districts previously treated. For this reason, we apply the Goodman-Bacon (2021) decomposition to determine the underlying source of identification.

Event study analyses also partially address this concern and provide evidence about parallel trends, intercept shifts and dynamic effects. However, with staggered treatment start dates, the coefficient on a given lead or lag can be contaminated by effects from other periods due to effect heterogeneity (De Chaisemartin and d'Haultfoeuille 2020, Callaway and Sant'Anna 2021, Sun and Abraham 2021). To address this concern, we implement Sun and Abraham's (2021) interaction-weighted (IW) estimator for dynamic

[16] We specifically match on the first year values of the variables listed in Table 1. Table A4 in Appendix presents the probit regression results of the matching model.

treatment effects.[17] Equations (3) and (4) are specifications of event study for graduation rate and test scores.

$$\mathrm{GR}_{\mathrm{it}} = \alpha + \sum_{r=-m}^{q} \beta_r \big(\mathrm{T_i} \cdot \mathrm{d_{i,r}}\big) + \mathrm{X_{it}}\gamma + \mu_i + \lambda_t + \varepsilon_{it} \tag{3}$$

$$\mathrm{Test}_{\mathrm{ijt}} = \alpha + \sum_{r=-m}^{q} \beta_r \big(\mathrm{T_i} \cdot \mathrm{d_{i,r}}\big) + \mathrm{X_{it}}\gamma + \mu_i + \lambda_t + \omega_j + \varepsilon_{ijt} \tag{4}$$

where $d_{i,r}$ is a dummy of the *r* years of leads or lags since district *i* initiated the first charter school[18]. The vector β_r represents measures of cohort-specific effects compared with the comparison group. We drop always-treated districts in the event study, as suggested in previous studies (Sun and Abraham 2021, Meinhofer et al. 2021). Five or more years leads/lags for graduation rate sample (three years or more for test score sample) are assigned into a single indicator given the length of sample period.

3.2 Threats to Identification

The main assumption in DD analysis is that the comparison group would have followed the same trend as the treatment group if the latter had not been treated. The standard initial test is whether the two groups follow parallel trends prior to treatment. We test this as follows:

$$\mathrm{GR}_{\mathrm{it}} = \alpha + \gamma(\mathrm{T_i} \cdot \mathrm{Time_t}) + \mathrm{X_{it}}\gamma + \mu_i + \lambda_t + \varepsilon_{it} \tag{5}$$

$$\mathrm{Test}_{\mathrm{ijt}} = \alpha + \gamma(\mathrm{T_i} \cdot \mathrm{Time_t}) + \mathrm{X_{it}}\gamma + \mu_i + \lambda_t + \omega_j + \varepsilon_{ijt} \tag{6}$$

where $Time_t$ is a linear calendar time trend (pre-treatment years only) and other terms are defined as before. The coefficient of interest is γ. We present our estimate of γ in

[17] The IW estimator is implemented in three steps. First, estimate the interacted regression, where the interactions are between relative time indicators and cohort indicators. Second, estimate the cohort shares underlying each relative time period. Third, take the weighted average of estimates from the first step, with weights set to the estimated cohort shares (Sun and Abraham 2021). We also considered other methods, especially the Callaway and Sant'Anna (2021) and Chaisemartin and D'Haultfoeuille (2020) methods, but these are focused on recovering a single treatment effect and we are interested in dynamic treatment effects.
[18] Table A5 in Appendix presents the number of districts by the year the first charter was initiated.

every table with DD specifications. While our results usually pass this test, we note exceptions and interpret these results with caution.

The pre-trends test is insufficient to establish parallel trends. Specifically, we still have to assume: (1) that charter location was conditionally exogenous (i.e., that treatment was not assigned based on unobserved factors that are correlated with student outcome trends); and (2) and that there were no other idiosyncratic shocks that happened to coincide with treatment timing.[19] Put differently, the DD analysis described above addresses selection on observable time-invariant unobservable factors, but it does not address time-varying unobservable factors.

It is worth considering specific scenarios under which either of these assumptions might be violated. First, especially given the investment required to open a charter school, it may be that charter schools locate in districts where decreases in TPS student outcomes are expected in the future, yielding a downward bias in effect estimates. Charter schools are more likely to locate in districts with low contemporaneous student outcomes (Glomm, Harris, and Lo 2005), signaling that we might expect charter schools to also locate where expected future performance in TPS is lower, in ways that are unobservable in our analysis.

Second, charter schools might open where districts experience idiosyncratic shocks, in which case future outcomes regress toward the mean. If charter schools prefer to locate near low-performing school, as noted above, then this would yield an upward bias (i.e., charter schools enter because of the negative shock in existing schools, but then bounce

[19] Since we are interested here in long run effects, it is worth noting that idiosyncratic shocks occurring after the start of treatment, and disproportionately affecting either one group, could also create bias if the shocks are correlated with treatment status. The longer the analysis goes into the future, the more likely this is to occur.

back with positive shocks in the next period). We do not see this second scenario as very likely because it takes several years to create an organization that can put together a charter application, submit the application and gain approval, hire personnel, purchase necessary capital, and recruit students. Also, with such a long-term investment, charter organizations are likely to consider multiple years of information in making their entry decisions, which reduces the chance of regression to the mean.

Third, local or state policy changes beyond charter schools might coincide with the timing of charter entry and disproportionately affect the outcomes of either the treated or untreated districts. For example, since school districts are often charter authorizers, it may be that a change in school board politics leads to a variety of simultaneous policy changes in a district. Since charter schooling is only one of these policies, we might falsely attribute outcome changes to charter schools that were actually caused by other policies. If the other policies had positive effects, then this would yield an upward bias, though the bias could also work in the other direction.

To account for these possibilities, we estimate four variations of the DD. First, we restrict the comparison group to districts located in states that do not have charter laws up to the year 2016 and whose districts always, therefore, have charter shares of zero.[20] This requires the assumption that states do not pass and implement laws at the same that unobserved changes are occurring in the specific districts where charter schools end up locating. This is plausible in the sense that state lawmakers do not know ex ante where charter schools will open within their states. They might expect them to locate in “failing districts” or urban districts, but these factors are observable and accounted in the analysis.

[20] These states are Kentucky, West Virginia, Montana, Nebraska, North Dakota, South Dakota, and Vermont. We chose the year 2016 because this is the last year of our data.

Still, we cannot rule out that this assumption is violated. Therefore, using similar logic, we also estimate versions of equations (1) and (2) that compare districts in early and late adopting states.[21] In this case, even if there are unobserved factors associated with whether states eventually adopt these policies, the assumption is still satisfied so long as the timing of the charter law passage is somewhat arbitrary. This is also possible given that it usually takes multiple legislative attempts to pass a charter law and that passage depends on an alignment of political conditions whose timing can be arbitrary. In that case, we can expect the treated districts (in the early-adopting states) to be similar to the comparison districts (in late-adopting states) on even time-varying unobserved factors.

Third, we carry out two placebo analyses. Our main estimates are based on the aligned charter market share, across all grades. But it is implausible that some charter schools would affect certain outcomes. In particular, the entry of charter high schools cannot directly influence test scores in grades 3-8; likewise, the entry of elementary charter schools cannot influence charter high school graduation until many years later, after students have passed through middle and high school grades. The placebo test therefore involves restricting the treatment variable to charter schools that are mis-aligned with the grades of the dependent variables. In another placebo test, we analyze the "effect" of subsequent high charter market share on student outcomes in periods prior to the entry of charters. Since later entry cannot impact current student outcomes, any apparent effect would be another measure of potential bias.

[21] The logic here is similar to the Goodman-Bacon method. The early adopter states are those passed state charter law before 1997 (27 states), and the late adopter states are those passed state charter law during 1997 to 2016 (17 states). See Table 2 for the list of states.

Fourth, treatment effect estimates can be bounded by adding available covariates (Oster, 2019). If the effect size changes relatively little compared with the increase in the variation explained, it is unlikely that the bias caused by unobservables is large. Others have proposed similar methods (Altonji, Elder, and Taber 2005, Oster 2019, Gibbons, Overman, and Sarvimäki 2021, Gibbons, Heblich, and Timmins 2021).

3.3 Fixed Effects Model

The advantages of the DD approach include providing visual and formal tests for parallel trends, facilitating the estimation of dynamic effects, and testing for non-linear effects based on market share. A disadvantage of the DD is that the treatment gradually increases in dosage (i.e., districts do not instantly shift to the highest market share we observe), yet we must choose s specific year for the treatment to start. To address this, we also employ a FE model using only those districts that have at least one charter school at some point during the panel. In this case, the charter market share is used as a continuous measure of charter enrollment share. We estimate the FE effects based on the following equations:

$$\mathrm{GR}_{\mathrm{it}} = \alpha + \beta \mathrm{Charter}_{\mathrm{it}} + \mathrm{X}_{\mathrm{it}}\gamma + \mu_{\mathrm{i}} + \lambda_{\mathrm{t}} + \varepsilon_{\mathrm{it}} \qquad (7)$$

$$\mathrm{Test}_{\mathrm{ijt}} = \alpha + \beta \mathrm{Charter}_{\mathrm{ijt}} + \mathrm{X}_{\mathrm{it}}\gamma + \mu_{\mathrm{i}} + \lambda_{\mathrm{t}} + \omega_{\mathrm{j}} + \varepsilon_{\mathrm{ijt}} \qquad (8)$$

where the $Charter_{ijt}$ is the continuous measure of charter enrollment share of district *i* year *t*. In our baseline analyses, we estimate the FE model using the current year charter enrollment share. In one robustness check, we estimate the FE model for the graduation rate using the average charter enrollment share in the last four years (because it takes four years for a high school student to graduate and student outcomes may be affected by charter share through the four years).[22] For test scores, we estimate the FE model using the prior year's charter enrollment share (same cohort), as this may better reflect the timing of the TPS responses.[23]

The identifying assumptions are similar between the DD and the FE, but not identical and these differences in assumptions provide additional tests for identification. Both the DD and FE account for fixed unobserved factors that are correlated with treatment status and with student outcomes. However, with the DD, we also have ways to account for time-varying unobserved factors, so long as these are similar in the comparison group. For this reason, the DD is our preferred approach, but if the DD and FE results differ, this would suggest that time-varying unobserved factors are an issue. Also, the FE assumes that each unit increase in charter market share has an immediate effect on outcomes whereas the DD assumes the timing coincides with the first entry. Again, the comparison of the DD and FE provides evidence about whether the results are sensitive to that decision.

4. Results

4.1 Difference-in-Differences and Event Studies

We find consistent evidence that charter entry improves geographic-district-level outcomes. Figure 2 shows event study analyses for the graduation rate, Math, and ELA, using both the traditional and Sun and Abraham (2021) methods, with the latter being our most preferred estimates. The results also suggest an immediate small positive effect on

[22] For example, we use the average charter enrollment share of high school (grades 9-12) in 2006-2010 to estimate the effects on graduation rate in 2010 .
[23] For example, we use the charter enrollment share of Grade 7 in 2010 to estimate the test score of Grade 8 in 2011.

graduation, which roughly doubles one year later and then plateaus. One possible reason for the immediate effect is that TPSs, which still comprise the vast majority of schools when the first charter school enters, anticipate charter entry and work on improving prior to, or during, the first treatment year.[24] Also, if the effect operates through the participant effect, it could be that students who were going to drop out of their TPS instead moved to a charter school that offered a different opportunity immediately. While it is difficult to determine the reason, we are not the first to find this result. Harris and Larsen (forthcoming) also find an immediate effect on high school graduation rates than test scores in New Orleans. The graduation estimates are also robust with the Sun and Abraham (2021) method, and to other robustness checks discussed below. The results are also positive with ELA scores, though these taper off in the Sun and Abraham (2021) specification in the third year post-treatment. With Math, we find positive effects, although we do not pass the pre-trends tests.

Table 3 therefore presents the DD estimates of the effect of charter treatment (i.e., having a charter enrollment share ever above 10 percent) on graduation rate, Math, and ELA test scores, which averages the pre- and post-treatment periods. Columns (1) and (2) are estimates of the DD model with districts from all states; Columns (3) and (4) are estimates of the DD-PSW model; Columns (5) and (6) are estimates of the DD-PSM model.

[24] Measurement error in the graduation rate is also a possibility. Recall that the AFGR is the number of graduates divided by the lagged number 8th, 9th, and 10th graders. This means, if there is a trend of in-migration, then the number of graduates will increase mechanically and faster than the (lagged) denominator. Thus, this could be an artifact of changing demographics. However, when we use the enrollment level as the dependent variable in the DD analysis, we find no evidence of enrollment change.

Here, we find positive and precisely estimated effects for all three outcomes. The magnitudes of the coefficients imply that having a large share of charter enrollment increases the graduation rate by 3-4 percentage points and improves Math and ELA scores by 0.08-0.16 standard deviations.[25] These estimates also generally pass parallel trends tests, presented at the bottom of Table 3, except in Math. In ELA, two of six the tests reject the null, but these two test coefficients are small relative to effect estimates.[26]

To better understand the source of variance underlying these estimates, we follow Goodman-Bacon (2021) in decomposing the DD model and showing the weights attached to each comparison. Table 4 shows that the majority of the weight (60-70 percent) comes from the treated versus never-treated comparison as well as the treated versus never-treated effects; moreover, for these highly weighted comparisons, the effect estimates are positive for all three outcomes.[27]

One advantage of the DD method is that it allows us to estimate non-linear effects. We therefore also estimated the model allowing a wide range of thresholds, from 1-20 percent charter share. Figure 3 plots the point estimates along with the associated thresholds. When raising the threshold, note that a larger number of districts have $0 < \mathrm{Charter} < \tau$, such that they are dropped entirely from the estimation, while at τ =1 percent, almost no districts are dropped.

---

[25] Note that the sample size drops dramatically in the DD-PSM because we match each treatment district to only one control district.
[26] We conduct additional analysis in appendix Table A6 which add state linear trends. The graduation rate results are very similar. The point estimates remain positive and sometimes statistically significant, but they are smaller and less precise. It could be that neighboring districts are responding to the threat of competition from potential charter schools. In that case we are "over-controlling" with the state trends; that is, if we are interested in the total effect, then the state trends absorb part of the charter effect--that part coming from neighboring districts that don't contribute to the estimation of in the other specifications. The same general pattern emerges when we switch from state trends to state-by-year fixed effects.
[27] See Figure A3 in the Appendix for the plots of Goodman-Bacon decomposition.

The expected pattern of results across threshold levels is unclear as this depends on: the source of the effects (participant versus competitive effects) at each threshold; whether the marginal charter school is as effective as the prior ones; how quickly charter market share reaches equilibrium; and whether there is effect heterogeneity that correlates with the long-run district charter share.[28]

As Figure 3 indicates, there is a fairly consistent rise in effect estimates on graduation rate and test scores as the threshold rises, though noticeable effects begin emerging at 5 percent charter market share and continue to about 15 percent market share, after which they plateau. It could be that, beyond a certain point, either the marginal charter entrant is less efficient than earlier entrants or that TPS efficiency wanes when they have to cope with large reductions in resources due to a loss of enrollment to charter schools. This pattern is consistent across all the outcomes and DD strategies. It also does not appear to be due to missing data.[29] We provide additional robustness checks below in the FE analysis, which further reinforce the presence of non-linear effects by market share.

One of the treated districts with $\tau > 20$ is New Orleans with the largest charter enrollment share among districts with at least 10,000 total students. It is also arguably the most successful charter reform effort documented to date (Harris and Larsen forthcoming). To test whether the $\tau > 20$ is driven by New Orleans, we drop this district and re-run the analysis. The results are very similar to those in Figure 3, suggesting that

[28] Also, note that we mark the start of treatment when the first charter school enters. In larger districts, more enter later and therefore probably have later effects that are not captured in these estimates.
[29] As Table A2 in the Appendix shows, the districts with the largest market shares are all included in the test score analysis. Six of these 20 districts are omitted in the graduation analyses, but we see the same plateauing pattern with both outcomes.

the pattern of results by threshold level is not driven by New Orleans (see Appendix Figure A4).

To understand the effects of the charter-heavy districts generally, we also estimated effects for the subset of districts with at least one charter school, but not more than 10-percent charter penetration (see Appendix Table A7). Here, we find small but significant positive effect on graduation rate (increase 1-2 percentage points) and positive but imprecise effects on test scores, as expected given the smaller dosage and associated smaller effects in low-charter-share districts shown in Figure 3.[30]

4.2 Tests for Endogeneity

The above results only address endogeneity caused by time-invariant factors and time-varying observables. In this section, we present results that address endogeneity from time-varying unobserved factors. First, we restrict the control sample to districts in states without charter schools as of 2016 to minimize the potential endogeneity of charter reform in our main results. Table 5 presents these results. The results are very similar to our main results (Table 3) discussed above. Second, we restrict the treated sample to early adopter states and the comparison group to late adopter states to minimize potential endogeneity of statewide charter reform in our main results. As shown in Table A9, the results are, again, very similar to Table 3.

---

[30] As one additional robustness check, we also switched from the charter enrollment share to charter school share as the dependent variable. The results are similar to that of charter enrollment share (see Table A8 in the Appendix), except that the magnitude of estimates is somewhat smaller. It is not obvious which should be preferred. The charter share measure may better reflect the number of TPSs that are under pressure (more small charter schools might be spread across a district, competing with more TPS), but, to the degree that school size is similar across places, the enrollment share (i.e., the number of students who have left) may be a better indicator of effective competitive pressure. We also considered other treatment measures that have been used in prior studies, but these studies have focused just on competition; specifically, competition between individual schools. Given that we are interested in net effects of all mechanisms (participant, competition, and closure) on entire districts, the measures of competition used in prior studies do not seem appropriate here.

Our third approach to testing for endogeneity involves two placebo methods. The above analyses are based on the charter share that aligned to grade levels and we now conduct analyses using charter share for grades that are "misaligned" with the outcomes. Again, the charter entry of elementary schools should not affect the graduation rate at least in the short run, before those students reach the age of potential high school graduation. Table 6 presents the placebo results for the "effect" of elementary grade charter market share on the high school graduation rate and the high school grade market share on elementary Math and ELA scores. These estimates are noticeably smaller (and less precise) than the main results above. Only two of 18 specifications suggest positive effects, but even these estimates are no longer precise when controlling district covariates. Most of the estimates for test scores are negative, suggesting that our main results may be biased downward.

In the other placebo test, we analyze the "effect" of subsequent high charter market share on student outcomes in periods prior to the entry of charters. Specifically, we falsely assume that charter entry occurred two years earlier and redo the analyses.[31] This is similar to the parallel trends test, which we generally passed for graduation and ELA, but not Math. We see this same pattern in the placebo; the placebo again fails for Math.[32]

In addition, we follow Oster (2019) and use the sensitivity of the estimated coefficients to the added controls to assess the potential for bias due to unobservables.[33]

[31] We also tried versions with one, three, and four years earlier than charter entry and got similar results.
[32] As Table A10 in the Appendix shows, the "effects" are 0.12-0.19 s.d. for Math and 0.03-0.04 s.d. for ELA.
[33] The Oster Bound analysis compares the estimate that does not include controls with estimate that includes controls, scaled by the change in R^2. The exact formula is $\beta^* = \tilde{\beta} - \tilde{\delta}\frac{(\beta_0-\tilde{\beta})(R_{max}-\tilde{R})}{\tilde{R}-R_0}$. Following Oster (2019), we assume $\tilde{\delta}=1$ and $R_{max}=1.3R_0$ (but no greater than 1), where β_0 and R_0 are, respectively, the point estimate and the R^2 for the baseline model without control variable; and $\tilde{\beta}$ and $\tilde{R}$ are the point estimate and the R^2 of a model with control variables.

We carried out an Oster bounds analysis for DD and present results in Table A11 in the Appendix. The Oster bounds analysis estimates the treatment effect if selection on unobservables is as important as selection on observables (assume $\tilde{\delta}$= 1 in their model). All of our estimates of Oster bounds are positive and slightly larger in magnitude than our baseline estimates. Further, the observable characteristics included in our data explain more than 80 percent of the variation in the change of student outcomes.[34]

In summary, our analysis of graduation rates passes all of these endogeneity tests. The analysis of ELA and Math scores also pass the majority of these tests, but we have noted some exceptions.

4.3 Tests for Enrollment and Compositional Changes

Our analysis focuses on data at the geographic district level, i.e., outcomes are a weighted average of those in the TPS and charter schools located within the same school district boundaries. This has the advantage of avoiding the usual problem caused by sorting of enrollment between charter and traditional public schools within districts. However, it is still possible that the timing and location of charter entry could be correlated with compositional changes in the student population across districts.

To address this issue, we conduct several event study analyses of the effects of charter entry on enrollment and its compositional changes. As Figure 4 shows, there are no statistically significant changes in enrollment, or its composition, after charter entry. The results do suggest some possible increases in the percentage of students who are Black and Hispanic and decreases in the percentage who are eligible for free or reduced-

[34] We also carried out a version of the Oster method where the fixed effects are added in the baseline model and only (time-varying) covariates are added. This naturally yields a smaller change in the R^2 and a much larger point estimate. We have chosen to report the more conservative estimates in the main text.

price lunches. However, the latter do not appear large enough to change the conclusions. The largest effect on FRL suggests a decrease of four percentage points, which we predict would increase test scores by only 0.01 s.d., a small fraction of the main charter entry effect reported above.[35] This reinforces that demographic change cannot explain our results.

4.4 Fixed Effects Results

We use FE estimates as an additional robustness check. This approach leverages within-district longitudinal variation in the charter market share (within the previously described treatment group). The first two columns of Table 7 use the contemporaneous charter market share and implicitly assume an immediate effect of charter entry on student outcomes. With the graduation rate, the latter two columns average the last four years of charter market share (because it takes four years of data to calculate a single graduation rate). With test scores, the last two columns lag the charter market share by one year, in case there is a delayed effect.

The results, again, indicate a significant effect of charter entry on the graduation rate. The results are similar between the contemporaneous charter market share and four-year averages. A 10 percent increase in average charter enrollment share increases the graduation rate by 1-2 percentage points. Also, as in the DD, we generally see positive effects on student test scores with the FE, and all estimates of Math are statistically significant. Our results are robust when using contemporaneous or lagged charter market shares.

[35] The coefficient on FRL in our DD estimates is about -0.3. This implies that decreasing the FRL percentage by four percentage point would increase test scores by only -0.3*-0.04=0.01 s.d.

To compare these FE with the DD results, note that the average eventual charter share using the 10 percent threshold is 18.4 percent.[36] We can therefore take the FE coefficient, which captures the effect of increasing the market share by one percentage point and multiply by 18.4 percent the average eventual max share of the DD, which yields 0.03. This is almost identical to the DD point estimates reported in Tables 3. Given that these various methods rest on different assumptions, the robustness of the results reinforces a causal interpretation.

Earlier, in the DD analyses, we reported that a plateauing of the effects at about 15 percent charter market share. One potential explanation is that the DD “turns on” the treatment for each district when the first charter school opens, but it may take many years for districts to reach their maximum charter shares, which we used to assign district treatment status. Therefore, it could be that the effects for larger market share districts reflect their small actual market shares in the early treatment years. The FE analysis is well-suited to test this theory because we leverage within-district variation in actual charter market share. These results reinforce that the idea of a plateau. Using a simple quadratic, we see that the coefficient on the linear term is still consistently positive and that the coefficient on the squared term is consistently (though not uniformly) negative.[37]

5. Effect Heterogeneity

Prior research has suggested that charter school participant effects vary by urbanicity, baseline achievement level, and grade level. Table 8 presents the charter effects in Metropolitan Areas (MA) and non-MA. The results show that the charter effects mostly

---

[36] Table A12 in Appendix list the average charter share across different models.
[37] The results are presented in Table A13 in the Appendix.

come from the MA and little effects are detected in non-MA districts. The magnitudes of the coefficients of DD models imply that having a large share of charter schools (ever above 10 percent) increases the graduation rate in MAs by four percentage points and improves Math and ELA scores in MAs by 0.13-0.21 standard deviations. This pattern is consistent with prior charter school participant effects by urbanicity (CREDO 2013, Chabrier, Cohodes, and Oreopoulos 2016).

The reasons for this form of effect heterogeneity remain speculative. One possible explanation is that it is simply easier to induce improvement in districts that are very low-performing to start with (Harris and Larsen forthcoming) and urban schools have lower outcome levels. We test this by estimating versions of equations (1) and (2) but adding an interaction between treatment and baseline achievement level (see Table A14 in the Appendix). Most of these interaction coefficients are negative (including eight of the nine coefficients from models including covariates), which is consistent with the above theory, though they are imprecisely estimated. Another possibility is that certain types of charter schools, such as those using the "No Excuses" approach (Angrist, Pathak, and Walters 2013), are more likely to locate in urban areas, (Chabrier, Cohodes, and Oreopoulos 2016), but this is difficult to test in these data.

We also evaluate the heterogeneous effects by school grade levels for test scores and report the results in Table 9. The results show precisely estimated positive effects on both Math and ELA test scores for middle school students (grades 6-8), but smaller and more sporadic effects for elementary school students (grades 3-5). Specifically, districts with a large share of charter schools (ever above 10 percent) improve Math and ELA scores of middle schools by 0.11-0.22 standard deviations.

Our average treatment effects are also a weighted average across states with different charter policies (e.g., how well funded charter schools are relative to traditional public schools, charter authorization and shutdown policies). Since there are many dimensions to these policies we interacted treatment with state charter law scores reported by the National Alliance for Public Charter Schools (NAPCS) in 2020 (Ziebarth 2020).[38] Results in Table 10 suggests that charter schools in states with higher scores from NAPCS do more to improve student outcomes relative to those in states with lower scores.

6. Mechanisms of the Total Charter Effect

The above analyses estimate the net effects of charter entry on entire school districts, capturing the participant, competition, and closure effects, collectively. Prior research provides ample examples of participant effects (Brewer et al. 2003, Hoxby and Rockoff 2004, Bifulco and Ladd 2006, Sass 2006, Booker et al. 2007, Hanushek et al. 2007, Abdulkadiroğlu et al. 2011, Furgeson et al. 2012, CREDO 2013, Curto and Fryer 2014) and competition effects (Hoxby 2003, Bettinger 2005, Bifulco and Ladd 2006, Sass 2006, Ni 2009, Zimmer and Buddin 2009, Imberman 2011, Linick 2014, Cordes 2018, Griffith 2019, Han and Keefe 2020). In this section, we switch from district- to school-level data to provide new evidence on a mechanism that has received less attention in the literature: how charter schools induce closure of low-performing TPS, as well as private schools.

6.1 Charter Entry Effects on TPS Closures

[38] NAPCS rated each state on 21 key components of state law, such as accountability, authorization, flexibility, performance-based contracts, and funding equity. NAPCS gave each component a weight of 1 to 4 and multiplied these by the component rating to obtain an index for each state. We divided these scores by the total score and therefore, the value of scores ranges from zero to one with the mean of 0.53 and the standard deviation of 0.23.

We evaluate the effects of charter entry on TPS closure during the sample period (1995-2016) using the TPS school closure measure created by Harris and Martinez-Pabon (2021). A closed TPS occurs when the school building is no longer used as a school. We therefore switch the dependent variable in the above DD analyses from test scores and graduation rates to closure.

Table 11 presents the results.[39] We find increases in TPS closure when charters enter. The magnitudes of the coefficients imply that having a charter enrollment share ever above 10 percent increases the school closure rate by about 0.5 percentage points (a 40 percent increase over the baseline rate). In two additional analyses, we also carry out the placebo version, examining "effects" of elementary charter school entry on high school closure (vice versa). Panels B and C show that there are positive but mostly insignificant placebo effects, especially when controlling district covariates, reinforcing that the main estimates reflect causal effects.

Whether these induced TPS closures improve district performance depends on whether the closed schools are low-performing relative to the entering charter schools and the other TPSs to which students in closed schools may be forced to move (Bross, Harris, and Liu 2016). The SEDA school-level data provides overall performance measure and overall achievement growth in Math and ELA for grades 3-8 in the sample period. Our primary interest is in the SEDA school achievement growth measure, which is based on the difference between average scores of all students in a specific grade in a school and the average scores of students in the previous grade in the prior year. For example, the SEDA growth measure captures how much student test scores changed, on

---

[39] Figure A5 in the Appendix presents the event study result of charter effects on TPSs closure.

average, from 3rd grade in one year to 4th grade in the following year (i.e., cohort growth). On average, SEDA-style cohort growth measures are useful proxies for longitudinal growth measures similar to school value-added (Reardon et al. 2019).[40]

These data suggest that the closed TPSs had lower performance than non-closed TPSs. The SEDA overall growth measure of Math & ELA test scores is 0.0041 s.d. for non-closed TPSs in treated districts, but -0.0023 s.d. for closed TPSs in treated districts. The practical effects of these differences might be smaller because the low-performing teachers in closed TPS may end up moving to the remaining TPS and charter schools. However, we note: (a) prior research suggests that teachers in closed/takeover schools are more likely than other teachers to leave the profession entirely and/or move into non-teaching positions (Lincove, Carlson, and Barrett 2019); and (b) schools and school leaders vary in their ability to convert individual teacher skill into student outcomes (Branch, Hanushek, and Rivkin 2012), so, even if teachers were all re-sorted to other schools, average school performance would still improve to some degree.

6.2 Effects of Charter Entry on Private School Enrollment

So far, the above analyses only look at the charter effects on student outcomes in public school systems, however, charter schools may also have some impacts on enrollment patterns in private schools, which also matter from a social welfare perspective. Also, if some private school students end up in publicly funded schools as a result of the charter entry, then our estimates on district-level graduation rate and test scores (which only capture TPSs and charter schools) might be biased upwards. For example, Toma, Zimmer, and Jones (2006) found that, in Michigan, approximately 17 percent of the

---

[40] In our own analysis of the Louisiana SEDA measures, we find a correlation of at least +0.6 between SEDA growth and conventional school value-added measures based on student-level data.

students who enroll in charter schools were previously enrolled in private schools. Buddin (2012) conducted a national evaluation and found similar results.[41] However, Chakrabarti and Roy (2016) found no causal evidence that charter schools in Michigan led to declines in private school enrollment.[42]

We focus on the districts that have at least one private school at baseline, and some small districts without any private schools are not included in the analysis.[43] The sample period is 1996-2016 and note that the private school data are only available biannually. We generally find no significant effects of charter entry on the share of private schools' enrollment as Table A15 in the Appendix presents.[44] A further implication is that our earlier estimates on district-level graduation rate and test scores are not biased by shifts in enrollments to or from private schools.

6.3 Other Mechanisms

Our data only allow suggestive evidence regarding participant and competition effects, which we present briefly. To understand the participant effects, we identified the single nearest TPS to every charter serving the same grades. The average SEDA growth measure for these nearest TPS schools is 0.0018, compared with the overall district average of 0.0055, in districts that have charter schools. In other words, it appears that charter schools tend to locate near TPS that are slightly lower performing, which likely contributes to the net positive effects reported earlier. The average SEDA growth measure of charter schools (0.0167) and is also larger than that for the nearest TPS

---

[41] This was especially true in urban districts.
[42] They used a fixed-effects model as well as an instrumental variables strategy that exploits exogenous variation from Michigan charter law.
[43] About 4,588 school districts have one private school during the sample period.
[44] One specification shows statistically significant reductions in private school enrollment, but that effect disappears when including district controls as column (4) shows. We also carried out a placebo analysis in Panel B and Panel C, which reinforces that the effect is null.

schools (0.0018), implying a participant effect of 0.0149 s.d. in annual growth. We note that this figure is very similar to what CREDO (2013) found using microdata and analysis of "virtual twins." This may be an upper-bound in the present context, however. While prior research suggests that school value-added measures generally have limited bias on average (Deming 2014, Angrist et al. 2017), Reardon et al. (2019) warn that their charter school measures are upwardly biased relative to TPS. In the additional personal communication (Kilbride, 2021), the SEDA charter-TPS bias appears to be 0.018 s.d., which would make the participant effect almost zero.

The school-level SEDA growth measures can also be used to gauge competitive effects. We would expect that having a larger charter market share would create more competition and do more to increase performance of TPS. Focusing on those school districts with both charter schools and TPS, we therefore take the weighted average achievement growth among TPS only in each district and do the same for charters and present a scatter plot of achievement growth for each sector by district charter share.[45] We then create a linear fit (linear, quadratic, and cubic) to gauge the average slope and then to test for non-linearities.

We are mainly interested in the slope of the TPS line. If competition is occurring, then we expect more charter schools to be associated with improved TPS growth. That slope is relatively flat, but generally upward sloping, as predicted.[46] This suggests that the competitive effect is relatively small. For example, if we take the fitted linear fit line

---

[45] Charter share is the average charter enrollment share during the sample period. The linear fitted line in Figure 5(a) is higher for charter schools than TPSs at all market shares (consistent with our earlier results), reflecting the previously described positive participant effect.

[46] In Figure 5, the slope of achievement growth for charter schools (red line in Figure 5a) is 0.0112, whereas that for TPSs (blue line in Figure 5a) is 0.0063.

literally, going from zero to 50 percent charter market share would increase annual TPS achievement growth by 0.0031 s.d.. This is consistent with the prior literature, which also finds small competitive effects using rigorous methods applied to smaller samples. Again, this evidence on participant and competitive effects is only descriptive and our main contribution with respect to mechanisms is with respect to school closures.

7. Conclusion

This study makes many new contributions to the research on charter schools and market-based school reforms generally. First, we provide the first analysis of the total combined effect of charter school entry on student outcomes. Using the DD strategy combined with various methods for addressing endogeneity, our analysis suggests that districts with more than 10 percent charters have increased elementary and middle school test scores by 0.06-0.16 standard deviations.

Second, we also contribute to a small but growing literature to study effects on student outcomes other than test scores (Imberman 2011, Booker et al. 2011, Furgeson et al. 2012, Angrist et al. 2013, Angrist, Pathak, and Walters 2013, Dobbie and Fryer 2013, Wong et al. 2014, Booker et al. 2014). We find that increased charter market share above 10 percent increases high school graduation rate by about 2-4 percentage points. This is important given how strongly predictive high school graduation is for long-term life outcomes.

Third, this is the first study to our knowledge to identify, and simultaneously study and compare, all of the mechanisms of the effects of charter entry: participant, competitive, and closure effects. In particular, we find that a portion of the system effects discussed above are due to charter schools replacing low-performing TPS—the closure effect. Understanding why charter effects emerge is just as important as estimating the effects themselves.

Fourth, we provide one of the first studies of charter schools using a sample that includes almost half of all charter schools, and nearly all TPS, in the country. CREDO (2013) is the other national study. National analysis is important given the increasing national, and federal, debates about charter schools.

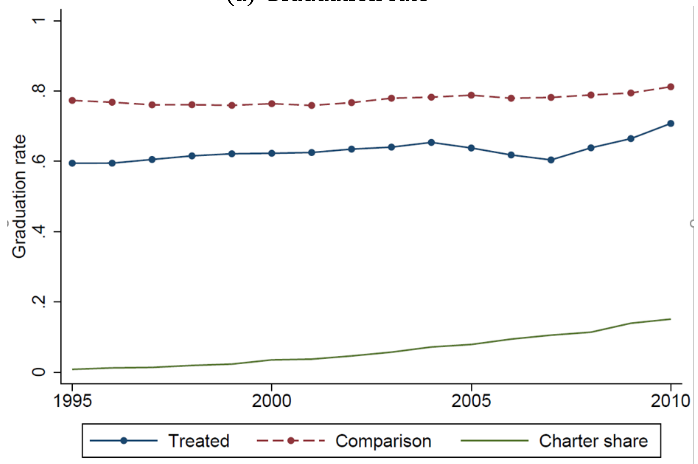
Fifth, with such a large sample, we are able to provide new evidence on effect heterogeneity. In addition to reinforcing past evidence regarding the concentration of effects in urban/MSA areas, we provide some initial evidence on why this might be. Specifically, we find suggestive evidence that charter effects are larger in lower-performing districts. The effects also seem to be larger in middle schools.

Sixth, the rising number of districts, such as Detroit, New Orleans and Washington, DC, that are majority-charter also raises the question: Is there some limit regarding how much charter market share is good for students? We find that there may be such as plateau. While the results in New Orleans have been especially positive (Harris and Larsen forthcoming), with near 100 percent charter market share, this may be an outlier. We find that noticeable effects begin emerging at 5 percent charter market share and rise and plateau at about 15 percent market share.

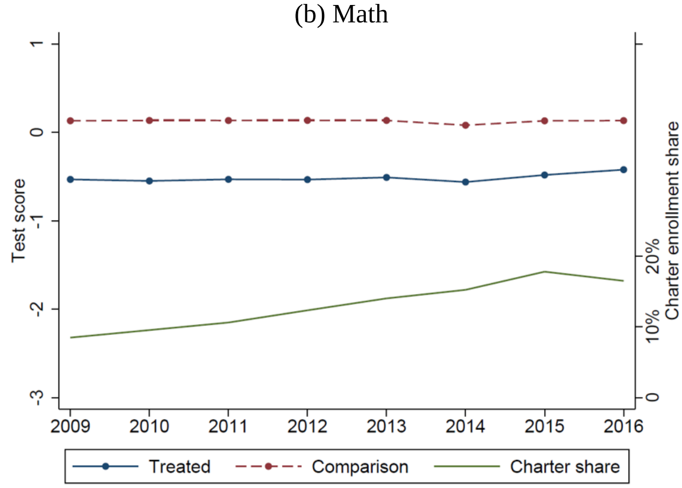
Charter school has been arguably the most influential school reform efforts of the past several decades. Informing these ongoing debates, our analysis suggests that charter schools improve student two important outcomes: test scores and high school graduation. Understanding welfare effects also requires, however, understanding when, how, and for whom charter school effects operate. Looking at data from across the country, our analysis makes many contributions toward end.

