# The MIT Undergraduate Journal of Economics

**Volume XX**

**2020-2021**

# The MIT Undergraduate Journal of Economics Volume XX

## 2020-2021

*Mailing Address:*

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

# Foreword

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

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

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

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

# The MIT Undergraduate Journal of Economics Volume XX

## 2020-2021

Contents

# The Effect of Common Core Standards on Racial Achievement Gaps

Isabelle Yen*

February 2021

**Abstract**

In the United States, there is a persistent academic achievement gap between white and Black students. While this may stem from a variety of factors, there is evidence that current education practices may further contribute to this gap. For the past several decades, one of the goals of standards-based education reform has been to address disparities in student performance. Using data from the Educational Opportunity Project at Stanford University, I perform a difference-in-difference analysis to identify whether implementation of the Common Core State Standards affected the white-Black achievement gap on standardized tests. I find that implementation had a small but inconsistent effect on the gap in mathematics and no significant effect on the gap in English Language Arts.

## Introduction

In this paper, I perform a difference-in-difference analysis to measure the effect of Common Core standards on achievement gaps between white and Black students in mathematics and English Language Arts, with the goal of measuring whether educational standards can help diminish racial disparities in performance.

In 1848, Horace Mann, education reformer and first Secretary of the Massachusetts State Board of Education, called education "the great equalizer" (Rhode et al., 2012). As the United States continues to grapple with longstanding racial inequality, the education system seems to be not so much equalizing as dividing. In 2017, data from the National Assessment of Educational Progress (NAEP) showed that, on average, Black students were performing two to two-and-a-half grade levels behind white students (Reardon and Fahle, 2017). On the 2019 NAEP, only twenty percent of Black fourth-graders were performing at the "Proficient" level, compared to over half of white fourth-graders (NAEP Report Card: Mathematics, 2019).

There is certainly a variety of other social, cultural, economic, and political factors that take effect well before students even enter school and extend far beyond the classroom that lead to gaps in academic achievement. Even if education in the United States did fulfill its promise, many of these obstacles

*Thank you to Ro'ee Levy, Jonathan Cohen, and Hannah Ruebeck for their guidance on this paper.





would likely still remain. This paper, however, focuses specifically on the ways in which the education system itself currently entrenches achievement gaps and how standards-based reform may impact these trends.

## Background

### Educational Inequality in the United States

Although educational inequality exists in other countries, the United States stands out among developed countries in that students often have completely different experiences in the education system based on their backgrounds. Unlike other countries, the United States does not fund schools equally from a centralized source. In fact, the wealthiest ten percent of school districts spend nearly ten times more than the poorest ten (Darling-Hammond, 2001).

Schools with fewer resources find it difficult to compete with more well-funded schools to attract the best teachers: a 1999 study in California found that schools with the highest percentage of students receiving free or reduced-price lunch were also the schools with the highest percentages of underqualified teachers (Shields et al., 1999). Levin (2007) visited several inner-city elementary schools with a high concentration of poverty as well as suburban and middle-class schools. He found that in the suburban schools, the assignments were much more challenging and students were given problems that required a deeper understanding of the material. While unequal funding affects students of all races, the average household income for Black or African American individuals is among the lowest of all racial and ethnic groups (Noël, 2018). This means that the negative consequences of unequal funding are more likely to affect Black students than other groups.

Within schools, tracking, or the placement of students into different levels of classes, can play a role in generating inequality between white and Black students. A study in North Carolina found that Black students were overrepresented in remedial-level courses and underrepresented in advanced courses compared to white students (Clotfelter et al., 2005). Oakes (1985) observed similar trends and also found that tracking programs did not seem to help students overall, while negatively affecting the behavior and motivation of students placed in lower tracks. While tracking is not inherently harmful, in many places, it has become a method of informal segregation and another obstacle to equity in education. This is especially alarming considering that some forms of ability grouping begin as early as elementary school (Loveless, 2013).

### Standards-Based Education Reform and the Common Core

The persistence of achievement gaps has led to efforts to implement educational standards in order to level the playing field for all students. The history of standards-based education reform began in the 1980s, when the National Commission on Excellence in Education published a report called *A Nation at Risk: The Imperative for Educational Reform!* (Kenna and Russell, 2018). The report stated that "where there should be a coherent continuum of learning, we have none, but instead an often incoherent, outdated patchwork quilt", and recommended the implementation of new standards (National Commission on Excellence in Education, 1983). It also stated that "[all], regardless of race or class or economic status, are entitled to a fair chance" and noted that while functional illiteracy was alarmingly high for young people on average, it was much higher for minority students.

The report prefaced the next several decades of reform. No Child Left Behind, signed into law in 2002, required states to prove that all students were proficient in reading and mathematics by 2014 and also required schools to report assessment results by demographic group in order to identify achievement gaps. While No Child Left Behind was not reauthorized by Congress in 2007, the requirement of complete proficiency by 2014 remained in place. 2009 saw the introduction of Race to the Top, a competitive grant program wherein states would compete for funding based on their adoption of certain reforms. States could apply for waivers from No Child Left Behind's 2014 deadline if they fulfilled requirements that included adopting college and career readiness standards and creating accountability systems for low-performing schools and schools with high achievement gap (Kenna and Russell, 2018).

The Common Core State Standards initiative also began in 2009. It was a state-led effort to increase consistency in what students were learning, and resulted in a set of standards that drew on high-performing state and international programs. While the federal government did not develop the Common Core, Race to the Top played a major role in encouraging its adoption (Kenna and Russell, 2018). The Common Core represented a shift from the way students were being taught before, especially in math, where it introduced a new style of questions that tested "conceptual" math (Hartnett, 2016). This drew frustration from parents who were used to more traditional methods (McArdle, 2014). On the other hand, there was also support, with a 2011 survey of teachers in forty states finding that more than ninety percent liked that the Common Core mathematics standards provided "a clear understanding of what students were expected to learn" (Schmidt and Burroughs, 2013).

There have been limited attempts to evalute the impact of Common Core. Some critics have pointed to declining results on the 2013-2015 NAEP as evidence that the standards were not improving student performance (Polikoff, 2017). Loveless (2016) categorized states as strong adopters, medium adopters, or nonadopters based on if and when they planned for classroom implementation of Common Core standards. He found that all three groups saw small gains in reading and small declines in mathematics on the NAEP between 2009 to 2015. Some states, however, only implemented Common Core in 2014 or 2015. At this point, there is unlikely to be conclusive evidence as to whether Common Core is "working" overall (Polikoff, 2017).

I explore how the Common Core may have affected another key goal of standards-based reform: leveling the playing field for students of different races. Though this question faces the same challenge of not having extensive data, there is reason to believe that some effects on disparities may be immediately observable. For one, all teachers in states that adopted Common Core are expected to make sure that their curricula meet the same set of standards. While this is not a replacement for teacher quality, it will help address the fact that students in some schools, particularly ones with more low-income and minority students, have been taught using curricula that clearly do not prepare them for careers or further education.

Teachers may be incentivized to focus more on lower-performing students to help them meet the new expectations. In addition, since a significant part of standards-based education reform involves having states measure and address high achievement gaps in schools, this may lead to states and districts allocating additional resources toward decreasing those gaps. There is also evidence that Common Core may lead to decreased emphasis on tracking. In California, over seventy percent of eighth graders were





enrolled in advanced math courses in 2013, compared to under fifty percent in 2015 (Loveless, 2016). It is possible that this trend may have been put in motion before the Common Core. Nonetheless, the fact that Common Core promotes the same standards for all students in each grade does contrast with the practice of tracking (Schmidt and Burroughs, 2013).

This paper explores whether Common Core implementation has had any impact on these disparities by comparing trends in achievement gaps in states that implemented the Common Core and states that did not. While there are other achievement gaps that are important to study, for this paper, I focus specifically on the well-documented gap between white and Black students. I find that the Common Core has had a minimal on achievement gaps in mathematics, but not across all grades, and no significant effect on achievement gaps in English Language Arts.

## Data

The dataset used for this paper comes from the Stanford Education Data Archive (SEDA), which is part of the Educational Opportunity Project at Stanford University (Fahle et al., 2021). SEDA includes test score data for schools across the United States; this paper used data aggregated by geographic school district.[1]

The SEDA datasets are built from the ED*Facts* data system from the United States Department of Education, which includes results from each state's standardized tests in mathematics and English Language Arts (ELA) for students in third to eighth grade from 2008-2009 to 2017-2018. Each state is required to test all students in grades three to eight but is not required to use a specific test. In addition, states do not provide numeric scores. Instead, they provide "proficiency data", which counts the number of students at different proficiency levels (such as below basic, basic, proficient, or advanced). This makes it challenging to compare results across states, as each state's threshold for proficiency may be different, but the researchers from the Educational Opportunity Project were able to standardize the results by estimating the proficiency threshold for each state. In order to place the proficiency thresholds for all states on the same scale, they rescaled them based on data from the NAEP, which is taken by a random sample of students in each state in certain grades and years.

The researchers then standardized scores for each district to a Grade Cohort Standardized scale, based on a reference group consisting of the three cohorts of students who were in fourth grade in 2009, 2011, and 2013. On this scale, the average score for a certain grade is equal to that grade, so the average score for a student in third grade would be three and the average score for a student in eighth grade would be eight. If the average score for fourth graders in a district is five, that means that the fourth grade students in that district are performing a year ahead of the national reference group (Fahle et al., 2019).

For each district, achievement gaps were calculated for different pairs of demographic groups by subtracting the mean score of students in one group from another. The white-Black achievement gap was calculated by subtracting the mean score of Black students in a certain grade from the mean score of white students in that grade.

[1]The full data is from Reardon, S.F., Ho, A. D., Shear, B. R., Fahle, E. M., Kalogrides, D., Jang, H., and Chavez, B. (2021). Stanford Education Data Archive (Version 4.0). http://purl.stanford.edu/db586ns4974

Since states had different standards before the introduction of the Common Core, the effect may be different depending on if a state already had a rigorous curriculum or not. A report published by the Fordham Institute scored pre-Common Core standards in every state (and Washington, D.C.) and the Common Core in two areas, "Content and Rigor" and "Clarity and Specificity" (Carmichael et al., 2010). While ideological bias may be a concern when using sources published by think tanks, it is less of a risk in this case as the scores were assigned by education professionals who determined ratings based on a detailed grading metric. Math standards in California, D.C., Florida, Indiana, and Washington as well as ELA standards in California, D.C., Indiana, Massachusetts, Tennessee, and Texas, scored higher than the Common Core. Results from these states were excluded from the dataset, since it is possible that states that already had strong standards may have been on a different trajectory in terms of addressing disparities compared to other states.

Out of the remaining states, the ones that adopted the Common Core math standards comprised the treatment group, and the rest, the control group. Within the treatment group, I used information from the Common Core website to find the year of full implementation for each state, ranging from 2012 to 2015 (Common Core State Standards, 2013). I classified districts as belonging to the treatment or control group based on which states they are located in.

## Empirical strategy

I estimate the following:

$$Y_{i,t} = \beta Post_{i,t} + \gamma_i + \delta_t + \epsilon_{i,t}$$

where $Y_{i,t}$ is the white-Black gap in district $i$ in year $t$, and $Post_{i,t}$ is an indicator variable equal to one if the district is in a treated state and the current year is at least the treatment year. $\gamma_i$ and $\delta_t$ are state and time fixed effects, respectively. $\epsilon_{i,t}$ is an unobserved error term. $\beta$ estimates the effect of Common Core implementation on the white-Black achievement gap.

I include state fixed effects to control for the possibility that individual states may have different policies and characteristics that affect achievement gaps. Similarly, I include time fixed effects because there may be other overall trends besides Common Core adoption that impact achievement gaps. I do not include any other control variables as there do not seem to be significant characteristics that clearly divide states that adopted the Common Core and those that did not.

The possibility that the parallel trends assumption is not satisfied is a potential threat to identification using a difference-in-difference specification. In order to check for evidence of parallel trends as a robustness check, I also run the following event study specfication to generate an event study plot:

$$Y_{i,t} = \alpha + \sum_{j=1}^{J} \lambda_j (\text{Lag}j)_{i,t} + \sum_{k=1}^{K} \mu_k (\text{Lead}k)_{i,t} + \gamma_i + \delta_t + \epsilon_{i,t}$$

where lags and leads are binary variables indicating that district $i$ was a certain number of years away from implementing the Common Core.



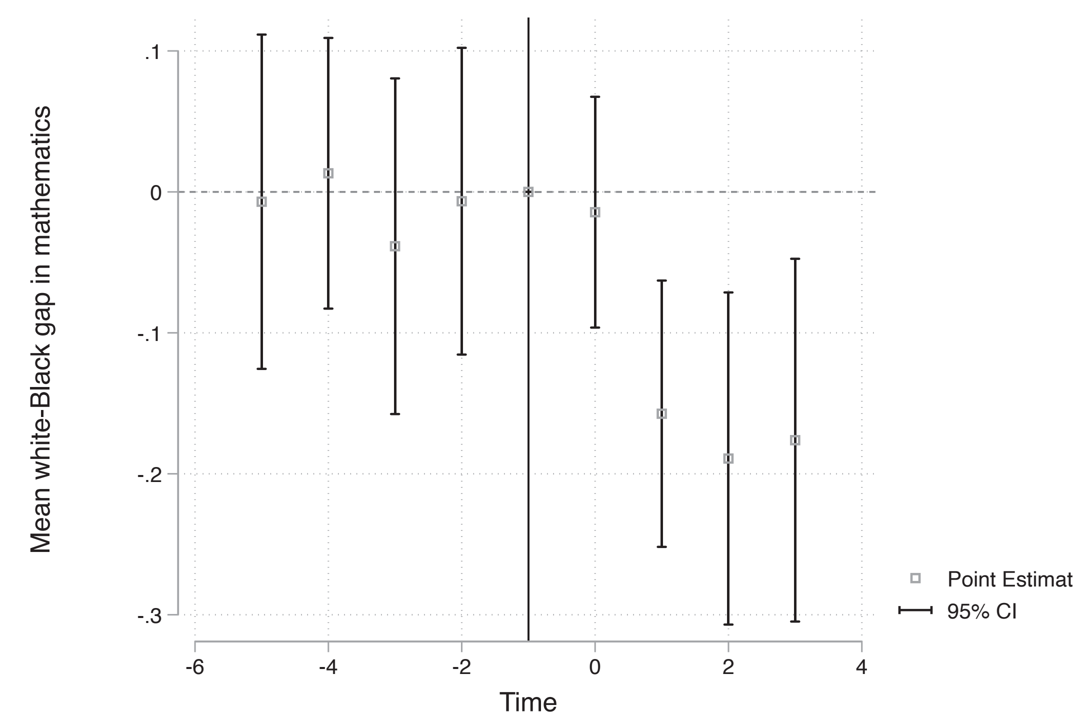


## Results

### Results for Mathematics and English Language Arts

I use a graph of coefficients on event lags and leads to check whether the parallel trends assumption is satisfied for the difference-in-difference analysis on Common Core implementation. Figure 1 shows that the pre-period coefficients for implementation of Common Core mathematics are close to zero, with confidence intervals crossing the x-axis. While this does not prove that the achievement gaps in treatment and control districts would have followed the same trends absent the Common Core, it provides evidence that the two groups were on similar trends before implementation. Figure 2 shows the event study plot for ELA. Similar to the graph for mathematics, the pre-period coefficients are all close to zero, and it does not suggest that the treatment and control groups were on different trends before Common Core implementation.[2]

Figure 1: Event study plot for Common Core implementation in mathematics

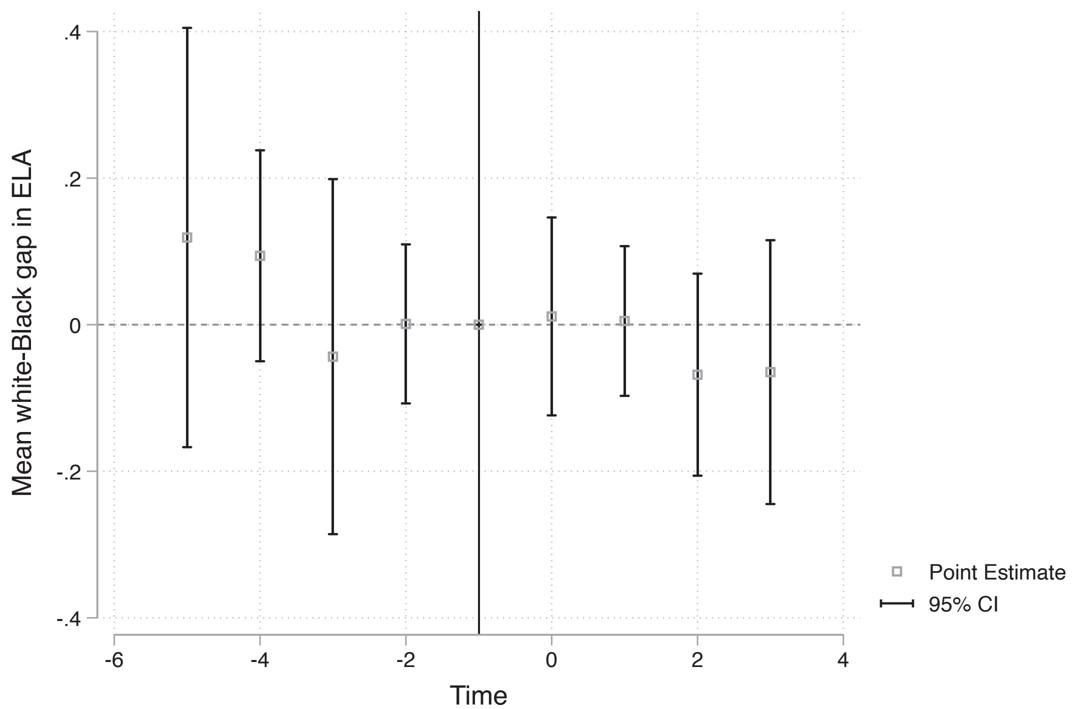



Note: the baseline (omitted) period is one year before Common Core implementation in each district, indicated by the vertical line in the plot. The sample for this plot was limited to states that implemented the Common Core in 2014 and 2015. Only balanced lags and leads are shown in the graph.

I run the difference-in-difference analysis for each grade, as well as on the pooled set of all grades. Table 1 summarizes the results of the analysis for Common Core mathematics, with the coefficient of *post* referring to the effect of implementing the Common Core standards on the mean white-Black achievement gap in mathematics. The column "all grades" refers to the pooled result for all grades.

[2] I use the eventdd package in Stata developed by Clarke and Schythe (2020) to generate the event study plots.

Figure 2: Event study plot for Common Core implementation in ELA




Note: the baseline (omitted) period is one year before Common Core implementation in each district, indicated by the vertical line in the plot. The sample for this plot was limited to states that implemented the Common Core in 2014 and 2015. Only balanced lags and leads are shown in the graph.

Fixed effects are hidden in the interest of space.

Table 1: Estimated effect of Common Core implementation on the white-Black gap in mathematics

| | All grades | Grade 3 | Grade 4 | Grade 5 | Grade 6 | Grade 7 | Grade 8 |
|---|---|---|---|---|---|---|---|
| post | -0.155*** | -0.155** | -0.160*** | -0.132*** | -0.0993 | -0.168* | -0.223*** |
| | (0.0495) | (0.0628) | (0.0447) | (0.0449) | (0.0747) | (0.0941) | (0.0718) |
| Constant | 1.518*** | 1.264*** | 1.472*** | 1.613*** | 1.938*** | 1.995*** | 2.057*** |
| | (0.0259) | (0.0242) | (0.0208) | (0.0276) | (0.0282) | (0.0368) | (0.0387) |
| state fixed effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| time fixed effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| grade fixed effects | Yes | No | No | No | No | No | No |
| N | 254365.00 | 45,138.00 | 45,182.00 | 43,942.00 | 43,427.00 | 39,699.00 | 36,977.00 |
| R-squared | 0.04 | 0.04 | 0.03 | 0.03 | 0.03 | 0.03 | 0.04 |

Notes: Standard errors (in parentheses) are clustered at the state level and robust to heterskedasticity. Significance at the 1, 5, and 10 percent levels indicated by ***, **, and *, respectively

Table 2 summarizes the results of the difference-in-difference analysis on the effect of Common Core reading on the achivement gap in ELA.





Table 2: Estimated effect of Common Core implementation on the white-Black gap in ELA

| | All grades | Grade 3 | Grade 4 | Grade 5 | Grade 6 | Grade 7 | Grade 8 |
|---|---|---|---|---|---|---|---|
| post | -0.0696 | -0.0798 | -0.0196 | -0.0691 | -0.118* | 0.000473 | -0.127** |
| | (0.0575) | (0.0926) | (0.0624) | (0.0703) | (0.0647) | (0.0520) | (0.0554) |
| Constant | 1.991*** | 2.140*** | 1.811*** | 2.001*** | 1.885*** | 2.226*** | 1.838*** |
| | (0.0408) | (0.0514) | (0.0436) | (0.0379) | (0.0444) | (0.0464) | (0.0474) |
| state fixed effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| time fixed effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| grade fixed effects | Yes | No | No | No | No | No | No |
| N | 233300.00 | 38,920.00 | 39,139.00 | 39,404.00 | 39,155.00 | 38,680.00 | 38,002.00 |
| R-squared | 0.04 | 0.04 | 0.05 | 0.04 | 0.04 | 0.03 | 0.03 |

Notes: Standard errors (in parentheses) are clustered at the state level and robust to heterskedasticity. Significance at the 1, 5, and 10 percent levels indicated by ***, **, and *, respectively

**Discussion of Results**

In 2009, the mean white-Black gap in mathematics for all grades was about 1.67, while the gap in ELA was about 1.58. This means that in the average school district, white students outperformed Black students by slightly over one and a half grade levels in both subjects before the Common Core.

The results of the difference-in-difference analysis show that there is a slight treatment effect in mathematics, but that it is not consistent across grades. The effect on the weighted average achievement gap across all grades was about -0.16, which is small compared to the existing achievement gaps. For ELA, implementation of the Common Core had no significant effect.

Figure 5 shows score trends in mathematics for white and Black fifth-graders in treatment and control districts. For reference, the mean score for fifth graders on the Grade Cohort Standardized scale is five. I chose to show only one grade in the interest of space, but the graphs for other grades are similar. The plot shows that, for both groups, trends in the treatment and control groups appear to be quite similar. While fifth graders in the treatment groups appear to have scored better than those in the control groups, the overall gap between the mean scores of white and Black students has not narrowed over time. Figure 6 shows score trends in ELA, reflecting the results showing that Common Core implementation does not appear to have narrowed the achievement gap.

Figure 3: Trendlines in math, grade 5

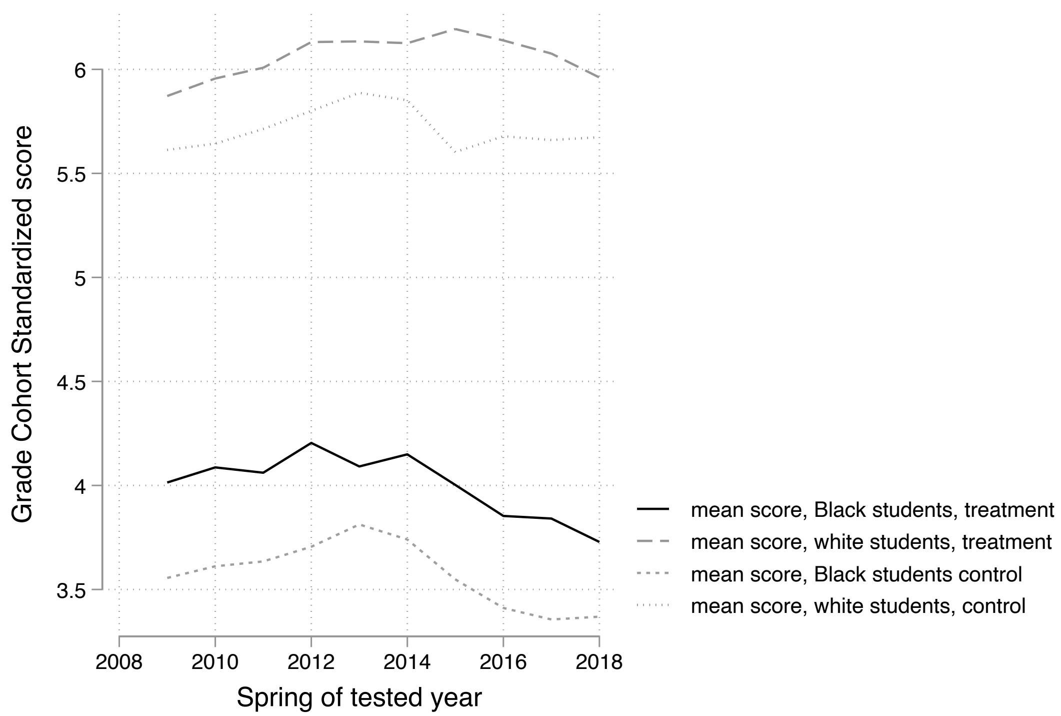
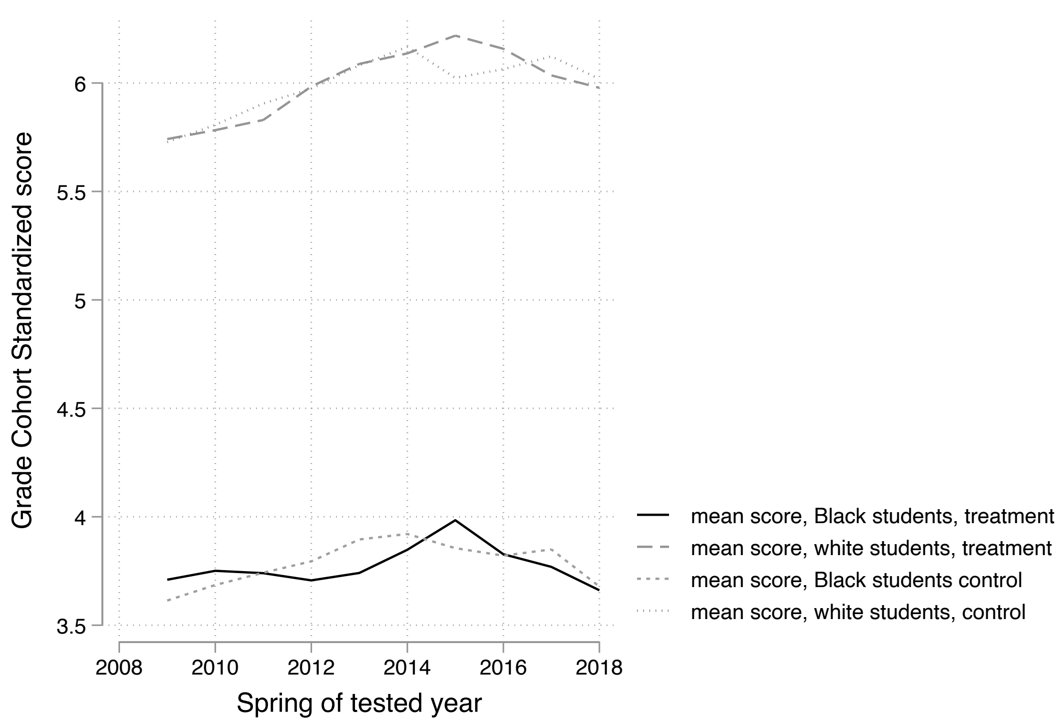



Figure 4: Trendlines in ELA, grade 5








## Conclusion

Based on the results, it appears that implementation of Common Core standards does not lead to a significant change in achievement gaps overall. While certain grades seem to see a small significant effect in mathematics, the change is small compared to the existing achievement gaps. This suggests that standards-based reform may not be the solution to achievement gaps, or that changing only what students are learning in schools is not an effective approach to tackling educational inequality.

It is important to consider certain limitations to the analysis. For one, the Common Core is a relatively recent development. It is still possible that it may have some effect on the inequalities that contribute to the white-Black achievement gap. Examples of potential ways in which the Common Core could impact achievement gaps include drawing attention to schools with high achievement gaps, increasing teachers' focus on lower-performing students, and discouraging the practice of tracking. It will be a while, however, until a more definitive evaluation of the effects of the Common Core can be made. Further research is needed later when more data are available and the changes have had longer to take hold.

I categorize school districts as being in the treatment or control groups based on state policies. One concern is that even if a state adopted the standards, implementation may not have been equally successful throughout the state. A number of states that adopted the Common Core did so mainly because they wanted to be eligible for certain federal grants through Race to the Top program, not because they were truly excited about the new standards (Siegfried, 2016). This may have impacted how successfully or comprehensively the Common Core was actually implemented in classrooms. In addition, there have been examples of pushback against the Common Core. For example, in 2014, an estimated 35,000 students in New York refused to take the Common Core assessments (McArdle, 2014). This resistance could mean that adoption of the standards may not have been able to enact its intended change in all states.

The results of this analysis do not allow for any sweeping evaluation of the effectiveness of standards-based reform, but they do indicate that implementing new standards is not enough to address achievement gaps. There are a variety of other factors outside of schools that contribute to disparities in student performance. Achievement gaps are unlikely to disappear as long as inequalities continue to exist in nearly every facet of life, from neighborhood safety to access to healthy food. The results of this paper highlight the importance of addressing these other factors before changes within schools can have a major impact on educational outcomes.

## References

Carmichael, S. B., G. Martino, K. Porter-Magee, and W. S. Wilson (2010). The State of State Standards - and the Common Core - in 2010. Technical report, Fordham Institute.

Clarke, D. and K. T. Schythe (2020). Implementing the Panel Event Study. pp. 32.

