Essays in Public and Environmental Economics

by

Mary Angela Quiroga

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Public Policy and Economics)
in The University of Michigan
2026

Doctoral Committee:

Professor Catherine Hausman, Co-Chair
Professor James Hines, Co-Chair
Professor Kevin Stange
Professor Mel Stephens

Mary Angela Quiroga

mquiroga@umich.edu

ORCID iD: 0009-0003-3694-1179

© Mary Angela Quiroga 2026

DEDICATION

This work is dedicated to Mary Alba Velasquez Kellogg, the woman who brought me into this world and to this country; to José David Quiroga, who I have been emulating for 31 years, going as far as starting a Ph.D. (though I finished mine first. Ha.); and to Will Nober, my partner through it all. Sí se pudo. Los quiero mucho.

ACKNOWLEDGEMENTS

To my committee: thank you for your support and mentorship. This dissertation would not be the same without you. Jim, I enjoyed our long chats about research, the state of the world, and many other seemingly random topics in your luxurious law school office. I'll never forget that good research is nimble. Catie, your white board could have been my fifth committee member. You are a real inspiration to me as a researcher and as a mentor. Kevin, I could always count on your good humor and positivity to carry me through. Mel, your thoughtful critiques and perspective shaped my papers and I truly appreciate how generous you are with your time.

I am grateful, also, to the larger Michigan community–particularly the labor, education, and PF researchers whose feedback has been enormously helpful. Kathy, Sue, Stephanie, Elizabeth, CJ, Katie: HAIL to the experimental variation! Thanks for bringing me on.

To my husband, Will, my dear friends in Michigan and beyond, and my family in Nevada* and Washington, D.C.: thank you for your love and support. I have sure needed those in spades over the last six years.

Many other people were, of course, crucial to my completion of a doctoral degree. Thank you to Lisa Cook, Domenico Giannone, and Todd Sorensen for writing my graduate school letters of recommendation. Thank you to my mentors and peers at the AEA Summer Program, McNair Scholars, Trio Scholars, and Upward Bound–particularly Monica Conlan, Ellen Houston, Jen Sims, Sandra Jimenez, and Chris Westin: thank you. There was a time I thought I might go to community college or skip college altogether and maybe I would have if it wasn't for the guide rails I got from my UB counselors. Thanks to my 9th grade English teacher, Susan Van Doren, for insisting I apply to Upward Bound in the first place.

Now, for the required disclaimers:

This material is based upon work supported by the National Science Foundation Grad-

*It's pronounced Ne-VAD-uh not Ne-VAH-da. Looking at you, east coasters.

uate Research Fellowship under Grant No. DGE-1841052. Any opinions, findings, and conclusions or recommendations expressed in this material are my own and do not necessarily reflect the views of the National Science Foundation. The views expressed are my own and do not necessarily reflect the positions or policies of the North Carolina Department of Public Instruction. The second chapter is registered as RCT ID AEARCTR-0001831, with 10.1257/rct.1831-4.0. A pre-analysis plan was initially filed in April 2017. This research used data structured and maintained by the MERI-Michigan Education Data Center (MEDC). MEDC data is modified for analysis purposes using rules governed by MEDC and are not identical to those data collected and maintained by the Michigan Department of Education (MDE) and/or Michigan's Center for Educational Performance and Information (CEPI). The research was supported by IES grant R305B150012 and the Smith Richardson Foundation. This research does not represent the views of IES or the US Department of Education.

TABLE OF CONTENTS

DEDICATION ii

ACKNOWLEDGEMENTS iii

LIST OF FIGURES ix

LIST OF TABLES xi

LIST OF APPENDICES xiv

ABSTRACT xv

CHAPTER

1 The Effects of Heat on Teachers and Their Students 1

1.1 Introduction 1

1.2 Conceptual Framework 5

1.3 Data and Empirical Strategy 6

1.3.1 Sample Selection 6

1.3.2 Weather Data and Extreme Heat Measure 7

1.3.3 Summary Statistics 9

1.3.4 Empirical Strategy 14

1.3.5 Comparing Older and Younger Teachers 15

1.4 Heat's Impacts on Teachers and Their Students 18

1.4.1 Test Scores and Teacher Age . . . 18
1.4.2 Teacher Absences . . . 22
1.4.3 Student Absences and Disciplinary Referrals . . . 25
1.4.4 Exploring the Returns to Age in the Presence of Extreme Heat . . . 27
1.4.5 Climate Control in Classrooms . . . 29
1.4.6 Climate Change . . . 29
1.5 Conclusion . . . 30

2 Degrees of Mobility: Shifting Low-Income Students to Higher Quality Colleges 32

2.1 Introduction . . . 32
2.2 Prior Literature . . . 35
2.2.1 What Is College Quality? . . . 35
2.2.2 Evidence on the Causal Effect of College Quality . . . 36
2.2.3 Contribution . . . 38
2.3 The Intervention and Randomization . . . 38
2.3.1 The HAIL Scholarship . . . 38
2.3.2 Randomization . . . 41
2.4 Data and Methods . . . 41
2.4.1 Data . . . 41
2.4.2 Empirical Strategy . . . 43
2.5 The Landscape of College Quality in Michigan . . . 44
2.6 Intent-to-Treat Effects . . . 45
2.6.1 Enrollment at a University of Michigan Campus . . . 45
2.6.2 Time to Degree and BA Selectivity . . . 46
2.6.3 Faster Credit Accumulation . . . 47
2.6.4 Reduced Attendance at Community Colleges and AA Degree Attainment . . . 48

2.6.5 Summary of Intent-to-Treat Effects 49
2.7 Whose College Choices Does a Tuition Guarantee Change? 50
2.8 Local Average Treatment Effects 51
2.8.1 The Quality Shock for Compliers 51
2.8.2 Educational Attainment and Degree Quality 52
2.8.3 The Satellite Pathway 53
2.8.4 Projected Earnings Effects 54
2.8.5 Plausibility of the Exclusion Restriction 55
2.9 Discussion 56
2.10 Conclusion 58
3 Gas Prices and the Safety Net 70
3.1 Introduction 70
3.2 Data and Empirical Strategy 72
3.2.1 Grocery price data 76
3.2.2 Empirical strategy 80
3.3 Results 82
3.4 Conclusion 86
APPENDICES 87
A.1 Teacher Age Bins Uncorrelated with Observables 88
A.2 Test Score and Teacher Age Using Alternate Heat Measures 89
A.3 Using Alternate Definitions of Teacher Age 94
A.4 Heterogeneity 96
A.5 Movers Analysis 101
B.1 Outcomes and Measurement 111

B.1.1 Enrollment 111
B.1.2 Degree Attainment 113
B.1.3 STEM Degrees 113
B.1.4 Projected Earnings 114
B.1.5 Graduate School Enrollment and Degree Attainment 114
B.1.6 Institution Characteristics 115
B.1.7 Credits 116
B.1.8 Randomization Inference 118
B.1.9 Complier Analysis 119
C.1 Robustness checks 123
BIBLIOGRAPHY 125

LIST OF FIGURES

FIGURE

1.1 Elementary Schools in North Carolina 8

1.2 Extreme Heat Day Temperature Cutoffs 10

1.3 Variation in Number of Extreme Heat Days 13

1.4 Percentage of older teachers 16

1.5 Extreme Heat and Scores Conditional on Teacher Age 21

1.6 Older teachers' attendance responds more in extreme heat 24

1.7 Test Scores by Teacher Age Above and Below Average Heat 28

2.1 Distribution of Characteristics of Colleges Attended by High-Achieving Students in Michigan, by Family Income 64

2.2 Estimated Effect of HAIL on First-Year College Enrollment 65

2.3 Estimated Effect of HAIL on Credit Accumulation at 4-year Institutions 66

2.4 Estimated Effect of HAIL on Degree Attainment 67

2.5 Average SAT Scores Relative to the Average SAT Score of Institutional Peers . . 68

2.6 Selectivity of Bachelor's Degrees Earned by High-Achieving Michigan Students, by Income and HAIL Treatment Status 69

3.1 Map of PADDs 73

3.2 SNAP participation per capita by state, 2001-2022 74

3.3 Gasoline prices, 2001-2022 75

3.4 Tornqvist index for full sample, 1999-2010 78

3.5 Laspeyres index for full sample, 1999-2010 79

3.6 Naive index for full sample, 1999-2010 79

A.1 Older teachers' quality suffers in heat above 90°F and below 30°F 91

A.2 No evidence of sorting by heat and teacher age 104

B.1 Effect of HAIL on STEM BA Degree Attainment Using Alternative STEM Definitions 111

B.2 Missingness by Treatment Status and Academic Term 112

B.3 Lee Bounds on Earned Four-Year Credits, by Academic Term: ITTs 112

C.1 Coefficient Stability Using Oster Method 123

LIST OF TABLES

TABLE

1.1 Summary Statistics 11

1.2 Percent of Absences By Description and Age 12

1.3 Difference Between Older and Younger Teachers 17

1.4 Regression of Test Scores on Extreme Heat Days and Teacher Age 21

1.5 Older teachers miss school more often during months with extreme heat . . . 23

1.6 No evidence of an effect of extreme heat on yearly absences 25

1.7 No evidence of extreme heat effect on disciplinary referrals 26

2.1 High-School-Level Summary Statistics and Treatment-Control Balance 60

2.2 Estimated Effects on Degree Attainment and Projected Earnings 61

2.3 Pre-Treatment Characteristics by Compliance Sub-Populations 62

2.4 Estimated Effects on Characteristics of Colleges Attended (College-Attending Students Only) 63

3.1 Summary Statistics 76

3.2 Grocery categories with largest expenditures 78

3.3 OLS Regressions Using Imputed Gasoline Prices 83

3.4 IV Regressions Using Imputed Gasoline Prices 84

3.5 First Stage Regressions 84

3.6 Regressions of SNAP on gasoline prices and grocery inflation 85

A.1 Regression of Teacher Age Bin Indicator on Observables 88

A.2 Older teachers' quality suffers in heat above 90°F and below 30°F 90
A.3 Younger teachers' students perform relatively better than older ones during hotter years 92
A.4 Findings robust to alternate definition of extreme heat using 85th Percentile . . 93
A.5 Findings robust to alternate definition of extreme heat using 95th Percentile . . 93
A.6 Similar Findings Using Age 50 Cutoff (85th Percentile of Teacher Age) 94
A.7 Similar Findings Using Age 59 Cutoff (95th Percentile of Teacher Age) 95
A.8 Similar Findings Excluding Youngest Teachers 95
A.9 Heterogeneity By Test Subject 96
A.10 Heterogeneity by Grade Level 97
A.11 Heterogeneity by Semester of Extreme Heat Days 97
A.12 Heterogeneity By Student Gender 98
A.13 Heterogeneity By Rurality 98
A.14 Heterogeneity By Student Race 99
A.15 Heterogeneity By County Median Income 100
A.16 Summary Table For Teachers Who Move 102
A.17 Differences Between Older and Younger Teachers Who Move 103

B.1 Summary of Literature on Effects of College Quality 106
B.2 Estimated Effect of HAIL on Financial Aid Receipt at the UM Flagship Adapted from Burland et al. (2023), p. 304 107
B.3 Estimated Effect of HAIL on Transfer Pathways 108
B.4 Estimated Effect of HAIL on Graduate School Outcomes 109
B.5 Estimated Effect of HAIL on Projected Earnings 8 years Following College Entry, using Alternate Populations to Project Earnings 110
B.6 Effect of HAIL on Selection into the Transcript Data Sample 117
B.7 Comparison of Sampling- and Randomization-Based Inference 119

C.1 Regressions of SNAP on gasoline prices using alternative lag structures 124

LIST OF APPENDICES

A Appendix to Chapter 1 . . . . . **87**

B Appendix to Chapter 2 . . . . . **105**

C Appendix to Chapter 3 . . . . . **122**

ABSTRACT

A central theme of my work is understanding how adverse economic and environmental conditions shape human capital formation and household well-being, as well as the role of policies that may mitigate these effects. In my dissertation, I exploit natural experiments, a randomized control trial, and administrative panel data to investigate both the direct consequences of adverse conditions, such as extreme heat and poverty, and the use of policies that can buffer their impacts. Though each chapter focuses on a different topic, all are concerned with understanding pressing policy issues.

The first chapter of my dissertation provides the first evidence that extreme heat directly affects teachers, with consequences for student learning. Using matched student–teacher administrative data from North Carolina linked to high-resolution weather records, the analysis shows that older teachers, who are more sensitive to heat stress, experience greater productivity losses and higher absence rates during hot years. Students of older teachers see test score declines with each additional extreme heat day, effectively erasing the benefits of teaching experience. Reversing heat-related test score losses would take a budget increase of $850 per pupil sustained over four years.

Earnings inequality among college graduates has grown, and returns to a bachelor's degree (BA) vary substantially across colleges. We experimentally evaluate a program that shifted high-achieving, low-income students into higher-quality colleges. Students randomly offered a guarantee of four years of tuition and fees at the University of Michigan were 21 percentage points more likely to enroll at a UM campus. The offer shifted students across colleges rather than into college: most compliers would otherwise have attended less-selective four-year colleges. HAIL increased four-year BA completion by 4 percentage points, but had no effect on attainment after six years. It substantially increased degree quality: students offered the scholarship were 12 percentage points more likely to earn a bachelor's degree from a highly or most selective college. IV estimates indicate that students induced to attend a University of Michigan campus were 56 percentage points more likely to earn such a degree. The increase in college quality predicts earnings gains of 18 to 25 percent. Faster degree completion adds about another 1 percent to the present

discounted value of lifetime earnings.

Together, these three chapters assess the effects of policy interventions and adverse shocks on students, teachers, and low-income households in the US context. These essays examine the unexpected costs of a changing climate, raise questions about the design of social safety net programs, and show that successful educational equity programs can come at little to no additional cost to institutions due to changes in financial aid packages.

CHAPTER 1

The Effects of Heat on Teachers and Their Students

Abstract

This paper provides the first evidence that extreme heat directly affects teachers, with consequences for student learning. Using matched student–teacher administrative data from North Carolina linked to high-resolution weather records, the analysis shows that older teachers, who are more sensitive to heat stress, experience greater productivity losses and higher absence rates during hot years. Students of older teachers see test score declines with each additional extreme heat day, effectively erasing the benefits of teaching experience. Reversing heat-related test score losses would take a budget increase of $850 per pupil sustained over four years.

1.1 Introduction

As climate change leads to more extreme weather, research suggests we will see increased crime rates, lower labor supply, and, especially among older adults, increased illness and death (Luber and McGeehin, 2008; Dell et al., 2014; Graff Zivin and Neidell, 2014; Baysan et al., 2019; Somanathan et al., 2021; Cohen and Gonzalez, 2024; Costa et al., 2024). Extreme temperatures hinder human capital formation and worker productivity in a range of settings, from manufacturing in India to education in the United States (Park et al., 2020b,a; Somanathan et al., 2021). Understanding these relationships in elementary schools is particularly important given the returns to early childhood education reported extensively in the literature–returns that may be diminished by exposure to extreme heat (Angrist and Krueger, 1991; Card, 1999; Heckman et al., 2010; Hendren and Sprung-Keyser, 2020, and many others).

Human capital development in schools depends on students and teachers, both of whom are susceptible to harmful environmental factors. While previous research shows that extreme temperatures reduce student performance on standardized tests, the mechanisms driving these effects remain underexplored in the economics literature (Park et al., 2020b; Roach and Whitney, 2022). From the literature on environmental stressors' effects on workers, it is reasonable to expect negative impacts on teachers' instructional quality. However, it can be difficult to disentangle effects on teachers separately from students given that the two groups experience heat exposure simultaneously. By taking advantage of differences in teachers' sensitivity to heat by age, this paper documents the teacher mechanism underlying heat's effects on student academic performance.

In this paper, I provide the first evidence of heat effects on teachers by exploiting variation in temperature and heat sensitivity to examine how extreme heat exposure affects teacher productivity and behavior. The analysis combines matched student-teacher administrative data from the North Carolina Education Research Data Center with a balanced panel of interpolated weather data from the PRISM climatology group (Schlenker, 2024). This analysis leverages physiological and behavioral differences between younger and older teachers. As humans age, we become more sensitive to heat-related stress and illness (Kenney, 1988). Additionally, older workers who are established in their careers may have more elastic responses to extreme heat due to the advantages of tenure, like extra vacation days. Using an interaction between cumulative heat shocks and teacher age, I compare the outcomes of students taught by older teachers, who are likely more negatively impacted by extreme heat, to those taught by less-impacted younger teachers. To examine a facet of teacher behavior with important spillovers to student learning, I investigate heat's impacts on teacher absences.

The analysis in this paper exploits cumulative weather shocks and examines the interaction of these shocks and teacher age to quantify the effects of a hotter learning environment on teachers and their students. To focus on yearlong learning rather than testing-day conditions, I use weather measures that span the entire school year and end-of-year standardized exam scores. The main independent variable, a measure of extreme heat, is defined as the number of days in a school year with temperatures above the 90th percentile, based on the empirical distribution of temperature at each location from 1997 to 2007, the decade prior to the sample period. This measure of extreme heat takes into account place-specific adaptation to heat–a factor that other heat measures, like number of days in a temperature range, cannot address. I use a two-way fixed effects model, regressing student test scores on this extreme heat measure and the interaction of extreme heat

and an indicator for advanced teacher age.

For causal identification, I assume that conditional on fixed effects and control variables, variation in extreme heat is exogenous. This assumption relies on the randomness in weather after accounting for location and can only be violated if individuals could predict local weather shocks. Additionally, to identify the difference between older and younger teachers, I assume that the interaction of heat and teacher age is conditionally exogenous. Using balance tables and a regression framework, I argue that this is a reasonable assumption. Moreover, the findings in this paper are robust to varying the definition of extreme heat and the method of classifying teacher age.

This paper presents two main findings showing heat disproportionately affects older teachers' productivity and attendance. First, I find that students taught by older teachers score 0.002 standard deviations lower on end-of-year standardized exams for each additional extreme heat day—equivalent to losing 22% of a school day. Putting this effect into dollar terms, it would take a school district an increase in spending of about $850 per pupil sustained over four years to undo the decrease in test scores due to heat over those years (on average, assuming an average number of extreme heat days each year) (Jackson and Mackevicius, 2024).[1]

Meanwhile, students of younger teachers are essentially unaffected. On average, students of older teachers score higher than those of younger teachers, a difference likely attributable to experience — but this advantage diminishes as the number of extreme heat days increases. The estimates imply that ten more extreme heat days would completely erase the benefits of having an older, more experienced teacher. These findings support those of previous studies on returns to experience for North Carolina teachers (Wiswall, 2013; Ladd and Sorensen, 2017). Secondly, I find that older teachers are twice as likely to be absent due to each additional extreme heat day, suggesting a direct behavioral response to heat exposure.

The current research on the effects of heat on student academic performance does not explore teacher effects, potentially due to the interconnectedness of student and teacher effects. This analysis uses heterogeneity in heat effects by teacher characteristics to elucidate the effects of heat on teachers. Since teacher productivity is often inferred from student test scores, isolating the effects on teachers themselves, independent of their stu-

[1]To calculate this dollar effect, I use a finding from Jackson and Mackevicius (2024)'s meta-analysis of the returns to school spending policies. I assume that each year has the average number of extreme heat days, about 3.3.

dents, can be difficult. Separating teacher and student effects is critical for policymakers to target mitigation strategies. I address this difficulty by proxying for teachers' heat sensitivity using teacher age and by analyzing teacher absenteeism. Additionally, this type of analysis requires matched student-teacher data with sufficient variation in temperature for credible identification. Along with rich longitudinal data, the state of North Carolina offers geographic variation missing in studies that focus on a single school district.

By analyzing the relationship between heat, learning outcomes, teachers' ages, and absences, this study aims to shed light on a mechanism underlying heat's effects. This research contributes to the growing body of work at the intersection of climate and education by highlighting an understudied mechanism, instructional quality, and how it varies with environmental conditions. Previous research indicates that students across the United States perform worse on standardized exams following periods of higher temperatures (Park et al., 2020b; Roach and Whitney, 2022). Extreme temperatures also influence students' disciplinary referrals and attendance, which likely further affect test scores (McCormack, 2023). However, these studies do not consider teachers or how their characteristics and behaviors might influence the relationship between heat and learning.

By analyzing teacher absences, this paper addresses a key yet understudied factor of worker productivity in education, which plays a crucial role in the development of human capital. Recent studies show that adverse environmental conditions such as heat and pollution negatively impact worker productivity across various sectors, including outdoor industries like construction and agriculture, as well as indoor settings such as manufacturing and call centers. Some of these productivity losses can be attributed to increased worker absences (Graff Zivin and Neidell, 2014; Chang et al., 2019; Somanathan et al., 2021; Casey et al., 2024; Costa et al., 2024). In manufacturing, for example, the impact of heat on absenteeism persists even in workplaces with climate control, a feature likely present in many classrooms in my study context (Somanathan et al., 2021).

While heat affects productivity and learning in many contexts, older individuals are especially susceptible to heat-related illness, mortality, and productivity loss. Older, more experienced teachers face a different set of career incentives compared to younger teachers and thus may be more likely to reduce productivity and labor supply in extreme heat. Further, the medical and public health literature consistently identifies advanced age as a major risk factor for heat-related illness (Bouchama and Knochel, 2002; Luber and McGeehin, 2008). Han et al. (2024) presents a meta-analysis of recent studies on heat and productivity among construction workers, finding that middle-aged and older workers are

particularly affected. My work bridges the public health and economics literature by examining the effects of heat on older workers in a new context. This work also speaks to the teacher productivity and absenteeism literature by highlighting how heat affects teacher absenteeism (Herrmann and Rockoff, 2012; Benhenda, 2022).

This works aims to inform evaluations of student and teacher performance, guide infrastructure investment decisions, and support policies related to school closures during heat waves. In doing so, this work helps address inequities between students attending heat-vulnerable schools and those in better-equipped environments. These considerations are increasingly urgent in light of climate change, which is likely to exacerbate disparities between under-resourced schools with aging infrastructure and well-resourced schools with modern facilities.

1.2 Conceptual Framework

In this section, I present a brief conceptual framework describing how teachers and students contribute to educational outcomes. The education production function in this framework depends on two forms of labor inputs: student inputs, S, teacher inputs, T, and total factor productivity, A. Student and teacher labor inputs have an output elasticity of α and $1-\alpha$, respectively. Student i with teacher type $j \in \{Y, O\}$ has output E_{ij} where

$$E_{ij} = AS_i^{\alpha} T_j^{1-\alpha} \tag{1.1}$$

Teacher inputs vary by teacher age in two age categories, younger, T_Y, and older, T_O. All inputs may be affected by heat. Heat, h, negatively impacts all teacher inputs, but may have a larger negative effect on older teacher inputs due to the physiological effects of heat as humans age and shifting career incentives. Written mathematically,

$$0 > \frac{\partial T_Y}{\partial h} > \frac{\partial T_O}{\partial h} \tag{1.2}$$

For ease of calculation, I take the natural logarithm of the production function. Taking the log of the production function and then finding the derivative with respect to h,

$$\frac{\partial \log E_{ij}}{\partial h} = \frac{\partial \log A}{\partial h}\frac{1}{A} + \frac{\partial \log S_i}{\partial h}\frac{\alpha}{S_i} + \frac{\partial \log T_j}{\partial h}\frac{1-\alpha}{T_j} \tag{1.3}$$

Combining Equations 1.2 and 1.3 implies that test scores decrease more for students of older teachers compared to students of younger teachers.

$$0 > \frac{\partial E_{iY}}{\partial h} > \frac{\partial E_{iO}}{\partial h} \tag{1.4}$$

However, it is clear from the literature on the returns to experience for teachers in North Carolina that older, and thus, more experienced teachers have more successful students (Wiswall, 2013; Ladd and Sorensen, 2017). These returns to experience may mean that older teachers are better able to make up for the detrimental effects of heat shocks. This would imply the following:

$$0 > \frac{\partial T_O}{\partial h} > \frac{\partial T_Y}{\partial h} > \Rightarrow 0 > \frac{\partial E_{iO}}{\partial h} > \frac{\partial E_{iY}}{\partial h} \tag{1.5}$$

Given this framework, it is unclear whether heat will have greater negative effects on students of older teachers compared to students of younger teachers. However, this is testable given data on heat, test scores, and teacher age.

1.3 Data and Empirical Strategy

I combine two main sources of data for this study: administrative records from the North Carolina Department of Public Instruction (with access provided by the North Carolina Education Research Data Center (NCERDC)) and a balanced panel of daily weather data from the PRISM climatology group (Schlenker, 2024).[2]

1.3.1 Sample Selection

The NCERDC compiles annual datasets tracking students and teachers across North Carolina's public schools using unique identifiers. Using these identifiers, I construct a panel of annual test scores, teacher assignments, and student demographics for all third through fifth grade students from 2008 to 2018.[3][4] After fifth grade, students typically rotate among

[2]Interpolated daily weather data is available at https://zenodo.org/records/10625288.

[3]Years refer to the spring of the academic year; for example, 2008 refers to the 2007–08 school year.

[4]Federal law, including the Every Student Succeeds Act and the preceding No Child Left Behind Act, requires states to administer annual standardized exams to students in grades three through eight. In North Carolina, these exams are conducted during the last ten days of each school year.

multiple classrooms rather than staying with one teacher all day, which reduces their exposure to any single teacher. For this reason, I focus on students in these grades. Within this student sample, I identify teachers who are assigned to third through fifth graders either for full-day, self-contained classrooms or for reading or math classes. I then match each student-teacher pair with the student's standardized test scores, total annual absences, and disciplinary incidents. The dataset does not directly include teacher age, so I calculate age based on the teacher's undergraduate graduation date, assuming individuals graduate at age 24. This estimate is based on an assumed six-year college trajectory beginning at age 18.

Following the sample restrictions I impose, the analysis sample remains large, with approximately 570,000 students and 22,000 teachers. While this sample generally spans 2008 to 2018, detailed records of teacher absences are only available to me for the 2006–2008 period. As a result, I am unable to link the teacher absence records to the larger, matched student-teacher sample used for the main analysis. However, these earlier data offer the advantage of monthly reporting frequency.

1.3.2 Weather Data and Extreme Heat Measure

To assign weather conditions to schools as accurately as possible, I use interpolated weather data rather than relying solely on information from weather stations. Although North Carolina has many weather stations, schools are not always located near one—the average distance from a school to the nearest station in this sample is 6.9 miles. The PRISM climatology group creates daily weather data on a 2.5-mile by 2.5-mile grid across the United States by interpolating between stations which is then processed into a balanced panel (Schlenker, 2024).[5] Using a balanced panel helps prevent bias that could occur if weather stations are added or removed in ways correlated with local economic conditions, which might in turn impact test scores. Each school in my dataset is matched to the weather data grid cell in which it is located. Not every PRISM grid has an elementary schools and some have multiple schools, as shown in Figure 1.1.

From these weather data, I construct a measure of extreme heat days, defining such a day as one with temperatures above the 90th percentile of the grid-level heat distribution in the decade prior to the main sample, 1997 to 2007. This approach captures the effects of extreme heat in a way that accounts for local adaptation to typical climate conditions. For

[5]Details about PRISM's modeling techniques are available at https://prism.oregonstate.edu/.

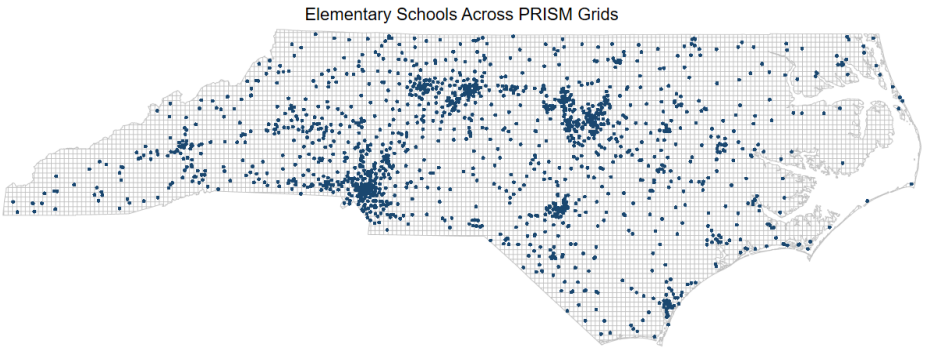


Figure 1.1: Elementary Schools in North Carolina

Note: This map of North Carolina shows the geographic distribution of elementary schools. Each dot represents an elementary school in the sample. The overlaid grid represents the 2.5 mile by 2.5 mile PRISM grid.

comparison with existing literature, I also categorize temperatures into ten-degree bins. In robustness checks, I vary the definition of an extreme heat day and use the number of days above the 85th or 95th percentile of the heat distribution.

This analysis uses extreme heat days to capture effects at the right tail of the temperature distribution, given the hot climate and the typical air-conditioning infrastructure present in North Carolina's buildings. In economics and related literature, researchers use a variety of measures of heat and temperature depending on the context and the relevant data frequency (Schlenker and Roberts, 2009; Auffhammer et al., 2013; Dell et al., 2014; Park et al., 2020a,b; McCormack, 2023; Sajid et al., 2024). Temperature measures include average maximum temperature (specified as a linear or polynomial function), cooling and heating degree days, extreme degree days, and days in temperature bins. Studies in this literature that examine the effects of cumulative heat on learning typically use average maximum temperature or days in temperature bins (Park et al., 2020b; Roach and Whitney, 2022). Importantly, their setting—the United States as a whole—includes many schools that categorically lack air conditioning, and, indeed, the findings are concentrated among those schools. However, given North Carolina's hot climate, buildings there are much more likely to have air conditioning. About 93–96% of homes in North Carolina's climate region reported having air conditioning at the beginning of my sample period, compared to 83% for the United States as a whole (U.S. Energy Information Administration, 2013). Although air-conditioning data on schools specifically are unavailable, the near ubiquity of climate control in North Carolina homes suggests a higher likelihood of climate control in schools.

The climate and infrastructure of the study setting motivate the use of extreme heat days as the main measure of heat. Because air conditioning is likely present in most North Carolina schools, average maximum temperature changes during the school year may not reflect variation in actual classroom heat exposure. By focusing the analysis on extreme heat days, I measure the impacts of particularly hot periods when air-conditioning systems may be inadequate, outdoor supervision may be particularly uncomfortable, and nighttime heat may disrupt sleep, all of which could reduce teaching quality.

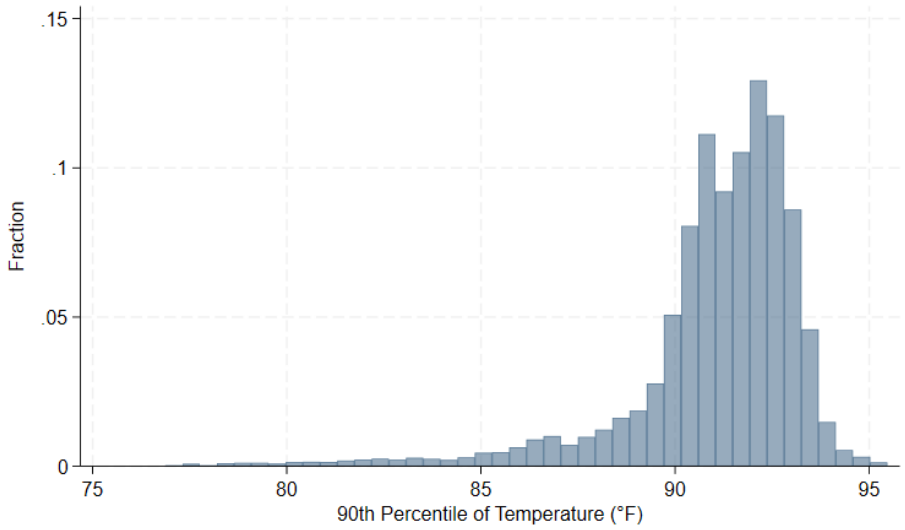
I measure extreme heat as the number of school days with daily maximum dry bulb temperature (DBT) above the local 90th percentile. Because DBT does not account for humidity, it may not fully capture physiological heat stress, which depends on both temperature and moisture. Wet bulb temperature (WBT), which incorporates humidity, is always less than or equal to DBT. Consequently, this measure may misclassify exposure: some dry days may exceed the DBT threshold without imposing meaningful heat stress, while humid days may impose substantial stress but fall below the threshold. Such misclassification is expected to attenuate the estimated effects of heat, implying that the results are conservative. WBT is not used in this analysis because it is unavailable in the interpolated, high-resolution weather data.[6]

1.3.3 Summary Statistics

The analysis sample contains more than 3.6 million individual third through fifth grade students. The sample contains a higher proportion of Black students (comprising 25% of the sample) and lower proportions of Hispanic and Asian students compared to the national average (U.S. Department of Education, National Center for Education Statistics, 2024). Table 1.1, Panel A presents summary statistics for the education dataset. On average, each class has between 28 and 29 students. For every ten students, there are approximately three disciplinary incidents per school year. In this sample, disciplinary incidents most frequently result from disruptive behavior, fighting, bus misbehavior, and aggression. Students are absent an average of 6 days per school year. The mean teacher age is 39 years old. The average school year temperature, excluding June through August, is 64.5°F (18 °C). Test scores are standardized to have a mean of zero and a standard deviation of one within each subject and year, as is standard practice in the education literature.

[6]WBT is available at the weather station level from NOAA's National Centers for Environmental Information Local Climatological Data service.

Figure 1.2: Extreme Heat Day Temperature Cutoffs



Note: This figure plots the 90th percentile (in °F) of each grid's temperature distribution. These percentiles are used to create the extreme heat days measure in this analysis: the number of days per school year above the 90th percentile. This figure shows that the majority of grids have a 90th percentile between 90 and 95°F (32-35°C).

Table 1.1, Panel B presents summary statistics for school year weather data, excluding summers. It includes weather in the months of September through April of the 2007-08 through 2017-18 school years. This panel shows that, on average, a school year has about three extreme heat days and the same number of days above 90°F (32 °C). In fact, for much of the state, the 90th percentile of maximum temperature is close to 90°F, though the western mountain region is generally cooler. Panel C includes summary statistics for the county-level unemployment rate and median income. Panel D summarizes the monthly teacher absence data with a shorter sample of 2006 through 2008.

The analysis in this paper relies on variation in the number of extreme heat days in North Carolina, net of year- and place-specific averages. Figures 1.3a through 1.3j show the variation in extreme heat used in this paper. For each year of the sample, each map shows the number of extreme heat in a grid minus the average number of extreme heat days for that year. Blues indicate that a grid had a cooler year than the state average while reds indicate that a grid had a warmer year than the state average. Focusing on a particular location and looking across years gives a sense of the variation. Squares without data indicate grid areas without elementary schools during the sample period.

Table 1.2 presents descriptions of teacher absences along with the percentage of each

type of absence by teacher age. Sick leave, the most common absence type, accounts for 54-58% of teacher absences. Annual leave and absences without deductions are the next most common types, each representing 17-22% of all absences. The remaining absences include absences with deduction or without pay, child involvement leave, and personal leave.

Table 1.1: Summary Statistics

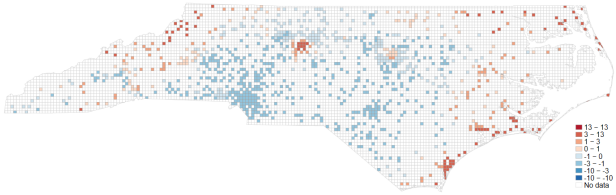
	Mean	SD
Panel A: Education Data		
Student Age	10.77	0.67
Percent Male	0.51	0.50
Percent Black	0.25	0.43
Class Size	28.87	16.75
Percent Native American	0.01	0.12
Incident Count	0.29	1.24
Days Absent	5.99	5.72
Teacher Age	38.87	10.14
Panel B: Weather Data (Excludes Summers)		
Max Temp °F	64.51	2.70
Extreme Heat Days	3.35	3.58
Days in 90s °F	3.30	3.95
Days in 80s °F	30.90	8.42
Panel C: County Labor Market Data		
Unemployment Rate	7.52	3.00
Median Income (2008 Dollars)	58,927.46	10,649.69
Panel D: Monthly Teacher Data		
Teacher Age	38.30	11.00
Absences Per Year	20.75	6.42
Observations, Panels A-C	3,663,375	
Observations, Panel D	165,841	

Note: Education and weather variables are at the school year level and span the 2007-08 through 2017-18 school years with the exception of student attendance which spans 2007-08 through 2012-13. Weather variables are calculated using September through April daily values. Monthly teacher variables span the 2005-06 through 2007-08 school years and exclude summer months.