Figure 1 Trends in graduation rate, Math, and ELA performance

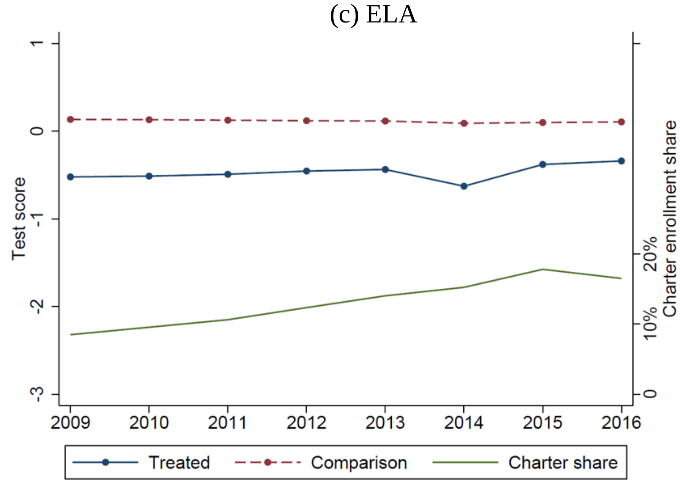
(a) Graduation rate



(b) Math

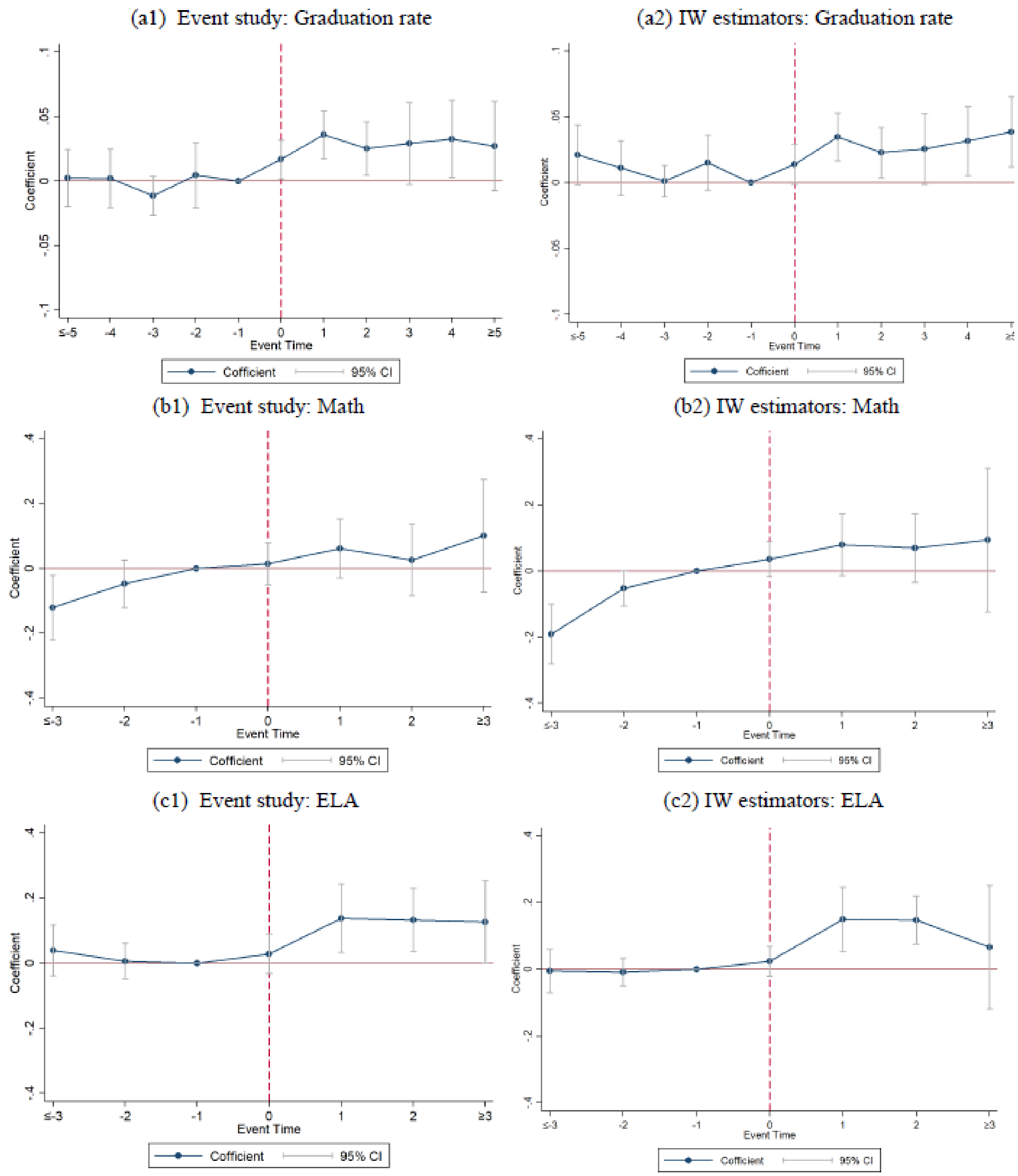


(c) ELA



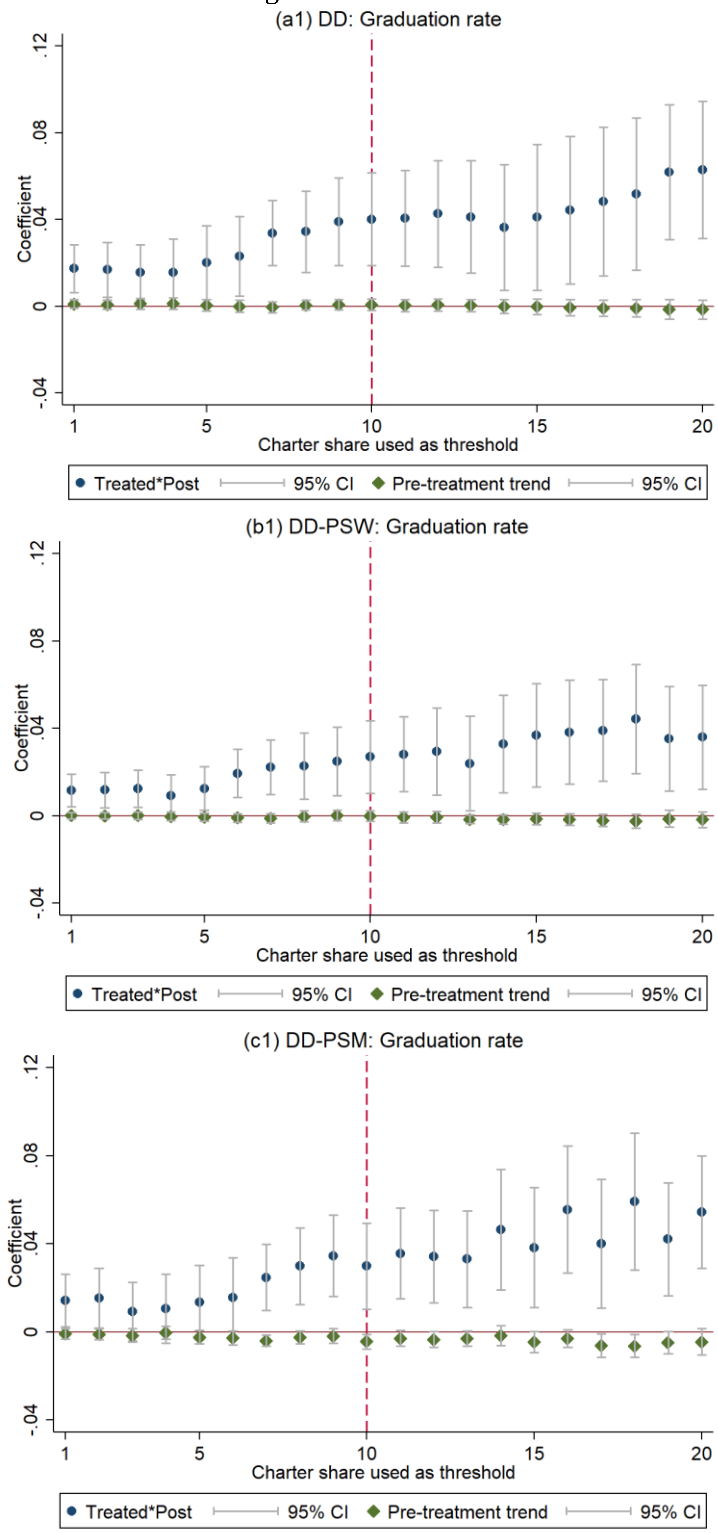
Notes: This figure plots the trends in Graduation rate, Math & ELA of treated districts (in the solid line), and comparison districts (in the dashed line). The green solid line plots the charter enrollment share of treated districts.

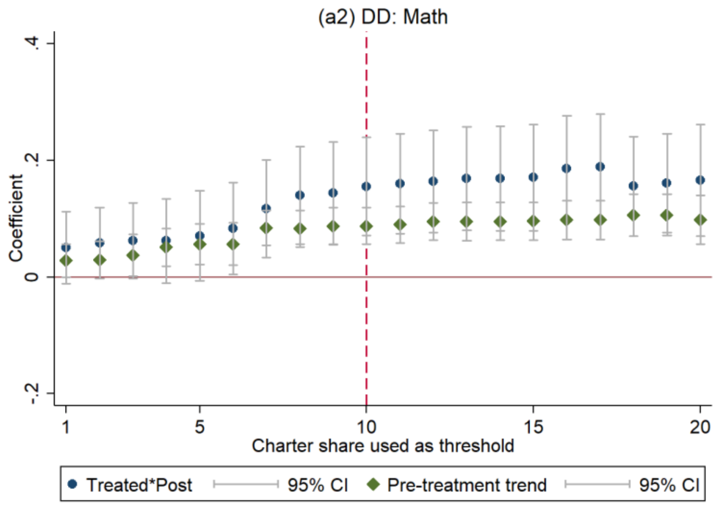
Figure 2 Event study results

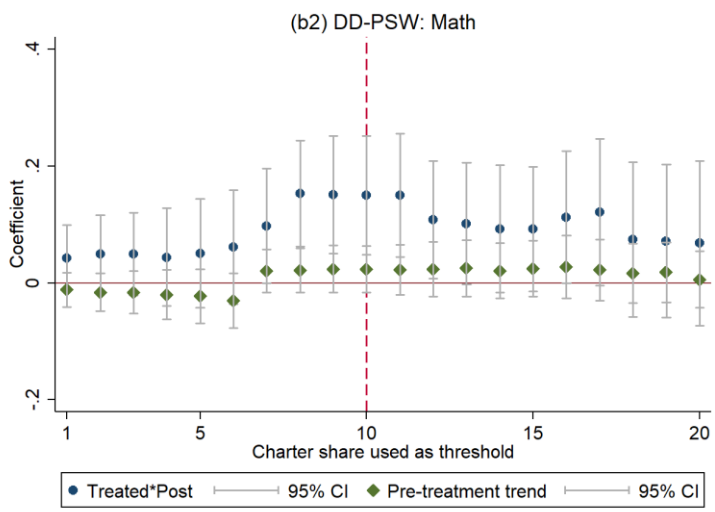


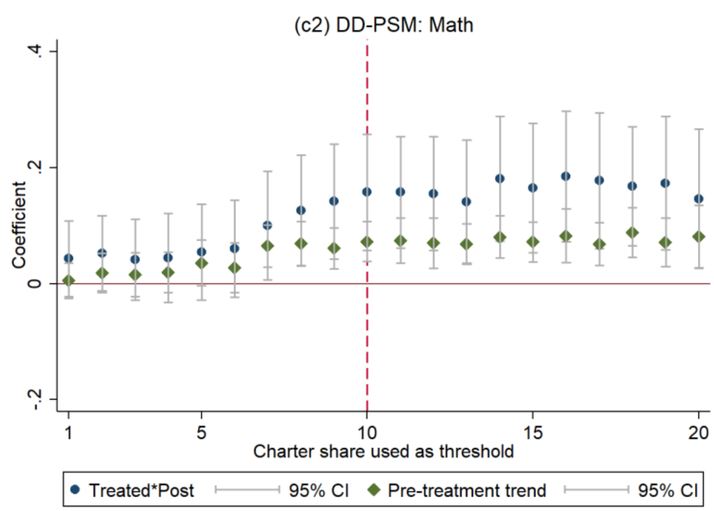
Notes: This figure presents results of event study and IW estimators for graduation rate, Math, and ELA. The event time zero is the first year of charter entry.

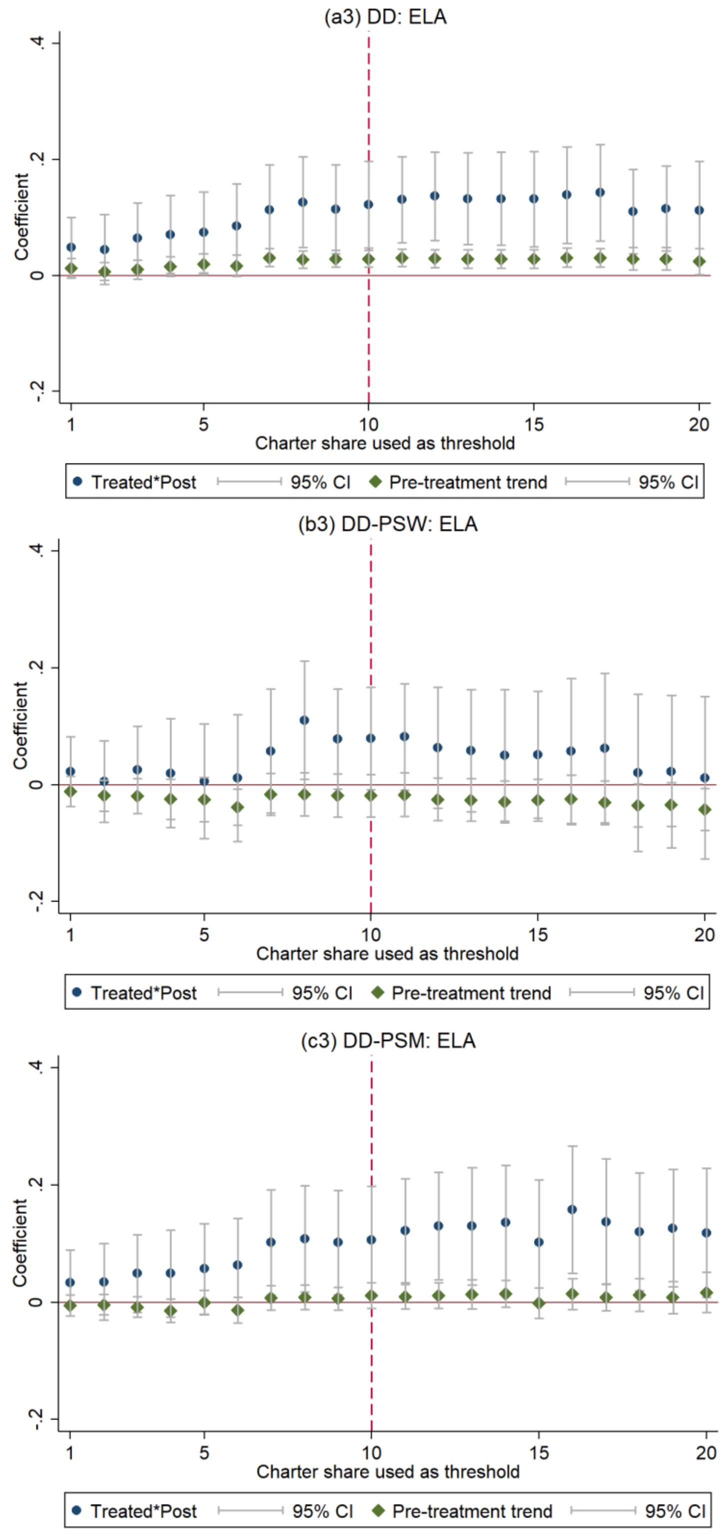
Figure 3 Plots of estimates using 1% to 20% as the threshold of treated districts




(a2) DD: Math
Coefficient
-.2
0
.2
.4
1
5
10
15
20
Charter share used as threshold
Treated*Post
95% CI
Pre-treatment trend
95% CI

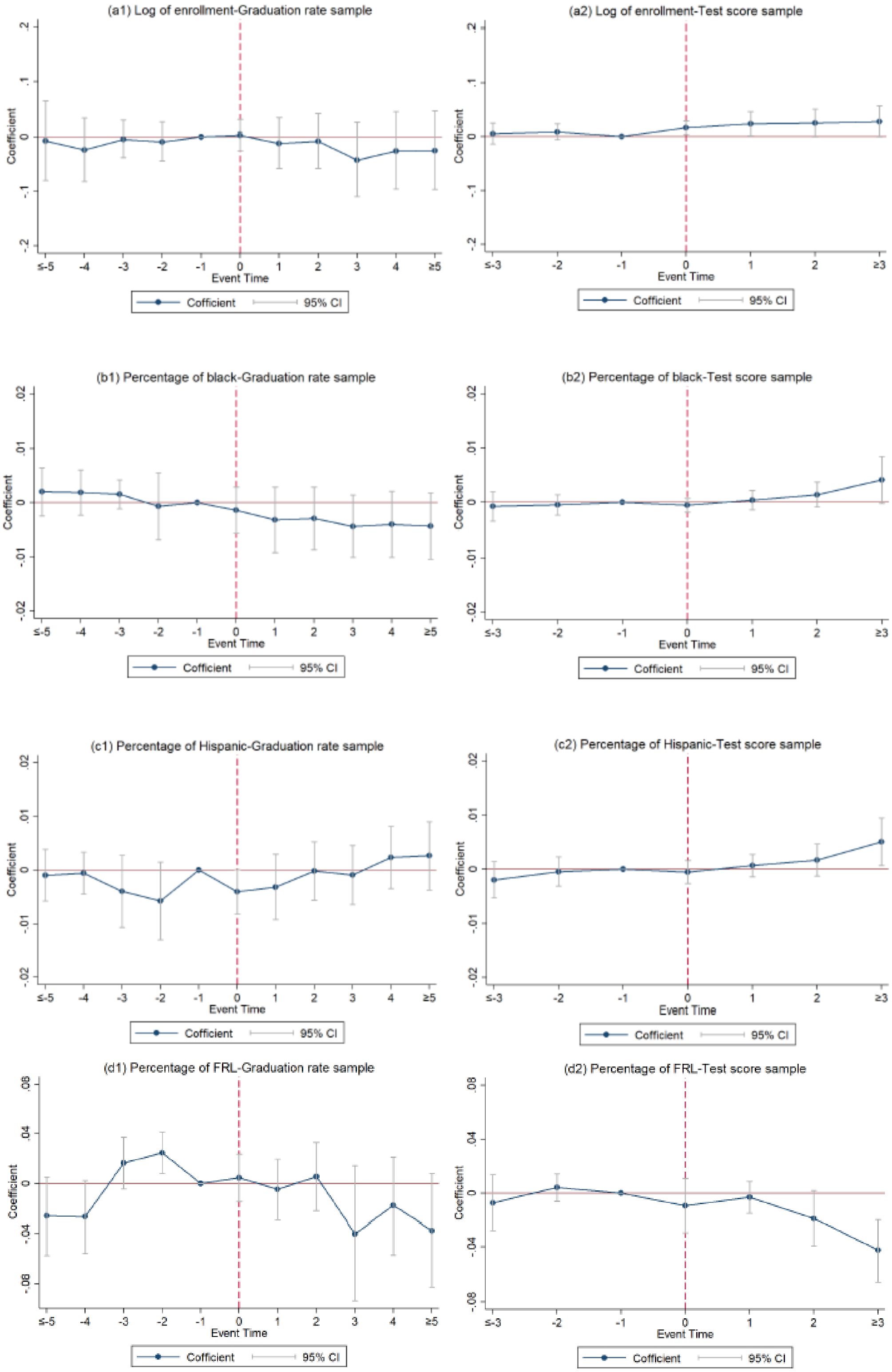

(b2) DD-PSW: Math
Coefficient
-.2
0
.2
.4
1
5
10
15
20
Charter share used as threshold
Treated*Post
95% CI
Pre-treatment trend
95% CI


(c2) DD-PSM: Math
Coefficient
-.2
0
.2
.4
1
5
10
15
20
Charter share used as threshold
Treated*Post
95% CI
Pre-treatment trend
95% CI



Notes: This figure plots the estimates in graduation rate, Math & ELA using the charter enrollment share of 1- 20% as the threshold of treated districts.

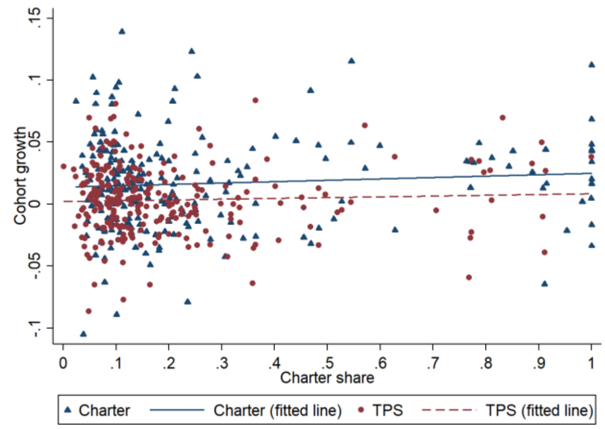
Figure 4 Event study results of enrollment and its compositions



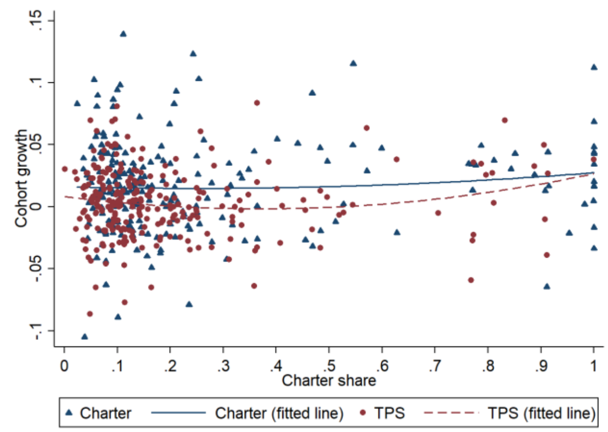
Notes: This figure presents the event study results enrollment and its compositions.

Figure 5 Charter share and cohort growth for test score

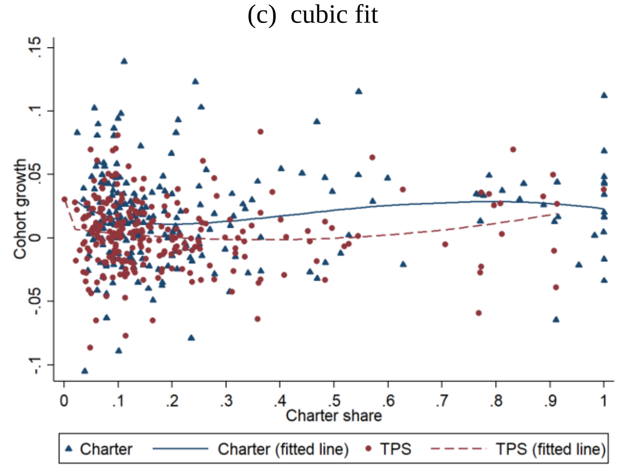
(a) linear fit



(b) quadratic fit



(c) cubic fit



Notes: This figure plot the (a) linear, (b) quadratic, and (c) cubic fit of scatters for the charter share and cohort growth measure of test score at district by charter status level. The sample is district with charter share ever above 10 percent in the sample period, and the charter share is the average charter enrollment share during 2009-2016. In figure (a), the slope of the top (charter) line is 0.0112 and 0.0063 for the bottom (TPS) line.

Table 1 Summary Statistics

Sample	Graduation rate (1995-2010)						Math & ELA (2009-2016)					
	ALL	DD &		DD-PSM		Fixed Effects	ALL	DD & DD-PSW		DD-PSM		Fixed Effects
		Treated	Control	Treated	Control			Treated	Control	Treated	Control	
Graduation rate	0.75	0.63	0.78	0.63	0.69	0.67	NA	NA	NA	NA	NA	NA
Math	NA	NA	NA	NA	NA	NA	0.00	-0.51	0.13	-0.51	-0.19	-0.40
ELA	NA	NA	NA	NA	NA	NA	0.00	-0.46	0.12	-0.46	-0.20	-0.36
White	65%	30%	71%	30%	33%	43%	56%	30%	63%	30%	46%	36%
Black	16%	23%	14%	23%	37%	22%	16%	23%	14%	23%	19%	23%
Hispanic	14%	38%	10%	38%	22%	26%	22%	42%	18%	42%	27%	34%
FRL	30%	49%	27%	49%	40%	37%	51%	65%	48%	65%	53%	61%
Special education	13%	12%	13%	12%	12%	12%	13%	12%	13%	12%	12%	12%
Ages 5–17 population	16%	24%	15%	24%	20%	19%	19%	23%	18%	23%	19%	22%
Ages 5–17 in poverty	19%	18%	19%	18%	19%	18%	17%	16%	18%	16%	17%	17%
Urban	27%	73%	22%	73%	60%	45%	27%	45%	22%	45%	40%	42%
Suburb	42%	21%	41%	21%	31%	48%	45%	46%	44%	46%	49%	51%
Town	11%	3%	13%	3%	6%	3%	12%	6%	15%	6%	7%	4%
Rural	20%	3%	24%	3%	4%	3%	17%	4%	21%	4%	7%	4%
Revenue per student	8,970	8,900	9,057	8,900	9,448	8,685	12,125	11,438	12,375	11,438	11,351	11,344
Expenditure per student	9,132	9,172	9,215	9,172	9,701	8,861	12,141	11,697	12,356	11,697	11,363	11,467
Teacher salary	79,686	91,449	78,534	91,449	88,776	83,465	99,038	102,014	98,660	102,014	103,108	100,225
Student teacher ratio	17	19	16	19	17	18	16	18	16	18	19	18
No. magnet school	5	52	1	52	4	17	6	35	1	35	6	19
No. public schools	61	311	29	311	135	168	62	290	25	290	79	179
Enrollment	46,690	264,722	18,961	264,722	89,521	137,604	42,851	198,433	16,748	198,433	55,331	124,652
Observation	144,219	2,193	137,780	2,193	2,199	6,439	410,077	10,109	388,064	10,109	11,073	22,013
N (district)	9,278	142	8,859	142	142	416	10,439	297	9,832	297	297	607

Notes: This table presents weighted means of outcome variables (graduation rate, math, and ELA) and control variables. graduation rate sample is weighted by high school enrollment, and Math & ELA sample is weighted by grade level enrollment. Treated group refers to the sample of districts that charter enrollment share ever above 10 percent during the sample period. Comparison group refers to the sample of districts without charter schools in all states for DD & DD-PSW, and it refers to the sample of matched districts (nearest neighbor) for DD-PSM. Fixed Effects sample refers to all districts with any charter share in the sample period. Data source: National Longitudinal School Database.

Table 2 Year charter law passed by state

Year	State
1991	Minnesota
1992	California
1993	Colorado, Massachusetts, Michigan, New Mexico, Wisconsin
1994	Arizona, Georgia, Hawaii, Kansas
1995	Alaska, Arkansas, Delaware, Louisiana, New Jersey, Rhode Island, Texas, Wyoming
1996	Connecticut, District of Columbia, Florida, Idaho, Illinois, New Hampshire, North Carolina, South Carolina
1997	Nevada, Ohio, Pennsylvania
1998	Missouri, New York, Utah, Virginia
1999	Oklahoma, Oregon
2001	Indiana
2002	Iowa, Tennessee
2003	Maryland
2010	Mississippi
2011	Maine
2015	Alabama
2016	Washington
2017	Kentucky
2019	West Virginia
NA	Montana, Nebraska, North Dakota, South Dakota, Vermont

Data source: National Longitudinal School Database.

Table 3 Effects of charter entry on student outcomes

	(1)	(2)	(3)	(4)	(5)	(6)
	DD		DD-PSW		DD-PSM	
Panel A: Graduation rate						
Treated*Post	0.037***	0.038***	0.034***	0.028***	0.039***	0.031***
	[0.009]	[0.011]	[0.011]	[0.009]	[0.010]	[0.010]
R-squared	0.805	0.808	0.817	0.825	0.886	0.894
Observations	139,973	139,973	139,973	139,973	4,383	4,383
N (district)	9,001	9,001	9,001	9,001	284	284
Pre-treatment trend	0.000	0.001	-0.001	-0.001	-0.001	-0.002
	[0.001]	[0.001]	[0.001]	[0.001]	[0.002]	[0.002]
Panel B: Math						
Treated*Post	0.152***	0.155***	0.154***	0.148***	0.136***	0.157***
	[0.040]	[0.043]	[0.055]	[0.052]	[0.048]	[0.051]
R-squared	0.846	0.847	0.863	0.865	0.869	0.873
Observations	398,173	398,173	398,173	398,173	21,007	21,007
N (district)	10,129	10,129	10,129	10,129	594	594
Pre-treatment trend	0.085***	0.088***	0.023	0.023	0.074***	0.073***
	[0.018]	[0.016]	[0.024]	[0.020]	[0.020]	[0.018]
Panel C: ELA						
Treated*Post	0.116***	0.121***	0.075	0.077*	0.075*	0.106**
	[0.035]	[0.038]	[0.049]	[0.045]	[0.043]	[0.046]
R-squared	0.882	0.883	0.884	0.886	0.914	0.917
Observations	398,173	398,173	398,173	398,173	21,007	21,007
N (district)	10,129	10,129	10,129	10,129	594	594
Pre-treatment trend	0.026***	0.029***	-0.025	-0.019	0.009	0.011
	[0.010]	[0.008]	[0.023]	[0.019]	[0.012]	[0.011]
District, (grade) & year FE	Yes	Yes	Yes	Yes	Yes	Yes
District control	No	Yes	No	Yes	No	Yes

Notes: The table shows DD estimates of the effects of the charter entry (enrollment share ever above 10%) on student outcomes. The treated group includes districts with charter enrollment share ever above 10%, and the comparison group includes districts without charter schools in all states; post in an indicator of the period after districts started first charter school. Controls include the log of district enrollment; the share of students who are Hispanic, black, white; the share of students who are in special education programs; the share of students on FRL programs; student-teacher ratio; average teacher salary; the number of magnet school; the total number of schools; the total revenue per student; the total expenditure per student; and whether the district is in an urban, suburban, town, or rural location; the estimate of the school-age population; the estimate poverty rate of the school-age population. Robust standard errors presented in parentheses are clustered at the district level. For DD and DD-PSM, regressions are weighted by high school enrollment for graduation rate and grade-level enrollment for Math and ELA; For DD-PSW, regressions are weighted by weight of DD times the inverse probability of propensity score.

Table 4 Weights and Estimates from the Goodman-Bacon (2021) Decomposition

	Weight	Estimate (1)	Estimate (2)
Panel A: Graduation rate			
Earlier Treated vs. Later Control	0.154	0.015	0.033
Later Treated vs. Earlier Control	0.116	0.022	0.049
Treated vs. Never Treated	0.683	0.011	0.024
Treated vs. Already Treated	0.047	0.028	0.062
Overall DD Estimate		0.014	0.031
Panel B: Math			
Earlier Treated vs. Later Control	0.023	-0.226	-0.645
Later Treated vs. Earlier Control	0.039	-0.018	-0.051
Treated vs. Never Treated	0.585	0.045	0.128
Treated vs. Already Treated	0.353	0.099	0.283
Overall DD Estimate		0.055	0.157
Panel C: ELA			
Earlier Treated vs. Later Control	0.023	-0.063	-0.142
Later Treated vs. Earlier Control	0.039	-0.006	-0.014
Treated vs. Never Treated	0.585	0.027	0.061
Treated vs. Already Treated	0.353	0.094	0.212
Overall DD Estimate		0.047	0.106
Sample weight, Covariates & FE		No	Yes

Notes: The Goodman-Bacon (2021) decomposition above displays the weights and components making up the overall DD estimates of charter entry on student outcomes. Decompositions are documented for graduation rate (panels A), Math (panels B), and for ELA (panels C) using the DD-PSM sample. Column (1) doesn't include sample weight, districts covariates, district fixed effects and year fixed effects while Column (2) does. Column (2) present the same estimates as Column (6) in Table 3. For the decomposition, each components' weight is given along with the point estimate for this comparison. The overall DD estimate is displayed at the foot of each panel.

Table 5 DD estimates: alternative comparison group (states no charter law)

	(1)	(2)	(3)	(4)	(5)	(6)
	DD		DD-PSW		DD-PSM	
Panel A: Graduation rate						
Treated*Post	0.041***	0.035***	0.034***	0.030***	0.037***	0.028***
	[0.009]	[0.010]	[0.013]	[0.011]	[0.010]	[0.011]
R-squared	0.859	0.867	0.808	0.823	0.861	0.874
Observations	13,948	13,948	13,948	13,948	4,420	4,420
N (district)	895	895	895	895	284	284
Pre-treatment	0.001	0.002	0.000	0.000	0.000	0.001
trend	[0.001]	[0.001]	[0.001]	[0.001]	[0.002]	[0.002]
Panel B: Math						
Treated*Post	0.129**	0.154***	0.183***	0.166***	0.123**	0.151***
	[0.051]	[0.047]	[0.057]	[0.055]	[0.052]	[0.048]
R-squared	0.806	0.812	0.865	0.869	0.820	0.826
Observations	32,128	32,128	32,128	32,128	21,481	21,481
N (district)	887	887	887	887	594	594
Pre-treatment	0.081***	0.084***	0.002	0.017	0.075***	0.083***
trend	[0.021]	[0.021]	[0.027]	[0.019]	[0.022]	[0.023]
Panel C: ELA						
Treated*Post	0.076*	0.101**	0.106**	0.102**	0.075	0.107**
	[0.044]	[0.042]	[0.051]	[0.046]	[0.046]	[0.043]
R-squared	0.868	0.872	0.874	0.877	0.883	0.887
Observations	32,128	32,128	32,128	32,128	21,481	21,481
N (district)	887	887	887	887	594	594
Pre-treatment	0.045***	0.044***	-0.025	-0.003	0.042***	0.043**
trend	[0.012]	[0.017]	[0.026]	[0.015]	[0.013]	[0.019]
District, (grade) & year FE	Yes	Yes	Yes	Yes	Yes	Yes
District control	No	Yes	No	Yes	No	Yes

Notes: The table shows DD estimates use an alternative comparison group. The comparison group includes districts in states without charter law as of the last year of the sample (2016). See notes of Table 3 for controls, clusters, and sample weight.

Table 6 DD estimates: Placebo grade level charter share

	(1)	(2)	(3)	(4)	(5)	(6)
	DD		DD-PSW		DD-PSM	
	Panel A: Graduation rate					
Treated*Post	0.016*	0.013	0.001	-0.009	0.017*	0.012
	[0.008]	[0.009]	[0.012]	[0.016]	[0.009]	[0.009]
R-squared	0.816	0.819	0.726	0.748	0.877	0.881
Observations	144,112	144,112	144,112	144,112	4,623	4,623
N (district)	9,268	9,268	9,268	9,268	300	300
Pre-treatment	0.000	-0.001	0.001	0.000	-0.002	-0.003*
trend	[0.001]	[0.002]	[0.001]	[0.001]	[0.002]	[0.002]
	Panel B: Math					
Treated*Post	-0.091	-0.084	-0.022	-0.049	-0.102*	-0.093*
	[0.056]	[0.053]	[0.059]	[0.048]	[0.058]	[0.055]
R-squared	0.851	0.852	0.872	0.875	0.877	0.880
Observations	409,129	409,129	409,129	409,129	24,268	24,268
N (district)	10,414	10,414	10,414	10,414	662	662
Pre-treatment	-0.006	-0.001	0.000	-0.009	-0.012	-0.018
trend	[0.013]	[0.012]	[0.013]	[0.009]	[0.014]	[0.013]
	Panel C: ELA					
Treated*Post	-0.006	-0.001	0.008	-0.009	-0.051	-0.034
	[0.048]	[0.046]	[0.059]	[0.052]	[0.050]	[0.048]
R-squared	0.887	0.888	0.895	0.897	0.917	0.919
Observations	409,129	409,129	409,129	409,129	24,268	24,268
N (district)	10,414	10,414	10,414	10,414	662	662
Pre-treatment	0.010	0.013	0.006	0.006	0.001	0.006
trend	[0.014]	[0.014]	[0.013]	[0.011]	[0.015]	[0.015]
District, (grade) & year FE	Yes	Yes	Yes	Yes	Yes	Yes
District control	No	Yes	No	Yes	No	Yes

Notes: The table shows DD estimates using placebo grade levels of charter share: elementary school grade levels (1-5) for graduation rate, and high school grade levels (9-12) for Math and ELA. See notes of Table 3 for controls, clusters, and sample weight.

Table 7 Estimates of fixed effects model for charter effects on student outcomes

	(1)	(2)	(3)	(4)
Panel A: Graduation rate				
	Same year share		Average last four years	
Charter share	0.130*	0.159***	0.115	0.125**
	[0.074]	[0.052]	[0.076]	[0.060]
R-squared	0.852	0.860	0.863	0.871
Observations	6,439	6,439	5,191	5,191
N(district)	416	416	416	416
Panel B: Math				
	Same year share		Same cohort last year	
Charter share	0.390***	0.339***	0.266**	0.211*
	[0.143]	[0.116]	[0.107]	[0.114]
R-squared	0.854	0.858	0.877	0.878
Observations	22,013	22,013	13,728	13,728
N(district)	607	607	604	604
Panel C: ELA				
	Same year share		Same cohort last year	
Charter share	0.195	0.043	0.138	0.062
	[0.129]	[0.107]	[0.103]	[0.102]
R-squared	0.909	0.912	0.922	0.924
Observations	22,013	22,013	13,728	13,728
N(district)	607	607	604	604
District, (grade) & year FE	Yes	Yes	Yes	Yes
District control	No	Yes	No	Yes

Notes: The table shows estimates of fixed effects model for charter effects on student outcomes for districts with any charter schools during the sample period. In columns (1) and (2), charter enrollment share of grades 9-12 is used for graduation rate, and charter enrollment share of grades 3-8 is used for test scores. In columns (1) and (2), the average last four years charter enrollment share of grade 9-12 is used for graduation rate, and the last-year same cohort grade enrollment share is used for test scores. See notes of Table 3 for controls and clusters. Regressions are weighted by high school enrollment for graduation rate and grade-level enrollment for Math and ELA.

Table 8 Effect heterogeneity: metropolitan areas VS non-metropolitan areas

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Graduation rate			Math			ELA		
	DD	DD-PSW	DD-PSM	DD	DD-PSW	DD-PSM	DD	DD-PSW	DD-PSM
Panel A Metropolitan areas									
Treated*Post	0.042***	0.042***	0.040***	0.170***	0.210***	0.162***	0.137***	0.130***	0.132***
	[0.011]	[0.010]	[0.011]	[0.045]	[0.056]	[0.052]	[0.040]	[0.036]	[0.047]
R-squared	0.838	0.865	0.909	0.873	0.887	0.865	0.907	0.907	0.912
Observations	67,462	67,462	2,424	228,532	228,532	14,743	228,532	228,532	14,743
N(district)	4,355	4,355	158	5,848	5,848	432	5,848	5,848	432
Pre-treatment	0.000	0.000	-0.001	0.101***	0.058***	0.087***	0.035***	-0.002	0.028***
trend	[0.001]	[0.001]	[0.002]	[0.016]	[0.018]	[0.019]	[0.006]	[0.023]	[0.010]
Panel B Non-metropolitan areas									
Treated*Post	-0.015	-0.007	-0.008	0.009	-0.039	-0.007	-0.028	-0.058	0.004
	[0.011]	[0.010]	[0.012]	[0.122]	[0.085]	[0.105]	[0.090]	[0.089]	[0.091]
R-squared	0.679	0.709	0.664	0.676	0.749	0.746	0.719	0.775	0.799
Observations	72,511	72,511	1,955	169,641	169,641	6,182	169,641	169,641	6,182
N(district)	4,646	4,646	126	4,281	4,281	162	4,281	4,281	162
Pre-treatment	0.000	-0.002	-0.004	-0.017	-0.016	-0.006	-0.033**	-0.031	-0.040
trend	[0.002]	[0.002]	[0.003]	[0.024]	[0.019]	[0.028]	[0.015]	[0.020]	[0.025]
District, (grade) & year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
District control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table shows DD estimates of heterogeneous effects of charter entry on student outcomes by (non-)metropolitan areas. See notes of Table 3 for controls, clusters, and sample weight.

Table 9 Effect heterogeneity: middle school VS elementary school

	(1)	(2)	(3)	(4)	(5)	(6)
		Math			ELA	
	DD	DD-PSW	DD-PSM	DD	DD-PSW	DD-PSM
Panel A: Middle school (Grade 6-8)						
Treated*Post	0.215***	0.131*	0.211***	0.117***	0.040	0.109**
	[0.048]	[0.076]	[0.054]	[0.042]	[0.054]	[0.048]
R-squared	0.888	0.898	0.925	0.903	0.901	0.936
Observations	183,088	183,088	8,897	183,088	183,088	8,897
N(district)	9,896	9,896	600	9,896	9,896	600
Pre-treatment	0.035*	-0.037	0.019	0.006	-0.041**	-0.019
trend	[0.020]	[0.023]	[0.024]	[0.015]	[0.020]	[0.014]
Panel B: Elementary school (Grade 3-5)						
Treated*Post	0.078	0.060	0.073	0.102**	0.045	0.091
	[0.053]	[0.071]	[0.062]	[0.047]	[0.045]	[0.058]
R-squared	0.860	0.878	0.888	0.902	0.910	0.938
Observations	216,913	216,913	11,699	216,913	216,913	11,699
N(district)	10,002	10,002	556	10,002	10,002	556
Pre-treatment	0.125***	0.052***	0.095***	0.048***	0.027	0.023*
trend	[0.020]	[0.018]	[0.023]	[0.010]	[0.019]	[0.012]
District, grade & year FE	Yes	Yes	Yes	Yes	Yes	Yes
District control	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The table shows DD estimates of heterogeneous effects of charter entry on student test scores by middle (elementary) schools. See notes of Table 3 for controls, clusters, and sample weight.