Clotfelter, C. T., H. F. Ladd, and J. Vigdor (2005). Who teaches whom? Race and the distribution of novice teachers. *Economics of Education Review 24*(4), 377–392.

Common Core State Standards (2013). Standards in Your State.

Darling-Hammond, L. (2001). Inequality in Teaching and Schooling: How Opportunity Is Rationed to Students of Color in America. In *The Right Thing to Do, The Smart Thing to Do: Enhancing Diversity in the Health Professions: Summary of the Symposium on Diversity in Health Professions in Honor of Herbert W. Nickens, M.D.* National Academies Press (US).

Fahle, E. M., B. Chavez, D. Kalogrides, B. R. Shear, S. F. Reardon, and A. D. Ho (2021, February). Stanford Education Data Archive: Technical Documentation (Version 4.0).

Hartnett, K. (2016, October). Meet the New Math, Unlike the Old Math. *Quanta Magazine*.

Kenna, J. L. and W. B. Russell, III (2018). The Culture and History of Standards-Based Educational Reform and Social Studies in America. *Journal of Culture and Values in Education 1*(1), 26–49.

Levin, H. M. (2007). On the Relationship between Poverty and Curriculum. *North Carolina Law Review 85*.

Loveless, T. (2013). The 2013 Brown Center Report on American Education: How Well are American Students Learning? Technical report, Brookings Institution.

Loveless, T. (2016). The 2016 Brown Center Report on American Education: How Well are American Students Learning? Technical report, Brookings Institution.

McArdle, E. (2014). What Happened to the Common Core? *Harvard Ed. Magazine Fall 2014*.

NAEP Report Card: Mathematics (2019). The Nation's Report Card.

National Commission on Excellence in Education (1983). A Nation At Risk: the Imperative for Educational Reform. Technical report, National Commission on Excellence in Education, Washington, D.C. Publisher: US Department of Education (ED).

Noël, R. A. (2018). Race, Economics, And Social Status. Technical report, U.S. Bureau of Labor Statistics.

Oakes, J. (1985). *Keeping Track: How Schools Structure Inequality*. New Haven, CT: Yale Univ. Press.

Polikoff, M. S. (2017). Is Common Core "Working"? And Where Does Common Core Research Go From Here? *AERA Open 3*(1), 1–6.

Reardon, S. F. and E. M. Fahle (2017). State of the Union 2017, Education. *Pathways Magazine 2017*, 20–23.

Rhode, D., K. Cooke, and H. Ojha (2012, December). The Decline of the 'Great Equalizer'. *The Atlantic*.

Schmidt, W. H. and N. A. Burroughs (2013). The Common Core State Standards address two tenacious problems in U.S. education. *Educational Leadership 70*(4).

Shields, P. M., C. E. Esch, D. C. Humphrey, V. M. Young, and H. Hunt (1999). Teaching and California's Future. The Status of the Teaching Profession: Research Findings and Policy Recommendations. Technical report, Center for the Future of Teaching and Learning.

Siegfried, J. B. (2016). It's Common Sense: Why the Common Core Is Not Coercive. *Cornell Journal of Law and Public Policy 25*(3).

# A Rising Tide for All or Wave for One?: The Effect of Charter School Competition on District Achievement

Paolo Adajar
Final Draft, 14.33

5/11/2020

**Abstract**

We analyze the impact of each additional charter school on the achievement of its local school district. We use a national dataset of 3rd–8th grade students in the 2015–2016 school year from the Stanford Educational Opportunity Project and the US Department of Education Common Core of Data. We instrument our specifications with a measure of district location and the strength of state charter laws measured by the Center for Education Reform. Our primary specifications indicate that each additional charter school decreases math achievement by 0.0202SD and decreases ELA achievement by 0.0115SD, declines that are statistically significant at the 5% level and are equivalent to 1-2 weeks of learning. Analysis by racial subgroups indicate that white students are harmed relative to all minority groups, but no differences by gender. These results indicate that while charter schools reduce racial disparities in education, they lower achievement in both math and ELA for all but one subgroup.

Since Minnesota passed the first charter school legislation in 1991, charter schools have grown to serve over 3 million students in the United States. As of 2019, 45 states (alongside DC, Guam, and Puerto Rico) have laws permitting charter schools to form, with nearly 7,500 charter schools serving over 5% of American students [2]. Charter schools are publicly-funded but privately-run, and often are exempt from many of the requirements of traditional public schools. While charters are often praised for their innovative approaches, they also receive criticism for their impact on public school finances, lack of regulation, and under-representation of students with disabilities. Charter schools are incredibly contentious — data from the 2019 Education Next Poll shows that parents are split 50/50 on whether they support charter schools or not [2]. The Harvard Graduate School of Education describes charter schools as "one of today's most contentious debates in education" [18].





What is important from the perspective of educational policy is not just whether charter schools have positive impacts on their own students, but whether they have positive impacts on education at-large, particularly in the districts where they are located. In this paper, we estimate the causal impact of each additional charter school on the academic performance *all* students in their district, not just those that attend the charter school.

We use data from the Educational Opportunity Project at Stanford and the US Department of Education Common Core of Data. Our dataset covers 3rd–8th grade students in 2016 from all public school districts in the United States. A naive estimation may have reverse causality issues, as the decision to create a charter school in a certain district may be in part due to the academic achievement of the existing students. To account for this, we instrument our regressions with the interaction between a district's locale (urban, suburban, and rural) and the strength of the state's charter policy laws, as quantified by the Center for Education Reform.

Our primary specifications show that the entrance of a charter school has small but statistically significant negative impacts on the achievement of its district. A new charter school will decrease the achievement in math of the local school district by 0.020S2D, and decrease achievement in English/Language Arts (ELA) by 0.0115SD. These effects are measured quite precisely, and translate to learning decreases on the order of 1-2 weeks of learning for the typical student in 3th–8th grade [16].

Analysis by racial and gender subgroups shows that charter schools may impact different populations in different ways. White students fare worse due to charter school competition than any group of minority students in both math and ELA. While minorities are relatively benefited, our point estimates of effects are negative for all but one of our specifications by race and subject; as such, we do not believe that charter schools are an optimal policy instrument for reducing racial achievement gaps. No statistically significant differences are found when analyzing the impacts by gender.

This study makes two primary contributions to the literature. First, it is the only study known to the authors to estimate the causal impact of charter competition on traditional public-school achievement using a national dataset. This study furthers existing evidence on the impacts of charter schools in their communities. Hoxby (2003) analyzes charter schools in Michigan and Arizona, showing that charter school competition raises math and reading test scores in the district, describing competition as a "tide that lifts all boats" [3]. Gilraine et al. (2019) find small gains from North Carolina charter school competition caused by increased performance in charters with similar curricula to public schools and increased performance in traditional public schools themselves [14]. Apperson (2010) uses New York City zoning laws to find that competition raises neighborhood student achievement by small amounts, but also causes substantial sorting by race into schools [7]. Griffith (2019) finds correlations between a higher share of students in charter schools and academic performance, especially among urban minorities, using the SEDA dataset that is also used in this paper (see Subsection 1.1) [15]. In a meta-analysis of 11 different studies, Epple et al. report that all 11 studies had neutral to positive effects [13].

Second, this paper furthers existing evidence that the impacts of competition in educational settings are quite limited. Perhaps the most famous paper in this literature is Hoxby (2000), who uses stream density as an instrument for natural school barriers, finding that Tiebout choice raises student achievement while lowering school spending. Card et al. (2010), studying competition between a public school system and a Catholic-only school system in Ontario, find that expanding school choice to all students would increase $6^{\text{th}}$ grade achievement by .06 – .08SD [10].

The remainder of the paper proceeds as follows. Section 1 discusses the sources of data used in this analysis. Section 2 describes our estimation strategy and the motivation behind our instrumental variables approach. Section 3 presents our results of our regressions, including various checks for robustness and discusses its implications. Section 4 concludes.





# 1 Data Sources

## 1.1 District and School Data

Our primary data source on school information comes from the Educational Opportunity Project at Stanford University, which created the Stanford Education Data Archive (SEDA) to "help scholars, policymakers, educators, and parents learn how to improve educational opportunity for all children" [20].

The dataset measures educational outcomes across the United States at the school, district, and state level. This data includes the average achievement of each public school (traditional and charter) with grades 3-8 through school years 2008–09 through 2015–16. We analyze data from the 2015-16 school year at the district-grade level – that is, each grade of each district is a separate observation in our analysis. Test score data has been normalized across states and years using data from the National Assessment of Educational Progress (NAEP), in such a way that the average test score for each observation is 0 and that a grade in a district with a score of 1 has students that achieve one standard deviation higher than a typical student in that grade. SEDA also includes test score data grouped by race and gender, enabling subgroup analysis along these dimensions.

SEDA has a rich set of covariates, including racial demographics and categorical variables for the school's location (urban, rural, town, or suburban). SEDA also computes a socioeconomic composite index on a scale with mean 0 and variance 1 with respect to all districts in the United States, using data from the American Community Survey (ACS) and incorporating median income, unemployment rate, SNAP receipt rates, and other similar information.

SEDA additionally redefines school districts in a way that allows us to estimate the impact of charters on district achievement. Many charters are not part of any school district; in SEDA, these schools are "considered to be part of the district in which they are physically located" [5]. While all data was already publicly available, SEDA combines these all into more succinct format.

SEDA does not report data for many district-grade observations for a variety of reasons, which limits the external validity of our analysis. Many "small" districts lack data, as SEDA suppresses data if it represents less than 20 students or if the score normalization created imprecise estimates of test score means, removing the test scores of 34% of districts. In addition, observations with data anomalies, such as those with errors in reporting, incomparable tests, schools that could not be located, and states with low participation rates, are all dropped from the dataset, observations that represent around 6% of the sample. As such, our analysis is done with the caveat that it may not generalize to all schools within the U.S., particularly those with few students.

We use data from the National Center for Education Statistics Common Core of Data (CCD) to find the number of charter and non-charter public schools for each district and grade in the 2015–2016 school year.[1] Of the 258,679 school-grade observations in CCD, all but 414 were able to be located in the SEDA dataset, which appears to be due to random errors in reporting. The remaining 258,265 school-grade observations are a part of 78,738 distinct district-grade observations. We additionally drop 234 virtual school districts, 3,967 with incomplete or missing covariates, and 20,947 with missing math and ELA test results.

[1]SEDA is based on the the ED*Facts* database, though both CCD and ED*Facts* are managed by the U.S. Department of Education. As such, CCD and SEDA can be linked using Locale Education Agency (LEA) numbers.





This leaves 53,590 districts with full controls and either math or ELA reported for the 2015-16 school year that we use for our analysis.

Summary statistics for the variables of interest in the SEDA dataset are included in Table 1. Note that the districts with charters tend to have lower test scores, be in urban areas, have a lower average socioeconomic status, and have larger minority populations.

## 1.2 State Charter Laws Data

The Center for Education Reform (CER) is an advocacy group that aims to "expand educational opportunities that lead to improved economic outcomes for all Americans" [6]. Since 1996, CER has analyzed charter school laws to see how well they facilitate the creation of new charter schools. In particular, CER has released annual rankings and scorecards of charter school laws every year since 1999 [11]. The policy score is assigned at the state level, not the district level.

These rankings evaluate each law on a scale from 0 to 55, with points awarded depending on how well the state's law allows multiple independent entities that can authorize charter schools, no caps on the number of schools, operational independence, fiscal equity with other schools, and the implementation of the law. These categories are chosen because they "have the most impact on the development and creation of charter schools" [1]. We use CER's 2013 report in this study, normalize scores to a scale from 0 to 1, and summarize its data in Figure 1 [4].

## 2 Methods

We estimate the causal impact of each additional charter school per grade on the achievement of the same grade in the district where it is located. We do so using the regression

$$TestScore_i = \alpha Charters_i + \beta \mathbf{X}_i + \varepsilon_i$$

Here, $i$ represents a district-grade observation, $TestScore_i$ is the average test score (either math or reading) of the district-grade observation, $Charters_i$ is the number of charter schools present in a district for a given grade, and $\mathbf{X}_i$ is a vector of district-level controls (racial demographics, number of students, socioeconomic status, urban/suburban/town/rural location, grade fixed effects, and state fixed effects). Note that state fixed effects control for $Policy$, as $Policy$ is fixed for each state.

In addition to our main analysis, we perform a subgroup analysis on two dimensions: race and gender. As shown in Table 1, charter schools have higher populations of minority students, and so we may expect effects of charter schools to differ across this subgroup. Note that because of SEDA's data suppression, as described in Subsection 1.1, not all districts will have data reported for each subgroup. For example, consider a district with 100 students in the 8$^{\text{th}}$ grade where 50 are white, 30 are black, 10 are Hispanic, and 10 are Asian.Test scores would be available in SEDA for the whole district, white students, and black students, but no data would be available for Hispanic and Asian students.

We restrict each subgroup analysis to observations that report test scores for the relevant subgroup. Our hypothetical district would be included in the analyses for white and black students, but would not be included in analyses for Asian and Hispanic students. As a result, the estimates for each subgroup represent different sets of districts, and the analysis performed for each subgroup may not be directly comparable to the results for other subgroups. These estimates represent the effect of an additional charter school on test scores for a specific subgroup when the population of that subgroup is at least 20.





However, there may be issues with reverse causality — charters may selectively enter into districts where existing schools have lower performance. Further, there may be additional unobserved characteristics of districts that correlate with test scores that impact whether charter school entry occurs. To account for these issues, we instrument $Charters_i$ in all of our specifications using the variable of $Urban_i \times Policy_i$ in a 2SLS approach. Here, $Urban_i$ is an indicator variable for if the district is located in an urban locale. $Policy_i$ is the strength of each state's public charter laws in 2013 (see Subsection 1.2). Note that the nature of this instrument means that we estimate the local average treatment effect (LATE) for urban school districts in states with a high $Policy$ score.

We expect this instrument to be relevant due to the nature of charter school entry. It is well-documented that most charter schools tend to be located in urban areas, as reported by the New York Times [12], the Atlantic [19], and the Brookings Institution [17]. Further, the strength (and existence) of charter policy laws has a clear impact on charter school development. We use a lagged variable for policy because founding charter schools is not an instantaneous process. Researchers from the Peabody College at Vanderbilt University report that founding a charter school is a process that can take 1–2 years [9]. As our school data covers the 2015–16 school year, this implies that the 2013 policy measure will have an appropriate lag. We use the interaction between $Urban_i$ and $Policy_i$, rather than just $Policy_i$, to increase the relevance of our instrument.

We can also explicitly test for relevance by estimating the first stage of this regression, and using an $F-$test of the first-stage to see whether this instrument are strong. We find that this instrument for strong for each of our specifications, and we report the result of relevance tests alongside the main results in Section 3.

We would expect these instruments to be exogenous of if holding the controls fixed, there is no correlation between $Urban_i \times Policy_i$ and $TestScores_i$ except through the number of charter schools. We control for the location of each district directly and $Policy_i$ indirectly through state fixed-effects. As $Policy_i$ is assigned at a state level, $Policy_i$ is collinear with state-fixed effects and including it in our specification is redundant.

The most direct channel by which exogeneity may be violated because of omitted variable bias is that our instrument may be related to other education policies. In particular, districts with a high $Urban \times Policy$ value may be correlated with public school funding, school accountability measures, and curricula choices. However, given the bipartisan support for public schools through 2012, it is plausible that states with a high $Urban \times Policy$ are uncorrelated with other partisan educational policies and thus performance on standardized tests [8]. As such, it seems implausible that this channel could violate the exogeneity assumption.

We also do not believe threats to exogeneity to occur due to reverse causality. First, low test scores in specific areas of a state cause the state to pass charter law policies to target those locations. Given that we are already controlling for the location of each district, this does not appear to be plausible. Second, urban areas with strong charter law policies leads to student sorting across districts, and likewise for different locales and weak charter laws. Given the challenges and costs associated with moving between districts (and doing so solely for a child's schooling), we do not believe this to be a significant factor. Finally, success of said laws *prior* to 2013 may impact both laws in 2013 and current educational outcomes. We believe the impact of this channel to be minimal, given the wide variety of factors that impact the passage of laws, such as government regulations, which party is in power, and the lobbying of advocacy groups.





# 3 Results

## 3.1 Primary Results

Results of our main specification is shown in Table 2. The impact of each additional charter school on the achievement of its school district is negative and statistically significant. A new charter school reduces the math achievement of its district by 0.020SD and decreases the district's ELA achievement by 0.012SD. To give context for these estimates, the typical $3^{\text{th}}$–$8^{\text{th}}$ grade student learns 0.30SD in a given school year [16]. As such, these learning decreases are equivalent to 2 weeks and 1 weeks of school, respectively. While the negative impacts of each additional charter school are relatively small, they are significant at the 5% level.

In addition, we note the validity of our instruments in both specifications. The math specification has a first-stage $F = 166.4$, while for the ELA specification has $F = 159.3$. These results show that our instrument is quite strong and causes much variation in the number of charter schools.

We can additionally perform a robustness check of these results using two additional instruments – $Suburban \times Policy$ and $Rural \times Policy$, defined in the same way as $Urban \times Policy$ but with a different indicator variable for location. Given these three instruments, we can test the consistency of results using all possible combinations of these three instruments. Results of the robustness checks for the math specification are reported in Table 3, while those for ELA are reported in Table 4.

For math, the results of our main specification are robust to using different sets of instruments; all specifications estimate that the impact of a new charter school is close to $-0.0202$SD, though not all are statistically distinguishable from 0. In addition, all specifications but one pass the $F-$test. The specification which fails is the one using only

$Suburban \times Policy$ as an instrument, which may be due to limited numbers of charters in suburban areas.

However, for ELA scores, our results are less consistent, with only 3 of the 6 new regressions having impacts of a new charter school close in magnitude to $-0.0115$SD. Note that the specifications using only $Suburban \times Policy$ and using only $Rural \times Policy$ both have positive results, indicating that there are heterogeneous effects of charter schools in urban, suburban, and rural locations. This result is not surprising, as some anecdotal evidence already exists for this phenomenon [17,19]. Again, the lone specification that fails the $F-$test is instrumented using only $Suburban \times Policy$, likely for the same reason as that this instrument failed in the robustness checks for math.

### 3.2 Effects by Race

In Table 5, we report results the effect of charter schools on Math and ELA achievement for white, black, Hispanic, and Asian students. These results are also depicted graphically in Figure 2.

We observe significant changes in the sample size of each specification. As noted in Subsection 1.1 and Section 2, this is because our data set only includes test scores for a racial subgroup when that subgroup has at least 20 students, among other factors. As such, each specification contains a different set of district-grade observations. In addition, we note that as $F > 10$ for each of our specifications, our instrument is strong even when performing our analysis by racial subgroups.

In math, the largest difference between two subgroups is between Asian and white students; white students are harmed by an additional charter school in math by nearly two weeks of achievement relative to Asian students, though this difference is not statistically





significant. Further, white students are harmed relative to all minority groups because of charter school competition, but no difference is statistically significant. Perhaps most importantly, for all subgroups, the impact of an additional charter school on district achievement is negative, indicating that charter schools lower achievement in mathematics across the board.

Turning to ELA achievement, we find even larger differences between racial subgroups. In particular, the difference in effects of charter school competition for black and white students is 0.0325SD ($p = 0.001$), a difference equivalent to a month of learning in ELA for the typical student. Further, the point estimate for the impact on black student achievement in ELA is *positive*, albeit statistically indistinguishable from 0. Smaller differences occur between all other subgroup pairs, but we find that on charter schools benefit minority students, particularly black students, relative to white students. However, the net effect on district achievement in ELA is negative, and all subgroups have either negative impacts or impacts statistically indistinguishable from 0.

In summary, we find that white students are harmed from charter school competition relative to minority students, indicating that charter schools may reduce existing racial achievement gaps. However, the net effect of charter schools (and for almost all subgroups) is *negative*. As such, we do not find that charter schools are an effective policy intervention on the basis of increasing educational equity.

### 3.3 Effects by Gender

Results for subgroup analysis by gender are shown in Table 6 and Figure 3. We again have that our instrument is relevant for each subgroup that we consider, with $F > 150$ for each specification. In addition, the difference in sample sizes for these specifications is much smaller than those for our subgroup analysis by gender.

The effect of each charter school on math and ELA achievement is remarkably similar for both male and female students. While point estimates indicate that females fare slightly worse than males because of charter schools, the differences are not statistically significant at the 5% level. Further, our estimated differences are incredibly small in magnitude and are equivalent to just around one day of learning in the typical 180-day school year for ELA, and one-fifth of a day of learning in math [16]. As such, policy-makers should not be concerned about any differential impacts of charters on each gender.

## 4 Conclusion

We find small but statistically significant declines in math and ELA scores caused by the entrance of each additional charter school. Subgroup analysis indicates that white students are impacted more negatively than students of other minority groups. Results indicate that charter schools reduce racial achievement gaps, but do so by lowering achievement for all students. No such differences exist when disaggregating test scores by gender.

Our results run counter to the bulk of research that exists for charter schools, which show small but positive gains when a charter school enters. Future research is needed to determine the factors which lead certain school districts to benefit positively from charter school competition, and whether this is something that can be influenced by policy. Further analysis is also needed regarding the equity impacts of charter school legislation. In this study, we find that charter schools negatively impact white students compared to minorities. Careful econometric analysis is needed to further understand the welfare impacts of charter school policy, especially given the high minority population that typically enrolls in charter schools.

It is imperative that the results of studies like these are not interpreted in a vacuum; the





performance of students on standardized is just one of many outcomes that all schools try to impact. Schools also try to develop pro-social behaviors and career motivation, impacts which we are unable to study. Further, given the nature of charter school finances (privately operated but publicly funded), charter schools may have desirable impacts on school finances. Analyzing the impact of test scores is but one factor that policy-makers should use when considering public charter legislation.

## References


[1] 2014 Charter School Laws Ranking and Scorecard; The Rationale Behind the Rankings. https://edreform.com/wp-content/uploads/2014/03/CharterLawsRankingMethodologyFINAL2014.pdf.

[2] What is a Charter School? — NCSRC. https://charterschoolcenter.ed.gov/what-charter-school.

[3] *School Choice and School Productivity: Could School Choice Be a Tide That Lifts All Boats?*, volume The Economics of School Choice. University of Chicago Press, Chicago, January 2003.

[4] Charter School Law Rankings and Scorecard. Technical report, Center for Education Reform, January 2013.

[5] Stanford Education Data Archive (SEDA), September 2019.

[6] Expanding Innovation & Opportunity in American Education. https://edreform.com/about/, 2020.

[7] Jarod Apperson. Beyond Their Walls: A Decade of Evidence from Spatial Variation in Access to New York City Charter Schools. page 22.

[8] Evie Blad. Democrats Once Proposed Tripling Charter Schools. Here's What's Changed. http://blogs.edweek.org/edweek/campaign-k-12/2019/06/charter-schools-democrats-party-platforms.html?cmp=SOC-SHR-FB.

[9] Marisa Cannata, Grant Thomas, and Zaia Thombre. Starting Strong: Best Practices in Starting a Charter School. Technical report, Vanderbilt Peabody College, October 2014.







[10] David Card, Martin D Dooley, and A. Abigail Payne. School Competition and Efficiency with Publicly Funded Catholic Schools. *American Economic Journal: Applied Economics*, 2(4):150–176, October 2010.

[11] Center for Education Reform. Annual Charter School Law Report Card Issued.

[12] Susan Dynarski. Urban Charter Schools Often Succeed. Suburban Ones Often Don't. *The New York Times*, November 2015.

[13] Dennis Epple, Richard Romano, and Ron Zimmer. Charter Schools: A Survey of Research on Their Characteristics and Effectiveness. Working Paper 21256, National Bureau of Economic Research, June 2015.

[14] Michael Gilraine, Uros Petronijevic, and John D Singleton. Horizontal Differentiation and the Policy Effect of Charter Schools. page 61, June 2019.

[15] David Griffith. Rising Tide: Charter School Market Share and Student Achievement. Technical report, Fordham Institute, Washington, DC, September 2019.

[16] Carolyn J. Hill, Howard S. Bloom, Alison Rebeck Black, and Mark W. Lipsey. Empirical Benchmarks for Interpreting Effect Sizes in Research. *Child Development Perspectives*, 2(3):172–177, December 2008.

[17] Terry Ryan and Paul T. Hill. In a changing rural America, what can charter schools offer?, May 2017.

[18] Zachary Jason. The Battle Over Charter Schools. https://www.gse.harvard.edu/news/ed/17/05/battle-over-charter-schools.

[19] Laura McKenna. Why Charter Schools Are Hard to Find in Suburbia. https://www.theatlantic.com/education/archive/2015/10/why-dont-suburbanites-want-charter-schools/408307/, October 2015.

[20] The Educational Opportunity Project at Stanford University. The Educational Opportunity Project at Stanford — Home. https://edopportunity.org/.





Table 1: Summary statistics of SEDA dataset for 3rd–8th grade students in the 2015–16 school year. Observations are at the district-grade level. Data construction is described in Subsection 1.1.

| | All Schools | | | No Charters | | | Contains Charters | | |
|---|---|---|---|---|---|---|---|---|---|
| | mean | sd | median | mean | sd | median | mean | sd | median |
| **Outcome Variables** | | | | | | | | | |
| Math Scores | 0.02 | 0.44 | 0.02 | 0.04 | 0.44 | 0.04 | -0.14 | 0.41 | -0.13 |
| ELA Scores | 0.05 | 0.39 | 0.06 | 0.07 | 0.39 | 0.07 | -0.08 | 0.36 | -0.07 |
| **Regressors** | | | | | | | | | |
| Number of Charters | 0.34 | 2.57 | 0.00 | 0.00 | 0.00 | 0.00 | 3.11 | 7.14 | 1.00 |
| Number of Schools | 3.68 | 11.24 | 1.00 | 2.29 | 3.89 | 1.00 | 14.84 | 29.64 | 7.00 |
| Total Enrollment (100s) | 3.47 | 11.07 | 1.17 | 2.14 | 4.22 | 1.02 | 14.16 | 28.85 | 6.06 |
| Socioeconomic Status | 0.20 | 0.91 | 0.26 | 0.23 | 0.89 | 0.28 | -0.08 | 1.02 | 0.01 |
| Urban | 0.07 | 0.25 | 0.00 | 0.04 | 0.19 | 0.00 | 0.29 | 0.45 | 0.00 |
| Suburban | 0.22 | 0.41 | 0.00 | 0.20 | 0.40 | 0.00 | 0.32 | 0.47 | 0.00 |
| Town | 0.21 | 0.41 | 0.00 | 0.22 | 0.41 | 0.00 | 0.17 | 0.37 | 0.00 |
| Rural | 0.50 | 0.50 | 1.00 | 0.54 | 0.50 | 1.00 | 0.22 | 0.42 | 0.00 |
| White (%) | 0.72 | 0.28 | 0.84 | 0.75 | 0.27 | 0.86 | 0.50 | 0.29 | 0.51 |
| Black (%) | 0.08 | 0.17 | 0.02 | 0.07 | 0.16 | 0.01 | 0.17 | 0.23 | 0.06 |
| Hispanic (%) | 0.15 | 0.22 | 0.06 | 0.14 | 0.20 | 0.05 | 0.28 | 0.26 | 0.19 |
| Asian (%) | 0.02 | 0.05 | 0.01 | 0.02 | 0.05 | 0.01 | 0.04 | 0.06 | 0.02 |
| Native American (%) | 0.02 | 0.10 | 0.00 | 0.02 | 0.10 | 0.00 | 0.01 | 0.05 | 0.00 |
| Observations | 53590 | | | 47644 | | | 5946 | | |

*Notes:* District-grade level observations. Does not include virtual school districts, incomplete covariates, or observations with no reported math or ELA test scores. More details about dropped observations in Subsection 1.1. Source: Stanford Education Data Archive (SEDA) [5].

Table 2: The impact of the number of charter schools per grade on district-wide academic achievement. Achievement is normalized nationally to mean 0 and each unit represents a standard deviation of achievement. Both specifications are instrumented with a measure of location and policy in a 2SLS approach, as described in Section 2. Results are interpreted in Subsection 3.1.

| | (1) Math | (2) ELA |
|---|---|---|
| Number of Charters | -0.0202*** | -0.0115** |
| | (0.00489) | (0.00416) |
| Total Enrollment | 0.0000157** | 0.00000576 |
| | (0.00000599) | (0.00000370) |
| Socioeconomic Status | 0.233*** | 0.210*** |
| | (0.00247) | (0.00201) |
| Number of Schools | 0.00213** | 0.00158* |
| | (0.000804) | (0.000621) |
| Urban | 0.0123 | 0.0362*** |
| | (0.00635) | (0.00494) |
| Suburban | 0.0293*** | 0.0560*** |
| | (0.00488) | (0.00399) |
| Rural | -0.0502*** | -0.0336*** |
| | (0.00349) | (0.00286) |
| White (%) | 0.340*** | 0.365*** |
| | (0.0268) | (0.0203) |
| Black (%) | -0.150*** | -0.0520* |
| | (0.0292) | (0.0219) |
| Hispanic (%) | -0.0236 | -0.00644 |
| | (0.0273) | (0.0206) |
| Asian (%) | 1.268*** | 1.089*** |
| | (0.0427) | (0.0359) |
| Observations | 50385 | 52988 |
| Adjusted $R^2$ | 0.541 | 0.576 |
| First-stage $F$ | 166.4 | 159.3 |
| State/Grade Fixed Effects | yes | yes |

*Notes:* District-grade level observations. Due to data suppression, each regression considers only district-grade observations which report scores; see Subsection 1.1 for more information. Each specification instrumented with $Urb \times Pol$ as described in Section 2. Robust standard errors in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.