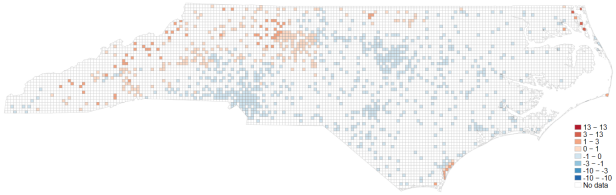
Table 1.2: Percent of Absences By Description and Age

	Teacher Age [24,54)	≥ 54
Absent with Deduction	0.03	0.04
Absent without Deduction	21.72	18.96
Absent without Pay	0.82	1.24
Annual Leave	17.61	21.80
Child Involvement Leave	0.06	0.01
Donated Shared Leave	0.05	0.05
Extended Sick Leave	0.32	0.05
Personal Leave	3.33	4.01
Sick Leave	56.04	53.81
Sick Leave Bank	0.02	0.02
Total	100.00	100.00

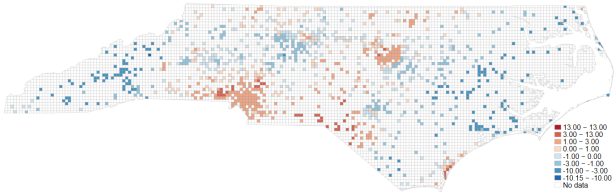
Figure 1.3: Variation in Number of Extreme Heat Days



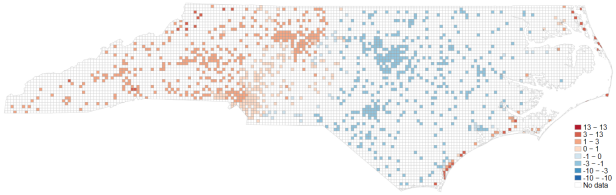
(a) 2009



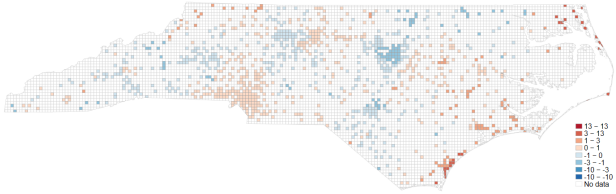
(b) 2010



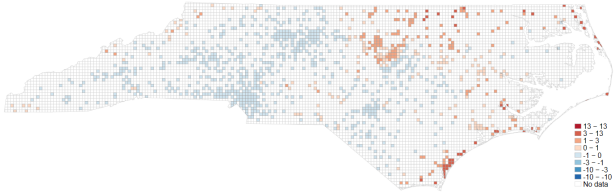
(c) 2011



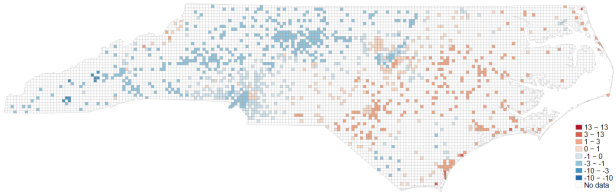
(d) 2012



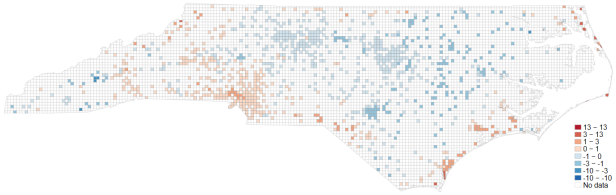
(e) 2013



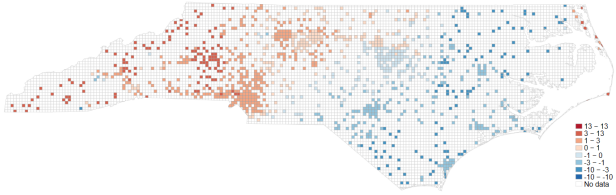
(f) 2014



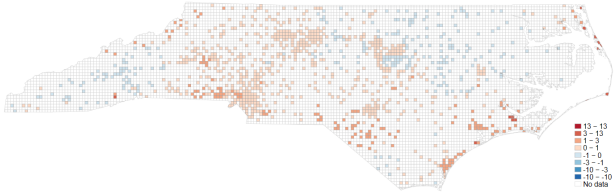
(g) 2015



(h) 2016



(i) 2017



(j) 2018

Note: This figure maps the number of extreme heat days by grid after subtracting that year's average. In this way, these maps represent the variation in extreme heat days after accounting for the yearly average. Figures a-j each show one year's variation from 2009 to 2018. Red and orange colors indicate a grid that had more extreme heat days than it had the previous year. Blues indicate the opposite. Missing data indicates grids in which there are no elementary schools.

1.3.4 Empirical Strategy

Students take annual exams at the end of each school year to measure their learning in mathematics and English language arts.[7] These exams are designed to test grade-level standards developed by North Carolina educators. Thus, the exams focus on content learned throughout the school year. Yearlong learning conditions and teacher attributes likely influence students' test performance.

To estimate the effect of extreme heat on student performance, I use a panel data model with year and school fixed effects. I regress standardized test scores on the number of extreme heat days in a school year and an indicator for teacher age above the 90th percentile. I include the interaction of these two explanatory variables to capture differences in heat effects by teacher age. This regression is of the following form:

$$y_{its} = \beta_1 H_{t\tilde{s}} + \beta_2 H_{t\tilde{s}} \times A_{tj} + \beta_3 A_{tj} + \mathbf{X_i}^{-1}\alpha + \delta_t + \lambda_{\tilde{s}} + \varepsilon_{its} \tag{1.6}$$

where

$$A_j = \begin{cases} 1 & \text{if teacher } j\text{'s age} \in [24, 54] \\ 0, & \text{otherwise} \end{cases}$$

and y_{its} is the outcome of interest: student i's standardized test score in year t and subject s. H_{ts} is the number of extreme heat days defined as the number of days in school year t that are above the 90th percentile for school $\tilde{s}$. $\mathbf{X_i}$ is a vector of control variables including lagged test score, student race, gender, and grade level. δ_t and $\lambda_{\tilde{s}}$ represent year and school fixed effects. School and year fixed effects are included to control for school-specific or year-specific unobservable factors that may affect test scores. In the model of teacher absences, I also include month fixed effects but cannot include most controls due to data availability during this shorter sample of 2006 through 2008.

I use the following panel data model to estimate the effects of extreme heat on teacher absences:

$$a_{jm} = \gamma_1 H_{m\tilde{s}} + \gamma_2 H_{m\tilde{s}} \times A_j + \gamma_3 A_j + \mathbf{C}^{-1}\kappa + \nu_y + \phi_m + \omega_{\tilde{s}} + u_{jm} \tag{1.7}$$

[7] End-of-grade science exams begin in the fifth grade and are excluded from this analysis because there are no past year scores.

where A_j is defined as in Equation A.2, a_{jm} is the number of absences that teacher j has in month m of school year t, $H_{m\tilde{s}}$ is the number of extreme heat days in the month at school $\tilde{s}$, and C is a vector of control variables including teacher education, county unemployment rate, real median income, CPI inflation, and yearly average precipitation.

1.3.5 Comparing Older and Younger Teachers

For causal identification, I assume that extreme heat days, conditional on fixed effects and observable characteristics, are uncorrelated with unobservables. To causually identify the effects of heat on younger and older teachers, I further assume that the interaction of age and extreme heat days, conditional on observables, is uncorrelated with unobservables. In this section, I provide suggestive evidence that these assumptions are reasonable.

Because the analysis compares outcomes between older and younger teachers, I must ensure that teacher age is not confounded with other systematic differences not captured by the fixed effects. For example, if older teachers are more likely to work in areas with more extreme heat days, any observed impacts might be attributable to location rather than age. However, I do not find evidence to support this concern.

Table 1.3 presents average heat, student demographics, and county-level economic variables by teacher age category and the difference between the two. Column 1 shows the average values of extreme heat days, days between 80 and 90°F and 90 and 100°F, average maximum temperature for the school year, the percentage of Black, Native American, and male, students, county-level unemployment rate, median income, and rural status. On average, older teachers experience 3.41 extreme heat days per school year, while younger teachers experience 3.39 extreme heat days. The difference between these figures is not significant. Older teacher work in schools with slightly fewer days in the 80s and 90s, with a slightly lower average maximum temperature. Though these differences are statistically significant, they are very small and represent less than a 1% difference in average maximum temperature, less than a 5% difference in days in the 80s and 90s. Older and younger teachers, on average, have very similar student demographics in terms of race and gender. However, older teachers tend to live in more rural areas with slightly higher unemployment rates and slightly lower median incomes. These differences are small but statistically significant. I include county economic variables in the main analysis to account for any differences in student outcomes stemming from economic differences that are not otherwise netted out.

Further, using a linear probability model, I show that teacher age categories are uncorrelated with extreme heat days conditional on school and year fixed effects. Appendix Table A.1 provides the full results of this analysis.

Teachers may also sort into schools based on the climate. Figure 1.4 displays the percentage of older teachers in each grid across North Carolina. The distribution appears generally even throughout the state. To further investigate this possibility, present descriptive results and an event study analysis to show that there is no evidence of differential climate sorting by teacher age. I focus on teachers who move from one school to another at least once during my sample. In a descriptive analysis presented in Appendix Table A.16, I find the mean temperature, extreme heat days, student demographics, unemployment rate, median income, and rurality of teachers' first, second, and third schools and show that there are very few significant differences between the change in characteristics from the one school to the next across teacher age, aside from median county-level income. Younger teachers move to schools in counties with higher median income relative to their prior school relative to older teachers (presented in balance table form in Appendix Tables A.17). Similarly, the event study plot show in Appendix Figure A.2 shows no difference in extreme heat between younger and older teachers who move.

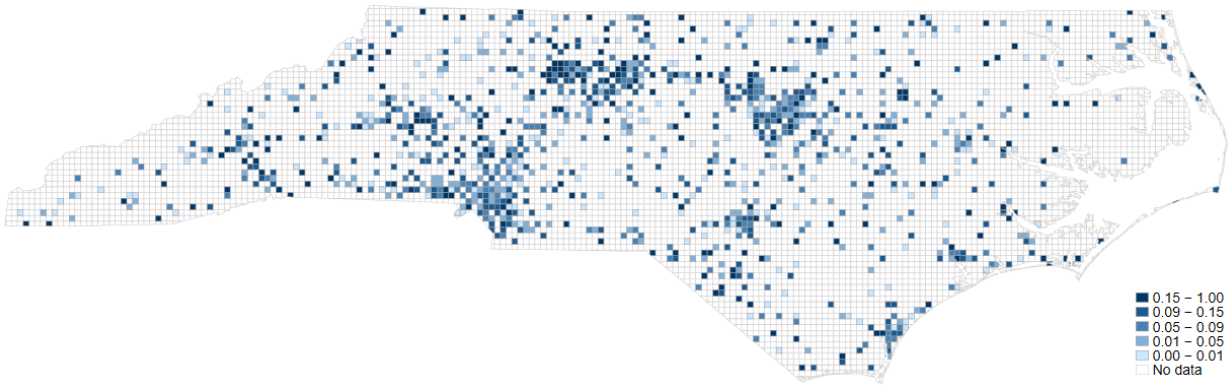


Figure 1.4: Percentage of older teachers

Note: This figure maps the average percentage of teachers above the age of 54 in each grid. Missing data indicates grids in which there are no elementary schools.

Table 1.3: Difference Between Older and Younger Teachers

	(1) Age [24,54)	(2) ≥ 54	(3) Difference (1)-(2)
Days Above 90th Percentile	3.39	3.41	-0.02
	(3.91)	(4.03)	(0.04)
Days in 80s °F	30.79	29.84	0.95***
	(8.21)	(8.46)	(0.09)
Days in 90s °F	3.28	3.14	0.14**
	(4.08)	(4.17)	(0.05)
Max Temp °F	64.43	64.10	0.33***
	(2.66)	(2.78)	(0.03)
Black Student %	0.28	0.28	0.00
	(0.28)	(0.29)	(0.00)
Male Student %	0.52	0.52	-0.01**
	(0.19)	(0.20)	(0.00)
Native American Student %	0.01	0.02	-0.00**
	(0.08)	(0.09)	(0.00)
Unemployment Rate	7.87	8.35	-0.48***
	(3.02)	(3.13)	(0.03)
Median Income	58811.03	57877.04	933.99***
	(10546.41)	(10370.08)	(113.43)
Rural %	0.22	0.26	-0.04***
	(0.42)	(0.44)	(0.00)
Observations	88,191	9,267	97,458

Columns 1 and 2 show means with standard deviations in parentheses. Column 3 shows the difference in means, with the standard error of the difference in parentheses and with stars specifying significance as follows: * for p<.05, ** for p<.01, *** for p<.001.

1.4 Heat's Impacts on Teachers and Their Students

1.4.1 Test Scores and Teacher Age

For students of older teachers, each additional day of extreme heat lowers test scores by an average of 0.002 standard deviations, a result that is significant at the 5% level and is consistent across choices of covariates. This decrease in scores is equivalent to a loss of 22% of a day's worth of instruction for an additional day of extreme heat.[8] The effect of extreme heat is significantly weaker for students taught by younger teachers, whose scores are 0.001 standard deviations higher than those taught by older teachers in response to an extreme heat day. This difference is statistically significant at the 1% level. Overall, students of younger teachers have a small and statistically insignificant response to extreme heat. Table 1.4 presents these findings in Column 6, including controls for past year test scores, the temperature on the day of the exam, student age, ethnicity, and gender, county-level unemployment and median income, and classroom size. Columns 1 presents results with no covariates, only year and school fixed effects. Columns 2 through 6 each add an additional covariate.

Moreover, additional extreme heat days reduce returns to teacher experience. At the average number of extreme heat days, students of older teachers achieve much higher test scores overall, even when accounting for prior achievement, exam-day temperature, student characteristics, and fixed effects. On average, students of younger teachers score 0.02 standard deviations lower than those of older teachers, a highly statistically significant finding that is stable across choice of covariates. I interpret this achievement gap as reflecting differences in teaching quality due to experience. Based on this interpretation, my findings suggest that teaching quality among more experienced teachers declines when they are exposed to extreme heat. Younger teachers are essentially unaffected, with a remarkably flat response to extreme heat days. Ten additional extreme heat days would completely erode the positive effect of having an older teacher. An additional ten extreme heat days would be a very large increase over the mean of three extreme heat days per school year; however, given estimates of climate change in the near-future, this increase may soon be in the realm of possibilities. (More on this in Section 1.4.6.)

[8] I calculate the average increase in test scores throughout a year using pre- and post-grade exams from third grade students that were administered for two years of the sample. This allows me to calculate the average increase for each day of instruction which makes for a useful benchmark against my estimated effects of an extra day of extreme heat.

Although the effects of extreme heat on scores are relatively stable across the choice of covariates, the magnitude of the effect for students of older teachers increases by nearly half when I do not include a student's past year test score as a covariate (as in Table 1.4, Column 1). At the same time, this specification results in a larger return to having an experienced teacher at the average of extreme heat days. These findings imply that when we do not account for a student's prior test scores, it appears that heat has greater effects on older teachers but at the same time, older teachers have larger returns. This result might be due to selection into older teachers' classes that is corrected by including past test scores and other student demographics in the regression specification.

Figure 1.5 plots mean test scores against bins of extreme heat days by teacher age category. This figure shows data residualized by the covariates included in regression Equation A.2: past test scores, exam day weather, student and classroom characteristics, county economic variables, and fixed effects. The figure shows that the test scores of students taught by older and younger teachers converge as exposure to extreme heat increases.

To test the robustness of these findings, I also conduct regressions using alternative measures of heat, such as cooling degree days, temperature bins, and a modified version of extreme heat days. The results remain qualitatively similar and indicate that temperature extremes, specifically days above 90°F and below 30°F, have the greatest impact on older teachers. The full results of these robustness checks appear in Appendix Tables A.2 through A.5. In Appendix Tables A.4 and A.5, I use alternate definitions in which one extreme heat day is one day above the 85th and 95th percentiles of heat rather than the 90th percentile. The results of these robustness checks are consistent with the main analysis.

Additionally, I test the sensitivity of the results to the definition of teacher age categories. I find similar results to those using the 90th percentile of teacher age when I use the 85th or 95th percentiles, though with somewhat noisier estimates. These results are presented in Appendix Tables A.7 and A.6. In another specification, I test whether the results are driven by the youngest, most inexperienced teachers having both worse outcomes and more heat resilience than other teachers. In Appendix Table A.8, I exclude the teachers below the tenth percentile of teacher age, that is, teachers below age 28. The results are again consistent with those in the main analysis.

Further, I investigate potential heterogeneous impacts to find whether any particular group drives the overall effects. If extreme heat affects some groups more than others, this finding provides a valuable insight to administrators and policy makers working to adapt classrooms to extreme heat.

To explore heterogeneous heat impacts, I conduct regression analyses of the form described in Equation A.2 after splitting the sample by test subject, grade level, student demographics, county-level median income, or semester-level heat and present the results in Appendix Tables A.9 through A.15. I find similar results for math and reading standardized test scores, though the results for math scores are noisy. Given the smaller point estimate for math scores, it is possible that there is not enough power to precisely estimate a smaller effect due to a decrease in the sample size. Splitting the sample by grade level offers similar results for each grade, though each subsample yields noisy results. Similarly, heterogeneity analysis by student gender, county rural status, or county median income yields no sizable differences between subgroups.

Table A.14 presents regressions by students' racial group. Black students of older teachers appear to be especially affected by extreme heat. For this group, an additional day of extreme heat leads to a 0.003 standard deviation decrease in test scores, which is precisely estimated at the 5% confidence level. Black student with younger teachers have an effect about half this size, though this difference is noisy. Native American students also appear to be more strongly affected than other groups, though these findings are noisy. Hispanic and White students have effects that are about half the size of those of Black students.

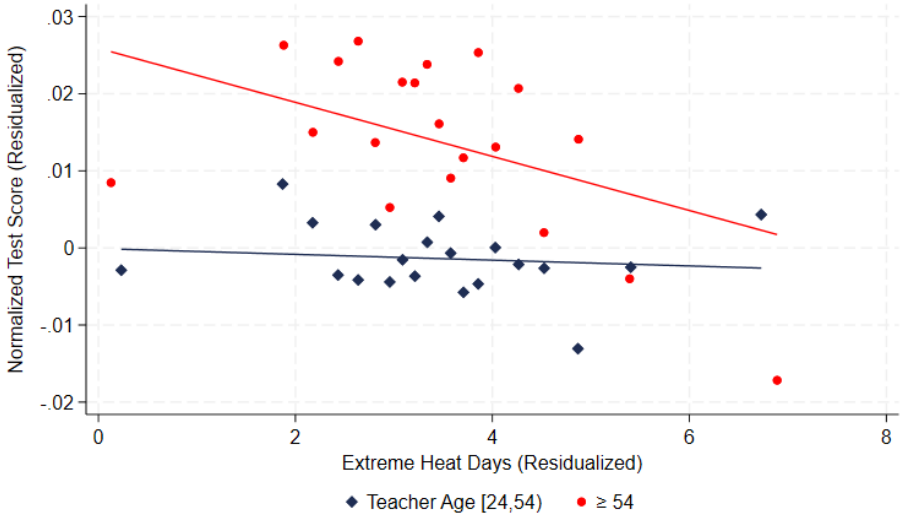
Finally, I conduct regressions in which I use only extreme heat days in the fall semester or spring semester of each year preceding end-of-year exams. I find that extreme heat days in the fall semester seem to drive the main result. This may be due to the fact that extreme heat days are much more common during the early fall semester months of September and October compared to any of the spring months of January through April.

Table 1.4: Regression of Test Scores on Extreme Heat Days and Teacher Age

	(1)	(2)	(3)	(4)	(5)	(6)
Extreme Heat Days	-0.00322**	-0.00153*	-0.00153*	-0.00167*	-0.00173*	-0.00185*
	(0.00119)	(0.000724)	(0.000724)	(0.000735)	(0.000735)	(0.000732)
Teacher Age [24,54)	-0.0539***	-0.0212***	-0.0212***	-0.0208***	-0.0208***	-0.0208***
	(0.00876)	(0.00372)	(0.00372)	(0.00375)	(0.00375)	(0.00363)
Teacher Age [24,54) × Extreme Heat Days	0.00269**	0.00147**	0.00147**	0.00142**	0.00140**	0.00135**
	(0.000971)	(0.000519)	(0.000519)	(0.000525)	(0.000524)	(0.000516)
Past Year Score	No	Yes	Yes	Yes	Yes	Yes
Day of Test Temperature	No	No	Yes	Yes	Yes	Yes
Student Demographics	No	No	No	Yes	Yes	Yes
County Economic Trends	No	No	No	No	Yes	Yes
Classroom Size	No	No	No	No	No	Yes
Observations	3,663,254	3,663,254	3,663,254	3,663,254	3,663,254	3,663,254

Note: Heteroskedasticity robust standard errors clustered at the grid level appear in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
This table presents the regression of normalized annual test scores on the number of extreme days in the school year, teacher age bins, and the interaction of the these explanatory variables. The first row presents the effect of one extreme heat day on students of older teachers on standard deviations of exam scores. The second row shows the effect of an indicator for younger teacher age at the average number of extreme heat days. The third row presents the difference in the effect of extreme heat days for students of younger versus older teachers. All regressions include year and school fixed effects. Column 1 does not include any covariates. Column 2 controls for a student's prior year test score. Column 3 adds the exam day temperature. Column 5 adds student gender, ethnicity, and age. Column 5 adds county-level yearly unemployment rates and average income. Column 6 adds the classroom size.

Figure 1.5: Extreme Heat and Scores Conditional on Teacher Age



Note: This figure plots average normalized test scores against the number of yearly extreme heat days by teacher age. Each point on the graph represents an average test score and an average number of extreme heat days for a range of extreme heat days, after both variables have been residualized by a set of controls. Test scores are normalized to have a mean of zero and standard deviation of one. Extreme heat days are defined as days above the 90th percentile of the empirical temperature distribution with a 2.5 mile by 2.5 mile grid.

1.4.2 Teacher Absences

Extreme heat may affect teachers through multiple channels. Teaching in a hot classroom or supervising outdoor activities during a heat wave can reduce teaching quality, though this effect is difficult to observe directly. Another channel is teacher absenteeism: extreme heat may exacerbate health conditions that cause teachers to miss school. At the same time, older, more experienced teachers who are advanced in their careers may face a smaller disincentive to miss work and may have more personal leave days. Such absences reduce instructional quality, as temporary substitute teachers typically fall in the 10th–20th percentile of teacher productivity (Herrmann and Rockoff, 2012).

To quantify the relationship between heat and teacher absences, I estimate a regression of teacher absences on extreme heat days as specified in Equation 1.7. On average, each additional extreme heat day results in 0.1 more absences per month for older teachers, a result that is highly statistically significant. This effect is about two times that of younger teachers–a result significant at the 0.01% level. Table 1.5 presents the full regression output. These results are stable across the choice of covariates. Column 1 presents this regression including only year, month, and school fixed effects. School fixed effects account for school- or location-specific factors that do not change over time, such as school culture, which may influence teacher absences. Year and month effects control for unobservable factors, like the adoption of new teacher leave policies, that can vary over time and affect teacher absences alongside the number of extreme heat days. Column 2 adds indicators for teacher education level. Column 3 introduces controls for county-by-month unemployment rates, real median income, and CPI inflation. The next two columns use subsamples to tease out whether these effects may differ by the type of absence. Column 5 uses the subsample of absences due to illness while column 6 uses the remainder of the sample, absences due to annual leave, personal leave, or unspecified reasons.

The effects of heat on absences for older teachers are stable across the choice of covariates, including using the subsample that includes only sick leave (shown in Table 1.5 Column 4. In this regression, the difference in the effects of heat between older and younger teachers becomes noisier and loses statistical significance though the point estimate and standard deviation are similar across models. This regression also shows that older teachers take more absences due to illness than younger teachers. Column 6, which uses the subsample that does not include absences due to illness echoes the previous findings that older teachers take about 1.5 times as many non-illness absences during months with one more extreme heat days compared to younger teachers.

Figure 1.6 shows the average number of absences against binned extreme heat days, separately by teacher age category. This figure shows data residualized by the covariates included in regression Equation 1.7: teacher education level, county economic variables, and fixed effects. This figure shows that older teachers are more likely to miss school when there are more extreme heat days in a month compared to younger teachers.

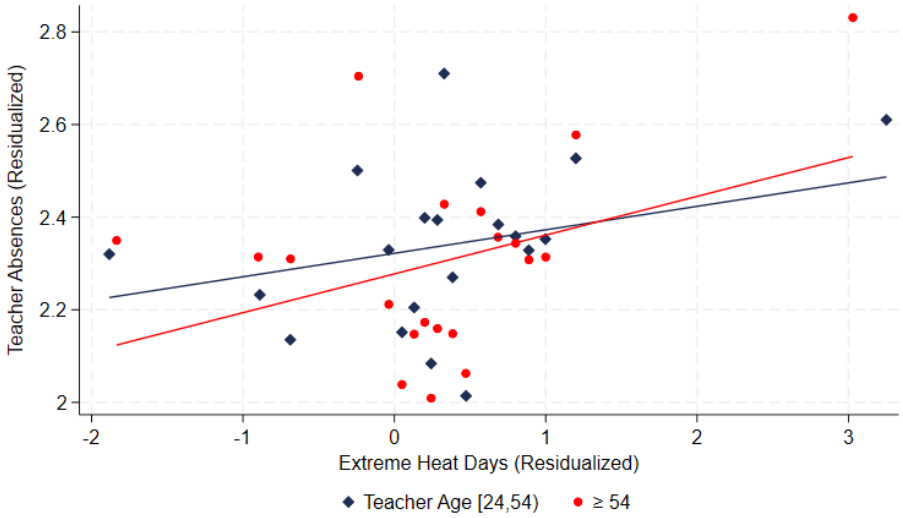
These findings indicate a pronounced behavioral response to heat, suggesting that additional mechanisms, such as reduced teaching effort on hot days, may also be at play. Given the magnitude of the observable behavioral effect, it is likely that other, less directly observable behaviors similarly contribute to the impact of heat on students' exam performance.

Table 1.5: Older teachers miss school more often during months with extreme heat

	(1)	(2)	(3)	(4)	(5)
Extreme Heat Days	0.0826***	0.0824***	0.0922***	0.111***	0.124***
	(0.0111)	(0.0111)	(0.0113)	(0.0207)	(0.0135)
Teacher Age [24,54)	0.0439*	0.0514**	0.0530**	-0.0687*	0.0184
	(0.0176)	(0.0176)	(0.0176)	(0.0270)	(0.0224)
Teacher Age [24,54) × Extreme Heat Days	-0.0406***	-0.0405***	-0.0414***	-0.0226	-0.0422**
	(0.00874)	(0.00874)	(0.00881)	(0.0185)	(0.0133)
Teacher Education	No	Yes	Yes	Yes	Yes
County Economic Trends	No	No	Yes	Yes	Yes
Sick Leave Indicator = 1	No	No	No	No	Yes
Sick Leave Indicator = 0	No	No	No	No	No
Observations	164,645	164,645	164,645	70,476	55,201

Note: Heteroskedasticity robust standard errors clustered at the grid level appear in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ This table presents regressions of teacher absences on the number of extreme days in the school year, teacher age bins, and the interaction of the these explanatory variables. All regression contain school, year, and month fixed effects. Column 3 includes a control for teacher education, county unemployment rate, real median income, and CPI inflation. Column 4 includes only absences due to sick leave and column 5 includes absences that are not due to illness.

Figure 1.6: Older teachers' attendance responds more in extreme heat



Note: This figure plots average residualized teacher absences over average residualized extreme heat days by teacher age. Each point on the graph represents the average absence and average extreme heat within a range of extreme heat days, after both variables are residualized by a set of controls. Extreme heat days are defined as days above the 90th percentile of the empirical temperature distribution with a 2.5 mile by 2.5 mile grid. Absences and heat days are at the monthly frequency.

1.4.3 Student Absences and Disciplinary Referrals

Previous literature indicates that extreme heat impacts student behaviors, though primarily in schools lacking air conditioning (McCormack, 2023). If this relationship holds in this setting, it may explain a separate channel through which heat impacts student outcomes.

In the following analysis, I regress student absences on extreme heat days, teacher age, and the interaction of these variables using a regression model akin to Equation A.2, including school and year fixed effects and a set of control variables. These analyses do not show evidence of significant effects of extreme heat on student absences or discipline, potentially because of data limitations–all student absences and disciplinary incidents are at the year level.

At face value, my findings suggest that one additional day of extreme heat in a school year leads to 0.003 fewer absences for students of older teachers and 0.01 fewer absences for students of younger teachers. However, these result is not precisely estimated, with large standard errors. Table 1.6, column 4 presents the regression results, which include controls for student demographics, prior year test scores, county economic trends, and year and school fixed effects.

Table 1.6: No evidence of an effect of extreme heat on yearly absences

	(1)	(2)	(3)	(4)	(5)
Extreme Heat Days	0.00425	0.00323	-0.00581	-0.00372	-0.00300
	(0.00823)	(0.00820)	(0.00870)	(0.00872)	(0.00873)
Teacher Age [24,54)	-0.0195	-0.0395	-0.0356	-0.0351	-0.0336
	(0.0304)	(0.0297)	(0.0297)	(0.0299)	(0.0299)
Teacher Age [24,54) $\times$ Extreme Heat Days	-0.00600	-0.00526	-0.00646	-0.00676	-0.00671
	(0.00523)	(0.00520)	(0.00514)	(0.00514)	(0.00514)
Past Year Score	No	Yes	Yes	Yes	Yes
Student Demographics	No	No	Yes	Yes	Yes
County Economic Trends	No	No	No	Yes	Yes
Classroom Size	No	No	No	No	Yes
Observations	3,658,774	3,658,774	3,658,774	3,658,774	3,658,774

Note: Heteroskedasticity robust standard errors clustered at the grid level appear in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ This table presents the regression of student absences on the number of extreme days in the school year, teacher age bins, and the interaction of the these explanatory variables. All regressions contain year and school fixed effects. Column 5 includes controls for lagged test score, student grade, ethnicity, gender, classroom size, and county-level unemployment rate and median income.

Table 1.7 presents the results from the regression of students' disciplinary referrals on the number of extreme heat days, teacher age, and the interaction of the two. I find very

small and noisy results, with confidence intervals that surround zero.

Table 1.7: No evidence of extreme heat effect on disciplinary referrals

	(1)	(2)	(3)	(4)	(5)
Extreme Heat Days	0.000217	-0.0000830	-0.000129	-0.000329	-0.0000792
	(0.00229)	(0.00228)	(0.00231)	(0.00231)	(0.00232)
Teacher Age [24,54)	0.00922	0.00340	0.00371	0.00363	0.00414
	(0.00843)	(0.00823)	(0.00810)	(0.00809)	(0.00810)
Teacher Age [24,54) $\times$ Extreme Heat Days	-0.000600	-0.000383	-0.000248	-0.000171	-0.000153
	(0.00141)	(0.00140)	(0.00139)	(0.00139)	(0.00139)
Past Year Score	No	Yes	Yes	Yes	Yes
Student Demographics	No	No	Yes	Yes	Yes
County Economic Trends	No	No	No	Yes	Yes
Classroom Size	No	No	No	No	Yes
Observations	3,663,254	3,663,254	3,663,254	3,663,254	3,663,254

Note: Heteroskedasticity robust standard errors clustered at the grid level appear in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ This table presents the regression of student disciplinary incidents on the number of extreme days in the school year, teacher age bins, and the interaction of the these explanatory variables. Column 5 includes controls for lagged test score, student grade, ethnicity, gender, and year and school fixed effects.

1.4.4 Exploring the Returns to Age in the Presence of Extreme Heat

To study the effects of heat on teachers, I rely on interactions between heat and aging, making a parametric choice to study the top 10% of teachers by age compared to their younger colleagues. To explore the non-parametric relationship between test scores and teacher age under different levels of heat exposure, I create a binned scatterplot with teacher age on the x-axis and standardized test scores on the y-axis, with two series: one for student-teacher pairs in years with an above-average number of extreme heat days and another for those below the average. Figure 1.7 includes data residualized by the covariates in regression Equation 1.7: teacher education level, county economic variables, and school and year fixed effects.

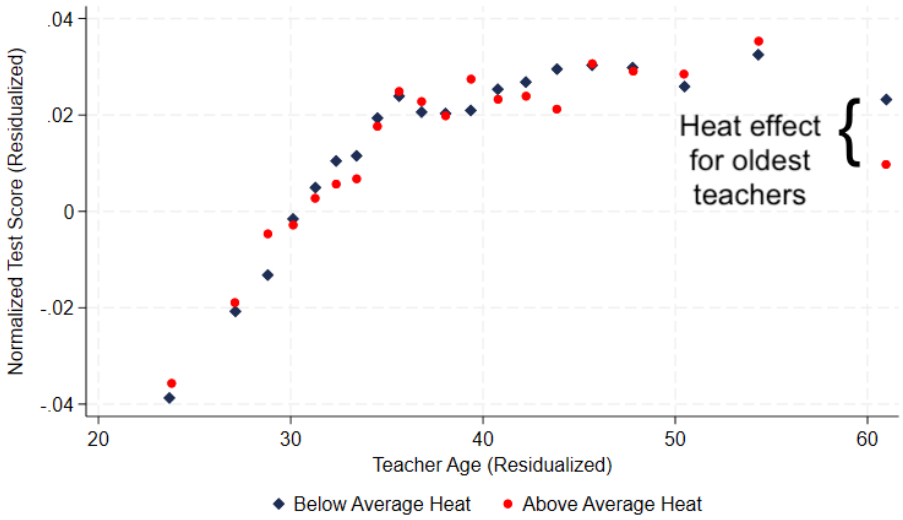
Figure 1.7 shows the descriptive relationship between test scores and teacher age by heat exposure level. The vertical distance between the red dots and blue diamonds illustrates how heat affects teacher performance by age. The largest difference appears for the oldest age group in the chart, teachers above age 54, validating the choice to focus on the top 10% of teachers by age.

The figure also shows that after about age 35, gains to experience flatten sharply, with the slope of the curve decreasing markedly in the early-to-mid 30s. While this is descriptive evidence only, it aligns with the literature on returns to teacher experience[9].

Further, at the right tail of teacher age, there appears to be a drop to the returns to teacher age, particularly for teachers with above average heat exposure. These teachers' students have test scores below that of other teachers above 35 years old. I infer from this relationship that the oldest, most heat-exposed teachers lose some of the gains to teacher experience, though not all. These teachers still have students who perform higher than those of the youngest teachers.

[9]Earlier conventional wisdom suggested that teacher experience yielded little improvement after the first two or three years of teaching (Hanushek et al., 2005). However, more recent work shows that teachers continue to make gains beyond the first five years, with potential improvements extending into later decades of experience (Wiswall, 2013; Papay and Kraft, 2015; Ladd and Sorensen, 2017).

Figure 1.7: Test Scores by Teacher Age Above and Below Average Heat



Note: This figure plots average normalized test scores against teacher age by heat exposure level. Each point on the graph represents an average test score and an average teacher age with a range of ages, after both variables have been residualized by a set of controls. Test scores are normalized to have a mean of zero and standard deviation of one. The blue diamonds represent student-teacher pairs who experienced a below-average number of extreme heat days in a year while the red dots are those who had above-average heat. Teacher age is calculated using undergraduate graduation date, assuming age 24 at graduation.

1.4.5 Climate Control in Classrooms

Though I do not explicitly account for air conditioning in classrooms throughout this analysis, the baseline climate and building policies in North Carolina suggest that most classrooms have climate control. There is no statewide database that details which schools have climate control or when air conditioning systems were installed or upgraded. Given the hot and humid climate in much of North Carolina and the years covered by my sample, I assume that all elementary classrooms are equipped with air conditioning. I have some evidence to support this assumption. First, I compile a dataset of school maintenance requests beginning in 2008 for several of the largest school districts in my sample, and I find maintenance requests for existing air conditioning systems in every district included.[10] Second, the state's 2014 Department of Public Instruction facilities guidelines specify the inclusion of central air conditioning, which suggests that schools built since at least 2014 are climate controlled. The analysis in this paper indicates that, despite the presence of air conditioning, teachers and students still experience declines in productivity and learning during periods of extreme heat. This may be due to a need for repairs and maintenance or it may be that older air-conditioning systems were not designed to handle current extremes.