Table 10 Effect heterogeneity: state charter law score

	(1)	(2)	(3)	(4)	(5)	(6)
	DD		DD-PSW		DD-PSM	
Panel A: Graduation rate						
Treated*Score *Post	0.058***	0.057***	0.057***	0.044***	0.060***	0.046***
	[0.016]	[0.018]	[0.021]	[0.017]	[0.017]	[0.017]
R-squared	0.807	0.810	0.819	0.826	0.886	0.895
Observations	126,285	126,285	126,285	126,285	4,336	4,336
N (district)	8,122	8,122	8,122	8,122	281	281
Pre-treatment trend	0.000	0.000	0.000	0.000	0.000	0.000
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Panel B: Math						
Treated*Score *Post	0.256***	0.257***	0.257***	0.247***	0.216***	0.249***
	[0.062]	[0.067]	[0.096]	[0.089]	[0.075]	[0.081]
R-squared	0.848	0.849	0.865	0.867	0.868	0.872
Observations	363,534	363,534	363,534	363,534	20,254	20,254
N (district)	9,261	9,261	9,261	9,261	577	577
Pre-treatment trend	0.003***	0.003***	0.001*	0.001**	0.002***	0.002***
	[0.001]	[0.000]	[0.001]	[0.001]	[0.001]	[0.000]
Panel C: ELA						
Treated*Score *Post	0.182***	0.185***	0.110	0.112	0.117*	0.162**
	[0.055]	[0.061]	[0.090]	[0.081]	[0.067]	[0.074]
R-squared	0.887	0.889	0.887	0.889	0.915	0.918
Observations	363,534	363,534	363,534	363,534	20,254	20,254
N (district)	9,261	9,261	9,261	9,261	577	577
Pre-treatment trend	0.001***	0.001***	-0.001	-0.001	0.000	0.000
	[0.000]	[0.000]	[0.001]	[0.001]	[0.000]	[0.000]
District, (grade) & year FE	Yes	Yes	Yes	Yes	Yes	Yes
District control	No	Yes	No	Yes	No	Yes

Notes: The table shows results for heterogeneous effects by state charter law score by interacting charter share with the state charter law score. *Score* is a variable from based on policy scores reported by the NAPCS in 2020. Higher values indicate a better policy as determined by the rating organization. All scores are divided by the highest score and therefore, the highest value of *Score* is one and the lowest value of *Score* is zero. See notes of Table 3 for controls, sample weight, and clusters.

Table 11 Effects of charter entry on TPS closure

	(1)	(2)	(3)	(4)	(5)	(6)
	DD		DD-PSW		DD-PSM	
Panel A: district charter share on all TPS closure						
Treated*Post	0.005***	0.006***	0.005**	0.005***	0.003**	0.004***
	[0.001]	[0.001]	[0.002]	[0.002]	[0.001]	[0.001]
R-squared	0.088	0.098	0.084	0.097	0.112	0.143
Observations	257,368	257,347	257,347	257,347	40,696	40,696
N(district)	11,725	11,725	11,725	11,725	1,852	1,852
Pre-treatment	0.000***	0.000	0.001**	0.000	0.000**	0.000
trend	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Panel B: high school charter share on elementary TPS closure						
Treated*Post	0.005*	0.003	0.002	0.002	0.003	0.002
	[0.003]	[0.003]	[0.002]	[0.003]	[0.002]	[0.002]
R-squared	0.070	0.086	0.075	0.093	0.089	0.257
Observations	260,012	259,991	259,991	259,991	8,611	8,611
N (district)	11,846	11,846	11,846	11,846	392	392
Pre-treatment	0.000	0.000	0.000	0.000	0.000	0.000
trend	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Panel C: elementary charter share on high school TPS closure						
Treated*Post	0.001**	0.001*	0.001	0.001	0.001*	0.000
	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]
R-squared	0.078	0.081	0.075	0.078	0.086	0.112
Observations	263,169	263,148	263,148	263,148	20,163	20,163
N (district)	11,989	11,989	11,989	11,989	918	918
Pre-treatment	0.000	0.000	0.000	0.000	0.000*	0.000*
trend	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
District, (grade) & year FE	Yes	Yes	Yes	Yes	Yes	Yes
District control	No	Yes	No	Yes	No	Yes

Notes: This table presents DD estimates of the effects of the charter entry (enrollment share ever above 10%) on the share of TPS closure. The sample period is from the year 1995 to 2016. For DD and DD-PSM, regressions are weighted by the total number of schools; For DD-PSW, regressions are weighted by the weight of DD times the inverse probability of propensity score. See notes of Table 3 for controls and clusters.

CHAPTER 3 SILENCE BREAKING: THE ROLE OF SEXISM ON SEXUAL CRIME REPORTING IN THE METOO ERA

1. Introduction

Sexual crime is a seriously under-reported crime category in the United States. In 1995, only 28% of rape and sexual assault victims, disproportionately women, reported to police. Although this reporting rate increased to 59% in 2003, the average reporting percentage declined to 36% during 2005 - 2010 (Planty et al. 2013). Reporting of sexual crime is a high-stakes personal decision due to some prejudicial and false beliefs called “rape myths” which stereotype sexual crime victims and inhibit reporting. For example, questioning motives of reporting, blaming victims for their victimization, exonerating perpetrators of blame, and downplaying assault’s consequences for victims will lead to revictimization and psychological traumas to survivors (Payne et al., 1999). More importantly, about 3 in 4 sexual crime incidents in the United States are committed by offenders known to victims such as intimate partners and casual acquaintances (Planty et al. 2013). The strong feeling of humiliation and fear of reprisal further discourage victims from coming forward.

Acceptance of rape myths can be partly explained by the hostile sexist attitude toward women. Glick and Fiske (1996) and Lee, Fiske, and Glick (2010) propose the concept of hostile sexism which represents a misogynic attitude that stereotypes women as sexually manipulative and inferior. Such a hostile sexism explains men’s greater tolerance of sexual misconducts, increases moral disengagement from sexual offenses, and even induces a higher proclivity to commit sexual assaults (Abrams et al. 2003, Masser, Viki, and Power 2006). Despite the contribution of these qualitative analyses, researchers currently know relatively little about how the hostile sexist attitude affects reporting decision of sexual crime victims. The lack of rigorous and quantitative analyses on this topic can be attributed to two rea- sons. For one, measuring sexist attitude about women is empirically challenging because sexism is multidimensional and socially subtle. Second, even if an accurate measure on sexism is obtainable, sexual crime rate in an area could be simultaneously affected by a variety of other factors which are closely correlated to sexism, making the true relation between sexual crime and sexism ambiguous. Figure 1 illustrates how the state-level sexism is correlated to the monthly average rape rate per 100,000 populations by FBI's Uniform Crime Report between January 2015 and September 2017. Here, sexism in these states is quantified by an index constructed by Charles, Guryan, and Pan (2020) who use data from the General Social Survey (GSS) between 1977 and 1998. Exploiting responses to questions related to respondents' sexist attitude about women and some standardization procedures, they construct a unidimensional index of residents' overall sexism in 44 states.[1] Figure 1 documents a weak relation between sexism and the prevalence of sexual crime, as the slope coefficient is highly insignificant.

In this paper, we attempt to study the role of local sexism on sexual crime by exploiting the recent MeToo movement. MeToo was sparked in October 2017 when Harvey Weinstein's allegation of sexual assaults was publicized, and is arguably one of the most far-reaching and consequential social media campaigns to date which seeks empathy and solidarity among survivors. In the MeToo era, the heightened public attention to sexual violence could potentially both encourage women victims to come forward and deter of- fenders from committing more crimes. In Figure 2, we compare the

[1] Charles, Guryan, and Pan (2020) pick eight commonly asked questions and separate them into three categories: Beliefs about women's and men's appropriate roles inside and outside the home, beliefs about women's capacities, and beliefs about working mothers. Responses to these questions are not available in District of Columbia, Hawaii, Idaho, Maine, Nebraska, Nevada, and New Mexico.

GSS sexism index with changes in the average monthly rape rate between the pre-MeToo (January 2015 - September 2017) and post-MeToo (October 2017 - December 2019) periods in these states. It provides prima facie evidence that, although rape rates increased in many states after MeToo, states with lower sexism witnessed a relative larger increase. The baseline estimate suggests that a unit decrease in sexism index leads to, on average, about 0.2 more rape per 100,000 populations. An interpretation on this significant result is that, in the MeToo era when the public attention was shifted to sexual violence, differential sexism levels would lead to heterogeneous responses.

However, the empirical relation between sexism and sexual crime displayed in Figure 2 faces two challenges. For one, low and high sexism states can be markedly different in many other observed or unobserved aspects, and these factors may account for the cross- state difference in the rape rates. For example, certain economic and cultural factors, as well as policy changes that are correlated to local sexism, may simultaneously affect victims' reporting decisions. Second, a measure of sexism at the state level may be of high granularity because local sexist attitude about women can be enormously different even within the same state, making our interpretation of the correlation between sexism and sexual crime rough and inaccurate.

The contribution of this study is fourfold. First, we construct a novel, media market level measure on sexism by making use of partial least squares (PLS) and some predictors of local sexism. PLS is a popular machine learning method and primarily used for dimensionality reduction as well as dealing with multicollinearity problems in linear regressions. The primary functionality of PLS is to generate a few linearly uncorrelated new variables that substantially explain the variation in the outcome variable from a large group of multicollineared predictors. Empirically, PLS has been adopted to identify the latent information (for example, investor's sentiment) that explains stock return (Kelly and Pruitt 2013, 2015, Huang et al. 2015) In this study, the predictors of local sexism contain the Google search indices of 12 derogatory terms which include instances of sexist slurs and gender insults based on someone's appearance, intellect, sexual experience, and mental stability. We also consider the share of Trump voters in the 2016 election as an additional sexism predictor. Through PLS, we economically separate out information in the predictors that largely explains Charles, Guryan, and Pan (2020)'s state-level GSS sexism index from measurement error and noise, thus construct a low granular estimate of sexism at a media market level. Second, exploiting MeToo as an exogenous shock, we study how sexual crime rate changes across areas with differential sexism levels. As will be shown later, as one of the most sweeping social movements in the past few decades, MeToo revitalizes discussions on sexual violence across the United States. After accounting for a variety of confounding factors, we find solid evidence that areas lower in sexism reported more sexual crimes than the counterparts higher in sexism after MeToo. Third, we shed light on the potential mechanism through which local sexism would affect sexual crime in the MeToo era. Our empirical results suggest that the surge in documented sexual crimes is unlikely caused by an increase in actual incidents but by reporting in areas with relatively low sexism indices. Finally, this study contributes not only to the small but growing literature on how MeToo affects the composition of public companies' boards and legislative bodies ((Heminway 2019), firm values (Lins et al. 2020), the criminal justice system (Conklin 2020), and sexual crime reporting in the OECD countries Levy and Mattsson (2022), but also to the literature on violence against women (Aizer and Dal Bo 2009, Iyengar 2009, Aizer 2010, Iyer et al. 2012, Miller and Segal 2019).

A study closely related to ours is by Levy and Mattsson (2022) who thoroughly examine the overall MeToo effect in both the OECD countries and United States. They conduct a triple-difference analysis by categorizing countries into the strong MeToo and weak MeToo groups based on these countries' Google search index of "MeToo", and then estimate the MeToo effect among the OECD countries. In a separate analysis on the United States, they directly compare the change in sexual crimes before and after MeToo but do not take the potentially heterogeneous MeToo effect in different areas into consideration. Our study is different from theirs. First, we construct a novel sexism index to differentiate local sexist attitude which closely correlates with sexual crime, and then provide the first stage evidence to show that MeToo has drawn differential levels of attention across these areas. Second, instead of estimating the overall MeToo effect directly, our study concentrates on investigating the role of sexism on sexual crime by exploiting the MeToo shock. Our results show that areas with low sexism indices have documented relatively more sexual crimes than the those with high indices since MeToo. Specifically, during 2015 - 2019, across the 1,184 counties in the 133 in-sample media markets, one standard deviation decrease in sexism index translated into 0.58 more documented sexual crime per 100,000 populations or an equivalent 3% increase relative to the pre-MeToo period. This result is robust to the inclusion of a set of socio-economic covariates, the size of police, and a sequence of other controls. Analyses on heterogeneity show that, relative to the high sexism areas, the low sexism areas document more incidents committed by known offenders, incidents not causing injury, and incidents involving forcible sexual violence. We conduct extensive robustness checks and obtain quite consistent results.

The seemingly surprising surge in documented sexual crimes could be sourced from two potential channels. First, an intuitive explanation is that there is an increase in actual sexual crimes in low sexism areas. This could be caused either by coincidentally deteriorated overall public safety in these areas after MeToo or by backlash from offenders who are inimical to social movements promoting women's rights because they feel their traditional gender role is undermined by MeToo (retaliation effect). Second, the relative more incidents could be also attributed to the change in reporting behavior of victims residing in the low sexism areas. As sexual violence drew more heightened attention in the low sexism areas after MeToo, the overall atmosphere may become more sympathetic to sexual violence survivors, and the role-model effects from the exposure of many high-profile cases will further inspire victims to come forward (reporting effect). On the other hand, the increased likelihood of being reported and ensuing legal action on sexual misconducts would deter potential offenders from committing more crimes (deterrence effect), and hence decrease the number of actual incidents across the country. In the meanwhile, strengthened public attention to sexual violence would press local law enforcement to conduct timely investigation to each reported incident (policing effect), securing more effective arrests which will incapacitate offenders. Together, the net effect of sexism on sexual crime in the MeToo era depends on which effect would dominate.

Our empirical results lend evidence to the reporting effect, as estimation based on a nation-wide survey on crime victimization suggests that the likelihood of reporting rapes is nearly doubled after MeToo. Although we cannot entirely dismiss the possibility of the retaliation and deterrence effects, we do not find statistically meaningful variation in non-sexual crimes against women in the low sexism areas after MeToo. We additionally examine the change of sexual crimes in circumstances of homicide and aggravated assault in which reporting bias is minimal, and find little evidence of increase in the low sexism areas. For the policing effect, we do not find significant difference in sexual crime arrest rates between the low and high sexism areas after MeToo. Finally, we provide two anecdotal evidence to explain why victims in low sexism areas are more responsive to MeToo: States with lower sexism tend to have more laws and policies in favor of women's rights, and more employees working for civic organizations which appeal for human rights. The former will empower women economically and increase their bargaining power, therefore lowering the probability of being victimized by intimate partners (Aizer 2010). The latter will increase societal representation for women, which would lower the concerns about gender bias in the criminal justice system (Hoekstra and Street 2021) and encourage women victims to come forward (Iyer et al. 2012, Miller and Segal 2019).

The rest of the paper proceeds as follows. Section 2 briefly describes the MeToo movement and presents a conceptual framework of its potential effects. In Section 3 we describe the procedure of constructing sexism index by using the Google search index and PLS. Section 4 discusses the data source and identification strategy. Sections 5 and 6 present the estimation results and robustness checks. Section 7 provides empirical results for the mechanism analysis. We conclude in Section 8.

2. Background and Conceptual Framework

2.1 The MeToo Movement

The MeToo movement was sparked on October 15, 2017, when American actress Alyssa Milano posted a tweet and called her followers to post sexual violence they experienced with a “#MeToo” hashtag after the exposure of sexual assault allegations against Harvey Weinstein. This hashtag became viral immediately on social media and induced around 1.6 million posts on Twitter in the first week of the movement (October 14 - 21), instigating a massive social movement which leads to public disclosures of sexual misconducts (Modrek and Chakalov 2019). Within a year, more than 200 high-profile men were top- pled down from their positions (Carlsen et al. 2018). Its impact was quickly spread to other countries and drew international attention (Levy and Mattsson, 2022).

To illustrate how Weinstein’s allegation and the ensuing MeToo shifted public attention to sexual violence, we refer to the search data from Google Trends (GT), an online interface that provides unfiltered samples of search requests made to Google. The GT index captures the intensity of Google searches containing certain terms, and measures the relative popularity of these terms: Given a selected time frame and geographic regions, the GT index is calculated as the quotient of the number of searches for that term divided by the total searches, and is then normalized to be ranging between 0 and 100, where 100 is the most searches for that topic and 0 indicates that a given period does not have sufficient search volume for that term. Compared with survey data, GT data are less susceptible to small-sample bias and could elicit Google users’ behaviors, subtle feelings, and socially sensitive attitude. In economics studies, the GT data have already been widely used since the seminal work by (Stephens-Davidowitz 2014). See, for ex- ample, Baker and Baker and Fradkin (2017), Hall, Palsson, and Price

(2018), Siliverstovs and Wochner (2018), Brodeur et al. (2021), and Deza, Maclean, and Solomon (2022). In our analysis, we consider two groups of search terms. One is “ metoo”, and the other contains three terms: “ sexual harassment + sexual abuse + sexual violence”, where “+” represents “or” in GT. Then, we aggregate the collected indices into quarterly level and scale them from 0 to 1. Figure 3 sketches the two terms’ GT index trajectories during 2015Q1 - 2019Q4. As can be seen therein, the search intensity of “metoo” stayed at a high level in 2018, but gradually died down in 2019. The search of keywords related to sexual crime reached its peak in 2017Q4 after the onset of MeToo, and the elevated general awareness to sexual violence sustained at a level around 0.5, indicating a far-reaching impact caused by MeToo.

2.2 Conceptual Framework

Hostile sexism is an important perspective to understand the consequence of MeToo because local sexist attitude, as argued above, can partly explain the behavior of sexual crime victims in different areas, and their responses to MeToo could be largely disproportionate due to two possible reasons. For one, as demonstrated by Charles, Guryan, and Pan (2020), women’s choice is usually affected by the social norms that prevail where they reside. In the low sexism areas, the entrenched norms that negatively or stereotypically treat women would be further weakened after MeToo, empowering more women to come forward. On the other hand, hostility and stereotype against women could be intact in extremely high sexism areas even after MeToo, as offenders believe their prejudice which objectifies women can be justified, and victims are still unwilling to report due to the perceived high costs (Cheng and Hsiaw 2021). Second, as will be demonstrated later, anecdotal evidence shows that low sexism areas usually have more laws and policies in favor of women's rights, and more employees working in civic groups and community organizations which appeal for women's rights. These policies and organizations can provide supplementary protections and supports to women in these areas.

Conceptually, the MeToo shock may lead to heterogeneous responses from victims, offenders, and police in areas with differential sexism levels. Specifically, the first mechanism is a reporting effect which makes victims in the low sexism areas more willing to come forward if MeToo and the ensuing exposed high-profile incidents indeed draw more attention in these areas. Reporting of sexual crime is a high-stakes personal decision, and victims' choice would be partly determined by the perceived costs of reporting, the expecting sanctions to offenders, and the behaviors of other victims. If victims anticipate that law enforcement is not responsive, they would be reluctant to contact police because their reporting decision would incur retaliatory costs. This under-reporting behavior would be further perceived by other victims, thus corroborating the prevalence of under- reporting and causing a coordination problem (Cheng and Hsiaw 2021). Since MeToo was sparked, a sequence of high-profile incidents have been exposed, leading to firings and indictments of the involved offenders. The national outcry against sexual violence would lower the costs of reporting and mitigate the coordination problem, especially in areas where the awareness of the prevalence of sexual violence is intensified. Therefore, the reporting effect would encourage victims to come forward and lead to an increase in the number of documented sexual crimes in these areas even when the number of actual incidents does not considerably change.

The second mechanism is the net effect of the deterrence and retaliation effects on offenders. Due to the increased real and perceived costs of being reported and arrested, potential offenders, especially those in the low sexism areas and known to victims, would respond by committing fewer or no longer committing any sexual crimes, which will decrease the number of actual incidents in these areas. On the other hand, it is not entirely impossible that there is a retaliation effect which could result in an increase in sexual crimes against women for other reasons. For example, if potential offenders, predominantly men, take a resentful attitude toward MeToo which promotes gender equality and women's rights, they would target women and commit more crimes against women in response to the movement. Therefore, the net change in sexual crimes from the offender side depends on the relative strength of the two opposing effects: If deterrence exhibits larger effect than retaliation in an area, we should expect a decline in the number of actual incidents. Otherwise, the number of actual incidents should increase.

The third mechanism is through improved policing activities, or the policing effect. After MeToo, local police agencies might take a more sympathetic attitude toward sexual crime victims and become more responsive to reported incidents. This should be especially true for police in the low sexism areas as the heightened public attention and intensified awareness can generate more pressure than the high sexism counterparts. For example, some states passed bills dealing with rape kits so that police can test each reported incident in a timely manner (Beitsch 2018). In addition, a proactive investigation on the reported incident would lead to more effective arrests and higher clearance rate, which will further enhance the deterrence effect. Overall, the policing effect should incapacitate offenders from committing more crimes, and thereby decrease the number of actual incidents. Taken together, the overall impact of local sexism on sexual crime after MeToo depends on the relative strength of these effects, and is ambiguous ex ante.

3. Measure of Sexism

3.1 Proxies of Sexist Attitude

Similar to prejudice toward race and age, sexist attitude toward women is implicit and difficult to measure. In this study, we rely on the Google Trends (GT) data to evaluate the regional sexist attitude. As mentioned above, GT yields a sequence of indices representing the search intensity of certain terms, and allows its users to compare the proportion of searches among different regions within a selected time frame.[2] We collect the GT indices of terms expressing sexist attitude from two dictionaries. The first dictionary contains a list of 206 primarily female-referential derogatory terms collected by James (1998).[3] Since most terms on this list are regional slangs with multiple meanings, we examine the GT indices of these terms one by one at the state and media market (MM) level between 2011 -2016.[4] This step helps us narrow the list down to four sexist slurs ([Words] 1 - 4) displayed in Table 1, as all other terms either have zero index across the country or are searched in very few areas. Due to their sparsity in GT, we exclusively concentrate on these four sexist slurs. To alleviate the concern that a vast

[2] A zero GT index means that the index of a term is below a Google-determined threshold that is unobservable to researchers. For terms that finally become the predictors, we adopt the algorithm by Stephens-Davidowitz (2014) to address the issue of comparability. Briefly, we first get the GT index for the word "weather." Then, we get the GT index for "weather + word(s)," and the index of searches that include either "weather" or "word(s)." Subtracting the index of "weather" from the index of "weather + word(s)" will give us an estimate for "word(s)."

[3] These words were based on a questionnaire to 125 English-native speaking students at the University of Toronto in 1995.

[4] Media markets in the United States are compiled by Nielsen Media Research. The full list of the 210 media markets can be retrieved from https://www.nielsen.com/wp-content/uploads/sites/3/ 2019/09/2019-20-dma-ranker.pdf.

majority of these searches are those directed to pornographic materials, we search "[Word] − pornhub" which returns results including searches containing that specific word but excluding searches with pornhub, one of the most-trafficked adult website in the world.[5] The first row of Table 2 summarizes the four derogatory terms, including their plurals, whose GT indices will be finally used. Furthermore, we cannot entirely exclude the possibility that the four words sometimes are used as joking rather than insult. But we note that this occurs primarily between individuals of the same sex (Sutton 1995, Eble 1996), and searches of these words made by female searchers should be second-order at best compared with non-female searchers.

Second, besides the four aforementioned sexist slurs, sexist attitude toward women could be multidimensional and socially subtle. Felmlee, Inara Rodis, and Zhang (2020) analyze more than 2.9 million tweets that contain gendered insults, and categorize the hostile contents into insulting someone's appearance, intellect, sexual experience, mental stability, and age. These tweets shame women by accusing them of falling short of the standards in the five categories. To extract negative adjectives that are commonly used to refer to women from these five aspects, we build our second dictionary by making use of Describing Words (https://describingwords.io/), an engine built to retrieve adjectives which commonly describe a noun based on a corpus including literature text files of about 10 gigabytes, mostly fiction and contemporary works. Its parser crawls through each book, returning their descriptions of nouns, and ranking adjectives by their usage frequency for that noun. We respectively search "woman" and "girl" in this engine and then identify the top two most frequently used adjectives in each of the four above

---

[5] www.similarweb.com/website/pornhub.com

semantic categories: “fat”, “ugly”, “emotional”, “mad”, “stupid”, “dumb”, “dirty”, “easy”.[6] Then, we document all possible permutations between the eight adjectives and the two nouns (including the plurals) in panel (a) of Table 2, and collect the GT indices of these terms.

We need to note that there does not exist a one-to-one correspondence between a searcher’s sexist attitude and her propensity to search the derogatory terms in Google. In fact, searchers can search these terms for many reasons other than sexism. If searchers’ motivations of searching these terms are independent of sexism but differ across areas, it will lead to measurement error in our area level sexism index and cause bias in our future econometric analysis. According to Describing Words, the most frequently used adjective to describe a woman is “beautiful”, which can not only transmit positive or neutral meanings compared with the prior eight derogatory adjectives, but also imply negative sentiment which objectifies women. Panel (b) of Table 2 lists three alternative terms, including [Word 5] in case searchers in some areas may be more prone to use vulgar language in Google. These terms, denoted as Google controls, are not added to our sexism dictionary, but will be considered in empirical analysis in Section 5 to alleviate concerns about measurement error.

Furthermore, we would like to emphasize that although the GT index has several drawbacks such as failing to reveal the actual search volume in different areas and adjusting indices of terms with searching intensity below an unknown, pre-determined threshold to zero, these technical adjustments will not seriously hamper the comparability of these indices across areas, because the state- or MM-level index could still reflect the

[6] We exclude the semantic category of age because “old girl” is rarely used and “old woman” does not necessarily reflect a negative sentiment.

relative popularity of a specific term. As will be shown next, the sexism index constructed in this study hinges on the relative popularity of the derogatory terms in different areas rather than the absolute search volume.

Finally, we additionally use the differential shares of Trump voters across states and MMs in the 2016 election as another potential predictor of sexism index. These state- and MM-level Trump voter shares are aggregated based on the county level voter turnout data retrieved from the MIT Election Lab (https://electionlab.mit.edu/).

3.2 Construction of Sexism Index by Partial Least Squares

Although we have collected several proxies of sexist attitude toward women (the GT indices of terms in the two dictionaries, and differential shares of Trump voters), it is inappropriate to employ them as sexism index directly. For one, there is no guarantee that these proxies directly link to local sexism as they contain information on the regional heterogeneity of other aspects. Second, as displayed in Appendix Table A1, some proxies such as [Word1] and [Word3] are highly linearly correlated, and thus bring in noise and redundant information. Therefore, we need to separate information that directly explains sexism from noise. To this end, we adopt the partial least squares (PLS) technique in the machine learning literature, and make use of the state-level GSS sexism measures discussed in the introduction section again. To conserve space, we relegate a brief introduction of PLS estimation to Appendix B, and refer interested readers to Garthwaite (1994) and James et al. (2009) for a more complete introduction. Briefly speaking, PLS has three advantages that help us construct a low granular sexism index. First, unlike the widely used principle component analysis (PCA) technique which yields linear combinations of the underlying proxies, known as principle components, that largely explain the variation within these proxies, PLS first regresses the variable of interest on these underlying proxies, and then extracts relevant information that best explain the variation in the outcome variable. This makes PLS a supervised machine learning method and assures the usefulness of the generated new variables. Second, PLS is a dimension-reduction procedure that notably condenses the dimension of these sexism proxies which contains 13 features (8 sexist slurs (both singular and plural), 4 sexism adjectives plus nouns, and the differential of Trump voter shares), and yields three linearly uncorrelated new variables that are most strongly related to the GSS sexism index.[7] Third, using the PLS estimates obtained from the state-level proxies and GSS index, we can predict the MM-level sexism which is the key identification of our main analysis. Compared with the state-level index, the MM-level data provide lower granularity and therefore can better shed light on the heterogeneity of local sexist attitude.

We first implement PLS on the state-level data. From the 13 sexism proxies, the first three generated new variables yield the lowest predicting error (by cross-validation) and explain 60% of the variation in the state-level GSS sexism index. Then, using the estimates of the first three new variables, we can predict the MM-level sexism index by fitting the PLS estimates to the MM-level sexism proxies, and then standardize the index to have mean 0 and standard deviation 1. Figure 4 visualizes the constructed sexism index across MMs (excluding Fairbanks-AK, Juneau-AK, Honolulu-HI and Augusta-GA).[8] It exhibits two salient features. First, sexism tends to be lower in areas of the west coast and northeast, New Mexico, Colorado, as well as scattered urban areas such as Chicago, Detroit, Miami-Ft. Lauderdale, and Springfield-MO. On the other hand, the highest sex-

[7] The number of new variables is determined by cross validation. See Appendix B for details.
[8] For the ease of comparison, we normalize the standardized sexism index between 0 and 1 in Figure 4.

ism appears in the southern areas such as Tennessee, Arkansas, Kentucky, and Alabama. Second, the level of sexism can be enormously different across MMs even within the same state. For example, both Texas and Mississippi have low sexism areas (for example, Austin-TX and Meridian-MS) and extremely high ones (for example, Tyler-Longview-TX and Greenwood-Greenville-MS).

We note that there would be some concerns about the adoption of this unidimensional, time-invariant PLS sexism index. First, if the sexist attitude about women in certain areas radically changed after MeToo, a time-invariant measure using the GT data prior to 2017 would fail to reflect this. We investigate this possibility by additionally collecting the GT indices of the same set of derogatory terms from October 2017 to 2019, and then construct the principal components of these post-MeToo proxies by conducting PCA. Here, PCA is more appropriate because it does not require the outcome variable (GSS sexism index), and thus can better capture the variation among these proxies before and after MeToo. Appendix Figure A3 compares the first three principal components of the pre- and post-MeToo periods. It shows that all three pairs are highly correlated and most points surround the 45-degree lines, indicating similar implicit patterns of these predictors before and after MeToo. Second, a time-invariant sexism index can be endogenously determined by many other factors which would also impact reporting of sexual crimes. For example, local sexist attitude about women could be simultaneously determined by local residents' educational backgrounds which would affect the overall sexual crime rate in that area. We defer this discussion to the next section, where we will consider the potential threats to our identifying strategy.

3.3 Comparison with Alternative Sexism Measures

We compare the PLS sexism index with two alternative sexism measures. The first is the state-level GSS sexism index by Charles, Guryan, and Pan (2020) which directly reflects the internal sexist attitude toward women. The second measure of sexism exploits the size of protests which promote women's rights. We collect data on 2017 women's marches from Count Love (https://countlove.org/), an online database collecting data on protests happened in the United States since 2016. Using reports from local newspaper and television outlets, it provides information on the location of protest and the number of protesters. The women's marches on January 21, 2017 were among the largest in recent years, with 1 to 1.6 percent of the populations in the United States participated in at least 408 marches on that day (Broomfield 2017). These local women's marches promoted several causes on human rights, women's rights included, and their sizes could partly reflect local residents' attention to topics on women's rights and gender inequalities.[9] Figure 5 depicts the relations among the three sexism measures in 43 states.[10] In panel (a), the horizontal and vertical axes respectively denotes the PLS and GSS sexism indices, and the blue bubbles represent the states covered by both measures. The size of bubble indicates the number of protesters per 1,000 state populations in 2017 women's marches. Panel (a) has two implications. First, the two indices are positively correlated with correlation coefficient $\rho = 0.77$, and such a relation is confirmed by the fitted linear regression (red line) with slope 1 which is statistically significant at the 1% level (t-statistic = 7.68). Second, protests in the low sexism states tend to draw more participants

[9] Using the number of protesters to measure the intensity of a social movement has been used in past studies. For example, Acemoglu et al. (2018) use the number of protesters of Egypt's Arab Spring movement to measure the relationship between intensity of protest and stock market valuations. Premkumar (2020) uses the number of Black Lives Matter protesters to measure heterogeneous degrees of attention to racial inequalities among different areas.
[10] Besides the 7 states omitted by the GSS sexism index, Wyoming is also excluded due to some missing Google Trends indices. The three state-level sexism indices are listed in Appendix Table A2.

with correlation coefficient $\rho = -0.65$. In panel (b), we compare rankings of these states based on the two indices and find similar patterns: States ranked low in sexism by PLS tend to be still ranked as low sexism ones by GSS with slope 0.73 (t-statistic = 6.82), and the number of protesters is also relatively larger in low sexism states.

4. Data and Empirical Model

4.1 Data and Summary Statistics

Our primary data source is the National Incident-Based Reporting System (NIBRS) which is a part of FBI's UCR. Compared with standard UCR crime data, for each documented incident, the NIBRS links information on victims, offenders, and arrestees such as age, sex, and race, circumstances such as location, date, and relationship between victim and offender, and whether the incident causes injury. This allows us to study the heterogeneity by crime types. We focus on incidents involving sexual offenses, including forcible (rape, sodomy, sexual assault with an object, and fondling) and non-forcible (incest and statutory rape) types.[11] Another reason that the NIBRS is ideally suited for this study is because it reports the date of occurrence of an incident, whereas UCR data are monthly aggregated and only roughly reflect when an incident was reported, not necessarily the month in which it occurred. In addition, compared with other incident-based data such as the National Crime Victimization Survey (NCVS) which only provides very limited records on sexual crimes and is subject to recall bias, the NIBRS provides a relatively complete and accurate record on sexual crimes.

[11] In the NIBRS, a documented incident may include multiple offenses simultaneously. We classify an incident as sexual crime if the record contains one of these sexual offense categories.

An important drawback of the NIBRS is its participating rate. Although participation in the NIBRS has increased steadily since it was launched in 1991, its geographic scope remains limited. As of 2015, jurisdictions reporting to the NIBRS represent coverage of more than 96 million populations in the United States, which is equivalent to 36.1% of the populations residing in the jurisdictions reporting to UCR.[12] In this study, we concentrate on jurisdictions that consistently reported to the NIBRS in every month from 2015 to 2019. This gives us 3,464 constituent jurisdictions that serve 85 million populations in 1,184 counties locating in 133 of the 210 MMs.[13] Our final data aggregate the number of incidents at the county-by-quarter level.

Panel (a) of Table 3 documents the summary statistics of sexual and non-sexual crimes in the 1,184 counties during the pre- and post-MeToo periods. For sexual crimes, besides considering the types of offenders and victims, we additionally consider the number of incidents cleared by arrests and the number of incidents accompanied with homicides or aggravated assaults. For the ease of comparison, these summary statistics are displayed separately for counties locating in the low and high sexism MMs, corresponding to below and above the 50th percentile of the sexism index. This distinction will be useful because differential sexism index will be the key variable for the heterogeneity in sexual crimes among areas. Panel (a) shows that low sexism areas' sexual crime rate, measured by the number of documented sexual crimes per 100,000

[12] Summary of NIBRS, 2015, Retrieved from https://ucr.fbi.gov/nibrs/2015/resource-pages/nibrs-2015_summary_final-1.pdf

[13] Appendix Figure A1 displays the distribution of the 133 MMs along the sexism index, and compares the number of MMs with and without NIBRS-reporting jurisdictions in each bin of the sexism index. It shows that, although the summation of counts of high sexism MMs is slightly larger than that of low sexism ones, the 133 in-sample MMs cover almost every sexism level, with both extremely low and high sexism MMs included. Figure A2 further demonstrates that the 133 in-sample MMs are geographically representative, although several large cities in California and eastern areas are not included.

populations, does not remarkably differ from the high sexism areas' before MeToo, but becomes slightly higher after MeToo. Notably, more sexual crimes were cleared by arrests in the low sexism areas in both periods. It is also evident that non-sexual crime rate is significantly lower in the low sexism areas during the whole sample period. In Figure 6, we further compare the quarterly average sexual crime rates in the low and high sexism areas between 2015 and 2019. Prior to MeToo, sexual crime rates in the two groups are comparable, with the high sexism areas' sexual crime rate slightly higher than the low sexism counterparts'. However, immediately after MeToo, their relative position switched as the sexual crime rate in the low sexism areas outpaced that in the high sexism areas in 2018 and 2019, although both trended down to the pre-MeToo level at the end of 2019.

To account for the potential bias driven by the heterogeneity across counties, we additionally collect the yearly county-level demographic and socioeconomic data from the Census Bureau. Specifically, we collect demographic characteristics (percentage of female, non-Hispanic white, non-Hispanic black, and Hispanic), educational attainment (percentage of population with college or higher education), labor participation rate, personal income, and poverty rate. In addition, we calculate the county-level police size to control for any potential influence on sexual crimes caused by differential police size across counties.[14] In this study, a county's covariates are generally representative to describe its constituent jurisdictions reporting to the NIBRS, as populations residing in these NIBRS-reporting jurisdictions on average account for 80% of the total county

[14] The number of sworn officers is from the UCR Law Enforcement Officers Killed and Assaulted. To make the police size to be aligned with our sample, we only consider the 3,464 in-sample jurisdictions and define the police size as the number of sworn officers per 100,000 jurisdiction populations.

population in our sample, indicating that unrepresentative counties are rare in our analysis. We keep this in mind and will further investigate the sensitivity of our analysis by dropping a few unrepresentative counties in robustness checks.

Panel (b) of Table 3 displays the summary statistics of the censuses in the 1,184 counties before and after MeToo. As can be seen therein, the low sexism areas have significantly larger population, higher percentage of population with college or higher education, higher percentage of Hispanic, higher labor participation rate, higher personal income, lower poverty rate, and, notably, fewer police per 100,000 populations in both periods. It is important to note that these level differences in county characteristics do not necessarily pose a threat to our identification strategy in the next section, because our empirical strategy only requires similarity in changes in these potential sexual crime determinants between the low and high sexism areas. However, the significant differences in these county-level characteristics do highlight the importance of adopting a credible research design to address these pre-existing differences.

4.2 Empirical Model

Our empirical strategy is designed to examine how differential sexism affects the number of documented sexual crimes across areas after MeToo. This strategy can be generalized to a regression framework

$$\text{Outcome}_{cmyq} = \beta_1 \text{ SexismIndex}_m \times \text{Post}_{yq} + \mathbf{X}_{cy}\, \boldsymbol{\gamma} + \alpha_c + \tau_{yq} + \epsilon_{cmyq} \quad (1)$$

Here, $Outcome_{cmyq}$ is a measure of sexual crime in county c locating in media market m in quarter q of year y. $SexismIndex_m$ denotes the standardized PLS sexism index ranging from -3.89 to 3.63 at the MM-level unless instructed otherwise. $Post_{yq}$ is an indicator variable that equals 1 from 2017Q4 when MeToo became viral on internet. $\boldsymbol{X}_{cy}$

is a vector of county-level time-varying covariates, including the demographic and socioeconomic controls mentioned earlier. α_c is the county fixed effects, accounting for the pre-existing differences among counties with differential sexism levels that correlate with both sexual crime rate and local sexism. τ_{yq} is the year-by-quarter fixed effects and controls for temporal changes in sexual crimes in all counties. Standard errors are clustered by MM. Here, the coefficient of interest is β_1, which corresponds to the interaction between the *Post* indicator and the sexism index *SexismIndex*. Intuitively, this coefficient gauges the differential response to MeToo in counties locating in MMs with varying levels of sexism. Unless otherwise instructed, all regressions are weighted by county populations residing in the constituent jurisdictions consistently reporting to the NIBRS within the sample period.

4.3 Threats to Identification

The key identifying assumption of equation (1) is that the sexism index is uncorrelated with the error terms after conditioning on the county-level covariates, county fixed effects, and year-by-quarter fixed effects. However, there are several concerns about the validity of our empirical strategy.