Table 3: Robustness checks for our main specification for math scores. Each column uses a different set of instruments in a 2SLS approach. Column (1) is identical to column (1) of Table 2. Results are interpreted in Subsection 3.1.

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| Number of Charters | -0.0202*** | -0.0183 | -0.200 | -0.0202*** | -0.0217*** | -0.0115 | -0.0216*** |
| | (0.00489) | (0.0160) | (0.193) | (0.00489) | (0.00489) | (0.0153) | (0.00489) |
| Total Enrollment | 0.0000157** | 0.0000146 | 0.000120 | 0.0000157** | 0.0000166** | 0.0000107 | 0.0000166** |
| | (0.00000599) | (0.0000104) | (0.000126) | (0.00000597) | (0.00000628) | (0.00000947) | (0.00000627) |
| Socioeconomic Status | 0.233*** | 0.233*** | 0.212*** | 0.233*** | 0.233*** | 0.234*** | 0.233*** |
| | (0.00247) | (0.00300) | (0.0232) | (0.00247) | (0.00247) | (0.00294) | (0.00247) |
| Number of Schools | 0.00213** | 0.00187 | 0.0268 | 0.00212** | 0.00233** | 0.000934 | 0.00233** |
| | (0.000804) | (0.00224) | (0.0273) | (0.000805) | (0.000826) | (0.00211) | (0.000826) |
| Urban | 0.0123 | 0.0127 | -0.0314 | 0.0123 | 0.0119 | 0.0144* | 0.0119 |
| | (0.00635) | (0.00736) | (0.0577) | (0.00635) | (0.00642) | (0.00706) | (0.00642) |
| Suburban | 0.0293*** | 0.0300*** | -0.0291 | 0.0294*** | 0.0289*** | 0.0322*** | 0.0289*** |
| | (0.00488) | (0.00695) | (0.0653) | (0.00488) | (0.00488) | (0.00679) | (0.00488) |
| Rural | -0.0502*** | -0.0503*** | -0.0390** | -0.0502*** | -0.0501*** | -0.0507*** | -0.0501*** |
| | (0.00349) | (0.00361) | (0.0129) | (0.00349) | (0.00349) | (0.00360) | (0.00349) |
| White (%) | 0.340*** | 0.339*** | 0.426*** | 0.340*** | 0.340*** | 0.336*** | 0.340*** |
| | (0.0268) | (0.0282) | (0.0983) | (0.0268) | (0.0268) | (0.0281) | (0.0268) |
| Black (%) | -0.150*** | -0.153*** | 0.0878 | -0.150*** | -0.148*** | -0.162*** | -0.148*** |
| | (0.0292) | (0.0363) | (0.260) | (0.0292) | (0.0292) | (0.0357) | (0.0292) |
| Hispanic (%) | -0.0236 | -0.0232 | -0.0615 | -0.0236 | -0.0239 | -0.0218 | -0.0239 |
| | (0.0273) | (0.0274) | (0.0513) | (0.0273) | (0.0273) | (0.0274) | (0.0273) |
| Asian (%) | 1.268*** | 1.271*** | 1.012*** | 1.268*** | 1.266*** | 1.281*** | 1.266*** |
| | (0.0427) | (0.0472) | (0.280) | (0.0427) | (0.0427) | (0.0468) | (0.0427) |
| Observations | 50385 | 50385 | 50385 | 50385 | 50385 | 50385 | 50385 |
| Adjusted $R^2$ | 0.541 | 0.542 | 0.124 | 0.541 | 0.541 | 0.544 | 0.541 |
| State/Grade Fixed Effects | yes | yes | yes | yes | yes | yes | yes |
| Instrumented with $Urb \times Pol$ | yes | no | no | yes | yes | no | yes |
| Instrumented with $Rur \times Pol$ | no | yes | no | yes | no | yes | yes |
| Instrumented with $Sub \times Pol$ | no | no | yes | no | yes | yes | yes |
| First-stage $F$ | 166.4 | 159.4 | 1.778 | 97.00 | 113.7 | 80.02 | 78.34 |

*Notes:* District-grade level observations. Due to data suppression, each regression considers only district-grade observations which report scores; see Subsection 1.1 for more information. Each specification instrumented with the set of instruments indicated. Robust standard errors in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 4: Robustness checks for our main specification for math scores. Each column uses a different set of instruments in a 2SLS approach. Column (1) is identical to column (2) of Table 2. Results are interpreted in Subsection 3.1.

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
|---|---|---|---|---|---|---|---|
| Number of Charters | -0.0115** | 0.0122 | 0.0881 | -0.0103* | -0.00895* | 0.00932 | -0.00888* |
| | (0.00416) | (0.0124) | (0.0516) | (0.00414) | (0.00405) | (0.0123) | (0.00405) |
| Total Enrollment | 0.00000576 | -0.00000741 | -0.0000497 | 0.00000511 | 0.00000435 | -0.00000581 | 0.00000431 |
| | (0.00000370) | (0.00000759) | (0.0000342) | (0.00000353) | (0.00000332) | (0.00000732) | (0.00000331) |
| Socioeconomic Status | 0.210*** | 0.213*** | 0.221*** | 0.210*** | 0.210*** | 0.212*** | 0.210*** |
| | (0.00201) | (0.00240) | (0.00640) | (0.00201) | (0.00201) | (0.00238) | (0.00201) |
| Number of Schools | 0.00158* | -0.00166 | -0.0120 | 0.00142* | 0.00123* | -0.00127 | 0.00122* |
| | (0.000621) | (0.00174) | (0.00759) | (0.000609) | (0.000588) | (0.00171) | (0.000588) |
| Urban | 0.0362*** | 0.0413*** | 0.0576*** | 0.0364*** | 0.0367*** | 0.0406*** | 0.0367*** |
| | (0.00494) | (0.00568) | (0.0169) | (0.00491) | (0.00487) | (0.00558) | (0.00487) |
| Suburban | 0.0560*** | 0.0638*** | 0.0890*** | 0.0563*** | 0.0568*** | 0.0629*** | 0.0568*** |
| | (0.00399) | (0.00559) | (0.0181) | (0.00398) | (0.00395) | (0.00556) | (0.00395) |
| Rural | -0.0336*** | -0.0351*** | -0.0398*** | -0.0337*** | -0.0338*** | -0.0349*** | -0.0338*** |
| | (0.00286) | (0.00294) | (0.00438) | (0.00286) | (0.00286) | (0.00293) | (0.00286) |
| White (%) | 0.365*** | 0.354*** | 0.320*** | 0.364*** | 0.363*** | 0.355*** | 0.363*** |
| | (0.0203) | (0.0213) | (0.0315) | (0.0204) | (0.0203) | (0.0212) | (0.0203) |
| Black (%) | -0.0520* | -0.0815** | -0.176* | -0.0534* | -0.0551* | -0.0779** | -0.0552* |
| | (0.0219) | (0.0269) | (0.0691) | (0.0220) | (0.0219) | (0.0267) | (0.0219) |
| Hispanic (%) | -0.00644 | -0.00259 | 0.00977 | -0.00625 | -0.00603 | -0.00306 | -0.00602 |
| | (0.0206) | (0.0207) | (0.0230) | (0.0206) | (0.0206) | (0.0207) | (0.0206) |
| Asian (%) | 1.089*** | 1.128*** | 1.254*** | 1.091*** | 1.093*** | 1.123*** | 1.093*** |
| | (0.0359) | (0.0407) | (0.0933) | (0.0359) | (0.0359) | (0.0405) | (0.0358) |
| Observations | 52988 | 52988 | 52988 | 52988 | 52988 | 52988 | 52988 |
| Adjusted $R^2$ | 0.576 | 0.575 | 0.469 | 0.576 | 0.577 | 0.576 | 0.577 |
| State/Grade Fixed Effects | yes | yes | yes | yes | yes | yes | yes |
| Instrumented with $Urb \times Pol$ | yes | no | no | yes | yes | no | yes |
| Instrumented with $Rur \times Pol$ | no | yes | no | yes | no | yes | yes |
| Instrumented with $Sub \times Pol$ | no | no | yes | no | yes | yes | yes |
| First-stage $F$ | 159.3 | 168.3 | 7.997 | 98.53 | 112.4 | 84.47 | 76.98 |

*Notes:* District-grade level observations. Due to data suppression, each regression considers only district-grade observations which report scores; see Subsection 1.1 for more information. Each specification instrumented with the set of instruments indicated. Robust standard errors in parentheses.
$^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$.





Table 5: The impact of the number of charter schools per grade on district-wide academic achievement, broken down by racial subgroups. Achievement is normalized nationally to mean 0 and each unit represents a standard deviation of achievement. Both specifications are instrumented with a measure of location and policy in a 2SLS approach, as described in Section 2. Results are interpreted in Subsection 3.2.

| | Math | | | | ELA | | | |
|---|---|---|---|---|---|---|---|---|
| | (1) White | (2) Black | (3) Hispanic | (4) Asian | (5) White | (6) Black | (7) Hispanic | (8) Asian |
| Number of Charters | -0.0192** | -0.0166 | -0.00503 | -0.00452 | -0.0243*** | 0.00821 | -0.00994 | -0.0156 |
| | (0.00652) | (0.00866) | (0.00602) | (0.0121) | (0.00719) | (0.00766) | (0.00618) | (0.0158) |
| Total Enrollment | 0.0000102* | 0.0000175* | -0.00000255 | 0.00000805 | 0.00000945 | 0.00000875 | 0.00000370 | 0.0000250 |
| | (0.00000514) | (0.00000711) | (0.00000494) | (0.0000124) | (0.00000567) | (0.00000571) | (0.00000507) | (0.0000140) |
| Socioeconomic Status | 0.274*** | 0.133*** | 0.122*** | 0.398*** | 0.242*** | 0.152*** | 0.119*** | 0.331*** |
| | (0.00273) | (0.00551) | (0.00436) | (0.0115) | (0.00233) | (0.00465) | (0.00402) | (0.0102) |
| Number of Schools | 0.00389** | 0.00130 | 0.000988 | 0.00131 | 0.00498*** | -0.00237* | 0.00141 | 0.00179 |
| | (0.00127) | (0.00127) | (0.000839) | (0.00142) | (0.00123) | (0.00121) | (0.000864) | (0.00223) |
| Urban | 0.0731*** | -0.0363*** | 0.0131 | 0.172*** | 0.0837*** | 0.00537 | 0.0256*** | 0.121*** |
| | (0.00765) | (0.0107) | (0.00845) | (0.0295) | (0.00677) | (0.00913) | (0.00764) | (0.0292) |
| Suburban | 0.00625 | 0.0161 | 0.0688*** | 0.120*** | 0.0196*** | 0.0789*** | 0.0899*** | 0.0892** |
| | (0.00549) | (0.0123) | (0.00875) | (0.0302) | (0.00490) | (0.0109) | (0.00809) | (0.0322) |
| Rural | -0.0674*** | -0.0302*** | 0.00719 | 0.0755* | -0.0503*** | 0.00139 | 0.0124 | 0.0546 |
| | (0.00374) | (0.00863) | (0.00755) | (0.0316) | (0.00312) | (0.00747) | (0.00640) | (0.0299) |
| White (%) | -0.217*** | 0.0571 | 0.229* | 5.294*** | -0.185*** | -0.0347 | 0.289*** | 5.247*** |
| | (0.0431) | (0.128) | (0.0940) | (0.703) | (0.0395) | (0.138) | (0.0857) | (0.658) |
| Black (%) | -0.0602 | -0.0569 | -0.00163 | 5.425*** | 0.0208 | -0.0577 | 0.145 | 5.417*** |
| | (0.0460) | (0.128) | (0.0944) | (0.704) | (0.0418) | (0.138) | (0.0864) | (0.655) |
| Hispanic (%) | -0.174*** | -0.105 | -0.0107 | 5.245*** | -0.139*** | -0.100 | 0.0397 | 5.221*** |
| | (0.0439) | (0.126) | (0.0927) | (0.692) | (0.0402) | (0.137) | (0.0846) | (0.651) |
| Asian (%) | 0.308*** | 0.346* | 0.342*** | 5.722*** | 0.304*** | 0.241 | 0.369*** | 5.321*** |
| | (0.0598) | (0.139) | (0.102) | (0.700) | (0.0579) | (0.146) | (0.0942) | (0.658) |
| Observations | 43070 | 10771 | 14793 | 5778 | 45377 | 11434 | 15569 | 5404 |
| Adjusted $R^2$ | 0.420 | 0.407 | 0.395 | 0.598 | 0.415 | 0.426 | 0.394 | 0.577 |
| First-stage $F$ | 127.0 | 60.87 | 116.4 | 51.71 | 138.4 | 58.72 | 101.2 | 34.02 |
| State/Grade Fixed Effects | yes | yes | yes | yes | yes | yes | yes | yes |

*Notes:* District-grade level observations. Due to data suppression, each regression considers only district-grade observations which report scores; see Subsection 1.1 for more information. Each specification instrumented with $Urb \times Pol$ as described in Section 2. Robust standard errors in parentheses.
$^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$.

Table 6: The impact of the number of charter schools per grade on district-wide academic achievement, broken down by gender subgroups. Achievement is normalized nationally to mean 0 and each unit represents a standard deviation of achievement. Both specifications are instrumented with a measure of location and policy in a 2SLS approach, as described in Section 2. Results are interpreted in Subsection 3.3.

| | Math | | ELA | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| | Male | Female | Male | Female |
| Number of Charters | -0.0192*** | -0.0195*** | -0.0119* | -0.0143*** |
| | (0.00522) | (0.00493) | (0.00464) | (0.00430) |
| Total Enrollment | 0.0000128* | 0.0000158** | 0.00000502 | 0.00000708 |
| | (0.00000620) | (0.00000596) | (0.00000399) | (0.00000423) |
| Socioeconomic Status | 0.247*** | 0.218*** | 0.208*** | 0.208*** |
| | (0.00273) | (0.00261) | (0.00223) | (0.00219) |
| Number of Schools | 0.00212* | 0.00199* | 0.00156* | 0.00204** |
| | (0.000857) | (0.000796) | (0.000694) | (0.000677) |
| Urban | 0.0237*** | 0.00484 | 0.0395*** | 0.0330*** |
| | (0.00669) | (0.00646) | (0.00535) | (0.00518) |
| Suburban | 0.0385*** | 0.0265*** | 0.0599*** | 0.0547*** |
| | (0.00528) | (0.00514) | (0.00438) | (0.00430) |
| Rural | -0.0472*** | -0.0397*** | -0.0264*** | -0.0279*** |
| | (0.00391) | (0.00383) | (0.00326) | (0.00317) |
| White (%) | 0.343*** | 0.251*** | 0.386*** | 0.326*** |
| | (0.0322) | (0.0290) | (0.0242) | (0.0231) |
| Black (%) | -0.199*** | -0.212*** | -0.0688** | -0.0805** |
| | (0.0342) | (0.0309) | (0.0257) | (0.0245) |
| Hispanic (%) | -0.0370 | -0.110*** | 0.00660 | -0.0672** |
| | (0.0321) | (0.0290) | (0.0242) | (0.0232) |
| Asian (%) | 1.286*** | 1.140*** | 1.082*** | 1.043*** |
| | (0.0487) | (0.0444) | (0.0398) | (0.0391) |
| Observations | 41567 | 40682 | 43592 | 43000 |
| Adjusted $R^2$ | 0.586 | 0.547 | 0.589 | 0.602 |
| First-stage $F$ | 157.6 | 160.8 | 150.1 | 154.0 |
| State/Grade Fixed Effects | yes | yes | yes | yes |

*Notes:* District-grade level observations. Due to data suppression, each regression considers only district-grade observations which report scores; see Subsection 1.1 for more information. Each specification instrumented with $Urb \times Pol$ as described in Section 2. Robust standard errors in parentheses.
$^{*}$ $p < 0.05$, $^{**}$ $p < 0.01$, $^{***}$ $p < 0.001$.





Figure 1: Distribution of Center for Education Reform charter school law scores (normalized) for each district-grade observation, as described in Subsection 1.2. Scores are assigned at the state level. Source: Center for Education Reform [4].

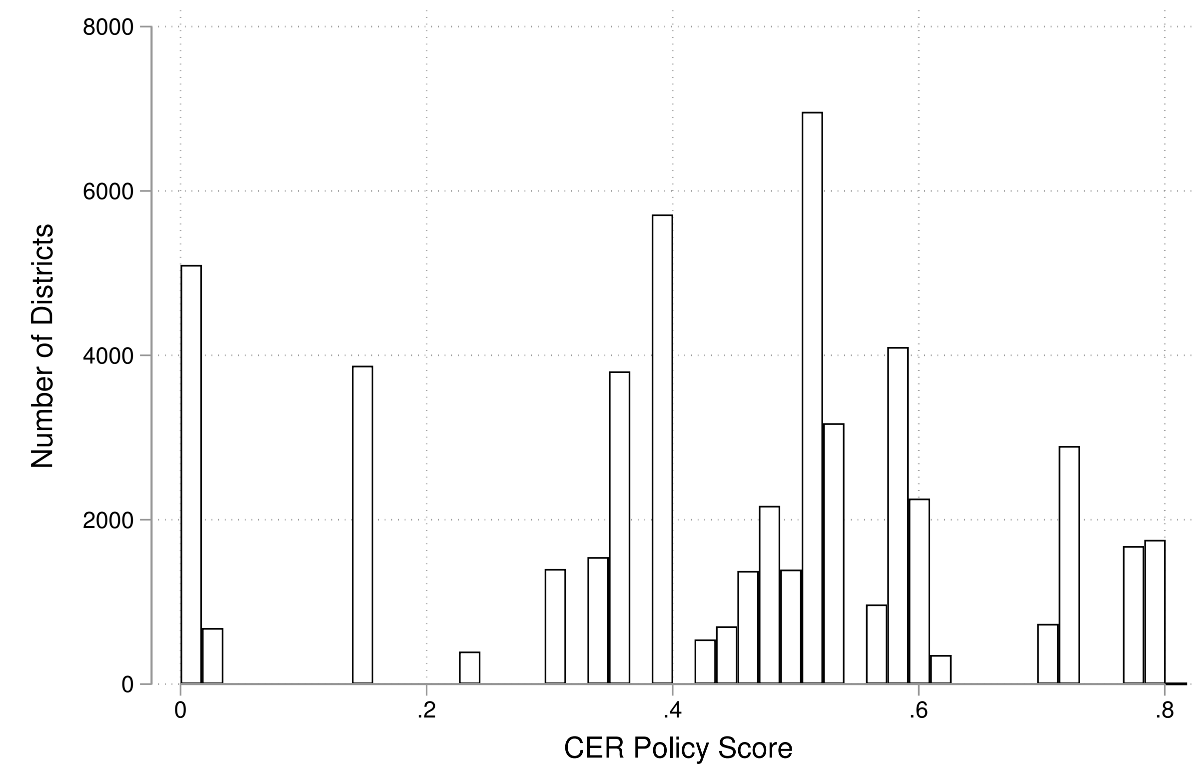
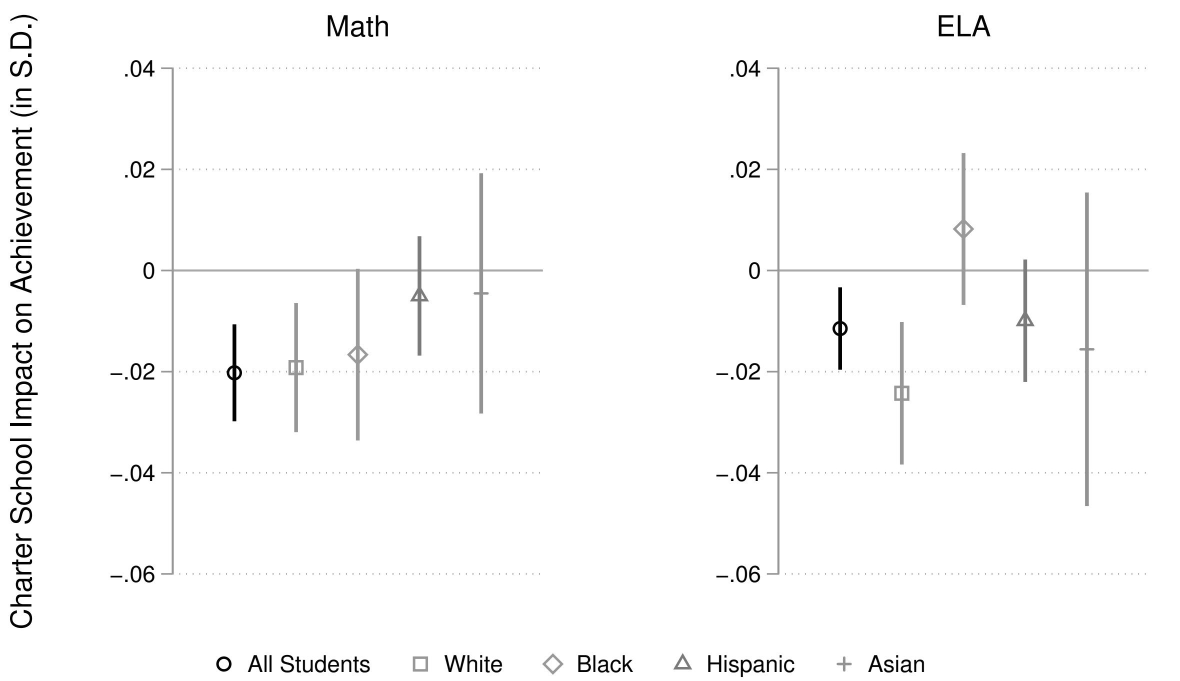


Figure 2: The 95% confidence interval for effect of new charter school on district achievement by subject and race, as reported in Table 5.

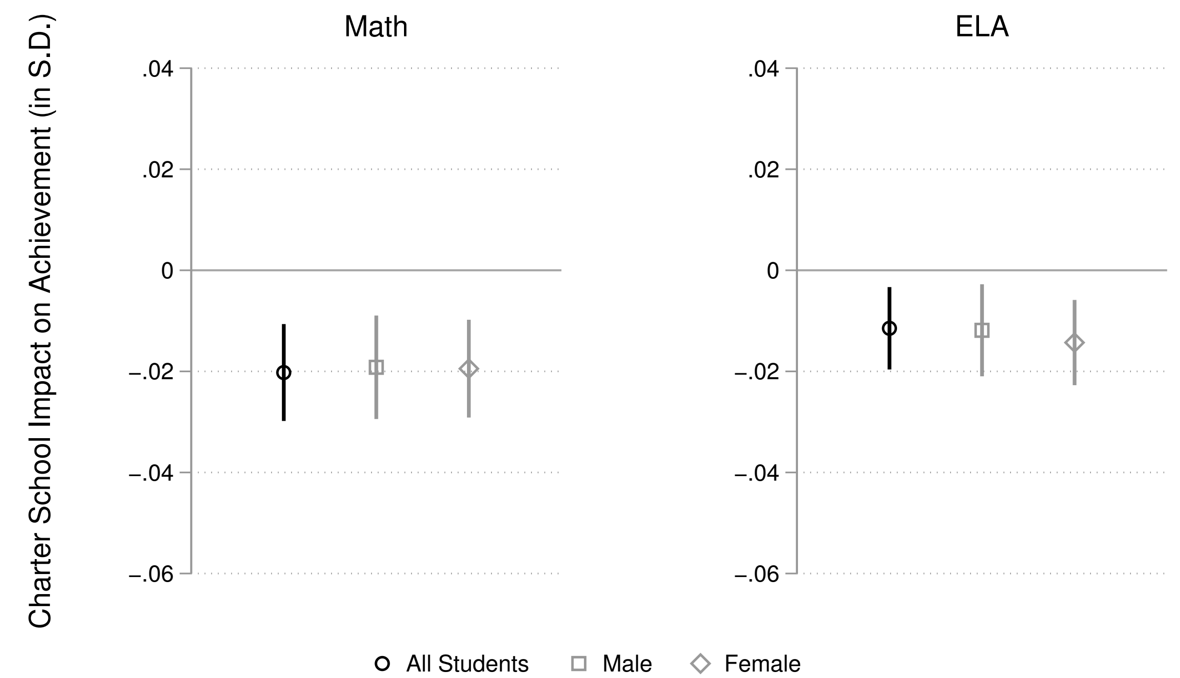



District−grade level observations. Due to data suppression, each regression considers only district−grade observations which report scores. Each specification instrumented with measure of location and policy. Sample sizes from left to right for Math are 50385, 43070, 10771, 14793, 5778; for ELA, 52988, 45377, 11434, 15569, 5404.





Figure 3: The 95% confidence interval for effect of new charter school on district achievement by subject and gender, as reported in Table 6.




District−grade level observations. Due to data suppression, each regression considers only district−grade observations which report scores. Each specification instrumented with measure of location and policy. Sample sizes from left to right for Math are 50385, 41567, 40682; for ELA, 52988, 43592, 43000.

# The Social Impacts of Public Sustainability Projects: Evidence from the EPA Brownfields Area Wide Planning Grant Program[1]

**Rikita Bansal**
rikita@mit.edu
Massachusetts Institute of Technology
77 Massachusetts Avenue
Cambridge, MA, 02139

**Abstract.** Neighborhood greening projects often instigate the gradual replacement of lower-income and vulnerable populations with more affluent persons, an example of environmental gentrification. Using panel data tracking EPA-designated Brownfields locations across the United States, this paper extracts the causal effect of sustainable public infrastructure on gentrification on a shorter-term time scale, as measured by Starbucks growth. Using a linear regression model controlling for ZIP Code and year fixed effects, this analysis indicates that such projects lead to early indicators of gentrification, emphasizing the importance of championing environmental justice alongside climate protection.

**Keywords:** Environmental Gentrification, Public Sustainability, Brownfields, Redevelopment, Greening.

## 1 Introduction

### *1.1 Motivation*

Climate change is the biggest threat to health in human history, with the World Health Organization conservatively estimating 250,000 additional deaths per year between 2030 and 2050 as a result of global greenhouse gas emissions (World Health Organization, 2018). In addition to increasing mortality, the World Bank estimates that climate change will force over 100 million people into poverty by 2030 (Rettner, Rachel. 2019). Although all communities feel the effects of a warming planet, environmental burdens and protections continue to be distributed unevenly along racial and socioeconomic lines.

While sustainability and green urbanism are necessary adaptation and mitigation measures in response to climate change, such projects often disenfranchise the working class. Environmen-

[1]All Code Written for this Analysis is Documented on https://github.com/rikitabansal under ”Sustainability and Gentrification: Starbucks in EPA Brownfields





tal Gentrification occurs when cleaning up contaminated areas or providing green amenities increases local property values and attracts wealthier residents to a previously polluted or disenfranchised neighborhood, thereby replacing original residents. The scarcity of housing opportunities in working-class and lower-income neighborhoods is further exacerbated in cities with large racial or ethnic minority populations. It has been long-recognized that urban green spaces are important to individuals' physical and mental health, as well as community well-being, cohesion, and resilience (Bemido-Rung et al., 2005). However, re-development of derelict land in New York City has often increased property values, enhanced neighborhood aesthetics, and reduced crime rates sufficiently for developers to invest in luxury housing projects that gentrify the very neighborhoods that the original sustainability projects sought to improve (Ottman et al., 2010).

In the process of constructing sustainable cities, it is critical to challenge the presumed inevitability of gentrification, and work to maintain diversity and equity. Existing research is limited to urban areas, where demographic changes instigated by sustainability projects have been documented on fairly gradual timescales (10+ years), given that they are not immediately visible. This research uniquely identifies systemic indicators of gentrification that hold in communities of all sizes throughout the United States. This paper explores the social impacts of public sustainability projects in polluted communities by using a combined entity and time fixed effects model to extract the causal effect of community greening projects on the short term growth in Starbucks frequency, where an increase in the number of Starbucks is an indicator of gentrification.

### *1.2 Background and Existing Literature*

Running from 2010 through 2019, the Environmental Protection Agency's Brownfields Area Wide Planning (AWP) Program provided grant funding and technical assistance to Brownfields communities selected via a national grant competition. A Brownfield is defined as a property, the expansion, redevelopment, or reuse of which may be complicated by the presence or potential presence of a hazardous substance, pollutant, or contaminant (Overview of EPA's Brownfields Program, 2020).

Cleaning up and reinvesting in any of the estimated 450,000 Brownfields properties has the potential to increase local tax bases, facilitate job growth, remove development pressures of open land, and both improve and protect the environment. Among the eight categories of grant funding, this paper focuses on AWP grants, where funding is designated for a specific project area, such as a neighborhood, downtown district, local commercial corridor, old industrial corridor, community waterfront or city block, and affects a single large Brownfield or multiple Brownfield sites. Other grant focus areas include community outreach or resident education, whereas AWP grants emphasize sustainable public infrastructure, which more concretely has the potential to trigger environmental gentrification.

Various studies have explored the greening gentrification phenomenon without extracting a causal effect, including a 2018 City University of New York study finding an association between proximity to community gardens and significant increases in per capita income over a five year study period (Maantay et al., 2018), highlighting how the environmental justice movement has often been appropriated to serve high-end development, sacrificing equity for profit. This paper also highlights how studies extracting shorter term gentrification are based in locations where gentrification is already evident in the pre-treatment period. Typically, studies uncover the instigation of gentrification through measuring urban displacement and affordable housing (Feldman, 2014), where the time scales need to be ten years or longer. This is because the quantitative approaches usually look at changes over time in household income, educational levels, rent or home prices,





percentage of non-Hispanic White population, unemployment rates, and sometimes percentage of the adult population employed in professional jobs. These are typically analyzed as individual variables, or used in some combination to create an index. Regardless of specifications, these changes are not immediate and therefore the effects can only be captured with an extended post-treatment period.