1.4.6 Climate Change

With a warming climate, teachers and their students face an increasing number of extreme heat days. To predict the impacts of climate change, climate models often use climate scenarios proposed by the Intergovernmental Panel on Climate Change (IPCC), such as Representative Concentration Pathway (RCP) 4.5, a moderate climate change scenario, and RCP 8.5, a more severe scenario. RCP 4.5 requires enough technological innovation or industrial and climate policy to reduce increases in greenhouse gases, reaching stable levels shortly after the year 2100. The more severe climate change scenario, RCP 8.5, assumes increasing global greenhouse gas emissions (United States Environmental Protection Agency, 2025). A peer-reviewed report on climate change in North Carolina predicts the prevalence of extreme weather in the coming decades using RCP 4.5 and 8.5 (Kunkel et al., 2020).[11] For the 2041-2060 period, under moderate climate change, the report esti-

[10] Data on school-level HVAC maintenance requests were collected through Freedom of Information Act requests.

[11] Then-governor Roy Cooper commissioned this report under Executive Order 80.

mates that there will generally be up to 15 to 25 more days above 95°F per year compared to the 1996-2015 average, with larger increases in isolated areas. Under more severe climate change, the report projects around 25 or more very hot days per year for much of the state with even more hot days in the limited areas. At the same time, under both climate scenarios, the report finds an increase throughout the state in the number of very warm nights–nights above a minimum temperature of 75°F.

Further, the North Carolina climate report highlights the engineering design problems we face due to climate change, given that infrastructure like air-conditioning systems is engineered using historical extreme temperature data. Much of the current infrastructure may already be out-of-date and incapable of handling current extreme temperatures. Through technological innovation and improved design, buildings may become better at handling very hot temperatures over time. However, recent work shows there may be a limited scope for climate adaptation to extreme heat, at least in the short term (Costa et al., 2024).

Climate scientists estimate that, even under a moderate climate change scenario, North Carolina will have an increase of 15 to 25 days of extreme heat per year in 15 years' time–implying that the classroom environment will likely become hotter during the school year. Assuming the same relationship holds between learning and heat that I present in this work, ten more extreme heat days per school year leads to a yearly decrease in test scores of 0.02 standard deviations. If these effects are additive across years and there is no further climate adaptation, this decrease results in a loss of learning similar to the estimated loss in the immediate aftermath of Hurricane Katrina every ten years (Sacerdote, 2012). However, this estimate does not capture the increased sensitivity to extreme heat due to an aging workforce. In 2018, adults above 45 years old made up about 42% the US population

1.5 Conclusion

While much of the literature on heat and learning focuses on students, this paper shows that teachers are themselves directly affected by extreme heat. I document how extreme heat undermines teacher productivity, particularly among older, more experienced educators, through both reduced instructional quality and increased absenteeism. In doing so, this work bridges two strands of research: the impact of climate on worker productivity and the economics of education, highlighting an overlooked mechanism through which environmental conditions shape student outcomes.

The results reveal a tension between teaching quality and sensitivity to heat. Older teachers, who typically deliver the greatest gains in student achievement in this analysis, are also the most sensitive to heat. In my setting, just ten additional extreme heat days are enough to erase the benefits of teacher experience. As climate change intensifies, this increased sensitivity threatens to erode one of the education system's key strengths—its experienced workforce. The fact that a relatively small group of teachers accounts for a large share of the climate-related loss in learning underscores both the scale of the challenge and the potential effectiveness of targeted mitigation strategies.

These findings also carry broader implications for inequality. Extreme heat effects are not distributed evenly: Black students and those in under-resourced schools with outdated infrastructure are more likely to learn in heat-vulnerable environments. If extreme heat disproportionately diminishes the effectiveness of experienced teachers in these schools, it could further widen achievement gaps across districts and communities. Investments in school infrastructure, workforce protections, and climate adaptation policies therefore have the potential to both improve resilience and reduce educational inequality.

Ultimately, this research shows that the costs of climate change extend beyond immediate health and labor market effects to the classroom, where they may compound over time by weakening the foundations of human capital. By identifying teachers as a critical channel through which heat influences learning, this paper provides new evidence to guide policy choices—whether in evaluating teacher performance, planning for climate-resilient schools, or deciding when to close schools during heat waves. These choices will shape whether the education system can continue to deliver high quality instruction in an era of rising temperatures.

CHAPTER 2

Degrees of Mobility: Shifting Low-Income Students to Higher Quality Colleges*

Abstract

Earnings inequality among college graduates has grown, and returns to a bachelor's degree (BA) vary substantially across colleges. We experimentally evaluate a program that shifted high-achieving, low-income students into higher-quality colleges. Students randomly offered a guarantee of four years of tuition and fees at the University of Michigan were 21 percentage points more likely to enroll at a UM campus. The offer shifted students across colleges rather than into college: most compliers would otherwise have attended less-selective four-year colleges. HAIL increased four-year BA completion by 4 percentage points, but had no effect on attainment after six years. It substantially increased degree quality: students offered the scholarship were 12 percentage points more likely to earn a bachelor's degree from a highly or most selective college. IV estimates indicate that students induced to attend a University of Michigan campus were 56 percentage points more likely to earn such a degree. The increase in college quality predicts earnings gains of 18 to 25 percent. Faster degree completion adds about another 1 percent to the present discounted value of lifetime earnings.

2.1 Introduction

Over the late 20th century, increasing returns to education drove rising earnings inequality. In the 21st century, by contrast, earnings inequality has primarily grown within edu-

*Co-authored with Elizabeth Burland, Nora Delaney, Sue Dynarski, Katie Leu, CJ Libassi, Kathy Michelmore, Stephanie Owen, and Liz Salinas

cation groups, especially among those with a bachelor's degree (BA) (Autor et al., 2020). Growing evidence suggests that the quality of a college is tied to the economic returns it delivers. Low-income students disproportionately attend open-access and less selective schools, while high-income students are far more likely to enroll in highly selective colleges. In the Ivy League, there are more students from the top one percent of the parental income distribution than there are students from the bottom 50 percent (Chetty et al., 2020). These enrollment patterns raise concerns that postsecondary education in the US may act more as a reinforcer of class rigidities than as an engine of economic mobility.

Michigan makes this national pattern concrete. Its public higher-education system includes a high-mobility flagship, less-selective four-year colleges, and community colleges that enroll many high-achieving, low-income students. Half of students from the bottom quintile of parental income who attend the University of Michigan flagship in Ann Arbor (hereafter, UM flagship) reach the top income quintile as adults, compared with just 10 percent at Grand Rapids Community College, 19 percent at Eastern Michigan University, and 34 percent at Michigan State University.[1]

But while the UM flagship has high upward mobility rates among low-income students, it enrolls few of them: as of 2014, just three percent of its undergraduates were from the bottom quintile of parental income. Three times as many (nine percent) were from the top one percent of the income distribution (Chetty et al., 2022).

These observed mobility patterns do not identify the causal effect of attending a more selective college. Since students select colleges based on family resources and academic preparation, simple comparisons confound selection with treatment. Quasi-experimental designs using admissions cutoffs and waitlists identify sizable returns to college quality for students at the margin of admission. Less is known about the broader set of low-income students who could plausibly attend highly selective colleges. We provide unusually clean experimental evidence on the effect of attending a higher-quality college for high-achieving, low-income students.

We fielded a randomized intervention designed to encourage high-achieving, low-income students to apply to and enroll at Michigan's highly selective flagship. The High Achieving Involved Leader (HAIL) Scholarship guaranteed to high school students four years of free tuition and fees at the flagship if they were admitted. Low-income students were pre-qualified using data on subsidized lunch eligibility from state administrative

[1]The colleges we use throughout the paper as examples are frequently attended by students in the control group.

records; no financial aid forms, including the FAFSA, were required. This was not an increase in aid, since the flagship already fully met financial need. Rather, it was a retiming of when the aid offer was made, from after admission to before applications.

HAIL more than doubled applications to the UM flagship and shifted enrollment from community colleges and less-selective four-year schools into the UM system (Dynarski et al., 2021). Because the scholarship did not change aid received conditional on enrollment at UM, random assignment of the offer generates clean variation in college choice for high-achieving, low-income students. We use the offer of the scholarship as an instrument for enrollment at any UM campus within one year of expected high school graduation, examining effects on completion and degree quality.

We estimate intent-to-treat effects of the offer on degree attainment, degree quality, and field of study. The offer changed where students enrolled, not whether they enrolled in college. The HAIL offer accelerated completion, raising four-year BA attainment by 4 percentage points (11 percent), with no effect on BA attainment after six years. Increased full-time enrollment and faster credit accumulation are the mechanisms for shorter time to degree. It also upgraded degree quality: treated students were 12 percentage points (81 percent) more likely to earn a BA from a highly selective institution. The treatment did not affect the share of students earning STEM degrees; treated students are just as likely as the control group to earn a STEM degree.

The students whose choices changed because of HAIL were academically strong and largely oriented toward four-year colleges. In the absence of HAIL, most would have attended four-year colleges, but few would have attended a college as selective as the UM flagship. HAIL therefore shifted compliers into colleges with stronger peers, greater resources, higher rates of full-time enrollment, and higher graduation rates.

Using assignment to treatment as an instrument, we find that attending any UM campus increased four-year BA attainment among compliers by 20 percentage points, relative to a control-complier mean of 42 percent. Enrolling at any UM campus as a freshman increases by 56 percentage points the likelihood of earning a BA from a college as selective as the UM flagship.[2] This is extremely large relative to the control-complier mean of five percent.

Based on this increase in BA quality, we project substantial increases in earnings. As the sample is too young to obtain meaningful estimates at the individual level, we project

[2]Barron's classifies the UM flagship as a "highly competitive" university, the second highest category.

earnings using college-level data from the College Scorecard (College Scorecard, 2021). Attending a UM campus increases projected earnings eight years after college entry by 26 percent, or $16,300 relative to a control-complier mean of $62,300.[3] Using earnings data for students from families earning less than $30,000, the projected gain is 31 percent. The overall projection is roughly equivalent to moving a graduate from the 52nd to the 66th percentile of the earnings distribution in the Current Population Survey for those with a BA.

On top of this quality return is an increase in lifetime earnings from an additional year in the labor force resulting from reduced time to degree. Among compliers, attending a UM campus increases the likelihood of graduating within four years by 20 percentage points. For these students, the earlier labor force entry adds 4 to 5 percent to the present discounted value of lifetime earnings, contributing roughly 1 percent to expected earnings across all compliers.[4]

2.2 Prior Literature

2.2.1 What Is College Quality?

"College quality" has no formal definition. Researchers, policymakers, and the media use terms such as "high quality," "highly selective," and "elite" to describe colleges that are distinctive along multiple dimensions. Colleges with higher achieving students, greater instructional resources, lower student-faculty ratios, and more selective admissions tend to have higher graduation rates, faster degree completion, and higher earnings (Black and Smith, 2006). Because these inputs are highly correlated, it is challenging to isolate their separate effects. Throughout the paper, we use "college quality" to refer broadly to this bundle of peer, resource, and institutional characteristics.

[3] We caution that these earnings predictions reflect both the causal effect of attending a particular college as well as non-random selection into higher-quality colleges. Recent work suggests that about 70–80% of the variation in earnings across institutions reflects the causal effect of college quality (Bleemer and Quincy, 2025).

[4] See Section 2.8.4 for these calculations.

2.2.2 Evidence on the Causal Effect of College Quality

Estimating the effect of college quality is difficult because students sort into colleges based on characteristics that also affect later outcomes. We focus in this review on studies that use quasi-experimental or experimental methods to isolate the causal effect of college quality.[5] Existing studies vary in the students they examine, the quality changes they identify, and the extent to which quality is bundled with supports supplied in the context of a particular intervention. We summarize the existing research in Appendix Table B.1.[6]

A small experimental literature studies interventions that shift students toward higher-quality colleges. Angrist et al. (2022) evaluate a scholarship that shifts students from community colleges into four-year colleges and find no effect on four-year BA attainment, but an eight-percentage-point increase in six-year attainment among students who planned to attend a four-year college. Barr and Castleman (2025) evaluate an advising intervention that shifts students toward higher-quality colleges and increases four-year degree attainment by six percentage points and six-year attainment by 7.6 percentage points. Cohodes et al. (2026) study a STEM summer program that induces students to enroll in highly selective colleges; they find a large increase in STEM degree attainment, an imprecise five-percentage-point increase in four-year attainment, and no effect on six-year attainment. These studies provide important evidence, but each combines movement to a higher-quality college with additional aid, advising, mentoring, or academic support supplied by the intervention itself.

A larger quasi-experimental literature uses admissions cutoffs to estimate the effect of college quality. Hoekstra (2009) estimates a large effect of attending a flagship on earnings; the paper provides no information about counterfactual colleges, so we cannot gauge the quality shock experienced by students. Zimmerman (2014), Goodman et al. (2017), and Kozakowski (2023) estimate the effect of gaining admission to a four-year public college in settings where the counterfactual is often a community college or no college. They

[5] We do not review the large literature that relies on observational comparisons or structural models rather than a clearly identified source of exogenous variation. Examples include Light and Strayer (2000) and Arcidiacono et al. (2014). We also do not review the extensive literature on affirmative action and college quality, which addresses a distinct set of questions regarding admissions policy and student sorting across colleges (e.g., Bleemer, 2022).

[6] A separate qualitative literature examines the experiences of low-income students in higher education. It documents challenges meeting basic needs such as food, housing, and transportation, as well as difficulties navigating the norms and expectations of highly selective colleges (Goldrick-Rab, 2016; Jack, 2019; Jack and Black, 2024). While this literature provides insight into why low-income students may face challenges in college, it does not provide causal evidence on how college quality affects their educational outcomes.

find positive effects on bachelor's degree completion, with point estimates summarized in Appendix Table B.1. Mountjoy (2026) studies admissions cutoffs at public four-year colleges in Texas, where the counterfactual includes both community colleges and other four-year colleges. He estimates increases of roughly one percentage point in four-year BA attainment, two percentage points in six-year BA attainment, and one percent in earnings.

These quasi-experimental designs provide credible causal evidence. However, by design, they focus on students at admissions thresholds. The resulting estimates apply to students who are academically marginal relative to their college peers. As a result, they confound the effect of college quality with the effect of low academic rank within a college.

In a closely-related paper, Cohodes and Goodman (2014) examine high-achieving students in Massachusetts induced by a small tuition discount to alter their college choices. The college landscape of Massachusetts is a mirror image of that in Michigan: the highest quality colleges in Massachusetts are private, while in Michigan the top college is the public flagship. Treated students were induced into public colleges of lower average quality than the private colleges they would have otherwise attended. As a result, their likelihood of graduating college was reduced by 2.5 percentage points.

A related literature focuses on elite private colleges. Chetty et al. (2026) estimate the effect of attending an Ivy-plus college and find effects on earnings and access to elite occupations. Dale and Krueger (2002) compare outcomes for students accepted by similarly selective colleges and conclude that college selectivity has little effect on earnings, though an update using administrative data finds positive earnings effects for Black, Hispanic, and first-generation students (Dale and Krueger, 2014). This evidence speaks to a different margin: movement among highly selective colleges, rather than movement from lower-quality colleges into flagship-quality options.

This literature highlights an important challenge: the college-quality "treatment" varies substantially across contexts. A student moving from a community college to a four-year college experiences a different change in peers, resources, and academic opportunities than a student moving from one elite college to another. Sound extrapolation therefore requires attention to both the students studied and the quality changes induced by treatment. Appendix Table B.1 reports, when available, the population and quality shock in each study. Missing information in the underlying papers makes it difficult to determine whether differences in estimated effects arise from differences in populations, differences in quality changes, or both.

2.2.3 Contribution

We make three contributions to the literature on college quality.

First, we isolate the consequences of attending a higher-quality college more cleanly than previous experimental studies. The HAIL intervention changed college enrollment but provided no additional financial aid, academic support, or advising once students enrolled. By contrast, the limited experimental evidence on college quality comes from interventions that provide treated students with additional grant aid, advising, mentoring, or other supports beyond those ordinarily provided by the colleges they attend.

Second, we provide an unusually detailed description of the quality shock. Most studies treat college quality as a black box, characterizing treatment as movement between broad categories such as community colleges and four-year colleges or across Barron's selectivity categories. Our administrative data follow students from high school into college, allowing us to measure how HAIL alters peer achievement, academic match, instructional resources, selectivity, and the socio-economic composition of peers. This provides a clearer picture of what changes when students move to a higher-quality college.

Third, we provide experimental evidence on the effect of college quality for academically strong students. Much of the existing literature identifies effects at admissions cutoffs, where students are academically marginal relative to their college peers. The students in our sample score higher than 85 percent of SAT-takers nationwide and have academic records typical of students at the highly selective UM flagship.

2.3 The Intervention and Randomization

2.3.1 The HAIL Scholarship

We partnered with the UM flagship to increase its enrollment of low-income students. We jointly designed and implemented a randomized intervention,[7] the HAIL Scholarship. HAIL offered selected students an up-front guarantee of four years of tuition and fees at the UM flagship if they were admitted. The offer conferred no admissions advantage.

[7]The randomized trial was approved by the IRB at the University of Michigan (HUM00096289). The present analysis is overseen by the University of Michigan (HUM00224676), Harvard University (IRB21-1221), Colby College (2021-093), and the University of Connecticut (X23-0649).

Students in both treatment and control schools applied through the standard admissions process, which reviews transcripts, test scores, essays, letters of recommendation, and extracurricular activities.

We identified sample students using state administrative records. Public high school students qualified if they were eligible for subsidized school meals and performed above test-score and GPA cutoffs chosen by the UM admissions office.[8] Qualifying SAT scores started at 1100 while qualifying GPAs started at 3.3. Students with higher test scores faced a lower GPA threshold (and vice versa).[9] Because Michigan public schools administer the SAT in school to all eleventh graders, we could identify high-achieving, low-income students who, in other states, may not have even attempted the SAT.[10] The distribution of student achievement in Michigan mirrors the national distribution, with students in 2017 averaging around 1000 on the SAT with a standard deviation of 195 (U.S. Department of Education, National Center for Education Statistics, 2017b).

Applicants from treatment and control schools were admitted to the flagship at similar rates. Forty-six percent of HAIL-eligible applicants were admitted, substantially higher than the overall admission rate of 29 percent (Dynarski et al., 2021). The intervention therefore did not simply induce marginally qualified students to apply. If anything, HAIL-eligible students were more academically qualified than the typical applicant.

The intervention consisted of targeted outreach that communicated the guarantee. Students in the treatment group were mailed a packet describing the HAIL offer and encouraging them to apply. Parents received a follow-up letter with the same information, while high school principals received a list of eligible students and information about the scholarship (see Dynarski et al., 2021 for materials and implementation details). Sample members at control schools received the UM flagship's standard outreach materials, which emphasized generous scholarships and high academic standards but made no concrete promise of financial aid.

The HAIL offer differed from typical financial aid in several ways. In the traditional system, students learn their net price only after applying and being admitted to a college. HAIL offered an aid guarantee before college applications were due and required no financial aid forms. In the traditional system, students reapply annually for aid, and aid

[8]The family income threshold for subsidized meal eligibility is 185 percent of the federal poverty line.

[9]Michigan switched from the ACT to the SAT in 2016, one year into the study; we convert ACT scores to SAT scores using ACT-SAT concordance tables provided by the College Board.

[10]Hyman (2017) shows universal testing led more disadvantaged students to take the college admissions test, with many appearing in the upper end of the score distribution.

can change from year to year. HAIL provided a four-year commitment in the fall of senior year of high school. By replacing uncertainty and administrative burden with an early, credible guarantee, the offer changed enrollment choices among high-achieving students from low-income families (Dynarski et al., 2021; Burland et al., 2023).

HAIL did not increase aid, but rather shifted when it was promised. Among those who attended the UM flagship, students in the treatment and control groups received the same amount of grant aid both annually and cumulatively (Appendix Table B.2). The two groups also received the same academic and social services at the flagship, with no special resources devoted to HAIL admits. Thus, HAIL altered college choice but did not increase the aid or services students received conditional on enrolling at UM.

We learned of an additional implementation detail after the publication of our first two studies of the intervention's impacts (Dynarski et al., 2021; Burland et al., 2023): the admissions office made an offer of deferred admission to some treatment-group students.[11] Affected students were promised transfer to the UM flagship after one year at a UM satellite campus (Dearborn or Flint), with the tuition guarantee applying to this pathway.[12]

The satellite campuses have lower peer achievement and graduation rates than the flagship. In principle, the deferred-admission offer could therefore reduce college quality for students diverted from other four-year colleges into a satellite campus.[13] Below, we show that the satellite campuses, while below the flagship on most quality dimensions, are similar to the colleges attended by the control group. We also find that many students entering through the satellite pathway ultimately earn highly competitive degrees, suggesting that the deferred-admission pathway served as a bridge into the flagship rather than a lower-quality alternative.

[11]The initial admissions review is "need blind," with scoring done by staff who do not know students' financial circumstances (or HAIL status). In the next stage, admissions decisions are made, with students admitted, rejected, deferred, or waitlisted. It is our understanding that at this stage satellite offers were assigned to marginal applicants.

[12]In the first cohort, 147 treatment students were offered deferred admission to the flagship if they first enrolled at Dearborn; half took the offer. In the second cohort, 214 treatment students were offered deferred admission if they first enrolled at Dearborn or Flint; a third took the offer. Across the two cohorts, students who took the offer mostly enrolled at Dearborn (60 percent vs 40 percent at Flint).

[13]The average SAT score for in-state students is 1360 at the flagship, 1160 at Dearborn, and 1061 at Flint. Six year graduation rates are 92 percent at the flagship, 62 percent at Dearborn, and 56 percent at Flint (College Scorecard, 2021).

2.3.2 Randomization

Random assignment occurred at the high school level, with all eligible students at a given school assigned the same treatment status. School randomization occurred within four strata defined by the number of HAIL-eligible students in a school. High schools with eligible students in both years of the experiment kept their treatment status in the second year.

We study 3,910 students who graduated from high school in 2016 and 2017: 1,932 treated students (from 289 unique schools) and 1,978 control students (from 299 schools). Table 2.1 shows the experimental balance on school-level characteristics. The treatment and control groups are balanced on these characteristics. In the control group, the average SAT score is 1254 and the average GPA is 3.82. Fifty-three percent of control-group schools are in rural areas, 40 percent are in Southeast Michigan, and average 93 miles from the University of Michigan. All students are eligible for free or reduced-price lunch; 71 percent qualify for free lunch and 29 percent for reduced-price lunch. The sample is primarily White and non-Hispanic and majority female. Forty-six percent of students sent their ACT scores to the UM flagship when they took the test in their junior year (before randomization).

2.4 Data and Methods

2.4.1 Data

We obtained longitudinal, student-level administrative data from the Michigan Center for Educational Performance and Information (CEPI) that contain the universe of students attending Michigan public schools. These data served as the base population from which we identified students for the original intervention, and also provide information on pre-treatment characteristics such as gender, race, school geography, and test scores. The ACT and SAT data files include indicators for up to four colleges that each student could choose to send their scores to at the time of test-taking (pre-treatment). The score-send data are incomplete for the second cohort; we therefore report statistics on score sending only for the first cohort.

To track postsecondary enrollment and degree completion, we link to National Student

Clearinghouse (NSC) data, which provide college attendance and credential records from virtually all U.S. colleges and universities. NSC data include the field of degree. Our primary educational outcomes are degree attainment, measured within four and six years after expected high school graduation, and the selectivity of the degree-granting college. We categorize degrees as "highly selective" if they are from colleges categorized as highly competitive or most competitive by the 2013 Barron's selectivity index. Detailed information on how we define our outcome measures is included in Appendix B.

To characterize the institutions attended by the students in our sample, we use several sources. We match NSC data to institutional data from College Scorecard and IPEDS. These sources allow us to track instructional spending per student, the proportion of Pell-eligible students, student-faculty ratios, and average graduation rates. By necessity, analyses using these data sources are limited to students who enroll in college. As we show below, there is no treatment-induced selection on the extensive margin of college. We also use the student-level CEPI data on SAT scores matched to NSC enrollment records to calculate the average score of all in-state students who attend each college. Due to universal test-taking and the full coverage of the administrative data, we can do this even for schools that do not require test scores.

We supplement NSC records with transcript data from the Student Transcript and Academic Record Repository (STARR), which includes credits attempted and earned at Michigan public colleges.[14] Because the intervention shifts enrollment toward Michigan public colleges, treatment status affects whether students appear in STARR. We therefore impute missing credits using the approach of Angrist et al. (2022), predicting credits based on the sector and selectivity of the college attended. As a robustness check, we also calculate Lee (2009) bounds. The two approaches (described in more detail in Appendix B.1.7) yield similar conclusions, though the Lee bounds are wide.

Data on adult earnings are not yet available for our cohorts. We therefore project earnings using publicly available College Scorecard data on median earnings eight years after college entry (College Scorecard, 2021). We assign students earnings based on the modal college they attend. For students who do not attend college, we use earnings estimates from Opportunity Insights (Chetty et al., 2022). We also calculate projections separately by family income at college entry. All dollar amounts are converted to 2025 values.

[14] We do not reliably observe credits transferred into a college, whether from advanced high school classes or from other colleges.

2.4.2 Empirical Strategy

We estimate the effects of the HAIL Scholarship on initial enrollment, degree attainment, degree selectivity, STEM degree completion, credit accumulation, transfers, and projected earnings. Because treatment was randomly assigned at the high school for each cohort, our analytic unit is the high school-by-cohort cell:

$$Y_{jt} = \beta_0 + \beta_1 HAIL_j + S_j + u_{jt} \tag{2.1}$$

where Y_{jt} is an outcome of interest for cohort t at high school j. $HAIL_j$ indicates assignment to the treatment group, and S_j is a vector of strata indicators. We cluster standard errors at the high school level. As an alternative method of inference, we also report exact p-values from randomization inference (described in Appendix B.1.8). β_1 measures the causal effect of being offered the HAIL Scholarship.

To estimate the causal effect of college quality, we compute local average treatment effects (LATEs) using instrumental variables (IV). The LATE captures the effect of attending a UM campus for students whose enrollment choices were altered by HAIL. We use the randomized HAIL offer as an instrument for enrollment at any UM campus within one year of expected high school graduation. This definition incorporates students induced into either the flagship or the satellite campuses through the deferred-admission pathway.

We estimate the following with two-stage least squares, with UM_{jt} representing the rate of enrollment at any UM campus for cohort t from high school j and γ_1 representing the LATE:

$$UM_{jt} = \alpha_0 + \alpha_1 HAIL_j + S_j + e_{jt} \tag{2.2}$$

$$Y_{jt} = \gamma_0 + \gamma_1 \widehat{UM}_{jt} + S_j + u_{jt} \tag{2.3}$$

The exclusion restriction requires that HAIL affects downstream outcomes only through changes in the college attended. The LATE estimates capture the effect of UM attendance induced by HAIL. These estimates attribute HAIL's effects on downstream outcomes to compliers induced by the offer into a UM campus.

We also estimate control complier means (CCMs), following Imbens and Rubin (1997a). The exclusion restriction, together with absence of defiers, allows us to estimate population shares and outcome means for always-takers (students in the control group

who attend UM) and never-takers (treated students who do not attend UM). Using these shares and means, we back out averages for control compliers. We use the same method to compare always-takers, never-takers, and compliers in Section 2.7. Appendix B.1.9 explains the procedure in more detail.

2.5 The Landscape of College Quality in Michigan

Michigan has no Ivy League college, but the state has substantial variation in college quality. The UM flagship is the best-resourced college in the state and the only large school classified as highly competitive by Barron's. Michigan State leads a second tier of selective public universities. A large set of broad-access colleges fills the rest of the distribution, including Eastern Michigan, Grand Valley State, regional four-year colleges, and 28 community colleges. Low-income high achievers in the HAIL sample are more likely to be in the lower part of this distribution. Upper-income students with the same academic credentials are concentrated nearer the top.

We compare the HAIL control group to students who have the same academic credentials but are ineligible for subsidized lunch ("upper-income"). Low-income and upper-income high achievers sort into different parts of the college-quality distribution (Figure 2.1). HAIL students attend colleges where their classmates average 1148 on the SAT, compared to 1182 for upper-income students. Instructional spending per student averages $16,388 at colleges attended by HAIL students compared to $18,062 for upper-income students. Faculty to student ratios are similar for the two groups. HAIL students go to colleges where 78 percent of students enroll full time, compared to 84 percent for upper-income students. There are also large differences in the likelihood of graduation. At colleges attended by HAIL controls, 54 percent of students graduate within four years compared to 62 percent for upper-income students (60 and 69 percent graduate within six years, respectively).

These are not small differences. Students who satisfy the same high-school achievement criteria enter colleges with meaningfully different peers, resources, and enrollment environments. Sorting across colleges by family income compounds the advantages that upper-income students carry into higher education.

The UM flagship anchors the top of this distribution and defines the quality margin in the paper. Enrolled students average 1360 on the SAT. Instructional spending exceeds

$30,000 per student. Only 3 percent attend part time. Median parental income exceeds $200,000. Gaining admission to the flagship means entering a very different educational environment for low-income, high achievers.

Values for the satellite campuses locate the deferred-admission pathway in this quality landscape. UM-Dearborn and UM-Flint lie well below the UM flagship but are near the middle of the distribution attended by HAIL controls. Treatment-group students steered toward the satellites through deferred admission were therefore not sent into unusually weak colleges but rather into colleges similar to those attended by the control group. This supports the monotonicity assumption underpinning the IV estimates in the paper.

2.6 Intent-to-Treat Effects

2.6.1 Enrollment at a University of Michigan Campus

HAIL shifted where students enrolled, not whether they enrolled in college. Overall college enrollment was 90 percent in both the treatment and control groups (Figure 2.2A). This differs from Dynarski et al. (2021), in which the HAIL offer was estimated to increase overall college attendance by 3.8 percentage points. Using updated NSC records, we no longer find an effect on college attendance.[15] HAIL increased enrollment at any UM campus within one year of expected high school graduation by 21 percentage points (standard error 2.1 pp, F-statistic 100). This first stage combines effects on both flagship and satellite enrollment.

The HAIL offer increased enrollment at the flagship as a freshman by 15 percentage points (Figure 2.2B). This is the enrollment effect we reported in Dynarski et al. (2021), published before we learned of the deferred-admission offer. The deferred-admission pathway increased freshman enrollment at UM-Dearborn or UM-Flint in our sample by 6 percentage points. Including this response raises the estimated effect on UM system enrollment from 15 to 21 percentage points.

[15]The state of Michigan receives data from National Student Clearinghouse each fall which it uses to update its longitudinal college attendance panel. Dynarski et al. (2021) used the first pull of data available for sample students. We later learned that NSC updates its data when colleges report an error in their previous submissions. Five percent of Michigan high school graduates had a correction to their NSC enrollment record for the year after high school. The schools making corrections are more likely to be open-access colleges. In the HAIL sample, the records of 260 students changed.

Half of those induced into the satellite campuses subsequently transferred to the flagship in their sophomore year (+2.9 pp, Appendix Table B.3). Transfers from the satellites to the flagship are essentially absent in the control group (0.1 percent in year two), indicating that the deferred-admission pathway expanded access to the UM flagship.

The treatment-control contrast sheds light on the college choices that students would have made in the absence of the HAIL offer. Of the 21 percentage-point increase in UM enrollment, 4 percentage points came from students who would otherwise have attended community colleges. The other 17 percentage points came from students who would otherwise have attended four-year colleges less selective than the UM flagship.[16] Later we show that these enrollment shifts generated large changes in peer achievement, institutional resources, and degree quality. These estimates also support the monotonicity assumption that underlies the IV analysis.

2.6.2 Time to Degree and BA Selectivity

The HAIL offer accelerated degree completion and increased degree quality. It had no effect on eventual bachelor's degree attainment or STEM completion.

The offer increased four-year BA completion by 4.1 percentage points (standard error 2.2 pp), relative to a control mean of 39 percent (Table 2.2). By six years after expected high school graduation, however, BA attainment is identical in the treatment and control groups at 63 percent. HAIL therefore changed the timing of degree completion rather than the likelihood of ultimately earning a bachelor's degree. This pattern is consistent with the enrollment results: HAIL shifted college choice among students already very likely to enroll in college. We return to these results in the Discussion.

The effect on degree quality, by contrast, is large and persistent. Within six years, treated students are 11.7 percentage points more likely to earn a BA from a highly selective or most selective college. Relative to a control mean of 14.4 percent, the offer increases by 81 percent the likelihood of earning such a selective degree. Within the treatment group, 43 percent of bachelor's degrees are earned from highly selective colleges, compared with 24 percent in the control group.

One concern is that students induced into more competitive academic environments

[16] Greater detail on the enrollment effects of HAIL is available in Dynarski et al. (2021) and Burland et al. (2023).

may avoid demanding, high-return fields of study (Arcidiacono et al., 2016). We find no evidence of this pattern. The HAIL offer has no effect on the likelihood of earning a STEM degree (Table 2.2). Roughly one quarter of students in both groups earn a STEM BA. This finding is unchanged under alternative STEM definitions, including classifications that incorporate business and economics (Appendix Figure B.1).

The four-year completion estimate is imprecise. As we show in the next subsection, transcript records corroborate the attainment results.

2.6.3 Faster Credit Accumulation

Nearly all students at the UM flagship enroll full time. Only 3 percent attend part time, compared with 9 percent at Michigan State University and 68 percent at Grand Rapids Community College (College Scorecard, 2016). These differences in enrollment intensity suggest a mechanism for shortened time-to-degree: full-time students tend to accumulate credits faster than part-time students.

Students in the treatment group accumulate credits at four-year colleges more quickly than those in the control group (Figure 2.3). This difference emerges in the summer after high school, with the treatment group earning 0.34 more four-year credits than the control group (standard error 0.09). The UM flagship has a pre-college summer program for students from disadvantaged backgrounds, which likely explains this result. The credit advantage grows to 1.3 after the freshman year and 1.98 after the sophomore year, then levels off at about 2.5 credits. The estimates for the first two spells are statistically significant at conventional levels; the confidence bands include zero for the later years.

These differences in credit accumulation may seem small, but they correspond to substantial differences in the share of students who are on track to earn a degree within four years. The typical BA in the United States requires 120 credits. Students therefore need to enroll in 30 credits per year to graduate on time. By the end of the first year of college, treated students are five percentage points more likely to have earned 30 credits at a four-year school (relative to a control mean of 31 percent). Over the first four years of college, treated students are three to five percentage points more likely to reach the 30-credit benchmark. This is in line with the estimated effect on four-year BA completion (4.1 pp, Table 2.2), pointing to faster credit accumulation as a central mechanism for the shift toward on-time graduation. This pattern aligns with evidence that early credit accumulation

strongly predicts degree completion (Angrist et al., 2022).

Our credit information is from the STARR data, which covers only students in Michigan's public colleges. Treatment status affects inclusion in STARR: 73 percent of the control group have transcripts in the first year after high school, compared with 77 percent of the treatment group (Appendix Table B.6). These marginal STARR students would otherwise have attended a private college or a college outside Michigan, but HAIL induced them to attend a Michigan public college.

This creates a risk of selection bias in the credit estimates since presence in STARR depends on an outcome downstream of random assignment. As described in the data section, we impute credits for these students for Figure 2.3. As a robustness check, we estimate effects in the unimputed sample and calculate Lee bounds (Appendix B.1.7). The Lee bounds are wide but consistent with the imputed-credit estimates (Appendix Figure B.3).

2.6.4 Reduced Attendance at Community Colleges and AA Degree Attainment

Community college tuition is well below that at four-year colleges (College Scorecard, 2016). The community-college-to-BA pathway is often promoted as a lower-cost route to a bachelor's degree. The largest state aid program for low-income students in Michigan (the Tuition Incentive Program, or TIP) paid full tuition only for community college (Burland et al., 2025).

Evidence suggests that this pathway can slow progress through college and lead to credit loss during transfer (Miller, 2025). In the United States, just 18 percent of those starting at a community college have earned an associate degree (AA) after six years (U.S. Department of Education, National Center for Education Statistics, 2017a). Even fewer (13 percent) earn a BA (U.S. Department of Education, National Center for Education Statistics, 2017a).

As shown earlier, the HAIL offer reduced the share of students starting at a community college by four percentage points (Figure 2.2). Subsequently, HAIL reduced transfers from community colleges into four-year colleges by 1.4 percentage points (Appendix Table B.3). This pattern suggests that some students induced into the UM system would otherwise

have started at a community college and then transferred to a four-year college.[17]

Consistent with this enrollment shift, HAIL reduced AA degree attainment by six percentage points relative to a control mean of 15 percent (Figure 2.4A). The decline reflects reductions in both the probability of earning only an AA and the probability of earning both an AA and BA, each by three percentage points (Figure 2.4B).

At the same time, the offer increased the share of students earning only a bachelor's degree by two percentage points (Figure 2.4B). Combined with the earlier evidence on credit accumulation and four-year completion, these patterns suggest that HAIL diverted students away from the community-college-BA pathway and toward direct enrollment in four-year colleges.