The first threat is the bias sourced from certain omitted variables that would affect both local sexism and sexual crime. For time-invariant factors such as political stance which do not shift within a short time interval, the included county-level fixed effects can capture them. For time-varying factors such as passage of laws which impose heavier punishment on sexual offenders and supplementary public funds for rape kits, the inclusion of the MM-by-year fixed effects should address this concern. In addition, changes in sexual crimes could respond to some county-level confounding factors such as labor participation rate by female which might have also historically shaped the sexist attitude toward women in those areas. We investigate the importance of these observable confounding factors from two perspectives. First, we regress sexism index on covariates included in equation (1) to check whether the sexism index could be explained by them. Here, because all in-sample counties are more or less impacted by MeToo, therefore we are testing for balance across counties with differential treatment intensity in the MeToo era. Put differently, we will test the conditional independence assumption. Second, following the strategy in Altonji, Elder, and Taber (2005) and Deza, Maclean, and Solomon (2022), we progressively add these covariates to a baseline equation (1) which only contains the county and time fixed effects, and then examine whether the coefficient of interest changes.

Second, the number of documented sexual crimes in different areas might have diverged earlier before MeToo due to some other high-profile incidents which would also affect the behaviors of sexual offense victims. For example, in April 2017, the New York Times reported that O’Reilly and Fox News had legal settlements with at least six women who accused him of sexual misconducts dating back to 2002 with a total amount of nearly $50 million (Steel and Schmidt 2017). O’Reilly was fired soon after the report. His law- suits were coincidentally publicized in the Sexual Assault Awareness Month of 2017, and drew notable public attention. To explore the potential impact of other events, we will conduct event studies to check the pre-existing parallel trends among these areas.

We will formally address these threats to the validity of our identifying assumption in Section 5.3. We want to note that, irrespective of the extent to which we can interpret our result as causal, our analysis in this study provides evidence to the heterogeneous responses to MeToo among areas with differential local sexism.

5. Results

5.1 First Stage Results

We expect that differential local sexism will lead to heterogeneous responses to MeToo, as public attention drawn by this social movement could be different across areas. To study this first stage hypothesis, we resort to GT again by collecting the search indices of “ metoo” and “ sexual harassment + sexual abuse + sexual violence” during the sample period. The former represents the relative popularity of MeToo, and the latter is a proxy of public awareness to sexual violence.

To compare the search intensify of these terms between the low and high sexism areas as defined in Section 4.1, we attempt to sketch their GT index trajectories separately. The challenge of making the GT indices among these areas comparable over the sample period is that Google does not provide a comparable panel of the GT indices at the MM level, but only allows users to get either a time series GT index for one specific MM or a comparable cross-sectional GT index across all MMs over a customized period. To address this concern, we first aggregate the cross-sectional GT indices in each quarter separately for the low and high sexism MMs, as well as MMs not included in our sample. Next, we compute the proportion of each group’s aggregated index to the summation of all MMs’ indices. Using this proportion as the weight of that group, we approximate the GT indices of the low and high MMs in each quarter by using the quarterly aggregated time series GT index of the whole country. Finally, we adjust the estimated low and high sexism MMs' time series GT indices by dividing their respective highest values during 2015Q1 - 2019Q4.

Figure 7 illustrates the comparison of the two terms' GT indices between the low and high sexism MMs between 2015 and 2019. For "metoo", panel (a) shows that the GT index in the low sexism MMs is twice as large as that in the high sexism MMs in 2018, although the popularity died down quickly from 2019Q1. We find similar observation in panel (b) which shows that sexual violence drew uniformly more attention in the low sexism MMs than in the high sexism ones after MeToo. In summary, we perceive these findings as a first stage analysis which provides suggestive evidence that why the low and high sexism areas would potentially experience different patterns in sexual crimes in the wake of MeToo.

5.2 Results of Baseline Model

Columns (1) - (5) in Table 4 present estimates of equation (1) with five specifications. In column (1), we consider a parsimonious specification that estimates sexual crime rate per 100,000 populations with the county and year-by-quarter fixed effects only. The point estimate suggests that there was a large and highly significant increase in sexual crime rate in counties locating in MMs lower in sexism after MeToo. The point estimate of -0.536 implies that one standard deviation decrease in sexism index translated into 0.536 or 2.5% increase in sexual crime rate after MeToo. While suggestive, estimate in column (1) does not seem to be convincing because its interpretation hinges on the identification assumption that there are no omitted time-varying or area-specific factors that might explain the rise in sexual crimes in areas with lower sexism. Therefore, in columns (2) and (3), we test the robustness of the result by sequentially adding a full set of time-varying county-level covariates and the police size control. These controls do not meaningfully change the magnitude of the coefficient which is still significant at the 5% level. To further reduce measurement error in the sexism index, column (4) adds the sequence of Google controls discussed in Section 3. The point estimate only slightly changes to -0.576 and is still significant at the 5% level. Furthermore, column (5) shows that the inclusion of the MM-by-year time trend in addition to the full set of controls does not appreciably affect the magnitude or significance of the coefficient on sexism index.

5.3 Tests for Identification

In this section, we investigate the threats to the validity of our identifying assumption discussed in Section 4.3.

5.3.1 Determinants of Sexism Index

First, we assess whether the sexism index can be predicted by our county-level covariates. As noted, sexist attitude about women could be determined by some confounding factors which would simultaneously affect sexual crimes. In Table 5, we investigate the cross- county relationship between the PLS sexism index and a variety of potential determinants, including percentage of female, percentage of population with college or higher education, labor participation rate, logarithmic personal income, and poverty rate. As displayed in Table 3, these variables remarkably differ between the low and high sexism areas. These right-hand-side variables are all calculated from the average of census in 2015 and 2016. Estimates in columns (1) - (4) show that variables related to educational attainment, labor market, and income are all individually significant. Specifically, hostile sexist attitude toward women tend to be less prevailing in areas with larger well-educated populations, higher labor participation rate, and higher income. This result is in line with Charles et al. (2020) who find strong correlation between sexism and economic outcomes such as labor participation and wage gap. In column (5) when all variables are included, the three are still significant. Interestingly, percentage of population with college or higher education has the wrong sign in this regression, but the reason seems to be the collinearity between educational attainment and income.

5.3.2 Correlation or Sexism?

The significant correlation between sexism index and indicators of educational attainment, labor force participation and income raises a concern that whether our estimates in Table 4 are capturing the causal effect of sexism on documented sexual crimes after MeToo, or instead differential trends in sexual crimes in well-educated, high-income areas. We consider two ways to explore the issue of endogeneity. First, we follow Altonji, Elder, and Taber (2005) and Deza, Maclean, and Solomon (2022) and examine the sensitivity of our estimates to different sets of covariates. Specifically, we begin with estimating equation (1) with county and year-by-quarter fixed effects only. Then, we sequentially add in one covariate and re- estimate the model. Results in Table 6 show that the estimated coefficients are robust as we add more county-level covariates to the model, and the estimate is largely unchanged when compared with the baseline result in the first row. This result assures us that the estimated effect of sexism on sexual crimes after MeToo is robust to separate county-level covariates. The second way to purge sexism index of potentially endogenous variation is to control directly for these measures in equation (1), thus exploiting only the variation in sexism index sourced from other related dimensions of culture. To this end, we use the data from the 2012 - 2016 American Community Survey (ACS) 5-year sample and calculate the gender gaps in average schooling years, labor participation rate, and wage before MeToo was sparked.[15] Since the smallest identifiable geographic unit for ACS is the Public Use Microdata Areas (PUMAs), which only contain areas with populations larger than 100,000, we finally find 787 matched counties from our sample. Column (1) of Table 7 presents the estimate based on the 787 counties with the full set of controls added, and the result echoes its counterpart in column (4) of Table 4: One standard deviation decline in sexism leads to 0.695 or 3% increase in sexual crime rate after MeToo. Columns (2) -(4) respectively add an interaction term between the post-MeToo dummy and the gender gaps in average schooling years, labor participation rate, and wage. None is individually significant, and these variables have little impact on the sexism index coefficient, which remains highly significant. In column (5), we include all three interactions and find the estimate on sexism index is still quite comparable to the benchmark in column (1).

5.3.3 Parallel Trends

Finally, to examine the parallel trends assumption, we consider the following event study regression:

$$\text{Outcome}_{cmyq} = \Sigma\, \beta_{yq}\, \text{SexismIndex}_m \times \text{Time}_{yq} + \mathbf{X}_{cy}\, \boldsymbol{\gamma} + \alpha_c + \tau_{yq} + \epsilon_{cmyq} \quad (2)$$

where $Time_{yq}$ is a set of 19 indicator variables that take the value of 1 for the combinations of $y \in (2015, 2016, 2017, 2018, 2019)$ and $q \in (1, 2, 3, 4)$, with $Time_{2017,3}$, the immediate period before MeToo, is left as the comparison group. Each coefficient β_{yq} can be interpreted as an estimate of the impact of sexism index on sexual crime rate in quarter q of year y. Thus, it is a generalization of equation (1) to estimate the quarter-by-

[15] The wage gap is defined as the difference in wages between women and men divided by the average wage, and the education gap is defined as the difference in schooling years between women and men divided by average years of schools.

quarter contrasts. If low sexism areas indeed witnessed a jump in documented incidents in the wake of MeToo, we should expect that the coefficients βyq are not significantly different from 0 until $\beta_{2017,4}$.

The 19 point estimates are plotted in Figure 8 to provide a visual summary, with the blue bars denote the 95% confidence intervals and standard errors are still clustered by MM. We do not find evidence of pre-trends, as most estimates before 2017Q4 are close to zero and none is significant at any conventional levels. After 2017Q3, all slope coefficients become negative and most are significant at the 5% level, indicating more documented sexual crimes in the low sexism areas than in the high sexism ones in 2018 and 2019.

5.4 Heterogeneous Effects

Thus far, we consider sexual crimes committed by all types of offenders. Exploiting the rich information in the NIBRS, we conduct several analyses on the heterogeneity by the offender and victim types.

First, we compare the effect on sexual crimes committed by offenders known or unknown to victims. We classify known and unknown offenders by relationship between victim and offender. Known offenders indicate that victims are acquaintances, neighbors, employees, employers, friends, family members, or otherwise known, while unknown offenders mean that victims are strangers or the relationship is unknown. Estimates of the two offender types are presented in panel (a) of Table 8. Columns (1) - (4) display the effect on sexual crimes committed by known offenders. In our preferred specification column (4), the coefficient estimate is -0.345, and is significant at the 5% level. This indicates that as the sexism index decreases by one standard deviation, sexual crime rate will increase by about 0.345 relative to the pre-MeToo era. On the other hand, estimates of effect on sexual crimes committed by unknown offenders, presented in columns (5)-(8), are only weakly significant at the 10% level in all four specifications, indicating suggestive evidence of increase in sexual crimes committed by unknown offenders in low sexism areas after MeToo. Notably, 78% of sexual crimes in the United States during 2005 - 2010 involved an offender who is a family member, intimate partner, friend, or acquaintance (Planty et al. 2013). When sexually abused by known offenders, victimsmight be hesitant to report due to the concerns of stigmatization and retaliation which will increase their own costs of reporting, especially when the offenders are intimate partners or supervisors to whom victims remain committed. Our estimates here suggest that low sexism leads to an increase in reporting known offenders in the MeToo era, whereas reporting of unknown offenders seems to be less susceptible to local sexism after MeToo. Second, we compare the effect on incidents with and without injury. Unlike severe sexual crimes such as rape which will cause serious physical injuries to victims, sexual misconducts such as verbal abuse and threat usually do not directly cause visible physical injury but less noticeable psychological and mental traumas. Therefore, some victims may decide not to report due to the lack of solid evidence. For non-injury incidents, estimates in columns (1) - (4) of panel (b) show that, relative to the pre-MeToo era, one standard deviation decrease in sexism index leads to an approximate 0.51 increase in sexual crime rate. The estimates are significant at the 5% level in all four specifications. However, we do not observe any statistically meaningful change in sexual crimes with injury, as shown in columns (5) - (8). The result suggest that the overall effect of low sexism is only significant on sexual crimes without injury after MeToo.

Next, we group sexual crimes into forcible and non-forcible ones. According to the NIBRS, forcible sexual crimes include forcible rape, forcible sodomy, sexual assault with an object, and forcible fondling. Non-forcible sexual crimes include incest and statutory rape. For the former, estimates in columns (1) - (4) of panel (c) imply that forcible sexual crime rate increases by 0.547 as the sexism index becomes one standard deviation lower after MeToo, and all estimates are significant at the 5% or better levels. Columns (5)-(8) display the effects on non-forcible incidents. It appears that the estimates in all four specifications are quite small and none is significant at any conventional levels.

In panels (d) - (f), we respectively separate victims into three different pairs. In panel (d), we group victims by residential status. Such a classification is sensible because local prevailing sexist attitude should generate larger impact on local residents who internalize these social norms which might be remarkably different from non-residents'.[16] Comparing the effect of sexism on resident and non-resident victims before and after MeToo, we observe that all coefficients are significant at the 5% level for resident victims, as shown in columns (1) - (4). The effect on non-resident victims, on the other hand, is insignificant at any conventional levels. These estimates suggest that the overall effect is mainly from resident victims. Estimates in panel (e) are all significant at the 5% level, suggesting that not only incidents with female victims increased in low sexism areas, incidents with male victims also increased in these areas after MeToo. The former is expected because sexual crime victims are predominantly women. For male victims, in this analysis more than 70% are under 16 years old, indicating that sexual crimes against child and teenager were more likely to be reported in low sexism areas after MeToo.

---

[16] Charles et al. (2020) introduce distinction between "background sexism" and "residential sexism" and argue that the former has lasting influence and is internalized when an adult is young.

Finally, in panel (f), we consider victims' races and examine the effect on white and black victims.[17] Columns (1) - (4) document a significant effect on incidents with white victims in low sexism areas after MeToo. For incidents with black victims, we only find suggestive evidence of the impact after MeToo, as all estimates in columns (5) - (8) are only marginally significant at the 10% level.

5.5 Falsification Tests

An alternative way to examine our identifying assumption is to check whether crimes against women that not ought to be affected by the local sexism were actually affected or not after MeToo. In other words, finding statistically significant effect on crimes that ought to be exogenous to sexism would invalidate our research design. Specifically, we investigate whether non-sexual UCR index crimes (homicide, robbery, aggravated assault, burglary, larceny theft, auto theft, and arson) against women meaningfully changed after MeToo across these areas. While some of these crimes such as burglary and robbery may accompany with sexual offense and are thus possibly affected by MeToo, we argue that such an effect would be second-order, at best.

The estimated effect on non-sexual crimes against women is shown in Table 9. All estimates are positive, indicating that non-sexual crimes against women decreased in the low sexism areas after MeToo. But the effect is imprecisely estimated and none is significant at any conventional levels. Results of this falsification test reassure us that the estimates documented in Table 4 are capturing the effect of local sexism after MeToo rather than the effect confounded by either concurrent policies or an overall increase in crimes against women in the low sexism areas.

[17] White and black victims together account for 93% of total sexual offense victims.

Next, to ensure that we are making a correct inference about statistical significance of the main finding, we conduct a test in the spirit of Bertrand et al. (2004), in which we permute the sexism indices among the 133 MMs. Then, we re-run equation (1) to estimate the coefficient of the interaction term. By conducting this procedure repeatedly, we generate a distribution of estimates and ask what proportion of the placebo estimates are lower than the actual estimated effect, or the empirical p-value of our estimate.

The resulting distribution of $\beta 1$ estimates based on 2,000 repeated sampling is illustrated by Figure 9. It corresponds to column (4) of Table 4, which is the preferred specification. It shows our estimate of -0.576 in Table 4 ranks below the 4th percentile, as only 71 placebo estimates are lower than the actual estimate in the 2,000 permutations.

6. Robustness Checks

6.1 Alternative Measures of Sexism and MeToo Awareness

Thus far, our discussions exclusively rely on the PLS sexism index to represent hostile attitude toward women in different areas. We also use Weinstein's allegation as the exogenous shock to separate the pre- and post-MeToo eras. To test the robustness of our finding based on this identification strategy, we consider several alternative sexism measures and MeToo awareness to construct new interaction terms in equation (1).

First, in panel (a) of Table 10, we substitute the continuous MM-level PLS sexism index with an indicator variable indicative of MMs with sexism indices below the 50th percentile, or low sexism MMs. This makes our identification the classic difference-in-differences design, and its estimate is in line with our main finding so far. In our preferred specification column (4), the point estimate is significant at the 5% level and indicates that sexual crime rate of low sexism counties increased by 1.191, or about 5%, relative to those in high sexism counties after MeToo.

Second, we interact the continuous MM-level PLS sexism index with the quarterly aggregated MeToo awareness index (*AwarenessSH*) which is proxied by the GT index of keywords related to sexual harassment. As displayed by Figure 3, there is a clear uptick in the awareness of sexual violence across the country since 2017Q4. All specifications in panel (b) exhibit significant and negative estimates which are still in line with the main finding. Specifically, the estimate in our preferred specification column (4) implies that one standard deviation decrease in the sexism index increases the sexual crime rate by 0.358 after MeToo when the awareness index is 0.428, the full sample average.

In the next two panels, we alternatively experiment with the two state-level sexism measures discussed in Section 3. In panel (c), we adopt the GSS sexism index in Charles et al. (2020) directly. Results in all columns are negative and significant at the 5% level. For our preferred specification in column (4), after accounting for the full set of controls, the estimate suggests that one standard deviation decrease in the GSS sexism index translated into 0.648 increase in sexual crime rate after MeToo. In panel (d), we use the number of protesters of the 2017 women's marches as the proxy of sexism in each state. All estimates are positive, and the result in our preferred specification column (4) suggests that an additional protester per 1,000 populations leads to 0.309 or 1.4% increase in sexual crime rate after MeToo.

6.2 Dropping Some Counties

As mentioned above, one drawback of the NIBRS is that it covers only about 30% populations in the United States, and many jurisdictions, cities with large populations in particular, have not participated in the program over the sample period. Readers may be concerned that many counties in this analysis are unrepresentative ones with only spotty coverage in the NIBRS, thereby the county-level covariates used above do not accurately reflect the actual socioeconomic status of these reporting jurisdictions.

We first test the sensitivity of our results by dropping counties whose populations residing in the constituent jurisdictions consistently reporting to the NIBRS are only a small fraction of the total population in those counties. Here, we define this fraction as population coverage rate. For example, we start with dropping counties whose population coverage rate is lower than 10%. This will lead to a removal of 23 constituent counties from the original 1,184 counties. Then we re-estimate equation (1) and obtain the estimate of $\beta 1$. We repeat this procedure by loosening the coverage rate from 10% to 90% so that more counties will be dropped from our sample. Figure 10 presents these estimates as well as their respective 95% confidence intervals, with the estimate in column (4) of Table 4 used as the reference (point and interval at the top). It shows the point estimates are robust to the omission of those counties, as their magnitude is largely unchanged and all is still significant at the 5% level.

In an additional robustness check, we examine the sensitivity of our main result by randomly dropping MMs. Specifically, we replicate the estimation of equation (1) and explore whether dropping all constituent counties in two randomly selected MMs leads to a different result. We exhaustively repeat this procedure 8,778 times ($C^{2}_{133} = 8778$) and then plot the empirical distribution of $\beta 1$ estimates in Figure 11. It shows the distribution is densely centered around -0.576, the point estimate in column (4) of Table 4, with its upper and lower bounds lie in the same direction as the estimate from our preferred specification. Overall, this result suggests that our main finding is unlikely to be severely biased by sexual crimes in a handful of counties which might have unusual characteristics.

7. Discussion

7.1 Evidence of Reporting Effect

As outlined, there are three plausible mechanisms explaining why low and high sexism areas respond to MeToo heterogeneously: The behavioral changes of victims (reporting effect), offenders (deterrence and retaliation effects), and police (policing effect). While it is likely that all three effects are operating, the reporting effect is of particular interest from policy makers' perspective considering the severe under-reporting of sexual crimes. Unlike the deterrence effect which could be realized through enhanced criminal justice such as Truth-in-Sentencing which incapacitates offenders from committing more crimes, greater willingness of reporting would not be easily achieved as long as the rape myths that stereotype women victims still prevail.

We explore the reporting effect through both direct and indirect ways. To directly examine the willingness of reporting sexual crimes, we consider the National Crime Victimization Survey (NCVS) which provides data of crime incidents on a nationally representative sample of approximately 49,000 to 77,400 households in the United States twice a year. Survey participants are asked some screening questions for possible crimes, and positive responses will be followed by additional questions, including report to police or not. We extract all incidents related to sexual offense happened between 2015 and 2019, and then estimate the linear probability model, logit model, and probit model with a binary dependent variable which equals 1 if a victim reported to police.[18] Due to the sparsity of recorded sexual crimes in the NCVS (about 200 incidents per year), in this analysis we concentrate on the coefficient of the indicator variable Post defined as be- fore, and add region dummies (Northeast, Midwest, South, and West) to these models.

Therefore, the coefficient on Post simply measures the change in probability of reporting across the country. Columns (1), (3) and (5) in Table 11 document estimates of the three models with region dummies as well as dummies of incident years (2014 - 2019) and inter- view waves (2015Q1 - 2019Q4). All estimates are significant at the 1% level and suggest the probability of reporting increased by about 15 - 30 percentage points after MeToo. Estimates in columns (2), (4) and (6) are still significant at the 1% level when both incident and victim controls are added, and indicate an increase of 24 - 46 percentage points in the probability of reporting after MeToo.[19] Compared with the pre-MeToo reporting rate, which is about 32%, the proportion of reporting has been doubled throughout the country after 2017Q4, although we want to emphasize that such a sizable estimated effect should be interpreted as suggestive evidence of the reporting effect because it only reflects rape incidents that would be reported in the NCVS.[20]

Next, we indirectly examine the reporting effect by investigating whether the net effect of the deterrence and retaliation effects are heterogeneous across areas with differential sexism after MeToo. This is crucial because if this were the case, the

---

[18] Incidents related to sexual offense in the NCVS include completed rape, attempted rape, sexual attack with serious assault, sexual attack with minor assault, sexual assault without injury, unwanted sexual contact without force, verbal threat of rape, and verbal threat of sexual assault.
[19] Incident controls include well known to victim, attempt/threat to rape, injury, single offender, forced or coerced unwanted sex, and in city of (S)MSA. Victim controls include age, schooling years, married, male, white, black, Asian, and Hispanic.
[20] 21Surveys such as NCVS still have under-reporting issues and scholars have noted the lower implied incidence rates of violence against women in the NCVS (see Tjaden and Thoennes, 2006, and Miller and Segal, 2019). We are unable to address this source of under-reporting in our sample.

estimated results should be also partly attributed to the behavioral changes of offenders between the low and high sexism areas. To this end, we focus on two types of sexual crimes which are less likely to be affected by the reporting effect: Homicide and aggravated assault of women under sexual crime circumstance. Homicide is arguably the least likely under-reported crime type, and aggravated assault usually leads to serious physical injury which needs immediate medical treatment. The first two panels in Table 12 present the results about homicide, with panel (a) considers homicide rate per 10 million populations and panel (b) uses count of homicide and zero-inflated Poisson regression to account for the excessive zeros. All estimates are insignificant at any conventional levels.[21] Since homicide is rare, panel (c) additionally checks aggravated assault and finds similar result. In summary, there does not seem to be a significant difference in sexual crime types that are less susceptible to the reporting effect between the low and high sexism areas after MeToo.

Finally, we evaluate the policing effect by examining the change in arrest rates which is defined as the percentage of incidents that are cleared by arrests out of the total number of sexual crime incidents known by police. In the United States, only 31% of the reported sexual crimes resulted in arrests during 2005 - 2010, and the overall conviction rate per 1,000 sexual crimes that reported to police - cases that end up with a felony conviction and incarceration - was only 0.02% between 2012 and 2016.[22]23 Thus, an increase in arrest rate would provide suggestive evidence of improved policing activity

---

[21] We also examine the data from UCR’s Supplementary Homicide Reports (SHRs) and find quite similar results. These results are available upon request.

[22] The estimates are from the Rape, Abuse & Incest National Network (RAINN). It is one of the largest anti-sexual violence organization in the United States and combines information from several federal government reports such as National Crime Victimization Survey and National Incident-Based Reporting System. Because it combines data from studies with different methodologies, it is an approximation, not a scientific estimate. For more details, please refer to https://www.rainn.org/statistics/ criminal-justice-system

and thereby enhanced policing effect. However, all four coefficients in panel (d) of Table 12 are imprecisely estimated, indicating an insignificant difference in arrest rates between the low and high sexism areas after MeToo.

Above all, although we cannot directly compare the willingness of reporting between the low and high sexism areas after MeToo, direct and indirect evidence suggests that there is a solid increase in the propensity of reporting throughout the country after MeToo.

7.2 Why Low Sexism Areas Are More Responsive?

Since our current empirical evidence suggests that the surge of incidents in the low sexism areas is primarily caused by reporting, here we provide two anecdotal observations that may shed some light on why victims in these areas tend to be more responsive after MeToo. First, low sexism states usually have broad laws and policies to protect women's rights and support families. Some of these laws, such as paid family leave and maternity leave, protect not only women's physical health, but also their economic status. For example, the paid family leave acts can lower job discontinuity of female workers Bertrand, Goldin, and Katz (2010) and increase female leadership (Bennett et al. 2020), and economically empowered women have been proved to be less likely victimized by domestic violence due to improved bargaining power (Aizer 2010, Anderberg et al. 2016). To investigate the relation between sexism index and policies related to family or medical leave at the state level, we use an evaluation of family-friendly policies reported by the National Partnership, a nonprofit, non-partisan organization which promotes fairness in workplace and quality health care. In the 2016 report, the National Partnership exhaustively searches " state websites, statutes, regulations and state government personnel handbooks to identify state laws and policies that guarantee access to family or medical leave to expecting and new parents, paid sick days, reasonable accommodations for pregnant workers and support for breastfeeding mothers" (Ness, Shabo, and Fink 2016), then awards each states scores based on the protections provided to private sector and state employees. Panel (a) of Appendix Figure A4 draws the scatter plot of the sexism index and the awarded scores of the 44 states covered by the GSS. As can be seen therein, there exists a clear negative correlation: States with lower sexism indices tend to earn higher scores. The red line denotes the estimated regression line, with the slope coefficient -12.11 which is significant at the 5% level (t-statistic = -2.02).

Second, victims in low sexism states can have greater representation in human rights organizations and more formal supports. Greater representation of disadvantage groups in elected officer or criminal justice system can alleviate concerns about gender bias (Hoekstra and Street 2021) and encourage women victims to come forward (Iyer et al. 2012, Miller and Segal 2019). Civic groups and community organizations promoting women's rights can provide assistance to women victims, decrease reporting costs by reducing stigma, and facilitate contact with law enforcement. Panel (b) draws the scatter plot of the GSS sexism index and the number of human rights organization employees (NAICS code=813311) per 1,000 populations in these states. It indicates that the two indicators are also negatively correlated, with the slope coefficient -0.04 which is significant at the 5% level (t-statistic=-2.16).

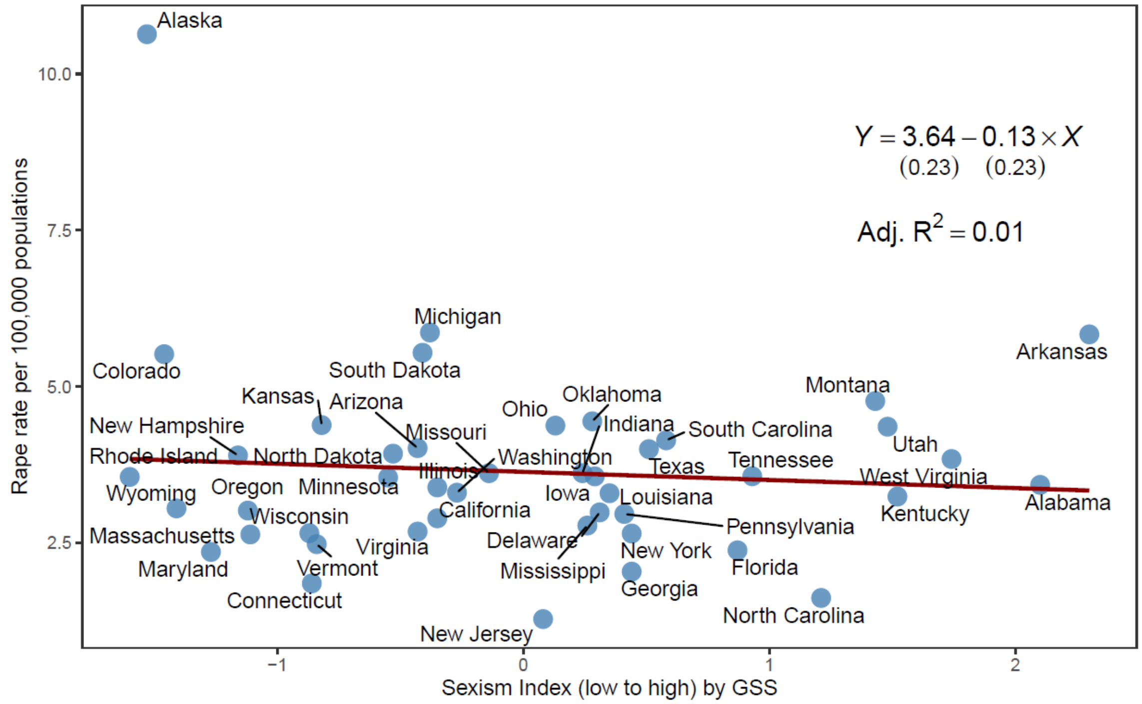
8. Conclusion

MeToo is a sweeping social movement which exposes many high-profile sexual misconducts committed by prominent men. Since sparked in 2017, this movement has shifted public attention to sexual violence survivors and induced heated discussions on gender discrimination and women's rights. Constructing a novel media market level measure on sexism, this study sheds light on how local sexism affected sexual crimes across areas in the United States in the MeToo era. In particular, we find that low sexism areas witnessed higher sexual crime rate than high sexism areas after MeToo. We investigate the possible channels through which local sexism would affect sexual crime rate, and demonstrate that the relative more incidents documented in these areas should be attributed to reporting rather than an increase in actual crimes.

We believe that our finding in this study can contribute to the current debates on MeToo. Sexual violence and its under-reporting have been a social problem in the United States for a long time, and MeToo reveals how prevailing sexual harassment is and how traumatic its consequence can be to survivors, predominantly women. Our study confirms that MeToo has indeed empowered women to come forward, but mainly those living in a less hostile environment. Although it would be difficult to reverse such a hostile social norm in the high sexism areas within a short time, we can empower women through certain concrete actions such as supplementary legal assistance.

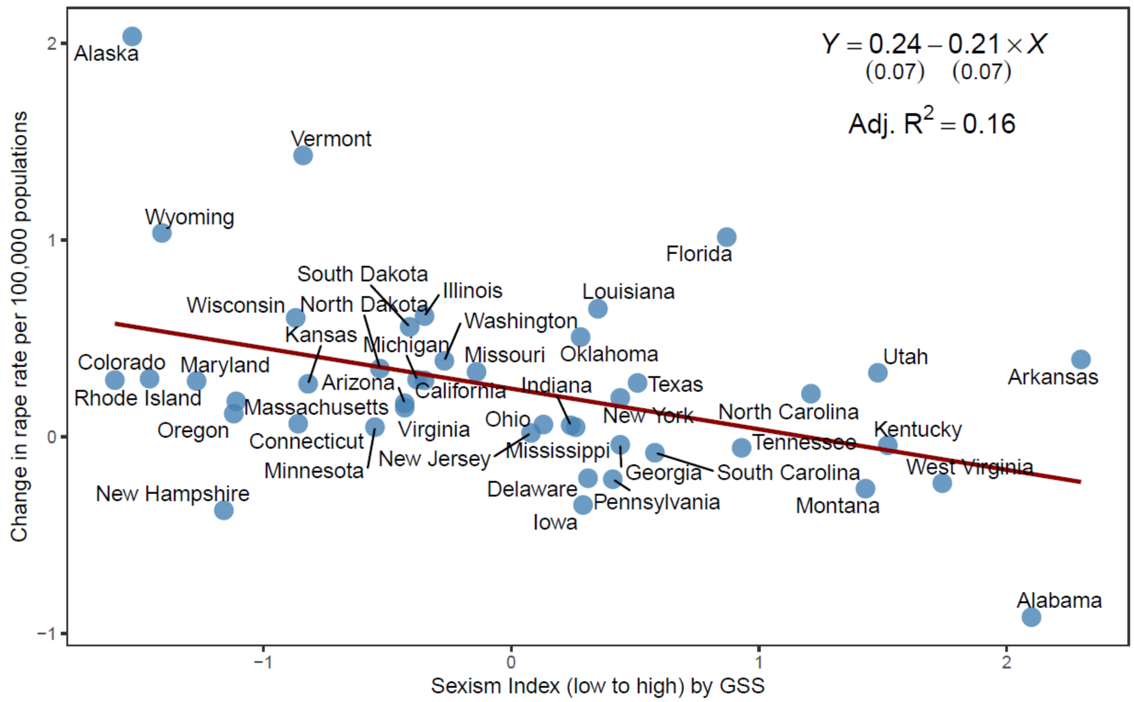
The conclusions in this study, of course, are subject to a few caveats. For one, we cannot completely exclude the possibility that the majority of the more documented incidents is from police recording. Besides designating a complaint as unfounded, police can also intentionally misclassify an incident as a less severe offense or, under certain extreme circumstances, ignore a complaint without any written record. If so, the relative more incidents in low sexism areas should be primarily attributed to the change in recording behaviors of police after MeToo. Second, although sexual harassment in workplace is the focal point of MeToo, it is still severely under-reported due the lack of evidence, and is not clearly defined in the NIBRS data. Finally, police agencies in many large cities on the east and west coasts still do not report to the NIBRS. A comprehensive analysis requires filing information requests to these local police agencies individually to collect data on sexual offenses, and we leave this study to future research.

Figure 1 State sexism index and monthly mean rape rates (2015 - 2017)



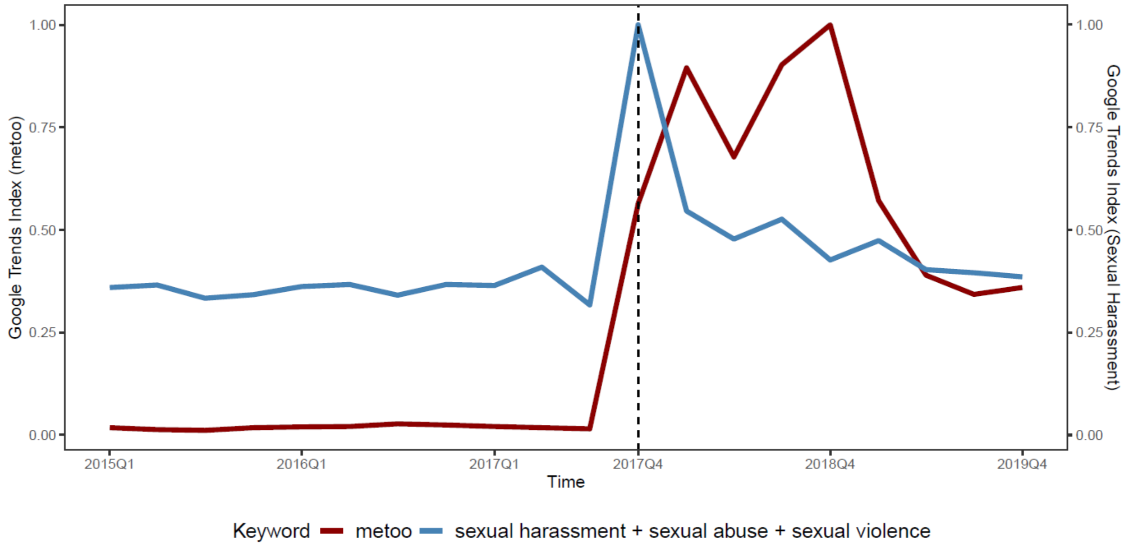
Notes. This figure plots the state sexism index constructed by Charles et al. (2020) and change in the UCR monthly average rape rate per 100,000 populations before (Jan 2015 - Sep 2017) and after MeToo (Oct 2017 - Dec 2019). The red line denotes the regression line. The sexism indices in District of Columbia, Hawaii, Idaho, Maine, Nebraska, Nevada and New Mexico are not available.

Figure 2 State sexism index and change in mean rape rates before and after MeToo



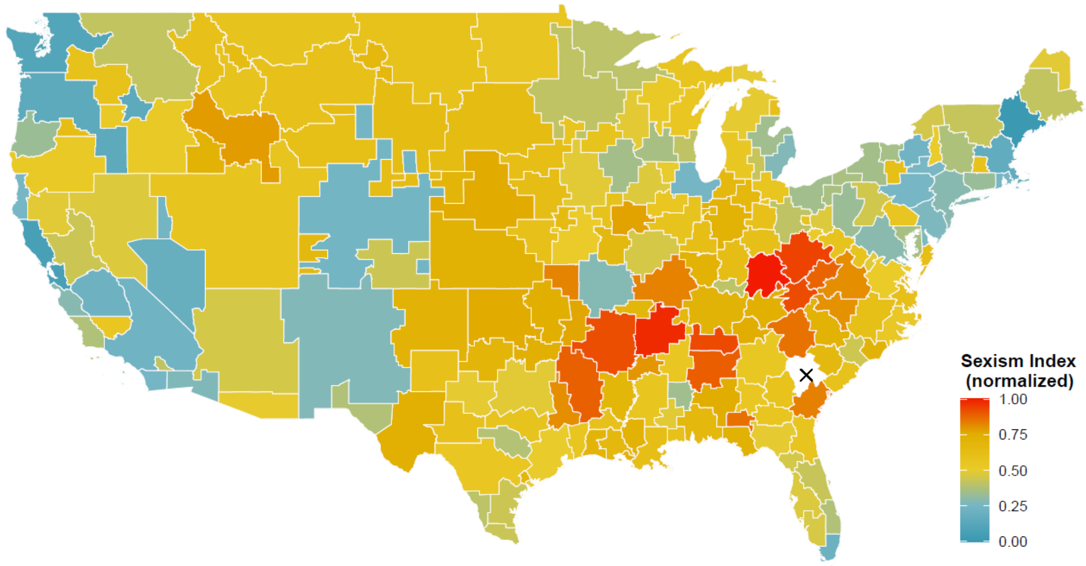
Notes. This figure plots the state sexism index constructed by Charles et al. (2020) and change in the UCR monthly average rape rate per 100,000 populations before (Jan 2015 - Sep 2017) and after MeToo (Oct 2017 - Dec 2019). The red line denotes the regression line. The sexism indices in District of Columbia, Hawaii, Idaho, Maine, Nebraska, Nevada and New Mexico are not available.