The motivation for using Starbucks as an indicator of gentrification is grounded in the findings of a 2018 Harvard Business School study, where the entry of Starbucks into already-gentrifying neighborhoods across the United States was significantly indicative of housing price growth and increases in the share of college educated persons within an area (Glaeser et al., 2018). The findings that businesses respond to contemporaneous gentrification as measured by housing prices, physical quality of neighborhoods, and demographic composition, is the basis for using a surrogate model for this study. Because the outcome of gentrification cannot be easily or directly quantified, the change in number of Starbucks locations is used instead as the outcome variable in this study.

## 2 Setting and Data

### *2.1 AWP Grant Recipients*

The EPA offered four rounds of Area-Wide Planning funding to a total of 83 recipients before officially terminating the program in 2019.

In Fall 2010, the EPA selected the first 23 AWP grant recipients, who each received $ 175,000 in funding and/or direct technical assistance from the agency. In Spring 2013, the EPA awarded 20 communities $200,000 to develop area-wide revitalization plans. In March of 2015, an additional 21 communities received $200,000 and the final funding cycle in January 2017 awarded 19 communities the same amount[2].

### 2.2 *Datasets*

ZIP Code level Panel data for the treatment and control groups between 2005 and 2020 was constructed by compiling AWP grant recipient data, AWP grant applicant data, and Starbucks opening-date data.

To populate the treatment group, the EPA's Grant Fact Sheet Search was used with appropriate filters[3] to extract the dataset containing the names of all grant recipients and their corresponding funding year. Each of the 595 treated ZIP Codes was then manually extracted by parsing each of the 83 Grant Fact Sheets as outlined in Appendix 7.2.

Data regarding Brownfields that applied for, but did not receive AWP funding in each funding cycle was not publicly available prior to this study. In order to populate the control group, the Freedom of Information Act Request (FOIA) as attached in Appendix 7.3 was submitted to the administrator of the EPA in order to obtain a list of all applicants in each funding year. Each of the 3,825 control ZIP codes was then extracted from the 303 grant applications using the CDXZip-Stream Microsoft Excel add-in.

Finally, a database aggregated by an independent third party, titled Starbucks Everywhere, was used to obtain the street name, city, and opening date for every Starbucks location in the United States. This database has been updated weekly by the creator since October of 2020. Using data updated until November 5th, 2020, the ZIP Code of each Starbucks was extracted using the ArcGIS Pro Geolocator tool with street name and city as the search values. The 5,743 Starbucks locations considered in the analysis can be seen in Figure 1.

[2]Recipient and Applicant Information is available on Github
[3]As outlined in Appendix 7.1





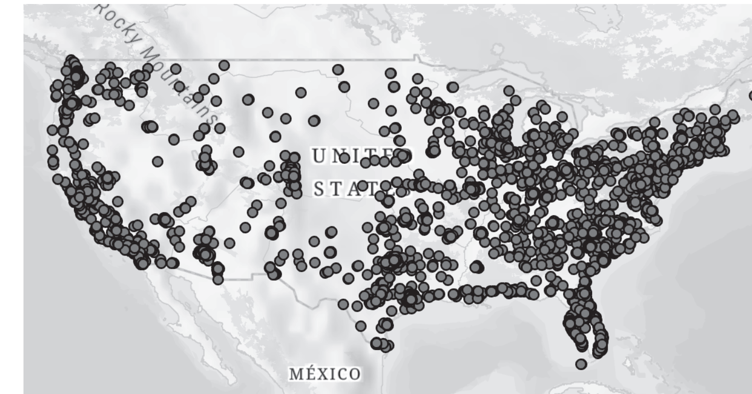



Figure 1: Starbucks Locations in the United States

The statistical software R was used to create panel data tracking ZIP Code level observations of grant treatment and Starbucks frequency each year between 2005 and 2020 by merging[4] the three data sets by ZIP Code and application year.

## 3 Empirical Strategy

### 3.1 *Research Design*

Disentangling the simultaneous relationship between public sustainability projects and gentrification poses a significant empirical challenge due to their co-determination (Autor et al., 2017). Using AWP grant funding as the treatment mitigates many of these issues by providing a clean exogenous measure of exposure to subsequent gentrification forces. Panel data on Starbucks frequencies at the ZIP Code level allows us to account for fixed differences across location — most importantly, the heterogeneity that exists in baseline Starbucks frequencies across Brownfields — and ascertain whether and how the frequency of Starbucks changed in response to post-sustainability project gentrification.

[4]Code available on Github, along with ZIP Code-populated datasets

The following entity and time fixed effects model is used in order to extract the casual effect of receiving an AWP grant on the number of Starbucks within a ZIP Code:

$$y_{it} = \sum_{m \in M} d_{it}^{m} \beta_m + \gamma_i + \mu_t + \epsilon_{it}$$

In this regression, the outcome $y_{it}$ is the number of Starbucks at ZIP Code $i$ at time $t$. $d_{it}^{m}$ is a dummy for ZIP Code $i$ receiving the AWP grant $m$ years relative to time $t$, with $M$ being the full set of dummies, where the value is 1 for grant recipients and 0 for applicants. This event study tracks changes over 10 year periods centered on funding year, such that $M = \{-5, -4, -3, -2, 0, 1, 2, 3, 4, 5\}$, where $m = -1$ is omitted as the base period and $d_{it}^{m}$ is defined as 0 (untreated) or 1 (treated) for each $m \in M$. $\gamma_i$ and $\mu_t$ are binary regressors for ZIP Code and year, respectively. Finally, $\epsilon_{it}$ is the residual, where standard errors are clustered at ZIP Code, given that Starbucks location frequency within a ZIP Code is not independent across years.

The coefficient of interest in this specification is $\beta_m$, which measures the differential change in the number of Starbucks for each post AWP funding period $m$. The combined entity and time fixed effects model allows us to eliminate bias from non observable factors that change between years but are constant over ZIP Codes and controls for factors that differ by location but are constant over time.

### *3.2 Assumptions*

In order for estimates of $\beta_m$ to represent the causal effect of AWP sustainability projects on gentrification as indicated by Starbucks opening frequency, we require that multiple assumptions hold. First, the grant treatment must be exogenous despite the grant assignment being a selection process that is non-random. We can still determine a causal effect if we assume that the treatment and control group exhibit parallel trends prior to treatment. This means that in the absence of funding, the two groups would have exhibited constant differences over time. As shown in Figure 2, Starbucks frequencies pre-intervention indicate that this assumption holds for all four funding years. Having the communities that applied for, but did not receive, funding as the control group provides a unique opportunity to control for communities that have similar needs and environmental hazard factors, across both rural and urban regions. Next, we assume that the regression measures only the changes in Starbucks locations that are caused by the Area-Wide Projects and not other factors. Finally, it is critical that we assume Starbucks locations are a valid surrogate for gentrification, a claim rooted in the aforementioned Harvard Yelp study.





## 4 Results

### *4.1 Regression Coefficient*

The results of regressing the number of Starbucks in a given ZIP on treated Brownfield locations five years post treatment are statistically significant as displayed in Table 1. The coefficient represents the changes five years post-intervention given that the full set of dummies ranged from -5 to 5, with $t = -1$ as the base year.

The coefficient $\beta_m$ indicates that Brownfield locations that received AWP grants have approximately 0.059 additional Starbucks open five years post funding year, as compared to untreated Brownfield locations. The clustered standard error was only 0.001353, making the results significant at the 0.1% level, indicating a very strong result. Finally, the adjusted $R^2$ value of 0.794 indicates that treatment status explains approximately 80% of the variation in Starbucks frequencies 5 years post treatment.

| | Coefficient | Clustered SE | Median Residual | Pr(>\|t\|) | Adjusted R-Squared |
|---|---|---|---|---|---|
| **Number of Starbucks ~ Time post grant receival** | 0.059091 | 0.001353 | -0.1414 | <2e-16*** | 0.794 |

***Reject the null at the 0.001 level of significance

Table 1: ZIP and Year Fixed Effects Linear Regression

### 4.2 *Event Study*

The event study graph displayed below is important in verifying the parallel trends assumption, where we see that the average new Starbucks frequency for the treated ZIP codes displayed in red follows the same pattern pre-base period as the control ZIP codes displayed in blue. The error bars capture the 95% confidence intervals. At $m = -1$, which is one year prior to funding cycle, the vertical line indicates the start of treatment, and the graph shows how treated ZIP codes display higher average Starbucks frequencies than the control ZIP codes. It is unclear what accounts for the trends pre-intervention, but it is likely that the store openings follow general macro-economic patterns and cyclical company behavior.





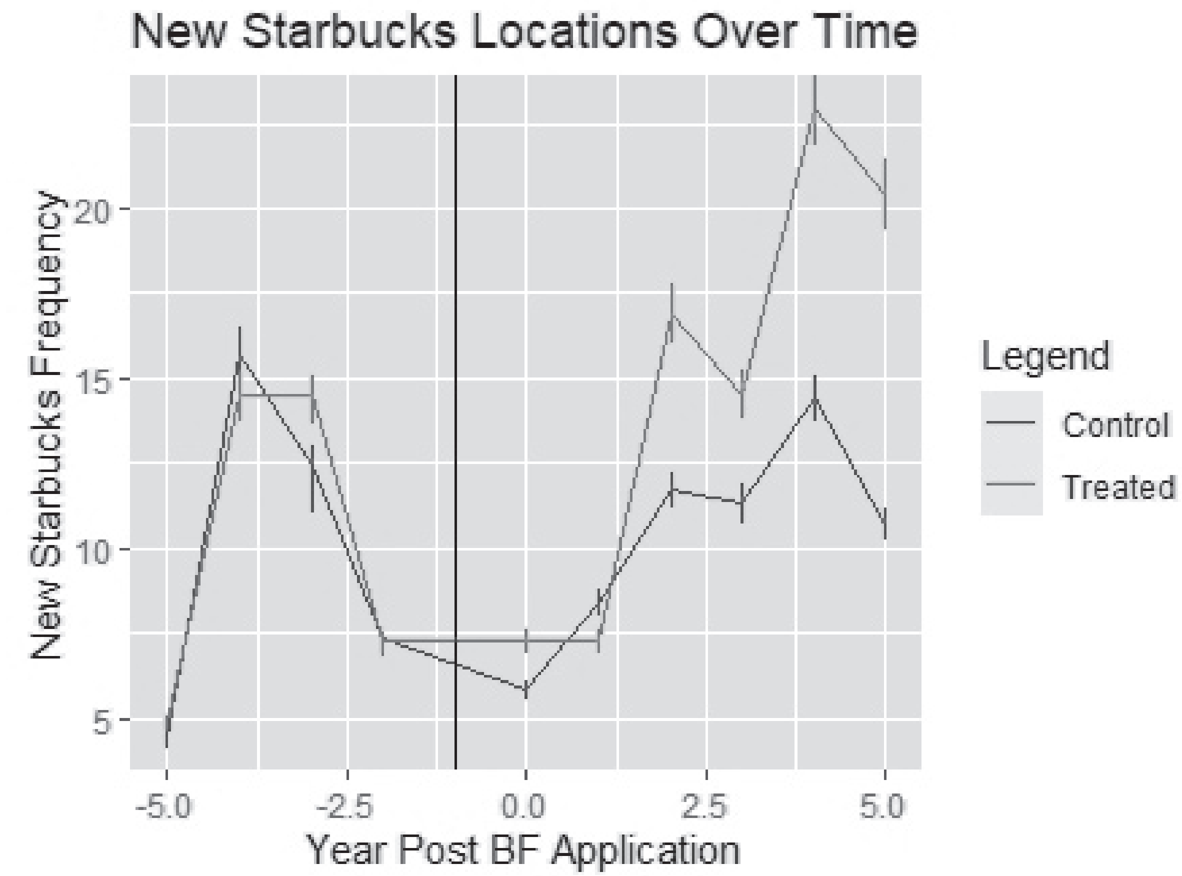



Figure 2: Event Study

### 4.3 *Economic Significance*

The causal effect of AWP grants ont the number of Starbucks locations is statistically significant, but it is equally important to consider whether or it is also economically significant. The upper bound of the 95% confidence interval is 0.06174288, which is still $\ll 1$. A 6% increase in the number of Starbucks over 5 years may seem small, but it is crucial to consider that the funded sustainability project typically concludes two years post funding, meaning the results typically represent a 6% increase over 3 years post-cleanup completion. Furthermore, with 40% of Starbucks locations being licensed (Luthor, Jason. 2019), there may be a further delay resulting from the time required for an individual to apply for a license and financially prepare for opening.

The economic significance of the results would be stronger if the statistical significance held over a longer post-funding period. However, with the earliest funding cycle being in 2010, the longest possible post-funding period is limited to 10 years, and includes only 25% of the treatment group. Given these considerations, and that the store growth measures only openings within a treatment area and does not account for adjacent ZIP Codes, the results of the regression are arguably also economically significant.

### *4.4 Robustness*

In order to check for robustness, we can re-run the regression with additional time varying controls and check if the results change. This is important because it is possible that the time effect is not necessarily constant throughout the five year post period. The regression with the additional control becomes:

$$y_{it} = \sum_{m \in M} d_{it}^m x_{it}' \beta_m + \gamma_i + \mu_t + \epsilon_{it}$$

where all terms in the equation are the same as the original regression, and the additional $x_{it}'$ term is a time-varying covariate. This new regression returns a slightly lower $\beta_m$ value of 0.0045865, but the standard error indicates that these results are still significant at the .1% confidence level. The results are displayed in Table 2 below:





| | **Coefficient** | **Clustered SE** | **Median Residual** | **Pr(>\|t\|)** | **Adjusted R-Squared** |
|---|---|---|---|---|---|
| **Number of Starbucks ~ (time-varying-control)*time post grant receival** | 0.0045865 | 0.0003301 | 0.00229 | <2e-16 *** | 0.7927 |

***Reject the null at the 0.0001 level of significance

Table 2: ZIP and Year Fixed Effects Linear Regression, with Time-Varying Controls

A robustness check with additional entity varying controls is unecessary, given insight into variance among ZIP codes is most-likely accounted for in the original regression, and further variability can be extracted from the standard errors, given that they are clustered at the ZIP code level.

## 5 Conclusions

### *5.1 Limitations*

The biggest limitations of the study are the small sample size of treated Brownfields and the accuracy of the ZIP Code data. Given that the last four digits of a US ZIP code may change as frequently as every four months, and the manual nature of treated ZIP code collection, it is possible that either (1) Starbucks that opened later in treated locations appeared under a different ZIP code than the Brownfields or (2) Starbucks opened in treated or control adjacent ZIP Codes were not included in the analysis.

*5.2 External Validity and Key Takeaways*

Measuring qualitative, contemporaneous results of gentrification is challenging, and analyses typically use census data that focuses on sociodemographic and economic changes that occur quite gradually. The effects of gentrification often reach beyond what is measurable, reshaping the cultural landscapes of communities to a point that original residents may feel like they no longer belong in their neighborhood even if they are financially able to stay (Maantay et al., 2018). Starbucks are a uniquely ubiquitous signal of housing price growth, which in turn exacerbates gentrification further down the road.

While the results of this study are externally valid for the green space subset of public sustainability projects, they cannot be firmly generalized to the broader sustainable development category, which includes other infrastructure such as eco-friendly aesthetics, renewable energy production, water-treatment plants, and waste-to-energy recycling among others. Furthering research regarding the social impacts of a broader range of non-cleanup projects would allow a better understanding of what the most "just" projects are, informing equitable funding.

Poorer and more minority neighborhoods often have less access to all of the aforementioned infrastructure in addition to open space, and are also less likely to have personal vehicles which can be used to access distant open spaces (Ferguson et al., 2018), making green spaces within disenfranchised communities even more important. Given that negative social implications of green spaces exist, it is imperative to enact policies that protect marginalized communities as part of the fight for environmental justice.

In the context of this study, implementing policy regulations in Brownfields communities may mitigate environmental gentrification. For example, limiting luxury development in Brownfield





areas or enforcing rent control in areas where green spaces have been constructed could help ensure that original residents reap the mental and physical health benefits of living sustainably.

Ultimately, in using random assignment of AWP grant status to Brownfields across four funding years, we identified that public greening projects significantly increased Starbucks frequency in an area, indicating gentrification of the treated neighborhood. In the five years immediately following project funding, the number of Starbucks changed on average by an estimated 0.059 additional stores.

The outcomes of greening public spaces are not unilaterally positive or negative, stressing the importance of factoring equity into the process of sustainability investing. In the quest to transform contaminated or otherwise unused urban land into beneficial green space, we must acknowledge and never lose sight of the fact that greening actions often pit the goals of environmental justice against the effects of environmental gentrification. This research ultimately illuminates the importance of fighting for equal access to environmental benefits and the means to achieve healthy lives, whilst also preventing vulnerable populations from bearing a disproportionate exposure to environmental burdens.

## 6 References

1. Althaus Ottman, M. M., Maantay, J. A., Grady, K., Cardoso, N., da Fonte, N. N. (2010). Community Gardens: An Exploration of Urban Agriculture in the Bronx, New York City. Cities and the environment, 3(1), 20. `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3160645/#__ffn_sectitle`

2. Autor, David H., Christopher J. Palmer, and Parag A. Pathak. 2019. "Ending Rent Control Reduced Crime in Cambridge." AEA Papers and Proceedings, 109 : 381-84. `https://www.`

nber.org/papers/w23914

3. Bedimo-Rung AL, Mowen AJ, Cohen DA. The significance of parks to physical activity and public health: a conceptual model. Am J Prev Med. 2005 Feb;28(2 Suppl 2):159-68. doi: 10.1016/j.amepre.2004.10.024. PMID: 15694524. https://pubmed.ncbi.nlm.nih.gov/15694524/

4. Feldman, Justin. ”Gentrification, urban displacement and affordable housing: Overview and research roundup”. August 15, 2014. https://journalistsresource.org/studies/economics/real-estate/gentrification-urban-displacement-affordable-housing-

5. M. Ferguson, H.E. Roberts, R.R.C. McEachan, M. Dallimer, Contrasting distributions of urban green infrastructure across social and ethno-racial groups, Landscape and Urban Planning, Volume 175, 2018,Pages 136-148, ISSN 0169-2046 https://www.sciencedirect.com/science/article/pii/S0169204618300951

6. Glaeser, Edward L., Hyunjin Kim, and Michael Luca. 2018. ”Nowcasting Gentrification: Using Yelp Data to Quantify Neighborhood Change.” AEA Papers and Proceedings, 108: 77-82. https://www.aeaweb.org/articles?id=10.1257/pandp.20181034

7. Luthor, Jason. *How to Go About Opening a Starbucks*. Chron, June 18, 2019. https://smallbusiness.chron.com/opening-starbucks-32199.html

8. Maantay, J. A., Maroko, A. R. (2018). Brownfields to Greenfields: Environmental Justice Versus Environmental Gentrification. International journal of environmental research and public health, 15(10), 2233. https://doi.org/10.3390/ijerph15102233

9. *Overview of EPA's Brownfields Program*

https://www.epa.gov/brownfields/overview-epas-brownfields-program

10. Rettner, Rachel. *More than 250,000 People May Die Each Year Due to Climate Change*. Life





Science, January 17, 2019.

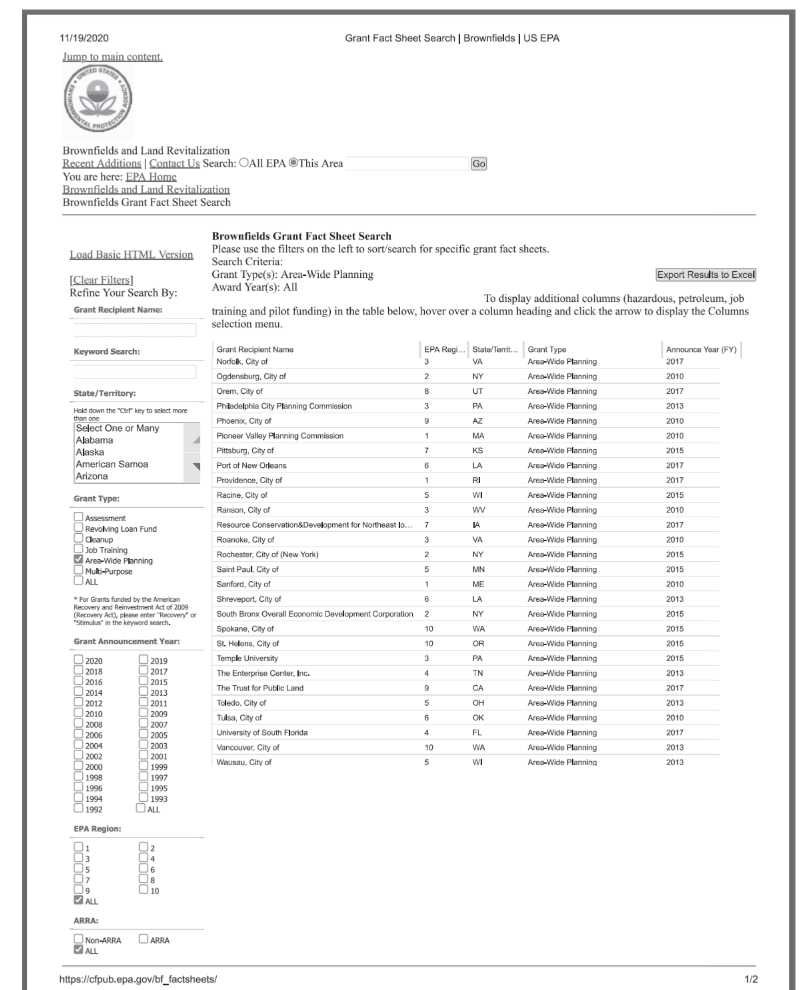
https://www.livescience.com/64535-climate-change-health-deaths.html

11. World Health Organization. *Climate Change and Health*. February 1, 2018. https://www.who.int/news-room/fact-sheets/detail/climate-change-and-health

## 7 Appendix

### *7.1 Brownfields Grant Fact Sheet Search*

All filters were cleared other than ”Grant Type”, which was set to ”Area-Wide Planning”, as shown in the following:



11/19/2020 Grant Fact Sheet Search | Brownfields | US EPA

Jump to main content.

Brownfields and Land Revitalization
Recent Additions | Contact Us Search: All EPA This Area Go
You are here: EPA Home
Brownfields and Land Revitalization
Brownfields Grant Fact Sheet Search

Load Basic HTML Version

[Clear Filters]
Refine Your Search By:

Grant Recipient Name:

Keyword Search:

State/Territory:

Hold down the "Ctrl" key to select more than one
Select One or Many
Alabama
Alaska
American Samoa
Arizona

Grant Type:

Assessment
Revolving Loan Fund
Cleanup
Job Training
Area-Wide Planning
Multi-Purpose
ALL

* For Grants funded by the American Recovery and Reinvestment Act of 2009 (Recovery Act), please enter "Recovery" or "Stimulus" in the keyword search.

Grant Announcement Year:

2020 2019 2018 2017 2016 2015 2014 2013 2012 2011 2010 2009 2008 2007 2006 2005 2004 2003 2002 2001 2000 1999 1998 1997 1996 1995 1994 1993 1992 ALL

EPA Region:

1 2 3 4 5 6 7 8 9 10 ALL

ARRA:

Non-ARRA ARRA ALL

**Brownfields Grant Fact Sheet Search**
Please use the filters on the left to sort/search for specific grant fact sheets.
Search Criteria:
Grant Type(s): Area-Wide Planning
Award Year(s): All

Export Results to Excel

To display additional columns (hazardous, petroleum, job training and pilot funding) in the table below, hover over a column heading and click the arrow to display the Columns selection menu.

| Grant Recipient Name | EPA Regi... | State/Territ... | Grant Type | Announce Year (FY) |
|---|---|---|---|---|
| Norfolk, City of | 3 | VA | Area-Wide Planning | 2017 |
| Ogdensburg, City of | 2 | NY | Area-Wide Planning | 2010 |
| Orem, City of | 8 | UT | Area-Wide Planning | 2017 |
| Philadelphia City Planning Commission | 3 | PA | Area-Wide Planning | 2013 |
| Phoenix, City of | 9 | AZ | Area-Wide Planning | 2010 |
| Pioneer Valley Planning Commission | 1 | MA | Area-Wide Planning | 2010 |
| Pittsburg, City of | 7 | KS | Area-Wide Planning | 2015 |
| Port of New Orleans | 6 | LA | Area-Wide Planning | 2017 |
| Providence, City of | 1 | RI | Area-Wide Planning | 2017 |
| Racine, City of | 5 | WI | Area-Wide Planning | 2015 |
| Ranson, City of | 3 | WV | Area-Wide Planning | 2010 |
| Resource Conservation&Development for Northeast Io... | 7 | IA | Area-Wide Planning | 2017 |
| Roanoke, City of | 3 | VA | Area-Wide Planning | 2010 |
| Rochester, City of (New York) | 2 | NY | Area-Wide Planning | 2015 |
| Saint Paul, City of | 5 | MN | Area-Wide Planning | 2015 |
| Sanford, City of | 1 | ME | Area-Wide Planning | 2010 |
| Shreveport, City of | 6 | LA | Area-Wide Planning | 2013 |
| South Bronx Overall Economic Development Corporation | 2 | NY | Area-Wide Planning | 2015 |
| Spokane, City of | 10 | WA | Area-Wide Planning | 2015 |
| St. Helens, City of | 10 | OR | Area-Wide Planning | 2015 |
| Temple University | 3 | PA | Area-Wide Planning | 2015 |
| The Enterprise Center, Inc. | 4 | TN | Area-Wide Planning | 2013 |
| The Trust for Public Land | 9 | CA | Area-Wide Planning | 2017 |
| Toledo, City of | 5 | OH | Area-Wide Planning | 2013 |
| Tulsa, City of | 6 | OK | Area-Wide Planning | 2010 |
| University of South Florida | 4 | FL | Area-Wide Planning | 2017 |
| Vancouver, City of | 10 | WA | Area-Wide Planning | 2013 |
| Wausau, City of | 5 | WI | Area-Wide Planning | 2013 |

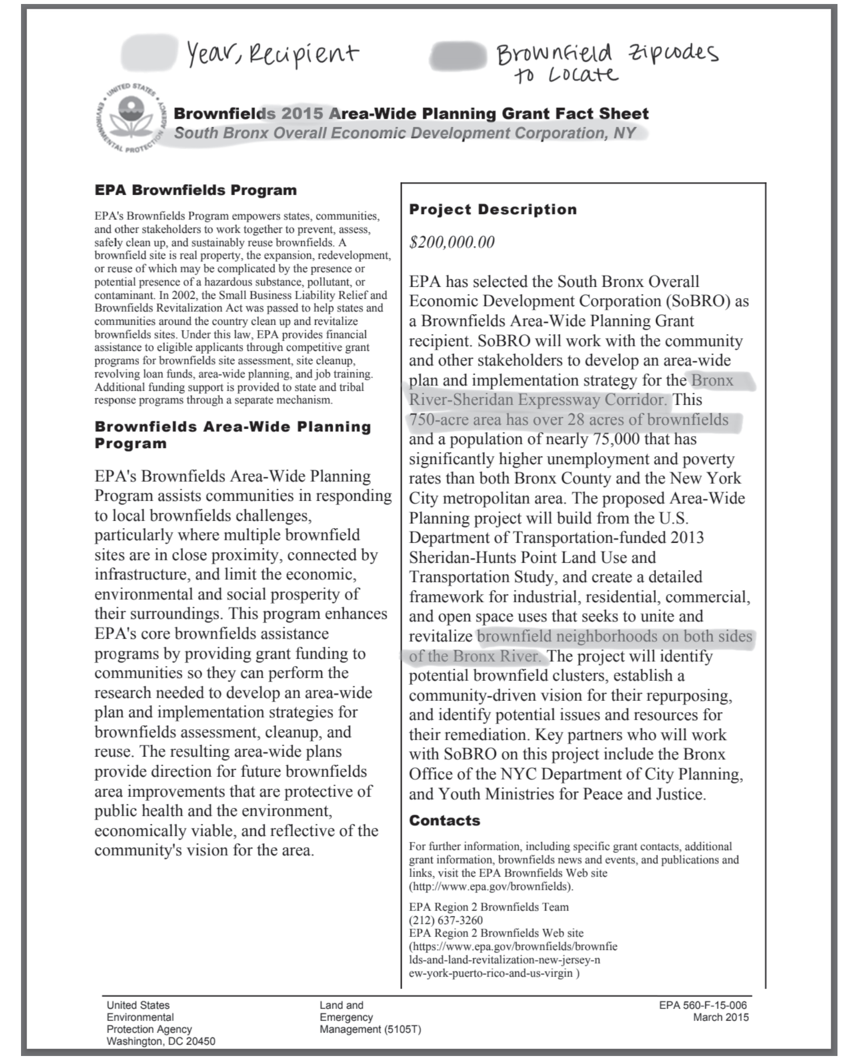
https://cfpub.epa.gov/bf_factsheets/ 1/2

### 7.2 Treated Zipcode Extraction

For each recipient, ZIP Codes were matched to the areas outlined in the project description. The following is an example: Highlighted in yellow is the grant funding year, and the recipient of the AWP award. Highlighted in pink are geographical regions in which the applicant indends to implement public infrastructure projects. Given this data, a google search of ”Year”, ”Applicant”, ”Geographical Region” was conducted in order to derive the more detailed community plans that the applicants presented at the state level. These plans included specfic addresses and/or maps that contained ZIP code level information on project areas. In this manner, all of the 595 affected zipcodes for the 83 recipients were formatted into Panel Data.