The decrease in AAs, combined with no change in BAs, has ambiguous welfare implications because we cannot observe which counterfactual pathway each affected student would have followed. Some students diverted from community college to the UM system may otherwise have earned an AA and then gone on to earn a less-selective BA. For these students, earlier enrollment in a four-year college in the UM system saves time and increases BA selectivity, predicting higher lifetime earnings.

Other students diverted from community colleges to the UM system may otherwise have earned an AA but now end up with some four-year credits and no BA. For these students, the welfare implications depend on the return to an AA compared with the return to four-year credits without a degree. We cannot assess these welfare implications until actual earnings are observed.

2.6.5 Summary of Intent-to-Treat Effects

Taken together, the intent-to-treat results indicate that HAIL substantially altered students' educational trajectories. The offer shifted students away from community colleges and less-selective four-year institutions, accelerated progress toward a bachelor's degree, increased the selectivity of the degrees they earned, and changed their pathways through higher education. We next ask whose choices changed and how large a quality shock HAIL induced.

[17] We also observe a small increase in transfers away from the flagship. In the second year after high school, transfers from the flagship to colleges outside the UM system increased by one percentage point. Relative to the 15 percentage-point increase in first-year flagship enrollment, this implies that roughly one in fifteen students induced into the flagship subsequently transferred elsewhere.

2.7 Whose College Choices Does a Tuition Guarantee Change?

The local average treatment effects we report later in the paper are identified by compliers, the students whose college choices were altered by the HAIL Scholarship. This section profiles that population, and compares them to the students who would have attended UM regardless of HAIL (always takers) and those who do not attend even when offered the scholarship (never takers).

Compliers differ from always-takers less in academic preparation than in college search and enrollment. They have an average SAT score of 1267, compared with 1288 among always-takers and 1246 among never-takers, and an average GPA of 3.84, compared with 3.89 and 3.81, respectively (Table 2.3).

Score-send data provide a window into the colleges students viewed as plausible options before receiving the HAIL offer. This logic parallels that of Dale and Krueger (2002, 2014), who use application behavior to capture dimensions of student ability and aspiration that are difficult to observe directly.

Before treatment, compliers actively engaged in the college search process. More than half sent scores to the University of Michigan flagship, compared with 36 percent of never-takers and 79 percent of always-takers. Similarly, 65 percent sent scores to a highly or most competitive college, a category that includes the UM flagship, compared with 49 percent of never-takers and 84 percent of always-takers. The average admit rate of colleges receiving score sends is 59 percent for compliers, compared with 66 percent for never-takers and 50 percent for always-takers. Across these measures, compliers fall between the two comparison groups.

Two features of the score-send patterns stand out. First, compliers showed substantial interest in selective colleges: 71 percent sent scores to very competitive colleges and 65 percent sent scores to highly or most competitive colleges. Second, their consideration sets were broader and less centered on the UM flagship than those of always-takers: they were less likely than always-takers to send scores to the UM flagship and sent scores to colleges with higher admit rates. These patterns suggest that compliers considered selective colleges, but with less certainty than always-takers about which selective colleges were within reach.

Geography helps explain whose choices HAIL changed. HAIL had the least effect on those who lived closest and furthest from the flagship. Compliers are more likely than

always-takers to attend rural high schools, more likely to attend high schools outside Southeast Michigan, and more likely to live farther from the UM flagship in Ann Arbor. Fifty-eight percent attend rural schools, compared with 32 percent of always-takers. The average distance from their high school to Ann Arbor is 82 miles, compared with 64 miles among always-takers. These students therefore come disproportionately from parts of the state where the UM flagship may be less salient. Compared with compliers, never-takers are similarly rural (57 percent) and even farther from Ann Arbor (106 miles). Research in other settings shows that distance to campus can shape college access (Acton et al., 2025).

The demographic composition of the complier population largely reflects these geographic patterns. Compliers are more likely than always-takers to be White and less likely to be Black or Hispanic. In Michigan, race and geography are closely linked. Black and Hispanic students are concentrated in urban Southeast Michigan, while the rural regions that contribute many compliers are overwhelmingly White.[18]

The complier analysis identifies a population that is academically strong, actively engaged in the college search process, and largely oriented toward four-year colleges. Yet absent HAIL, they rarely enroll in colleges as selective as the UM flagship. HAIL changes the choices of students who were well prepared for highly competitive colleges and considered selective options, but who would have enrolled in less-selective colleges without the offer.

2.8 Local Average Treatment Effects

2.8.1 The Quality Shock for Compliers

HAIL induced students to attend colleges with substantially stronger peers and greater resources. We measure this quality shock in Table 2.4 by estimating treatment effects on the characteristics of colleges attended. To capture transfers, including those induced by deferred admission, we average college characteristics over the first two years after high school. We report both ITT and LATE effects, focusing on LATEs to highlight the quality shock experienced by compliers.

[18]In the control group, Black urban students in Southeast Michigan have considerably higher rates of application to and enrollment in the UM flagship than do White rural students. As shown in Dynarski et al. (2021), HAIL reduced heterogeneity by region in these outcomes.

Compliers experienced a 187-point increase in peer achievement, shifting from colleges whose students averaged 1128 on the SAT to colleges whose enrolled students averaged over 1300. This change came with large changes in institutional resources, peer characteristics, and graduation rates. Full-time enrollment rises from 76 percent to 90 percent. Instructional spending per student nearly doubles (a gain of $13,000) while endowment per student increases seven-fold, rising $236,000 from a base of $39,000. Faculty resources also increase, with instructional faculty per 100 students rising from 6.5 to 8.

Peer composition changes markedly. For students induced to attend UM, median parental income rises from $133,573 to $196,527. The share of students with a college-educated parent increases from 70 percent to 78 percent, and the share not receiving a Pell Grant rises from 68 percent to 80 percent. The rejection rate at attended colleges increases from 23 percent to 67 percent.

Large differences also appear in institutional graduation rates. The average six-year graduation rate for Pell recipients rises by 32 percentage points, from 52 percent at the colleges attended by control compliers to 84 percent at colleges attended by treated compliers. This explains our null effects on overall degree attainment: control compliers complete college at rates (77 percent, Table 2.2) far above the Pell graduation rates of the colleges they would otherwise have attended.

Taken together, changes to enrollment (Figure 2.2) and institutional characteristics (Table 2.4) show that HAIL induces large changes in the colleges attended by compliers. Under treatment, compliers attend colleges with academically stronger peers, more resources, higher graduation rates, and more advantaged student bodies. The LATE estimates below therefore capture the consequences of a substantial increase in college quality.

2.8.2 Educational Attainment and Degree Quality

We now estimate the consequences of these quality changes for educational attainment and projected earnings. We estimate these effects using the IV strategy described in Section 2.4.2, instrumenting enrollment at any UM campus within one year of expected high school graduation with assignment to the HAIL offer.

Attending a UM campus as a freshman increases the likelihood of earning a BA within four years by 20 percentage points (standard error 9.9 percentage points, Table 2.2). This

represents a large increase over the control complier mean of 42 percent. We find no evidence that these gains translate into higher rates of six-year degree attainment.

A longer time spent attaining a given level of education is costly to both students and society. Tuition is charged by the year and not by the degree. The opportunity cost, in the form of forgone wages, accrues annually. And the later in life a degree is granted, the fewer years there are for it to pay off in the labor force. Researchers have therefore focused intently on the determinants of time-to-degree (e.g., Bound et al., 2012). Our results suggest that time-to-degree is heavily influenced by the quality of the college that a student attends.

The effect on degree quality is large. Students induced by HAIL to attend a UM campus are 56.4 percentage points more likely to earn a highly selective BA. The standard error is 6.2 percentage points. Relative to the control complier mean of 4.9 percent, this represents an eleven-fold increase in the likelihood of earning a highly selective or most selective degree.

As in the ITT estimates, HAIL changed degree timing and quality, not eventual BA attainment. In Appendix Table B.4, we further show no effect on graduate school enrollment or attainment.

2.8.3 The Satellite Pathway

To assess the satellite pathway, we construct joint enrollment-attainment outcomes that combine students' first UM campus with the degree they ultimately earn.[19] These outcomes, reported in the middle panel of Table 2.2, show whether students who begin at a UM satellite later earn any BA and whether they ultimately earn a BA from a highly competitive institution. The joint outcomes suggest that the satellite pathway was not simply a diversion into lower-quality colleges without later access to selective degrees. Among treated compliers, 19 percent start at a UM satellite and earn a BA within six years; 11 percent start at a satellite and ultimately earn a highly competitive BA. Thus, a substantial share of students entering through the satellite pathway ultimately complete high-selectivity degrees.

[19] Joint enrollment-attainment outcomes are unconditional on initial enrollment. They are set to one when both of the underlying outcomes are equal to one.

2.8.4 Projected Earnings Effects

Because our cohorts are still young, we do not yet observe their earnings. We therefore project earnings using institution-level earnings data from College Scorecard (College Scorecard, 2021). Eight years after college entry, students who attended the UM flagship earn an average of $86,000, compared with $69,000 at Michigan State University and $38,000 at Michigan's community colleges.

The ITT effect on projected earnings is $3,400 (last panel of Table 2.2). Attending any UM campus increases compliers' projected earnings by $16,300 (standard error $4,000). This is 26 percent of the control complier mean of $62,300. In national data, this difference is equivalent to moving from the 52nd to the 66th percentile of the earnings distribution among those with a BA.[20]

These projections are based on average earnings at each college. HAIL students, however, come from lower-income families than the typical college student. We therefore also project earnings using data on students from families with incomes below $30,000. These projections imply larger effects. The ITT effect is $3,600, and attending any UM campus increases projected earnings of compliers by $17,100 (standard error $3,500). This is 31 percent of the control complier mean of $55,402. In national data, this difference is equivalent to moving from the 45th to the 61st percentile of the earnings distribution among those with a BA.[21]

These projected earnings estimates should be interpreted with caution. Earnings differences across colleges reflect both college quality and non-random sorting of students into colleges. Recent estimates suggest that roughly 70 to 80 percent of the variation in institutional earnings can be attributed to the causal effects of college quality (Bleemer and Quincy, 2025). To the extent that these estimates apply in our setting, they imply a causal, annual earnings effect of 18 to 25 percent.

The reduced time to degree induced by the increase in college quality also boosts projected earnings. An additional year in the labor force at the very start of a worker's career

[20]These calculated percentiles are based on the earnings of those aged 28–32 with at least a bachelor's degree in the March Current Population Survey in the years 2014 through 2017. All values are converted to 2025$.

[21]The corresponding estimate based on earnings of students from families with incomes above $75,000 is a 22 percent increase relative to the complier mean (Appendix Table B.5). This pattern suggests that the payoff to attending a more selective college may be larger for low-income students than for their higher-income peers.

is particularly valuable, adding 4–5 percent to the present discounted value of lifetime earnings at age 21.[22] Since attending a UM campus increases the likelihood of graduating within four years by 20 percentage points, this implies that earlier graduation contributes roughly 1 percent to expected earnings across all compliers.

2.8.5 Plausibility of the Exclusion Restriction

The exclusion restriction requires that, among students who enroll at UM, HAIL affects attainment and earnings only through the UM enrollment decision, not through HAIL-specific differences in aid, services, or supports after enrollment. If HAIL changes outcomes for always-takers after they enroll at UM, the LATE estimates will attribute their response to compliers and therefore overstate the causal effect of college quality. The ITT estimates do not rely on the exclusion restriction and remain valid as average effects of the HAIL offer.

A natural concern is that treated students who enrolled at UM received more financial aid than comparable control students who enrolled at UM. They did not. Treated and control students who enrolled at the flagship received nearly identical grant aid in both the first year and over four years (Appendix Table B.2). HAIL students also received no special academic, advising, or social support after enrollment; they had access to the same programs and services as other UM students. The intervention changed when students learned about aid, not how much aid they received, and did not alter the services available to them after enrollment.

The main remaining concern is certainty about college costs. Always-takers in the treatment group received an up-front, four-year guarantee of tuition and fees, while always-takers in the control group learned about comparable aid through the standard admissions and financial-aid process and had to reapply every year. We argue in Burland et al. (2023) that this certainty, rather than greater generosity of aid, changed application and enrollment decisions.

The exclusion restriction assumes that any effect of certainty ends once students enroll. This assumption would hold if students learned about the generosity of UM financial aid

[22]This calculation assumes a discount rate of 3 percent, a career that ends at 67 and starts at either 21 or 22, and no real increase in annual earnings. If real earnings are instead allowed to rise by 2 percent a year, an extra year is worth 5 percent of lifetime discounted earnings (since the annual increase in earnings also starts a year earlier). Since attending a UM campus increases four-year graduation by 20 percentage points, the expected gain across all compliers is approximately 0.20 times 4 to 5 percent, or about 1 percent.

either through HAIL or through the standard aid process. The flagship meets the financial need of all admitted students, and aid is not conditioned on academic performance. A low-income student who enrolls and begins interacting with the financial aid office may therefore develop the same understanding of aid generosity regardless of HAIL status.

Alternatively, cost certainty may affect student behavior after enrollment. Students who receive a four-year guarantee may experience less financial stress, work fewer hours while enrolled, or focus more intensely on completing a degree on time. The guarantee may also encourage students to think in terms of a four-year completion horizon.

We think these channels are unlikely to explain the main results. Students who remain enrolled beyond four years are eligible for financial aid that is as generous as the HAIL offer. The guarantee therefore does not create a sharp financial incentive to graduate within four years.

We cannot separately identify effects for compliers and always-takers. If always-takers respond positively to the HAIL guarantee after enrollment, the LATE estimates should be interpreted as upper bounds on the effect of college quality for compliers.

2.9 Discussion

In earlier work, we predicted that HAIL would increase bachelor's degree attainment by roughly seven percentage points (Table 8 of Dynarski et al., 2021). Instead, we find large effects on four-year completion and degree selectivity, but no effect on bachelor's degree attainment after six years. This section explains why.

Our earlier prediction was not an impact estimate from HAIL. It used prior cohorts of high-achieving, low-income Michigan students to project how enrollment at more selective colleges might affect completion. The updated results show why that projection was too large: HAIL compliers are a selected subset of an already selected population. They were high-achieving, applied to UM, were admitted, and changed their enrollment decision in response to the offer.

Two new findings help explain why we overpredicted the effect of HAIL on bachelor's degree attainment. First, in Dynarski et al. (2021), HAIL appeared to increase overall college enrollment by 3.8 percentage points; using updated NSC records, we no longer

find an effect.[23] Given that more than three-quarters of enrollment compliers earn BAs even absent treatment, this revision alone implies that the earlier data underpredicted BA completion in the control group by roughly three percentage points.

Second, the complier analysis reveals that the students whose enrollment decisions changed because of HAIL completed degrees at much higher rates than we would have predicted based on the institutional graduation rates of their counterfactual schools. Within six years of entering college, 77 percent of control compliers earn a bachelor's degree (Table 2.2). Yet the colleges they attended had an average Pell-recipient graduation rate of just 52 percent (Table 2.4). This suggests that the students induced to enroll at a UM campus were highly likely to complete degrees, even in the absence of treatment. This likely reflects both their strong academic preparation and unobserved strengths that matter for college success, including characteristics observed by UM admissions officers.

Peer-achievement data help clarify why this result differs from a simple low-income-student narrative. Because Michigan universally administers the SAT in public high schools, we can calculate peer achievement even at open-access colleges. Figure 2.5A shows that low-income students statewide typically attend colleges where their SAT scores fall below the peer mean. The HAIL sample is different. These students are low-income but high-achieving. Figure 2.5B shows that 83 percent of control-group HAIL students scored above the average SAT score of their college peers, compared with 75 percent of treatment-group students. HAIL shifted students into colleges with stronger peers, but it did not push them into academic overmatch. It reduced undermatch while leaving most students academically well prepared relative to their peers.

For academically strong students, attending a higher-quality college may change the timing and quality of the degree rather than whether they complete a BA. This distinction matters for how we evaluate college quality: degree completion alone misses shifts in time-to-degree, degree selectivity, and projected earnings.

The credit and pathway results point to the mechanism: HAIL moved students away from slower routes through community colleges and less-selective institutions and into colleges where full-time enrollment and faster credit accumulation are more common.

This pattern helps reconcile our estimates with the broader literature. Studies that move students from no college or community college into four-year colleges often find gains in bachelor's degree completion. By contrast, HAIL compliers were already highly likely to

[23] See Section 2.6.1 for a detailed explanation.

attend and complete college. Our results are closer to studies that examine movement among four-year colleges or among highly prepared students: the gains appear less in eventual degree completion than in the quality, timing, and projected labor-market value of the degree earned. This comparison underscores why the population and quality shock matter when interpreting estimates of college quality.

2.10 Conclusion

We analyze a randomized intervention that shifted high-achieving, low-income students into higher-quality colleges. HAIL gave students a simple, early, credible guarantee of affordability before they applied. By moving information about aid to the front of the college-choice process, the intervention changed where students enrolled. Because HAIL left post-enrollment aid and services unchanged, the experiment provides unusually clean evidence on the consequences of attending a higher-quality college.

The students induced to enroll at a UM campus were already likely to attend college and complete a BA. HAIL therefore changed the quality and timing of the degree they earned, rather than whether they earned a degree. The intervention increased four-year BA completion, sharply increased the likelihood of earning a BA from a highly or most selective college, and raised projected earnings. The college quality premium implies causal gains of 18 to 25 percent; earlier graduation adds roughly 1 percent more. The intervention shifted high-achieving, low-income students toward the types of degrees typically earned by their upper-income counterparts (Figure 2.6).

The results also speak to concerns about academic mismatch. The students induced to enroll at a UM campus were not marginal admits. They were academically strong, scored above the average student at the colleges they would otherwise have attended, and resembled students who enrolled at the UM flagship. HAIL reduced undermatch while moving students into substantially higher-quality colleges. This pattern aligns with evidence that academic match and college quality can improve outcomes for well-prepared students (Dillon and Smith, 2020; Bleemer, 2022).

The lesson has limits. Better information, simpler applications, and clearer aid promises can move students toward higher-quality colleges. But they cannot create more high-quality seats. Abdulkadiroğlu et al. (2026) make a similar point in K12 school choice: allocation policy can improve who gets access to strong schools, but it works within a fixed

supply of those schools. HAIL reduced undermatch by reallocating access to high-quality colleges. Mobility at scale requires expanding that supply.

The constraint matters especially in higher education. Many elite colleges have responded to rising demand by becoming more selective rather than by adding seats (Blair and Smetters, 2021). UM Ann Arbor is a counterexample: its entering first-year class has grown by roughly one-third since HAIL began.[24] Mobility at scale requires more of this kind of capacity growth: not just better access to high-quality colleges, but more seats at institutions that reliably generate degrees, earnings, and upward mobility.

[24]UM Ann Arbor's first-year class grew from 6,071 students in fall 2015 to 8,178 in 2025 (University of Michigan Office of the Registrar, n.d.; University Record, 2025).

Table 2.1: High-School-Level Summary Statistics and Treatment-Control Balance

Characteristic	Control mean	Treated mean	P-value
High school in Southeast MI	0.401	0.394	0.864
	(0.029)	(0.030)	
Town/rural school	0.530	0.540	0.838
	(0.030)	(0.030)	
Suburban school	0.340	0.360	0.592
	(0.028)	(0.029)	
City school	0.129	0.100	0.273
	(0.020)	(0.018)	
Distance of high school from UM (miles)	93.2	96.4	0.597
	(4.691)	(4.824)	
Proportion eligible for free lunch	0.709	0.692	0.450
	(0.015)	(0.015)	
Proportion eligible for reduced-price lunch	0.291	0.308	0.450
	(0.015)	(0.015)	
Proportion female	0.571	0.605	0.102
	(0.016)	(0.015)	
Proportion White	0.772	0.787	0.566
	(0.017)	(0.016)	
Proportion Black	0.094	0.087	0.685
	(0.013)	(0.011)	
Proportion Asian American	0.061	0.057	0.718
	(0.008)	(0.009)	
Proportion Hispanic	0.053	0.057	0.681
	(0.008)	(0.009)	
Average SAT	1254	1259	0.200
	(2.805)	(3.017)	
Average GPA	3.82	3.83	0.263
	(0.006)	(0.006)	
Proportion sent ACT scores to UM flagship (first cohort only)	0.459	0.453	0.835
	(0.021)	(0.022)	
F-test p-value		0.449	
Number of school-years	526	500	1,026
Number of students	1,978	1,932	3,910

Notes: Analysis conducted at the high school-by-cohort level. Each p-value is from a t-test of the coefficient on treatment status from a regression of the characteristic on treatment and strata dummies. Race/ethnicity categories are mutually exclusive. The omitted race/ethnicity category includes students who are American Indian, Alaska Native, or Native Hawaiian. The first cohort took the ACT, while the second took the SAT; we convert to the SAT scale for consistency. Score send data is incomplete for the second cohort, so we calculate the share sending scores to UM for the first cohort only. F-test is from a joint significance test predicting treatment based on the characteristics listed here (except sending scores to UM) as well as strata dummies. Standard errors clustered at the high-school-level in parentheses.

Table 2.2: Estimated Effects on Degree Attainment and Projected Earnings

	ITT		LATE	
	Effect of HAIL offer	Control mean	Effect of Attending Any UM Campus	Control complier mean
First Stage				
Enroll Any UM Campus within 1 year	0.208	0.149		
	(0.021)			
F-statistic	100.141			
ITT and LATE				
BA Attainment				
Any BA within 4 years	0.041	0.391	0.197	0.420
	(0.022)		(0.099)	
Any BA within 6 years	−0.006	0.625	−0.030	0.765
	(0.021)		(0.104)	
Highly competitive+ BA within 6 years	0.117	0.144	0.564	0.049
	(0.019)		(0.062)	
STEM BA within 6 years	−0.010	0.257	−0.050	0.424
	(0.019)		(0.091)	
Joint Enrollment + BA Attainment (*Satellite Pathway*)				
Start at UM Satellite and earn any BA within 6 years	0.040	0.021	0.192	0.000
	(0.009)		(0.041)	
Start at UM Satellite and earn highly comp.+ BA within 6 years	0.023	0.001	0.112	0.000
	(0.005)		(0.022)	
Projected Earnings (*2025$*)				
Projected using institutional average	3,393	58,523	16,321	62,342
	(985)		(3,957)	
Projected using students with family income $\leq$ \$30K	3,557	52,569	17,109	55,402
	(912)		(3,524)	
N school-cohorts	1,026			

Notes: Analysis conducted at the high-school-by-cohort-level. Robust standard errors reported in parentheses. ITT coefficients are estimates of $\hat{\beta}_1$ in Equation (2.1), estimated using OLS. LATEs are estimates of $\hat{\gamma}_1$ in Equation (2.3), estimated using 2SLS. Degree outcomes measured by the end of the fourth or sixth summers following high school, and are not conditional on initial enrollment. Control complier means are calculated at the student level, inverse weighted by size of high school cohort. Institution earnings are median earnings 8 years post college-entry from College Scorecard (2013 entry cohort). Projected earnings are assigned to students based on their modal college attended. Students with two modal institutions are assigned the average for the two modal institutions. College Scorecard earnings are only available for students who enrolled in college. Students who did not enroll or who attend institutions where earnings are not available are assigned earnings from non-attendees in Opportunity Insights data. Earnings are all measured in 2025 values, adjusting for inflation using the consumer price index (CPI-U).

Table 2.3: Pre-Treatment Characteristics by Compliance Sub-Populations

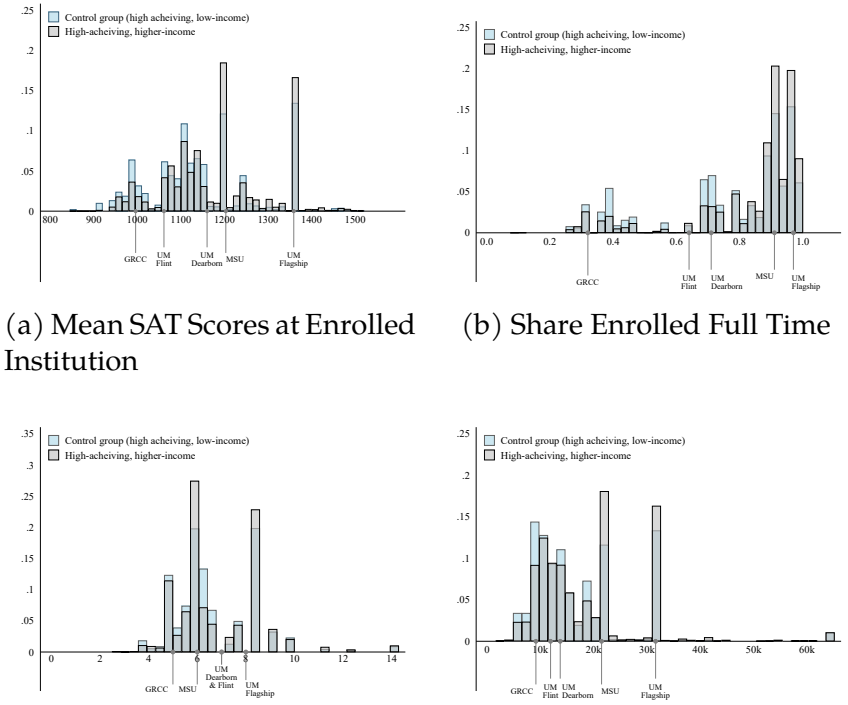
	Full sample	Always-takers	Never-takers	Compliers
Population share	1.000	0.149	0.644	0.207
Geography				
High school in Southeast	0.398	0.640	0.331	0.430
Distance of high school from UM (miles)	94.7	63.5	106.0	82.4
Urbanicity				
Town/rural school	0.535	0.317	0.571	0.580
Suburban school	0.350	0.459	0.344	0.291
City school	0.115	0.224	0.085	0.130
Demographics				
Female	0.588	0.612	0.600	0.532
White	0.779	0.597	0.802	0.840
Black	0.091	0.203	0.082	0.039
Asian American	0.059	0.105	0.044	0.072
Hispanic	0.055	0.078	0.059	0.025
Academic Preparation				
SAT composite	1257	1288	1246	1267
GPA	3.83	3.89	3.81	3.84
N students		3,910		
N school-cohorts		1,026		
Pre-treatment college signaling (first cohort only)				
Sent scores to UM flagship	0.456	0.786	0.361	0.544
Sent scores to very competitive school	0.614	0.587	0.589	0.706
Sent scores to highly competitive+ school	0.567	0.839	0.487	0.646
Mean admit rate of score sends	0.625	0.497	0.663	0.587
N students, first cohort		2,108		
N schools, first cohort		529		

Notes: This analysis is done at the student level and inverse weighted by size of high school cohort. Shares and means are calculated following Imbens and Rubin (1997a) and Marbach and Hangartner (2020), described in Appendix B.1.9. Always takers are defined as students in the control group who attend a UM campus within one year of high school graduation. Never takers are students assigned to treatment who do not attend UM. Score send data is incomplete for the second cohort; analysis using score send data uses the first cohort only (including the calculation of population shares).

Table 2.4: Estimated Effects on Characteristics of Colleges Attended (College-Attending Students Only)

	ITT		LATE	
	Effect of HAIL offer	Control mean	Effect of Attending Any UM Campus	Control complier mean
First Stage				
Enroll Any UM Campus within 1 year	0.208	0.149		
	(0.021)			
F-statistic	101.043			
ITT and LATE: Characteristics of the Institutions Attended				
Average SAT of enrolled Michigan students	40	1143	187	1128
	(6)		(21)	
Share enrolled full time	0.031	0.775	0.147	0.756
	(0.009)		(0.041)	
Instructional expenditures per FTE student	2,794	16,170	13,152	15,045
	(443)		(1,399)	
Endowment per FTE student	50,136	71,215	236,031	38,692
	(6,432)		(18,612)	
Instructional faculty per 100 students	0.324	6.658	1.528	6.479
	(0.072)		(0.293)	
Median parental income	13,372	140,125	62,954	133,573
	(2,018)		(6,470)	
Share with parent who graduated college	0.019	0.702	0.087	0.695
	(0.005)		(0.020)	
Share not receiving a Pell Grant	0.025	0.674	0.116	0.679
	(0.006)		(0.023)	
Rejection rate (1−acceptance rate)	0.094	0.284	0.441	0.226
	(0.012)		(0.034)	
Average 6-year BA rate, Pell recipients	0.068	0.559	0.317	0.519
	(0.013)		(0.047)	
N students			3,652	
N school-cohorts			999	

Notes: Analysis conducted at the high-school-by-cohort-level. Robust standard errors reported in parentheses. This analysis excludes students who do not enroll in college in at least one of the two years following high school. ITT coefficients are estimates of $\hat{\beta}_1$ in Equation (2.1), estimated using OLS. LATEs are estimates of $\hat{\gamma}_1$ in Equation (2.3), estimated using 2SLS. Control complier means are calculated at the student level following Imbens and Rubin (1997a), described in Appendix B.1.9, and inverse weighted by size of high school cohort. Students are assigned an average of the characteristics for their first attended institution in years 1 and 2. Dollar amounts are adjusted with the CPI-U to 2025 dollars. Variable construction is described in more detail in Appendix B.1.6.



(a) Mean SAT Scores at Enrolled Institution

(b) Share Enrolled Full Time

(c) Instructional Faculty per 100 Students

(d) Instructional Expenditures per FTE Student

Figure 2.1: Distribution of Characteristics of Colleges Attended by High-Achieving Students in Michigan, by Family Income

Notes: Higher-income, high achieving students are in the same cohorts and meet the same academic criteria as HAIL students. Distributions are at the student-level. Students are assigned the characteristics of the first college attended after high school. Institution SAT is the mean for pre-HAIL Michigan public high school students. All other institutional characteristics are from IPEDS and College Scorecard (2015–16), described in more detail in Appendix B.6. Dollar amounts are adjusted with the CPI-U to 2025 dollars.

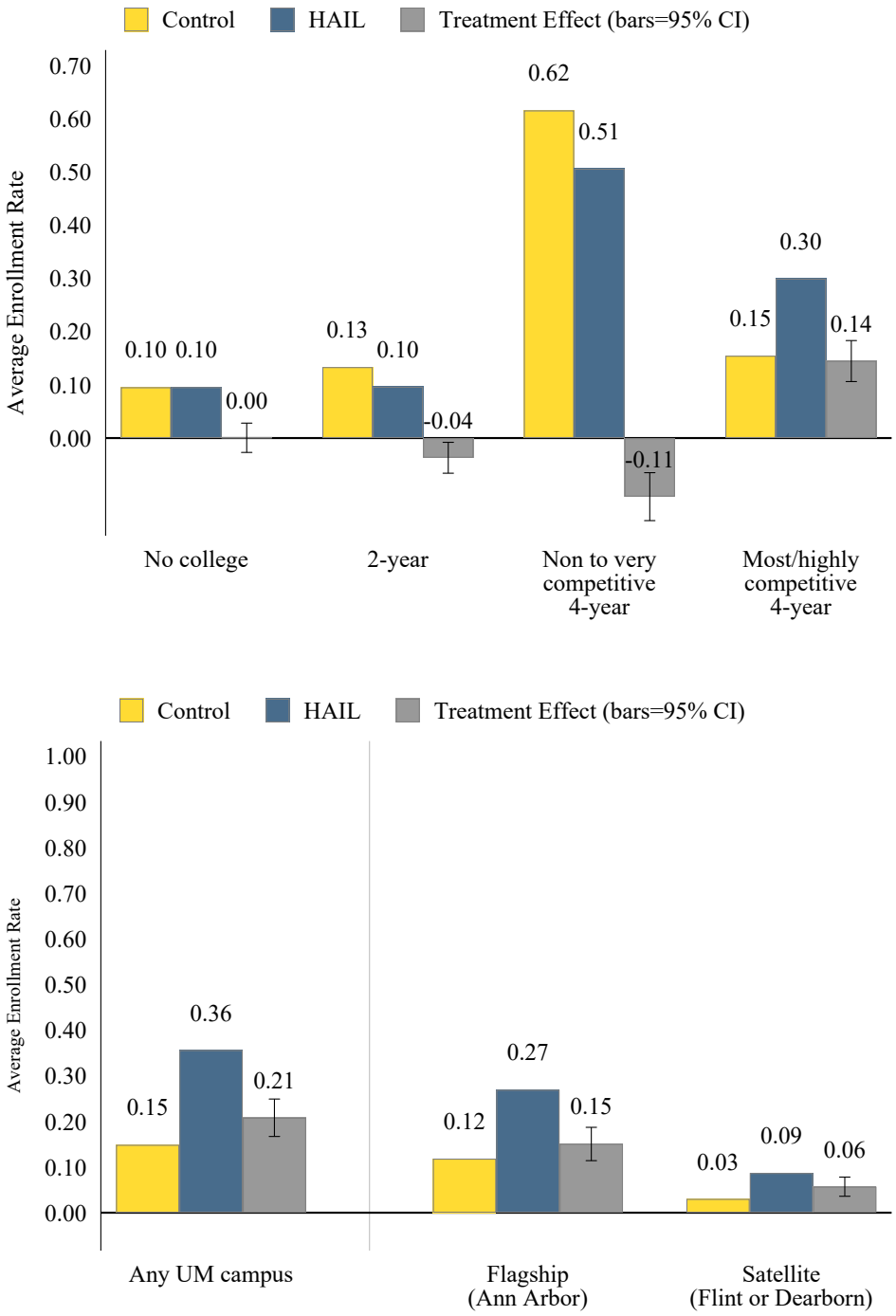
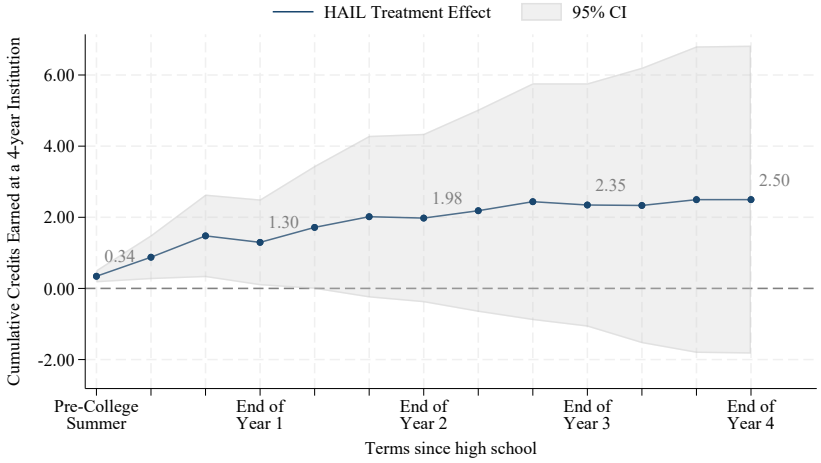
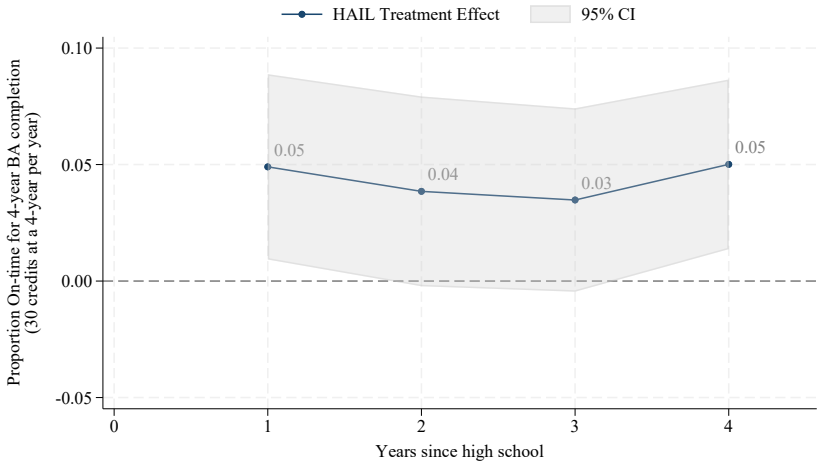


Figure 2.2: Estimated Effect of HAIL on First-Year College Enrollment

Notes: Figure plots the rate of enrollment in the first year following high school by Barron's institution competitiveness, and by UM flagship (Ann Arbor) and UM Satellites (Flint or Dearborn). This analysis is conducted at the high-school-by-cohort-level. Treatment effect coefficients are estimates of $\hat{\beta}_1$ in Equation (2.1). Ninety-five percent confidence intervals on treatment effects are based on robust standard errors.