Figure 3 Google Trends indices of "metoo" and keywords related to sexual harassment



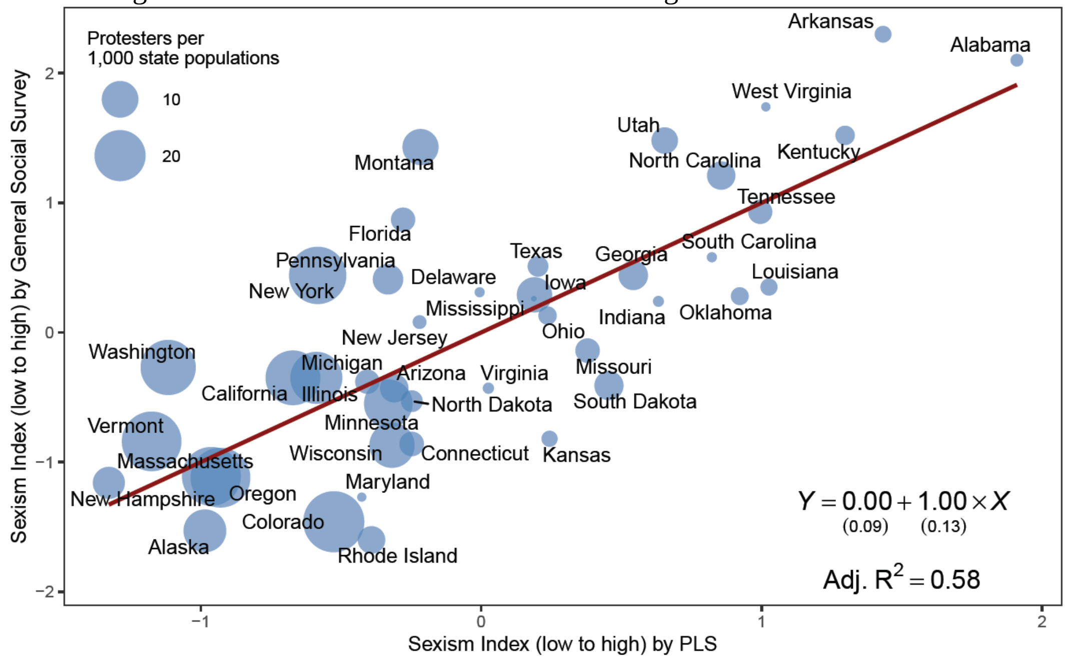
Notes. This figure plots the quarterly aggregated Google Trends indices for keyword "metoo" (red) and keywords "sexual harassment + sexual abuse + sexual violence" (blue) during 2015 - 2019. The vertical dashed line denotes 2017Q4 when the MeToo movement was sparked.

Figure 4 Media market-level sexism index in the United States

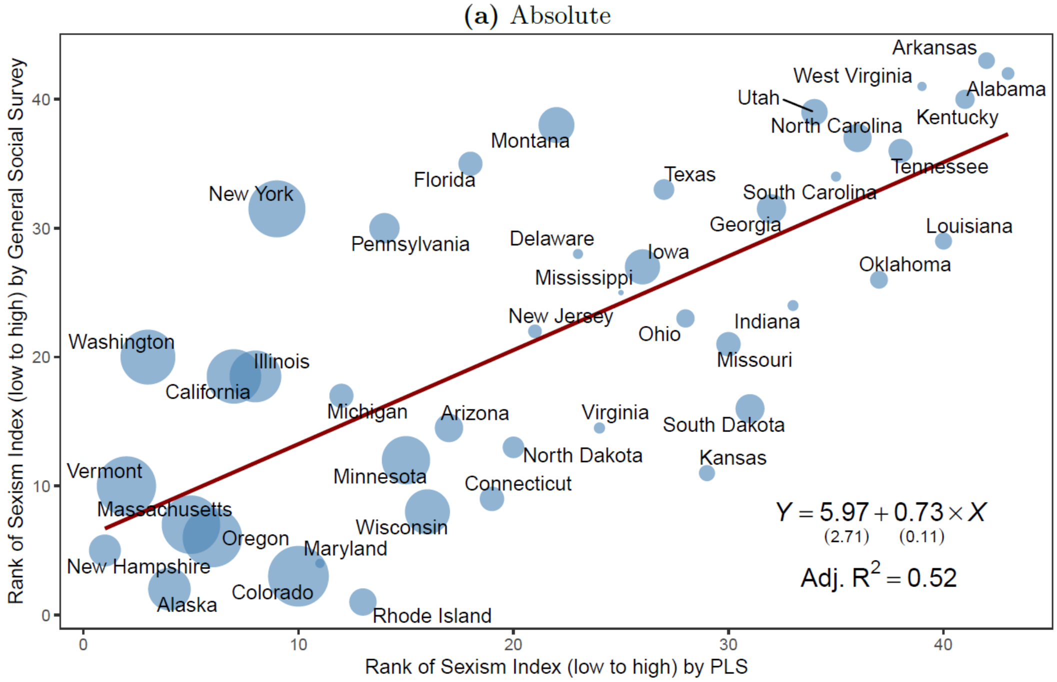


Notes. This figure plots the media market-level sexism index using partial least squares (PLS) discussed in Section 3. For the ease of comparison, the index is normalized between 0 and 1. Alaska and Hawaii are not displayed. Index in Augusta-GA (denoted by "×") is not available due to missing values in the Google Trends index. Fairbanks-AK, Juneau-AK and Honolulu-HI are not displayed.

Figure 5 Visualization of the correlations among three sexism measures



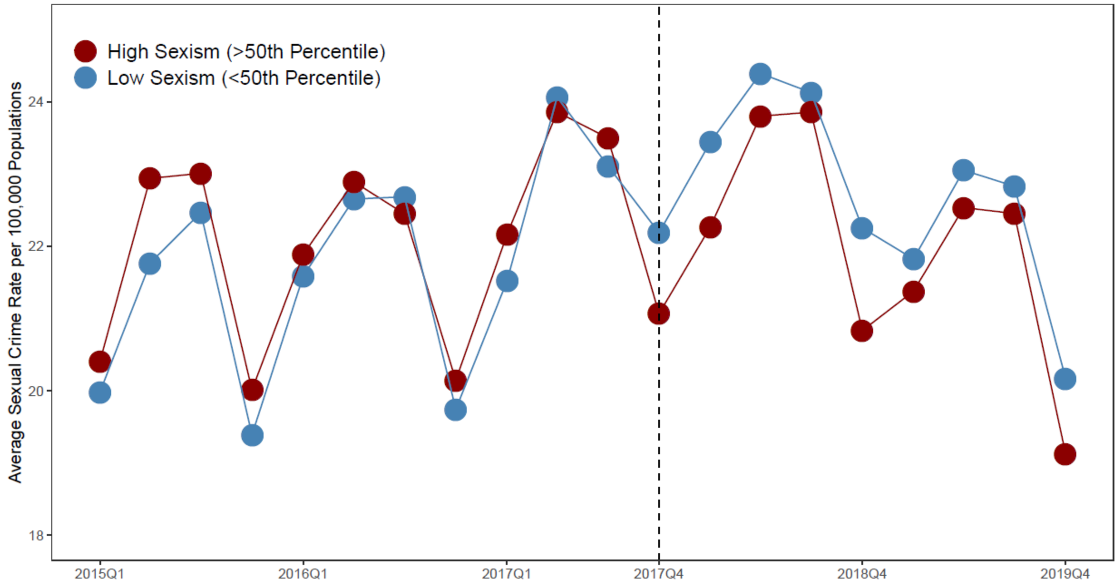
(a) Absolute



(b) Rank

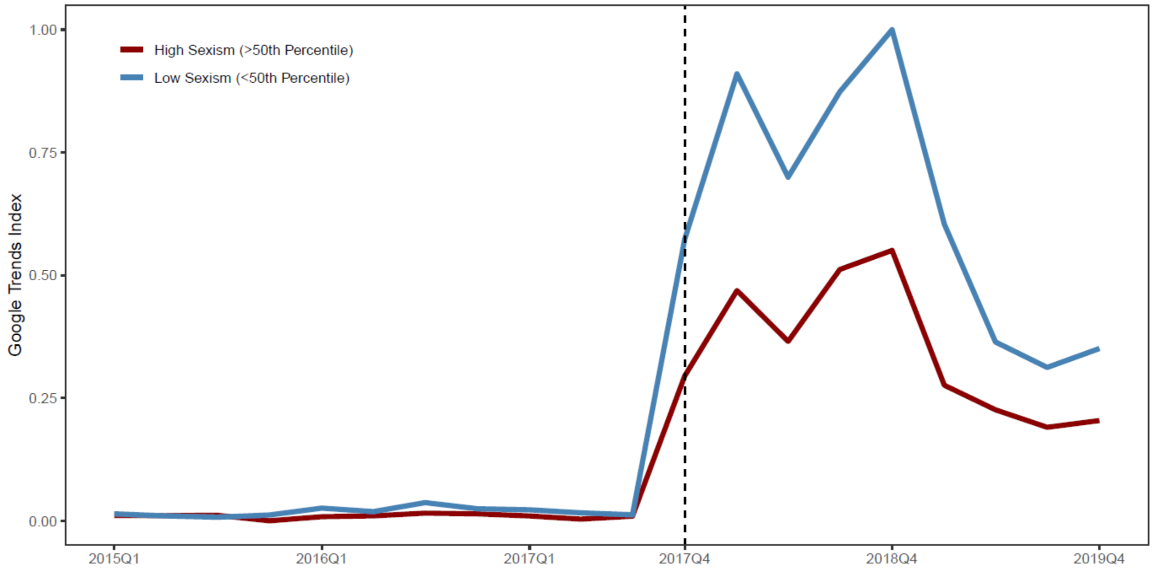
Notes. This figure plots three state-level sexism measures: PLS sexism, GSS sexism, and the number of protesters per 1,000 populations in 2017 women's marches. Panels (a) uses the computed sexism indices, and panels (b) uses the ranks of sexism indices. The red lines in each panel denote the regression lines. The size of blue bubble denotes the number of protesters per 1,000 populations in 2017 women's marches.

Figure 6 Comparison of sexual crime rates in low and high sexism areas

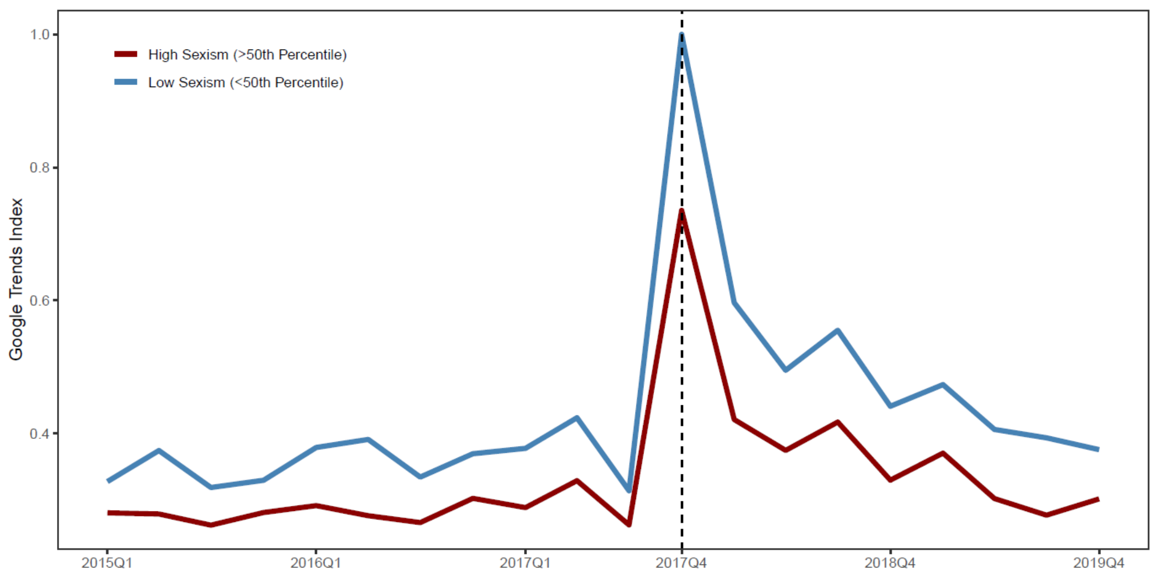


Notes. This figure plots quarterly average sexual crime rates in the low sexism (below the 50th percentile of sexism index) and high sexism (above the 50th percentile of sexism index) media markets.

Figure 7 Comparison of Google Trends indices between low and high sexism areas



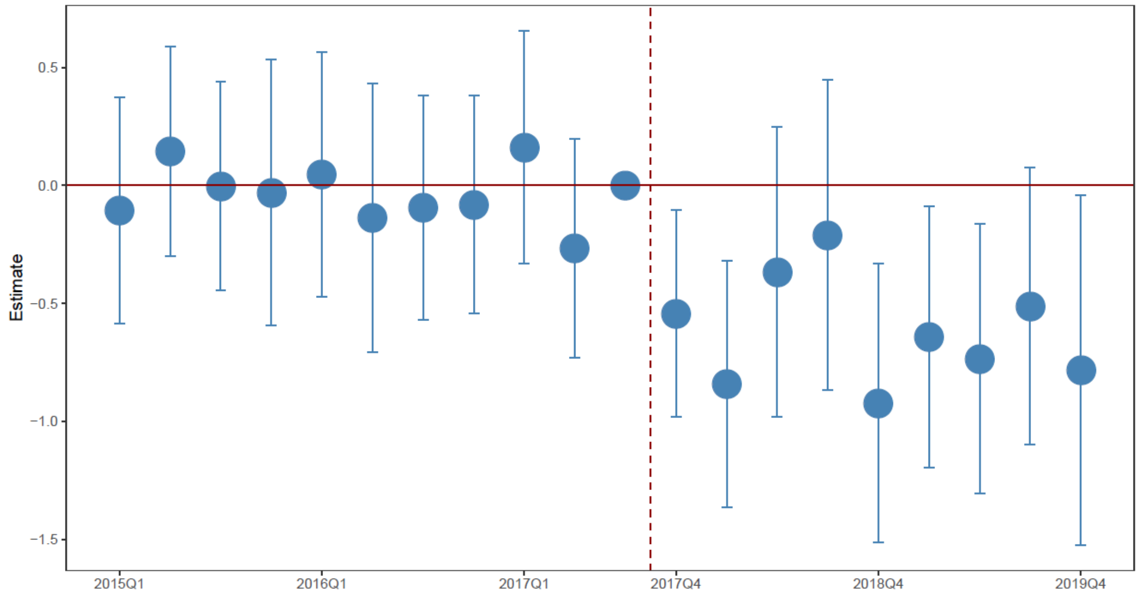
(a) Keyword = “metoo”



(b) Keyword = “sexual harassment + sexual abuse + sexual violence”

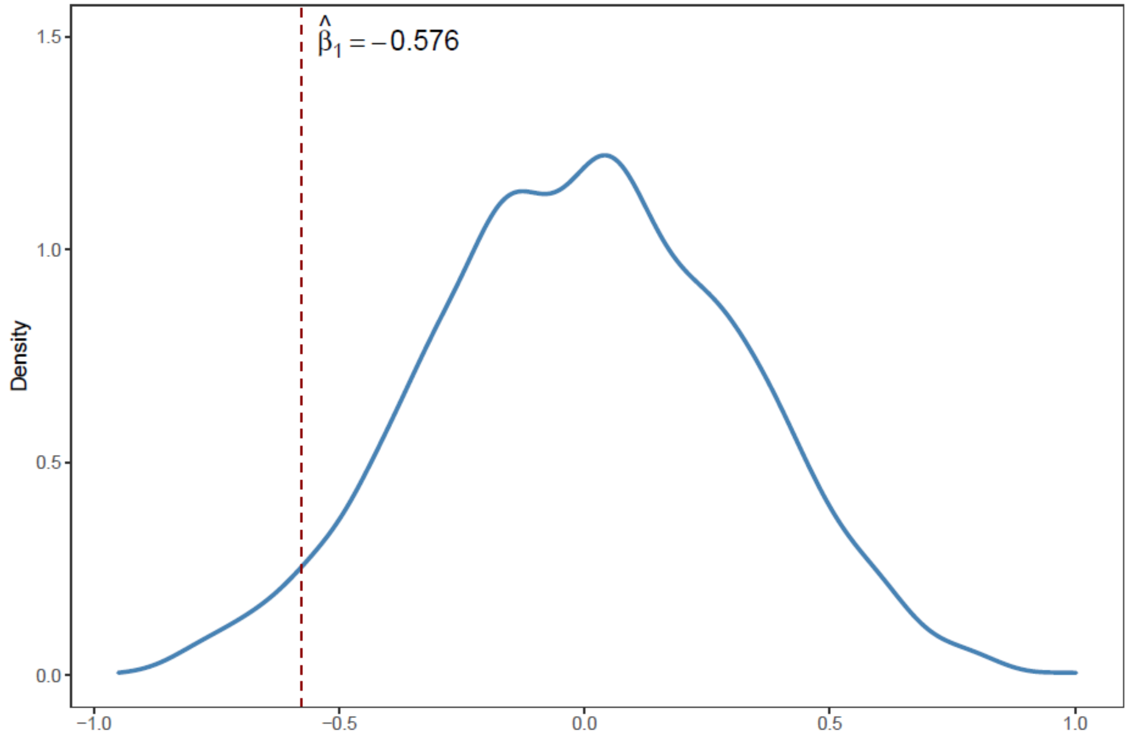
Notes. This figure plots the quarterly aggregated Google Trends indices for keyword “metoo” (panel (a)) and “sexual harassment + sexual abuse + sexual violence” (panel (b)) during 2015 - 2019. The vertical dashed line denotes 2017Q4 when the MeToo movement was sparked.

Figure 8 Coefficients on the interactions between sexism index and year-by-quarter dummies



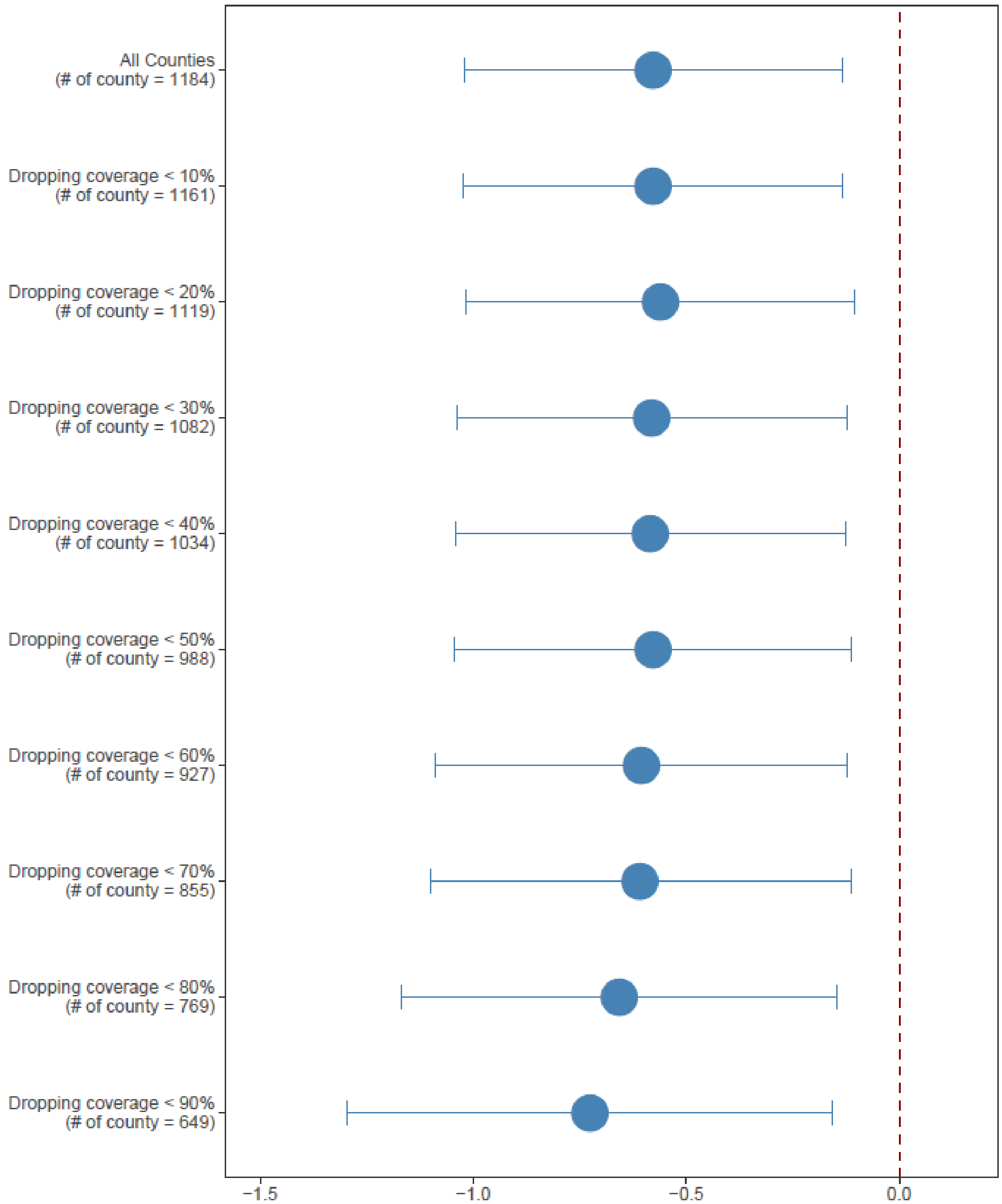
Notes. This figure plots the coefficients on the interaction terms between the year-by-quarter dummy and sexism index in equation (2) using observations between 2015Q1 and 2019Q4, after accounting for the county fixed effect, year-by-quarter fixed effect, county-level covariates, police size, and Google controls. The blue vertical bars denote the 95% confidence intervals for each estimate. 2017Q3 is dropped as the comparison group.

Figure 9 Empirical distribution of placebo sexual crime rate estimates



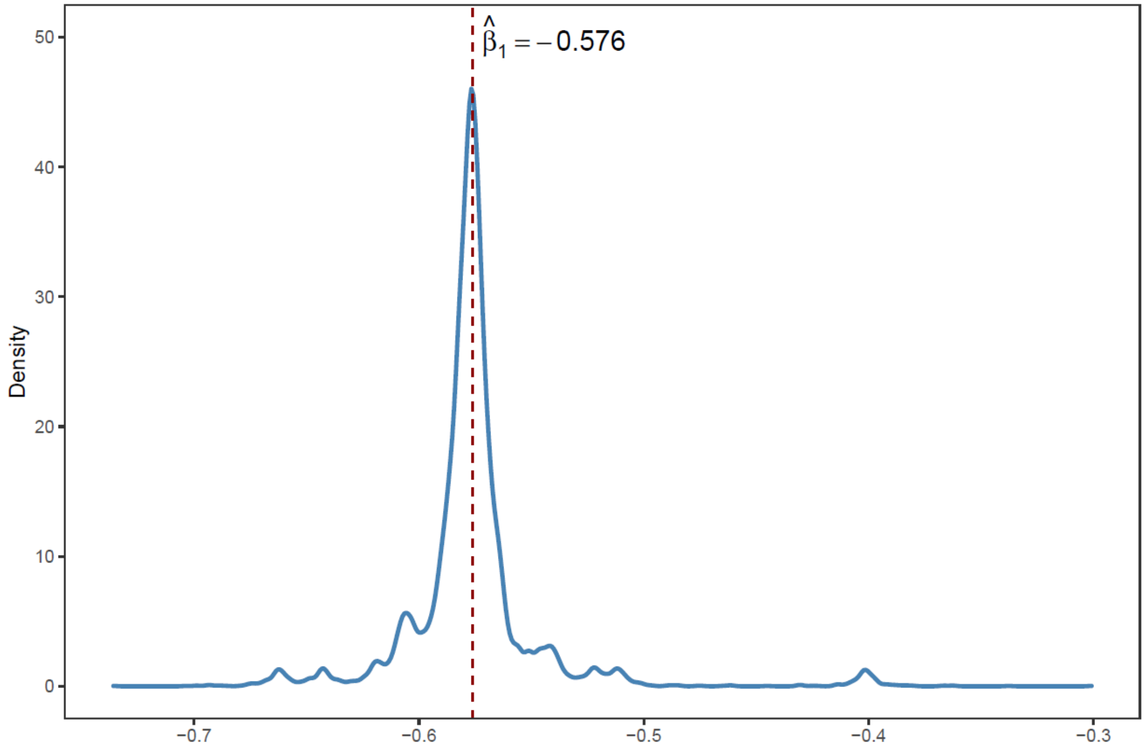
Notes. This figure plots the distribution of β1 estimates in equation (1) with the sexism indices among the 133 MMs permuted 2,000 times. The vertical dashed line corresponds to -0.576, the estimate of our preferred specification column (4) in Table 4. It ranks below the 4th percentile of the distribution.

Figure 10 Robustness check: Dropping counties with different population coverage rates



Notes. This figure plots the estimates of β1 in equation (1) and its 95% confidence intervals after dropping counties with a variety of population coverage rates.

Figure 11 Robustness check: Dropping all counties in two randomly selected MMs



Notes. This figure plots the distribution of β1 estimates in equation (1) after dropping all counties in two randomly selected MMs from the 133 MMs. We exhaustively repeat the estimation procedure 8,778 times. The vertical dashed line denotes -0.576, the estimate of our preferred specification column (4) in Table 4.

Table 1 Coding of offensive words

Word 1	Word 2	Word 3	Word 4	Word 5
bitch	cunt	slut	whore	fuck

Table 2 Google search terms

Panel (a): Dictionary of derogatory terms
Sexism Slurs
[Word 1](es) − pornhub, [Word 2](s) − pornhub, [Word 3](s) − pornhub, [Word 4](s) − pornhub
Appearance
fat woman(en) + fat girl(s) + ugly woman(en) + ugly girl(s)
Mental Stability
emotional woman(en) + emotional girl(s) + mad woman(en) + mad girl(s)
Intellect
stupid woman(en) + stupid girl(s) + dumb woman(en) + dumb girl(s)
Sex
dirty woman(en) + dirty girl(s) + easy woman(en) + easy girl(s)
Panel (b): Google Controls
(1) beautiful woman(en) + beautiful girl(s); (2) [Word 5]

Notes. "+" and "–" respectively represent "or" and "excluding" in Google Trends. Coding of the [Word 1] - [Word 5] is listed in Table 1.

Table 3 Summary statistics

	Pre-MeToo (2015Q1 - 2017Q3)				Post-MeToo (2017Q4 - 2019Q4)			
	All	Low	High	Low - High	All	Low	High	Low - High
Panel (a): Sexual crime data (per 100,000 populations; per 10 million populations for homicide)								
Sexual crime	21.87	21.73	22.12	-0.39	22.42	22.70	21.92	0.77*
	(12.71)	(12.43)	(13.19)	[-1.02]	(12.53)	(12.52)	(12.54)	[1.75]
Known	16.33	16.06	16.82	-0.76***	16.34	16.34	16.35	-0.01
	(10.55)	(10.01)	(11.45)	[-2.61]	(9.71)	(9.31)	(10.38)	[-0.02]
Unknown	4.03	3.92	4.22	-0.30**	4.35	4.36	4.33	0.03
	(4.51)	(4.23)	(4.96)	[-2.26]	(4.79)	(4.65)	(5.04)	[0.22]
Non-injury	16.83	16.68	17.10	-0.42	17.64	17.89	17.21	0.68*
	(10.87)	(10.56)	(11.39)	[-1.33]	(10.82)	(10.83)	(10.78)	[1.80]
Injury	5.04	5.05	5.02	0.03	4.78	4.81	4.71	0.10
	(4.13)	(4.07)	(4.24)	[0.29]	(3.86)	(3.69)	(4.15)	[0.86]
Resident	14.55	13.51	16.40	-2.88***	14.28	14.26	14.32	-0.06
	(10.15)	(9.17)	(11.46)	[-8.73]	(9.96)	(9.72)	(10.37)	[-0.15]
Non-resident	4.03	4.36	3.44	0.92***	3.91	4.10	3.56	0.54***
	(4.18)	(4.21)	(4.06)	[8.77]	(4.27)	(4.19)	(4.39)	[4.39]
Forcible	20.51	20.37	20.77	-0.40	21.24	21.51	20.76	0.76*
	(12.34)	(12.16)	(12.64)	[-1.08]	(12.29)	(12.36)	(12.17)	[1.72]
Non-forcible	1.36	1.36	1.35	0.02	1.18	1.18	1.17	0.02
	(2.18)	(2.14)	(2.27)	[0.32]	(1.92)	(1.83)	(2.07)	[0.34]
Arrest	4.10	4.20	3.94	0.26**	3.74	3.90	3.46	0.44***
	(3.99)	(3.95)	(4.04)	[2.57]	(3.61)	(3.61)	(3.61)	[4.28]
Homicide	0.12	0.08	0.18	-0.10	0.13	0.10	0.18	-0.08
	(3.83)	(2.05)	(5.78)	[-1.16]	(3.92)	(3.01)	(5.17)	[-0.87]
Agg. assault	0.05	0.05	0.05	0.00	0.06	0.06	0.05	0.01
	(0.25)	(0.21)	(0.31)	[0.44]	(0.29)	(0.24)	(0.37)	[0.84]
Non-sexual index crime	161.53	149.37	183.22	-33.84***	140.33	127.24	163.85	-36.62***
	(99.77)	(94.90)	(104.47)	[-7.16]	(90.43)	(81.93)	(99.75)	[-7.55]
Panel (b): County covariates								
% Female	50.72	50.64	50.86	-0.22***	50.70	50.60	50.87	-0.27***
	(1.16)	(1.14)	(1.19)	[-6.23]	(1.17)	(1.14)	(1.20)	[-6.94]
% White	72.73	72.43	73.27	-0.84	71.79	71.32	72.64	-1.33**
	(15.19)	(14.45)	(16.42)	[-1.38]	(15.33)	(14.66)	(16.44)	[-1.98]
% Black	11.04	9.08	14.53	-5.44***	11.13	9.22	14.55	-5.33***
	(12.17)	(9.77)	(14.94)	[-9.55]	(12.07)	(9.67)	(14.88)	[-8.56]
% Hispanic	9.53	10.84	7.19	3.65***	10.02	11.37	7.58	3.80***
	(8.72)	(8.89)	(7.87)	[13.37]	(8.91)	(9.09)	(8.01)	[12.29]
% College edu. or higher	60.17	62.85	55.39	7.46***	61.30	63.86	56.69	7.17***
	(10.00)	(9.14)	(9.66)	[23.94]	(9.84)	(9.04)	(9.55)	[20.77]
Labor participation rate	64.36	65.72	61.94	3.78***	64.23	65.60	61.77	3.83***
	(6.09)	(5.54)	(6.27)	[20.76]	(6.17)	(5.57)	(6.42)	[18.46]
Poverty rate	14.25	12.99	16.51	-3.52***	13.12	11.86	15.37	-3.51***
	(5.11)	(4.85)	(4.77)	[-19.40]	(4.89)	(4.53)	(4.68)	[-18.58]
log(Personal income)	10.97	11.03	10.84	0.19***	11.04	11.11	10.91	0.19***
	(0.21)	(0.21)	(0.14)	[26.15]	(0.21)	(0.21)	(0.14)	[23.20]
Officer/100,000 pop.	193.82	187.76	204.62	-16.86***	186.92	180.46	198.51	-18.04***
	(82.55)	(72.70)	(96.75)	[-5.66]	(82.43)	(71.32)	(98.26)	[-5.36]
Population	399716.67	509759.77	203351.89	49179.73***	407987.18	519818.99	207115.83	50438.28***
	(430350.35)	(476018.02)	(226289.13)	[18.52]	(437599.28)	(484209.39)	(227043.19)	[16.86]
Observations	13024	6116	6908	13024	10656	5004	5652	10656

Notes. The observation unit is county-year-quarter for panel (a), and county-year for panel (b). Values in round parentheses are standard deviations, and values in square parentheses are t-statistics. Low denotes counties locating in media markets whose sexism indices are below the 50th percentile, and High denotes counties locating in media markets whose sexism indices are above the 50th percentile. ***significant at the 1% level, **5% level, *10% level.

Table 4 The effect of sexism on sexual crime rate after MeToo

Dep. Var. = Sexual crimes/100,000 populations	(1)	(2)	(3)	(4)	(5)
SexismIndex×Post	-0.536***	-0.573**	-0.580**	-0.576**	-0.500***
	[0.195]	[0.223]	[0.222]	[0.225]	[0.133]
Pre-MeToo Mean			21.868		
R^2	0.739	0.739	0.740	0.740	0.753
Observations	23,680	23,680	23,680	23,680	23,680
County & Y-by-Q FEs	Yes	Yes	Yes	Yes	Yes
County covariates	No	Yes	Yes	Yes	Yes
Police officer	No	No	Yes	Yes	Yes
Google controls	No	No	No	Yes	Yes
MM-by-Year FE	No	No	No	No	Yes

Notes. Standard errors are in brackets and clustered at the media market-level. Post is an indicator and set to 1 from 2017Q4. SexismIndex is the MM-level PLS sexism index. County-level covariates include percentage of female, percentage of non-Hispanic white, percentage of non-Hispanic black, percentage of Hispanic, percentage of population with college education or higher, labor participation rate, poverty rate, and logarithmic personal income. Police size represents the number of police per 100,000 populations. Google controls are Google Trends indices for "beautiful woman(en) + beautiful girl(s)" and "[Word 5]". All regressions are weighted by the county populations residing in the constituent jurisdictions consistently reporting to the NIBRS. ***significant at the 1% level, **5% level, *10% level.

Table 5 Determinants of sexism index

Dep. Var. = PLS sexism index	(1)	(2)	(3)	(4)	(5)
% Female	-0.001				0.024
	[0.025]				[0.022]
% with college education		-0.017***			0.015**
		[0.004]			[0.006]
Labor participation rate			-0.037***		-0.021**
			[0.006]		[0.008]
log(personal income)				-1.130***	-1.492***
				[0.249]	[0.361]
Poverty rate				0.015	0.006
				[0.011]	[0.011]
R^2	0.705	0.723	0.734	0.748	0.754
Observations	1,184	1,184	1,184	1,184	1,184

Notes. Standard errors are in brackets and clustered at the media market-level. Each column is a separate regression of the PLS sexism index on the county-level characteristics, including percentage of female, percentage of population with college education or higher, labor participation rate, poverty rate, and logarithmic personal income are calculated from the average of census in 2015 and 2016. All regressions are weighted by average county population in 2015 and 2016. ***significant at the 1% level, **5% level, *10% level.

Table 6 Endogeneity test: The effects of sexism on sexual crime rate after MeToo using different covariates

Dep. Var. = Sexual crimes/100,000 populations	(1)
Baseline	-0.536***
	[0.195]
% Female	-0.543***
	[0.202]
% White	-0.618**
	[0.241]
% Black	-0.583**
	[0.223]
% Hispanic	-0.584***
	[0.222]
% with college education	-0.582***
	[0.221]
Labor participation rate	-0.572**
	[0.220]
Poverty rate	-0.574***
	[0.219]
log(personal income)	-0.573**
	[0.223]
Officer/100,000 populations	-0.580**
	[0.222]
Google Control 1	-0.582***
	[0.221]
Google Control 2	-0.576**
	[0.225]
Observations	23,680
County FEs	Yes
Y-by-Q FEs	Yes

Notes. Standard errors are in brackets and clustered at the media market-level. The baseline row reports the estimate of the baseline model which only contains the county and year-by-quarter fixed effects. Below the baseline result, each row documents the estimate of a regression that includes a covariate and all covariates included in the preceding row, including percentage of female, percentage of non-Hispanic white, percentage of non-Hispanic black, percentage of Hispanic, percentage of population with college education or higher, labor participation rate, poverty rate, and logarithmic personal income, number of police per 100,000 populations, and Google Trends indices for “beautiful woman(en) + beautiful girl(s)” and “[Word 5]”. All regressions are weighted by the county populations residing in the constituent jurisdictions consistently reporting to the NIBRS. ***significant at the 1% level, **5% level, *10%level.

Table 7 Endogeneity test: The potential effects of other sources on sexual crime rate

Dep. Var. = Sexual crimes/100,000 populations	(1)	(2)	(3)	(4)	(5)
SexismIndex×Post	-0.695***	-0.670***	-0.700***	-0.751***	-0.777***
	[0.241]	[0.234]	[0.252]	[0.268]	[0.263]
GapSchool×Post		-0.110			0.105
		[0.265]			[0.279]
GapLFP×Post			-0.020		0.548
			[0.289]		[0.444]
GapIncome×Post				-0.203	-0.680
				[0.242]	[0.427]
Pre-MeToo Mean			21.297		
R^2	0.770	0.770	0.770	0.770	0.770
Observations	15,740	15,740	15,740	15,740	15,740
County & Y-by-Q FEs	Yes	Yes	Yes	Yes	Yes
County covariates	Yes	Yes	Yes	Yes	Yes
Police officer	Yes	Yes	Yes	Yes	Yes
Google controls	Yes	Yes	Yes	Yes	Yes
MM-by-Year FE	No	No	No	No	No

Notes. Standard errors are in brackets and clustered at the media market-level. Post is an indicator and set to 1 from 2017Q4. SexismIndex is the MM-level PLS sexism index. GapSchool, GapLFP, and GapIncome respectively denote the gender gap in schooling years, labor participation rate, and personal income. County-level covariates include percentage of female, percentage of non-Hispanic white, percentage of non-Hispanic black, percentage of Hispanic, percentage of population with college education or higher, labor participation rate, poverty rate, and logarithmic personal income. Police size represents the number of police per 100,000 populations. Google controls are Google Trends indices for "beautiful woman(en) + beautiful girl(s)" and "[Word 5]". All regressions are weighted by the county populations residing in the constituent jurisdictions consistently reporting to the NIBRS. ***significant at the 1%level, **5% level, *10% level.

Table 8 Heterogeneous MeToo effect by offender and crime types

Dep. Var. = Sexual crimes/100,000 populations								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel (a)	Known Offener				Unknown Offender			
SexismIndex×Post	-0.361***	-0.342**	-0.343**	-0.345**	-0.131	-0.188*	-0.193*	-0.188*
	[0.121]	[0.142]	[0.142]	[0.142]	[0.101]	[0.109]	[0.109]	[0.111]
Pre-MeToo Mean	16.335				4.030			
R^2	0.697	0.697	0.697	0.697	0.649	0.65	0.65	0.65
Panel (b)	No Injury				With Injury			
SexismIndex×Post	-0.438**	-0.514**	-0.521**	-0.510**	-0.098*	-0.059	-0.058	-0.066
	[0.184]	[0.204]	[0.204]	[0.207]	[0.058]	[0.066]	[0.066]	[0.065]
Pre-MeToo Mean	16.830				5.038			
R^2	0.717	0.718	0.718	0.718	0.518	0.519	0.519	0.519
Panel (c)	Forcible				Non-Forcible			
SexismIndex×Post	-0.509***	-0.548**	-0.557***	-0.547**	-0.027	-0.024	-0.022	-0.030
	[0.183]	[0.211]	[0.210]	[0.213]	[0.028]	[0.031]	[0.031]	[0.031]
Pre-MeToo Mean	20.511				1.358			
R^2	0.743	0.744	0.744	0.744	0.46	0.462	0.462	0.462
Panel (d)	Resident				Non-Resident			
SexismIndex×Post	-1.254**	-1.304**	-1.323**	-1.330**	0.042	0.102	0.104	0.102
	[0.527]	[0.527]	[0.525]	[0.531]	[0.131]	[0.144]	[0.145]	[0.148]
Pre-MeToo Mean	14.547				4.032			
R^2	0.662	0.663	0.664	0.664	0.571	0.573	0.573	0.573
Panel (e)	Female				Male			
SexismIndex×Post	-0.411**	-0.444**	-0.448**	-0.449**	-0.118***	-0.120**	-0.123**	-0.120**
	[0.165]	[0.184]	[0.184]	[0.186]	[0.042]	[0.048]	[0.048]	[0.049]
Pre-MeToo Mean	18.860				2.952			
R^2	0.710	0.710	0.710	0.710	0.430	0.430	0.431	0.431
Panel (f)	White				Black			
SexismIndex×Post	-0.272***	-0.291**	-0.296**	-0.295**	-0.200*	-0.227*	-0.229**	-0.229*
	[0.102]	[0.118]	[0.117]	[0.115]	[0.106]	[0.115]	[0.116]	[0.117]
Pre-MeToo Mean	16.458				3.818			
R^2	0.728	0.728	0.728	0.728	0.861	0.862	0.862	0.862
Observations	23,680	23,680	23,680	23,680	23,680	23,680	23,680	23,680
County & Y-by-Q FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
County covariates	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Police officer	No	No	Yes	Yes	No	No	Yes	Yes
Google controls	No	No	No	Yes	No	No	No	Yes

Notes. Standard errors are in brackets and clustered at the media market-level. Post is an indicator and set to 1 from 2017Q4. SexismIndex is the MM-level PLS sexism index. County-level covariates include percentage of female, percentage of non-Hispanic white, percentage of non-Hispanic black, percentage of Hispanic, percentage of population with college education or higher, labor participation rate, poverty rate, and logarithmic personal income. Police size represents the number of police per 100,000 populations. Google controls are Google Trends indices for "beautiful woman(en) + beautiful girl(s)" and "[Word 5]". All regressions are weighted by the county populations residing in the constituent jurisdictions consistently reporting to the NIBRS. ***significant at the 1% level, **5% level, *10% level.