Year, Recipient
Brownfield Zipcodes to Locate

**Brownfields 2015 Area-Wide Planning Grant Fact Sheet**
*South Bronx Overall Economic Development Corporation, NY*

**EPA Brownfields Program**

EPA's Brownfields Program empowers states, communities, and other stakeholders to work together to prevent, assess, safely clean up, and sustainably reuse brownfields. A brownfield site is real property, the expansion, redevelopment, or reuse of which may be complicated by the presence or potential presence of a hazardous substance, pollutant, or contaminant. In 2002, the Small Business Liability Relief and Brownfields Revitalization Act was passed to help states and communities around the country clean up and revitalize brownfields sites. Under this law, EPA provides financial assistance to eligible applicants through competitive grant programs for brownfields site assessment, site cleanup, revolving loan funds, area-wide planning, and job training. Additional funding support is provided to state and tribal response programs through a separate mechanism.

**Brownfields Area-Wide Planning Program**

EPA's Brownfields Area-Wide Planning Program assists communities in responding to local brownfields challenges, particularly where multiple brownfield sites are in close proximity, connected by infrastructure, and limit the economic, environmental and social prosperity of their surroundings. This program enhances EPA's core brownfields assistance programs by providing grant funding to communities so they can perform the research needed to develop an area-wide plan and implementation strategies for brownfields assessment, cleanup, and reuse. The resulting area-wide plans provide direction for future brownfields area improvements that are protective of public health and the environment, economically viable, and reflective of the community's vision for the area.

**Project Description**

*$200,000.00*

EPA has selected the South Bronx Overall Economic Development Corporation (SoBRO) as a Brownfields Area-Wide Planning Grant recipient. SoBRO will work with the community and other stakeholders to develop an area-wide plan and implementation strategy for the Bronx River-Sheridan Expressway Corridor. This 750-acre area has over 28 acres of brownfields and a population of nearly 75,000 that has significantly higher unemployment and poverty rates than both Bronx County and the New York City metropolitan area. The proposed Area-Wide Planning project will build from the U.S. Department of Transportation-funded 2013 Sheridan-Hunts Point Land Use and Transportation Study, and create a detailed framework for industrial, residential, commercial, and open space uses that seeks to unite and revitalize brownfield neighborhoods on both sides of the Bronx River. The project will identify potential brownfield clusters, establish a community-driven vision for their repurposing, and identify potential issues and resources for their remediation. Key partners who will work with SoBRO on this project include the Bronx Office of the NYC Department of City Planning, and Youth Ministries for Peace and Justice.

**Contacts**

For further information, including specific grant contacts, additional grant information, brownfields news and events, and publications and links, visit the EPA Brownfields Web site (http://www.epa.gov/brownfields).

EPA Region 2 Brownfields Team
(212) 637-3260
EPA Region 2 Brownfields Web site
(https://www.epa.gov/brownfields/brownfie
lds-and-land-revitalization-new-jersey-n
ew-york-puerto-rico-and-us-virgin )

United States Environmental Protection Agency Washington, DC 20450
Land and Emergency Management (5105T)
EPA 560-F-15-006
March 2015

### 7.3 FOIA Request

Andrew R. Wheeler
Environmental Protection Agency
1200 Pennsylvania Avenue, N.W.
Washington, DC 20460

Re: Freedom of Information Act Request

Dear Mr. Wheeler,

This is a request under the Freedom of Information Act. I hereby request copies of the following records [or all records containing the following information]:

> Documentation of all EPA Brownfields locations that **applied** for Area-Wide Planning (AWP) Grants in each funding year between 2010 and 2017 (FY 2010, FY 2013, FY 2015, FY 2017).

As the FOIA requires, please release all reasonably segregable nonexempt portions of documents.

In order to help to determine my status to assess fees, you should know that I am affiliated with the Massachusetts Institute of Technology, and this request is made for a scholarly and scientific purpose and not for commercial use.

I request a waiver of all fees for this request. Disclosure of the requested information to me is in the public interest because it is likely to contribute significantly to public understanding of the operations or activities of the government and is not primarily in my commercial interest. I am conducting a research study examining the efficacy of receiving Brownfields grants and in order to conduct said economic analysis, I need to be able to compare communities that received funding with those that applied for, but did not receive funding.

I am willing to pay fees for this request up to a maximum of $25. If you estimate that the fees will exceed this limit, please inform me before processing my request.

If you have any questions regarding this request, please contact me at (425) 802-6945 and/or rikita@mit.edu. I look forward to receiving your response within the twenty day statutory time period. Thank you for your consideration of this request.

Sincerely,
Rikita Bansal
rikita@mit.edu
497 Commonwealth Avenue,
Unit 3
Boston, MA 02215





# The Effect of Anti-Chinese Sentiment on Mental Health Outcomes of Asians in the United States

Catherine Huang*

December 2020

**Abstract**

I use an instrumental variables approach to estimate the effect of an increase in anti-Chinese sentiment on Asian mental health outcomes. Using a measure of interest in trade as an instrument for anti-Chinese sentiment, I find no significant effect of an increase in anti-Chinese sentiment on poor mental health days experienced by Asians in the U.S. As measured by my index, the average increase in anti-Chinese sentiment in U.S. states from 2013 to 2018 was 4.9 points out of 100 possible points. My analysis shows that a 4.9 point increase in anti-Chinese sentiment increased poor mental health days experienced by Asians by at most 0.5 days, a 20% increase from the mean number of poor mental health days.

## Introduction

Asians are the fastest growing major racial or ethnic group in the U.S. Between 2000 and 2015, the total U.S. Asian population grew 72% from 11.9 million to 20.4 million.[1] Yet mental health outcomes for Asians in the U.S. are still poorly understood. In addition, recent data suggest a confluence of negative sentiment towards China due to U.S.-China tensions and an increase in anti-Asian sentiment directed towards Asian Americans. The percentage of Americans who have an unfavorable view of China rose from 66% in March 2020 to 73% in July 2020.[2] Across the same time period, the COVID-19 pandemic triggered a wave of 2,583 hate incidents against Asian Americans between March 19th, 2020 and August 5th, 2020. These hate incidents included instances of virulent animosity, anti-immigrant nativism, racist characterizations of Chinese, and racial slur usage.[3] It is more important than ever to understand the relationship between racial discrimination and the mental health of Asian Americans. Accordingly, this paper investigates the effects of anti-Chinese sentiment on the mental health of Asian Americans across the 2013-2018 five year period.

Interest in Asian-American mental health outcomes has increased with the spotlight brought by COVID-19, though there were a few studies in the early 2000's. Vachuska (2020) shows that COVID-19 has

*Code and data available at https://github.com/chuang1326/1433finalproject.
[1] López et al. (2017)
[2] Silver et al. (2020)
[3] Jeung et al. (2020)

caused increases in both anti-Chinese and anti-Hispanic sentiment. Separately, Ahrens (2020) reviews evidence from the psychology literature that racial discrimination against Asian Americans is correlated with increased depression and anxiety in Asian American populations. In work published before 2020, Gee et al. (2007) shows that self-reported racial discrimination is associated with increased rates of mental health disorders; the authors find these results even when controlling for a variety of sociodemographic characteristics and indicators of wealth. Chatterji et al. (2007) shows evidence that Asians' labor market outcomes are affected by mental health disorders, but the results are not significant. I find it interesting that there are relatively few papers published on Asian American mental health between 2007 and 2020, when research groups like the U.S. Center for Disease Control (CDC) were collecting better data on Asian American populations and Asian populations were growing further.[4] This paper's contributes to the literature by contributing knowledge about the missing period between 2013 and 2018.

In this paper, I estimate the effect of an increase in anti-Chinese sentiment on mental health outcomes for Asian Americans. My analysis exploits variation in Google searches for an anti-Chinese slur, Google searches for an anti-Chinese expletive, and poor mental health days experienced per month by Asian adults from 2013 to 2018. I instrument for increases in searches for the slur or expletive using both trade balance and searches for the search word "trade." (The latter is referred to in this paper as the trade interest index.) I also control for unemployment, median income, state fixed effects, and year fixed effects to assuage concerns that trade balance is only exogenous conditional on these omitted variables.

My analysis does not find a significant effect of anti-Chinese sentiment on mental distress in Asian Americans. In a variety of specifications using different instruments and measures of anti-Chinese sentiment, I find that the 95% confidence intervals include 0. The effect of a 4.9 point increase in the slur index is bounded by -0.3 days on the low end and +0.5 days on the high end. This result is consistent when an additional control, population share of Asians, is added to the analysis.

These results suggest that anti-Chinese sentiment related to economic tensions does not translate strongly into discriminatory acts in absentia of exogenous shocks like the COVID-19 pandemic.

The paper is organized as follows. Section I describes the trade balance, sentiment, and mental health data used in this paper. Section II describes the empirical instrumental variables strategy and argues for the validity of the trade interest index as an instrument. Section III provides an overview of the results. Section IV concludes and makes recommendations for future research on this topic.

# 1 Background and Data

## 1.1 Google Trends Sentiment Data: Slur and Swear Indices

This paper uses Google Trends data on searches for an anti-Chinese slur as one measure of anti-Chinese sentiment. The state-year level panel dataset covers most U.S. states from the year 2013 to the year 2018. Data are initially reported by Google Trends as values between 0 and 100. These values

[4]Starting in 2011, the CDC's National Health and Nutrition Examination Survey began over-sampling non-Hispanic Asians in a primary sample design change. NHANES (2016)





are proportional to searches for the slur divided by total searches for any word in the state and year. Since Google scales the data within years so that one state in each year has an index of 100, it is necessary to rescale the data such that each state follows its time series trend across years. I perform this normalization using California as a benchmark state; I create a dataset that ranges from 0 to 115. I rescale this data once more so the values fall between 0 and 100 for ease of interpretation.

I test whether the Google Trends slur index is an effective measure of anti-Chinese sentiment by checking its correlation with reputable survey data on attitudes towards Asians. The American National Election Studies (ANES) surveys solicit "temperature" ratings of different racial and ethnic groups. In particular, survey respondents are asked to rate Asians on a scale of 0 to 100, with 100 being the highest level of approval and 0 being the lowest level of approval. I find that in a regression of the 2016 and 2018 ANES ratings on the slur index, the coefficient on the slur index is positive and significant at the 5% level (see Table 6 in the Appendix). These results are opposite what one might expect, because the higher the measure of anti-Chinese sentiment using the slur index, the lower the measure of anti-Chinese sentiment using the ANES temperate ratings. These results persist when I control for Asian population share, which one might believe to be an omitted variable; Table 6 shows that greater slur usage is with correlated with a denser Asian population, but the inclusion of this population share variable does not generate a negative relationship between the slur index and ANES temperature ratings. I am not sure why there is an inverted relationship between the slur index and ANES ratings. It is possible that the positive relationship exists between 2016 and 2018, but not in a larger dataset with more than two years. Further research can establish whether the positive correlation represents a major flaw in the slur index data, but I will not do that analysis here because it would involve additional data downloading and cleaning that is outside the scope of the project.

For robustness, I generate a second measure of anti-Chinese sentiment, which will be called the swear index. The swear index is created from Google Trends data on searches for the phrase "f*** China." This dataset is less complete due to the relative rarity of the phrase; Google Trends removes datapoints if there are few enough results in the state and time period to endanger the privacy of Google's users. As a result, 24% of the datapoints available for the slur index are missing in the dataset for the swear index. I replace these missing datapoints with 0 values since datapoints are only removed if the search volume is very small.

To maximize the power of my regressions, I use the more complete slur index as my endogenous variable of interest. I use the swear index as an alternate variable of interest for robustness checks.

## 1.2 KFF Mental Health and Population Share Data

I use the Kaiser Family Foundation (KFF) reports on the average number of poor mental health days per month. The data is availabe by race, state, and year. The original data comes from the CDC's Behavioral Risk Factor Surveillance System, which uses telephone surveys to interview 400,000 American adults each year. The KFF aggregated data are available from 2013 to 2018 for all states except for Arkansas, Mississippi, Montana, West Virginia, and Wyoming, since these states have insufficient data.

I also use KFF data on the fraction of the population that is Asian (hereafter "population share") in regression robustness analysis. The original population share data come from the U.S. Census Bureau's

American Community Survey and are restricted to the civilian, non-institutionalized popoulation. The data are available for all states and the years from 2013 to 2018.

### 1.3 Google Trends Trade Interest Index and Trade Balance Data

This paper uses Google Trends data on searches for the word "trade" as an instrument for instrumental variables analysis. Similarly to the slur index dataset, the trade interest dataset covers most states at the state-year level from 2013 to 2018. The data are first processed to be consistent with time series trends, then rescaled so the final panel dataset ranges from 0 to 100.

I use trade balance as a secondary instrument. The trade balance data come from the U.S. Census Bureau website, where the dataset is downloadable as exports from U.S. states to China and imports to U.S. states from China. I calculate a simple measure of trade balance in billions of dollars by subtracting imports from exports. This dataset is available for every state and every year from 2013 to 2018.

Table 1 provides several statistics about the data used in this paper.

Table 1: Summary of Data

| | Mean | Std. Dev | Minimum | Maximum |
|---|---|---|---|---|
| Slur Index | 41.0 | 13.9 | 16.6 | 100 |
| Trade Index | 66.7 | 9.3 | 38.8 | 100 |
| Poor Mental Health Days/Month | 2.6 | 1.0 | .3 | 8.5 |
| Swear Index | 29.3 | 13.7 | 7.2 | 100 |
| Trade Balance (Billions of 2018 USD) | -7.5 | 19.8 | -144.8 | 12.3 |
| Asian Population Share | .04 | .06 | .007 | .4 |

Note: Index variables come from Google Trends. Mental health data and population share data come from the Kaiser Family Foundation. Trade balance data come from the U.S. Census Bureau website.

## 2 Empirical strategy

My empirical strategy is to instrument for anti-Asian sentiment using the trade interest index. I estimate the reduced form regression and first stage regression, respectively:

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \beta_2 U_{it} + \beta_3 I_{it} + \sum_{i=1}^{n} \alpha_{Yi} S_i + \sum_{t=2013}^{2018} \gamma_{Yt} Y_t + \epsilon_{it}$$

$$X_{it} = \beta_4 + \beta_5 Z_{it} + \beta_6 U_{it} + \beta_7 I_{it} + \sum_{i=1}^{n} \alpha_{Xi} S_i + \sum_{t=2013}^{2018} \gamma_{Xt} Y_t + u_{it}$$

$Y_{it}$ is the number of poor mental health days per month for Asians in state $i$ in year $t$, $X_{it}$ is an index of Google searches for an anti-Chinese slur in state $i$ in year $t$, $U_{it}$ is unemployment in state $i$ in year $t$, and $I_{it}$ is median income in state $i$ in year $t$. $S_i$ and $Y_t$ are indicators for state $i$ and year $t$, respectively.





$Z_{it}$ is the trade interest index, an instrument for $X_{it}$ which is proportional to searches for the word "trade" in state $i$ and year $t$. $\epsilon_{it}$ and $u_{it}$ represent error terms.

The trade interest index is an effective instrument for anti-Chinese sentiment because it is correlated with searches for an anti-Chinese slur. In addition, conditional on controls, trade interest is arguably correlated with few other inputs to Asian American mental health. Figure 1 shows that there is a nontrivial positive correlation between the trade interest index and the slur index after partialing out the unemployment rate, median income, state fixed effects, and year fixed effects.[5] Table 2 confirms that the correlation between the trade interest index and slur index is significant at the 5% level after controlling for unemployment, income, state, and year. This correlation satisfies the relevance condition of instrumental variables analysis. Table 2 also shows the validity of the concern that median income could be an omitted variable. Column 6 of the table shows median income is predicted by the trade interest index at the 1% level. Though Column 5 shows I cannot reject that trade interest does not predict unemployment, I will include unemployment in my regression since there is evidence that a trade deficit causes unemployment (Alawin (2013)).

Trade balance is a relatively ineffective instrument compared to the trade interest index. Table 2 shows that trade balance predicts neither the slur index nor the swear index. Since the instrumental variables strategy requires a strong correlation in the first stage, I use trade balance as an instrument only in robustness checks of my analysis.

An item of note in Table 2 is that the trade interest index is very closely correlated with the swear index after controlling for unemployment, income, state, and year. One might suggest that the combination of the trade interest index and swear index are a superior, more powerful first stage for the instrumental variables regression. However, I will continue to use the slur index as my endogenous variable of interest because the slur index dataset is more complete than the swear index dataset.

Table 2: Trade Variables as Predictors of Slur, Swear Indices

| | (1) Slur Index | (2) Swear Index | (3) Slur Index | (4) Swear Index | (5) Unemp. | (6) Med. Income | (7) Pop. Share |
|---|---|---|---|---|---|---|---|
| Trade Interest Index | 0.406** | -1.033*** | | | -0.004 | 405.657*** | 0.001*** |
| | (0.18) | (0.19) | | | (0.01) | (58.48) | (0.00) |
| Trade Balance | | | -0.321 | -0.244 | | | |
| | | | (0.31) | (0.34) | | | |
| N | 282 | 282 | 282 | 282 | 282 | 282 | 282 |
| R-squared | 0.62 | 0.72 | 0.61 | 0.69 | 0.00 | 0.15 | 0.03 |

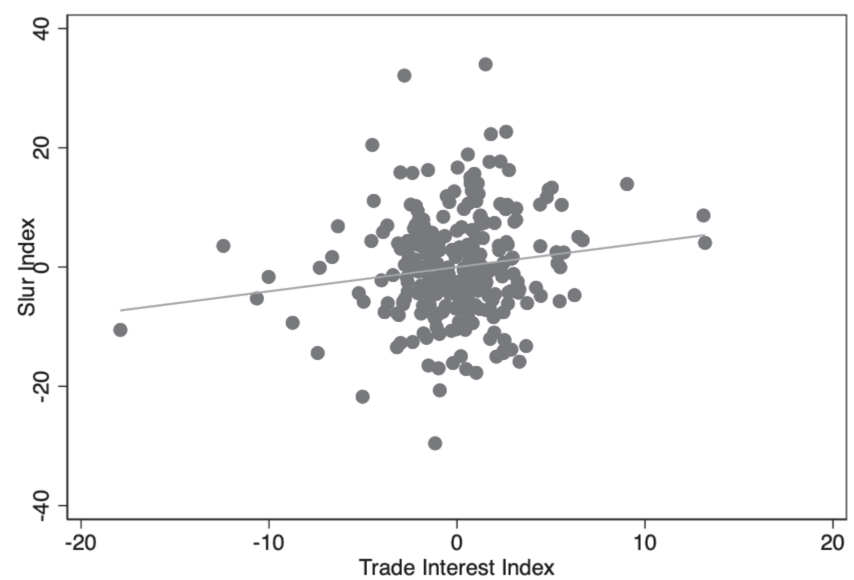
Notes: Standard errors in parentheses below each estimate. Significance at the 1, 5, and 10 percent levels indicated by ***, **, and *, respectively. In the first four columns, unemployment, median income, state fixed effects, and year fixed effects have been partialed out.

I control for unemployment, income, state fixed effects, and year fixed effects because I believe these are the major candidates for variables that might violate the exclusion restriction, but it is possible that ICE deportation rates and measures of xenophobia are additional omitted variables. For example, it is known that ICE deportation rates differ by state.[6] It is possible that deportation rates are correlated

[5] I control for these variables to avoid bias in my estimates: state and year fixed effects control for differences in baselines between states and years, while controls for unemployment rate and median income ward off omitted variables bias. In particular, I control for unemployment rate because general unemployment may affect mental health outcomes for specific ethnicities in addition to being correlated with trade interest. Similarly, I control for median income because income may be correlated with both trade concerns and mental health.

[6] TRAC (2019).

Figure 1: Positive Correlation Between Slur Index and Trade Interest Index




Note: unemployment, income, state fixed effects, and year fixed effects have been partialed out both variables.

with trade interest and mental health outcomes outside of the anti-Asian sentiment channel: states and year combinations with greater deportation rates might have greater trade deficits, and greater deportation rates might cause heightened mental distress. Similarly, heightened levels of xenophobia in certain states and years might be correlated with greater trade deficits and poor mental health in a set of causal channels that are separate from those involving sinophobia in particular. However, as measures of xenophobia and deportation rates for Asians are difficult to obtain, analysis of these possible omitted variables is outside the scope of this project.

If the trade interest index is correlated with the slur index, but uncorrelated (conditional on the controls) with other variables affecting poor mental health days for Asian-Americans, then the remaining variation in poor mental health days due to trade interest can be attributed to anti-Chinese sentiment. Since the trade interest index is not obviously correlated with other determinants of Asian-American mental health outcomes and I have controlled for the omitted variables for which I have data, I will make the assumption that the exogeneity condition holds for instrumental variables analysis.

## 3 Results

### 3.1 The Effect of Anti-Asian Sentiment on Asian American Mental Health Outcomes

I find no significant effect of an increase in anti-Chinese sentiment on the number of poor mental health days experienced by Asian Americans. Column 1 of Table 3 shows the results of the main regression. The dependent variable is the number of poor mental health days experienced each month by Asian respondents. The endogenous variable of interest is the slur index, a measure of Google searches for





an anti-Chinese slur; the index ranges between 0 and 100. Note that all columns in Table 3 control for state fixed effects and year fixed effects in addition to the unemployment rate, median income, and Asian population share. I find that an increase in slur usage does not predict a signficant change in mental distress when significance is measured at the 5% level: the 95% confidence interval of the coefficient on the slur index is equal to [-.0589495, .1014993] and contains 0.

Table 3: Impact of Sentiment on Poor Mental Health Days

| Instrument: | Trade Interest Index | | | |
|---|---|---|---|---|
| | (1) Days | (2) Days | (3) Days | (4) Days |
| Slur Index | 0.021 | | 0.024 | |
| | (0.04) | | (0.04) | |
| Swear Index | | -0.008 | | -0.008 |
| | | (0.01) | | (0.01) |
| Unemployment | -0.195* | -0.184* | -0.197* | -0.185* |
| | (0.11) | (0.11) | (0.11) | (0.11) |
| Med. Income | 0.000 | 0.000 | 0.000 | 0.000 |
| | (0.00) | (0.00) | (0.00) | (0.00) |
| Pop. Share | | | -67.956** | -51.335* |
| | | | (34.29) | (30.84) |
| N | 264 | 264 | 264 | 264 |
| R-squared | 0.41 | 0.43 | 0.42 | 0.43 |

Notes: Standard errors in parentheses below each estimate. Significance at the 1, 5, and 10 percent levels indicated by ***, **, and *, respectively. All columns control for state and year fixed effects in addition to displayed variables.

I find that a variation of the measure of anti-Chinese sentiment does not suggest a signficant effect of sentiment on mental health outcomes. Column 2 of Table 3 shows the results of an alternate regression in which the endogenous variable of interest is changed to the swear index, a measure of Google searches for the phrase "f*** China." There is no significant effect of a increase in the swear index on mental distress.

The results in Columns 1 and 2 of Table 3 are robust under the addition of a potential omitted variable, population share. Column 7 of Table 2 shows that the Asian population share in each state is predicted by the trade interest index at the 1% level. One might argue that population share may also be correlated with mental health outcomes because areas with concentrated Asian populations may have better or worse mental health than less concentrated areas. In the case that population share is correlated with both trade interest and poor mental health days, population share would be an omitted variable in our instrumental variables analysis. I address this concern by running the regressions in Columns 3 and 4 of Table 3, which show that none of the coefficients on the slur and swear indices become significant with the addition of population share to the model of the main regression. The 95% confidence interval on the coefficient of the slur index in Column 3 of Table 3 does not change compared to Column 1 of Table 3; the new confidence interval is [-.0572654, .1054896], which is very similar to the previous confidence interval of [-.0589495, .1014993]. The main conclusions in Columns 1 and 2 are robust under the addition of population share as an omitted variable.

Using the main regression in Column 1 of Table 3, I estimate that in the average state, changes in anti-Chinese sentiment between 2013 and 2018 increased poor mental health days by up to 0.5 days per month. A 0.5 day increase in poor mental health days constitutes a notable 20% increase in mental distress given that the mean number of poor mental health days is 2.6 days (see Table 1). To see this, I calculate for each state the difference between the slur index in 2018 and the slur index in 2013. I report these differences in Table 4. The table shows that on average, the slur index increased in states by 4.9 points from 2013 to 2018. By multiplying the endpoints of the 95% confidence interval of Table 3, Column 1 by 4.9, I obtain an estimate that in the average U.S. state, anti-Chinese sentiment increased by 4.9 points between 2013 and 2018 and affected poor mental health days among Asians by between -0.3 days and +0.5 days. Since the mean number of poor mental health days is 2.6 days, a +0.5 day shift in mental health outcomes constitutes a moderate effect of a 20% increase in mental distress. Table 4 shows the slur index decreased the most in Delaware, where it decreased by 36 points; the index increased the most in Indiana, where it increased by 24 points. Then the confidence interval for the effect of the 2013-2018 change in anti-Chinese sentiment in Delaware is [-3.6, 2.1], while the confidence interval for the effect of the 2013-2018 change in sentiment in Indiana is [-1.4, 2.5]. These results suggest that changes in anti-Chinese sentiment in the 2013-2018 period could have increased poor mental health days by up to 2.5 days per month, or decreased poor mental health days by up to 3.6 days per month, depending on the state. Given that poor mental health days ranged from 0.3 days per month to 8.5 days per month, with a mean of 2.6 days, the results suggest that the 5-year change in anti-Chinese sentiment might have had quite a large effect on mental health outcomes.

Table 4: Change in Slur Index from 2013 to 2018

| State | Change in Index | State | Change in Index |
|---|---|---|---|
| AVERAGE | 4.87 | Kentucky | 7.94 |
| Delaware | -35.83 | Washington | 8.63 |
| Arkansas | -33.39 | Tennessee | 8.69 |
| South Carolina | -19.28 | Maryland | 9.41 |
| West Virginia | -18.55 | Wisconsin | 11.14 |
| South Dakota | -17.79 | Mississippi | 11.17 |
| Nevada | -11.93 | Texas | 11.20 |
| New Mexico | -10.37 | Connecticut | 11.97 |
| Idaho | -9.38 | Arizona | 12.75 |
| Alaska | -7.86 | Georgia | 12.83 |
| Montana | -6.44 | Pennsylvania | 13.51 |
| North Carolina | -4.50 | Ohio | 13.58 |
| Louisiana | -4.39 | Hawaii | 14.45 |
| Oklahoma | -2.89 | Rhode Island | 15.24 |
| Missouri | -2.83 | Utah | 16.07 |
| Kansas | -1.29 | Virginia | 16.87 |
| Vermont | -.82 | Massachusetts | 17.64 |
| Michigan | .44 | Wyoming | 18.37 |
| North Dakota | 2.03 | District of Columbia | 18.50 |
| Colorado | 2.05 | Illinois | 18.51 |
| Alabama | 2.98 | Maine | 18.51 |
| Nebraska | 5.26 | Oregon | 19.38 |
| New Jersey | 5.29 | California | 20.11 |
| Florida | 5.44 | New York | 22.59 |
| New Hampshire | 7.75 | Minnesota | 23.44 |
| Iowa | 7.94 | Indiana | 24.38 |

Note: Slur index ranges from 0 to 100.
Change is defined as slur index in the state in 2018 minus slur index in the state in 2013.





### 3.2 Robustness

For robustness, I repeat the previous analysis using trade balance as the instrument instead of the trade interest index. As a reminder, trade balance is a weaker instrument than the trade interest index because trade balance is not strongly correlated with the slur and swear indices. In Columns 1 and 2 of Table 5, I find that the coefficients on the slur and swear indices are not significant. In Columns 3 and 4, I find that the null results of Columns 1 and 2 do not change under the addition of Asian population share as a control. (As above, I control for population share because Column 7 of Table 2 gives evidence that population share may violate the exclusion restriction.) Finally, the confidence intervals in Table 5 are about four times larger than in Table 3; this is consistent with the fact that the trade interest index is a stronger instrument than trade balance in this setting. To conclude, I find use of trade balance as an instrument neither refutes nor improves upon the conclusions of the Results section.

Table 5: Impact of Sentiment on Poor Mental Health Days

| Instrument: | Trade Balance | | | |
|---|---|---|---|---|
| | (1) Days | (2) Days | (3) Days | (4) Days |
| Slur Index | -0.141 | | -0.142 | |
| | (0.14) | | (0.16) | |
| Swear Index | | -0.162 | | -0.212 |
| | | (0.20) | | (0.39) |
| Unemployment | -0.275 | 0.247 | -0.276 | 0.392 |
| | (0.20) | (0.61) | (0.20) | (1.15) |
| Med. Income | 0.000 | -0.000 | 0.000 | -0.000 |
| | (0.00) | (0.00) | (0.00) | (0.00) |
| Pop. Share | | | 3.140 | 100.284 |
| | | | (87.77) | (304.58) |
| N | 264 | 264 | 264 | 264 |
| R-squared | . | . | . | . |

Notes: Standard errors in parentheses below each estimate. Significance at the 1, 5, and 10 percent levels indicated by ***, **, and *, respectively. All columns control for state and year fixed effects in addition to displayed variables.