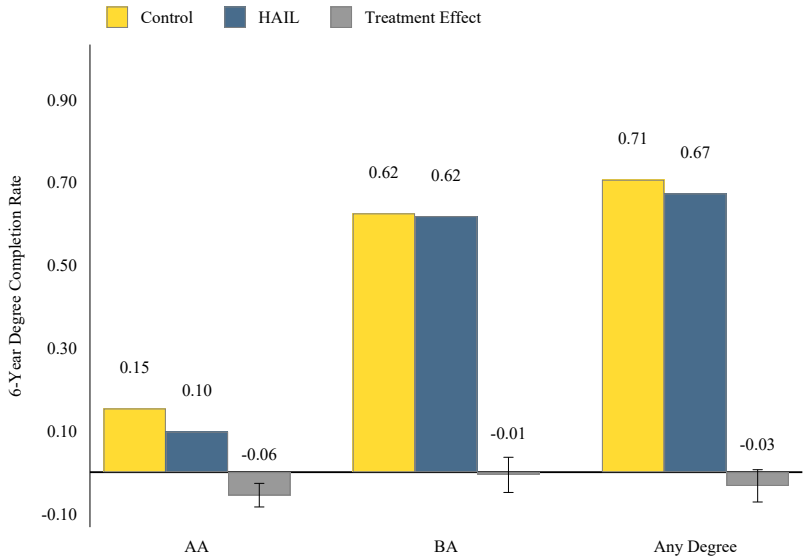
(a) Cumulative Credits Earned



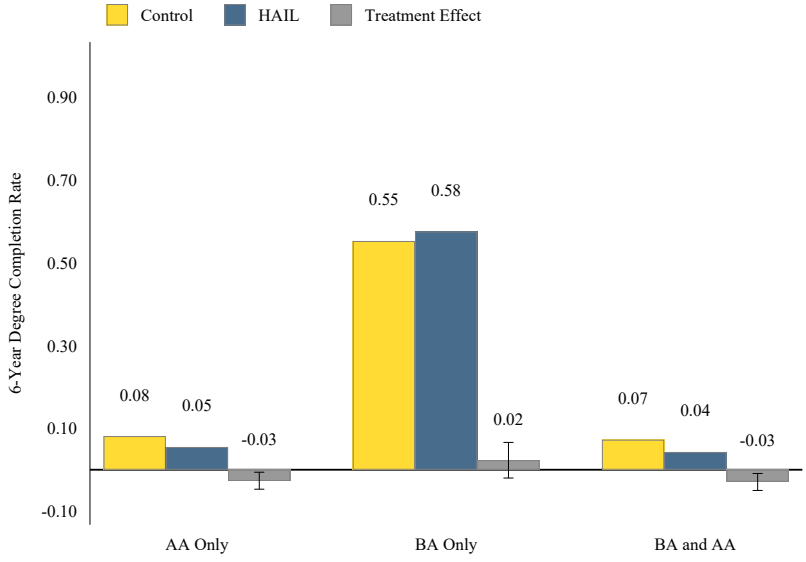
(b) On-track for 4-year BA Completion

Figure 2.3: Estimated Effect of HAIL on Credit Accumulation at 4-year Institutions

Notes: This analysis is conducted at the high-school-by-cohort-level. Robust standard errors are reported in parentheses. Treatment effect coefficients are estimates of $\hat{\beta}_1$ in Equation (2.1). Credits are calculated at the term-level then aggregated to the student-year level. Measured credits include undergraduate coursework taken at a given institution. Annual credits are measured at the end of the summer term and include fall, spring, and summer credits. Year 0 includes the summer between high school and fall 1. Credits are imputed for students who have an undergraduate enrollment record in NSC but are missing credit information in STARR using a regression of the credit outcome on selectivity, enrollment-status, years since high school, and semester-type (fall, spring, summer) for control students with non-missing credits, then summed by year.



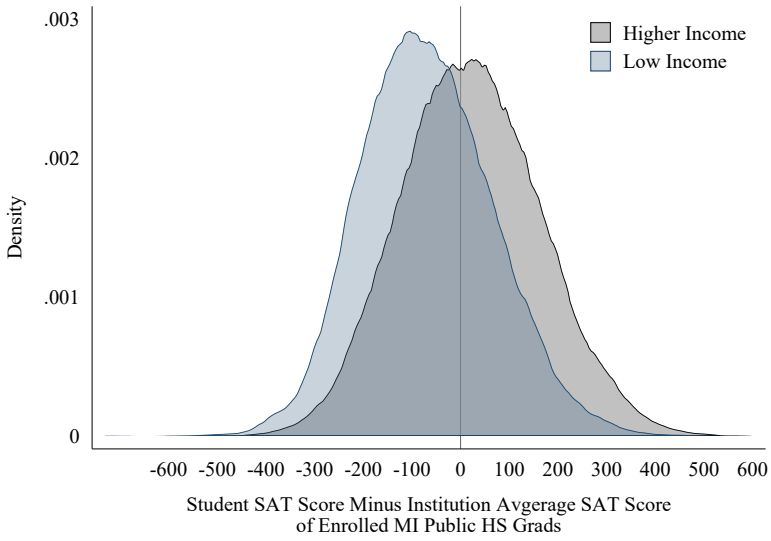
(a) Any Degree Attainment



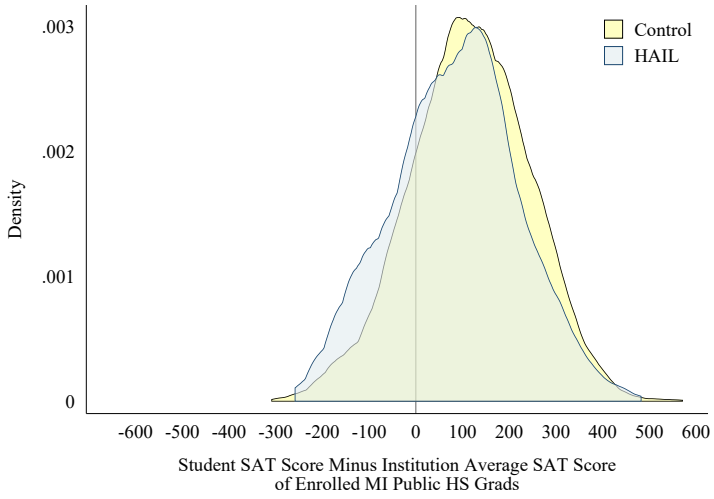
(b) Degree Combinations

Figure 2.4: Estimated Effect of HAIL on Degree Attainment

Notes: Figure plots the rate of degree attainment within 6 years of high school. This analysis is conducted at the high-school-by-cohort-level. Treatment effect coefficients are estimates of $\hat{\beta}_1$ in Equation (2.1). Ninety-five percent confidence intervals on treatment effects are based on robust standard errors.



(a) All Michigan Students



(b) HAIL Treated and Control Students

Figure 2.5: Average SAT Scores Relative to the Average SAT Score of Institutional Peers

Notes: Figures plot the difference between a student's SAT score and the average SAT at the institution they attend. Distributions are at the student-level. Institution mean SAT is calculated using the test scores of Michigan public school students finishing high school in 2013–2014 who attend a given institution in the first year following high school. All ACT scores converted to SAT scale using concordance tables provided by the College Board. For institutions with fewer than 20 Michigan public high school students enrolled, the average score for institutions grouped by selectivity and public/private status. Students not enrolled during the first year are excluded. Panel A includes all Michigan public high school students expected to graduate in the same cohorts as the HAIL sample (2016–2017), excluding those who received the HAIL treatment, not restricted by academic achievement.

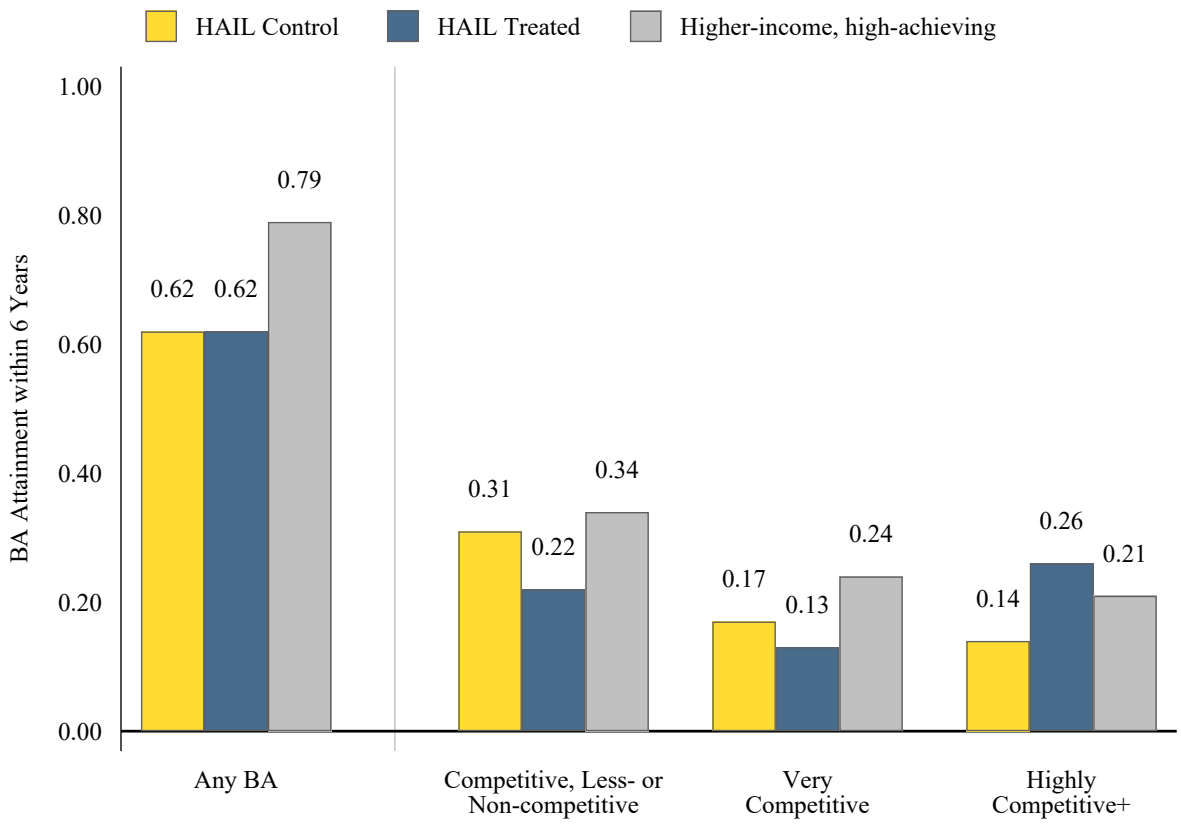


Figure 2.6: Selectivity of Bachelor's Degrees Earned by High-Achieving Michigan Students, by Income and HAIL Treatment Status

Notes: Sample is eleventh grade students in Michigan public schools in 2015 and 2016 who meet HAIL GPA and ACT/SAT criteria. Higher-income students are those not eligible for free- or reduced-price lunch. Degree selectivity is based on Barron's competitiveness index. Degree attainment rates for HAIL treatment and control groups use our high-school-by-cohort-level measures. Average degree attainment for higher-income students is calculated at the student level.

CHAPTER 3

Gas Prices and the Safety Net

Abstract

In this paper, I examine how changes in gasoline prices affect households' use of government support, in particular, the Supplemental Nutrition Assistance Program (SNAP). Using publicly available administrative data on SNAP use linked with state-level prices for gasoline, I estimate how fluctuations in household expenses shape enrollment. Results show that a ten-cent increase in gasoline prices is associated with a 0.8% rise in SNAP enrollment in the following month, controlling for grocery costs and local labor market conditions. These findings suggest that households respond quickly and directly to changes in volatile, unavoidable expenses. This work expands our understanding of program dynamics by showing that demand for assistance is not only determined by income and employment, but also by the shocks households face in meeting basic needs.

3.1 Introduction

The Supplemental Nutrition Assistance Program (or SNAP, formerly known as food stamps) is among the largest social safety net programs in the United States, serving 42 million participants per month in 2023. Although participation among eligible households is estimated to be high on average (81% nationally in 2019, the last available pre-COVID-19 pandemic year), this estimate implies that approximately 10 million Americans who qualify do not participate in SNAP every month (Cunnyngham, 2023). To better understand the factors preventing higher social safety net program participation, research has focused on administrative burden, stigma, lack of information, and business cycles (Moffitt, 1983;

Daponte et al., 1999; Currie, 2006; Bitler and Hoynes, 2010; Ganong and Liebman, 2018; Gray, 2019; Herd and Moynihan, 2019). Experimental evidence suggests that these barriers impede SNAP participation and, indeed, facilitating application to SNAP increases take-up (Finkelstein and Notowidigdo, 2019).

While researchers have studied how business cycles and unemployment affect social safety net program participation, the role of non-food budget shocks is understudied. Macroeconomists have found that gasoline prices in particular affect consumer sentiment and spending (Edelstein and Kilian, 2009; Coibion and Gorodnichenko, 2015; Gelman et al., 2020; Kilian and Zhou, 2022). Further, Hastings and Shapiro (2018) find that when gasoline prices increase, households receiving SNAP increase fuel expenditures, indicating a tightening budget for low-income households. Gasoline price shocks may affect SNAP application behavior due to these documented effects on low-income households' budgets.

The federal baseline for SNAP eligibility is 130% of the poverty line, though states often have higher income limits in addition to asset limits.[1] The majority of households who are likely to be at the margin of SNAP eligibility, that is, near 130% of the federal poverty rate, own personal vehicles (U.S. Department of Transportation, Bureau of Transportation Statistics, n.d.). Only 11-30% households at the lowest two quintiles of the income distribution lack a car, with this rate decreasing by income quintile. Gasoline is an expense that is essential for most households but has volatile prices, is not directly covered by a social safety net program, and has an outsized influence on consumer sentiment.

It is not clear whether increasing prices should drive eligible households to apply for SNAP or discourage them from doing so. Low-income households face costs to applying to SNAP that are increasing in prices of necessities like energy, leading to the "Heat or Eat" relationship found in the literature (Bhattacharya et al., 2003; Beatty et al., 2014). Further, utility from receiving SNAP is also decreasing in prices of necessities, i.e., SNAP is less valuable in real terms if the price of necessities has gone up while SNAP benefits remain unchanged. In the case of gasoline, a change in price may also be especially salient and change individual perceptions of their own economic health. Nevertheless, the net effect of rising gasoline prices is an empirical question.

In this paper, I measure the relationship between changes in the price of gasoline and the use of SNAP, controlling for unemployment rates, inflation, and grocery prices us-

[1] Many states also have automatic eligibility for individuals enrolled in other social safety net programs or for seniors and people with disabilities (U.S. Department of Agriculture, Food and Nutrition Service, n.d.).

ing SNAP participation counts from the Census Bureau, gasoline prices from the Energy Information Administration, and quarterly food prices from the US Department of Agriculture. For the primary analysis, I regress SNAP counts in a month and state on gasoline prices and covariates in the previous month. I include a variety of location and time fixed effects across regression models to account for state- and time-invariant factors that may affect use of food assistance programs. Additionally, I employ an instrumental variables approach, using state gasoline taxes to instrument for gasoline prices. Finally, I conduct an analysis that directly includes changes in grocery prices because these price changes are likely to impact SNAP participation and may fluctuate independently of overall inflation.

The results of the main analysis show that a one standard deviation increase in gasoline prices (equal to $0.83) is associated with a 7% or 55,000 person increase in SNAP use in the following month. This finding is statistically precise and robust to the choice of covariates and fixed effects. The instrumental variables approach results in imprecise estimates with confidence intervals that include the OLS estimate. The regression analyses that include food inflation indicate that the relationship between gasoline prices and SNAP use does not flow strictly through food prices, having estimates of similar magnitude and precision to those in the main specification.

This study suggests that SNAP supports households facing energy price shocks. Having a clearer understanding of this driver of application behavior can serve policymakers working on to reach SNAP-eligible households, particularly during periods with high gasoline prices.

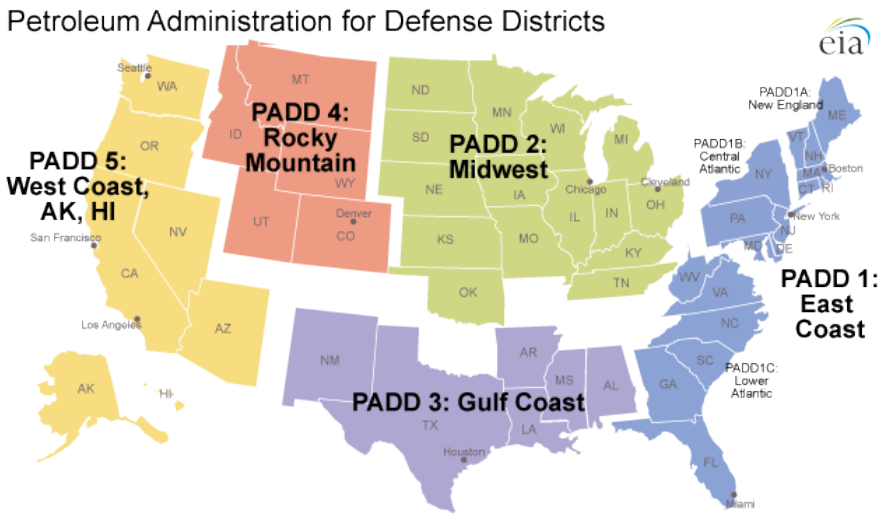
3.2 Data and Empirical Strategy

The analysis sample combines gasoline price and tax data from the Energy Information Administration (EIA) and food assistance and economic variables from the Census Small Area Income and Poverty Estimates. Gasoline prices from the EIA are at the Petroleum Administration for Defense District (PADD) (or sub-PADD) and month level. The United States has five PADDs encompassing multiple states as shown in Figure 3.1 (U.S. Energy Information Administration, 2011): East Coast (which is further divided into three sub-PADDs), Midwest, Gulf Coast, Rocky Mountain, and West Coast (including Alaska and Hawaii). To impute the state-level gasoline price, I first subtract the monthly PADD-level average gasoline tax. I do this because the PADD-level gasoline prices include taxes. Then, for each state, I add the monthly state-level gasoline tax. This imputation is done due to

data availability and is likely to introduce measurement bias. Figure 3.3a shows the raw time series of nominal gasoline prices at the PADD level. Figure 3.3b shows this series with price inflation included while Figure 3.3c displays the estimated state-level time series.

SNAP participant counts and unemployment rates per state and month come from the US Census Bureau Small Area Income and Poverty Estimates. Per capita SNAP enrollees are plotted in Figure 3.2. This figure shows the cyclicality of SNAP enrollment, with increases during the 2008 and 2020 recessionary periods. Each line in this figure represents the per capita SNAP rate for a state, showing the large differences in SNAP enrollment between states.

Figure 3.1: Map of PADDs



Note: This map shows the five Petroleum Administration for Defense Districts (PADDs) in the United States. Source: Energy Information Administration, https://www.eia.gov/todayinenergy/detail.php?id=4890

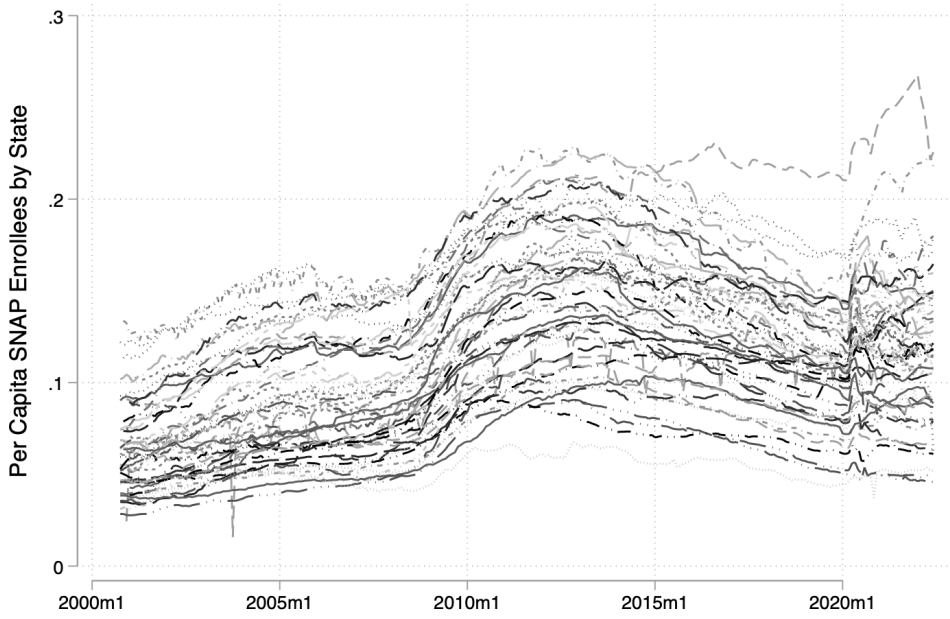
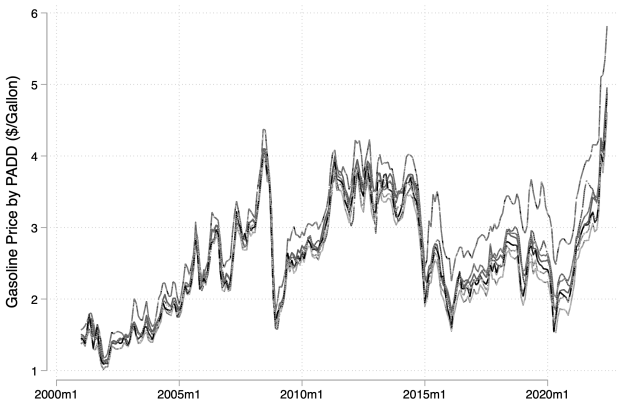
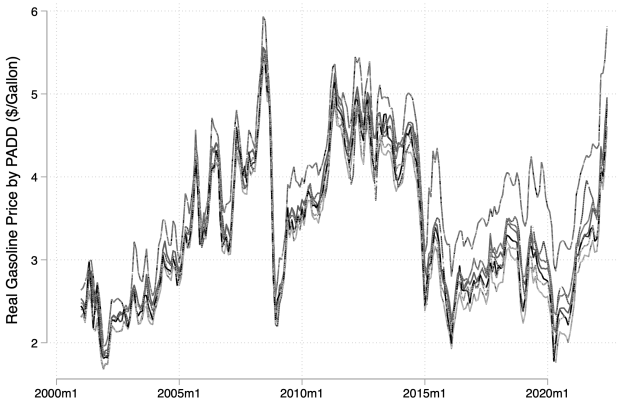


Figure 3.2: SNAP participation per capita by state, 2001-2022

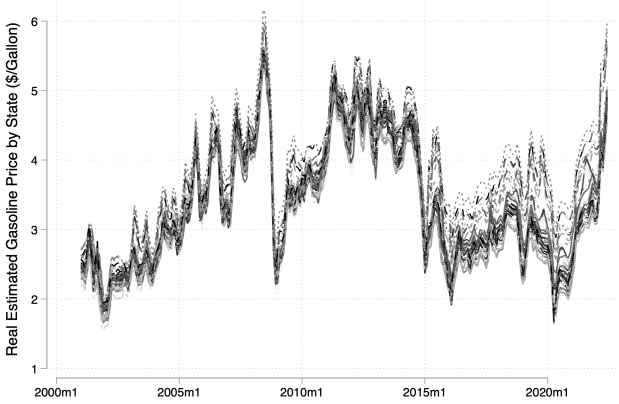
Note: This figure plots the number of individuals participating in SNAP per capita in each state from 2000 through 2022. The data are from the Census Bureau Small Area Income and Poverty Estimates accessed through FRED, Federal Reserve Bank of St. Louis.



(a) Nominal gasoline prices by PADD



(b) Real gasoline prices by PADD



(c) Real gasoline prices by state (estimated)

Figure 3.3: Gasoline prices, 2001-2022

Note: These figures plot the time series of gasoline prices at the PADD and state levels. Figure (a) plots nominal prices while figures (b) and (c) plot real prices in 2022 dollars.

The sample contains approximately 13,000 observations at the state and month level, covering January 2001 to June 2022 for all fifty states and the District of Columbia. In Table 3.1, I provide the means and standard deviations of the SNAP enrollees, SNAP enrollees per capita, gasoline prices (at both the state and PADD level), gasoline tax rates, and unemployment rates. The average state and month has nearly 700,000 SNAP enrollees, or 11% of the population, though there is wide variance in these figures, both between states and over time.

	Mean	Standard Deviation	Minimum	Maximum
SNAP Enrollees	696,642	800,921	10,210	4,775,267
Per Capita SNAP	.11	.043	.016	.27
Gasoline Price (State)	3.4	.86	1.5	6.2
Gasoline Price (PADD)	2.6	.77	1	5.8
Gasoline Tax	.23	.073	.075	.58
Unemployment Rate	5.6	2.2	1.4	30
N	13,158			

Table 3.1: Summary Statistics

This table displays the means, standard deviations, minimums, and maximums of the listed variables from a sample comprising data from January 2001 to June 2022. All variables are at the state-month level.

3.2.1 Grocery price data

Inflation in general, and grocery price inflation in particular, may influence households' SNAP application behavior. Because grocery inflation rates are not publicly available at the subnational level, I construct inflation measures using the Quarterly Food-At-Home Prices (QFAHP) data set.

The US Department of Agriculture, which administers SNAP, released the QFAHP data set from 1999 to 2010 at the Census division level in two sub-samples, 1999-2006 and 2004-2010. I create two main food indexes: a Tornqvist index and a Laspeyres index. The Tornqvist index is constructed using budget share weights that change over time, a feature meant to capture substitution effects. Meanwhile, the Laspeyres index uses fixed weights, making it more akin to a price index than a cost-of-living index (note that the Consumer Price Index is a modified Laspeyres index) (Blisard et al., 1999).

The Tornqvist price index is a geometric mean of relative prices that weighs the price

of a good by its mean share of total expenditures. This price index is calculated as follows:

$$P_{Tornqvist} = \prod_{i=1}^{n} \left(\frac{p_{i,t}}{p_{i,t-1}} \right)^{\frac{1}{2}\left[\frac{p_{i,t-1}q_{i,t-1}}{\sum_{j=1}^{n}\left(p_{j,t-1}q_{j,t-1}\right)} + \frac{p_{i,t}q_{i,t}}{\sum_{j=1}^{n}\left(p_{j,t}q_{j,t}\right)} \right]} \tag{3.1}$$

where $p_{i,t}$ refers to the price of good i in quarter t.

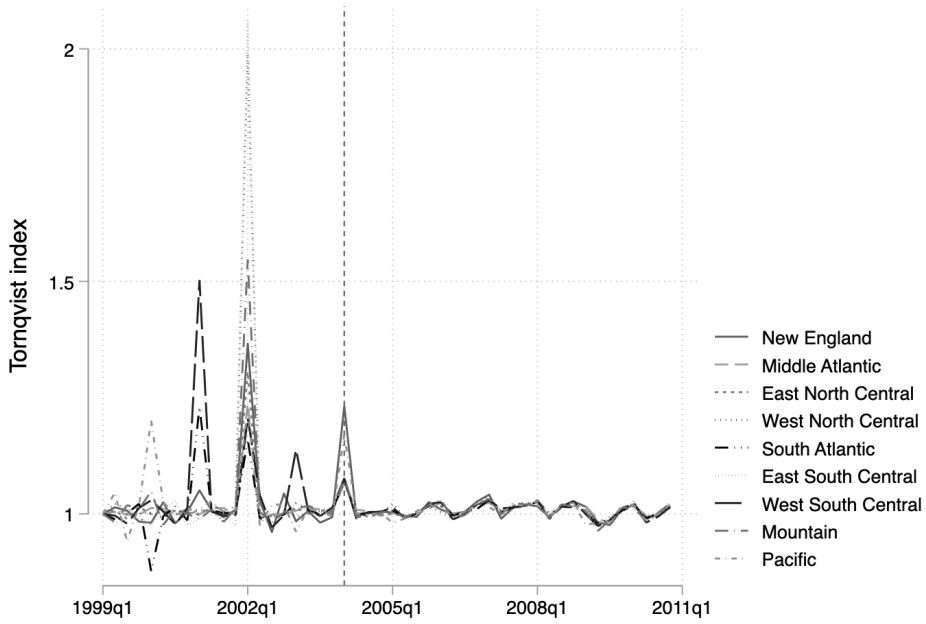
The Laspeyres index is a ratio of prices weighted by the quantity in the first period, t_0. It is calculated in the following way:

$$P_{Laspeyres} = \frac{\sum_{i=1}^{n} p_{i,t}q_{i,t_0}}{\sum_{i=1}^{n} p_{i,t_0}q_{i,t_0}} \tag{3.2}$$

As a final measure of grocery inflation, I also construct a simple price change using only the grocery categories with the largest expenditures at the Census division-quarter level. I restrict this price change to the grocery categories with largest expenditures as a straightforward way to capture the most important price changes for households. For over 90% of the sample, one of only two grocery categories has the greatest expenditure. Table 3.2 shows that "Frozen entrees and sides" and "Fresh/frozen regular fat meat" are the most common largest expenditure category by an order of magnitude.

Although the QFAHP data set is available for 1999 through 2010, there is a change in the underlying product aggregation that makes prices in the two sub-samples not comparable to each other. This change is apparent if we calculate the price indexes using the full sample. The variation in prices decreases significantly following 2004 as shown in Figures 3.4-3.6. Due to this, I split the analysis into the two sub-samples and calculate the grocery indexes within the sub-sample.

Figure 3.4: Tornqvist index for full sample, 1999-2010



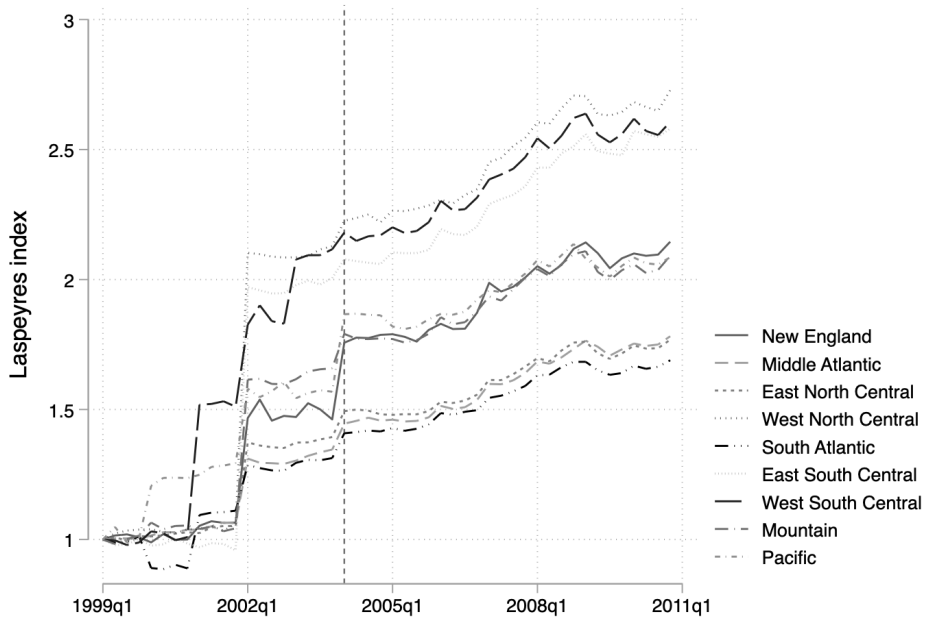
Note: This figure plots the time series of the Tornqvist index for grocery prices, a measure of food price inflation, from quarter 1 of 1999 to quarter 4 of 2010. It shows the change in variation of prices following a change in the underlying food price data in 2004.

Table 3.2: Grocery categories with largest expenditures

	Frequency	Percent
Other bread, rolls, rice, pasta, cereal	56	3.478261
Fresh/frozen regular fat meat	590	36.64596
Non-alcoholic carbonated beverages	2	.1242236
Packaged sweets/baked goods	22	1.36646
Frozen entrees and sides	924	57.3913
Packaged snacks	15	.931677
Ready to eat deli items (hot and cold)	1	.0621118
Total	1610	100

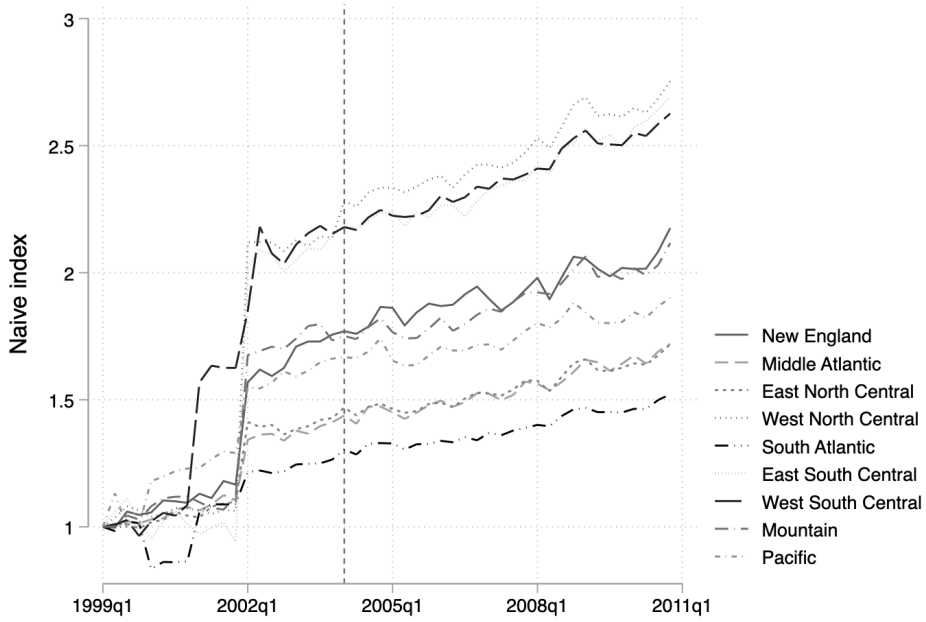
Note: This table displays the top grocery categories by expenditure using the Department of Agriculture's Quarterly Food-At-Home Prices dataset spanning 1999 through 2010.

Figure 3.5: Laspeyres index for full sample, 1999-2010



Note: This figure shows the Laspeyres index for grocery price inflation from quarter 1 of 1999 to quarter 4 of 2010. The vertical line indicates quarter 1 of 2004, when a change in the underlying food data occurred.

Figure 3.6: Naive index for full sample, 1999-2010



Note: This figure plots a naive index of grocery price inflation, calculated using the change in price of each Census division's largest food expenditure category from one quarter to the next.

3.2.2 Empirical strategy

To investigate the relationship between gasoline prices and SNAP use, I regress the natural log of SNAP participation counts on gasoline prices in the prior month.[2] I include state, month, and year fixed effects to account for state and time-invariant factors that contribute to program participation. I also include the monthly unemployment rate for a given state and CPI inflation to account for the cyclical nature of program participation (Ganong and Liebman, 2018). Standard errors are robust to heteroskedasticity and are clustered at the state level. I estimate the following regression,

$$SNAP_{s,t} = \beta_1 p_{s,t-1} + \beta_2 u_{s,t-1} + \beta_3 i_{s,t-1} + \gamma_s + \upsilon_t + \varepsilon_{s,t}, \tag{3.3}$$

where $SNAP_{t,s}$ refers to the natural log of SNAP participation in month t and state s. $p_{s,t-1}$ is the gasoline price in the prior month $t-1$ and state s, $u_{s,t-1}$ is the prior month's unemployment rate in state s, and $i_{s,t-1}$ is the inflation rate. γ_s and υ_t represent state and year fixed effects.

While Equation 3.3 allows me to recover the relationship between SNAP participation and gasoline prices, changes in both of these variables may be caused by unobserved underlying economic forces. To address possible endogeneity issues like reverse causality or omitted variable bias, I use state-level changes in gasoline taxation rates as instruments for gasoline prices.

Gasoline prices are unlikely to be influenced by take-up of SNAP; however, perhaps there are omitted variables that influence both SNAP use and gasoline prices that are not adequately addressed in a fixed effects regression. I use a plausibly exogeneous instrument to address this potential issue. I instrument for gasoline prices using state and federal gasoline tax rates. The exclusion restriction postulates that gasoline tax rates should only affect SNAP use through the price of gasoline. Since gasoline tax rates ought to only affect the price of gasoline, this seems plausible. If gasoline tax rates affect other tax rates, say the sales tax, this may be a threat to this instrumental variables strategy.