Table 9 Falsification tests

Dep. Var. = Non-sexual index crimes against women/100,000 populations				
	(1)	(2)	(3)	(4)
SexismIndex×Post	1.809	0.834	0.765	0.898
	[1.952]	[1.456]	[1.451]	[1.453]
Pre-MeToo Mean		161.529		
R^2	0.938	0.938	0.938	0.938
Observations	23,680	23,680	23,680	23,680
County & Y-by-Q FEs	Yes	Yes	Yes	Yes
County covariates	No	Yes	Yes	Yes
Police size	No	No	Yes	Yes
Google controls	No	No	No	Yes

Notes. Standard errors are in brackets and clustered at the media market-level. Post is an indicator and set to 1 from 2017Q4. SexismIndex is the MM-level PLS sexism index. County-level covariates include percentage of female, percentage of non-Hispanic white, percentage of non-Hispanic black, percentage of Hispanic, percentage of population with college education or higher, labor participation rate, poverty rate, and logarithmic personal income. Police size represents the number of police per 100,000 populations. Google controls are Google Trends indices for "beautiful woman(en) + beautiful girl(s)" and "[Word 5]". All regressions are weighted by the county populations residing in the constituent jurisdictions consistently reporting to the NIBRS. ***significant at the 1% level, **5% level, *10% level.

Table 10 Robustness checks: Alternative measures of sexism

Dep. Var. = Sexual crimes/100,000 populations	(1)	(2)	(3)	(4)
Panel (a): LowSexism (binary) = 1 if low sexism MM (below median of sexism index)				
LowSexism×Post	1.152**	1.191**	1.198**	1.191**
	[0.470]	[0.494]	[0.493]	[0.496]
Pre-MeToo Mean		21.868		
R^2	0.739	0.739	0.739	0.739
Observations	23,680	23,680	23,680	23,680
Panel (b): Awareness (continuous) = Google Trends index for "sexual harassment" keywords				
SexismIndex×Awareness	-0.883***	-0.852***	-0.854***	-0.837***
	[0.292]	[0.288]	[0.288]	[0.287]
Pre-MeToo Mean		21.868		
R^2	0.739	0.739	0.739	0.739
Observations	23,680	23,680	23,680	23,680
Panel (c): SexismGSS (continuous) = Sexism index by GSS in Charles et al. (2020)				
SexismGSS×Post	-0.651**	-0.650**	-0.652**	-0.648**
	[0.254]	[0.254]	[0.251]	[0.254]
Pre-MeToo Mean		21.808		
R^2	0.746	0.746	0.746	0.746
Observations	22,240	22,240	22,240	22,240
Panel (d): WomenMarch (continuous) = Women's march protesters per 1,000 populations				
WomenMarch×Post	0.340**	0.321**	0.315**	0.309**
	[0.131]	[0.143]	[0.143]	[0.143]
Pre-MeToo Mean		21.868		
R^2	0.739	0.739	0.739	0.739
Observations	23,680	23,680	23,680	23,680
County & Y-by-Q FEs	Yes	Yes	Yes	Yes
County covariates	No	Yes	Yes	Yes
Police officer	No	No	Yes	Yes
Google controls	No	No	No	Yes

Notes. Standard errors are in brackets and clustered at the media market-level. Post is an indicator and set to 1 from 2017Q4. SexismIndex is the MM-level PLS sexism index. County-level covariates include percentage of female, percentage of non-Hispanic white, percentage of non-Hispanic black, percentage of Hispanic, percentage of population with college education or higher, labor participation rate, poverty rate, and logarithmic personal income. Police size represents the number of police per 100,000 populations. Google controls are Google Trends indices for "beautiful woman(en) + beautiful girl(s)" and "[Word 5]". All regressions are weighted by the county populations residing in the constituent jurisdictions consistently reporting to the NIBRS. ***significant at the 1% level, **5% level, *10% level.

Table 11 Sexual crime victims' willingness to report after MeToo

Dep. Var. = Reported to Police or Not (Reported = 1)	Linear Probability		Logit		Probit	
	(1)	(2)	(3)	(4)	(5)	(6)
Post	0.177***	0.305***	1.017***	1.840***	0.601***	1.034***
	[0.016]	[0.014]	[0.048]	[0.260]	[0.017]	[0.096]
R^2	0.102	0.216	-	-	-	-
Pseudo-R^2	-	-	0.083	0.191	0.083	0.188
Marginal Effect	-	-	0.209	0.352	0.207	0.342
Pre-MeToo Mean			0.315			
Observations	1,077	1,077	1,077	1,077	1,077	1,077
Region FEs	Yes	Yes	Yes	Yes	Yes	Yes
Incident year FEs	Yes	Yes	Yes	Yes	Yes	Yes
Interview wave FEs	Yes	Yes	Yes	Yes	Yes	Yes
Incident controls	No	Yes	No	Yes	No	Yes
Victim controls	No	Yes	No	Yes	No	Yes

Notes. The incidents related to sexual offense are from the 2015 - 2019 NCVS data. Post is an indicator and set to 1 for years 2018 and 2019. Region FEs include dummies for Northeast, Midwest, South, and West. Incident year FEs include dummies for 2014 - 2019. Interview wave FEs include dummies for 2015Q1 - 2019Q4. Incident controls include well known to victim, attempt/threat to rape, injury, single offender, forced or coerced unwanted sex, and in city of (S)MSA. Victim controls include age, schooling years, married, male, white, black, Asian, and Hispanic. Heteroscedasticity-consistent standard errors are in brackets. ***significant at the 1% level, **5% level, *10% level.

Table 12 Deterrence effect on homicides and aggravated assault related to sexual crimes

	(1)	(2)	(3)	(4)
Panel (a): Dep. Var. = (Homicides related to sexual crime/100,000 populations)×100				
SexismIndex×Post	0.014	-0.004	-0.005	-0.015
	[0.055]	[0.069]	[0.069]	[0.072]
Pre-MeToo Mean		0.119		
R^2	0.056	0.057	0.057	0.057
Observations	23,680	23,680	23,680	23,680
Panel (b): Dep. Var. = Count of homicide related to sexual crime				
SexismIndex×Post	0.106	0.121	0.147	0.077
	[0.301]	[0.330]	[0.351]	[0.358]
Pre-MeToo Mean		0.003		
Observations	23,680	23,680	23,680	23,680
Panel (c): Dep. Var. = Aggravated assault related to sexual crime/100,000 populations				
SexismIndex×Post	0.002	0.002	0.002	0.002
	[0.003]	[0.003]	[0.003]	[0.003]
Pre-MeToo Mean		0.047		
R^2	0.084	0.084	0.085	0.085
Observations	23,680	23,680	23,680	23,680
Panel (d): Dep. Var. = (Arrests/Total sexual crimes)×100%				
SexismIndex×Post	0.124	0.102	0.105	0.071
	[0.249]	[0.257]	[0.256]	[0.259]
Pre-MeToo Mean		20.105		
R^2	0.474	0.475	0.475	0.475
Observations	20,417	20,417	20,417	20,417
County & Y-by-Q FEs	Yes	Yes	Yes	Yes
County covariates	No	Yes	Yes	Yes
Police officer	No	No	Yes	Yes
Google control	No	No	No	Yes

Notes. Standard errors are in brackets and clustered at the media market-level. Post is an indicator and set to 1 from 2017Q4. SexismIndex is the MM-level PLS sexism index. County-level covariates include percentage of female, percentage of non-Hispanic white, percentage of non-Hispanic black, percentage of Hispanic, percentage of population with college education or higher, labor participation rate, poverty rate, and logarithmic personal income. Police size represents the number of police per 100,000 populations. Google controls are Google Trends indices for “beautiful woman(en) + beautiful girl(s)” and “[Word 5]”. All regressions are weighted by the county populations residing in the constituent jurisdictions consistently reporting to the NIBRS. ***significant at the 1% level, **5% level, *10% level.

REFERENCES

Abdulkadiroğlu, Atila, Joshua D Angrist, Susan M Dynarski, Thomas J Kane, and Parag A Pathak. 2011. "Accountability and flexibility in public schools: Evidence from Boston's charters and pilots." *The Quarterly Journal of Economics* 126 (2):699-748.

Abrams, Dominic, G Tendayi Viki, Barbara Masser, and Gerd Bohner. 2003. "Perceptions of stranger and acquaintance rape: The role of benevolent and hostile sexism in victim blame and rape proclivity." *Journal of personality and social psychology* 84 (1):111.

Aizer, Anna. 2010. "The gender wage gap and domestic violence." *American Economic Review* 100 (4):1847-59.

Aizer, Anna, and Pedro Dal Bo. 2009. "Love, hate and murder: Commitment devices in violent relationships." *Journal of public Economics* 93 (3-4):412-428.

Almond, Douglas, Hilary W Hoynes, and Diane Whitmore Schanzenbach. 2011. "Inside the war on poverty: The impact of food stamps on birth outcomes." *Review of Economics and Statistics* 93 (2):387-403.

Altonji, Joseph G, Todd E Elder, and Christopher R Taber. 2005. "Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools." *Journal of political economy* 113 (1):151-184.

Anderberg, Dan, Helmut Rainer, Jonathan Wadsworth, and Tanya Wilson. 2016. "Unemployment and domestic violence: Theory and evidence." *The Economic Journal* 126 (597):1947-1979.

Angrist, Joshua D, Sarah R Cohodes, Susan M Dynarski, Parag A Pathak, and Christopher D Walters. 2013. "Charter Schools and the Road to College Readiness: The Effects on College Preparation, Attendance and Choice. Understanding Boston." *Boston Foundation*.

Angrist, Joshua D, Sarah R Cohodes, Susan M Dynarski, Parag A Pathak, and Christopher R Walters. 2016. "Stand and deliver: Effects of Boston's charter high schools on college preparation, entry, and choice." *Journal of Labor Economics* 34 (2):275-318.

Angrist, Joshua D, Peter D Hull, Parag A Pathak, and Christopher R Walters. 2017. "Leveraging lotteries for school value-added: Testing and estimation." *The Quarterly Journal of Economics* 132 (2):871-919.

Angrist, Joshua D, Parag A Pathak, and Christopher R Walters. 2013. "Explaining charter school effectiveness." *American Economic Journal: Applied Economics* 5 (4):1-27.

Appelbaum, Eileen, and Ruth Milkman. 2015. Leaves That Pay: Employer and Worker Experiences with Paid Family Leave in California. Washington DC: Center for Economic and Policy Research.

Baker, Michael, and Kevin Milligan. 2008. "Maternal employment, breastfeeding, and health: Evidence from maternity leave mandates." *Journal of Health Economics* 27 (4):871-887.

Baker, Michael, and Kevin Milligan. 2010. "Evidence from maternity leave expansions of the impact of maternal care on early child development." *Journal of Human Resources* 45 (1):1-32.

Baker, Michael, and Kevin Milligan. 2015. "Maternity leave and children's cognitive and behavioral development." *Journal of Population Economics* 28 (2):373-391.

Baker, Scott R, and Andrey Fradkin. 2017. "The impact of unemployment insurance on job search: Evidence from Google search data." *Review of Economics and Statistics* 99 (5):756-768.
Bana, Sarah, Kelly Bedard, and Maya Rossin-Slater. 2018. "Trends and Disparities in Leave Use under California's Paid Family Leave Program: New Evidence from Administrative Data." *AEA Papers and Proceedings* 108:388-91. doi: doi: 10.1257/pandp.20181113.
Bartel, Ann P, Maya Rossin-Slater, Christopher J Ruhm, Jenna Stearns, and Jane Waldfogel. 2018. "Paid family leave, fathers' leave-taking, and leave-sharing in dual-earner households." *Journal of Policy Analysis and Management* 37 (1):10-37.
Baum, Charles L, and Christopher J Ruhm. 2016. "The effects of paid family leave in California on labor market outcomes." *Journal of Policy Analysis and Management* 35 (2):333-356.
Beitsch, Rebecca. 2018. "# MeToo has changed our culture. Now it's changing our laws." *The Pew Charitable Trusts*.
Bennett, Benjamin, Isil Erel, Léa H Stern, and Zexi Wang. 2020. *(Forced) Feminist Firms*: National Bureau of Economic Research.
Berger, Lawrence M, Jennifer Hill, and Jane Waldfogel. 2005. "Maternity leave, early maternal employment and child health and development in the US." *Economic Journal* 115 (501):F29-F47.
Bergman, Peter, and Jr McFarlin, Isaac. 2020. Education for all? A nationwide audit study of school choice. National Bureau of Economic Research.
Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2004. "How much should we trust differences-in-differences estimates?" *Quarterly Journal of Economics* 119 (1):249-275.
Bertrand, Marianne, Claudia Goldin, and Lawrence F Katz. 2010. "Dynamics of the gender gap for young professionals in the financial and corporate sectors." *American economic journal: applied economics* 2 (3):228-55.
Bettinger, Eric P. 2005. "The effect of charter schools on charter students and public schools." *Economics of Education Review* 24 (2):133-147.
Bifulco, Robert, and Christian Buerger. 2015. "The influence of finance and accountability policies on location of New York state charter schools." *journal of education finance*:193-221.
Bifulco, Robert, and Helen F Ladd. 2006. "The impacts of charter schools on student achievement: Evidence from North Carolina." *Education Finance and Policy* 1 (1):50-90.
Booker, Kevin, Brian Gill, Tim Sass, and Ron Zimmer. 2014. Charter high schools' effects on educational attainment and earnings. Mathematica Policy Research.
Booker, Kevin, Scott M Gilpatric, Timothy Gronberg, and Dennis Jansen. 2007. "The impact of charter school attendance on student performance." *Journal of Public Economics* 91 (5-6):849-876.
Booker, Kevin, Scott M Gilpatric, Timothy Gronberg, and Dennis Jansen. 2008. "The effect of charter schools on traditional public school students in Texas: Are children who stay behind left behind?" *Journal of Urban Economics* 64 (1):123-145.
Booker, Kevin, Tim R Sass, Brian Gill, and Ron Zimmer. 2011. "The effects of charter high schools on educational attainment." *Journal of Labor Economics* 29 (2):377-415.

Branch, Gregory F, Eric A Hanushek, and Steven G Rivkin. 2012. Estimating the effect of leaders on public sector productivity: The case of school principals. National Bureau of Economic Research.

Brewer, Dominic, Ron Zimmer, Richard Buddin, Derrick Chau, and Brian Gill. 2003. "Charter school operations and performance: Evidence from California."

Brodeur, Abel, Andrew E Clark, Sarah Fleche, and Nattavudh Powdthavee. 2021. "COVID-19, lockdowns and well-being: Evidence from Google Trends." *Journal of public economics* 193:104346.

Broomfield, Matt. 2017. "Women's march against Donald Trump is the largest day of protests in US history, say political scientists." *The Independent* 23.

Bross, Whitney, Douglas N Harris, and Lihan Liu. 2016. "The effects of performance-based school closure and charter takeover on student performance." *Education Research Alliance for New Orleans*.

Buddin, Richard. 2012. "The impact of charter schools on public and private school enrollments." *Cato Institute Policy Analysis* (707).

Bullinger, Lindsey Rose. 2019. "The Effect of Paid Family Leave on Infant and Parental Health in the United States." *Journal of Health Economics* 66:101-116.

Callaway, Brantly, and Pedro HC Sant'Anna. 2021. "Difference-in-differences with multiple time periods." *Journal of Econometrics* 225 (2):200-230.

Carlsen, Audrey, Maya Salam, Claire Cain Miller, Denise Lu, Ash Ngu, Jugal K Patel, and Zach Wichter. 2018. "MeToo brought down 201 powerful men. Nearly half of their replacements are women." *New York Times* 29.

Carlson, Deven, Lesley Lavery, and John F Witte. 2012. "Charter school authorizers and student achievement." *Economics of Education Review* 31 (2):254-267.

Carneiro, Pedro, Katrine V Løken, and Kjell G Salvanes. 2015. "A flying start? Maternity leave benefits and long-run outcomes of children." *Journal of Political Economy* 123 (2):365-412.

Chabrier, Julia, Sarah Cohodes, and Philip Oreopoulos. 2016. "What can we learn from charter school lotteries?" *Journal of Economic Perspectives* 30 (3):57-84.

Chakrabarti, Rajashri, and Joydeep Roy. 2016. "Do charter schools crowd out private school enrollment? Evidence from Michigan." *Journal of Urban Economics* 91:88-103.

Charles, Kerwin Kofi, Jonathan Guryan, and Jessica Pan. 2020. The effects of sexism on American women: The role of norms vs. discrimination. National Bureau of Economic Research.

Cheng, Ing-Haw, and Alice Hsiaw. 2021. "Reporting sexual misconduct in the# metoo era." *American Economic Journal: Microeconomics* (3506936).

Conklin, Michael. 2020. "# MeToo Effects on Juror Decision Making." *Calif. L. Rev. Online* 11:179.

Conley, Timothy G, and Christopher R Taber. 2011. "Inference with "difference in differences" with a small number of policy changes." *Review of Economics and Statistics* 93 (1):113-125.

Cordes, Sarah A. 2018. "In pursuit of the common good: The spillover effects of charter schools on public school students in New York City." *Education Finance and Policy* 13 (4):484-512.

CREDO, Center for Research on Education Outcomes. 2013. "National charter school study." 54 (6):1079-1116.

Currie, Janet, Matthew Neidell, and Johannes F Schmieder. 2009. "Air pollution and infant health: Lessons from New Jersey." *Journal of Health Economics* 28 (3):688-703.

Curto, Vilsa E, and Jr Fryer, Roland G. 2014. "The potential of urban boarding schools for the poor: Evidence from SEED." *Journal of Labor Economics* 32 (1):65-93.

Dahl, Gordon B., Katrine V. Løken, Magne Mogstad, and Kari Vea Salvanes. 2016. "What Is the Case for Paid Maternity Leave?" *Review of Economics and Statistics* 98 (4):655-670. doi: 10.1162/REST_a_00602.

Danzer, Natalia, and Victor Lavy. 2018. "Paid parental leave and children's schooling outcomes." *Economic Journal* 128 (608):81-117.

De Chaisemartin, Clément, and Xavier d'Haultfoeuille. 2020. "Two-way fixed effects estimators with heterogeneous treatment effects." *American Economic Review* 110 (9):2964-96.

Deming, David J. 2014. "Using school choice lotteries to test measures of school effectiveness." *American Economic Review* 104 (5):406-11.

Deza, Monica, Johanna Catherine Maclean, and Keisha Solomon. 2022. "Local access to mental healthcare and crime." *Journal of Urban Economics* 129:103410.

Dobbie, Will, and Roland G Fryer. 2013. The medium-term impacts of high-achieving charter schools on non-test score outcomes. National Bureau of Economic Research.

Donald, Stephen G, and Kevin Lang. 2007. "Inference with difference-in-differences and other panel data." *Review of Economics and Statistics* 89 (2):221-233.

Dustmann, Christian, and Uta Schönberg. 2012. "Expansions in maternity leave coverage and children's long-term outcomes." *American Economic Journal: Applied Economics* 4 (3):190-224.

Eble, Connie C. 1996. *Slang & sociability: In-group language among college students*: Univ of North Carolina Press.

Ely, Danielle M, and Anne K Driscoll. 2020. "Infant Mortality in the United States, 2018: Data From the Period Linked Birth/Infant Death File." *National Vital Statistics Reports: From the Centers for Disease Control and Prevention, National Center for Health Statistics, National Vital Statistics System* 69 (7):1-18.

Felmlee, Diane, Paulina Inara Rodis, and Amy Zhang. 2020. "Sexist slurs: Reinforcing feminine stereotypes online." *Sex roles* 83 (1):16-28.

Ferman, Bruno, and Cristine Pinto. 2019. "Inference in differences-in-differences with few treated groups and heteroskedasticity." *Review of Economics and Statistics* 101 (3):452-467.

Figlio, David, and Cassandra Hart. 2014. "Competitive effects of means-tested school vouchers." *American Economic Journal: Applied Economics* 6 (1):133-56.

Furgeson, Joshua, Brian Gill, Joshua Haimson, Alexandra Killewald, Moira McCullough, Ira Nichols-Barrer, Natalya Verbitsky-Savitz, Bing-ru Teh, Melissa Bowen, and Allison Demeritt. 2012. "Charter-school management organizations: Diverse strategies and diverse student impacts." *Mathematica Policy Research, Inc.*

Galiani, Sebastian, and Brian Quistorff. 2017. "The synth_runner package: Utilities to automate synthetic control estimation using synth." *The Stata Journal* 17 (4):834-849.

Gibbons, Stephen, Stephan Heblich, and Christopher Timmins. 2021. "Market tremors: Shale gas exploration, earthquakes, and their impact on house prices." *Journal of Urban Economics* 122:103313.
Gibbons, Stephen, Henry Overman, and Matti Sarvimäki. 2021. "The local economic impacts of regeneration projects: Evidence from UK's single regeneration budget." *Journal of Urban Economics* 122:103315.
Glick, P, and ST Fiske. 1996. "The ambivalent sexism inventory: Dif."
Glomm, Gerhard, Douglas N Harris, and Te-Fen Lo. 2005. "Charter school location." *Economics of Education Review* 24 (4):451-457.
Goldhaber, Dan D, and Eric R Eide. 2003. "Methodological thoughts on measuring the impact of private sector competition on the educational marketplace." *Educational Evaluation and Policy Analysis* 25 (2):217-232.
Goodman-Bacon, Andrew. 2021. "Difference-in-Differences with Variation in Treatment Timing." *Journal of Econometrics*.
Griffith, David. 2019. Rising Tide: Charter School Market Share and Student Achievement. Washington, DC: Thomas B. Fordham Institute.
Hall, Jonathan D, Craig Palsson, and Joseph Price. 2018. "Is Uber a substitute or complement for public transit?" *Journal of urban economics* 108:36-50.
Han, Eunice S, and Jeffrey Keefe. 2020. "The Impact of Charter School Competition on Student Achievement of Traditional Public Schools after 25 Years: Evidence from National District-level Panel Data." *Journal of School Choice*:1-39.
Hanushek, Eric A, John F Kain, Steven G Rivkin, and Gregory F Branch. 2007. "Charter school quality and parental decision making with school choice." *Journal of public economics* 91 (5-6):823-848.
Harris, Douglas N. 2020. *Charter school city: What the end of traditional public schools in New Orleans Means for American education*: University of Chicago Press.
Harris, Douglas N, and Matthew Larsen. forthcoming. "The Effects of the New Orleans Post-Katrina Market-Based School Reforms on Medium-Term Student Outcomes." *Journal of Human Resources*
Harris, Douglas N, and Valentina Martinez-Pabon. 2021. "Does “Creative Destruction” Apply to Schools? a National Analysis of Closure and Takeover of Traditional Public, Charter, and Private Schools." *Tulane University working paper*.
Heft-Neal, Sam, Jennifer Burney, Eran Bendavid, Kara Voss, and Marshall Burke. 2019. Air pollution and infant mortality: Evidence from Saharan dust. National Bureau of Economic Research.
Heminway, Joan MacLeod. 2019. "Me, Too and# MeToo: Women in Congress and the Boardroom." *Geo. Wash. L. Rev.* 87:1079.
Ho, Andrew D. 2020. What Is the Stanford Education Data Archive Teaching Us About National Educational Achievement? : SAGE Publications Sage CA: Los Angeles, CA.
Hoekstra, Mark, and Brittany Street. 2021. "The Effect of Own-Gender Jurors on Conviction Rates." *The Journal of Law and Economics* 64 (3):513-537.
Hoxby, Caroline Minter. 2003. "School choice and school productivity. Could school choice be a tide that lifts all boats?" In *The economics of school choice*, 287-342. University of Chicago Press.

Hoxby, Caroline Minter, and Jonah E Rockoff. 2004. *The impact of charter schools on student achievement*: Department of Economics, Harvard University Cambridge, MA.
Huang, Dashan, Fuwei Jiang, Jun Tu, and Guofu Zhou. 2015. "Investor sentiment aligned: A powerful predictor of stock returns." *The Review of Financial Studies* 28 (3):791-837.
Huang, Rui, and Muzhe Yang. 2015. "Paid maternity leave and breastfeeding practice before and after California's implementation of the nation's first paid family leave program." *Economics & Human Biology* 16:45-59.
Imberman, Scott A. 2011. "The effect of charter schools on achievement and behavior of public school students." *Journal of Public Economics* 95 (7-8):850-863.
Iyengar, Radha. 2009. "Does the certainty of arrest reduce domestic violence? Evidence from mandatory and recommended arrest laws." *Journal of public Economics* 93 (1-2):85-98.
Iyer, Lakshmi, Anandi Mani, Prachi Mishra, and Petia Topalova. 2012. "The power of political voice: women's political representation and crime in India." *American Economic Journal: Applied Economics* 4 (4):165-93.
Jabbar, Huriya. 2015. "“Every kid is money” market-like competition and school leader strategies in New Orleans." *Educational Evaluation and Policy Analysis* 37 (4):638-659.
James, Deborah. 1998. "Gender-linked derogatory terms and their use by women and men." *American Speech* 73 (4):399-420.
Kelly, Bryan, and Seth Pruitt. 2013. "Market expectations in the cross-section of present values." *The Journal of Finance* 68 (5):1721-1756.
Kelly, Bryan, and Seth Pruitt. 2015. "The three-pass regression filter: A new approach to forecasting using many predictors." *Journal of Econometrics* 186 (2):294-316.
Kiross, Girmay Tsegay, Catherine Chojenta, Daniel Barker, and Deborah Loxton. 2020. "The effects of health expenditure on infant mortality in sub-Saharan Africa: evidence from panel data analysis." *Health Economics Review* 10 (1):1-9.
Kniesner, Thomas J., and W. Kip Viscusi. 2019. The Value of a Statistical Life. Oxford University Press.
Ladd, Helen F, and John D Singleton. 2020. "The fiscal externalities of charter schools: Evidence from North Carolina." *Education Finance and Policy* 15 (1):191-208.
Lee, Tiane L, Susan T Fiske, and Peter Glick. 2010. "Next gen ambivalent sexism: Converging correlates, causality in context, and converse causality, an introduction to the special issue." *Sex Roles* 62 (7):395-404.
Levy, Roee, and Martin Mattsson. 2022. "The effects of social movements: Evidence from# MeToo." *Available at SSRN 3496903*.
Lichtman-Sadot, Shirlee, and Neryvia Pillay Bell. 2017. "Child health in elementary school following California's paid family leave program." *Journal of Policy Analysis and Management* 36 (4):790-827.
Lincove, Jane Arnold, Deven Carlson, and Nathan Barrett. 2019. "The Effects of School Closure on the Teacher Labor Market: Evidence from Portfolio Management in New Orleans." *Working Paper*.
Linick, Matthew Allen. 2014. "Measuring Competition: Inconsistent definitions, inconsistent results." *education policy analysis archives* 22 (16):n16.
Lins, Karl V, Lukas Roth, Henri Servaes, and Ane Tamayo. 2020. "Gender, culture, and firm value: Evidence from the Harvey Weinstein scandal and the# MeToo movement."

Liu, Qian, and Oskar Nordstrom Skans. 2010. "The duration of paid parental leave and children's scholastic performance." *BE Journal of Economic Analysis & Policy* 10 (1).
Loeb, Susanna, Jon Valant, and Matt Kasman. 2011. "Increasing choice in the market for schools: Recent reforms and their effects on student achievement." *National Tax Journal* 64 (1):141.
Lubienski, Christopher. 2007. "Marketing schools: Consumer goods and competitive incentives for consumer information." *Education and Urban Society* 40 (1):118-141.
Masser, Barbara, G Tendayi Viki, and Clair Power. 2006. "Hostile sexism and rape proclivity amongst men." *Sex Roles* 54 (7):565-574.
Meinhofer, Angélica, Allison E Witman, Jesse M Hinde, and Kosali Simon. 2021. "Marijuana liberalization policies and perinatal health." *Journal of Health Economics* 80:102537.
Mettetal, Elizabeth. 2019. "Irrigation dams, water and infant mortality: Evidence from South Africa." *Journal of Development Economics* 138:17-40.
Miller, Amalia R, and Carmit Segal. 2019. "Do female officers improve law enforcement quality? Effects on crime reporting and domestic violence." *The Review of Economic Studies* 86 (5):2220-2247.
Modrek, Sepideh, and Bozhidar Chakalov. 2019. "The# MeToo movement in the United States: text analysis of early twitter conversations." *Journal of medical Internet research* 21 (9):e13837.
NCHS, NVSS. 2008. User guide to the natality public use file. In *National Center for Health Statistics*.
NCHS, NVSS. 2020. "Linked Birth and Infant Death Data 2000–2008." *National Center for Health Statistics*.
Ness, Debra L., Vicki Shabo, and Sarah Fleisch Fink. 2016. Expecting Better: A State-by-State Analysis of Laws that Help New Parents. Washington, DC: National Partnership for Women & Families.
Ni, Yongmei. 2009. "The impact of charter schools on the efficiency of traditional public schools: Evidence from Michigan." *Economics of Education Review* 28 (5):571-584.
Oster, Emily. 2019. "Unobservable selection and coefficient stability: Theory and evidence." *Journal of Business & Economic Statistics* 37 (2):187-204.
Pac, Jessica E, Ann P Bartel, Christopher J Ruhm, and Jane Waldfogel. 2019. Paid family leave and breastfeeding: Evidence from California. National Bureau of Economic Research.
Pihl, Ariel Marek, and Gaetano Basso. 2019. "Did California Paid Family Leave Impact Infant Health?" *Journal of Policy Analysis and Management* 38 (1):155-180.
Planty, Michael, Lynn Langton, Christopher Krebs, Marcus Berzofsky, and Hope Smiley-McDonald. 2013. *Female victims of sexual violence, 1994-2010*: US Department of Justice, Office of Justice Programs, Bureau of Justice ….
Pongou, Roland. 2013. "Why is infant mortality higher in boys than in girls? A new hypothesis based on preconception environment and evidence from a large sample of twins." *Demography* 50 (2):421-444.
Raub, Amy, Arijit Nandi, Alison Earle, Nicolas De Guzman Chorny, Elizabeth Wong, Paul Chung, Priya Batra, Adam Schickedanz, Bijetri Bose, and Judy Jou. 2018. "Paid parental leave: A detailed look at approaches across OECD countries." *WORLD Policy Analysis Center*.

Reardon, Sean F, John P Papay, Tara Kilbride, Katherine O Strunk, Joshua Cowen, Lily An, and Kate Donohue. 2019. "Can Repeated Aggregate Cross-Sectional Data Be Used to Measure Average Student Learning Rates? A Validation Study of Learning Rate Measures in the Stanford Education Data Archive. CEPA Working Paper No. 19-08." *Stanford Center for Education Policy Analysis*.

Ridley, Matthew, and Camille Terrier. 2018. Fiscal and education spillovers from charter school expansion. National Bureau of Economic Research.

Rossin-Slater, Maya. 2011. "The effects of maternity leave on children's birth and infant health outcomes in the United States." *Journal of Health Economics* 30 (2):221-239.

Rossin-Slater, Maya, Christopher J Ruhm, and Jane Waldfogel. 2013. "The effects of California's paid family leave program on mothers' leave-taking and subsequent labor market outcomes." *Journal of Policy Analysis and Management* 32 (2):224-245.

Ruhm, Christopher J. 2000. "Parental leave and child health." *Journal of Health Economics* 19 (6):931-960.

Sass, Tim R. 2006. "Charter schools and student achievement in Florida." *Education Finance and Policy* 1 (1):91-122.

Siliverstovs, Boriss, and Daniel S Wochner. 2018. "Google Trends and reality: Do the proportions match?: Appraising the informational value of online search behavior: Evidence from Swiss tourism regions." *Journal of Economic Behavior & Organization* 145:1-23.

Stearns, Jenna. 2015. "The effects of paid maternity leave: Evidence from Temporary Disability Insurance." *Journal of Health Economics* 43:85-102.

Steel, Emily, and Michael S Schmidt. 2017. "Bill O'Reilly settled new harassment claim, then Fox renewed his contract." *New York Times*.

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

Sun, Liyang, and Sarah Abraham. 2021. "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects." *Journal of Econometrics* 225 (2):175-199.

Sutton, Laurel. 1995. "A. 1995. Bitches and Skanky Hobags: The Place of Women in Contemporary Slang." *Gender Articulated: Language and the Socially Constructed Self*.

Tanaka, Sakiko. 2005. "Parental leave and child health across OECD countries." *Economic Journal* 115 (501):F7-F28.

Tanaka, Shinsuke. 2015. "Environmental regulations on air pollution in China and their impact on infant mortality." *Journal of Health Economics* 42:90-103.

Toma, Eugenia F, Ron Zimmer, and John T Jones. 2006. "Beyond achievement: Enrollment consequences of charter schools in Michigan." *Advances in Applied Microeconomics* 14 (2):241-255.

Wong, Mitchell D, Karen M Coller, Rebecca N Dudovitz, David P Kennedy, Richard Buddin, Martin F Shapiro, Sheryl H Kataoka, Arleen F Brown, Chi-Hong Tseng, and Peter Bergman. 2014. "Successful schools and risky behaviors among low-income adolescents." *Pediatrics* 134 (2):e389-e396.

Ziebarth, Todd. 2020. "Measuring up to the model: A ranking of state public charter school laws." *National Alliance for Public Charter Schools*.

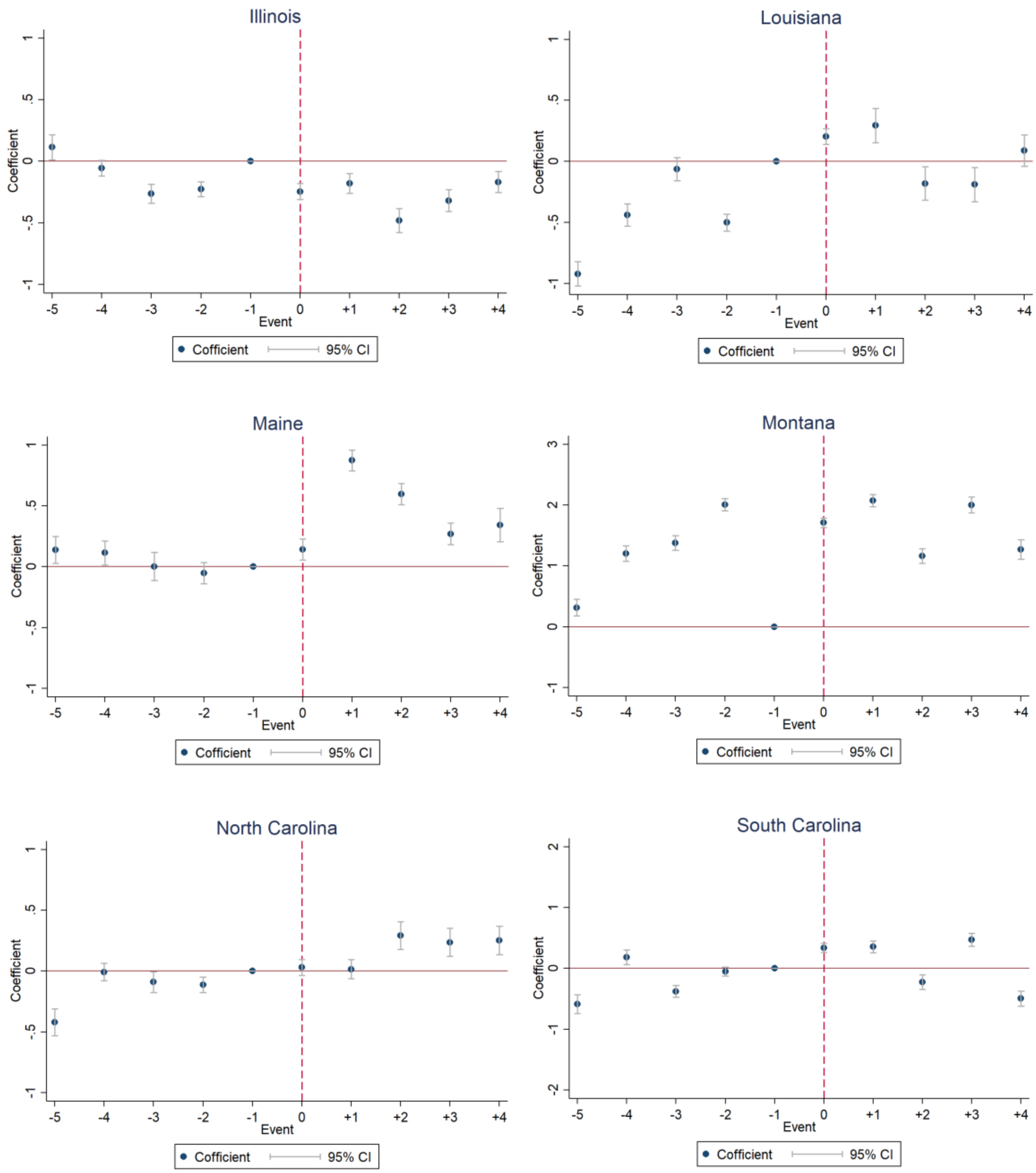
Zimmer, Ron, and Richard Buddin. 2009. "Is charter school competition in California improving the performance of traditional public schools?" *Public Administration Review* 69 (5):831-845.

Zimmer, Ron, Brian Gill, Jonathon Attridge, and Kaitlin Obenauf. 2014. "Charter school authorizers and student achievement." *Education Finance and Policy* 9 (1):59-85.

APPENDICES

Chapter 1

Figure A1 Event study estimates of selected placebo tests



Notes: This figure displays coefficients from event study regressions of selected placebo tests with F-P p-values less than 0.1. Event time is a dummy of the year(s) of leads or lags since the CA-PFL is effective, for example, the event time 0 is a dummy of the year PFL effective (July 2004 to June 2005).

Table A1 Effects of CA-PFL for all plurality

	(1)	(2)	(3)
CA*Post	-0.147	-0.161	-0.137
P-value	0.000	0.000	0.000
F-P p-value	0.175	0.098	0.016
R-squared	0.474	0.477	0.479
Observations	5,508	5,508	5,508
State FE, Time FE	Y	Y	Y
Birth control	N	Y	Y
Maternal control	N	N	Y

Notes: The table presents the DD estimates of the effects of the CA-PFL on PNMR for all plurality. See notes to table 2 for other details.

Table A2 List of states of two alternative comparison groups

Top 25 family-friendly states			Bottom 25 family-friendly states		
Rank	State	Grade	Rank	State	Grade
2	District of Columbia	140	27	Florida	20
3	New York	135	27	Iowa	20
4	Rhode Island	125	27	Kansas	20
5	Connecticut	120	27	New Hampshire	20
6	Hawaii	110	27	North Carolina	20
7	New Jersey	100	27	Ohio	20
8	Oregon	95	27	Virginia	20
9	Vermont	85	34	Indiana	15
10	Illinois	70	34	New Mexico	15
10	Massachusetts	70	34	North Dakota	15
12	Minnesota	65	37	Kentucky	10
12	Washington	65	37	Pennsylvania	10
14	Maine	60	37	Texas	10
15	Colorado	50	40	Alabama	0
16	Louisiana	45	40	Arizona	0
16	Wisconsin	45	40	Georgia	0
18	Maryland	40	40	Idaho	0
19	Arkansas	35	40	Michigan	0
20	Alaska	30	40	Mississippi	0
20	Delaware	30	40	Missouri	0
20	Montana	30	40	Nevada	0
20	Nebraska	30	40	Oklahoma	0
20	Utah	30	40	South Carolina	0
25	Tennessee	25	40	South Dakota	0
25	West Virginia	25	40	Wyoming	0

Note: This table presents grades of state policies that support expecting and new parents just before and soon after the arrival of a new child. California ranked first with a grade of 155. See Ness, Shabo, and Fink (2016) for detailed methodologies that they used to calculate the grades. The top 25 family-friendly states (other than California) and the bottom 25 family-friendly states are two alternative comparison groups used in this article.