## 4 Conclusion

This paper quantifies the effect of anti-Chinese sentiment on Asian mental health outcomes in the U.S. Methodologically, I control for confounding variables in the sentiment-mental health relationship by utilizing trade interest as an instrument for anti-Chinese sentiment. I find that anti-Chinese sentiment may have had a moderate effect on mental distress. I bound the true effect of an increase in negative sentiment by multiplying the endpoints of the 95% confidence interval in Table 3, Column 1 by my estimates of real changes to sentiment. I find that in the average state, poor mental health days changed by between -0.3 and +0.5 days per month from 2013 to 2018 owing to a +4.9 point change in the slur index. Since the average number of poor mental health days per month in my dataset is 2.6 days, a 0.5 day increase in poor mental health days amounts to a substantial 20% increase in mental distress. Therefore, I cannot reject that an increase in anti-Chinese sentiment had no effect on Asians' mental health outcomes, but I also cannot reject that an increase in anti-Chinese sentiment increased poor mental health days by a moderate amount. This result is consistent under modifications to the regression variables and under the addition of the population share variable.

The null result likely comes from imprecise confidence intervals. In particular, my dataset has a small sample size and noisy data: the outcome variable is limited to N=264 owing to restrictions on the amount of mental health data available on Asians, and the endogenous variable of interest comes from Google Trends, which tends to be noisy for rarer search terms. Exact values for Google searches for the slur are also obfuscated by searches for the American musician Chink Santana and the English phrase "chink in the armor." With better data on both mental health and anti-Chinese sentiment, one could narrow the confidence interval to obtain a precise result on the coefficient on sentiment. I would recommend using the ANES temperature rating data for the sentiment variable, as long as there is enough mental health data to match the ANES data.

To conclude, the null result suggests that anti-Chinese sentiment related to economic tensions translates moderately but not significantly into discriminatory action that affects Asian mental health outcomes. Yet there is strong anecdotal evidence from the 2,583+ hate incidents against Asian Americans this year that discriminatory beliefs translate into physical and verbal violence. It very likely that anti-Chinese sentiment is expressed more or less depending on the state of American culture. In times of increased division and xenophobia, sinophobes feel more comfortable attacking the targets of their hate. As mentioned earlier in the paper, I suggest that further research takes these concerns into account by controlling for deportation rates and a measure of xenophobia in the model. With additional study, it may be possible to better understand the determinants of Asian American mental health outcomes in the U.S.





## References

**Ahrens, Rita Pin**, "COVID-19 Impact on Asian American and Pacific Islander Mental and Physical Health," 2020, p. 5.

**Alawin, Mohammad**, "Trade balance and unemployment in Jordan," *European Scientific Journal*, 2013, *9* (7). Publisher: Citeseer.

**Chatterji, Pinka, Margarita Alegria, Mingshan Lu, and David Takeuchi**, "Psychiatric disorders and labor market outcomes: evidence from the National Latino and Asian American Study," *Health economics*, 2007, *16* (10), 1069–1090. Publisher: Wiley Online Library.

**Gee, Gilbert C., Michael Spencer, Juan Chen, Tiffany Yip, and David T. Takeuchi**, "The association between self-reported racial discrimination and 12-month DSM-IV mental disorders among Asian Americans nationwide," *Social science & medicine*, 2007, *64* (10), 1984–1996. Publisher: Elsevier.

**Jeung, Russell, Tara Popovic, Richard Lim, and Nelson Lin**, "Anti-Chinese Rhetoric Employed by Perpetrators of Anti-Asian Hate," October 2020, p. 3.

**López, Gustavo, Neil Ruiz, and Eileen Patten**, "Key facts about Asian Americans," September 2017.

**NHANES**, "NHANES 2015-2016 Overview," 2016.

**Silver, Laura, Kat Devlin, and Christine Huang**, "Americans Fault China for Its Role in the Spread of COVID-19," July 2020.

**TRAC**, "Ten-Fold Difference in Odds of ICE Enforcement Depending Upon Where You Live," April 2019.

**Vachuska, Karl**, "Initial Effects of the Coronavirus Pandemic on Racial Prejudice in the United States: Evidence from Google Trends," 2020. Publisher: SocArXiv.

## Appendix

Table 6: Multiple Measures of Anti-Chinese Sentiment

| | (1) Asian Approval Rating | (2) Asian Approval Rating | (3) Asian Approval Rating | (4) Asian Approval Rating |
|---|---|---|---|---|
| Slur Index | 0.094* | | 0.084* | |
| | (0.05) | | (0.05) | |
| Swear Index | | 0.012 | | -0.008 |
| | | (0.03) | | (0.03) |
| Pop. Share | | | 26.331*** | 28.486*** |
| | | | (9.52) | (10.03) |
| N | 100 | 100 | 100 | 100 |
| R-squared | 0.04 | 0.00 | 0.11 | 0.08 |

Notes: Standard errors in parentheses below each estimate. Significance at the 1, 5, and 10 percent levels indicated by ***, **, and *, respectively.



# Dual-Threat: How the Quarterback Position Drives Competitive and Business Success in the NFL

Spencer Hylen

May 10, 2020

**Abstract**

In the NFL, quarterbacks have the largest individual impact on team wins and revenue through their on-field performance and marketability. However, excess investment in a quarterback for revenue generation can restrict a team's ability to invest in other positions, hurting the team's on-field performance. To investigate this trade-off, I use a pair of two-stage least squares regressions with a draft-based instrument. This paper suggests diminishing returns to quarterback cost such that spending more than 13.6% of the salary cap on a quarterback decreases a team's win total. The effect of quarterback cost on wins and revenue was not found to be significant.

# Introduction

One of the hardest questions to answer in professional sports is the value of an individual player. This question is of particular interest to teams that operate in leagues with restrictions on player payroll, known as a "salary cap". The salary cap restriction is most notable in the NFL, where the hard cap and significant position diversity force teams to make difficult investment decisions to optimize their payroll. The greatest debate revolves around the quarterback (QB) position, which is consistently the highest paid[1], highest profile, and most valuable[2] player on a team's roster. This creates a difficult dichotomy for franchises, who, as businesses, need to balance between winning games and generating revenue. Paying the QB strikes the heart of this dichotomy. Signing a QB for a large sum forces less investment in other positions, weakening the team's roster, but makes the team more marketable to the national audience, bringing in more revenue. This paper will empirically investigate the impact of expensive QBs on their respective team's on-field and financial success.

To fix ideas, suppose there is an NFL team with a cheap rookie QB in the final year of their deal. The team can either pay the player market rate or attempt to select a new QB in the amateur draft that takes place every year. If we assume a competitive market, the team will need to compensate the current QB for their expected production. This could be the same expected production as another QB available in the draft but is often more consistent than the expected production of the drafted rookie QB. Because of the NFL's passing premium (Rockerbie 2008) and the position's importance, consistent QB play is expensive and costs much more than the volatile QB play available through the draft. With a hard limit on how much teams can spend, this higher cost restricts the team's ability to invest in other positions. If they choose to go with the cheaper QB available through the draft, they may find themselves with more volatile QB play but more money to invest in the rest of the team[3].

[1]Of the 25 highest paid players in the NFL in 2019, 16 were QBs, including all 11 highest paid players.
[2]Of the last 20 NFL MVPs, 16 were QBs.
[3]Experienced NFL fans will note that once the QB is selected, their performance can be volatile, consistently good, or consistently bad. A priori, it is impossible to know which level of performance they will obtain. As a result, for teams considering drafting a QB, rookie QB play can be considered volatile in expectation.





There isn't a clear answer to this dilemma for a team focused purely on winning. Literature in the space supports paying less for a QB, as signing an expensive QB leads to greater pay inequality which has a negative effect on wins (Borghesi 2007). However, it doesn't consider the impact of salary cap constraint issues created by having an expensive QB. To complicate the decision further, the team is also a business and wants to generate revenue in addition to winning. A low-cost, high-variability rookie could be good for the team's on-field success, but this new rookie likely won't have the star power of the team's expensive QB. Without a big-name, expensive QB, the team will attract less media attention, fan interest, and sponsorship opportunities, all of which earn money for the team. The dilemma presented here raises the core question this paper addresses: should the team re-sign the current QB to an expensive deal or take a chance on the cheaper, high-variability new rookie?

Based on current theory, the effect of signing an expensive QB on winning is ambiguous but suggests diminishing returns to spending on a QB. Spending marginally more for greater consistency should promote winning until it restricts investment too heavily or creates too large of a pay inequality. Theory around which QB draws more revenue is clearer and suggests re-signing the expensive QB at any cost to drive more revenue, but is unsupported empirically. This paper will argue that there is a spending ceiling, where any spending above that ceiling provides more revenue but costs the team wins.

The empirical tests used to investigate this dilemma are a pair of two-stage least squares (2SLS) regressions. Both use QB percentage of the salary cap as the independent variable. They also share an instrument derived from QB prospect availability in the draft and each team's draft pick that year to address endogeneity concerns. The measures of on-field success include regular season, playoff, and championship victories to capture the impact on both regular and postseason success. Profit is not available, so annual revenue is used as a proxy for business success. A number of control variables are included to isolate the other drivers of a team's wins and revenue.

This paper does not find statistically significant results for regular season wins, postseason wins, or revenue. The results for regular season wins suggest diminishing returns to QB cost, with negative returns for spending above 13.6% of the salary cap. A lack of publicly available datapoints on QB contracts, team revenue, and other controls is believed to be the cause of the insignificant results. These results don't directly answer the question of expensive QB vs. rookie QB, but they encourage decision makers to think twice about signing an expensive QB based on the belief it will lead to postseason success or increased revenue.

The paper proceeds in sections as follows:

1: Provides a review of related literature and discussion of data sources.

2: Describes the empirical context.

3: Presents the empirical results.

Section 3 will be followed by a brief conclusion that further contextualizes the results and presents opportunities for future work.

## 1 Literature Review

Based on a review of related literature, most of the work done involving either NFL team wins or salary cap focuses broadly on roster construction. A large portion of the analyses examined were from senior term papers for economic (or similar) degrees. This section will highlight three of these papers and discuss their implications for the context of this paper.

The most similar work to this paper was done by Borghesi (2007), who examined how variations in player compensation influence team success. Specifically, he investigated the fairness of pay distributions and found that teams with greater pay equity (whether justified or unjustified) outperformed their competitors more often than teams that took a higher-paid superstar approach. Since QBs are most often the highest paid player on the roster,





having an expensive QB inherently creates more spending inequality. This suggests that teams who pay more for a QB may have less wins because of both investment restrictions and unequal pay among teammates. Initially, this appears to be an endogeneity concern, as it is capturing two methods through which expensive QBs impact wins. However, as this paper is only concerned with the overall impact, rather than separating out the investment restriction impact, this adds to the supporting theory rather than present a concern.

Kowalski and Leeds (2001) provide support for properly-compensated QBs being classified as "expensive" through their analysis of the change in the Collective Bargaining Agreement (CBA). Their work found that the CBA change in 1993 greatly increased the reward to performance for players in the NFL. Under the hard cap, higher reward to performance exacerbates challenges with player investment decisions. Moreover, even a properly compensated QB will be expensive due to the inflated importance of their production. Together, these provide further justification for the salary cap optimization challenges teams face and the cost of the QB. They also provide additional motivation for investigating the impact of this increased compensation on team performance.

Work has also been done relating the salary cap to team success (Zimmer 2016), but its empirical evidence contains endogeneity concerns. The paper uses a fixed effects regression on winning percentage and includes controls such as salary cap concentration, lagged winning percentage, and the ratio of QB to running back (RB) salary. It does not include any consideration of defense, which leads to the unaddressed endogeneity issue. Defensive spending is inversely correlated with the QB percentage of the cap and correlated (in some way) with winning percentage. This paper was a common example of the senior term papers that comprise most of the relevant work done on this topic. Unfortunately, this slightly limits the usefulness of a large portion of relevant work.

Collectively, the works discussed here help motivate and add additional dimensions to this paper. They also focus primarily on teams' on-field success rather than the team as a business. This is an important consideration when analyzing player compensation decisions, as football teams are also businesses that need to generate revenue. By offering a different perspective of the sports team as a business, this paper will add to and expand on the literature on roster construction.

# 2 Data

The data and empirical tests covered in this section and Section 3 (respectively) are divided into two parts to properly examine each of the drivers for on-field and off-field success. Both sets of panel data are sourced from the same public resources, but they each contain different attributes which are available over different time periods. For example, while QB salaries are available until 2006, they become sporadically available before 2011, while social media data from Facebook and Twitter is only available since 2012. As social media data is an important control in the off-field success tests, this compresses the data set available for this regression. This section describes the data sources and discusses the difference in data structure for each of the empirical tests.

## 2.1 Sources

The tests performed in this paper involve team financials, team on-field performance, and individual performance. To collect sufficient information, data was manually gathered from three main sources:

1. Over The Cap: An independent website that aggregates detailed contract information on NFL players. While their data on recent years is extensive, its player contract coverage is sparse for years prior to 2009. It is the best publicly available resource for NFL contract information.

2. Pro Football Reference: Another independent website, Pro Football Reference aggregates detailed statistics for NFL players and teams dating back to pre-1970. Pro





Football Reference provides both team information (like wins, playoff appearances, championships, etc) as well as information on individual players (awards, etc).

3. Statista: A leading provider of market and consumer data, Statista is a helpful aggregator of business data, including the NFL. Along with standard business information like revenue, they provide information about home game attendance and social media followers.

Details on which specific source provided each variable can be found in Table 1, which contains descriptions and measures of each variable used in the empirical tests. Table 1 also contains information on the other sources used to collect all of the data. As this data was manually gathered, there is no one way to easily draw it from each of these sites. For reproduction of results, each data set is available online as a public Google Drive document[4].

The primary weakness of the data is the accessibility of the sources it was gathered from. The sources are all reliable, but because the data was gathered manually there are a large number of man hours involved in reproducing the data set. Moreover the data available is likely not as extensive as would be available from a different private source. Nonetheless, the points available are both precise and accurate.

## 2.2 Transformations

Data transformations are done sparingly throughout the data sets but are applied consistently across each set. Most transformations are done to adjust units so the coefficients can be interpreted more easily. In each set, points that don't contain all of the attributes necessary for that regression are dropped. For this reason, each set is a different size and may contain different points (although there is significant overlap). The difference in units can be visualized between the raw data sets available online and in Table 1.

[4]https://docs.google.com/spreadsheets/d/1i_71gdlNzCmt-TJFJKOeZwB1-sjLVmRgZDnUgqK44ew/edit?usp=sharing

## 2.3 Data Sets

### 2.3.1 Common Elements

While most of the data gathered in each set is different, they overlap with their measure of QB cost (the independent variable) and their instrument. To properly measure QB cost, the QB with the highest cost that counts towards the hard cap ("cap hit") is used, regardless of if that player starts or plays regularly. Using the highest cost captures the level of investment a team was willing to make in the QB position. While injuries do occur, they are not frequent for QBs and can sometimes result in improved team performance[5]. Signing a QB for that amount means the team felt comfortable paying that much for a QB, even if the team ultimately plays a cheaper one. Since the hard salary cap increases every year, the highest QB cap hit is divided by that year's salary cap to create a measure that is consistent over time. With this consistent measure, conclusions about QB spending and the cap will continue to be relevant into the future.

The instrument used in both regressions is derived from the randomness in the draft and is made up of two key pieces. The first is simply a team's exposure to top prospects available in the draft, measured by the team's draft pick number that year. The second is the strength of QB prospects in the draft. Historical prospect rankings are not often available from professional analysts, so amateur analyst prospect rankings are averaged to create a top prospect consensus. To determine the strength of the QBs in that class, a cutoff is placed after 10 prospects. If more than one QB is present in that top 10 prospect consensus, that class is considered a strong class. These pieces are interacted to create the instrument used in both regressions, which is motivated by the empirical test construction as discussed in Section 3.

[5]Anecdotal evidence for this claim includes Tom Brady over Drew Bledsoe, Nick Foles over Carson Wentz, Ben Roethlisberger over Tommy Maddox.





### 2.3.2 Data Set 1: On-Field Success

The panel data used to investigate on-field success is unique due to its size, different measures of the dependent variable, and small number of control variables. At 418 team instances, this set is much larger than the off-field set (224 team instances). It covers 2006-2019, although some points were dropped because they lacked reliable QB cap information or other relevant controls. Three unique measures of the on-field success dependent variable are used in this regression. Each offers a different perspective of a team's competitive performance that year. The first is regular season wins, which are available for all teams and are a baseline indicator for success. Postseason wins, the second measure, are an intermediate indicator that judge how more successful teams fare when playing each other. Since sports are ultimately about the overall winner, championship (Super Bowl) victories are used to see if the amount spent on a QB sets championship teams apart. Together, these three measures provide a tiered perspective of how QB cost impacts on-field performance.

In addition to the dependent variable measures described above, this set contains three control variables. The first is team strength of schedule, which is used to capture how strong the team's opponents were that year[6]. All else being equal, a team with harder opponents will win less games than a team with easier opponents. The second is relative home attendance, which is used to capture the environment the team plays in. Teams that play in relatively fuller stadiums are generally more enthused and perform better than those who play in emptier stadiums. As Dohmen (2003) finds, they also gain the benefit of favorable calls from officials. The last control is team strength outside of the QB position. As with team strength of schedule, teams with better players should win more than teams with worse players. Table 1 describes the specific quantities used to measure these controls in detail as well as their sources. It also contains the dependent and independent variables used for the on-field empirical test. Table 2 contains summary statistics for this data set.

[6]As each team only plays 13 distinct teams out of 31 possible teams, this number is not mathematically trivial and differs for each team.

#### 2.3.3 Data Set 2: Off-Field Success

Team's primary revenue drivers are usually more complicated than the drivers of on-field success, requiring more control variables and resulting in a smaller data set. The six controls used include:

1. Team historical success: Many fans associate with teams that have a long history of winning. More fans equates to more spending. Additionally, teams with more successful histories can take advantage of fan nostalgia to sell more merchandise.
2. Home market size: Larger home markets attract more fans and more local spending.
3. Team recent success: A number of fans (commonly referred to as "bandwagoners") support and spend on teams that have more recent success.
4. Presence of star players: Star players attract fans and help sell team merchandise.
5. Fan engagement: Some fans follow teams from afar and will still spend on merchandise even if they aren't in the same geographic area.
6. Team local popularity: Teams with higher attendance at their games will bring in more revenue.

This panel data is composed of 224 team seasons from 2012-2018 and was restricted by the availability of social media data, which has only been recorded consistently since 2012. As with the previous set, Table 1 describes the measures and details of these variables, along with a summary statistics table (Table 3).

## 3 Empirical Context

Using the two data sets above, two 2SLS regressions were constructed to evaluate the impact of QB cost on the success of their respective teams measured in wins and revenue. This





section begins by describing the intuition behind each of the empirical tests. I then examine endogeneity concerns present in each of the tests and finally conclude by motivating the instrumental variable solution to those concerns.

### 3.1 Intuition

As discussed in the introduction and Section 1, the current theory and empirical evidence suggests QB cost has an impact on the on and off-field success of the team. The impact of QB cost on competitive performance can be modeled through the following cross-sectional OLS regression:

$$W_{it} = \beta_0 + \beta_1 QB_{it} + \beta_2 QB_{it}^2 + \beta_3 SOS_{it} + \beta_4 ATTEN_{it} + \beta_5 SPRT_{it} + \epsilon_{it}$$

where $W_{it}$ is the number of wins (either regular season, postseason, or championship) for team $i$ in season $t$.

$QB_{it}$ is the independent variable of interest, measured by QB percentage of the salary cap. It is represented linearly and quadratically to capture diminishing returns to QB spending. $SOS_{it}$, $ATTEN_{it}$, and $SPRT_{it}$ are the strength of schedule, home attendance, and strength of team $i$ in season $t$. All these variables are described in Section 2 and in Table 1.

I define a similar cross-sectional OLS regression model for the impact of QB cost on team revenue:

$$R_{it} = \beta_0 + \beta_1 QB_{it} + \beta_2 HIST_{it} + \beta_3 CITY_{it} + \beta_4 REC_{it} + \beta_5 STAR_{it} + \beta_6 FAN_{it} + \beta_7 POP_{it} + \epsilon_{it}$$

where $R_{it}$ is the revenue earned by team $i$ in season $t$.

Business information provided by the NFL is limited, so revenue is the only measure of a franchise's off-field success. $HIST_{it}$, $CITY_{it}$, and $REC_{it}$ are the team's historical success, home market size, and recent success. $STAR_{it}$, $FAN_{it}$, and $POP_{it}$ represent the presence of other star players on the team, fan engagement, and the team local popularity measured via attendance at games. As with the variables in the on-field tests, these are described in Section 2 and in Table 1.

### 3.2 Endogeneity Concerns

The cross-sectional OLS regressions above aren't well specified due to endogeneity in the QB market. Assuming the QB market is informed and competitive, it is reasonable to believe all non-rookie QBs will be "expensive" due to more valuable production and increased reward to performance in the NFL (Kowalski, Leeds 2001). Consistently successful teams won't have access to cheap QBs through the draft and will only sign expensive QBs. Worse teams are stuck drafting a cheap rookie QB. As a result, there is a correlation between a team's performance last year and the cost of their QB. Since success and failure are often persistent in the NFL, there is a causal impact of last year's performance on this year's wins and revenue. This creates endogeneity in both regressions that is addressed by the 2SLS regression.

### 3.3 Instrumental Variables Solution

Correcting the endogeneity concerns expressed above is done through the randomness introduced via the rookie talent pool, where prospect abilities fluctuate wildly and depend on the inherent ability of the athletes available in that draft. The selection order in the draft dictates a team's ability to obtain a cheap QB, as most top prospects are unavailable after the first 10 picks. Obtaining one of these QBs thus depends on two factors: proximity of a team's pick to the beginning of the draft (i.e., having a good draft pick) and the availability of a good QB in that draft. As discussed previously, teams with expensive QBs don't usually have good picks, so having a good pick could have a potential causal relationship with having an expensive QB. This makes a good pick relevant. Interacting this with the availability of a good QB, which is dependent on the random amateur talent pool, creates an instrument that is relevant and exclusive.

As I aim to use both QB cap percentage and QB cap percentage squared, I will use





both "good QB interacted with pick number" and "good QB interacted with pick number" squared. As the unsquared versions of these values are considered relevant and exclusive, it stands to reason that the squared versions are relevant and exclusive as well.

The arguments above are theoretically supported and all but one can be empirically tested. The endogeneity concern is examined with a Hausman test, and the relevance condition for each of the instruments is tested with an additional regression for both data sets. Since the instruments are just numerical manipulations of each other and do not have unique underlying interpretations, an exclusion restriction cannot be used. I am confident the arguments supporting the endogeneity, exclusion, and relevance issues are strong enough that limited empirical testing is not a concern for the validity of those assumptions.

## 4 Results

This section discusses the results of the empirical tests in Section 3 and is divided into three parts. The first discusses the results of the relevance tests. The second examines the empirical results for on-field performance. The third and final part evaluates the empirical test for off-field performance.

### 4.1 Background: Endogeneity and Relevance

As described in Section 3 above, there are a few assumptions required for the 2SLS test to be necessary and valid. For the test to be necessary, there must be endogeneity present. This was examined for both data sets via a Hausman test. On both sets, the Hausman test did not demonstrate that the difference in OLS and 2SLS coefficients was systematic, with $\chi^2$ values of 1.08 and 0.09 (respectively). Despite this, the limited nature of the data set and strong theoretical argument in favor of the 2SLS test with the draft instrument supports its use throughout the remainder of the paper.

For the test to be valid, it must satisfy relevance and exclusion. The results of the OLS regressions to test for relevance can be seen in Table 4 (for the on-field tests) and Table 5 (for the off-field tests). The results of this test for the on-field regressions show that the draft instrument is a relevant, albeit weak, instrument. The F-statistic for the linear version of the instrument is about 3.1, while the F-statistic for the squared version is 1.87. Results for the off-field regression are not much more promising, with F-statistics of 3.43 and 3.23 for the linear and quadratic versions of the test (respectively). Intuitively, weakness in this instrument seems likely. The high correlation of the draft pick with a team's performance is diluted by the random variation of talent in the draft through the interaction of the variables. Since this interaction is necessary for the test to satisfy exclusion, the 2SLS tests use the draft instrument despite its weakness.

Together, these tests demonstrate limited empirical support for the assumptions behind the use of 2SLS tests. Because of the strong theoretical support and limited data availability, it is likely these tests are not truly representative of the population. As a result, this paper believes the 2SLS regression is necessary and that the draft instrument is valid for performing it.

## 4.2 On-Field Performance Results

The results of the 2SLS test for on-field success can be found in Table 6. In the test for regular season wins, only strength of schedule was found to have a statistically significant impact (at the $p < 0.01$ level or below). Additionally, none of the variables were found to have significant effects on postseason play. Because the regular season sample set had the most variance in team performances (as all teams play in the regular season), this section will focus on the results from that test.

Among the coefficients in the regular season version of the test, the largest is on QB cap percentage, which has an estimated impact of 4.49 wins per additional percentage of the cap. Looking at the relationship with the squared variable is even more notable. Holding





controls constant, QB impact is:

$$W_{it} = -0.331QB_{it}^2 + 4.485QB_{it}^2$$

Spending on a QB adds about 4.1 wins per percentage point of cap. However, this effect decreases as the team spends more, validating the diminishing returns theory. The effect on wins actually turns negative for teams spending more than 13.6%. There are some caveats to this threshold. The 95% confidence interval for both coefficients is wide and includes negative and positive values, making it unclear if this specification is accurate. The large standard errors for both give further reason to distrust these results. Despite these potential issues, the regression demonstrates that there could be a spending threshold for QBs where too much spending can make the team worse. Considering this threshold is essential for decision makers when making roster investment decisions.

For the postseason and Super Bowl, significant results are nonexistent. For the Super Bowl, I believe this is due to a lack of data on teams that won. Only 3.1% of teams in the on-field data set won the Super Bowl. The relatively high standard errors support this. Since the Super Bowl is another postseason game, one would expect the relative standard errors to be about the same as in the postseason test. However, they are twice as large, suggesting that there is a need for more data to get a more precise estimate. While there is less evidence to support a lack of data on playoff teams, they only make up 38.5% observations on what is already a small data set. A larger data set could provide more precise tests across the board but especially at the Super Bowl level.

## 4.3 Off-Field Performance Results

The results of the 2SLS test for off-field success can be found in Table 7. In both tests, the amount spent on a QB had no statistically significant effect on revenue. Despite this, there were a number of other significant factors. The population of a team's home city was shown to increase a team's revenue by about \$6M for every additional one million people, and the social media followers of a team increase revenue by about \$41M for every additional one million followers (both significant at the $p < 0.001$ level). The most surprising significant outcome was historical championships *decreasing* revenue by \$12M for each additional win, as one would expect more historical success to increase revenue. Overall, this confirms my assumption that the control variables are important drivers of revenue for a team. I still believe that a high-profile (and highly paid) QB is important to bring in revenue for a team. However, I believe that the QB percentage of the salary cap is not a good measure for this.

## Conclusion

This paper investigated the impact of an expensive QB on a team's on-field success (through wins) and off-field success (through revenue). To do this, two 2SLS tests were used with an instrumental variable based on the variability of talent available in the NFL draft. The results of the test indicated that spending more on a QB had no significant impact on a team's wins or revenue. The test on wins suggested a diminishing returns effect on regular season wins that occurred at around 13.6% of the salary cap, implying that teams spending much more than this threshold should reevaluate their QB and explore alternative options.

As the data set used in this paper was quite limited due to constraints on publicly available data, future work could include reproducing these empirical results with a larger data set. Further investigation into the impact of an expensive QB on revenue through use of other independent variables could help better answer the trade-off question this paper intended to answer.