I estimate a two stage least squares model using the following estimating equations,

[2]It is a statutory requirement for eligible SNAP applicants to receive benefits within 30 days following application, though the actual timeliness varies from state to state (U.S. Department of Agriculture, Food and Nutrition Service, 2026).

$$p_{gas,t-1,s} = \pi_1 tax_{gas,t-1,s} + X_{s,t-1} + u_{t-1,s} \qquad \text{(First stage)} \tag{3.4}$$

$$SNAP_{t,s} = \beta_1 p_{gas,t-1,s} + \varepsilon_{t-1,s} \qquad \text{(Structural equation)} \tag{3.5}$$

Given the exogeneity of gasoline tax rates, this approach allows me to test for endogeneity of gasoline prices by including the residual from the first stage, $p_{gas,t-1,s} - \widehat{p_{gas,t-1,s}} = \widehat{u_{t-1,s}}$ in an OLS regression in the following way:

$$SNAP_{t,s} = \gamma_1 p_{gas,t-1,s} + \gamma_2 \widehat{u_{t-1,s}} + X_{s,t-1} + \varepsilon_{t-1,s} \tag{3.6}$$

I test whether the coefficient on the first stage residual, γ_2, is different from zero. If γ_2 is not equal to zero, then the variation in gasoline prices that is not explained by the tax rates is correlated with SNAP use. In that case, this unexplained variation, $\widehat{u_{t-1,s}}$, needs to be included in the structural equation to guard against omitted variable bias.

I fail to reject the null hypothesis that $\gamma_2 = 0$ and thus find no evidence of endogeneity. This finding indicates that the instrumental variables approach is not necessary, though it is included for completeness.

Perhaps gasoline prices drive SNAP enrollment due to increased food prices stemming from higher transportation costs. To investigate this, the final analysis includes grocery store food prices as an additional covariate to the regression of the number of SNAP enrollees on gasoline prices. Due to data limitations, this analysis is conducted at the Census division by quarter level. The regressions are specified as follows,

$$SNAP_{c,q} = \alpha_1 p_{c,q-1} + \alpha_2 u_{c,q-1} + \alpha_3 f_{c,q-1} + \nu_c + \theta_q + w_{c,q}, \tag{3.7}$$

where $SNAP_{c,q}$ refers to the natural log of SNAP enrollees in Census division c and quarter q. The terms $u_{c,q-1}$ and $f_{c,q-1}$ refer, respectively, to the unemployment rate and

grocery inflation in division q and quarter $q-1$. This specification also includes Census division and quarter fixed effects, ν_c and θ_q, to control for division- and quarter-invariant factors.

3.3 Results

I find that a 10 cent increase in the gasoline price is associated with a 0.8% increase in SNAP participation in the following month while controlling for unemployment, CPI inflation, and state and month fixed effects. This relationship is precisely estimated at the 99.9% confidence level and is shown in Table 3.3 Column 5. This estimate implies that one standard deviation increase in gasoline prices (equal to $0.83) is associated with a 7% or 55,000 person increase in SNAP use in the following month. Columns 1-4 show that excluding a variety of fixed effects, the unemployment rate, and CPI inflation modestly changes the estimate. Further, using the Oster method to test for omitted variable bias, I find that my preferred specification (Column 5) is stable to changes in the relative degree of selection on observed and unobserved variables, as shown in Appendix Figure C.1 (Altonji et al., 2005; Oster, 2019).

For completeness, I also report the findings from the IV strategy in Table 3.4. Under this approach, a find a very noisy and large effect of gasoline prices on SNAP use. Table 3.5 shows the results of the first stage of the IV regression, where gasoline prices are regressed on state-level gasoline taxes and controls. The results in Column 3 imply that a $1 tax increase corresponds with a $1.04 increase in gasoline prices, a more than 100% pass-through rate. While the literature consistently finds very high pass-through rates between 80 and 100%, my finding is likely due to data quirks stemming from the imputation from PADD-level to state-level prices (Doyle and Samphantharak, 2008; Marion and Muehlegger, 2011; Alm et al., 2009).

Because it may take applicants more than a month to from SNAP application to approval, I have alternate specifications that include longer lags for gasoline prices and find a similar, if modestly larger, effect. In Appendix Table C.1, I show that the relationship between gasoline prices and future SNAP enrollment is robust to alternate lag structures. Appendix Table C.1 (in columns 4-6) also includes the results of the regression of *per capita* SNAP enrollment on gasoline prices and covariates. I find that a $0.10 increase in gasoline prices is associated with an increase of about 70 SNAP enrollees per 100,000 people in the following month. Given an average enrollment rate of 11%, this finding implies a 0.64%

increase in enrollment from a $0.10 increase in gasoline prices. All together, these analyses generally bolster the main result–when gasoline prices increase, SNAP enrollment follows.

Table 3.3: OLS Regressions Using Imputed Gasoline Prices

	(1) log(SNAP)	(2) log(SNAP)	(3) log(SNAP)	(4) log(SNAP)	(5) log(SNAP)	(6) log(SNAP)	(7) log(SNAP)
Gasoline price (lag)	0.129**	0.160***	0.0328*	0.0578***	0.0876***	0.182**	0.393*
	(0.0392)	(0.00707)	(0.0133)	(0.0149)	(0.0237)	(0.0671)	(0.191)
Unemployment (lag)				0.0324***	0.0312***	0.0398***	0.225***
				(0.00420)	(0.00418)	(0.00623)	(0.0545)
CPI (lag)					-0.0102**	0.00405**	-0.0149
					(0.00324)	(0.00124)	(0.0303)
State FE	N	Y	Y	Y	Y	Y	N
Month FE	N	Y	Y	Y	Y	Y	Y
Year FE	N	N	Y	Y	Y	Y	Y
Month-of-Year FE	N	N	N	N	N	Y	N
PADD FE	N	N	N	N	N	N	Y
Obs	13107	13107	13107	13107	13107	13107	13107

Each numbered column in this table represents a regression of the number of SNAP participants (log scale) in a state and month regressed on gasoline prices in the previous month. Column 1 shows this regression with no covariates or fixed effects. Columns 2-7 add fixed effects, the level of unemployment, and CPI. Standard errors in parentheses are robust to heteroskedasticity and are clustered at the state level. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Furthermore, I find that the relationship between gasoline prices and SNAP use is of the same magnitude and precision given the inclusion of grocery price inflation, particularly in the more recent sample including the years between 2004 and 2010. For those years, I find that, when gasoline prices in a Census division increase by $0.10, SNAP use increases by 0.6% the following quarter, controlling for the unemployment rate and division and quarter fixed effects (as shown in Table 3.6). These coefficients are precisely estimated at the 99.9% confidence level with standard errors clustered at the division level. In an earlier sample from 1999 to 2006, the effect is larger, with an increase in SNAP enrollment around 2% following a $0.10 increase in gasoline prices. This disparity between the two samples may be due to the underlying difference in the data or it could point to a flattening of the relationship between SNAP use and gasoline prices over time.

Table 3.4: IV Regressions Using Imputed Gasoline Prices

	(1) log(SNAP)	(2) log(SNAP)	(3) log(SNAP)	(4) log(SNAP)	(5) log(SNAP)	(6) log(SNAP)
Gasoline price (lag)	2.358	-7.070	0.368	0.393	0.383	2.142
	(2.405)	(12.90)	(0.237)	(0.222)	(0.219)	(1.252)
Unemployment (lag)				0.0468***	0.0339***	0.231***
				(0.0100)	(0.00439)	(0.0547)
CPI (lag)					-0.0528	-0.269
					(0.0320)	(0.184)
State FE	N	Y	Y	Y	Y	N
Month FE	N	Y	Y	Y	Y	Y
Year FE	N	N	Y	Y	Y	Y
PADD FE	N	N	N	N	N	Y
Effective F Stat	2.673	0.324	23.57	20.40	23.76	316.1
Obs	13107	13107	13107	13107	13107	13107

Each numbered column in this table represents an instrumental variables regression of the number of SNAP participants (log scale) in a state and month regressed on gasoline prices in the previous month. Changes in state-level gasoline taxes instrument for gasoline prices. Column 1 shows this regression with no covariates or fixed effects. Columns 2-6 add fixed effects, the level of unemployment, and CPI. Standard errors in parentheses are robust to heteroskedasticity and are clustered at the state level. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3.5: First Stage Regressions

	(1) Gas. price	(2) Gas. price	(3) Gas. price	(4) Gas. price	(5) Gas. price	(6) Gas. price
State gasoline tax	0.803	-0.355	1.035***	1.012***	1.042***	1.231***
	(0.491)	(0.623)	(0.213)	(0.224)	(0.214)	(0.0692)
Unemployment				-0.0426***	-0.00840*	-0.00524*
				(0.00486)	(0.00373)	(0.00248)
CPI					0.145***	0.145***
					(0.000910)	(0.000687)
State FE	N	Y	Y	Y	Y	N
Month FE	N	Y	Y	Y	Y	Y
Year FE	N	N	Y	Y	Y	Y
PADD FE	N	N	N	N	N	Y
Obs	13107	13107	13107	13107	13107	13107

Each numbered column in this table represents a regression of the gasoline prices on state-level gasoline taxes in the previous month. Column 1 shows this regression with no covariates or fixed effects. Columns 2-6 add fixed effects, the level of unemployment, and CPI. Standard errors in parentheses are robust to heteroskedasticity and are clustered at the state level. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3.6: Regressions of SNAP on gasoline prices and grocery inflation

	(1) Log(SNAP)	(2) Log(SNAP)	(3) Log(SNAP)	(4) Log(SNAP)	(5) Log(SNAP)	(6) Log(SNAP)	(7) Log(SNAP)
Gasoline price (lag)	0.182***	0.0611**	0.0721**	0.0612**	0.228***	0.161***	0.165***
	(0.00921)	(0.0159)	(0.0145)	(0.0171)	(0.0128)	(0.0174)	(0.0150)
Unemployment (lag)	0.116***	0.0129	0.0108	0.0120	0.137***	0.0760**	0.0748**
	(0.00896)	(0.0129)	(0.0116)	(0.0120)	(0.0164)	(0.0189)	(0.0193)
Tornqvist index (2004-10)		0.325					
		(0.345)					
Laysperes index (2004-10)			-0.354				
			(0.206)				
$\Delta Price$ (2004-10)				0.0134			
				(0.0216)			
Tornqvist index (1999-06)					-0.0583		
					(0.0365)		
Laysperes index (1999-06)						0.204**	
						(0.0585)	
$\Delta Price$ (1999-06)							0.176**
							(0.0443)
Observations	351	99	99	99	207	207	207

Columns 1-7 each show the results of a regression of the log of SNAP enrollees in a Census division on the previous quarter's average gasoline price and other covariates. All regressions include division and quarter fixed effects. Standard errors in parentheses are robust to heteroskedasticity and are clustered at the division level. * p<.05, ** p<.01, *** p<.001.

3.4 Conclusion

This paper shows that SNAP enrollment increases during the month following increases in gasoline prices, even after controlling for inflation, unemployment, and state- and time-invariant factors.

The literature on SNAP application behavior has so far focused largely on stigma, information interventions, and assistance with applications. The findings in this paper demonstrate that the use of food assistance programs is driven, in part, by the price of gasoline, a salient necessity for many households. Policymakers and SNAP administrators are likely to find increased demand for SNAP during periods with high gasoline prices, resulting in increased benefit outlays and a need for more administrative staff. These findings take on additional relevance given recent legislative changes that tighten SNAP eligibility, as households facing energy price shocks may find fewer options available to them.[3]

The analyses in this paper are largely limited by data availability. The analysis lacks application data, which would much more directly speak to the impact of gasoline price changes on application behavior. Using caseload data conflates new applicants with incumbents. Individual-level data on SNAP application, receipt, household income, gasoline prices, and spending would also make it possible to distinguish between mechanisms. In particular, increased gasoline prices may constitute a true income shock or negatively impact consumer sentiment. Further, aggregated data can mask within-state variation such as differences between rural and urban populations who may differ in both their gasoline use and access to SNAP.

This area of research remains underexplored. Future work is needed to examine how energy price volatility shapes demand for public assistance as well as the adequacy of existing safety net programs in buffering low-income households against commodity price shocks.

[3]The One Big Beautiful Bill Act of 2025 (Pub. L. No. 119-21) applies more restrictive criteria for SNAP eligibility in addition to changing the cost-sharing structure between the federal and state governments and numerous other provisions.

APPENDIX A

Appendix to Chapter 1

A.1 Teacher Age Bins Uncorrelated with Observables

In this section, I present the results of regressions of a teacher age bin indicator on extreme heat days, school and year fixed effects, and a set of control variables including students' past year test scores, student demographics, county unemployment rate, county median income, and classroom size. These regressions are based on linear probability models of the following form:

$$A_{jt} = \beta_0 + \beta_1 Heat + \mathbf{X_i}^{-1}\alpha + \gamma_k + \upsilon_t + \varepsilon_{jt} \tag{A.1}$$

where

$$A_j = \begin{cases} 1 & \text{if teacher } j\text{'s age} \in [24, 54] \\ 0, & \text{otherwise} \end{cases}$$

and $Heat$ refers to the number of extreme heat days at a school during a school year. The variables γ_k and υ_t refer to school and year fixed effects. Table A.1 presents the full results of these regression analyses. I find that teacher age bins are not correlated with extreme heat days. This result is stable across the choice of observable characteristics. Because teacher age and extreme heat days do not appear to be correlated, it is plausible that they are uncorrelated conditional on unobservables, an assumption needed for causal claims.

Table A.1: Regression of Teacher Age Bin Indicator on Observables

	(1)	(2)	(3)	(4)
Extreme Heat Days	0.000802	0.000804	0.000846	0.000899
	(0.000632)	(0.000632)	(0.000632)	(0.000633)
Past Year Score	Yes	Yes	Yes	Yes
Student Demographics	No	Yes	Yes	Yes
County Economic Trends	No	No	Yes	Yes
Classroom Size	No	No	No	Yes
Observations	3663375	3663322	3663322	3663322

Note: Heteroskedasticity robust standard errors clustered at the grid level appear in parentheses. $^{*}\ p < 0.05$, $^{**}\ p < 0.01$, $^{***}\ p < 0.001$ This table presents the regression of teacher age bins on extreme heat days and observable characteristics. All regressions include year and school fixed effects. Column 1 includes the prior year test scores for a teacher's current students. Column 2 adds student demographics including ethnicity, gender, and age. Column 3 adds county-level unemployment rates and median income. Finally, column 4 adds class size.

A.2 Test Score and Teacher Age Using Alternate Heat Measures

In the main analysis, the independent variable, the number of days above the 90th percentile of historic temperature, captures heat relative to baseline to account for adaptation by climate such as air-conditioning infrastructure. However, it may be useful to look at effects of absolute heat to compare to past literature and to investigate possible nonlinearities. To do this, I use the number of days in each school year below 30°F, between 30 and 40 °F, and so on, using the number of days between 60 and 70°F as the excluded category.

In this appendix, I regress standardized test scores on the number of days in each temperature bin in a school year and an indicator for teacher age above the 90th percentile. I include the interaction of these two explanatory variables to capture differences in heat effects by teacher age. This regression is of the following form:

$$y_{its} = \beta_{\mathbf{a}} \sum_{a=1}^{N} DaysInBin_{at\tilde{s}} + \gamma_{\mathbf{a}} \sum_{a=1}^{N} DaysInBin_{at\tilde{s}} \times A_{tj} + \upsilon A_{tj} + \mathbf{X_i}^{-1}\alpha + \delta_t + \lambda_{\tilde{s}} + \varepsilon_{its} \quad \text{(A.2)}$$

where

$$A_j = \begin{cases} 1 & \text{if teacher } j\text{'s age} \in [24, 54] \\ 0, & \text{otherwise} \end{cases}$$

and y_{its} is the outcome of interest: student i's standardized test score in year t and subject s. $DaysInBin_{ats}$ is the number of days in bin a in school year t for school $\tilde{s}$. $\mathbf{X_i}$ is a vector of control variables including lagged test score, student race, gender, and grade level. δ_t and $\lambda_{\tilde{s}}$ represent year and school fixed effects. School and year fixed effects are included to control for school-specific or year-specific unobservable factors that may affect test scores.

The results of this regression indicate that all teachers' students scores decrease with each additional day above 90°F and below 30°F–but older teachers' students seem to be especially affected. Table A.2, Column 6 presents the full regression output and Figure A.1 plots the coefficients.

Table A.2: Older teachers' quality suffers in heat above 90°F and below 30°F

	(1)	(2)	(3)	(4)	(5)	(6)
Days in 90s °F	-0.00195	-0.00158*	-0.00159*	-0.00167*	-0.00193**	-0.00205**
	(0.00124)	(0.000719)	(0.000718)	(0.000718)	(0.000725)	(0.000728)
Days in 80s °F	0.00148	0.000400	0.000385	0.000477	0.000244	0.000311
	(0.00112)	(0.000586)	(0.000586)	(0.000574)	(0.000583)	(0.000584)
Days in 70s °F	0.000857	0.0000172	0.0000192	0.000146	0.0000718	0.0000360
	(0.000874)	(0.000440)	(0.000440)	(0.000431)	(0.000433)	(0.000433)
50s °F	0.000852	0.0000558	0.0000408	0.0000453	-0.0000118	-0.0000968
	(0.000813)	(0.000393)	(0.000392)	(0.000385)	(0.000385)	(0.000388)
40s °F	0.000580	-0.000751	-0.000770	-0.000840*	-0.000839*	-0.000963*
	(0.000794)	(0.000432)	(0.000432)	(0.000426)	(0.000426)	(0.000428)
30s °F	0.00220	-0.000242	-0.000276	-0.000338	-0.000222	-0.000441
	(0.00134)	(0.000666)	(0.000664)	(0.000658)	(0.000663)	(0.000663)
$<$ 30 °F	-0.00869**	-0.00535***	-0.00536***	-0.00527***	-0.00560***	-0.00584***
	(0.00288)	(0.00150)	(0.00150)	(0.00148)	(0.00148)	(0.00147)
Teacher Age [24,54) $\times$ $<$ 30 °F	0.00720*	0.00306*	0.00306*	0.00275	0.00268	0.00271
	(0.00290)	(0.00151)	(0.00151)	(0.00147)	(0.00147)	(0.00147)
Teacher Age [24,54) $\times$ 30s °F	-0.000403	-0.000388	-0.000390	-0.000321	-0.000276	-0.000306
	(0.00122)	(0.000536)	(0.000536)	(0.000531)	(0.000531)	(0.000531)
Teacher Age [24,54) $\times$ 40s °F	-0.000383	-0.000152	-0.000153	-0.000201	-0.000191	-0.000158
	(0.000697)	(0.000351)	(0.000351)	(0.000342)	(0.000342)	(0.000344)
Teacher Age [24,54) $\times$ 50s °F	-0.000391	-0.000355	-0.000358	-0.000365	-0.000345	-0.000337
	(0.000788)	(0.000367)	(0.000367)	(0.000359)	(0.000359)	(0.000361)
Teacher Age [24,54) $\times$ Days in 70s °F	-0.0000889	0.0000297	0.0000269	-0.0000716	-0.0000473	-0.0000436
	(0.000878)	(0.000430)	(0.000431)	(0.000422)	(0.000421)	(0.000421)
Teacher Age [24,54) $\times$ Days in 80s °F	-0.000220	-0.0000549	-0.0000544	-0.0000477	-0.0000443	-0.0000397
	(0.00110)	(0.000524)	(0.000524)	(0.000512)	(0.000513)	(0.000513)
Teacher Age [24,54) $\times$ Days in 90s °F	0.00150	0.00110*	0.00109*	0.00103	0.00101	0.00101
	(0.00108)	(0.000540)	(0.000540)	(0.000533)	(0.000532)	(0.000531)
Teacher Age [24,54)	-0.00945	0.00385	0.00417	0.0119	0.00872	0.00674
	(0.127)	(0.0617)	(0.0617)	(0.0605)	(0.0604)	(0.0606)
Past Year Score	No	Yes	Yes	Yes	Yes	Yes
Day of Test Temperature	No	No	Yes	Yes	Yes	Yes
Student Demographics	No	No	No	Yes	Yes	Yes
County Economic Trends	No	No	No	No	Yes	Yes
Classroom Size	No	No	No	No	No	Yes
Observations	3,663,254	3,663,254	3,663,254	3,663,254	3,663,254	3,663,254

Note: Heteroskedasticity robust standard errors clustered at the grid level appear in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
This table presents the regression of normalized test scores on ten-degree temperature bins, teacher age bins, and the interaction of the these explanatory variables. Column 6 includes controls for lagged test score, the temperature on the day of the exam, student grade, ethnicity, gender, and year and school fixed effects.

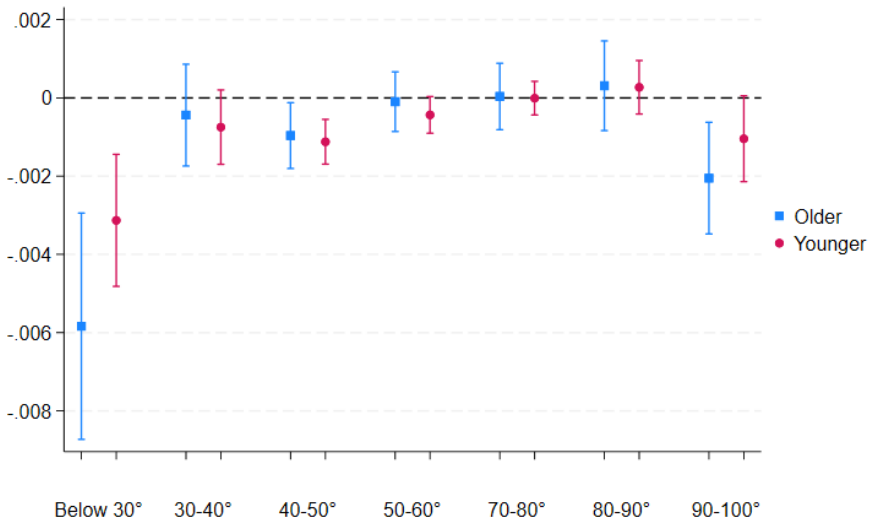


Figure A.1: Older teachers' quality suffers in heat above 90°F and below 30°F

Note: This figure plots the coefficients and standard errors from the regression of standardized test scores on the number of days in each temperature bin and teacher age.

Table A.3: Younger teachers' students perform relatively better than older ones during hotter years

	(1)	(2)	(3)	(4)	(5)
Cooling Degree Days (100s)	-0.000992	-0.000949	-0.00138	-0.000986	-0.00207
	(0.00466)	(0.00466)	(0.00456)	(0.00457)	(0.00459)
Heating Degree Days (100s)	-0.00489**	-0.00515***	-0.00587***	-0.00605***	-0.00630***
	(0.00150)	(0.00152)	(0.00149)	(0.00151)	(0.00150)
Teacher Age [24,54)	-0.0633*	-0.0636*	-0.0702**	-0.0727**	-0.0717**
	(0.0256)	(0.0255)	(0.0247)	(0.0249)	(0.0248)
Teacher Age [24,54) × Cooling Degree Days (100s)	0.00738*	0.00741*	0.00786**	0.00797**	0.00786**
	(0.00305)	(0.00305)	(0.00298)	(0.00299)	(0.00299)
Teacher Age [24,54) × Heating Degree Days (100s)	0.000701	0.000708	0.000882	0.000930	0.000913
	(0.000570)	(0.000570)	(0.000548)	(0.000552)	(0.000551)
Past Year Score	Yes	Yes	Yes	Yes	Yes
Day of Test Temperature	No	Yes	Yes	Yes	Yes
Student Demographics	No	No	Yes	Yes	Yes
County Economic Trends	No	No	No	Yes	No
Classroom Size	No	No	No	No	Yes
Observations	3663375	3663375	3663322	3663322	3663322

Note: Heteroskedasticity robust standard errors clustered at the grid level appear in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
This table presents the regression of normalized test scores on heat and cooling degree days, teacher age bins, and the interactions of the these explanatory variables. All regressions include year and school fixed effects. Column 5 includes controls for lagged test score, the temperature on the day of the exam, student grade, ethnicity, and gender.

Table A.4: Findings robust to alternate definition of extreme heat using 85th Percentile

	(1)	(2)	(3)	(4)	(5)	(6)
Extreme Heat Days (85th Percentile)	-0.00301**	-0.00195***	-0.00194***	-0.00182**	-0.00189***	-0.00199***
	(0.00101)	(0.000559)	(0.000560)	(0.000562)	(0.000561)	(0.000563)
Teacher Age [24,54)	-0.0568***	-0.0228***	-0.0229***	-0.0221***	-0.0222***	-0.0224***
	(0.0101)	(0.00437)	(0.00437)	(0.00430)	(0.00430)	(0.00429)
Teacher Age [24,54) × Extreme Heat Days (85th Percentile)	0.00182*	0.00100*	0.00101*	0.000962*	0.000972*	0.000938*
	(0.000769)	(0.000408)	(0.000408)	(0.000405)	(0.000405)	(0.000405)
Past Year Score	No	Yes	Yes	Yes	Yes	Yes
Day of Test Temperature	No	No	Yes	Yes	Yes	Yes
Student Demographics	No	No	No	Yes	Yes	Yes
County Economic Trends	No	No	No	No	Yes	Yes
Classroom Size	No	No	No	No	No	Yes
Observations	3,663,254	3,663,254	3,663,254	3,663,254	3,663,254	3,663,254

Note: Heteroskedasticity robust standard errors clustered at the grid level appear in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
This table presents the regression of normalized annual test scores on the number of extreme days beyond the 85th percentile in the school year, teacher age bins, and the interaction of the these explanatory variables. All regressions include year and school fixed effects. Column 1 includes no other covariates. Column 2 controls for a student's prior year test score. Column 3 adds the exam day temperature. Column 4 adds student gender, ethnicity, and age. Column 5 adds county-level yearly unemployment rates and average income. Column 6 adds the classroom size.

Table A.5: Findings robust to alternate definition of extreme heat using 95th Percentile

	(1)	(2)	(3)	(4)	(5)	(6)
Extreme Heat Days (95th Percentile)	-0.00551*	-0.00157	-0.00155	-0.00156	-0.00154	-0.00157
	(0.00256)	(0.00142)	(0.00142)	(0.00139)	(0.00140)	(0.00140)
Teacher Age [24,54)	-0.0511***	-0.0194***	-0.0194***	-0.0189***	-0.0189***	-0.0193***
	(0.00851)	(0.00350)	(0.00350)	(0.00341)	(0.00342)	(0.00341)
Teacher Age [24,54) × Extreme Heat Days (95th Percentile)	0.00552*	0.00275*	0.00274*	0.00275*	0.00268*	0.00267*
	(0.00238)	(0.00120)	(0.00120)	(0.00118)	(0.00117)	(0.00118)
Past Year Score	No	Yes	Yes	Yes	Yes	Yes
Day of Test Temperature	No	No	Yes	Yes	Yes	Yes
Student Demographics	No	No	No	Yes	Yes	Yes
County Economic Trends	No	No	No	No	Yes	Yes
Classroom Size	No	No	No	No	No	Yes
Observations	3,663,254	3,663,254	3,663,254	3,663,254	3,663,254	3,663,254

Note: Heteroskedasticity robust standard errors clustered at the grid level appear in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ This table presents the regression of normalized annual test scores on the number of extreme days beyond the 95th percentile in the school year, teacher age bins, and the interaction of the these explanatory variables. All regressions include year and school fixed effects. Column 1 includes no other covariates. Column 2 controls for a student's prior year test score. Column 3 adds the exam day temperature. Column 4 adds student gender, ethnicity, and age. Column 5 adds county-level yearly unemployment rates and average income. Column 6 adds the classroom size.

A.3 Using Alternate Definitions of Teacher Age

Table A.6: Similar Findings Using Age 50 Cutoff (85th Percentile of Teacher Age)

	(1)	(2)	(3)	(4)	(5)	(6)
Extreme Heat Days	-0.00189	-0.000901	-0.000897	-0.000968	-0.00104	-0.00121
	(0.00104)	(0.000653)	(0.000654)	(0.000657)	(0.000657)	(0.000661)
Teacher Age [24,50)	-0.0538***	-0.0229***	-0.0229***	-0.0225***	-0.0224***	-0.0226***
	(0.00724)	(0.00305)	(0.00305)	(0.00297)	(0.00296)	(0.00296)
Teacher Age [24,50) × Extreme Heat Days	0.00136	0.000848*	0.000848*	0.000764	0.000747	0.000720
	(0.000788)	(0.000431)	(0.000431)	(0.000429)	(0.000429)	(0.000430)
Past Year Score	No	Yes	Yes	Yes	Yes	Yes
Day of Test Temperature	No	No	Yes	Yes	Yes	Yes
Student Demographics	No	No	No	Yes	Yes	Yes
County Economic Trends	No	No	No	No	Yes	Yes
Classroom Size	No	No	No	No	No	Yes
Observations	3,663,254	3,663,254	3,663,254	3,663,254	3,663,254	3,663,254

Note: Heteroskedasticity robust standard errors clustered at the grid level appear in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
This table presents the regression of normalized annual test scores on the number of extreme days in the school year, teacher age bins, and the interaction of the these explanatory variables, where the cutoff age between younger and older teachers is age 50, or the 85th percentile of teacher age. All regressions include year and school fixed effects. Column 1 does not include any covariates. Column 2 controls for a student's prior year test score. Column 3 adds the exam day temperature. Column 4 adds student gender, ethnicity, and age. Column 5 adds county-level yearly unemployment rates and average income. Column 6 adds the classroom size.

Table A.7: Similar Findings Using Age 59 Cutoff (95th Percentile of Teacher Age)

	(1)	(2)	(3)	(4)	(5)	(6)
Extreme Heat Days	-0.00350*	-0.00177	-0.00176	-0.00176	-0.00181	-0.00202*
	(0.00151)	(0.000927)	(0.000928)	(0.000941)	(0.000941)	(0.000937)
Teacher Age [24,59)	-0.0567***	-0.0182***	-0.0182***	-0.0167***	-0.0168***	-0.0171***
	(0.0114)	(0.00503)	(0.00503)	(0.00495)	(0.00494)	(0.00492)
Teacher Age [24,59) × Extreme Heat Days	0.00287*	0.00166*	0.00165*	0.00150	0.00147	0.00147
	(0.00133)	(0.000783)	(0.000783)	(0.000779)	(0.000778)	(0.000772)
Past Year Score	No	Yes	Yes	Yes	Yes	Yes
Day of Test Temperature	No	No	Yes	Yes	Yes	Yes
Student Demographics	No	No	No	Yes	Yes	Yes
County Economic Trends	No	No	No	No	Yes	Yes
Classroom Size	No	No	No	No	No	Yes
Observations	3,663,254	3,663,254	3,663,254	3,663,254	3,663,254	3,663,254

Note: Heteroskedasticity robust standard errors clustered at the grid level appear in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
This table presents the regression of normalized annual test scores on the number of extreme days in the school year, teacher age bins, and the interaction of the these explanatory variables, where the cutoff age between younger and older teachers is age 59, or the 95th percentile of teacher age. All regressions include year and school fixed effects. Column 1 does not include any covariates. Column 2 controls for a student's prior year test score. Column 3 adds the exam day temperature. Column 4 adds student gender, ethnicity, and age. Column 5 adds county-level yearly unemployment rates and average income. Column 6 adds the classroom size.

Table A.8: Similar Findings Excluding Youngest Teachers

	(1)	(2)	(3)	(4)	(5)	(6)
Extreme Heat Days	-0.003*	-0.001	-0.001	-0.001*	-0.001*	-0.002*
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Teacher Age [28,54)	-0.042***	-0.015***	-0.015***	-0.014***	-0.014***	-0.014***
	(0.009)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
Teacher Age [28,54) × Extreme Heat Days	0.002*	0.001**	0.001**	0.001*	0.001*	0.001*
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Past Year Score	No	Yes	Yes	Yes	Yes	Yes
Day of Test Temperature	No	No	Yes	Yes	Yes	Yes
Student Demographics	No	No	No	Yes	Yes	Yes
County Economic Trends	No	No	No	No	Yes	Yes
Classroom Size	No	No	No	No	No	Yes
Observations	3,241,374	3,241,374	3,241,374	3,241,374	3,241,374	3,241,374

Note: Heteroskedasticity robust standard errors clustered at the grid level appear in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
This table presents the regression of normalized annual test scores on the number of extreme days in the school year, teacher age bins, and the interaction of the these explanatory variables, excluding teachers below the age of 28. All regressions include year and school fixed effects. Column 1 does not include any covariates. Column 2 controls for a student's prior year test score. Column 3 adds the exam day temperature. Column 4 adds student gender, ethnicity, and age. Column 5 adds county-level yearly unemployment rates and average income. Column 6 adds the classroom size.

A.4 Heterogeneity

Table A.9: Heterogeneity By Test Subject

	(1) Math	(2) Reading	(3) All
Extreme Heat Days	-0.000882	-0.00270***	-0.00185*
	(0.00100)	(0.000719)	(0.000732)
Teacher Age [24,54)	-0.0156**	-0.0251***	-0.0208***
	(0.00521)	(0.00335)	(0.00363)
Teacher Age [24,54) × Extreme Heat Days	0.00103	0.00153**	0.00135**
	(0.000705)	(0.000554)	(0.000516)
Observations	1,845,537	1,817,679	3,663,254

Note: Heteroskedasticity robust standard errors clustered at the grid level appear in parentheses: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. This table presents the regression of normalized annual test scores on the number of extreme days beyond the 90th percentile in the school year, teacher age bins, and the interaction of the these explanatory variables. All regressions include year and school fixed effects and controls for student's prior year test score, exam day temperature, student gender, ethnicity, and age, county-level yearly unemployment rates and average income, and classroom size. Column 1 includes only math standardized test scores. Column 2 includes only reading standardized test scores. Column 3 includes all test scores.

Table A.10: Heterogeneity by Grade Level

	(1) Grade 4	(2) Grade 5	(3) All
Extreme Heat Days	-0.00169	-0.00218*	-0.00185*
	(0.000937)	(0.00104)	(0.000732)
Teacher Age [24,54)	-0.0218***	-0.0189***	-0.0208***
	(0.00481)	(0.00468)	(0.00363)
Teacher Age [24,54) × Extreme Heat Days	0.00158*	0.00113	0.00135**
	(0.000731)	(0.000720)	(0.000516)
Observations	1,760,149	1,875,729	3,663,254

Note: Heteroskedasticity robust standard errors clustered at the grid level appear in parentheses: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. This table presents the regression of normalized annual test scores on the number of extreme days beyond the 90th percentile in the school year, teacher age bins, and the interaction of the these explanatory variables. All regressions include year and school fixed effects and controls for student's prior year test score, exam day temperature, student gender, ethnicity, and age, county-level yearly unemployment rates and average income, and classroom size. Column 1 includes only fourth grade. Column 2 includes only fifth grade. Column 3 includes all test scores.

Table A.11: Heterogeneity by Semester of Extreme Heat Days

	(1) Fall	(2) Spring	(3) All
Extreme Heat Days	-0.00202**	0.0130	-0.00185*
	(0.000734)	(0.00842)	(0.000732)
Teacher Age [24,54)	-0.0209***	-0.0157***	-0.0208***
	(0.00360)	(0.00325)	(0.00363)
Teacher Age [24,54) × Extreme Heat Days	0.00140**	-0.00956	0.00135**
	(0.000509)	(0.00793)	(0.000516)
Observations	3,663,322	3,663,322	3,663,322

Note: Heteroskedasticity robust standard errors clustered at the grid level appear in parentheses: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. This table presents the regression of normalized annual test scores on the number of extreme days beyond the 90th percentile in the school year, teacher age bins, and the interaction of the these explanatory variables. All regressions include year and school fixed effects and controls for student's prior year test score, exam day temperature, student gender, ethnicity, and age, county-level yearly unemployment rates and average income, and classroom size. Column 1 includes heat days from each fall semester. Column 2 includes heat days from the spring semesters only. Column 3 includes all extreme heat days.

Table A.12: Heterogeneity By Student Gender

	(1) Girls	(2) Boys	(3) All
Extreme Heat Days	-0.00183*	-0.00187*	-0.00185*
	(0.000811)	(0.000818)	(0.000732)
Teacher Age [24,54)	-0.0171***	-0.0245***	-0.0208***
	(0.00386)	(0.00400)	(0.00363)
Teacher Age [24,54) × Extreme Heat Days	0.00146*	0.00127*	0.00135**
	(0.000592)	(0.000599)	(0.000516)
Observations	1,845,537	1,817,679	3,663,254

Note: Heteroskedasticity robust standard errors clustered at the grid level appear in parentheses: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. This table presents the regression of normalized annual test scores on the number of extreme days beyond the 90th percentile in the school year, teacher age bins, and the interaction of the these explanatory variables. All regressions include year and school fixed effects and controls for student's prior year test score, exam day temperature, student gender, ethnicity, and age, county-level yearly unemployment rates and average income, and classroom size. Column 1 includes only female students' test scores. Column 2 includes only male students' test scores. Column 3 includes all test scores.