Table A3 Neonatal mortality rate

	(1)	(2)	(3)
CA*Post	-0.049	-0.059	-0.027
P-value	(0.291)	(0.231)	(0.582)
F-P p-value	[0.701]	[0.641]	[0.832]
R-squared	0.602	0.604	0.606
Observations	5,508	5,508	5,508
State FE, Time FE	Y	Y	Y
Birth control	N	Y	Y
Maternal control	N	N	Y

Notes: The table presents the DD estimates of the effects of the CA-PFL on neonatal mortality rate. The neonatal mortality rate is the number of deaths during the first 28 days of life per 1,000 live births. See notes to table 2 for other details.

Table A4 Robustness check: exclude IL, LA, ME, MT, NC, and SC

	(1)	(2)	(3)
CA*Post	-0.143	-0.153	-0.129
P-value	0.000	0.000	0.000
F-P p-value	0.044	0.002	0.000
R-squared	0.453	0.456	0.457
Observations	4,860	4,860	4,860
State FE, Time FE	Y	Y	Y
Birth control	N	Y	Y
Maternal control	N	N	Y

Notes: The table presents the DD estimates of the effects of the CA-PFL on PNMR excluding states IL, LA, ME, MT, NC, and SC from the analysis. See notes to table 2 for other details.

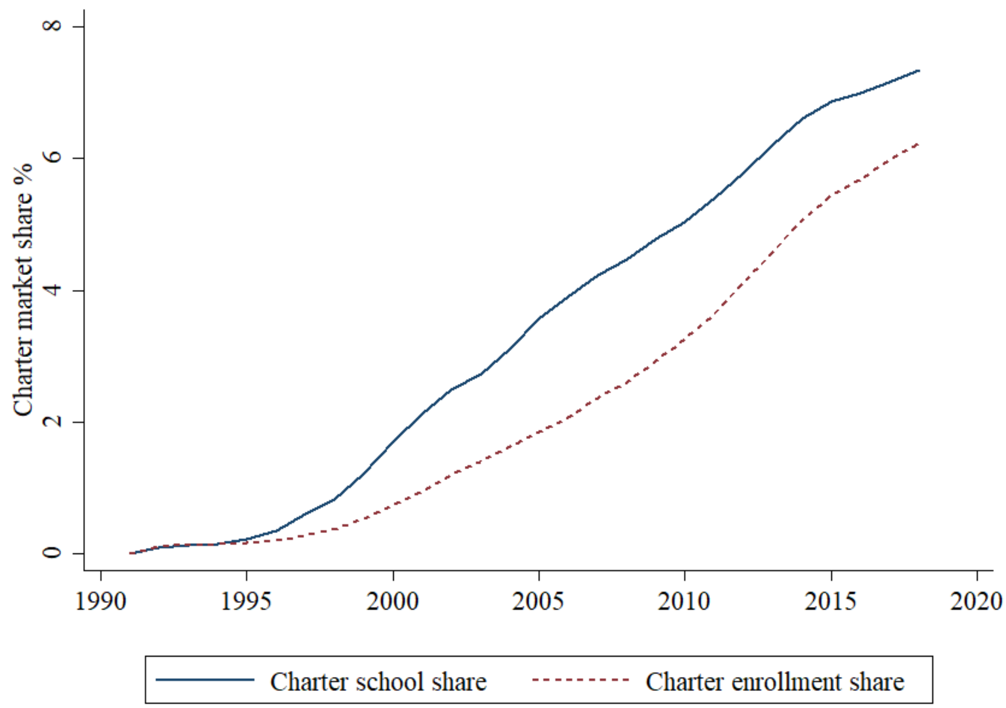
Table A5 Robustness check: exclude the year 2004

	(1)	(2)	(3)
CA*Post	-0.162	-0.171	-0.142
P-value	0.000	0.000	0.000
F-P p-value	0.108	0.066	0.005
R-squared	0.450	0.453	0.455
Observations	4,896	4,896	4,896
State FE, Time FE	Y	Y	Y
Birth control	N	Y	Y
Maternal control	N	N	Y

Notes: The table presents the DD estimates of the effects of the CA-PFL on PNMR excluding the year 2004 from the analysis. See notes to table 2 for other details.

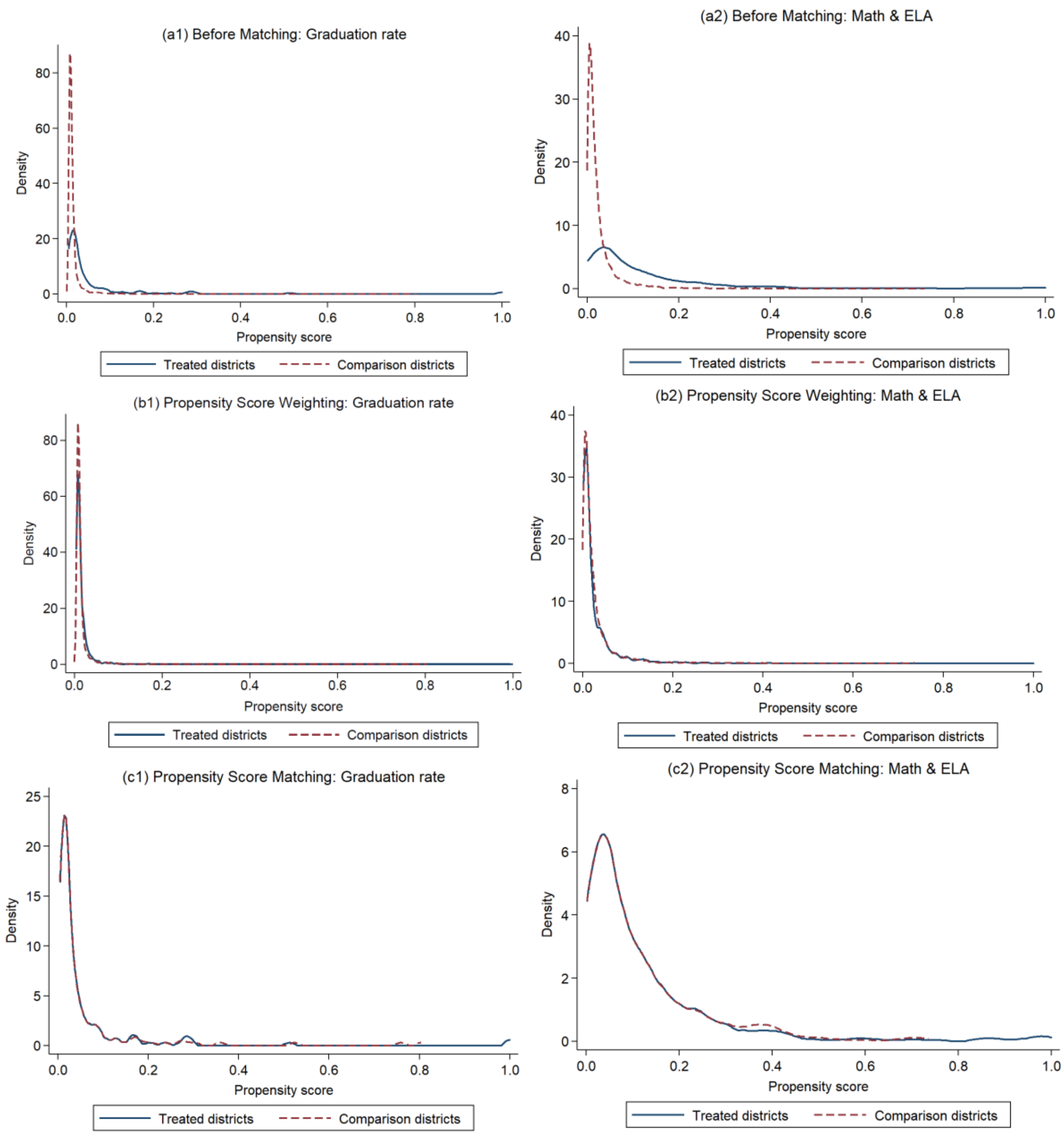
Chapter 2

Figure A1 Trends in charter school share and charter enrollment share



Notes: This figure plots the trends in charter school share (in the dashed line) and charter enrollment share (in the solid line).

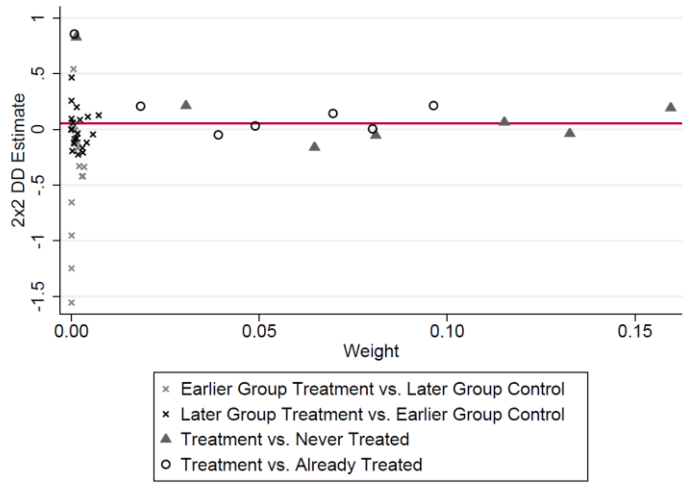
Figure A2 Density of propensity score before and after matching



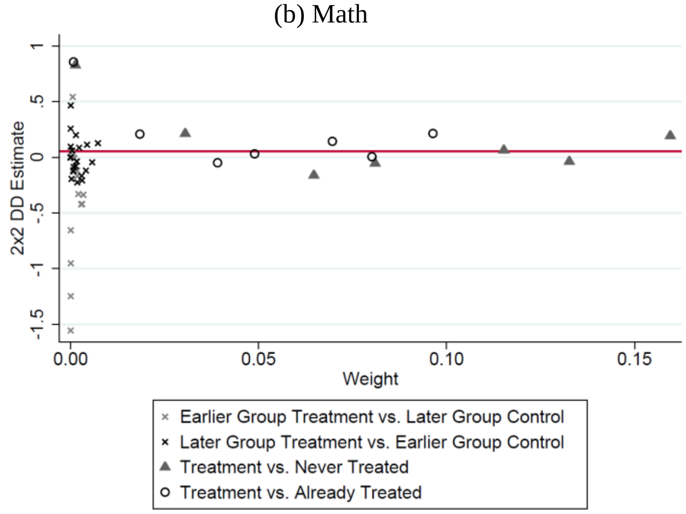
Notes: This figure plots the density of the propensity score of covariates of treated districts (in the solid line) and comparison districts (in the dashed line).

Figure A3 Goodman-Bacon (2021) Decomposition

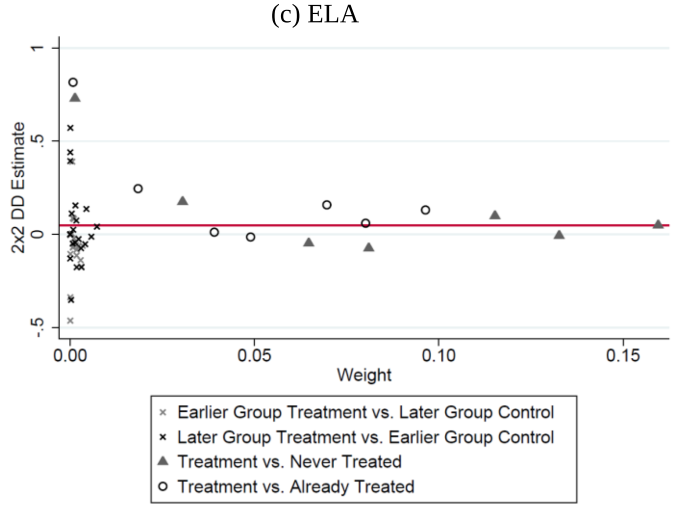
(a) Graduation rate



(b) Math

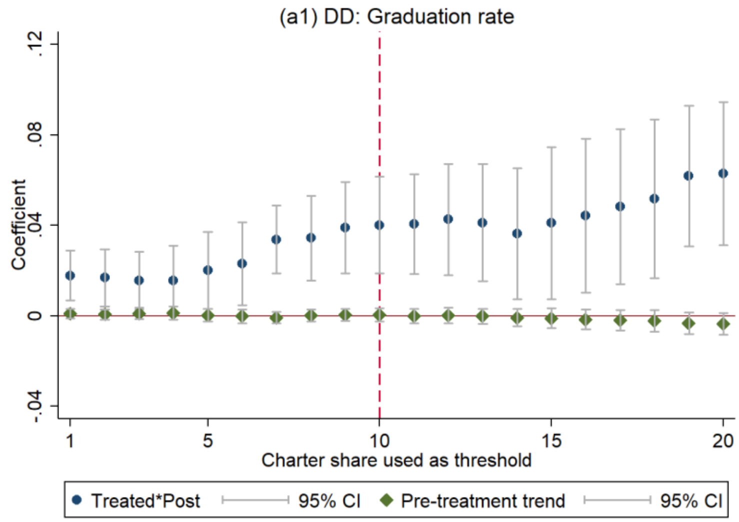


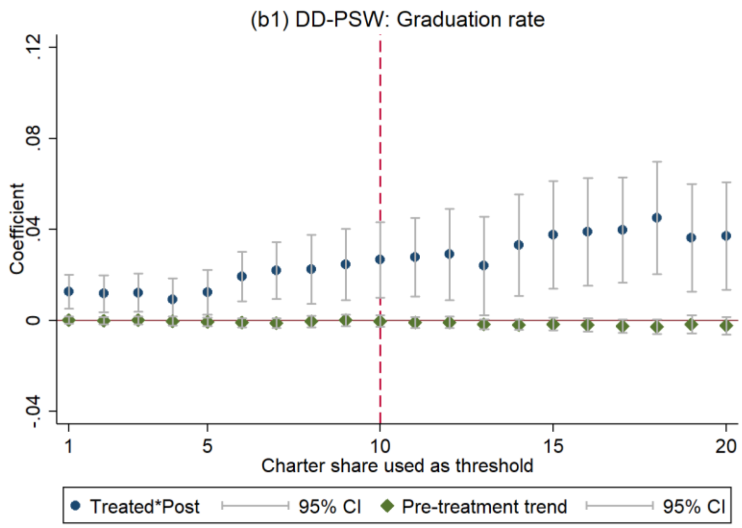
(c) ELA

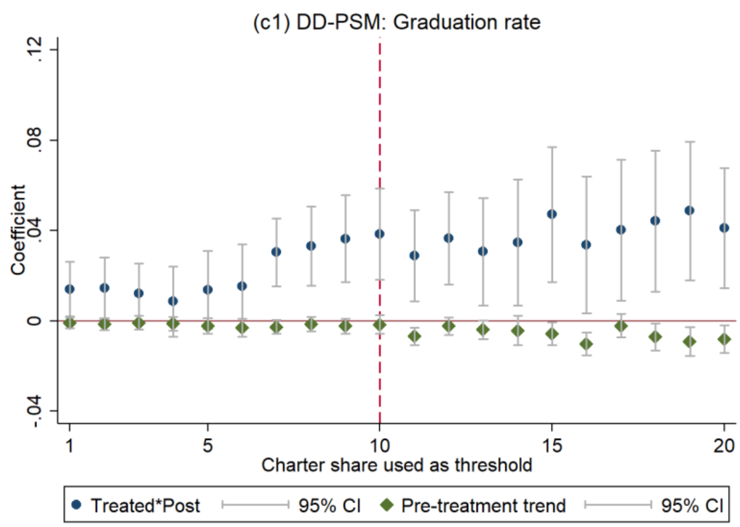


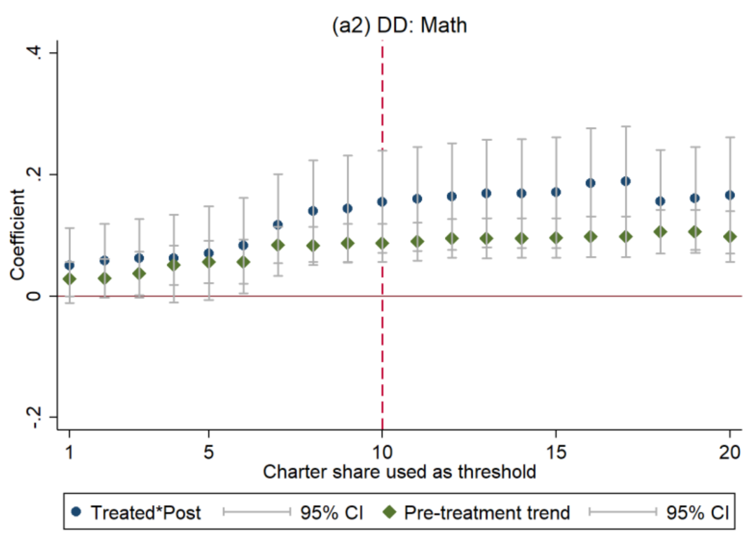
Notes: The above figures document the Goodman-Bacon decomposition into a series of 2×2 difference-in-differences models depending on the type of comparison unit.

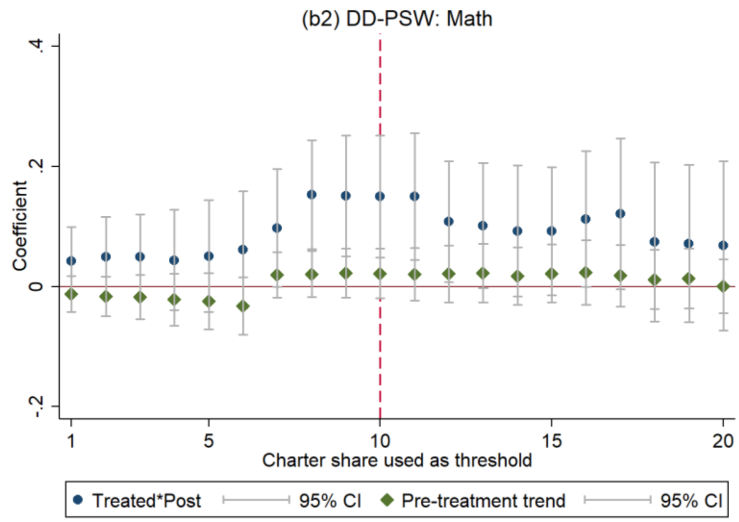
Figure A4 Plots of estimates using 1% to 20% as the threshold of treated districts without New Orleans

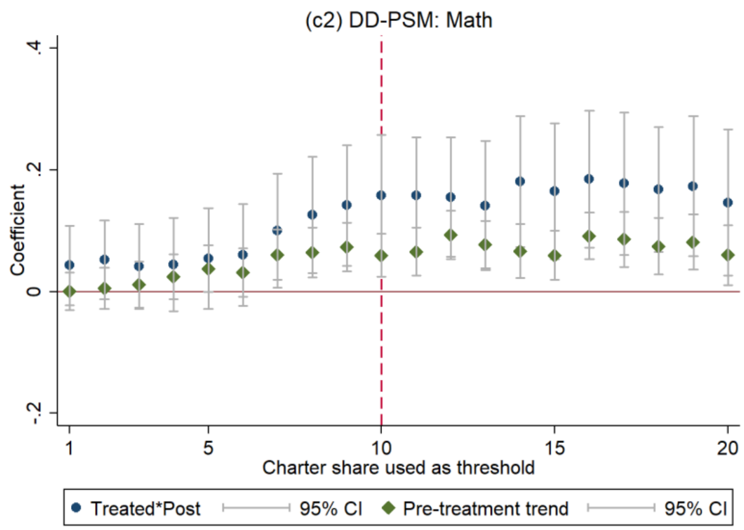


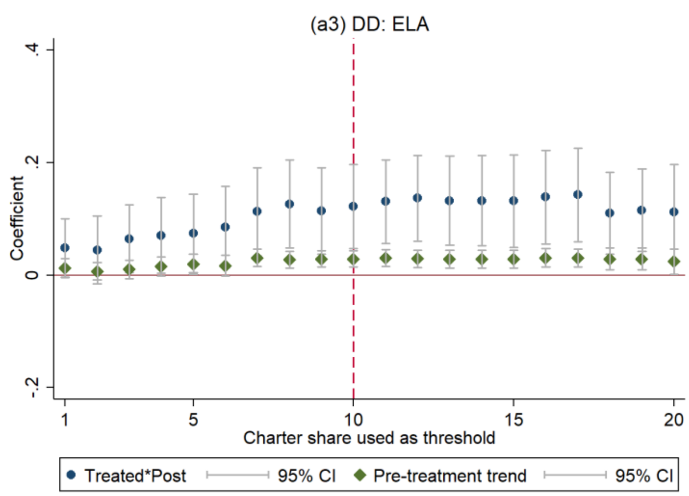

(b1) DD-PSW: Graduation rate
Coefficient
-.04
0
.04
.08
.12
1
5
10
15
20
Charter share used as threshold
Treated*Post
95% CI
Pre-treatment trend
95% CI

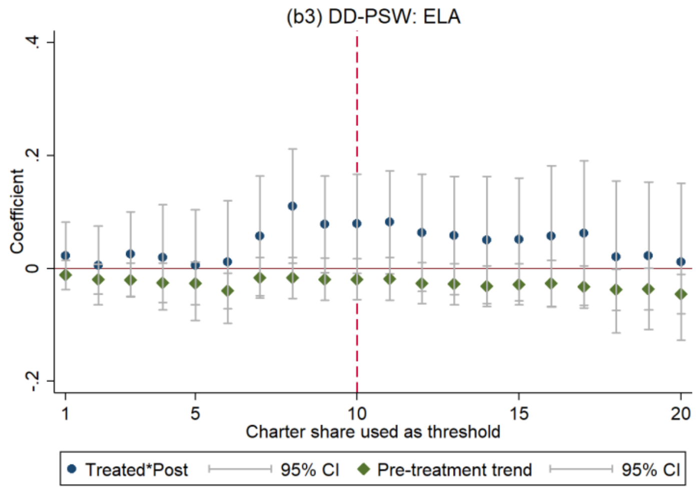

(c1) DD-PSM: Graduation rate
Coefficient
-.04
0
.04
.08
.12
1
5
10
15
20
Charter share used as threshold
Treated*Post
95% CI
Pre-treatment trend
95% CI

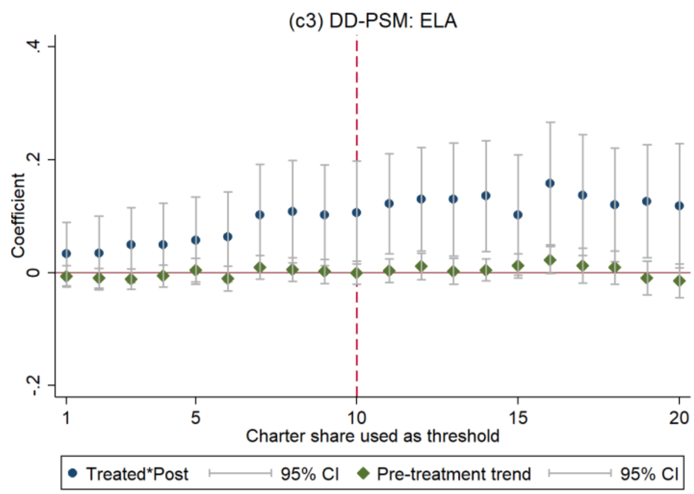

(a2) DD: Math
Coefficient
-.2
0
.2
.4
1
5
10
15
20
Charter share used as threshold
Treated*Post
95% CI
Pre-treatment trend
95% CI


(b2) DD-PSW: Math
Coefficient
Charter share used as threshold
Treated*Post
95% CI
Pre-treatment trend
95% CI


(c2) DD-PSM: Math
Coefficient
Charter share used as threshold
Treated*Post
95% CI
Pre-treatment trend
95% CI


(a3) DD: ELA
Coefficient
Charter share used as threshold
Treated*Post
95% CI
Pre-treatment trend
95% CI

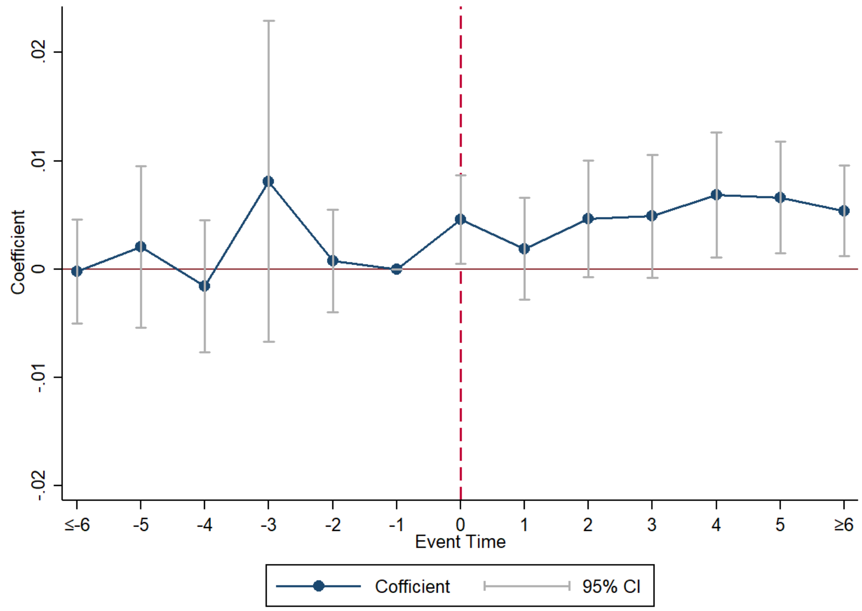




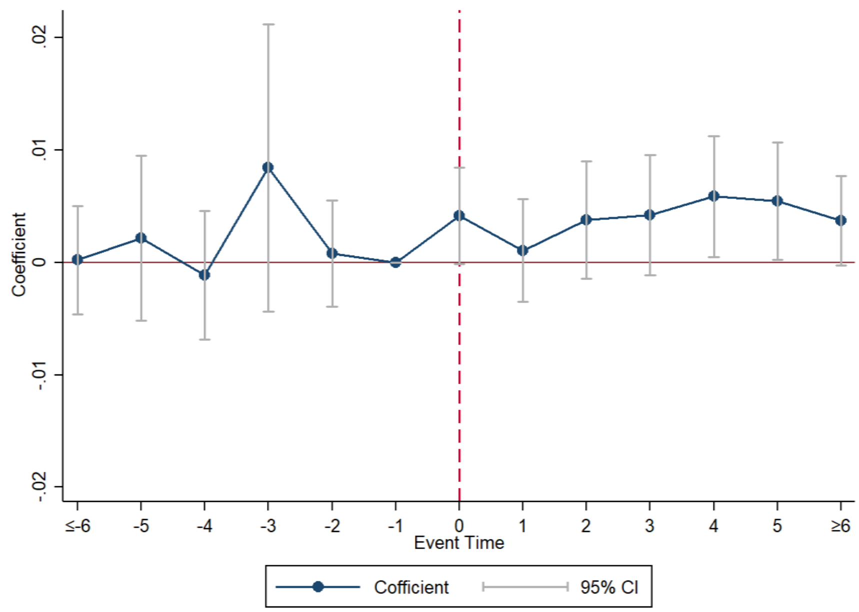
Notes: This figure plots the estimates (without New Orleans school district) in graduation rate, Math & ELA using the charter enrollment share of 1- 20% as the threshold of treated districts.

Figure A5 Event study results of TPS closure

(a) Event study



(b) IW estimators



Notes: This figure presents results of event study and IW estimators for TPS closure. The event time zero is the first year of charter entr

Table A1 Summary Statistics (unweighted)

Sample	Graduation rate (1995-2010)						Math & ELA (2009-2016)					
	ALL	DD & DD-PSW		DD-PSM		Fixed Effects	ALL	DD & DD-PSW		DD-PSM		Fixed Effects
		Treated	Control	Treated	Control			Treated	Control	Treated	Control	
Graduation rate	0.81	0.75	0.812	0.75	0.79	0.75	NA	NA	NA	NA	NA	NA
Math	NA	NA	NA	NA	NA	NA	0.07	-0.47	0.09	-0.47	-0.37	-0.38
ELA	NA	NA	NA	NA	NA	NA	0.10	-0.39	0.12	-0.39	-0.32	-0.35
White	81%	65%	82%	65%	64%	66%	75%	59%	76%	59%	58%	57%
Black	7%	8%	7%	8%	8%	11%	9%	10%	9%	10%	10%	13%
Hispanic	7%	18%	7%	18%	16%	17%	12%	26%	11%	26%	26%	25%
FRL	30%	35%	30%	35%	34%	31%	48%	54%	47%	54%	56%	55%
Special education	13%	12%	13%	12%	13%	13%	14%	12%	14%	12%	12%	13%
Ages 5–17 population	16%	18%	16%	18%	17%	17%	18%	20%	18%	20%	21%	20%
Ages 5–17 in poverty	19%	19%	19%	19%	20%	19%	17%	15%	17%	15%	16%	16%
Urban	5%	16%	5%	16%	14%	23%	6%	16%	5%	16%	16%	25%
Suburb	22%	22%	22%	22%	20%	32%	28%	28%	27%	28%	29%	31%
Town	17%	19%	17%	19%	22%	19%	22%	28%	22%	28%	23%	25%
Rural	56%	43%	57%	43%	44%	25%	45%	32%	47%	32%	36%	23%
Revenue per student	9,276	8,975	9,307	8,975	9,194	8,621	12,882	11,235	12,969	11,235	10,899	11,360
Expenditure per student	9,354	9,068	9,382	9,068	9,312	8,759	12,794	11,277	12,873	11,277	10,933	11,404
Teacher salary	70,445	74,705	70,214	74,705	73,025	75,372	92,218	94,213	92,005	94,213	94,147	95,975
Student teacher ratio	15	17	15	17	16	17	15	18	15	18	18	18
No. magnet school	0	3	0	3	1	2	0	2	0	2	1	2
No. public schools	8	29	7	29	18	33	8	26	7	26	17	31
Enrollment	4,083	18,827	3,236	18,827	10,888	22,219	4,221	15,664	3,373	15,664	10,515	19,184
Observation	144,219	2,193	137,780	2,193	2,199	6,439	410,077	10,109	388,064	10,109	11,073	22,013
N (district)	9,278	142	8,859	142	142	416	10,439	297	9,832	297	297	607

Notes: This table presents unweighted summary statistics.

Table A2 Top 20 districts with largest charter enrollment share

School District	State	NAPCS Charter Enrollment	NAPCS Total Enrollment	NAPCS Enrollment Share	NLSD Charter Enrollment	NLSD Total Enrollment	NLSD Enrollment Share	Sample Graduation rate	Sample Test score
Orleans Parish School District	LA	46,932	49,646	95%	48,495	51,100	95%	Y	Y
Gary Community School Corporation	IN	5,060	10,288	49%	5,060	10,288	49%	Y	Y
Queen Creek Unified District	AZ	6,776	13,858	49%	5,070	12,166	42%	N	Y
District of Columbia Public Schools	DC	43,393	91,528	47%	38,696	86,330	45%	Y	Y
Detroit Public Schools	MI	38,667	83,504	46%	37,235	87,045	43%	Y	Y
Kansas City Public Schools	MO	11,420	26,630	43%	12,602	27,769	45%	Y	Y
Southfield Public School District	MI	4,543	10,697	42%	4,543	10,674	43%	N	Y
Inglewood Unified School District	CA	5,193	13,594	38%	5,453	13,854	39%	Y	Y
Camden City School District	NJ	4,731	12,672	37%	4,892	12,616	39%	Y	Y
Indianapolis Public Schools	IN	15,244	42,874	36%	15,466	42,383	36%	N	Y
Franklin-McKinley School District	CA	3,866	11,152	35%	3,305	10,591	31%	N	Y
Dayton City School District	OH	6,652	19,745	34%	6,828	19,850	34%	Y	Y
Natomas Unified School District	CA	4,952	14,880	33%	4,952	14,880	33%	N	Y
Philadelphia City School District	PA	64,393	195,631	33%	64,970	192,172	34%	Y	Y
Newark City School District	NJ	17,501	53,215	33%	17,204	52,917	33%	Y	Y
Alum Rock Union Elem School District	CA	4,623	14,265	32%	5,089	14,731	35%	N	Y
St. Louis City School District	MO	11,082	34,936	32%	11,022	33,958	32%	Y	Y
Cleveland Municipal School District	OH	16,352	54,641	30%	20,076	58,301	34%	Y	Y
San Antonio Independent School District	TX	18,515	62,119	30%	17,979	58,901	31%	Y	Y
Oakland Unified School District	CA	18,502	52,457	30%	16,070	53,018	30%	Y	Y

Notes: This table compares the top 20 districts (with the largest charter enrollment share among districts with at least 10,000 total students in the 2018 spring year) from a report of the National Alliance for Public Charter Schools with data from NLSD. Source: A Growing Movement: America's Largest Charter School Communities, Thirteenth Edition, January 2019. Please notes that this table only list districts with enrollment more than 10,000 students, there are still some other districts with enrollment less than 10,000 students but also have high charter enrollment share.

Table A3 Effects of charter entry on student outcomes (cluster at state level)

	(1)	(2)	(3)	(4)	(5)	(6)
	DD		DD-PSW		DD-PSM	
Panel A: Graduation rate						
Treated*Post	0.037***	0.038***	0.034***	0.028**	0.039***	0.031***
	[0.010]	[0.010]	[0.012]	[0.011]	[0.011]	[0.008]
R-squared	0.805	0.808	0.817	0.825	0.886	0.894
Observations	139,973	139,973	139,973	139,973	4,383	4,383
N (district)	9,001	9,001	9,001	9,001	284	284
Pre-treatment	0.000	0.001	-0.001	-0.001	-0.001	-0.002
trend	[0.001]	[0.001]	[0.002]	[0.001]	[0.002]	[0.001]
Panel B: Math						
Treated*Post	0.152**	0.155**	0.154***	0.148***	0.136**	0.157***
	[0.067]	[0.063]	[0.053]	[0.048]	[0.059]	[0.052]
R-squared	0.846	0.847	0.863	0.865	0.869	0.873
Observations	398,173	398,173	398,173	398,173	21,007	21,007
N (district)	10,129	10,129	10,129	10,129	594	594
Pre-treatment	0.085***	0.088***	0.023	0.023	0.074***	0.073***
trend	[0.022]	[0.019]	[0.030]	[0.024]	[0.020]	[0.015]
Panel C: ELA						
Treated*Post	0.116**	0.121**	0.075	0.077	0.075	0.106**
	[0.048]	[0.051]	[0.050]	[0.050]	[0.056]	[0.049]
R-squared	0.882	0.883	0.884	0.886	0.914	0.917
Observations	398,173	398,173	398,173	398,173	21,007	21,007
N (district)	10,129	10,129	10,129	10,129	594	594
Pre-treatment	0.026**	0.029***	-0.025	-0.019	0.009	0.011
trend	[0.010]	[0.008]	[0.018]	[0.013]	[0.012]	[0.008]
District, (grade) & year FE	Yes	Yes	Yes	Yes	Yes	Yes
District control	No	Yes	No	Yes	No	Yes

Notes: The table replicates Table 3 but clusters standard errors at the state level.

Table A4 Estimates of Propensity Score (Probit model)

Matching variables	Graduation rate sample	Test score sample
Log of enrollment	-0.176***	0.058
	[0.056]	[0.040]
Perc. of white	-0.895***	-0.729**
	[0.298]	[0.292]
Perc. of Black	-0.653*	-0.521
	[0.377]	[0.329]
Perc. of Hispanic	0.212	0.172
	[0.314]	[0.307]
Perc. of FRL	-0.292	0.882***
	[0.329]	[0.273]
Perc. of special edu.	1.022	-3.031***
	[0.859]	[0.803]
Perc. of poverty	-0.319	-2.388***
	[0.528]	[0.626]
Perc. of school age	2.765**	-10.270***
	[1.178]	[0.947]
Urban	0.329*	0.065
	[0.189]	[0.195]
Suburb	0.004	0.173
	[0.126]	[0.175]
Town	0.217**	0.388**
	[0.107]	[0.173]
Rural	0.000	0.333*
	[.]	[0.174]
Revenue per student	0.000	0.000
	[0.000]	[0.000]
Expenditure per student	0.000	0.000
	[0.000]	[0.000]
Student teacher ratio	0.097***	0.139***
	[0.016]	[0.013]
Teacher salary	0.000	-0.000***
	[0.000]	[0.000]
No. magnet school	0.000	-0.021**
	[.]	[0.009]
No. public schools	0.015***	0.015***
	[0.002]	[0.002]
Constant	-2.459***	-1.385***
	[0.622]	[0.517]
N(district)	9001	10129

Note: This table shows the results of Probit model to estimate the propensity score.

Table A5 Number of districts by year first charter initiated

Graduation rate (high school)		Math & ELA (grade 3-8)	
Year first charter started	Number of districts	Year first charter started	Number of districts
Before 1995	10	Before 2009	406
1995	11	2009	20
1996	12	2010	35
1997	11	2011	30
1998	18	2012	32
1999	23	2013	27
2000	46	2014	31
2001	40	2015	17
2002	42	2016	9
2003	30	Total	607
2004	19		
2005	27		
2006	30		
2007	31		
2008	20		
2009	25		
2010	21		
Total	416		

Notes: The table presents the number of districts by year first charter initiated.

Table A6 Effects of charter entry on student outcomes (include state time trends)

	(1)	(2)	(3)	(4)	(5)	(6)
	DD		DD-PSW		DD-PSM	
Panel A: Graduation rate						
Treated*Post	0.037***	0.036***	0.030***	0.024***	0.018	0.014
	[0.008]	[0.009]	[0.010]	[0.009]	[0.015]	[0.010]
R-squared	0.811	0.814	0.826	0.832	0.900	0.909
Observations	139,973	139,973	139,973	139,973	4,383	4,383
N (district)	9,001	9,001	9,001	9,001	284	284
Pre-treatment	0.000	0.000	-0.002	-0.001	-0.001	-0.001
trend	[0.001]	[0.001]	[0.001]	[0.001]	[0.002]	[0.002]
Panel B: Math						
Treated*Post	0.016	0.027	0.110**	0.110**	0.000	0.017
	[0.038]	[0.040]	[0.052]	[0.051]	[0.047]	[0.052]
R-squared	0.851	0.852	0.870	0.871	0.882	0.885
Observations	398,173	398,173	398,173	398,173	21,007	21,007
N (district)	10,129	10,129	10,129	10,129	594	594
Pre-treatment	0.032*	0.033**	-0.012	-0.006	0.001	0.007
trend	[0.018]	[0.016]	[0.016]	[0.014]	[0.019]	[0.017]
Panel C: ELA						
Treated*Post	0.033	0.039	0.067	0.065	0.032	0.045
	[0.032]	[0.032]	[0.046]	[0.043]	[0.035]	[0.037]
R-squared	0.885	0.886	0.888	0.889	0.921	0.924
Observations	398,173	398,173	398,173	398,173	21,007	21,007
N (district)	10,129	10,129	10,129	10,129	594	594
Pre-treatment	0.001	0.002	-0.045**	-0.039**	-0.010	-0.012
trend	[0.011]	[0.009]	[0.019]	[0.017]	[0.013]	[0.015]
District, (grade) & year FE	Yes	Yes	Yes	Yes	Yes	Yes
District control	No	Yes	No	Yes	No	Yes

Notes: The table replicates Table 3 but includes state specific linear time trends.