# Tables

Table 1: On-Field Variable Descriptions

| Variable | Measure | Description | Source |
|---|---|---|---|
| $QB_{it}^{1,2}$ | % of the salary cap | The highest QB salary cap hit on the roster divided by the salary cap for that year. | OTC |
| $W_{it}{}^1$ | Wins | The number of games the team won. | PFR |
| $W_{it}{}^1$ | Playoff wins | The number of playoff games the team won. | PFR |
| $W_{it}{}^1$ | Indicator | Indicator for if the team won the Super Bowl. | PFR |
| $R_{it}{}^2$ | \$ (millions) | Revenue earned by the team for that year. | Statista |
| $SPRT_{it}{}^1$ | Defensive All-Pros | Measures teammate ability and avoids simultaneity by using only defensive players. All-Pros are only given to the best 22 players. | PFR |
| $ATTEN_{it}{}^1$ /POP$_{it}{}^2$ | Average home attendance (10,000s) | "Home field advantage" is crowd energy and a function of attendance. Similarly, more popular teams usually have higher attendance. | Statista |
| $SOS_{it}{}^1$ | Sum of opponent wins | Measures the strength of opponents on a team's schedule. | PFR |
| $HIST_{it}{}^2$ | Championships prior to year $t$ | Fans often follow teams with a history of winning and can be nostalgically inspired to spend more. | PFR |
| $REC_{it}{}^2$ | Last season's playoff wins | "Bandwagon" fans support teams with more recent success, usually measured by playoff wins. | PFR |
| $CITY_{it}{}^2$ | Population (millions) | Location motivates sports team fandom, so a larger local population should drive more revenue. | U.S. Census |
| $STAR_{it}{}^2$ | Non-QB Pro Bowlers | The Pro Bowl is partially determined by a fan vote and is generally a popularity contest. | PFR |
| $FAN_{it}{}^2$ | Total FB /Twitter followers (millions) | Remotely engaged fan populations will spend more money (on merchandise or other products), which can be measured via social media. | Statista |
| $PICK_{it}{}^{1,2}$ | Team draft pick number | Highest 1st round draft pick number (33 if they didn't pick, as there are 32 picks in the 1st round). | PFR |
| $PROS_{it}{}^{1,2}$ | Average # of QBs in top 10 prospects | Averaging the number of QBs ranked in the top 10 provides a consensus rating for each year. | Amateur Scouts* |

OTC: Over the Cap, PFR: Pro Football Reference
* Amateur Scouts includes FootballsFuture.com, NFLDraftGeek.com, and WalterFootball.com.
$^1$: Used in on-field regression, $^2$: Used in off-field regression

Table 2: On-Field Summary Statistics

| | Mean | Median | Std. Dev. | Min | Max |
|---|---|---|---|---|---|
| Wins | 8.0646 | 8 | 3.0587 | 0 | 16 |
| Highest QB Cap Percentage | 7.3586 | 7.1148 | 4.5414 | .3342 | 20.88 |
| Strength of Supporting Cast | .39474 | 0 | .69577 | 0 | 4 |
| Average Home Attendance | 6.7931 | 6.8403 | .84446 | 2.5335 | 9.2721 |
| Strength of Schedule | 127.49 | 128 | 8.2634 | 106 | 151 |
| Top QB Prospect Indicator | .57895 | 1 | .49432 | 0 | 1 |
| Draft Pick Number | 17.679 | 17 | 10.215 | 1 | 33 |
| Top QB Prospect X Pick Number | 10.158 | 5 | 11.549 | 0 | 33 |
| Observations | 418 | | | | |

Table 3: Off-Field Summary Statistics

| | Mean | Median | Std. Dev. | Min | Max |
|---|---|---|---|---|---|
| Revenue | 371.88 | 364.5 | 100.01 | 229 | 950 |
| Highest QB Cap Percentage | 7.8032 | 7.8825 | 4.6072 | .38075 | 20.88 |
| Team Historical Championships | 1.5313 | 1 | 1.7815 | 0 | 6 |
| Recent Success | .34375 | 0 | .75919 | 0 | 4 |
| Teammate Popularity | 3.317 | 3 | 2.2134 | 0 | 10 |
| Facebook Fans and Twitter Followers | 3.1603 | 2.475 | 2.3338 | .33 | 12.43 |
| Average Home Attendance | 6.8031 | 6.8508 | .8613 | 2.5335 | 9.2721 |
| Home City Population | 1.2536 | .60172 | 1.93 | .10406 | 8.1751 |
| Top QB Prospect Indicator | .57143 | 1 | .49598 | 0 | 1 |
| Draft Pick Number | 17.388 | 17 | 10.059 | 1 | 33 |
| Top QB Prospect X Pick Number | 9.8839 | 5 | 11.405 | 0 | 33 |
| Observations | 224 | | | | |





Table 4: On-Field Relevance Test Results

| | (1)<br>Highest QB Cap % | (2)<br>Highest QB Cap % Squared |
|---|---|---|
| Top QB Prospect X Pick Number | 0.0470* | |
| | (0.0183) | |
| Strength of Supporting Cast | -0.443 | -7.098 |
| | (0.314) | (4.969) |
| Average Home Attendance | 0.166 | 2.492 |
| | (0.288) | (4.633) |
| Strength of Schedule | -0.0553* | -0.695 |
| | (0.0253) | (0.383) |
| Prospect X Pick Number Squared | | 0.0165 |
| | | (0.0100) |
| Constant | 12.99*** | 145.4* |
| | (3.832) | (59.42) |
| Observations | 418 | 418 |
| F | 3.101 | 1.875 |

Standard errors in parentheses

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

*Notes*: Estimates from standard OLS models. Standard errors shown above are robust. Clustered errors were considered but not implemented due to inexperience and time constraints. Includes all observations in the on-field data set. Detailed variable description can be found in Table 1. Results show that the average number of top QB prospects interacted with a team's first-round pick number provides a relevant (albeit weak) instrument for QB cap percentage.

Table 5: Off-Field Relevance Test Results

| | (1) Highest QB Cap % | (2) Highest QB Cap % Squared |
|---|---|---|
| Top QB Prospect X Pick Number | 0.0499 | |
| | (0.0265) | |
| Team Historical Championships | 0.0649 | 1.942 |
| | (0.238) | (4.188) |
| Recent Success | -0.376 | -9.322 |
| | (0.412) | (6.438) |
| Teammate Popularity | -0.0664 | -1.072 |
| | (0.130) | (2.098) |
| Facebook Fans and Twitter Followers | 0.601** | 10.58** |
| | (0.230) | (3.531) |
| Average Home Attendance | -0.0501 | -2.288 |
| | (0.435) | (7.202) |
| Home City Population | 0.116 | 1.725 |
| | (0.157) | (2.896) |
| Prospect X Pick Number Squared | | 0.0167 |
| | | (0.0150) |
| Constant | 5.858* | 61.98 |
| | (2.916) | (47.19) |
| Observations | 224 | 224 |
| F | 3.429 | 3.229 |

Standard errors in parentheses

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

*Notes*: Estimates from standard OLS models. Standard errors shown above are robust. Clustered errors were considered but not implemented due to inexperience and time constraints. Includes all observations in the off-field data set. Detailed variable description can be found in Table 1. Results show that the average number of top QB prospects interacted with a team's first-round pick number provides a relevant (albeit weak) instrument for QB cap percentage.





Table 6: 2SLS Results: On-Field Success

| | (1) Wins | (2) Playoff Wins | (3) Championships |
|---|---|---|---|
| Highest QB Cap Percentage | 4.485 | 0.692 | 0.00421 |
| | (4.678) | (0.703) | (0.146) |
| Highest QB Cap Percentage Squared | -0.331 | -0.0470 | 0.00109 |
| | (0.403) | (0.0610) | (0.0123) |
| Strength of Supporting Cast | 1.213 | 0.280 | 0.0374 |
| | (0.915) | (0.147) | (0.0310) |
| Average Home Attendance | 0.635 | 0.128 | 0.0130 |
| | (0.506) | (0.0719) | (0.0109) |
| Strength of Schedule | -0.156** | -0.00656 | 0.000536 |
| | (0.0550) | (0.00909) | (0.00165) |
| Constant | 14.89 | -1.368 | -0.252 |
| | (9.522) | (1.559) | (0.298) |
| Observations | 418 | 418 | 418 |

Standard errors in parentheses

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

*Notes*: All estimates from a two-stage least squares (2SLS) model. Includes all observations in the on-field data set. Standard errors shown above are robust. Clustered errors were considered but not implemented due to inexperience and time constraints. Detailed variable descriptions can be found in Table 1. Strength of supporting cast is measured in All-Pro awards given to defensive teammates. Home attendance is measured per 10,000s of people, and strength of schedule is measured in wins. Cap percentage is measured as a whole number (i.e. $3.5 = 3.5\%$). The difference in coefficients between QB cap percentage and QB cap percentage squared supports the diminishing returns hypothesis and suggests there are some initial benefits to paying more for a QB, but investing too much can be detrimental to the overall success of the team.

Table 7: 2SLS Results: Off-Field Success

| | (1) Revenue |
|---|---|
| Highest QB Cap Percentage | -0.801 |
| | (6.947) |
| Team Historical Championships | -12.36*** |
| | (3.375) |
| Recent Success | -11.49 |
| | (5.902) |
| Teammate Popularity | 0.424 |
| | (1.661) |
| Facebook Fans and Twitter Followers | 38.06*** |
| | (5.898) |
| Average Home Attendance | 23.15** |
| | (7.139) |
| Home City Population | 4.798** |
| | (1.801) |
| Constant | 115.8 |
| | (61.21) |
| Observations | 224 |

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

*Notes*: All estimates from a two-stage least squares (2SLS) model. Both models include all observations in the off-field dataset. Standard errors shown above are robust. Clustered errors were considered but not implemented due to inexperience and time constraints. Detailed variable descriptions can be found in Table 1. Home city population, social media followers, and team revenue are measured in millions of units. Home attendance is measured in 10,000s of people. Recent success is the number of playoff wins the team had a year prior. Teammate popularity is the number of Pro Bowl awards given to teammates that year. Cap percentage is measured as a whole number (i.e. 3.5 = 3.5%). In both regressions the amount spent on a QB had no statistically significant effect on revenue, although other controls were found to be significant.





# Figures

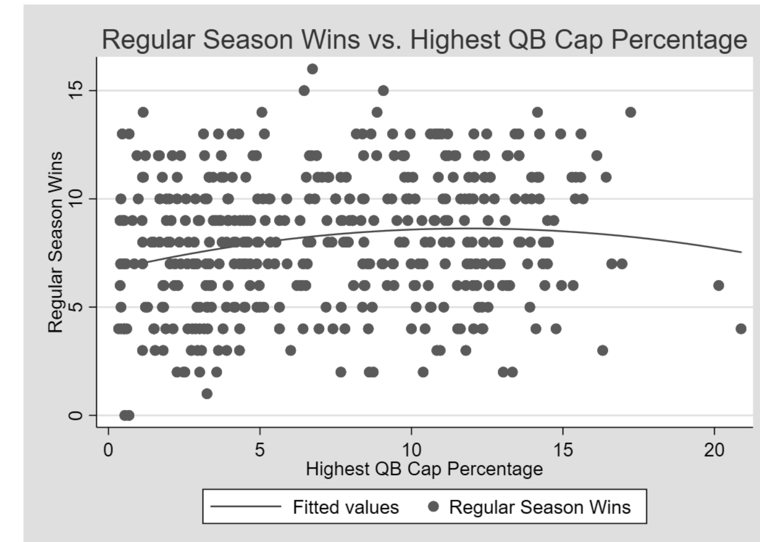



Figure 1: Best fit of regular season wins with highest QB cap percentage and highest QB cap percentage squared. Note the diminishing returns trend that becomes negative as more of the cap is invested in a single player at QB. This supports the results of the regression that show QB cost has a causal impact on a team's ability to win.

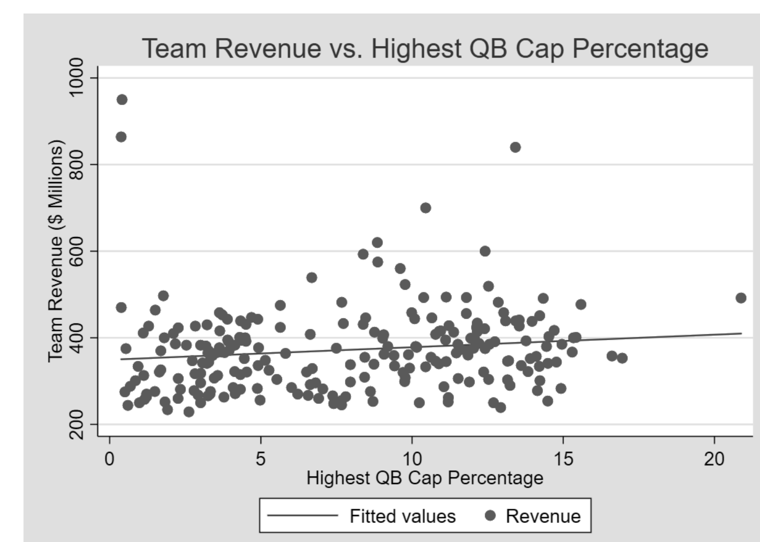



Figure 2: Best fit of revenue data with highest QB cap percentage. Note a slight positive trend that isn't present in the results of the regression (Table 7). This is likely due to confounders that are removed by the instrument and control variables. The positive trend could then be a result of reverse causality: teams that make more money have a bigger brand to represent, and are willing to spend more on a QB to help represent that brand nationally.

Works Cited

Borghesi, Richard. "Allocation of Scarce Resources: Insight from the NFL Salary Cap." *Journal of Economics and Business*, vol. 60, no. 6, 2008, pp. 536–550., doi:10.1016/j.jeconbus.2007.08.002.

Dohmen, Thomas J. "In Support of the Supporters? Do Social Forces Shape Decisions of the Impartial?" *Institute for the Study of Labor (IZA)*, Apr. 2003.

Leeds, Michael A., and Sandra Kowalewski. "Winner Take All in the NFL." *Journal of Sports Economics*, vol. 2, no. 3, 2001, pp. 244–256., doi:10.1177/152700250100200304.

Rockerbie, Duane W. "The Passing Premium Puzzle Revisited." *Journal of Quantitative Analysis in Sports*, vol. 4, no. 2, 2008, doi:10.2202/1559-0410.1093.

Zimmer, Timothy E. "The Impact of NFL Salary Cap Concentration on Team Success." *Choregia*, vol. 12, no. 1, 2016, pp. 53–66., doi:10.4127/ch.2016.0108.



# Effect of Public Library Access on K-12 English Language Arts Performance

Whitney Zhang

May 11 2020

Does public library access affect K-12 students' English Language Arts performance? Using data from the California Standardized Testing and Reporting exam and the Institute of Museum and Library Services, this paper estimates the impact of having a public library nearer or farther from a school. I measure that for schools that do not have a public library within a 10 mile radius, introducing a public library within 5 miles increases the percentage of students that are proficient in English Language Arts in a school by around 7 percentage points. However, for schools that already have a library nearby, marginal changes in distance of a school to a library have little to no effect. Combined with other empirical specifications, these results suggest that introducing a nearby public library is an effective public investment for schools very distant from a library, but ineffective otherwise.

Public libraries are an important social institution that support learning and literacy. Today, there are thousands of public libraries in the US, over 171 million Americans have library cards, and over 81% of Americans have visited a public library. Libraries are built and supported by the public fisc and are touted by advocates as critical to improving educational outcomes and ensuring universal access to books and services. There have been some studies on the effect of libraries on educational outcomes in non-US countries, but studies of libraries in the US have only examined correlations between public library quality and access and educational performance.

The free services that libraries provide have the potential to counteract credit constraints to education faced by families. Nevertheless, the question of their impact on educational outcomes still remains to be answered. Does simply providing access to these resources result in increased utilization of materials and services? What is the return to public libraries, considering that many schools have their own in-school libraries? This paper provides empirical evidence on the effect of public libraries on education outcomes by exploiting the opening and closing of libraries in California.

For this investigation, I construct a new data set tracking the straight-line distance of schools in California to their nearest public library in school years 2002-2003 to 2012-2013. I first conduct a pooled OLS regression and find that, counter-intuitively, increasing the distance from a school to its nearest library is associated with an increase in the percentage of students that are at or above proficiency on the California English Language Arts Standardized Testing and Reporting exam. An alternative specification that uses dummies for different distance ranges shows that increasing the distance from a school to its nearest library is negatively associated with performance when the distance increase is large (over 5 miles). However, both of these specifications are subject to omitted variables bias. Even after controlling for demographic factors, areas with and without libraries introduced or closed are not perfect comparison groups; those that have libraries introduced could have more educational resources, or having found that English Language Arts performance has been poor, be more likely to build a public library.

Therefore, to control for such bias, I conduct a difference-in-difference analysis by conducting a two-way fixed effect regression with school and year fixed effects. I argue for parallel trends by showing that, controlling for demographic covariates, the distance to the nearest library two years into the future is not associated with whether a school will have its nearest public library closer or farther in the current year. My difference-in-difference estimates show that for the average school, increasing the distance from a school to the nearest library by one mile results in a statistically insignificant effect during and two years





after the treatment year. In the first year after treatment, there is a slight decrease in the percentage of students that are proficient in English Language Arts in a school on the order of 0.1 percentage points. However, for schools that do not have a public library within a 10 mile radius, introducing a public library within 5 miles increases the percentage of students that are proficient in English Language Arts by around 7 percentage points.

This paper joins Rodríguez-Lesmes, Trujillo, and Valderrama (2014)'s difference-in-difference case-study of the introduction of two large public libraries in low-income areas in Bogotá, Colombia, which compares results in national standardized test scores before and after the libraries' opening for schools close and far from the libraries. They find no statistically significant impact of the libraries' introduction, with point estimates showing an average score increase of 0.02 to 0.06 standard deviations for being within 1500 meters of a public library.

The paper proceeds as follows. Section I provides a brief institutional background about public libraries and K-12 schools in California. Section II provides a description of my data construction. Section III provides empirical results. Section IV presents concluding remarks.

## 1 Background

In California, the number of physical public library outlets increased from 1084 to 1116 between 2003 and 2013, according to the Institute of Museum and Library Services' Public Library Survey. In this period, 87 library outlets were introduced and 55 library outlets were closed, resulting in changes in the distance to their nearest library for many schools. Table 1 shows between years how many schools had their distance to nearest library change, as well as summary statistics for the sign and magnitude of the change. Most changes are within a mile or less.

There is a large literature on libraries *in* schools. Kachel (2013) compiled a series of studies of school libraries in a variety of states around the US showing positive correlations between reading and writing standardized test scores and having full-time certified librarians, library support staff, flexible scheduling, more computers and technology in libraries, higher library circulation, newer collections, higher library budgets, and more professional development. Additionally, these factors are also positively correlated with the closing of achievement gaps. For example, students who are poor, minority, or disabled are over twice as likely to have "advanced" writing scores if they have a full-time librarian at their school (Lance and Schwarz 2012). Although many of these studies include controls for socioeconomic status, they are still likely subject to omitted variables bias. The only study that is robust to such bias is by Borkum, He, and Linden (2013), who conduct a randomized control trial of an Indian school library program in Bangalore. Students who receive the treatment receive access to well-equipped libraries that have regular educational activities. They find no effect on language exam scores or students' attendance rates from providing a library directly to a school, and a negative effect from providing a library through a visiting librarian.

There is some literature on specific aspects of public libraries, such as types of circulated material and specific programs. Celano and Neuman (2001) find that attendance at a library summer reading program in Pennsylvania is correlated with higher student reading levels. Lance and Marks (2008) find that there is a correlation of 0.514 between the amount of childrens' materials circulated by public libraries and fourth-grade reading scores on the National Assessment of Educational Progress, a national assessment of student educational achievement. The Oceano Branch of the San Luis Obispo City-County Public Library system implemented a Raising a Reader program (The Urban Libraries Council, 2007). After three months of program participation, parents reported increased time reading to their children and use of the library system. The study does not report whether student achievement improved. The aforementioned Rodríguez-Lesmes, Trujillo, and Valderrama (2014) study produces the only causal, rather than merely correlative, study of the effect of access to public libraries on student achievement.





## 2 Data

Library data is from the Institute of Museum and Library Services' Public Library Survey. The survey is issued to all public libraries, as defined by state law, in the US and contains information on library visits, circulation, size of collections, public service hours, staffing, electronic resources, operating revenues and expenditures and number of service outlets for over 9000 libraries over fiscal years 2002-2012, which approximately align with school years 2002-2003 to 2012-2013. While there are unit and item nonresponse issues for libraries in outlying areas, this is not an issue for California, our area of interest. Library locations are geocoded 2007 and onwards, but from inspection the provided geocodes are inaccurate. I re-geocode them using the provided addresses with Geocodio. Some addresses are not tagged as fully accurate by Geocodio; these, I re-geocode with the Bing Maps API.

My measure of educational achievement is from the California Department of Education's (CADOE) English-Language Arts Standardized Testing and Reporting examination results for the 2002-2003 through 2012-2013 school years for grades 2 through 11 at the school $\times$ grade $\times$ student subgroup level. All students are required to take the exam, unless they have significant cognitive disabilities or their parent submits a written request to exempt a student from the test. Student subgroups are categories for variables like gender, English language fluency, and socioeconomic status, as well as a category for all students in that grade and school. Each grade $\times$ school $\times$ student subgroup observation includes information on the number of students tested, the percentage of students tested, the percentage of students tested at an advanced level, the percentage of students tested at a proficient level, the percentage of students tested at a basic level, the percentage of students tested at a below basic level, and the percentage of students tested at a far below basic level. Data is not reported for subgroups with fewer than 10 students. The CADOE also has a dataset of school geocodes.

To control for ommitted variabes bias, I control for factors relating to student school performance. I calculate expenditures per student using data on total spending in a school district from the CADOE's Current Cost of Education data set. I also use data on student race and gender from the CADOE's school enrollment data set, which includes school enrollment per grade. Lastly, I use data on the percentage of students eligible for free or reduced lunch from the CADOE's Free and Reduced Price Meal Eligibility data set. Data is available for all years of interest.

I keep only schools that are present in the dataset in all years and are only part of one school district in the entire period. I consider a school to be present in the dataset even if it does not have any scores presented due to having fewer than ten students in each grade. I remove all library outlets that are bookmobiles. California does not have any mail-only libraries in 2002-2013. I sum the percentages of all students that test at an "Advanced" or "Proficient" level to obtain the percentage proficiency for each school, as students who test at an "Advanced" level also meet proficiency. I geocode schools and libraries that have missing geocodes using Geocodio and calculate for each school the straight-line, or Haversine, distance to its nearest library. Haversine distance is not the true distance to the location based on the best route to get from the school to the library and does not account for transportation in the area, but nevertheless serves as a proxy for ease of access to a public library.

Table 2 presents summary statistics for the variables. On average, schools are within 1.5 miles from a public library. The distribution is very left skewed, with over 75% of schools within 2 miles from a public library, and few schools very far from a library. Figure 1 presents boxplots of the distribution of the percentage of students in a school that are proficient over 2003-2013. There is a significant upward time trend, with the median percentage proficiency rising from 28% to 43%. The distribution of the percentage of students that are proficient in schools over time does not change substantially.





# 3 Empirical Results

To test the effect of library distance on student performance, I begin with a pooled OLS estimate, and then add school and year fixed effects. I argue that my fixed effects regressions provide difference-in-difference estimates through establishing the parallel trends assumption. The point estimates of coefficients in the difference-in-difference estimates and pooled OLS estimates are of the same sign, but the difference-in-difference estimates are of smaller magnitude.

## 3.1 Pooled OLS Estimates

First, I estimate the following pooled OLS model:

$$Y = \beta_0 + \beta_1(DIST) + \beta_2(DIST^2) + \beta_3(DIST^3) + \lambda X + \epsilon$$

where $Y$ is the percentage of students that are proficient in each school $\times$ year observation, $DIST$ is the distance from a school to its nearest library, and $X$ are covariates as described above. The coefficient on $DIST$ is the main estimate of interest. Estimates are displyed in Table 3. I find that increasing the distance from a school to its nearest library by one mile is associated with an increase in the percentage of students proficient at that school by 0.538 percentage points, controlling for the fraction of students eligible for free or reduced price meals, the fraction of students that are Hispanic, and the expenditures per student of that school district. The coefficients on $DIST^2$ and $DIST^3$ are precisely measured to be very small. The coefficients on $DIST$ are very small in columns 3, 4 and 5 of Table 3; the distance from a school to its nearest library has little association with its the percentage of students at that school that reach proficiency.

Figure 2 illustrates the spread of the data and the estimates. The vast majority of schools are within around 0 to 7 miles of a library, with a large spread of scores. As such, the regressions mostly focus on this cluster of schools. However, from visual inspection, it's clear that the regression lines do not match well the proficiency rates of schools that are further from a library.

Therefore, I also conduct a piecewise regression estimate using dummies $D_{mile}$ to represent schools with distances to the nearest library between 0 and 2.5 miles, 2.5 to 5 miles, 5 to 7.5 miles, and 7.5 to 10 miles:

$$Y = \beta_0 + \beta_1 D_{2.5} + \beta_2 D_5 + \beta_3 D_{7.5} + \beta_4 D_{10} + \lambda X + \epsilon$$

Table 4 displays results. The variables of interest here are $\beta_1$ through $\beta_4$, with each coefficient representing the percentage point increase of students proficient in a school within that mile range of distance to a nearest library as compared to a school that is over 10 miles away from a library, controlling for covariates.

The variables of interest are both statistically significant and economically significant. Having a library within 10 miles of a school, as opposed to further than 10 miles from a school, is associated with a 5.6 point increase in the percentage of students that are proficient in English Language Arts at that school. Similarly, having a library be within 7.5 or 5 miles rather than further than 10 miles provides a 2.1 point or 1.9 point increase in the percentage of students proficient, respectively. However, having a library within 2.5 miles versus 5 miles of a school provides no statistically significant change. This matches the continuous fixed effects findings above, since most of the observations that the regression relies on for variation are of schools within the 5 mile range. Thus, the results suggest that having a library within 5 miles of a school could bring benefits to student English Language Arts proficiency, but any closer makes little to no difference.

Notably, it is important to be wary of these results, since the number of observations of schools whose nearest library is over 5 miles away is relatively small. Moreover, there is likely still significant omitted variables bias present in the above regressions. The coefficients and standard errors are not robust to the addition of more explanatory variables, as is clear





by comparison of the coefficients on the variables of interest in columns 4 and 5 in Table 3 and in columns 3 and 4 in Table 4.

### 3.2 Fixed Effect/Difference-in-Difference Estimates

Therefore, to correct for such bias, I estimate the following two-way fixed effects model for each school:

$$\begin{aligned} Y_{it} =& \beta_0 + \beta_{t-2}(DIST_{it-2}) + \beta_{t-1}(DIST_{it-1}) + \beta_t(DIST_{it}) + \beta_{t+1}(DIST_{it+1}) + \beta_{t+2}(DIST_{it+2}) \\ &+ \lambda X + \alpha_i + \nu_t + \epsilon_{it} \end{aligned}$$

where $t$ represents the treatment year and $i$ represents each school. Results are displayed in Table 5 and Figure 3. The coefficients on $DIST_{t+1}$ and $DIST_{t+2}$ in the second column, have 95% confidence intervals of [-0.231, 0.005] and [-0.207, 0.181], respectively, which are tightly around 0. Conditional on FFRPM, fraction Hispanic, and expenditures per student, the coefficients on $DIST_{t+1}$ and $DIST_{t+2}$ in the fifth column are even closer to 0 and more tightly bound with 95% confidence intervals of [-0.220, 0.004] and [-0.195, 0.189] respectively. Whether a library will get a library one or two years in the future is not associated with their current proficiency percentage, indicating that the treatment and control groups have similar trends in the past two years of proficiency rates.

Additionally, the coefficients on all $DIST$ variables are robust to the addition of controls, as is seen in comparing column 2 to columns 3-5, unlike in the OLS regression in Table 2, indicating that the year and school fixed effects have accounted for most of the omitted variables bias. Therefore, given parallel trends and the accounting of omitted variables bias, I argue that column 5 provides a causal difference-in-difference estimate.

The coefficients on $DIST_t$ and $DIST_{t-2}$ have 95% confidence intervals of [-0.146, 0.046] and [-0.159, 0.123], respectively; the effect of having a library closer or father from a school is precisely measured to around zero. The coefficient on $DIST_{t-1}$ is estimated to be -0.155 with a 95% confidence interval of [-0.275, -0.035]. If the distance from a school to its nearest library reduces by one mile, the percentage of students that are proficient in English Language Arts decreases by 0.155 in the following year, although this is still potentially an overestimate, given that the coefficient on $DIST_{t-2}$ is negative and of similar magnitude. That is, in a school of 1000 students, only 1.5 students would move from below to at or above proficiency if the nearest public library were a mile farther. Taken together, these results indicate that the distance of a school to a public library has little to no effect on its students' English Language Arts proficiency. However, it is worth noting that as illustrated before, most of the data and variation is in schools that already are relatively close to a library.

I next repeat the cross-sectional regressions with distance indicators and fixed effects for a two-way fixed effects model:

$$Y_{it} = \beta_0 + \beta_1 D_{2.5it} + \beta_2 D_{5it} + \beta_3 D_{7.5it} + \beta_4 D_{10it} + \lambda X + \alpha_i + \nu_t + \epsilon_{it}$$

As before, from examining Table 6 we find statistically significant effects of having a library within 0 to 2.5 miles, within 2.5 to 5 miles, and within 7.5 to 10 miles. The effect of having a library within 5 to 7.5 miles is not significant, due to a lack of schools that change from having a library within 5 miles to having a library within 7 to 7.5 miles or from having a library beyond 7.5 miles to having a library within 5 to 7.5 miles. All of these effects have large standard errors, much larger than in the OLS regression, due to a lack of schools whose nearest library distances vary such that they are able to "flip" an indicator. Additionally, the effects (column 4 in Table 6) are smaller than in the OLS regression (column 4 in Table 4), for having a library within 0 to 2.5 miles, within 2.5 to 5 miles, and within 5 to 7.5 miles, but the effect is larger for having a library within 7.5 to 10 miles. Specifically, having a library within 0 to 2.5 miles rather than beyond 10 miles increases the percentage proficiency of a school by 7.405 percentage points — versus the OLS estimate of 9.301 percentage points, having a library within 2.5 to 5 miles rather than beyond 10 miles increases percentage proficiency by 7.293 percentage points — versus the OLS estimate of 9.423 points, and





having a library within 7.5 to 10 miles rather than beyond 10 miles increases the proficiency rate by 8.401 percentage points — versus the OLS estimate of 7.754 percentage points. All results are controlling for at each school the fraction of students eligible for free or reduced price meals, the fraction of students that are Hispanic, and expenditures per student. The smaller magnitude of these effects as opposed to the OLS regression indicate that there was omitted variables bias that was controlled for by the school and time fixed effects. Moreover, as expected, the bias was positive: schools with libraries nearer to them perform better on standardized tests in English Language Arts, perhaps due to confounding factors like a community that greater values education or differences in socioeconomic factors that cannot fully be captured through the demographic controls.