Table A.13: Heterogeneity By Rurality

	(1) Rural	(2) Suburban	(3) Urban
Extreme Heat Days	-0.000699	-0.00181	-0.00228*
	(0.00130)	(0.00201)	(0.00101)
Teacher Age [24,54)	-0.00203	-0.0219	-0.0280***
	(0.00791)	(0.0117)	(0.00425)
Teacher Age [24,54) × Extreme Heat Days	0.000373	0.00109	0.00180**
	(0.00110)	(0.00149)	(0.000640)
Observations	819,807	469,295	2,374,152

Note: Heteroskedasticity robust standard errors clustered at the grid level appear in parentheses: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. This table presents the regression of normalized annual test scores on the number of extreme days beyond the 90th percentile in the school year, teacher age bins, and the interaction of the these explanatory variables. All regressions include year and school fixed effects and controls for student's prior year test score, exam day temperature, student gender, ethnicity, and age, county-level yearly unemployment rates and average income, and classroom size. Column 1 includes only schools in rural counties. Column 2 includes only suburban counties and Column 3 uses only urban counties.

Table A.14: Heterogeneity By Student Race

	(1) American Indian	(2) Asian and Pacific Islander	(3) Hispanic	(4) Black	(5) White	(6) Multi-Racial	(7) N/A
Extreme Heat Days	-0.00235	-0.00124	-0.00120	-0.00313*	-0.00139	-0.00170	0.00560
	(0.00424)	(0.00245)	(0.00128)	(0.00131)	(0.000853)	(0.00223)	(0.120)
Teacher Age [24,54)	0.0170	-0.0310***	-0.0172**	-0.0152**	-0.0227***	-0.0277**	-0.126
	(0.0291)	(0.00911)	(0.00585)	(0.00517)	(0.00426)	(0.00926)	(0.139)
Teacher Age [24,54) × Extreme Heat Days	0.00267	0.00204	0.000970	0.00170	0.000999	0.00216	0.0101
	(0.00309)	(0.00203)	(0.000986)	(0.000901)	(0.000679)	(0.00185)	(0.0107)
Observations	50,185	96,919	567,961	883,830	1,862,873	144,499	56,936

Note: Heteroskedasticity robust standard errors clustered at the grid level appear in parentheses: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. This table presents the regression of normalized annual test scores on the number of extreme days beyond the 90th percentile in the school year, teacher age bins, and the interaction of the these explanatory variables. All regressions include year and school fixed effects and controls for student's prior year test score, exam day temperature, student gender, ethnicity, and age, county-level yearly unemployment rates and average income, and classroom size.

Table A.15: Heterogeneity By County Median Income

	(1)	(2)	(3)	(4)	(5)
	\$30-40K	\$40-50K	\$50-60K	\$60-70K	>\$70-80K
Extreme Heat Days	-0.00296	-0.00298	-0.000131	-0.00293	-0.00442
	(0.00490)	(0.00165)	(0.00115)	(0.00155)	(0.00292)
Teacher Age [24,54)	0.0199	-0.0126	-0.0206***	-0.0284***	-0.0294***
	(0.0237)	(0.00814)	(0.00560)	(0.00735)	(0.00816)
Teacher Age [24,54) × Extreme Heat Days	-0.000108	0.00128	0.00112	0.00194*	0.00138
	(0.00222)	(0.00129)	(0.000885)	(0.000948)	(0.00146)
Observations	92,096	603,716	1,481,712	880,954	604,776

Note: Heteroskedasticity robust standard errors clustered at the grid level appear in parentheses: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. This table presents the regression of normalized annual test scores on the number of extreme days beyond the 90th percentile in the school year, teacher age bins, and the interaction of the these explanatory variables. All regressions include year and school fixed effects and controls for student's prior year test score, exam day temperature, student gender, ethnicity, and age, county-level yearly unemployment rates and average income, and classroom size. Column 1-5 each include schools in counties a given median income level, with Column 1 including only counties with a median income between \$30,000 and \$40,000, for example.

A.5 Movers Analysis

This analysis explores whether younger and older teachers geographically sort by extreme heat in different ways. If older teachers choose to live in hotter climates compared to younger teachers, we would expect to see these teachers moving from a relatively cooler place to a warmer one, compared to younger teachers. Table A.17 presents the average changes in weather between a teacher's school and the next school where she works. For example, on average, younger teachers relocate to schools that have 1.57 fewer extreme heat days compared to their prior workplace. Older teachers on average relocate to schools that have 2.88 fewer extreme heat days than their prior school. The difference between these two figure is statistically significant, indicating that older teachers who relocate indeed move to cooler places than younger teachers. Table A.17 also shows these average differences between schools by teacher age for the number of days in the 80-90°F range and 90-100°F range, the average maximum temperature, student demographics, the county-level unemployment rate, and the median income. On average, younger teachers move to schools in slightly hotter places, with lower unemployment rates and higher median income. Younger teachers relocate to schools in counties with a median income that is on average $2,470 higher than their prior school; older teachers relocate to areas with a median income that is $1,280 higher on average. This difference hints at a potential reason for the move. Teacher may switch schools for better-resourced school districts earlier in their careers.

The descriptive analysis shown in Tables A.16 and A.17 indicate underlying differences between younger and older teachers who move. In addition, I implement an event study to investigate whether there is any causal evidence of geographic sorting by teacher age. This event study framework uses a regression of the following form,

$$y_{st} = \alpha + \sum_{j=2}^{J} \beta_j (\text{Lag}\, j)_{st} + \sum_{k=1}^{K} \gamma_k (\text{Lead}\, k)_{st} + \mu_s + \lambda_t + X'_{st}\Gamma + \varepsilon_{st}, \tag{A.3}$$

where y_{st} is the number of extreme heat days at school s in year t and $\text{Lag}\, j_{st}$ and $\text{Lead}\, k_{st}$ are indicators for the number for years preceding or following a move. The regression includes school and year fixed effects, μ_s and λ_t and a vector of covariates, X_{st} including percentage of a teacher's students who are male, percentage of students who are Black, student age, and student grade-level.

I find no evidence that older teachers move to cooler places compared to younger teachers. I present these finding in Figure A.2, which has the number of school years relative to a move on the horizontal axis and the estimated effect of that year on the number of extreme heat days on the vertical axis. These point estimates correspond with β_j and γ_k in Equation A.3. The point estimates in red are those for younger teachers while those in yellow are those for older teachers. The figure also includes the 95% confidence intervals around each of point estimate.

Table A.16: Summary Table For Teachers Who Move

	First School		Second School		Third School	
	mean	sd	mean	sd	mean	sd
Panel A: Teacher Age <54						
Extreme Heat Days	5.08	4.11	3.29	2.54	3.28	2.32
Days in 80s	30.64	6.52	31.72	7.24	32.38	7.81
Days in 90s	5.24	4.56	3.36	2.91	3.33	2.61
Average Maximum Temperature	64.50	2.27	64.58	2.27	64.83	2.32
Black Student Percent	0.30	0.26	0.29	0.25	0.30	0.26
Female Student Percentage	0.01	0.09	0.00	0.07	0.01	0.07
Unemployment Rate	8.11	2.52	6.59	2.62	5.50	1.93
Median Income	58,927.17	10,305.55	61,407.93	10,808.20	62,968.37	10,879.01
Rural	0.19	0.39	0.17	0.37	0.12	0.33
Number of Teachers, First School					4,083	
Second School					4,083	
Third School					569	
	First School		Second School		Third School	
	mean	sd	mean	sd	mean	sd
Panel B: Teacher Age ≥54						
Extreme Heat Days	6.03	4.12	2.89	2.64	4.11	3.36
Days in 80s	31.31	7.56	30.45	8.37	35.64	6.45
Days in 90s	6.16	4.41	2.94	2.96	5.04	3.77
Average Maximum Temperature	64.73	2.71	64.23	2.56	65.27	2.22
Black Student Percent	0.27	0.27	0.34	0.30	0.40	0.33
Female Student Percentage	0.02	0.14	0.01	0.12	0.00	0.00
Unemployment Rate	7.90	2.41	7.78	3.20	6.13	3.48
Median Income	57,869.21	9,964.29	58,990.05	10,943.22	60,207.16	11,562.44
Rural	0.22	0.42	0.22	0.42	0.41	0.51
Number of Teachers, First School					209	
Second School					209	
Third School					17	

Note: This table presents summary statistics, including the mean, standard deviation, and number for teachers who change schools throughout the sample. All variables are at the school year level and span the 2007-08 through 2017-18 school years.

Table A.17: Differences Between Older and Younger Teachers Who Move

	(1) Age [24,54)	(2) ≥54	(3) Difference (1)-(2)
Change in Extreme Heat Days	-1.57	-2.80	1.23***
	(5.01)	(5.35)	(0.36)
Change in Days in 80s	1.25	-0.48	1.73**
	(7.46)	(8.35)	(0.57)
Change in Days in 90s	-1.66	-2.82	1.15**
	(5.30)	(5.51)	(0.37)
Change in Average Maximum Temperature	0.16	-0.42	0.58**
	(2.56)	(2.61)	(0.18)
Change in Black Student Percent	-0.01	0.06	-0.07**
	(0.28)	(0.36)	(0.02)
Change in Male Student Percent	0.00	0.00	0.00
	(0.15)	(0.20)	(0.01)
Change in Unemployment Rate	-1.64	-0.25	-1.39***
	(2.94)	(3.34)	(0.23)
Change in Median Income	2467.06	1283.71	1183.35**
	(8322.48)	(5961.31)	(414.58)
Change in Rural	-0.03	-0.00	-0.02
	(0.33)	(0.24)	(0.02)
Observations	4734	226	4960

Note: The sample in this table includes teachers who change schools at least once. Columns 1 and 2 show the mean difference in each variable between a teacher's prior and following school, with standard deviations in parentheses. Column 3 shows the difference in mean differences, with the standard error of this difference in parentheses and with stars specifying significance as follows: $^{*}\ p < 0.05$, $^{**}\ p < 0.01$, $^{***}\ p < 0.001$.

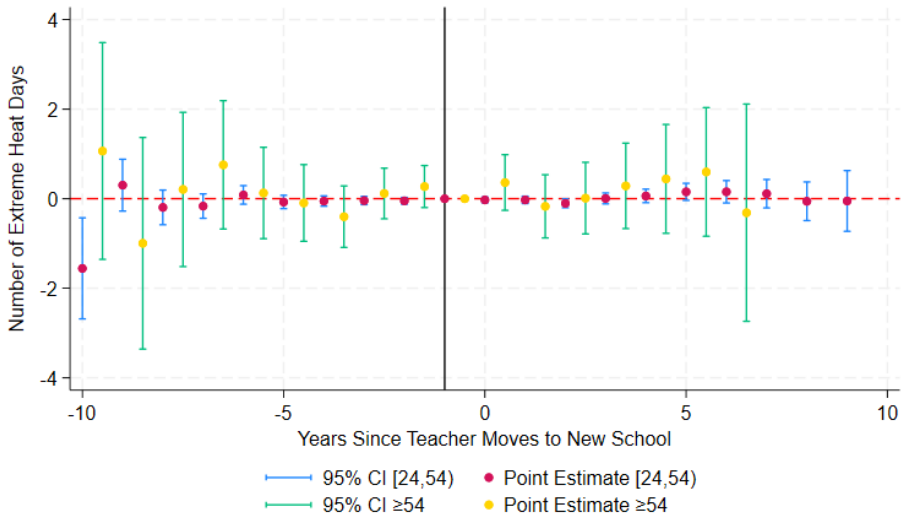


Figure A.2: No evidence of sorting by heat and teacher age

Note: This event study plot presents estimates for the number of extreme heat days relative to a move from one school to another. The x-axis shows the number of years since the move to a new school.

APPENDIX B

Appendix to Chapter 2

Table B.1: Summary of Literature on Effects of College Quality

Citation	Avg. SAT of sample	T vs C (SDs, ITT)	4-yr BA attainment		6-yr BA attainment		Earnings		Context
			Control mean	Treatment effect (ITT)	Control mean	Treatment effect (ITT)	Control mean	Treatment effect (%)	
Present Analysis	1254	0.270	0.390	0.040	0.620	−0.010	37,907	0.072	MI
Chetty et al. (2026)	1374							0.040	U.S.
Cohodes et al. (2026)	1436		0.560	0.053	0.770	0.025	104,213	0.118	U.S.
Mountjoy (2026)	916	0.075	0.100	0.008	0.300	0.018	40,815	0.013	TX
Barr & Castleman (2025)			0.268	0.062	0.528	0.076			MA, NY, IL
Bleemer (2024)	1043	0.220			0.600	0.022	70,000	0.039	CA
Kozakowski (2023)	878	0.060			0.390	0.036	30,577	0.064	MA
Angrist et al. (2022)	1100		0.320	−0.028	0.636	0.081			NE
Goodman et al. (2017)	906	0.020			0.370	0.009			GA
Dale & Krueger (2014) — Overall	1210						139,698	−0.008	U.S.
Dale & Krueger (2014) — Black & Hisp.	1024						98,472	0.076	U.S.
Cohodes & Goodman (2014)$^{+}$	1279	0.018	0.462	0.019	0.631	0.025			MA
Zimmerman (2014)	943		0.050	−0.007	0.200	0.057	28,963	0.051	FL
Hoekstra (2009)	1000							0.135	?

Notes: $^{+}$ signifies the effects are reverse-coded to align with the direction of the other results for easier comparison. Notes on average SAT scores of the sample: We convert Angrist et al. (2022) reported ACT scores to SAT scores. For estimates from Cohodes and Goodman (2014) and Cohodes et al. (2026), we convert reported standardized test scores (in SD units) to SAT scores using NCES-reported SAT means and SDs. Cohodes and Goodman (2014) report that those in the RD sample score about 0.90 standard deviations above average MCAS scores relative to all high school grads in Massachusetts; we use Massachusetts-specific estimates to convert this to an equivalent SAT score. Cohodes et al. (2026) report a standardized math score (a mix of PSAT, SAT, and ACT scores), an average of 1.929 sd, which we convert to an equivalent SAT score using the national mean and standard deviation from 2017. For the paper from Hoekstra (2009), the SAT score we report is based on this description in the text: "the average SAT scores of the entering classes at the flagship over the period of time studied were between 1000 and 1100" (p. 717). We use the lower number since that range is the average, and the RD is for the marginal student. The T vs C contrast column represents the average test scores of students attending institutions in the treated group relative to the comparison group. Unless otherwise noted, these reflect differences in SAT scores between treatment and comparison groups. Goodman et al. (2017) report PSAT scores rather than SAT scores, Mountjoy (2026) reports 11th grade standardized test scores, and Kozakowski (2023) reports MCAS standardized test scores from Massachusetts. Cohodes and Goodman (2014) only report SAT math scores, not composite. All other papers use SAT composite test scores (Zimmerman, 2014; Dale and Krueger, 2014; Chetty et al., 2026; Bleemer, 2024; Barr and Castleman, 2025). SAT score treatment effects converted into standard deviation units using our institution distribution, where a standard deviation unit is approximately 133 points.

Table B.2: Estimated Effect of HAIL on Financial Aid Receipt at the UM Flagship
Adapted from Burland et al. (2023), p. 304

	Control Mean	HAIL mean	P-value
Grants at UM Flagship, First Year	\$24,729	\$24,207	0.360
	(497)	(235)	
Grants at UM Flagship, Years 1–4	\$106,643	\$105,735	0.720
	(2230)	(1011)	
N (Enrolled at UM Flagship)	237	465	

Notes: Adapted from Table 3 in "The power of certainty: Experimental evidence on the effective design of free tuition programs," by Burland, E., Dynarski, S., Michelmore, K., Owen, S., and Raghuraman, S., 2023. American Economic Review: Insights, 5(3): 293–310 (https://doi.org/10.1257/aeri.20220094). Analysis done at the student-level. Includes only students enrolled at the UM Flagship full time and who have financial aid data reported. Standard errors are clustered at the high-school-by-cohort-level. Includes zeros for students who receive no aid. "Grants" includes all institutional and departmental scholarships and grants, federal grants, state grants and scholarships, private scholarships, and other departmental aid.

Table B.3: Estimated Effect of HAIL on Transfer Pathways

	Transfer in Fall 2		Transfer in Fall 3	
Outcome	ITT: Effect of HAIL Offer	Control mean	ITT: Effect of HAIL Offer	Control mean
Any Transfer	0.023	0.080	−0.021	0.070
	(0.012)		(0.011)	
Institution type				
2-year to 4-year	−0.007	0.020	−0.014	0.031
	(0.005)		(0.008)	
4-year to 2-year	−0.002	0.027	−0.004	0.011
	(0.007)		(0.004)	
2-year to 2-year	0.005	0.002	−0.002	0.003
	(0.003)		(0.002)	
4-year to 4-year	0.028	0.031	−0.001	0.023
	(0.009)		(0.006)	
Transfer into UM Flagship				
From 2-year or non-UM 4-year	−0.002	0.004	−0.003	0.004
	(0.002)		(0.002)	
From UM Satellite	0.029	0.001	0.001	0.001
	(0.005)		(0.001)	
Transfer from UM Flagship				
To a 2-year or non-UM 4-year	0.010	0.000	0.003	0.002
	(0.004)		(0.003)	
To a UM Satellite	0.003	0.000	0.001	0.000
	(0.002)		(0.001)	
N school-cohorts	1,026			

Notes: Analyses are conducted at the high-school-by-cohort-level. Robust standard errors reported in parentheses. Treatment effect coefficients are estimates of $\hat{\beta}_1$ from Equation (2.1). Transfers are defined as having different enrollment from one fall to the following fall (fall 1 to fall 2, fall 2 to fall 3). Transfer outcomes are unconditional on initial enrollment. For example, a student who started at a four-year school would be coded as a zero for transfer from two- to four-year, as would a student who started at a two-year and did not transfer.

Table B.4: Estimated Effect of HAIL on Graduate School Outcomes

	ITT		LATE	
	Effect of HAIL offer	Control mean	Effect of Attending Any UM Campus	Control complier mean
First Stage				
Enroll Any UM Campus within 1 year	0.208	0.149		
	(0.021)			
F-statistic	100.141			
ITT and LATE: Graduate School, within 6 Years				
Enrolled in Graduate School	−0.004	0.180	−0.021	0.225
	(0.016)		(0.077)	
Earned Graduate Degree	−0.004	0.053	−0.020	0.066
	(0.009)		(0.044)	
N school-cohorts	1,026			

Notes: Analyses are conducted at the high-school-by-cohort-level. Robust standard errors reported in parentheses. ITT coefficients are estimates of $\hat{\beta}_1$ in Equation (2.1). LATEs are estimates of $\hat{\gamma}_1$ in Equation (2.3), estimated using 2SLS. Control complier means are calculated at the student level, inverse weighted by size of high school cohort. Graduate school enrollment reflects terms that a student is enrolled after they have earned a bachelor's degree, or in terms flagged in the transcript data as graduate enrollments. Graduate degree completion is measured in the summer following the 6th spring and includes any graduate degrees awarded.

Table B.5: Estimated Effect of HAIL on Projected Earnings 8 years Following College Entry, using Alternate Populations to Project Earnings

	ITT		LATE		
	Effect of HAIL offer	Control mean	Effect of Attending Any UM Campus	Control complier mean	Percent change
First Stage					
Enroll Any UM campus within 1 year	0.208	0.149			
	(0.021)				
F-statistic	100.141				
ITT and LATE: Projected earnings (2025$), Overall and by Family Income					
Projected using institutional average	3,393	58,523	16,321	62,342	26%
	(985)		(3,957)		
Projected using students with family income $\leq$ \$30K	3,557	52,569	17,109	55,402	31%
	(912)		(3,524)		
Projected using students with family income \$30K–75K	3,020	58,801	14,528	62,629	23%
	(902)		(3,639)		
Projected using students with family income $>$ \$75K	3,191	64,437	15,350	68,825	22%
	(989)		(4,047)		
N (School-cohorts)			1,026		

Notes: Analyses are conducted at the high-school-by-cohort-level. Robust standard errors reported in parentheses. ITT coefficients are estimates of $\hat{\beta}_1$ in Equation (2.1). LATEs are estimates of $\hat{\gamma}_1$ in Equation (2.3), estimated using 2SLS. Control complier means are calculated at the student level, inverse weighted by size of high school cohort. Institution earnings (both overall and by family income) reflect the median earnings of enrolled students 8 years post college-entry from College Scorecard (2013 entry cohort). Projected earnings are assigned to students based on their modal college attended. Students with two modal institutions are assigned the average for the two modal institutions. Scorecard earnings are only available for students who enrolled in college. Students who did not enroll or attend institutions where earnings are not available are assigned earnings from Opportunity Insights. Earnings are all measured in 2025 values, adjusting for inflation using the consumer price index (CPI-U).

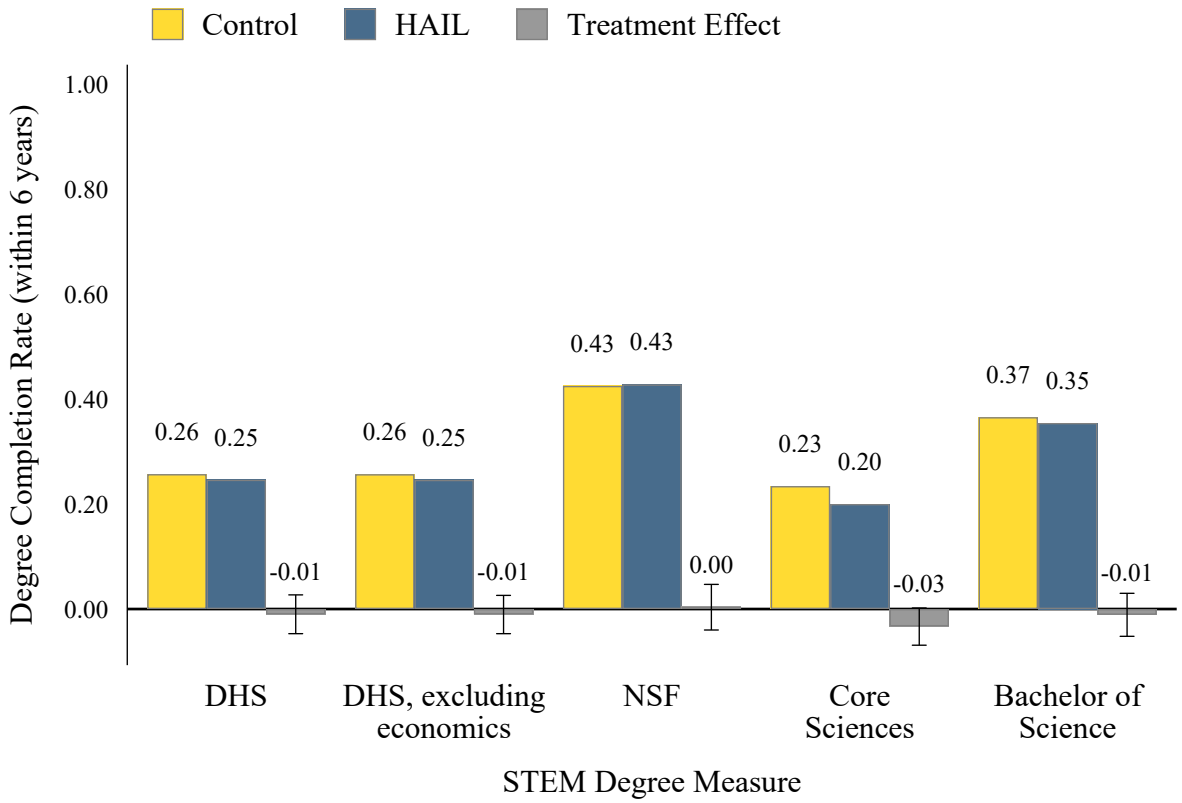


Figure B.1: Effect of HAIL on STEM BA Degree Attainment Using Alternative STEM Definitions

Notes: The first column, "STEM BA (DHS)" is the original definition used in the main analyses. With the exception of "Bachelor of Science" (which is exclusively from NSC), we start with STEM degree information from the STARR dataset and then merge on NSC data. We count a student as having a STEM degree if they are coded as having a STEM degree in either dataset.

B.1 Outcomes and Measurement

B.1.1 Enrollment

We use enrollment from NSC, which we structure into three terms per academic year—fall, spring, and summer—using NSC enrollment dates. We measure enrollment from expected high school graduation. The first HAIL cohort graduated high school in the spring of 2016, therefore "fall 1" is fall 2016, "spring 1" is spring 2017, and "summer 1" is summer 2017. In the transcript data, we enforce the same three-term structure that aligns

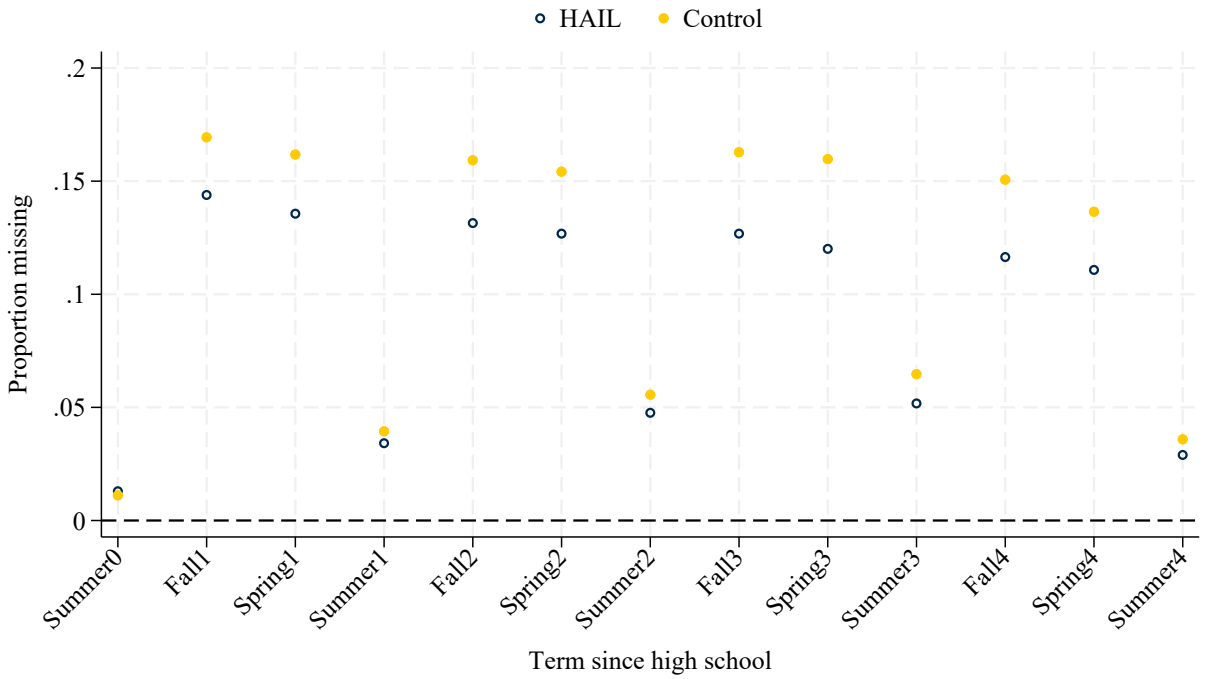


Figure B.2: Missingness by Treatment Status and Academic Term

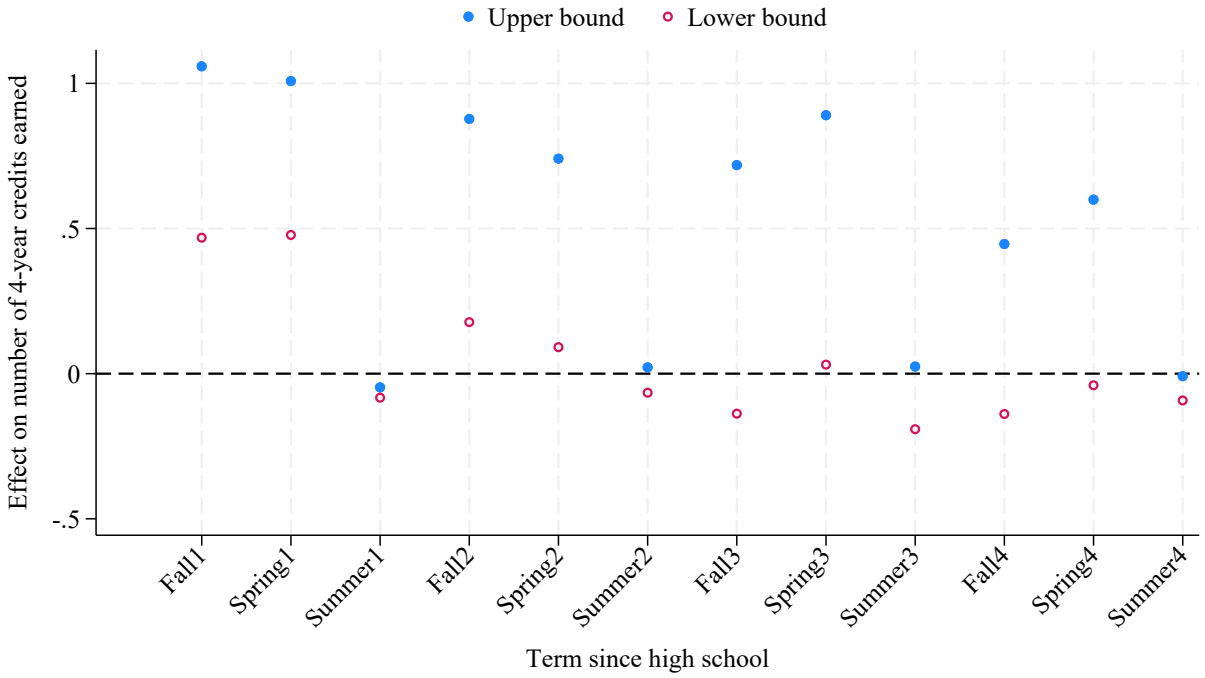


Figure B.3: Lee Bounds on Earned Four-Year Credits, by Academic Term: ITTs

with the NSC enrollment outcomes, which allows us to merge these data at the student-by-term level. For each year, we create mutually-exclusive enrollment outcomes based on the first institution enrolled in a given year. We only capture undergraduate enrollment, which we define as any enrollment terms that occur prior to a student earning their first bachelor's degree in NSC. We also exclude terms defined as "Graduate" in the student transcript data. We also exclude terms where a student was enrolled in a non-degree program according to the student transcript data, or terms where a student was enrolled for fewer than 15 days, which corresponds with the timing to withdraw and still receive a refund on tuition paid.

B.1.2 Degree Attainment

We use degree attainment from NSC, which we also structure into three terms per academic year using dates provided by NSC for degree attainment to align with the enrollment terms. We supplement this with degree attainment from the STARR transcript data. We measure time to degree from expected high school graduation. Degrees "by" a given term include degrees earned in that term, or any prior term. In cases where a student earns multiple bachelor's degrees, we report their first bachelor's degree earned. We measure 4- and 6-year degree attainment in the summer following the 4th and 6th spring. For example, for a student in the first HAIL cohort (expected to graduate high school in 2016), 4-year BA attainment is measured in the summer of 2020 and 6-year BA attainment in the summer of 2022.

B.1.3 STEM Degrees

Our primary measure of STEM uses the Department of Homeland Security's list of STEM CIP codes (United States Department of Homeland Security, 2022). This includes all 2-digit CIP codes classified under the broad categories of biological and biomedical sciences, mathematics and statistics, physical sciences, and engineering, as well as certain 6-digit CIP codes under other categories. We test the robustness of our findings to different definitions of STEM:

- DHS excluding economics: excludes economics from the DHS definition due to the subjectivity of defining only certain economics degrees as "STEM."

- NSF: National Science Foundation's "science and engineering (S&E) fields" (National Science Board, 2024). Using a crosswalk provided by NCES, we identify all of the CIP codes which crosswalk with these broader categories of STEM, at the 4-digit CIP level, and supplement with "degree major" information from NSC when CIP code is not available.
- Core Sciences: Includes any degree in biological sciences, atmospheric sciences, computer sciences, mathematics, chemistry, physics, and engineering (e.g., Griffith, 2010; Riegle-Crumb et al., 2012; Hatfield et al., 2022).
- Bachelor of Science: National Student Clearinghouse degree type is Bachelor of Science.

Appendix Figure B.1 shows the results using each of these definitions of STEM. Regardless of which definition we use, we find no effect of HAIL on STEM degree attainment.

B.1.4 Projected Earnings

To project earnings, we use data from College Scorecard, which reports median adult earnings for all Title IV funded students. We use data for the cohort who entered college in 2013 measured eight years post college entry (College Scorecard, 2021). We calculate each student's modal institution attended between year 1 and year 6 following high school, and merge on the institutional earnings from College Scorecard using OPEID code. Students with multiple modal institutions are assigned the average earnings for their two modal institutions. We measure median earnings in 2025 values, adjusted for inflation using the consumer price index (CPI-U). These data are only available for students who ever enroll. While HAIL does not affect college enrollment overall, we still assign these students to the projected earnings from Opportunity Insights data, measured at age 30 (Chetty et al., 2022), for students who do not enroll in college. We also use these data for students who enroll in institutions with missing earnings data.

B.1.5 Graduate School Enrollment and Degree Attainment

Graduate school enrollment is measured using NSC, supplemented with transcript data, structured the same way as our undergraduate enrollment measures. Students are considered enrolled in graduate school during terms they are enrolled in NSC after they have

already earned a bachelor's degree or if their enrollment is flagged in the transcript data as enrollment in a graduate degree program. Graduate degree completion is measured using NSC, and supplemented with transcript data, and includes any degrees earned that are classified as master's degrees or doctoral degrees (including an M.D., J.D., or Ph.D.). We measure 6-year graduate degree attainment in the summer following the 6th spring. For example, for a student in the first HAIL cohort (expected to graduate high school in 2016), 6-year graduate degree attainment is counted if a student earned a graduate degree during the summer of 2022 or a prior term.

B.1.6 Institution Characteristics

We calculate institution SAT scores by taking the average SAT score (converted from ACT scores using ACT-SAT concordance tables provided by the College Board) of Michigan public high school students who finished high school between 2013–2015 and who enrolled at each institution in their first year. Enrollment for these cohorts is measured the same as for our sample, described above. Institutions with fewer than 20 students enrolled in previous cohorts are assigned the average SAT score for institutions with their same Barron's selectivity and public/private designation. We assign HAIL students the average institution quality for the first school they attend in each of their first and second years following expected high school graduation. Figure 2.1a uses just the first year. Table 2.4 uses an average of the institutions attended in years 1 and 2. Median parental income comes from Opportunity Insights (Chetty et al., 2022). Endowment data comes from the IPEDS finance survey, which we download using Urban Institute's Education Data Portal (U.S. Department of Education, National Center for Education Statistics, 2023). The remaining characteristics use College Scorecard data (College Scorecard, 2016). Dollar amounts are measured in 2025 values, adjusted for inflation using the consumer price index (CPI-U). We merge on the characteristics for the institutions students enrolled in for their first year and for their second year (measured as described above). For Figure 2.1b–d, we just use the characteristics of the first-year enrollment. For Table 2.4, individuals are assigned an average of the institution characteristics for their year 1 institution and their year 2 institution. For students who only enrolled in year 1 following high school, they are assigned the values for their year 1 institution. The same logic is applied for students who only enroll in year 2. Students who do not enroll in years 1 or 2 receive missing values.

B.1.7 Credits

We construct measures of cumulative credits earned while enrolled at four-year institutions and credits earned while enrolled at community colleges. The transcript data does not allow us to distinguish between transferred credits, AP credits, and dual enrollment credits. Counting these credits would result in counting the same credits multiple times, first at the initial institution and again at the institution students transfer to. Therefore, our measures exclusively focus on credits attempted and earned at a given school.[1] We also focus exclusively on undergraduate credits, using our enrollment measure to exclude any graduate credits. Students not enrolled anywhere in a given term are assigned zero credits. For those who do not attend public colleges within Michigan (i.e., are enrolled, but we do not have transcript data for them), we impute credits attempted and credits earned (described below).