Table A7 Alternative treated group: 0-10 percent charter share

	(1)	(2)	(3)	(4)	(5)	(6)
	DD		DD-PSW		DD-PSM	
Panel A: Graduation rate						
Treated*Post	0.012**	0.009	0.018***	0.014***	0.012*	0.007
	[0.005]	[0.006]	[0.005]	[0.004]	[0.006]	[0.006]
R-squared	0.805	0.808	0.818	0.822	0.853	0.857
Observations	142,026	142,026	142,026	142,026	8,462	8,462
N (district)	9,133	9,133	9,133	9,133	548	548
Pre-treatment	0.001	0.000	0.000	0.000	0.000	-0.001
trend	[0.001]	[0.001]	[0.002]	[0.001]	[0.001]	[0.001]
Panel B: Math						
Treated*Post	0.008	0.006	0.029	-0.015	0.009	0.005
	[0.045]	[0.041]	[0.045]	[0.027]	[0.046]	[0.041]
R-squared	0.848	0.849	0.863	0.867	0.884	0.886
Observations	399,968	399,968	399,968	399,968	23,972	23,972
N (district)	10,142	10,142	10,142	10,142	620	620
Pre-treatment	0.029	0.032	-0.006	-0.020	0.022	0.022
trend	[0.032]	[0.032]	[0.019]	[0.015]	[0.033]	[0.034]
Panel C: ELA						
Treated*Post	0.044	0.036	0.030	0.006	0.023	0.015
	[0.035]	[0.030]	[0.038]	[0.032]	[0.036]	[0.031]
R-squared	0.884	0.885	0.886	0.888	0.912	0.914
Observations	399,968	399,968	399,968	399,968	23,972	23,972
N (district)	10,142	10,142	10,142	10,142	620	620
Pre-treatment	0.035	0.035*	0.009	0.005	0.028	0.026
trend	[0.023]	[0.020]	[0.015]	[0.014]	[0.024]	[0.021]
District, (grade) & year FE	Yes	Yes	Yes	Yes	Yes	Yes
District control	No	Yes	No	Yes	No	Yes

Notes: The table shows results using districts with charter enrollment share greater than 0 but less than 10 percent as the treated group. See notes of Table 3 for controls, sample weight, and clusters.

Table A8 Alternative measure of charter share: charter school share

	(1)	(2)	(3)	(4)	(5)	(6)
	DD		DD-PSW		DD-PSM	
Panel A: Graduation rate						
Treated*Post	0.016***	0.013**	0.016***	0.011**	0.014**	0.010*
	[0.005]	[0.005]	[0.004]	[0.004]	[0.006]	[0.006]
R-squared	0.816	0.819	0.794	0.800	0.856	0.860
Observations	143,855	143,855	143,855	143,855	12,137	12,137
N (district)	9,252	9,252	9,252	9,252	786	786
Pre-treatment	0.001	0.001	0.000	-0.001	0.000	-0.001
trend	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]
Panel B: Math						
Treated*Post	0.093**	0.095**	0.091**	0.065*	0.074*	0.058
	[0.036]	[0.038]	[0.043]	[0.036]	[0.040]	[0.038]
R-squared	0.847	0.847	0.863	0.866	0.865	0.867
Observations	404,114	404,114	404,114	404,114	32,629	32,629
N (district)	10,280	10,280	10,280	10,280	904	904
Pre-treatment	0.040**	0.041***	0.007	-0.003	0.029	0.019
trend	[0.016]	[0.015]	[0.013]	[0.011]	[0.017]	[0.017]
Panel C: ELA						
Treated*Post	0.084***	0.089***	0.067	0.049	0.069**	0.064*
	[0.030]	[0.032]	[0.042]	[0.038]	[0.032]	[0.034]
R-squared	0.882	0.884	0.886	0.888	0.900	0.902
Observations	404,114	404,114	404,114	404,114	32,629	32,629
N (district)	10,280	10,280	10,280	10,280	904	904
Pre-treatment	0.026***	0.028***	0.004	0.003	0.017*	0.009
trend	[0.009]	[0.008]	[0.012]	[0.011]	[0.010]	[0.010]
District, (grade) & year FE	Yes	Yes	Yes	Yes	Yes	Yes
District control	No	Yes	No	Yes	No	Yes

Notes: The table shows DD estimates of the effects of the charter entry (charter school share ever above 10%) on student outcomes. See notes of Table 3 for controls, clusters, and sample weight.

Table A9 DD estimates: Early adopter vs. later adopter

	(1)	(2)	(3)	(4)	(5)	(6)
	DD		DD-PSW		DD-PSM	
Panel A: Graduation rate						
Treated*Post	0.044***	0.045***	0.041***	0.035***	0.052***	0.046***
	[0.009]	[0.011]	[0.013]	[0.010]	[0.011]	[0.013]
R-squared	0.816	0.819	0.817	0.828	0.884	0.891
Observations	63,075	63,075	63,075	63,075	3,562	3,562
N (district)	4,068	4,068	4,068	4,068	232	232
Pre-treatment trend	0.002	0.003*	0.000	0.000	0.002	0.003*
	[0.001]	[0.002]	[0.002]	[0.001]	[0.002]	[0.002]
Panel B: Math						
Treated*Post	0.119***	0.130***	0.159***	0.143***	0.129**	0.165***
	[0.042]	[0.047]	[0.059]	[0.055]	[0.052]	[0.050]
R-squared	0.831	0.834	0.868	0.872	0.857	0.864
Observations	169,357	169,357	169,357	169,357	19,004	19,004
N (district)	4,297	4,297	4,297	4,297	528	528
Pre-treatment trend	0.081***	0.083***	0.023	0.023	0.071***	0.077***
	[0.019]	[0.014]	[0.026]	[0.019]	[0.021]	[0.019]
Panel C: ELA						
Treated*Post	0.094**	0.105**	0.076	0.073	0.078*	0.117**
	[0.038]	[0.042]	[0.051]	[0.047]	[0.047]	[0.048]
R-squared	0.871	0.873	0.885	0.888	0.906	0.909
Observations	169,357	169,357	169,357	169,357	19,004	19,004
N (district)	4,297	4,297	4,297	4,297	528	528
Pre-treatment trend	0.023**	0.031***	-0.028	-0.016	0.009	0.029*
	[0.010]	[0.009]	[0.024]	[0.018]	[0.012]	[0.015]
District, (grade) & year FE	Yes	Yes	Yes	Yes	Yes	Yes
District control	No	Yes	No	Yes	No	Yes

Notes: The table shows DD estimates of treated districts in early adopter states and comparison districts in late adopter states. The early adopter states are those passed state charter law before 1997 (27 states), and the late adopter states are those passed state charter law during 1997 to 2016 (17 states). See Table 2 for the list of states. See notes of Table 3 for controls, clusters, and sample weight.

Table A10 Placebo test: Effect of subsequent charter share on student outcomes

	(1)	(2)	(3)	(4)	(5)	(6)
	DD		DD-PSW		DD-PSM	
	Panel A: Graduation rate					
Treated*Post	0.005	0.005	-0.002	-0.003	-0.001	-0.005
	[0.008]	[0.008]	[0.009]	[0.008]	[0.008]	[0.008]
R-squared	0.790	0.793	0.839	0.843	0.915	0.921
Observations	138,889	138,889	138,889	138,889	3,299	3,299
N (district)	8,986	8,986	8,986	8,986	269	269
	Panel B: Math					
Treated*Post	0.181***	0.188***	0.067	0.065	0.160***	0.163***
	[0.037]	[0.033]	[0.049]	[0.043]	[0.043]	[0.039]
R-squared	0.842	0.842	0.858	0.861	0.891	0.895
Observations	389,555	389,555	389,555	389,555	12,389	12,389
N (district)	9,924	9,924	9,924	9,807	389	389
	Panel C: ELA					
Treated*Post	0.034*	0.043**	-0.037	-0.024	-0.002	0.010
	[0.020]	[0.017]	[0.051]	[0.043]	[0.025]	[0.027]
R-squared	0.877	0.878	0.889	0.890	0.922	0.926
Observations	389,555	389,555	389,555	389,555	12,389	12,389
N (district)	9,924	9,924	9,924	9,807	389	389
District, (grade) & year FE	Yes	Yes	Yes	Yes	Yes	Yes
District control	No	Yes	No	Yes	No	Yes

Notes: The table shows the “effect” of subsequent charter share on student outcomes in periods prior to the entry of charters by falsely assuming that charter entry occurred two years earlier. See notes of Table 3 for controls, sample weight, and clusters.

Table A11 Oster bound analysis

Outcome	Estimate	R^2	Oster bound (R^2=0.95)	Oster bound (R^2=1)
Graduation rate	0.038	0.808	0.073	0.086
Math	0.155	0.847	0.257	0.306
ELA	0.121	0.883	0.176	0.217

Note: This table reports point estimates, the R-squares and Oster bounds of our main estimates of DD and FE. Panel A correspond to the Column 2 of Table 3, and panel B to the Column 2 of Table 7 of the paper. All Oster bounds are calculated based on the comparison with estimates that do not include controls and fixed effects. We follow Oster (2019) to assume $\tilde{\delta}$ = 1 and $R_{max} = 1.3R_0$, however, in our case, $1.3R_0$ for all estimates are greater than 1, so, we instead, calculates Oster bounds for $R_{max} = 0.95$ and $R_{max} = 1$, respectively. R^2 refers to the R-squared of the specification with full control variables and fixed effects.

Table A12 Average charter enrollment share across models

	Graduation rate			Math & ELA		
	ALL	DD-Treated	FE	ALL	DD-Treated	FE
(a) Average charter share	0.6%	6.3%	2.7%	1.5%	12.3%	6.3%
(b) Average max charter share	1.9%	18.4%	8.0%	2.3%	18.8%	9.6%

Notes: The table average charter enrollment share across DD and FE models. (a) is the charter shares averaged across years; and (b) is the averaged eventual max share (it is what we're using to place districts in the treatment group).

Table A13 Estimates of non-linear effects of charter share on student outcomes

	(1)	(2)	(3)	(4)
Panel A: Graduation rate				
	Same year share		Average last four years	
Charter share	0.291***	0.323***	0.267**	0.251**
	[0.096]	[0.091]	[0.126]	[0.109]
Charter share square	-0.314**	-0.317***	-0.334**	-0.274
	[0.122]	[0.111]	[0.154]	[0.174]
R-squared	0.854	0.861	0.864	0.871
Observations	6,439	6,439	5,191	5,191
N (district)	416	416	416	416
Panel B: Math				
	Same year share		Same cohort last year	
Charter	0.733**	0.741***	0.301	0.173
	[0.339]	[0.274]	[0.272]	[0.288]
Charter share square	-0.659	-0.745*	-0.065	0.069
	[0.431]	[0.384]	[0.336]	[0.348]
R-squared	0.854	0.858	0.877	0.878
Observations	22,013	22,013	13,728	13,728
N (district)	607	607	604	604
Panel C: ELA				
	Same year share		Same cohort last year	
Charter	0.513*	0.253	0.241	0.035
	[0.300]	[0.251]	[0.262]	[0.264]
Charter share square	-0.61	-0.39	-0.188	0.049
	[0.392]	[0.355]	[0.323]	[0.323]
R-squared	0.909	0.912	0.922	0.924
Observations	22,013	22,013	13,728	13,728
N (district)	607	607	604	604
District, (grade) & year FE	Yes	Yes	Yes	Yes
District control	No	Yes	No	Yes

Notes: The table shows estimates of non-linear effects of charter effects on student outcomes for districts with any charter schools during the sample period. In columns (1) and (2), charter enrollment share of grades 9-12 is used for graduation rate, and charter enrollment share of grades 3-8 is used for test scores. In columns (1) and (2), the average last four years charter enrollment share of grade 9-12 is used for graduation rate, and the last-year same cohort grade enrollment share is used for test scores. See notes of Table 3 for controls and clusters. Regressions are weighted by high school enrollment for graduation rate and grade-level enrollment for Math and ELA.

Table A14 Effect heterogeneity: baseline performance

	(1)	(2)	(3)	(4)	(5)	(6)
	DD		DD-PSW		DD-PSM	
Panel A: Graduation rate						
Treated*post*	-0.001	-0.034	-0.042	-0.032	0.008	-0.016
baseperformance	[0.038]	[0.044]	[0.037]	[0.034]	[0.039]	[0.040]
R-squared	0.805	0.808	0.818	0.825	0.886	0.894
Observations	139,973	139,973	139,973	139,973	4,383	4,383
N (district)	9,001	9,001	9,001	9,001	284	284
Pre-treatment	-0.003	-0.004	0.001	0.001	-0.001	-0.003
trend	[0.003]	[0.004]	[0.004]	[0.004]	[0.003]	[0.003]
Panel B: Math						
Treated*post*	-0.171	-0.226	-0.056	-0.116	-0.181	-0.273*
baseperformance	[0.161]	[0.167]	[0.187]	[0.160]	[0.160]	[0.162]
R-squared	0.846	0.847	0.863	0.865	0.869	0.874
Observations	398,173	398,173	398,173	398,173	21,007	21,007
N (district)	10,129	10,129	10,129	10,129	594	594
Pre-treatment	0.004	-0.019	0.034	0.017	0.005	-0.026
trend	[0.071]	[0.064]	[0.072]	[0.067]	[0.070]	[0.068]
Panel C: ELA						
Treated*post*	-0.094	-0.131	0.090	0.096	-0.092	-0.170
baseperformance	[0.136]	[0.157]	[0.193]	[0.166]	[0.140]	[0.177]
R-squared	0.882	0.883	0.884	0.886	0.914	0.917
Observations	398,173	398,173	398,173	398,173	21,007	21,007
N (district)	10,129	10,129	10,129	10,129	594	594
Pre-treatment	0.015	-0.018	-0.012	-0.041	0.015	-0.018
trend	[0.039]	[0.035]	[0.089]	[0.076]	[0.039]	[0.042]
District, (grade) & year FE	Yes	Yes	Yes	Yes	Yes	Yes
District control	No	Yes	No	Yes	No	Yes

Notes: The table shows DD estimates of heterogeneous effects of the charter share on student outcome by baseline performance. *baseperformance* is the variable of baseline performance, it ranges from 0 to 1 by decile, and 0 refers to the least 10 percentile base performance, 0.1 refers the 10-20 percentile base performance, and so force. See notes of Table 3 for controls, clusters, and sample weight.

Table A15 Effects of charter entry on private school enrollment share

	(1)	(2)	(3)	(4)	(5)	(6)
	DD		DD-PSW		DD-PSM	
Panel A: district charter share on private school enrollment						
Treated*Post	0.000	0.002	-0.007***	0.001	-0.003	0.001
	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]	[0.002]
R-squared	0.904	0.912	0.907	0.920	0.905	0.919
Observations	39,973	39,973	39,973	39,973	10,932	10,932
N(district)	3,779	3,779	3,779	3,779	1,013	1,013
Pre-treatment	0.000	-0.001*	-0.001***	-0.001***	-0.001**	-0.001**
trend	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Panel B: high school charter share on elementary private school enrollment						
Treated*Post	0.014	0.007	-0.011	-0.004	0.011	0.004
	[0.014]	[0.009]	[0.007]	[0.005]	[0.012]	[0.006]
R-squared	0.900	0.911	0.863	0.897	0.840	0.934
Observations	40,786	40,786	40,786	40,786	1,720	1,720
N (district)	3,857	3,857	3,857	3,857	161	161
Pre-treatment	0.001	0.001	0.000	0.000	0.001	0.001
trend	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]
Panel C: elementary charter share on high school private school enrollment						
Treated*Post	0.004	0.001	-0.003	0.002	0.001	0.001
	[0.004]	[0.003]	[0.003]	[0.004]	[0.004]	[0.003]
R-squared	0.899	0.907	0.900	0.913	0.870	0.903
Observations	41,217	41,217	41,217	41,217	5,119	5,119
N (district)	3,897	3,897	3,897	3,897	477	477
Pre-treatment	-0.001	-0.001**	-0.001***	-0.001**	-0.002**	-0.002**
trend	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]
District, (grade) & year FE	Yes	Yes	Yes	Yes	Yes	Yes
District control	No	Yes	No	Yes	No	Yes

Notes: This table presents DD estimates of the effects of the charter entry (enrollment share ever above 10%) on the share of private school enrollment. The sample period is from the year 1996 to 2016 biannually. For DD and DD-PSM, regressions are weighted by the district total enrollment; For DD-PSW, regressions are weighted by the weight of DD times the inverse probability of propensity score. See notes of Table 3 for controls and clusters.

Appendix B Sample Description

This section describes how we create our sample for analyses. For graduation sample, we start from a conservative sample with schools have enrollments of high school grades (9-12). This sample contains 35,764 public schools (charters included), and 3,222 charter schools of them located in 1,218 geographic school districts. Then, we merge the schools data to district graduation rate data, and schools in districts with missing outcomes are dropped. Apart from removing districts with missing outcomes across the whole sample period, a few districts only have limited waves of data were also removed to ensure more balanced analyses of DD. Specifically, we removed districts with 13 or less waves (full wave is 16) of observations (account for 9% observations). The remined sample contains 25,038 public schools (charters included), and 1,074 charter schools located in 416 districts.

For test score sample, similarly we start from a conservative sample with schools have enrollments of grades 3-8, which has 84,918 public schools (charters included), and 6,719 charter schools of them located in 1,610 geographic school districts. Then, we merge the schools data to district test score data, and schools in districts with missing outcomes are dropped. Similarly, a few districts only have limited waves of data were also removed to ensure more balanced analyses of DD. Specifically, we removed districts with four or less waves (full wave is eight) of observations (account for 6% observations). The remined sample contains 70,493 public schools (charters included) and 2,453 charter schools located in 607 districts. Table B1 presents the summary statistics of in sample schools versus missing ones.

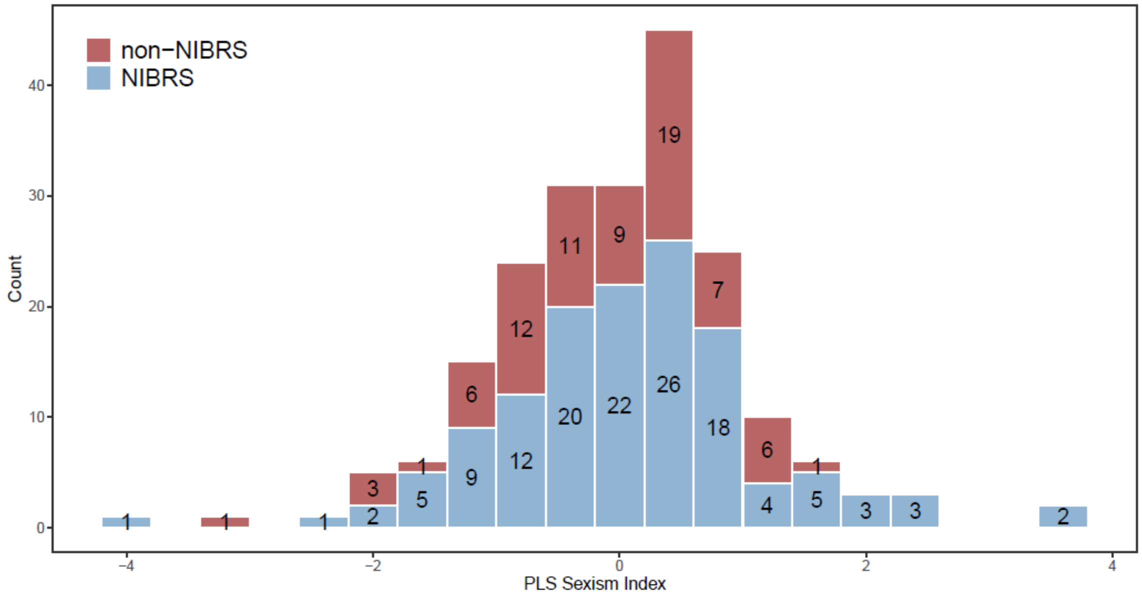
Table B1 Summary statistics for sample schools/districts versus missing ones

	Graduation sample		Test score sample	
	In sample	Missing	In sample	Missing
Panel A Charter schools				
Average student enrollment	336	274	404	411
Black%	20	24	20	33
Hispanic%	24	23	21	16
White%	46	45	39	36
FRL%	40	46	51	59
Urban%	83	84	84	87
Non-LEA authorizer %	5	61	4	76
N(school)	1,074	2,148	2,453	4,266
Obs	5,279	11,094	12,969	22,010
Panel B Districts with charter schools				
Average student enrollment	21,669	6,326	19184	9889
Black%	11	15	13	12
Hispanic%	17	17	25	29
White%	64	58	57	53
FRL%	32	38	55	55
Urban%	75	74	81	80
Non-LEA authorizer %	18	60	14	61
Charter share (ever max)	18	81	18	26
N(district)	416	802	607	1,003
Obs	6,920	7,386	22,013	74,505

Notes: This table presents summary statistics of in sample schools versus missing ones. In Panel B, if the district has at least one Non-LEA authorized charter school, then it is defined as Non-LEA authorizer. The charter share is the max charter share during sample period for a given district.

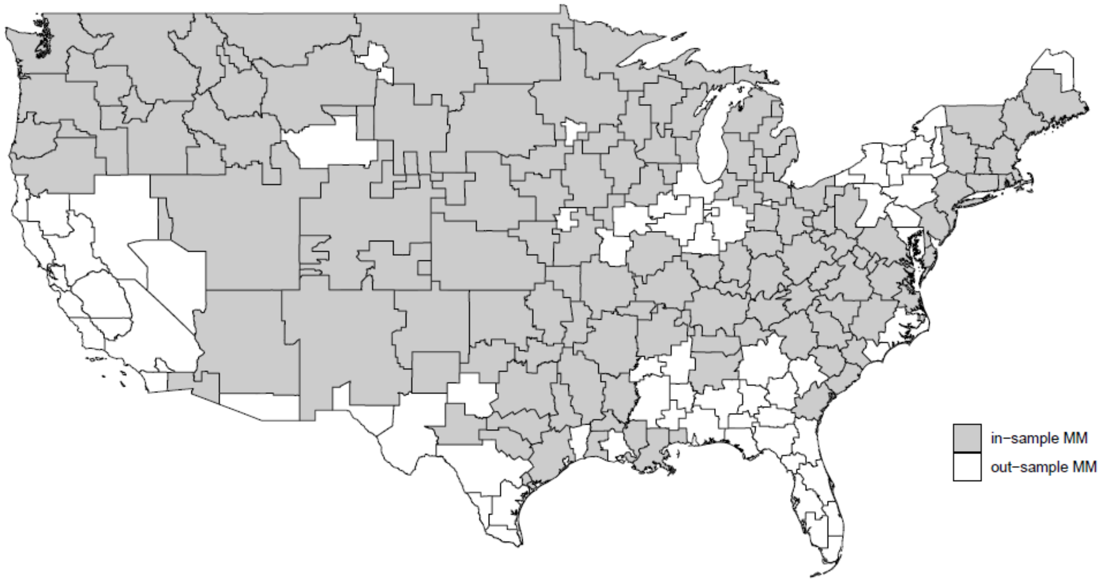
Chapter 3

Figure A1: Counts of MMs with and without NIBRS-reporting jurisdictions



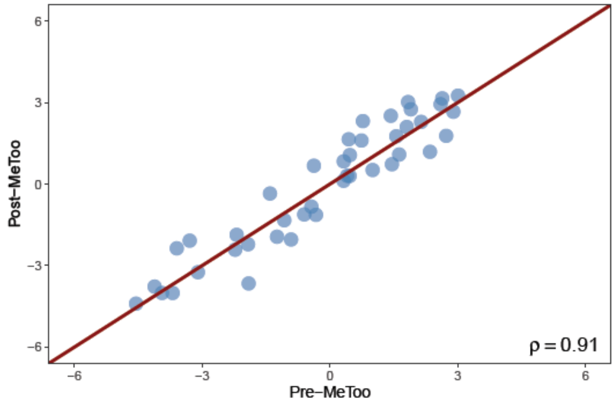
Notes. This figure plots counts of MMs with and without NIBRS-reporting jurisdictions. Values in each bin denote the number of NIBRS MMs and non-NIBRS MMs. The width of each bin is 0.4.

Figure A2 Geographic distribution of in-sample MMs

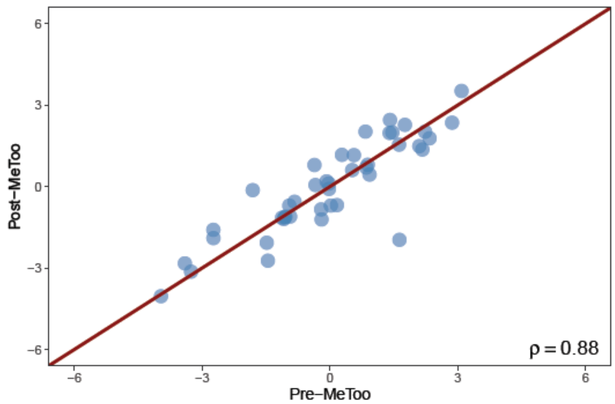


Notes. The shaded areas denote MMs with jurisdictions reporting to the NIBRS consistently in every month from 2015 to 2019.

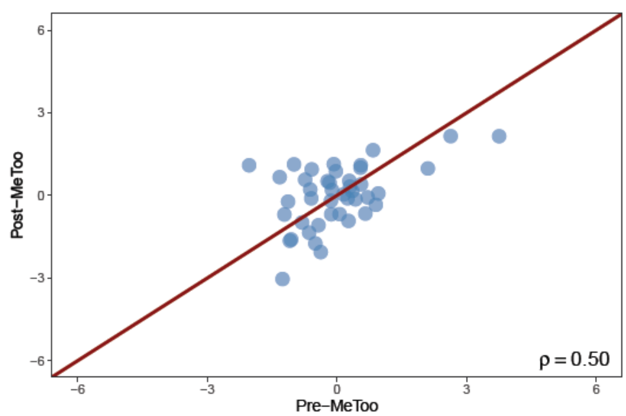
Figure A3 Comparison of pre- and post-MeToo principal components



(a) 1st Principal Component



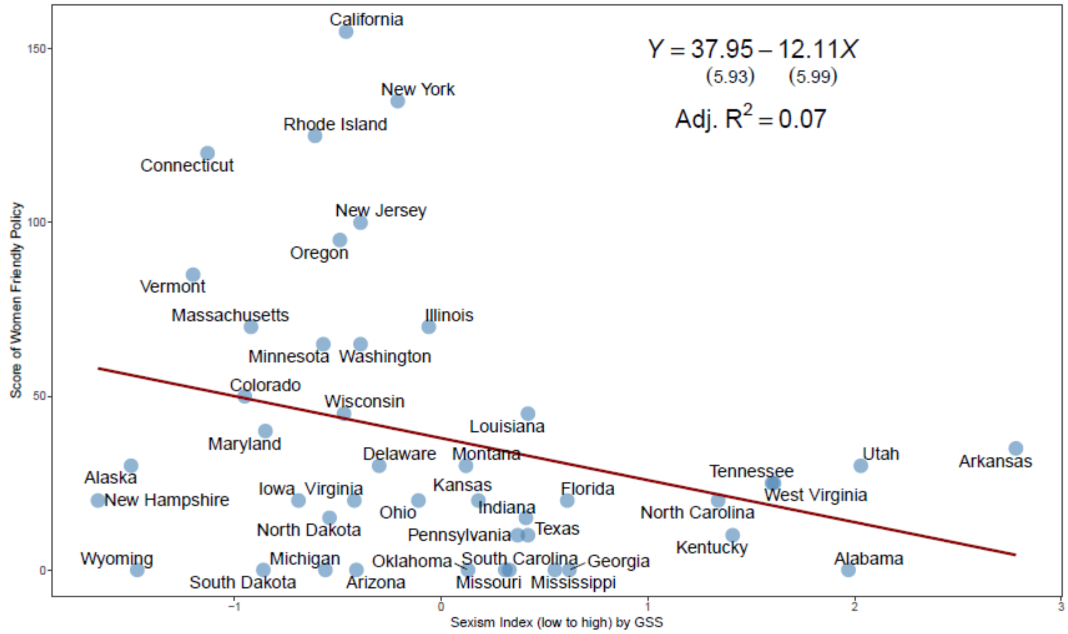
(b) 2nd Principal Component



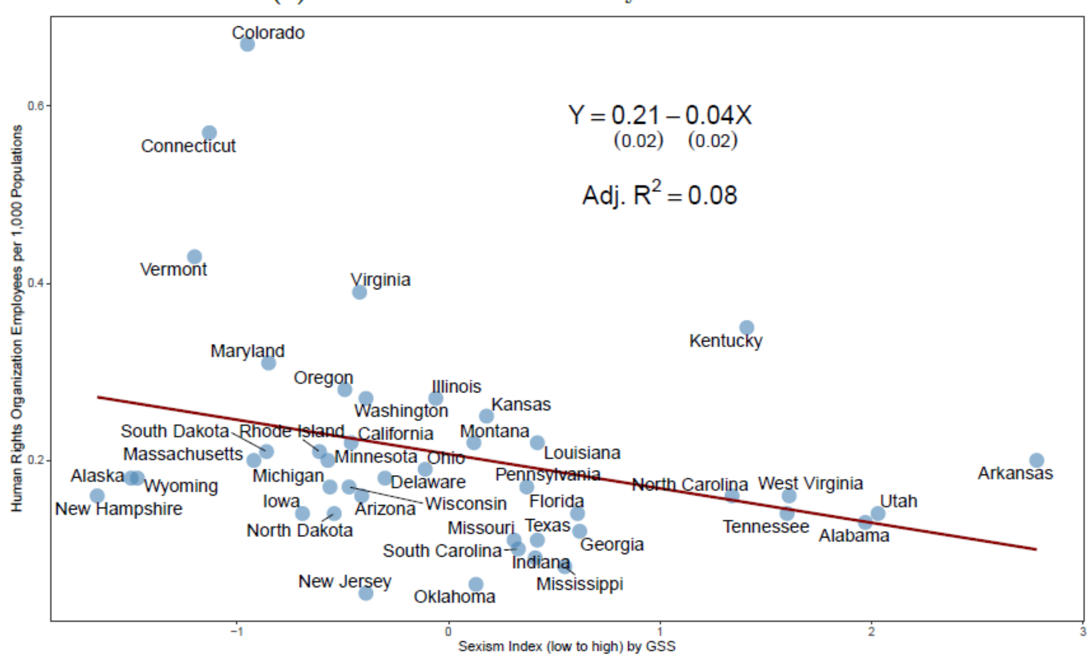
(c) 3rd Principal Component

Notes. The pre-MeToo principal components are obtained through the 2011-2016 Google Trends data, and the post-MeToo principal components are obtained through the 2017:Q4-2019 Google Trends data. The three components are obtained from principal component analysis (PCA). The red lines denote the 45-degree lines with zero intercept. The correlation coefficient of each pair is denoted by ρ.

Figure A4 Correlations of women friendly policies scores and human rights organization employees with the GSS sexism index



(a) State-Level Women Friendly Policies Scores



(b) State-Level Human Rights Organization Employees

Notes. Panel (a) plots the scatter plot between the GSS sexism index and the scores of women friendly policies by the National Partnership for Women & Families (2016). Panel (b) plots the scatter plot between the GSS sexism index and the state-level human rights organization employees (NAICS Code = 813311) per 1,000 populations in 2017 (excluding District of Columbia (27.91) and New York (1.01)). The red lines in both panels denote the regression lines.

Table A1 Correlation matrix of the 13 predictors in PLS

	[Word 1]	[Word 1](es)	[Word 2]	[Word 2](s)	[Word 3]	[Word 3](s)	[Word 4]	[Word 4](s)	Appearance	Mental	Intellect	Sex	Trump share
[Word 1]	1.00	-0.02	0.39	-0.46	0.71	0.81	0.12	0.28	0.64	0.46	0.01	-0.21	0.02
[Word 1](es)	-0.02	1.00	0.68	0.66	0.13	0.14	0.08	0.23	-0.13	0.12	0.44	0.48	0.53
[Word 2]	0.39	0.68	1.00	0.32	0.55	0.58	0.02	0.26	0.37	0.33	0.52	0.38	0.55
[Word 2](s)	-0.46	0.66	0.32	1.00	-0.31	-0.38	-0.10	-0.05	-0.35	-0.16	0.28	0.43	0.16
[Word 3]	0.71	0.13	0.55	-0.31	1.00	0.92	0.02	0.33	0.74	0.36	0.45	0.05	0.47
[Word 3](s)	0.81	0.14	0.58	-0.38	0.92	1.00	0.09	0.38	0.68	0.40	0.36	0.04	0.43
[Word 4]	0.12	0.08	0.02	-0.10	0.02	0.09	1.00	0.25	0.05	0.13	0.02	0.02	0.04
[Word 4](s)	0.28	0.23	0.26	-0.05	0.33	0.38	0.25	1.00	0.37	0.15	0.24	0.22	0.23
Appearance	0.64	-0.13	0.37	-0.35	0.74	0.68	0.05	0.37	1.00	0.41	0.27	-0.14	0.25
Mental	0.46	0.12	0.33	-0.16	0.36	0.40	0.13	0.15	0.41	1.00	0.17	0.07	0.18
Intellect	0.01	0.44	0.52	0.28	0.45	0.36	0.02	0.24	0.27	0.17	1.00	0.44	0.56
Sex	-0.21	0.48	0.38	0.43	0.05	0.04	0.02	0.22	-0.14	0.07	0.44	1.00	0.47
Trump share	0.02	0.53	0.55	0.16	0.47	0.43	0.04	0.23	0.25	0.18	0.56	0.47	1.00

Notes. Coding of the [Word 1] - [Word 4] is listed in Table 1.

Table A2 state-level measures of sexism

State	GSS Index	PLS Index	Women's March Protesters
Alabama	2.10	1.91	1.71
Alaska	-1.53	-0.99	13.80
Arizona	-0.43	-0.31	6.01
Arkansas	2.30	1.43	2.38
California	-0.35	-0.67	23.16
Colorado	-1.46	-0.53	29.11
Connecticut	-0.86	-0.25	4.51
Delaware	0.31	-0.01	1.41
District of Columbia	NA	NA	1057.1489
Florida	0.87	-0.28	4.41
Georgia	0.44	0.54	6.40
Hawaii	NA	NA	10.86
Idaho	NA	NA	7.24
Illinois	-0.35	-0.59	20.77
Indiana	0.24	0.63	1.49
Iowa	0.29	0.19	9.28
Kansas	-0.82	0.24	2.25
Kentucky	1.52	1.30	2.97
Louisiana	0.35	1.03	2.44
Maine	NA	NA	14.20
Maryland	-1.27	-0.43	1.35
Massachusetts	-1.11	-0.96	26.91
Michigan	-0.38	-0.41	4.42
Minnesota	-0.55	-0.33	17.97
Mississippi	0.26	0.19	1.22
Missouri	-0.14	0.38	4.50
Montana	1.43	-0.22	9.71
Nebraska	NA	NA	9.51
Nevada	NA	NA	6.64
New Hampshire	-1.16	-1.33	7.62
New Jersey	0.08	-0.22	1.85
New Mexico	NA	NA	11.02
New York	0.44	-0.58	25.49
North Carolina	1.21	0.86	5.95
North Dakota	-0.53	-0.25	3.61
Ohio	0.13	0.24	2.73
Oklahoma	0.28	0.92	2.61
Oregon	-1.12	-0.93	27.84
Pennsylvania	0.41	-0.33	6.87
Rhode Island	-1.60	-0.39	5.60
South Carolina	0.58	0.82	1.42
South Dakota	-0.41	0.46	6.29
Tennessee	0.93	0.99	4.38
Texas	0.51	0.20	3.33
Utah	1.48	0.65	5.23
Vermont	-0.84	-1.18	27.71
Virginia	-0.43	0.03	1.52
Washington	-0.27	-1.12	23.55
West Virginia	1.74	1.01	1.35
Wisconsin	-0.87	-0.32	15.48
Wyoming	-1.41	NA	7.45

Notes. The sexism index by GSS is constructed Charles et al. (2020). The number of protesters of 2017 women's marches are collected by authors from https://countlove.org/.

Appendix B Algorithm of Partial Least Squares (PLS)

In this appendix, we first briefly introduce the idea of the partial least squares (PLS) algorithm. Suppose Y denotes a $n \times 1$ outcome variable and X denotes a $n \times p$ matrix, where n and p respectively indicate sample size and the number of features or predictors. To explore how the p features in X predict the variation in Y , we consider the following linear model:

$$Y = X\theta + \varepsilon, \tag{A1}$$

where θ is a $p \times 1$ vector of slope coefficients and ε is a $n \times 1$ vector of error terms.

In practice, when p is large and some features are highly correlated, forming linear combinations of features will help reduce noise and extract useful information from these features. The PLS algorithm condenses the p features to a remarkably smaller number of K linear combinations of features, or PLS components. That is, the predictive model in equation (A1) could be rewritten as

$$Y = (X\Omega_K)\theta_K + \tilde{\varepsilon}, \tag{A2}$$

where Ω_K is a $p \times K$ matrix with column ω_i, $i = 1, 2, ..., K$ and $K \leq p$, denoting the set of linear combination weights (also known as factor loadings) used to construct the i_{th} components, so $X\Omega_K$ is the dimension-reduced version of the predictive feature set. Likewise the slope coefficient θ_K is now becomes a $K \times 1$ vector and estimable using OLS. The PLS algorithm then seeks K linear combinations of X that have maximal predictive association with Y . The weights used to construct the i_{th} PLS component solve

$$\omega_i = \arg\max_{\omega} \mathrm{Cov}^2(Y, X\omega) \quad \text{s.t.} \quad ||\omega|| = 1 \text{ and } \mathrm{Cov}(X\omega, X\omega_j) = 0, \quad j = 1, 2, ..., i-1. \tag{A3}$$

In other words, a PLS component has the highest possible covariance with Y under the constraint of being uncorrelated with all previous components. The problem in equation (A3) has no closed-form solution for all components, but can be efficiently solved using a number of prominent algorithms such as the NIPALS by Martens and Martens (2001) [1]and the SIMPLS by de Jong (1993)[2]. In practice, to avoid overfit which will lead to poor out-of-sample prediction, the optimal number of component K will be determined by the widely used cross-validation (CV) method.

An alternative, unsupervised dimension reduction method is principal component regression (PCR). Similar to PLS, PCR also seeks the K linear combinations of X. However, unlike PLS, the first step in PCR conducts principal component analysis (PCA) to combine features and construct the K linear combinations that best reserve the covariance structure among the p features. In other words, PCR does not incorporate Y when constructing principal components, but condenses data into components based on the covariation among the features. If these features indeed contain information explaining social norms on sexist attitudes, the constructed components by PCA should be closely related to Y , and slope coefficients on low variance components will be zeroed out. Specifically, PCR estimates ΩK and the ith linear combination by solving

$$\omega_i = \arg\max_{\omega} \mathrm{Var}(X\omega) \quad \text{s.t.} \quad \|\omega\| = 1 \text{ and } \mathrm{Cov}(X\omega, X\omega_j) = 0, \quad j = 1, ..., i-1. \quad \text{(A4)}$$

Equation (A4) could be solved via singular value decomposition (SVD) of X, and the constructed principal components extract the most possible common variation within X, measured by eigenvalues. The optimal choice of K will be similarly determined by the CV method.

---

[1] Martens, H., and Martens, M. (2001). Multivariate Analysis of Quality: An Introduction, J.Wiley Sons.
[2] de Jong, S. (1993). SIMPLS: An Alternative Approach to Partial Least Squares Regression.