To check whether the parallel trends assumption is met for these fixed effects estimates, I test each of these dummies using the same test on leads as in the continuous distance difference-in-difference estimate. That is, for all $m \in \{2.5, 5, 7.5, 10\}$, I perform the following estimate:

$$\begin{aligned} Y_{it} = &\beta_0 + \beta_{m,t+2} D_{m,it+2} + \beta_{m,t+1} D_{m,it+1} + \beta_{m,t} D_{m,it} + \beta_{m,t-1} D_{m,it-1} + \beta_{m,t-2} D_{m,it-2} + \\ &\beta_{m' \neq m} D_{m' \neq m,it} + \lambda X + \alpha_i + \nu_t + \epsilon_{it} \end{aligned}$$

Examining Figure 4, I find that parallel trends do not hold for $D_{7.5it}$ and $D_{10it}$. The coefficients on $D_{7.5it+2}$ and $D_{10it+2}$ are significant at the 0.05 level and are -3.215 and 3.541, respectively — certainly not 0. Schools who obtain a closest library between 5 to 7.5 miles have decreasing performance in the two years prior to treatment relative to their peer institutions (ie., after controlling for demographic factors) who do not obtain a closest library between 5 to 7.5 miles. On the other hand, schools who obtain a closest library between 7.5 to 10 miles have increasing performance in the two years prior to treatment relative to their peer institutions (ie., after controlling for demographic factors) who do not obtain a closest library between 7.5 to 10 miles. (Alternatively, the terms "decreasing" and "increasing" can be reversed if libraries are thought of as closing, rather than opening.) As such, the coefficients on $D_{7.5it}$ and $D_{10it}$ cannot be interpreted as causal estimates. The data and empirical tests specified unfortunately cannot inform us of the effect of having a library within 5 to 10 miles, rather than 10 miles, from a school.

However, parallel trends do hold for $D_{2.5it}$ and $D_{5it}$. The coefficients for $D_{2.5it+1}$ and $D_{2.5it+2}$ have 95% confidence intervals [-1.13, 0.771] and [-0.573, 1.20], respectively. The coefficients for $D_{5it+1}$ and $D_{5it+2}$ have 95% confidence intervals [-0.543, 1.36] and [-1.17, 0.632], respectively. There is no additional statistically significant effect in the first or second year post-treatment. As such, we can state with high confidence that the coefficients on $D_{2.5it}$ and $D_{5it}$ are indeed causal estimates.

There is a notable difference between the statistically and economically significant effect using the binned measure of distance and the near-zero estimate using the continuous measure of distance. Prior, I argue that the linear difference-in-difference fixed effect model shows there is little to no effect of changing the distance from a library by one mile. In particular, during and two years after the treatment year the effect is very closely bounded around zero, and in the year following the treatment year the effect of having a public library one mile further is a mere 0.155 percentage point decrease in the percentage of students that are at or above proficiency in that school. Here, however, I find that having a library within 5 miles of a school as opposed to within 10 miles of a school increases the percentage of students in a school that are at or above proficient by over 7 percentage points, a major increase: in a school of 1000 students, over 70 students would come to meet proficiency. Recall that most schools in the data are schools that are already very close to a library. As such, we learn that minor shifts in library distance for schools that are already reasonably close to a library do not have an impact on students' performance, potentially because once a library is nearby — here, within around 5 miles — its resources become accessible to schools and students. On the other hand, if a school does not have a library within an accessible distance, introducing a public library can be a great benefit to the learning of students in that school; likewise, for a school that only has a single public library nearby, closing a public





library can be a derailment to many students.

## 4 Conclusion

Libraries are an important public institution. At a basic level, they provide literature and other forms of media. Today, many also provide other resources and services, such as internet access, tutoring for English learners, and children's reading programs. Yet, little is known about how these institutions effect K-12 educational outcomes in the communities they serve. The introduction and closure of libraries provides a useful empirical context, as each introduction and closure may affect the library access of many nearby schools, each to a different extent. I find that for schools that are further than 10 miles from a public library, introducing a public library within 5 miles of that school increases the percentage of students that are proficient in English Language Arts by around 7 percentage points. Yet, for schools in general, a marginal change in the distance to a public library has little to no effect, as most schools already have a public library within 5 miles.

There are some caveats to this interpretation. First, I use straight line distance to the nearest library as a proxy for access to a library. This does not consider actual driving distance or the availability of public transit, which may affect the true ease of access to a library and thus bias the estimate through measurement error. For example, in an area with few cars and poor public transit, a distance of five miles may be quite far for some students, whereas in an area with many cars or good public transit, a library five miles away would be more accessible. Areas with fewer cars and poorer public transit may have worsened academic performance, regardless of its public libraries. Second, I do not control for library quality. It is possible that libraries that are closer to schools are different in quality compared to libraries that are further from schools. Some future possibilities for improvement include using actual driving or public transportation distance, considering the effects of different qualities of libraries, such as through measuring the number of books available or the hours of operations, and using student-level national data, which would provide a larger sample and allow for student-level fixed effects for more precise estimates.

Nevertheless, this paper provides an interesting contribution to the literature on public libraries by finding that the marginal effect of having a closer public library quickly diminishes once a school has one library that is relatively close. Additionally, it has multiple implications for public policy. First, funding for introducing more public libraries should be targeted at schools who are very far from a public library, insofar as one focuses on the educational effects of a public library. Second, for areas that already have a public library relatively close, merely having closer access in terms of distance to a public library is not enough to increase student performance. Libraries in these areas should consider more targeted programs towards families and developing better school partnerships, which have been shown to be effective.





# References


Borkum, E., F. He, and L. L. Linden (2013, Mar). The Effects of School Libraries on Language Skills: Evidence from a Randomized Controlled Trial in India. Technical report, Institute of Labor Economics (IZA).

Celano, D. and S. B. Neuman (2013). The Role of Libraries in Childrens' Literacy Development.

Kachel, D. (2013). School library research summarized: A graduate class project. Mansfield, PA: Mansfield University.

Lance, K. C. and R. B. Marks (2008). The Link Between Public Libraries and Early Reading Success. *School Library Journal*.

Lance, K. C. and B. Schwarz (2013). How Pennsylvania School Libraries Pay Off: Investments in Student Achievement and Academic Standards.

Rodríguez-Lesmes, P., J. D. Trujillo, and D. Valderrama (2014). Are Public Libraries Improving Quality of Education? When the Provision of Public Goods is not Enough. *Desarrollo y Sociedad 74*(5), 225–274.

The Urban Libraries Council (2007). Making Cities Stronger.

# Appendix

Table 1: Nonzero Change in Distance to Nearest Library (miles), 2003-2013

| Statistic | N | Mean | St. Dev. | Min | Pctl(25) | Pctl(75) | Max |
|---|---|---|---|---|---|---|---|
| 2003-2004 | 0 | * | * | * | * | * | * |
| 2004-2005 | 239 | 0.026 | 1.149 | −6.286 | −0.173 | 0.353 | 3.288 |
| 2005-2006 | 227 | −0.546 | 3.483 | −28.946 | −0.443 | 0.280 | 2.033 |
| 2006-2007 | 179 | −0.343 | 0.915 | −7.873 | −0.546 | 0.034 | 1.876 |
| 2007-2008 | 260 | −0.377 | 1.427 | −5.726 | −0.705 | 0.064 | 12.242 |
| 2008-2009 | 192 | −0.204 | 0.673 | −4.443 | −0.272 | 0.096 | 1.378 |
| 2009-2010 | 34 | −0.234 | 1.570 | −3.250 | −1.222 | 0.857 | 2.372 |
| 2010-2011 | 231 | 0.246 | 3.335 | −2.335 | −0.264 | 0.091 | 28.946 |
| 2011-2012 | 0 | * | * | * | * | * | * |
| 2012-2013 | 0 | * | * | * | * | * | * |

*Note:* * are for instances in which the statistic does not apply due to a lack of observations. School-level observations for school years 2002-2003 through 2012-2013. Schools that are introduced, closed, switch districts, or are part of multiple districts during 2003-2013 are not included. Only central and branch libraries are included.

Table 2: Summary Statistics for School-Level Data, 2003-2013

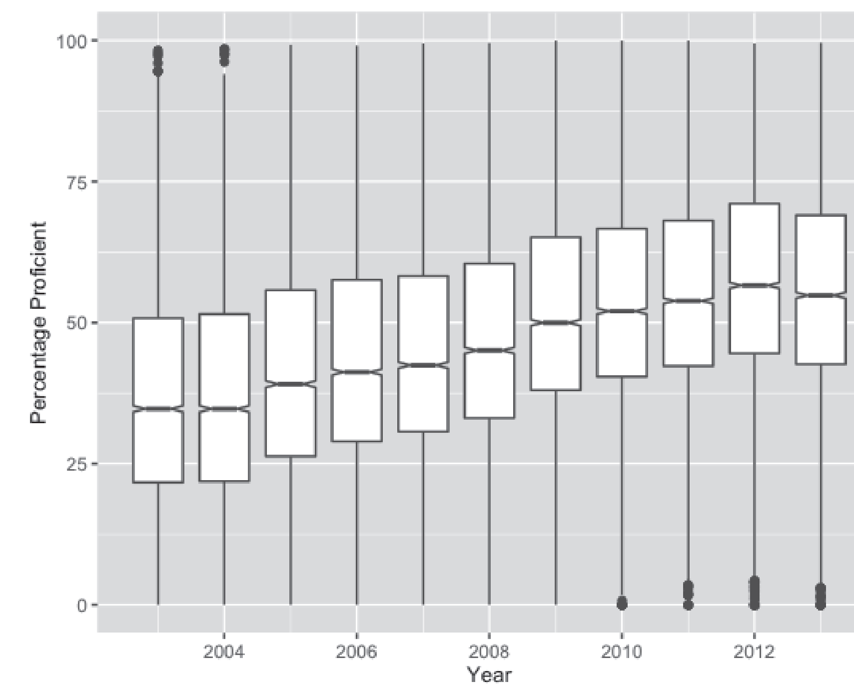
| Statistic | N | Mean | St. Dev. | Min | Pctl(25) | Median | Pctl(75) | Max |
|---|---|---|---|---|---|---|---|---|
| Percentage proficient | 74,151 | 47.897 | 20.166 | 0.000 | 33.086 | 46.801 | 62.418 | 100.000 |
| Distance to nearest library (miles) | 74,151 | 1.490 | 2.058 | 0.0001 | 0.604 | 1.034 | 1.696 | 49.455 |
| Expenditures per student (100s of dollars) | 74,151 | 74.434 | 12.287 | 3.454 | 66.892 | 73.400 | 80.320 | 218.519 |
| FFRPM | 74,151 | 0.537 | 0.298 | 0.000 | 0.274 | 0.568 | 0.806 | 1.000 |
| Fraction female | 74,151 | 0.486 | 0.032 | 0.000 | 0.471 | 0.487 | 0.502 | 1.000 |
| Fraction White | 74,151 | 0.310 | 0.265 | 0.000 | 0.064 | 0.243 | 0.534 | 1.000 |
| Fraction Hispanic | 74,151 | 0.472 | 0.295 | 0.000 | 0.202 | 0.445 | 0.735 | 1.000 |
| Fraction African American | 74,151 | 0.068 | 0.102 | 0.000 | 0.012 | 0.031 | 0.081 | 0.992 |
| Fraction Asian | 74,151 | 0.082 | 0.130 | 0.000 | 0.010 | 0.032 | 0.092 | 1.000 |
| Fraction Native American | 74,151 | 0.009 | 0.031 | 0.000 | 0.001 | 0.004 | 0.008 | 0.981 |
| Fraction Filipino | 74,151 | 0.025 | 0.044 | 0.000 | 0.003 | 0.011 | 0.027 | 0.812 |

*Note:* School-level observations for school years 2002-2003 through 2012-2013. Schools that are introduced, closed, switch districts, or are part of multiple districts during 2003-2013 are not included. Central and branch libraries are included. Bookmobiles are not included. There are no mail-order-only libraries. Percentage proficient is the percentage of students at the school who score at a proficient or above level on the English Language Arts section of the California Standardized Test. Distance to nearest library is the straight line distance in miles to a school's nearest library. Expenditures per student are annual estimates and are provided by the CADOE at the district rather than school level. FFRPM is the fraction of students eligible for free or reduced price meals.





Figure 1: Boxplot of Percentage Proficient Over 2003-2013, School-Level Data




*Note:* Percentage proficient is the percentage of students at the school who score at a proficient or above level on the English Language Arts section of the California Standardized Test. Center hinge corresponds to median. Lower and upper hinges correspond to the first and third quartiles. Whiskers extend to 1.5 times the interquartile range from hinges. Notches roughly correspond to a 95% confidence interval for comparing medians. Points are observations of schools that are beyond 1.5 times the interquartile range from hinges.

Table 3: OLS Regression of Percentage of Students Proficient on Continuous Distance

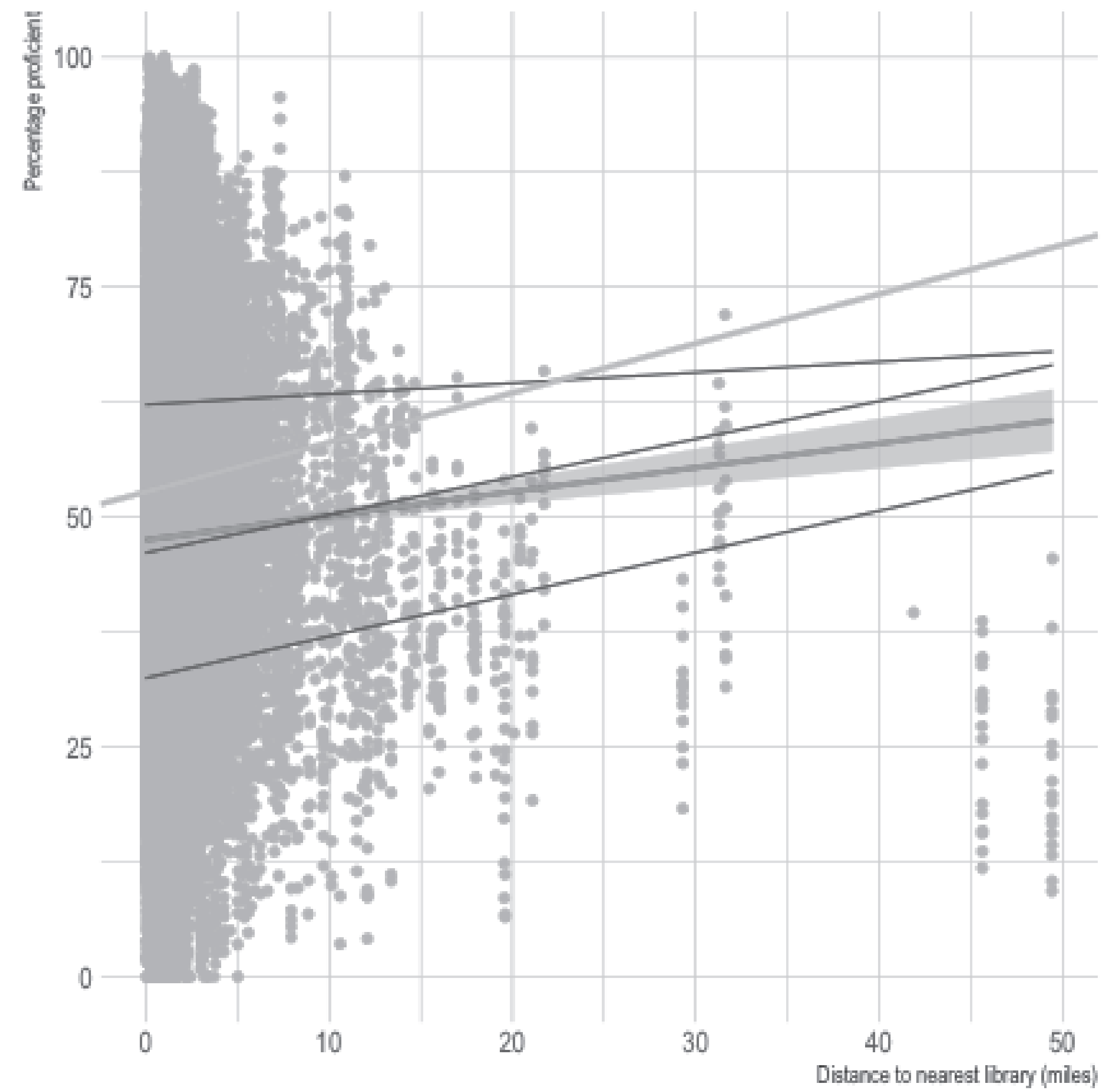
| | Percentage Proficient | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| *DIST* (miles) | 0.263*** | 2.722*** | 0.136* | −0.062 | 0.538*** |
| | (0.036) | (0.095) | (0.070) | (0.069) | (0.068) |
| $DIST^2$ | | −0.233*** | −0.019*** | −0.026*** | −0.083*** |
| | | (0.010) | (0.007) | (0.007) | (0.007) |
| $DIST^3$ | | 0.004*** | 0.0002 | 0.0004*** | 0.001*** |
| | | (0.0002) | (0.0001) | (0.0001) | (0.0001) |
| FFRPM | | | −46.423*** | −34.575*** | −36.499*** |
| | | | (0.182) | (0.284) | (0.277) |
| Fraction Hispanic | | | | −15.502*** | −14.530*** |
| | | | | (0.288) | (0.280) |
| Expenditures per student (100s of dollars) | | | | | 0.285*** |
| | | | | | (0.004) |
| Constant | 47.506*** | 44.974*** | 72.734*** | 74.004*** | 52.742*** |
| | (0.091) | (0.127) | (0.143) | (0.142) | (0.346) |
| Observations | 74,151 | 74,151 | 74,151 | 74,151 | 74,151 |
| $R^2$ | 0.001 | 0.012 | 0.473 | 0.493 | 0.522 |

*Note:* DIST is the distance to the nearest library in miles. FFRPM is the fraction of students eligible for free or reduced price meals. Standard errors in parentheses. Blank means variable not entered. *p<0.1; **p<0.05; ***p<0.01





Figure 2: OLS Regression of Percentage of Students Proficient on Continuous Distance




*Note:* Gray points are school x year observations. The red line represents regression 2 in Table 2, with a shaded 95% confidence interval. The blue lines represent regression 2, but as 0.25, 0.5, and 0.75 quantile regressions, rather than OLS. The orange line represents regression 5.

Table 4: OLS Regression of Percentage of Students Proficient on Distance Indicators

| | Percentage Proficient | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| $D_{2.5}$ | 3.762*** | 2.622*** | 5.689*** | 9.301*** |
| | (0.784) | (0.570) | (0.562) | (0.549) |
| $D_5$ | 8.144*** | 2.021*** | 4.399*** | 9.423*** |
| | (0.819) | (0.596) | (0.586) | (0.575) |
| $D_{7.5}$ | 4.127*** | 1.912*** | 3.660*** | 7.754*** |
| | (0.948) | (0.689) | (0.677) | (0.661) |
| $D_{10}$ | 4.320*** | 2.070** | 2.889*** | 5.616*** |
| | (1.217) | (0.885) | (0.868) | (0.845) |
| FFRPM | | −46.535*** | −34.558*** | −36.468*** |
| | | (0.181) | (0.283) | (0.277) |
| Hispanic | | | −15.626*** | −14.764*** |
| | | | (0.288) | (0.280) |
| Expenditures per student (100s of dollars) | | | | 0.282*** |
| | | | | (0.004) |
| Constant | 43.775*** | 70.367*** | 68.370*** | 44.318*** |
| | (0.780) | (0.576) | (0.567) | (0.660) |
| Observations | 74,151 | 74,151 | 74,151 | 74,151 |
| $R^2$ | 0.004 | 0.473 | 0.493 | 0.522 |

*Note:* $D_{2.5}$ is 1 if a school is between 0 to 2.5 miles of a library, 0 otherwise. $D_5$ is 1 if a school is between 2.5 to 5 miles of a library, 0 otherwise. $D_{7.5}$ is 1 if a school is between 5 to 7.5 miles of a library, 0 otherwise. $D_{10}$ is 1 if a school is between 7.5 to 10 miles of a library, 0 otherwise. FFRPM is the fraction of students eligible for free or reduced price meals. Standard errors in parentheses. Blank means variable not entered. *p<0.1; **p<0.05; ***p<0.01



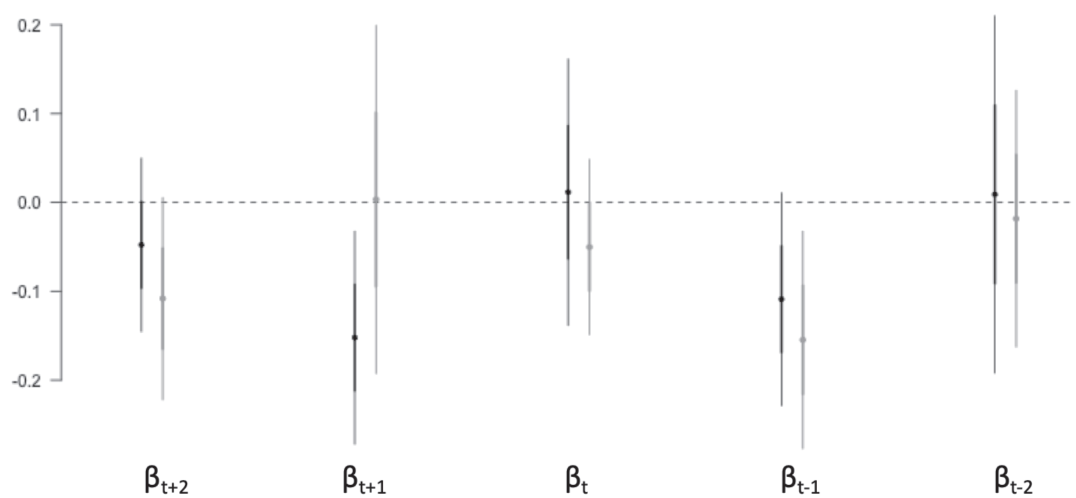


Table 5: FE Regression of Percentage of Students Proficient on Continuous Distance

| | Percentage Proficient | | | | |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| $DIST_t$ (miles) | −0.180*** | −0.048 | −0.055 | −0.052 | −0.050 |
| | (0.048) | (0.049) | (0.049) | (0.049) | (0.049) |
| $DIST_{t-1}$ | | −0.152** | −0.150** | −0.155** | −0.155** |
| | | (0.060) | (0.059) | (0.061) | (0.061) |
| $DIST_{t-2}$ | | 0.012 | −0.008 | −0.017 | −0.018 |
| | | (0.075) | (0.072) | (0.072) | (0.072) |
| $DIST_{t+1}$ | | −0.109* | −0.113* | −0.109* | −0.108* |
| | | (0.060) | (0.060) | (0.057) | (0.057) |
| $DIST_{t+2}$ | | 0.009 | 0.013 | 0.003 | 0.003 |
| | | (0.100) | (0.099) | (0.098) | (0.098) |
| FFRPM | | | −6.353*** | −4.788*** | −4.790*** |
| | | | (0.494) | (0.443) | (0.442) |
| Fraction Hispanic | | | | −14.337*** | −14.313*** |
| | | | | (1.129) | (1.130) |
| Expenditures per student (100s of dollars) | | | | | 0.011 |
| | | | | | (0.010) |
| Observations | 44,799 | 44,799 | 44,799 | 44,799 | 44,799 |
| Adjusted $R^2$ | −0.175 | −0.175 | −0.162 | −0.143 | −0.143 |

*Note:* $DIST_t$ is the distance to the nearest library in miles at time t. Heteroskedasticity-robust clustered within standard errors in parentheses. FFRPM is the fraction of students eligible for free or reduced price meals. Blank means variable not entered. The sample size is smaller than in OLS due to the dropping of observations with missing leads or lags. *p<0.1; **p<0.05; ***p<0.01

Figure 3: FE Regression of Percentage of Student Proficient on Continuous Distance




*Note:* Black corresponds to estimates without controls. Red corresponds to estimates with all covariates added. Percentage proficient is the percentage of students at the school who score at a proficient or above level on the English Language Arts section of the California Standardized Test. Whiskers represent 95% confidence intervals with heteroskedasticity-robust clustered standard errors.



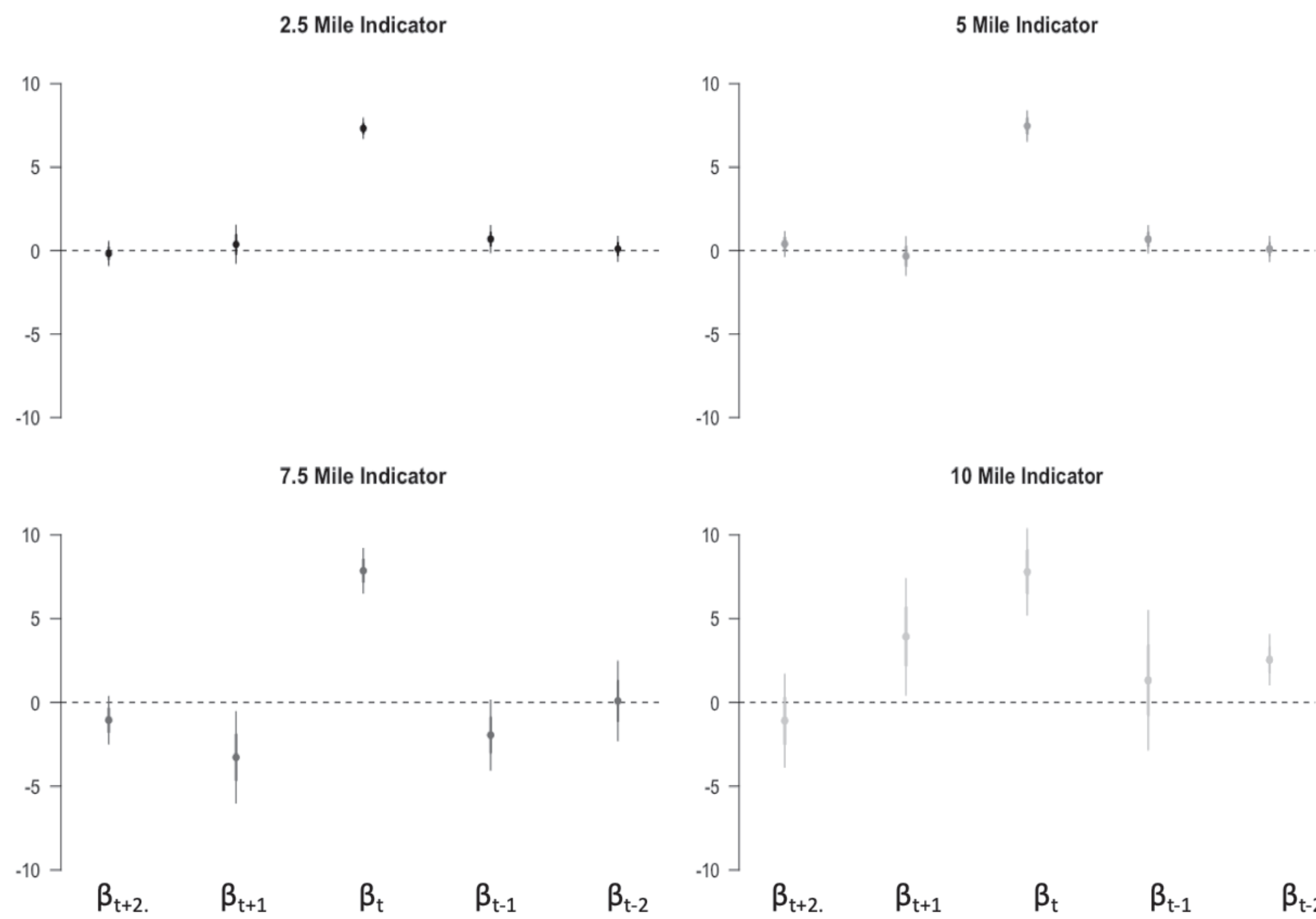


Table 6: FE Regression of Percentage of Students Proficient on Distance Indicators

| | Percentage Proficient | | | |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| $D_{2.5}$ | 7.363*** | 7.519*** | 7.397*** | 7.405*** |
| | (0.065) | (0.066) | (0.065) | (0.066) |
| $D_5$ | 7.301*** | 7.384*** | 7.286*** | 7.293*** |
| | (0.508) | (0.505) | (0.494) | (0.494) |
| $D_{7.5}$ | 5.266*** | 5.489*** | 5.236*** | 5.237*** |
| | (1.194) | (1.151) | (1.153) | (1.156) |
| $D_{10}$ | 8.718*** | 8.849*** | 8.437*** | 8.501*** |
| | (1.190) | (1.176) | (1.173) | (1.167) |
| FFRPM | | −6.342*** | −4.776*** | −4.777*** |
| | | (0.493) | (0.442) | (0.441) |
| Hispanic | | | −14.338*** | −14.313*** |
| | | | (1.129) | (1.130) |
| Expenditure per student (100s of dollars) | | | | 0.012 |
| | | | | (0.010) |
| Observations | 44,799 | 44,799 | 44,799 | 44,799 |
| Adjusted $R^2$ | −0.175 | −0.162 | −0.143 | −0.143 |

*Note:* $D_{2.5}$ is 1 if a school is between 0 to 2.5 miles of a library, 0 otherwise. $D_5$ is 1 if a school is between 2.5 to 5 miles of a library, 0 otherwise. $D_{7.5}$ is 1 if a school is between 5 to 7.5 miles of a library, 0 otherwise. $D_{10}$ is 1 if a school is between 7.5 to 10 miles of a library, 0 otherwise. FFRPM is the fraction of students eligible for free or reduced price meals. School and year fixed effects are included. Heteroskedasticity-robust clustered within standard errors in parentheses. Blank means variable not entered. *p<0.1; **p<0.05; ***p<0.01

Figure 4: FE Regression of Percentage of Student Proficient on Distance Indicators with Leads and Lags




*Note:* Results are with all covariates added. Percentage proficient is the percentage of students at the school who score at a proficient or above level on the English Language Arts section of the California Standardized Test. Whiskers represent 95% confidence intervals with heteroskedasticity-robust clustered standard errors.