Selection into the Transcript Sample. For those students who attend public colleges within Michigan, we have access to student transcript data. While these transcripts are far richer than the NSC data, we only have them for students attending public colleges in Michigan. Students attending private colleges, and those attending college outside Michigan, are missing from these data. This could induce selection bias: the HAIL offer plausibly could have affected the likelihood that a student appears in the STARR data by increasing the share of students attending a public college in Michigan. We therefore use non-missing STARR data combined with universal NSC enrollment data to impute missing credit information for those students who do not attend public colleges within Michigan. Appendix Table B.6 shows the effect of HAIL on selection into the transcript sample by comparing NSC enrollment (with universal coverage) to enrollment calculated from the transcript data. The estimated effect of the HAIL offer on attending a public college in Michigan, and therefore having any credits reported in the transcript data, is 4.1 percentage points, with a standard error of 2.0 percentage points. Since HAIL induced students into the STARR sample, a simple comparison of credits attempted and earned between the treatment and control groups is likely biased in favor of finding a positive treatment effect, since we are more likely to observe credits for students in the treatment

[1] After this cleaning process, 99 percent of student-term observations have 18 or fewer credits attempted or earned, the standard upper-bound for full-time enrollment. However, the cleaning process is imperfect, and one percent of student-terms have credits attempted or earned that exceed 18. We top-code credits in any given semester (regardless of the type of institution) at 22 credits (changing the credits attempted/earned for 0.1 percent of cases). This does not meaningfully change the average credits attempted or earned.

group.

Table B.6: Effect of HAIL on Selection into the Transcript Data Sample

Outcome	Effect of HAIL offer (ITT)	Control mean
Enrolled at any college, Year 1 (NSC)	−0.001	0.905
	(0.014)	
Attempted any credits, Year 1 (transcript data)	0.041	0.730
	(0.020)	
N school-cohorts	1,026	

Notes: All analyses done at the high-school-by-cohort-level. Robust standard errors reported in parentheses. Treatment effect coefficients are from estimating Equation (2.1), reporting estimates of $\hat{\beta}_1$. Credits calculated at the term-level and include undergraduate coursework taken at a given institution and collapsed to the student-year level. The difference between the treatment effect on enrollment in NSC and the treatment effect on credits attempted in the transcript data represents selection into the transcript sample.

Credit Imputation. To address this, we impute credits attempted and earned for students who are enrolled in either a private institution within Michigan, or an institution outside the state of Michigan. We follow a similar procedure to that of Angrist et al. (2022). For terms where students attend Michigan public institutions, we have their actual credits attempted and earned. For terms where students attend out-of-state institutions or in-state private institutions, we impute credits (attempted and earned) at the student-by-term level, separately for 2- and 4-year enrollments. To do so, we regress credits on enrollment term (fall, spring, summer), year (1–6 post high school graduation), and enrollment status (full-time, half-time, less than half-time, withdrawn/deceased). For four-year enrollments, we also include indicators for Barron's competitive index (highly or most competitive, very competitive or competitive, less-competitive or non-competitive four-year school) and the interaction between Barron's category and enrollment status. We then assign student-terms with missing credits their predicted values from the regression. Credits are imputed for 8.7 percent of student-terms (7.3 percent for treated students, 9.2 percent for control-group students).

Lee Bounds. In addition to our primary specification, we also estimate Lee Bounds (Lee, 2009). Lee Bounds allow us to bound our treatment effects on credits by modeling the likelihood of missingness from the sample as a latent feature that is predicted by treatment assignment. The lower bound assumes all additional treated students who select into the

sample are the top performers, and the upper bound assumes they are the bottom performers. In practice, this means trimming the highest values of earned credits for a portion of the sample equivalent to the differential attrition rate, then estimating treatment effects on the trimmed sample (lower bound); the upper bound is estimated on a sample with the equivalent proportion of the lowest credits trimmed. Lee Bounds do not bound the average treatment effect on all students. Rather, the approach bounds the effect for individuals for whom we would have observed credit outcomes regardless of treatment. For each semester, we calculate the treatment-control difference in missingness for each credit outcome, at the student-level. We define missing as a term where a student is observed as enrolled in NSC, and missing in the transcript data. In each semester, we drop the top and bottom N_{drop} observations, where N_{drop} is equal to the differential missingness rate times the number of treated students. For example, in Fall 1, the missing rate for control is 16.94 percent, and the rate for treated is 14.39 percent, for a difference of 2.55 percent. There are 1,932 treated students. Thus in Fall 1 $N_{\text{drop}} = 0.0255 \times 1{,}932 = 49$ students. Appendix Figure B.2 shows the missingness by treatment status and academic term. The difference between these two is the differential missingness rate. We estimate treatment effects with the trimmed variables at the school-by-cohort level, using Equation (2.1) (i.e., controlling for strata and clustering at the high-school level). The lower bound estimates the effect of HAIL on the credit outcome, dropping the top N_{drop} observations from the credit outcome. The upper bound does the same, dropping the lower N_{drop} observations. Appendix Figure B.3 shows the bounds on earned credits from four-year institutions in each academic term. The treatment effects using Lee Bounds are largely consistent with the results of our primary approach using imputation: effects on credits earned in fall and spring of years 1 through 4 are likely positive, effects on credits earned in years 5 and 6 are likely negative, and credits effects in summer terms are close to zero.

B.1.8 Randomization Inference

Along with traditional sampling-based inference, for our primary intent-to-treat effects in Table 2.2 we also report exact p-values from randomization-based inference. This approach calculates the likelihood of obtaining the observed treatment effects by random chance, where the randomness comes from assignment of a fixed number of units (in our case, high schools) to treatment, rather than from random sampling from a population. Randomization inference tests the sharp null hypothesis of no treatment effect for any unit (Athey and Imbens, 2017). Using the same procedure used in the original randomization,

we randomly reassign treatment status to the 1,026 school-cohorts in our sample. We then estimate "treatment effects" based on this reassignment, controlling for randomization strata as in Equation (2.1). We repeat this procedure 1,000 times to generate a distribution of potential treatment effects that could be due to baseline differences between schools assigned to treatment and control. For each outcome, we calculate the share of the 1,000 simulated treatment-control differences that are larger in absolute value than the difference observed in the actual random assignment discussed throughout the paper. This proportion represents the randomization-based p-value. Appendix Table B.7 compares sampling-based p-values to randomization-inference based p-values. None of the p-values for a given outcome differ by more than 0.02, and the conclusions are unchanged by the inference approach.

Table B.7: Comparison of Sampling- and Randomization-Based Inference

	Effect of HAIL Offer (ITT)	Sampling-based p-value	Randomiz. Inference-based p-value
First Stage			
Enroll any UM campus within 1 year	0.208	0.000	0.000
Outcomes			
Any BA within 4 years	0.041	0.060	0.070
Any BA within 6 years	−0.006	0.768	0.758
Highly competitive+ BA within 6 years	0.117	0.000	0.000
STEM BA within 6 years	−0.010	0.583	0.596
Start at UM satellite and earn any BA within 6 years	0.040	0.000	0.000
Start at UM Satellite and earn highly comp.+ BA within 6 years	0.023	0.000	0.000
Projected earnings, using institution average	3,393	0.001	0.002
Projected earnings, using family income $\leq$ \$30K	3,557	0.000	0.000

Notes: Analysis done at the school-cohort level. Table reports standard sampling-based p-values, which correspond to the estimates and robust standard errors in Table 2.2, compared to randomization-inference-based p-values. Randomization inference p-values are from 1000 treatment assignment iterations, and test a sharp null hypothesis of no effect for any unit.

B.1.9 Complier Analysis

To calculate control complier outcome means (Tables 2.2 and 2.4), as well as average pre-treatment characteristics for the latent sub-populations defined by the LATE framework (Table 2.3), we use the methods from Imbens and Rubin (1997b) and formally extended by Marbach and Hangartner (2020). By relying on the same assumptions required to estimate LATEs—namely, the exclusion restriction and the absence of defiers—we can directly estimate means for always takers, never takers, and compliers. First, we estimate the population shares using the sample analogues. For this to work, we must assume there are no "defiers"—students who are less likely to enroll at UM when they receive a HAIL offer relative to no offer (Imbens and Rubin, 1997b). We calculate the proportion of always-takers as the share of the control group who enroll at UM:

$$p_{\text{at}} = E[\text{EnrollUM} \mid \text{Treat} = 0] \tag{B.1}$$

Similarly, the proportion of never-takers is the share of the treatment group who do not enroll:

$$p_{\text{nt}} = 1 - E[\text{EnrollUM} \mid \text{Treat} = 1] \tag{B.2}$$

Then, the proportion of compliers is simply the remainder (assuming no defiers):

$$p_{\text{cm}} = 1 - p_{\text{at}} - p_{\text{nt}} \tag{B.3}$$

Next, we can calculate outcome means for always- and never-takers. By assuming their behavior is not affected by treatment, we are assuming control and treated always (and never) takers have the same outcomes in expectation. The always-taker outcome mean is:

$$\bar{Y}_{\text{at}} = E[Y \mid \text{Treat} = 0,\ \text{EnrollUM} = 1] \tag{B.4}$$

And the never-taker mean is the outcome mean among treated students who do not enroll:

$$\bar{Y}_{\text{nt}} = E[Y \mid \text{Treat} = 1,\ \text{EnrollUM} = 0] \tag{B.5}$$

We consider the overall control mean for any outcome of interest to be a weighted average across three types of students: always-takers, never-takers, and compliers. We directly observe outcomes for the always-takers—those students in the control group who enroll at UM. Assuming control and treated always-takers (and never-takers) have the same outcomes in expectation, we also observe the outcomes for the never-takers—those students in the treatment group who do not enroll at UM. We can then use the fact that the overall control mean is a weighted average across the three types of students to solve for the CCM as a function of observable quantities:

$$\bar{Y}_{\text{cm}} = \frac{\bar{Y}_{\text{control}} - p_{\text{at}}\bar{Y}_{\text{at}} - p_{\text{nt}}\bar{Y}_{\text{nt}}}{p_{\text{cm}}} \tag{B.6}$$

We estimate each of these at the student-level, since compliance is an individual-level behavior. However, we weight all averages by the inverse of the school size to be comparable with our high-school-by-cohort-level treatment effects. Estimated CCMs for outcomes that should be bounded at zero can be negative, and are in some instances. Treated and control students (and all sub-populations, including compliers, which are defined by pre-treatment characteristics) are by assumption identical in expectation but may not be in our finite sample. Kling et al. (2007) describe this in detail: for binary outcomes, sampling variation can produce negative estimates of the CCM. Their method assumes that there would have been the same fraction of noncompliers in the control group as were observed

in the treatment group, and that these noncompliers had exactly the same outcome prevalence as treatment noncompliers. In any one sample, this method may produce a negative estimate of the CCM if a particular realization of the treatment noncomplier mean is higher than the realization of the control noncomplier mean, even though the method is unbiased in repeated sampling.

APPENDIX C

Appendix to Chapter 3

C.1 Robustness checks

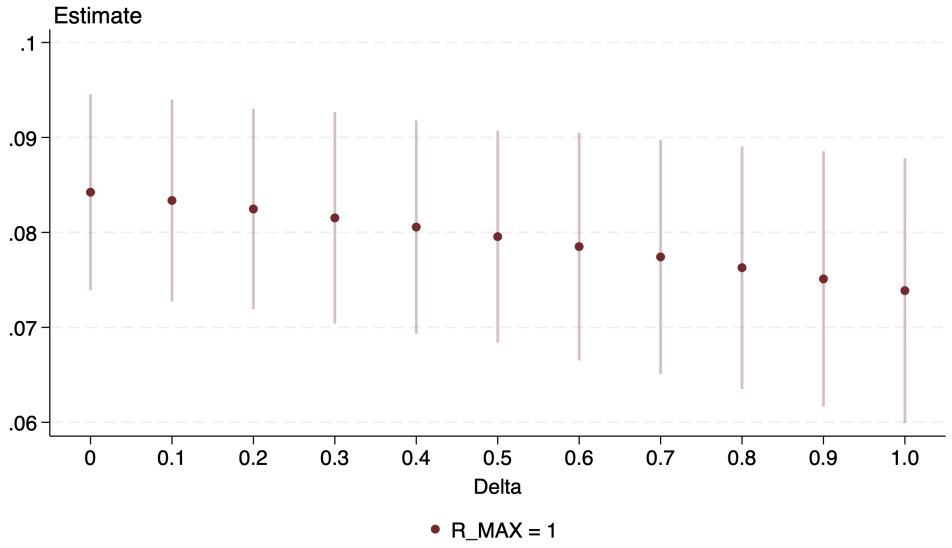


Figure C.1: Coefficient Stability Using Oster Method

This figure plots the estimate of the regression of SNAP enrollees on gasoline prices and covariates as specified in Equation 3.3. These estimates are plotted against the levels of Delta, a parameter indicating the relative degree of selection on observed and unobserved variables (Altonji et al., 2005; Oster, 2019).

Table C.1: Regressions of SNAP on gasoline prices using alternative lag structures

	(1) log(SNAP)	(2) log(SNAP)	(3) log(SNAP)	(4) Per Capita SNAP	(5) Per Capita SNAP	(6) Per Capita SNAP
Gasoline price (lag)	0.0876***			0.00674**		
	(0.0237)			(0.00225)		
Unemployment (lag)	0.0312***			0.00389***		
	(0.00418)			(0.000538)		
CPI (lag)	-0.0102**			-0.000587		
	(0.00324)			(0.000375)		
Gasoline price (lag 2)		0.103***			0.00817**	
		(0.0265)			(0.00250)	
Unemployment (lag 2)		0.0305***			0.00379***	
		(0.00407)			(0.000525)	
CPI (lag 2)		-0.0109**			-0.000658	
		(0.00345)			(0.000395)	
Gasoline price (lag 3)			0.108***			0.00874***
			(0.0269)			(0.00252)
Unemployment (lag 3)			0.0293***			0.00366***
			(0.00399)			(0.000510)
CPI (lag 3)			-0.0130**			-0.000852*
			(0.00386)			(0.000431)
Obs	13107	13056	13005	13107	13056	13005

Standard errors in parentheses are robust to heteroskedasticity and are clustered at the state level. All regressions include state and month fixed effects. * p<.05, ** p<.01, *** p<.001.

BIBLIOGRAPHY

Abdulkadiroğlu, Atila, Parag A. Pathak, and Christopher R. Walters, "Who Gets What in Education: Can School Matching Improve Student Achievement?," NBER Working Paper w34936, National Bureau of Economic Research 2026.

Acton, Riley, Kalena E. Cortes, Lester Miller, and Camila Morales, "Distance to degrees: How college proximity shapes students' enrollment choices and attainment across race-ethnicity and socioeconomic status," *Economics of Education Review*, 2025, p. 102724.

Alm, James, Edward Sennoga, and Mark Skidmore, "Perfect Competition, Urbanization, and Tax Incidence in the Retail Gasoline Market," *Economic Inquiry*, January 2009, *47* (1), 118–134.

Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber, "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools," *Journal of Political Economy*, 2005, *113* (1), 151–184.

Angrist, Joshua D. and Alan B. Krueger, "Does compulsory school attendance affect schooling and earnings?," *Quarterly Journal of Economics*, 1991, *106* (4), 979–1014.

Angrist, Joshua, David Autor, and Amanda Pallais, "Marginal Effects of Merit Aid for Low-Income Students," *The Quarterly Journal of Economics*, 2022, *137* (2), 1039–1090.

Arcidiacono, Peter, Esteban Aucejo, Patrick Coate, and V. Joseph Hotz, "Affirmative action and university fit: Evidence from Proposition 209," *IZA Journal of Labor Economics*, 2014, *3* (1), 7.

_ , Esteban M. Aucejo, and V. Joseph Hotz, "University Differences in the Graduation of Minorities in STEM Fields: Evidence from California," *American Economic Review*, 2016, *106* (3), 525–562.

Athey, Susan and Guido W. Imbens, "The Econometrics of Randomized Experiments," *Handbook of Economic Field Experiments*, 2017, *1*, 73–140.

Auffhammer, Maximilian, Solomon M. Hsiang, Wolfram Schlenker, and Adam Sobel, "Using Weather Data and Climate Model Output in Economic Analyses of Climate Change," *Review of Environmental Economics and Policy*, July 2013, *7* (2), 181–198.

Autor, David, Claudia Goldin, and Lawrence F. Katz, "Extending the Race between Education and Technology," *AEA Papers and Proceedings*, 2020, *110*, 347–351.

Barr, Andrew and Benjamin Castleman, "Increasing Degree Attainment among Low-Income Students: The Role of Intensive Advising and College Quality," *American Economic Review*, 2025, *115* (11), 4075–4103.

Baysan, Ceren, Marshall Burke, Felipe González, Solomon Hsiang, and Edward Miguel, "Non-economic factors in violence: Evidence from organized crime, suicides and climate in Mexico," *Journal of Economic Behavior & Organization*, December 2019, *168*, 434–452.

Beatty, Timothy K. M., Laura Blow, and Thomas F. Crossley, "Is there a 'heat-or-eat' trade-off in the UK?," *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, 2014, *177* (1), 281–294.

Benhenda, Asma, "Absence, substitutability and productivity: Evidence from teachers," *Labour Economics*, June 2022, *76*, 102167.

Bhattacharya, Jayanta, Thomas DeLeire, Steven Haider, and Janet Currie, "Heat or Eat? Cold-Weather Shocks and Nutrition in Poor American Families," *American Journal of Public Health*, 2003, *93* (7).

Bitler, Marianne and Hilary W. Hoynes, "The State of the Safety Net in the Post-Welfare Reform Era," *Brookings Papers on Economic Activity*, 2010, *41* (2), 71–127.

Black, Dan A. and Jeffrey A. Smith, "Estimating the Returns to College Quality with Multiple Proxies for Quality," *Journal of Labor Economics*, 2006, *24* (3), 701–728.

Blair, Peter Q. and Kent Smetters, "Why Don't Elite Colleges Expand Supply?," NBER Working Paper w29309, National Bureau of Economic Research 2021.

Bleemer, Zachary, "Affirmative Action, Mismatch, and Economic Mobility after California's Proposition 209," *The Quarterly Journal of Economics*, 2022, *137* (1), 115–160.

_ , "Top Percent Policies and the Return to Postsecondary Selectivity," 2024. Unpublished manuscript.

_ **and Sarah Quincy**, "Changes in the College Mobility Pipeline since 1900," NBER Working Paper w33797, National Bureau of Economic Research 2025.

Blisard, Noel, David Smallwood, and Steve Lutz, "Food Cost Indexes for Low-Income Households and the General Population," Technical Report, United States Department of Agriculture, Economic Research Service 1999.

Bouchama, Abderrezak and James P. Knochel, "Heat Stroke," *New England Journal of Medicine*, Jun 2002, *346* (25), 1978–1988.

Bound, John, Michael F. Lovenheim, and Sarah Turner, "Increasing Time to Baccalaureate Degree in the United States," *Education Finance and Policy*, 2012, *7* (4), 375–424.

Burland, Elizabeth, Jasmina Camo-Biogradlija, Kate Ferrara-Gerson, Xandria Fields, Katherine Michelmore, Nicholas Sotherland, Kevin Stange, Melinda Thompson, and Megan Tompkins-Stange, "Bureaucracy and Burden: Understanding Take-Up of a Need-Based Financial Aid Program," EdWorkingPaper 25-1283, Annenberg Institute at Brown University 2025.

_ **, Susan Dynarski, Katherine Michelmore, Stephanie Owen, and Shwetha Raghuraman**, "The Power of Certainty: Experimental Evidence on the Effective Design of Free Tuition Programs," *American Economic Review: Insights*, 2023, *5* (3), 293–310.

Card, David, "The causal effect of education on earnings," in Orley Ashenfelter and David Card, eds., *Handbook of Labor Economics*, Vol. 3A, Elsevier, 1999, pp. 1801–1863.

Casey, Gregory, Stephie Fried, and Matthew Gibson, "Impact of U.S. Labor Productivity Losses from Extreme Heat," *Federal Reserve Bank of San Francisco Economic Letter 2024-14*, Jul 2024.

Chang, Tom Y., Joshua Graff Zivin, Tal Gross, and Matthew Neidell, "The Effect of Pollution on Worker Productivity: Evidence from Call Center Workers in China," *American Economic Journal: Applied Economics*, January 2019, *11* (1), 151–172.

Chetty, Raj, David J. Deming, and John N. Friedman, "Diversifying Society's Leaders? The Determinants and Causal Effects of Admission to Highly Selective Private Colleges," *The Quarterly Journal of Economics*, 2026, *141* (1), 51–145.

_ **, John Friedman, Emmanuel Saez, Nicholas Turner, and Danny Yagan**, "Income Segregation and Intergenerational Mobility Across Colleges in the United States," *The Quarterly Journal of Economics*, 2020, *135* (3), 1567–1633.

_ **,** _ **,** _ **,** _ **, and** _ , "Replication Data for: Income Segregation and Intergenerational Mobility Across Colleges in the United States," 2022. Version 1.0 [Dataset].

Cohen, François and Fidel Gonzalez, "Understanding the Link between Temperature and Crime," *American Economic Journal: Economic Policy*, May 2024, *16* (2), 480–514.

Cohodes, Sarah R. and Joshua S. Goodman, "Merit Aid, College Quality, and College Completion: Massachusetts' Adams Scholarship as an In-Kind Subsidy," *American Economic Journal: Applied Economics*, 2014, *6* (4), 251–285.

_ **, Helen Ho, and Silvia C. Robles**, "Diversifying the STEM Pipeline: Evidence from STEM Summer Programs for Underrepresented Youth," *Journal of Human Resources*, 2026.

Coibion, Olivier and Yuriy Gorodnichenko, "Is the Phillips Curve Alive and Well after All? Inflation Expectations and the Missing Disinflation," *American Economic Journal: Macroeconomics*, January 2015, *7* (1), 197–232.

College Scorecard, "College Scorecard Institution-Level Data," 2016. [Dataset]. https://collegescorecard.ed.gov/data/ (accessed December 17, 2025).

_ , "College Scorecard Institution-Level Data," 2021. [Dataset]. https://collegescorecard.ed.gov/data/ (accessed December 17, 2025).

Costa, Hélia, Guido Franco, Filiz Unsal, Sarath Mudigonda, and Maria Paula Caldas, "The heat is on: Heat stress, productivity and adaptation among firms," *OECD Economics Department Working Papers*, Dec 2024.

Cunnyngham, Karen, "Reaching Those in Need: Estimates of State Supplemental Nutrition Assistance Program Participation Rates in 2020," Final report submitted to the U.S. Department of Agriculture, Food and Nutrition Service, Mathematica August 2023.

Currie, Janet, "The Take-Up of Social Benefits," in Alan J. Auerbach, David Card, and John M. Quigley, eds., *Public Policy and the Income Distribution*, New York: Russell Sage Foundation, 2006, pp. 80–148.

Dale, Stacy Berg and Alan B. Krueger, "Estimating the Payoff to Attending a More Selective College: An Application of Selection on Observables and Unobservables," *The Quarterly Journal of Economics*, 2002, *117* (4), 1491–1527.

_ **and** _ , "Estimating the Effects of College Characteristics over the Career Using Administrative Data," *Journal of Human Resources*, 2014, *49* (2), 323–358.

Daponte, Beth Osborne, Seth Sanders, and Lowell Taylor, "Why Do Low-Income Households not Use Food Stamps? Evidence from an Experiment," *Journal of Human Resources*, None 1999, *34* (3), 612–628.

Dell, Melissa, Benjamin F. Jones, and Benjamin A. Olken, "What Do We Learn from the Weather? The New Climate-Economy Literature," *Journal of Economic Literature*, September 2014, *52* (3), 740–798.

Dillon, Eleanor Wiske and Jeffrey A. Smith, "The Consequences of Academic Match between Students and Colleges," *Journal of Human Resources*, 2020, *55* (3), 767–808.

Doyle, Joseph J. and Krislert Samphantharak, "$2.00 Gas! Studying the Effects of a Gas Tax Moratorium," *Journal of Public Economics*, April 2008, *92* (3–4), 869–884.

Dynarski, Susan, C.J. Libassi, Katherine Michelmore, and Stephanie Owen, "Closing the Gap: The Effect of Reducing Complexity and Uncertainty in College Pricing on the Choices of Low-Income Students," *American Economic Review*, 2021, *111* (6), 1721–1756.

Edelstein, Paul and Lutz Kilian, "How Sensitive Are Consumer Expenditures to Retail Energy Prices?," *Journal of Monetary Economics*, 2009, *56* (6), 766–779.

Finkelstein, Amy and Matthew J Notowidigdo, "Take-Up and Targeting: Experimental Evidence from SNAP," *The Quarterly Journal of Economics*, August 2019, *134* (3), 1505–1556.

Ganong, Peter and Jeffrey B. Liebman, "The Decline, Rebound, and Further Rise in SNAP Enrollment: Disentangling Business Cycle Fluctuations and Policy Changes," *American Economic Journal: Economic Policy*, 2018, *10* (4), 153–176.

Gelman, Michael, Yuriy Gorodnichenko, Shachar Kariv, and Dmitri Koustas, "The Response of Consumer Spending to Changes in Gasoline Prices," *American Economic Journal: Macroeconomics*, 2020, *12* (2), 1–35.

Goldrick-Rab, Sara, *Paying the Price: College Costs, Financial Aid, and the Betrayal of the American Dream*, The University of Chicago Press, 2016. D. M. Anderson and P. Kinsley, Eds.

Goodman, Joshua, Michael Hurwitz, and Jonathan Smith, "Access to 4-Year Public Colleges and Degree Completion," *Journal of Labor Economics*, 2017, *35* (3), 829–867.

Gray, Colin, "Leaving benefits on the table: Evidence from SNAP," *Journal of Public Economics*, 2019, *179*, 104054.

Griffith, Amanda L., "Persistence of Women and Minorities in STEM Field Majors: Is It the School That Matters?," *Economics of Education Review*, 2010, *29* (6), 911–922.

Han, Shurong, Lin Dong, Yulei Weng, and Jianjun Xiang, "Heat exposure and productivity loss among construction workers: a meta-analysis," *BMC Public Health*, November 2024, *24* (1). Publisher: Springer Science and Business Media LLC.

Hanushek, Eric, Steven Rivkin, and John Kain, "Teachers, Schools, and Academic Achievement," *Econometrica*, 02 2005, *73*, 417–458.

Hastings, Justine and Jesse M. Shapiro, "How Are SNAP Benefits Spent? Evidence from a Retail Panel," *American Economic Review*, December 2018, *108* (12), 3493–3540.

Hatfield, Neil, Nathan Brown, and Chad M. Topaz, "Do Introductory Courses Disproportionately Drive Minoritized Students Out of STEM Pathways?," *PNAS Nexus*, 2022, *1* (4), pgac167.

Heckman, James J., Seong Hyeok Moon, Rodrigo Pinto, Peter A. Savelyev, and Adam Q. Yavitz, "The rate of return to the HighScope Perry Preschool Program," *Journal of Public Economics*, 2010, *94* (1–2), 114–128.

Hendren, Nathaniel and Ben Sprung-Keyser, "A Unified Welfare Analysis of Government Policies," *Quarterly Journal of Economics*, 2020, *135* (3), 1209–1318.

Herd, Pamela and Donald P. Moynihan, *Administrative Burden: Policymaking by Other Means*, New York: Russell Sage Foundation, 2019.

Herrmann, Mariesa A. and Jonah E. Rockoff, "Worker Absence and Productivity: Evidence from Teaching," *Journal of Labor Economics*, 2012, *30* (4), 749–782.

Hoekstra, Mark, "The Effect of Attending the Flagship State University on Earnings: A Discontinuity-Based Approach," *Review of Economics and Statistics*, 2009, *91* (4), 717–724.

Hyman, Joshua, "ACT for All: The Effect of Mandatory College Entrance Exams on Postsecondary Attainment and Choice," *Education Finance and Policy*, 2017, *12* (3), 281–311.

Imbens, Guido W. and Donald B. Rubin, "Estimating Outcome Distributions for Compliers in Instrumental Variables Models," *The Review of Economic Studies*, 1997, *64* (4), 555–574.

_ **and** _ , "Estimating Outcome Distributions for Compliers in Instrumental Variables Models," *The Review of Economic Studies*, 1997, *64* (4), 555–574.

Jack, Anthony Abraham, *The Privileged Poor: How Elite Colleges Are Failing Disadvantaged Students*, Harvard University Press, 2019.

_ **and Zoila Black**, "Belonging and Boundaries at an Elite University," *Social Problems*, 2024, *71* (4), 996–1013.

Jackson, C. Kirabo and Claire L. Mackevicius, "What Impacts Can We Expect from School Spending Policy? Evidence from Evaluations in the United States," *American Economic Journal: Applied Economics*, 2024, *16* (1), 412–446.

Kenney, W. Larry, "Control of heat-induced cutaneous vasodilatation in relation to age," *European Journal of Applied Physiology and Occupational Physiology*, 1988, *57* (1), 120–125.

Kilian, Lutz and Xiaoqing Zhou, "Oil prices, gasoline prices, and inflation expectations," *Journal of Applied Econometrics*, 2022, *37* (5), 867–881.

Kling, Jeffrey R., Jeffrey B. Liebman, and Lawrence F. Katz, "Experimental Analysis of Neighborhood Effects," *Econometrica*, 2007, *75* (1), 83–119.

Kozakowski, Whitney, "Are Four-Year Public Colleges Engines for Economic Mobility? Evidence from Statewide Admissions Thresholds," EdWorkingPaper 23-727, Annenberg Institute at Brown University 2023.

Kunkel, K.E., D.R. Easterling, A. Ballinger, S. Bililign, S.M. Champion, D.R. Corbett, K.D. Dello, J. Dissen, G.M. Lackmann, R.A. Luettich Jr., L.B. Perry, W.A. Robinson, L.E. Stevens, B.C. Stewart, and A.J. Terando, "North Carolina Climate Science Report," Technical Report, North Carolina Institute for Climate Studies 2020.

Ladd, Helen F. and Lucy C. Sorensen, "Returns to Teacher Experience: Student Achievement and Motivation in Middle School," *Education Finance and Policy*, April 2017, *12* (2), 241–279.

Lee, David S., "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects," *The Review of Economic Studies*, 2009, *76* (3), 1071–1102.

Light, Audrey and Wayne Strayer, "Determinants of College Completion: School Quality or Student Ability?," *The Journal of Human Resources*, 2000, *35* (2), 299.

Luber, George and Michael McGeehin, "Climate Change and Extreme Heat Events," *American Journal of Preventive Medicine*, November 2008, *35* (5), 429–435. Publisher: Elsevier BV.

Marbach, Moritz and Dominik Hangartner, "Profiling Compliers and Noncompliers for Instrumental-Variable Analysis," *Political Analysis*, 2020, *28* (3), 435–444.

Marion, Justin and Erich Muehlegger, "Fuel Tax Incidence and Supply Conditions," *Journal of Public Economics*, October 2011, *95* (9–10), 1202–1212.

McCormack, Kristen, "Education Under Extremes: Temperature, Student Absenteeism, and Disciplinary Infractions," *Job Market Paper*, 2023.

Miller, Lester, "Switching Schools: Effects of College Transfers," EdWorkingPaper 25-1159, Annenberg Institute at Brown University 2025.

Moffitt, Robert, "An Economic Model of Welfare Stigma," *The American Economic Review*, 1983, *73* (5), 1023–1035.

Mountjoy, Jack, "Marginal Returns to Public Universities," *The Quarterly Journal of Economics*, 2026, *141* (1), 429–497.

National Science Board, "The State of U.S. Science and Engineering 2024," Technical Report, National Science Foundation 2024.

Oster, Emily, "Unobservable Selection and Coefficient Stability: Theory and Evidence," *Journal of Business & Economic Statistics*, 2019.

Papay, John P. and Matthew A. Kraft, "Productivity returns to experience in the teacher labor market: Methodological challenges and new evidence on long-term career improvement," *Journal of Public Economics*, October 2015, *130*, 105–119.

Park, R. Jisung, A. Patrick Behrer, and Joshua Goodman, "Learning is inhibited by heat exposure, both internationally and within the United States," *Nature Human Behaviour*, October 2020, *5* (1), 19–27.

_ , Joshua Goodman, Michael Hurwitz, and Jonathan Smith, "Heat and Learning," *American Economic Journal: Economic Policy*, May 2020, *12* (2), 306–339.

Riegle-Crumb, Catherine, Barbara King, Eric Grodsky, and Chandra Muller, "The More Things Change, the More They Stay the Same? Examining Gender Equality in Prior Achievement and Entry into STEM College Majors over Time," *American Educational Research Journal*, 2012, *49* (6).

Roach, Travis and Jacob Whitney, "Heat and learning in elementary and middle school," *Education Economics*, January 2022, *30* (1), 29–46.

Sacerdote, Bruce, "When the Saints Go Marching Out: Long-Term Outcomes for Student Evacuees from Hurricanes Katrina and Rita," *American Economic Journal: Applied Economics*, January 2012, *4* (1), 109–135.

Sajid, Osama, Jennifer Ifft, and Ariel Ortiz-Bobea, "The impact of extreme heat on farm finances: farm-level evidence from Kansas," 2024.

Schlenker, Wolfram, "Daily Weather Data - Continental USA (Version 2018)," 2024. data retrieved from https://zenodo.org/records/10625288.

_ **and Michael J. Roberts**, "Nonlinear temperature effects indicate severe damages to U.S. crop yields under climate change," *Proceedings of the National Academy of Sciences*, September 2009, *106* (37), 15594–15598.

Somanathan, E., Rohini Somanathan, Anant Sudarshan, and Meenu Tewari, "The Impact of Temperature on Productivity and Labor Supply: Evidence from Indian Manufacturing," *Journal of Political Economy*, June 2021, *129* (6), 1797–1827.

United States Department of Homeland Security, "DHS STEM Designated Degree Program List," Technical Report, United States Department of Homeland Security 2022.

United States Environmental Protection Agency, "Changes over Time — EnviroAtlas," https://www.epa.gov/enviroatlas/changes-over-time 2025. Last updated July 14, 2025.

University of Michigan Office of the Registrar, "Enrollment Summary, Fall 2015–Fall 2019," n.d. Accessed June 18, 2026. https://ro.umich.edu/sites/default/files/attachments/19enrollmentsummary.pdf.

University Record, "U-M Breaks Enrollment Record, Welcomes 53,000 Students for Fall Term," 2025. September 26, 2025. Accessed June 18, 2026. https://record.umich.edu/articles/u-m-breaks-enrollment-record-welcomes-53000-students-for-fall-term/.

U.S. Department of Agriculture, Food and Nutrition Service, "SNAP Application Processing Timeliness," 2026. Accessed: May 27, 2026. Page last updated March 12, 2026.

_ , "SNAP Eligibility," https://www.fns.usda.gov/snap/recipient/eligibility. Accessed: May 22, 2026.

U.S. Department of Education, National Center for Education Statistics, "Table 1.2-A: Among 2011–12 First-Time Postsecondary Students Who Began at a 2-Year Public College, Percentage Distribution of 6-Year Attainment and Retention Status at Any Institution Attended, by Selected Personal Characteristics: 2012–17," 2017. Data Lab Tables Library. https://nces.ed.gov/datalab/table/library/detail/14326.

_ , "Table 226.40: SAT Mean Scores, Standard Deviations, and Score Ranges for High School Seniors, and Percentage of the Graduating Class Taking the SAT, by State: 2017," 2017. https://nces.ed.gov/programs/digest/d17/tables/dt17_226.40.asp.

_ , "Integrated Postsecondary Education Data System," 2023. Accessed via the Urban Institute's Education Data Portal.

_ , "Racial/Ethnic Enrollment in Public Schools," https://nces.ed.gov/programs/coe/indicator/cge/racial-ethnic-enrollment 2024. Condition of Education. U.S. Department of Education, Institute of Education Sciences. Last updated May 2024.

U.S. Department of Transportation, Bureau of Transportation Statistics, "Transportation Economic Trends," https://data.bts.gov/stories/s/Transportation-Cost-Burden-Transportation-Spending/v67s-yiqd/. Accessed: May 22, 2026.

U.S. Energy Information Administration, "Consumers Spent More on Gasoline in 2011 Than in Any Other Year," https://www.eia.gov/todayinenergy/detail.php?id=4890 December 2011. Accessed: March 6, 2026.

_ , "2009 Residential Energy Consumption Survey (RECS) — Housing Characteristics, Consumption, and Expenditures," https://www.eia.gov/consumption/residential/data/2009/index.php?view=characteristics 2013. Final release date: May 6, 2013.

Wiswall, Matthew, "The dynamics of teacher quality," *Journal of Public Economics*, April 2013, *100*, 61–78.

Zimmerman, Seth D., "The Returns to College Admission for Academically Marginal Students," *Journal of Labor Economics*, 2014, *32* (4), 711–754.

Zivin, Joshua Graff and Matthew Neidell, "Temperature and the Allocation of Time: Implications for Climate Change," *Journal of Labor Economics*, January 2014, *32* (1), 1–26. Publisher: University of Chicago Press.