UNIVERSITY OF CALIFORNIA SAN DIEGO

Essays in Applied Microeconomics

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Economics

by

Zachary A. Goodman

Committee in charge:

Professor Gordon Dahl, Chair
Professor Jeffrey Clemens
Professor Melissa Famulari
Professor Craig McIntosh
Professor Katherine Meckel

2021

The dissertation of Zachary A. Goodman is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2021

DEDICATION

To my parents, for believing in me and encouraging me to pursue my dreams.

EPIGRAPH

Give a man a fish and feed him for a day. Teach a man to fish and feed him for a lifetime. Teach a man to cycle and he will realize fishing is stupid and boring.

—Desmond Tutu

TABLE OF CONTENTS

Dissertation Approval Page . . . iii

Dedication . . . iv

Epigraph . . . v

Table of Contents . . . vi

List of Figures . . . viii

List of Tables . . . ix

Acknowledgements . . . x

Vita . . . xiii

Abstract of the Dissertation . . . xiv

Chapter 1 The effect of a supplementary video book on learning in intermediate microeconomics . . . 1
- 1.1 Introduction . . . 1
- 1.2 Related Literature . . . 5
- 1.3 Study Design . . . 9
  - 1.3.1 Description of the sample and institution . . . 9
  - 1.3.2 Description of the IMVH . . . 11
  - 1.3.3 Experiment Design . . . 13
  - 1.3.4 Empirical strategies . . . 16
- 1.4 Results . . . 22
  - 1.4.1 Attrition and balance . . . 22
  - 1.4.2 Relevancy of the encouragement instrument . . . 25
  - 1.4.3 Effects on exam scores . . . 26
  - 1.4.4 Spillovers during concurrent term . . . 28
  - 1.4.5 Spillovers to subsequent term . . . 29
  - 1.4.6 Treatment Effect Heterogeneity . . . 30
- 1.5 Models of Studying Behavior . . . 31
- 1.6 Discussion . . . 37
  - 1.6.1 Contributions . . . 37
  - 1.6.2 Experiment design choices and *local* treatment effects . . . 39
  - 1.6.3 Limitations . . . 42
- 1.7 Conclusion . . . 43
- 1.8 Acknowledgements . . . 44
- 1.9 Appendix . . . 60
  - 1.9.1 Additional experiment details . . . 60
  - 1.9.2 LATE estimators using Neyman's repeated sampling approach . 62

Chapter 2 The Effect of the Cook County, Illinois Sweetened Beverage Tax on Sugar Demand . . . 78
2.1 Introduction . . . 79
2.2 Background and Related Literature . . . 82
2.2.1 Beverage taxes in the United States . . . 82
2.2.2 Cook County's beverage tax . . . 84
2.2.3 Theoretical rational for beverage taxes . . . 86
2.3 Data . . . 88
2.4 Methods . . . 90
2.4.1 Effect on food and nutrient purchases . . . 90
2.4.2 Welfare . . . 93
2.5 Results . . . 96
2.5.1 Effect on nutrients . . . 96
2.5.2 Effect on beverage purchases . . . 97
2.5.3 Effect on sugar from select sources . . . 98
2.5.4 Effect on shopping and scanning behaviors . . . 98
2.5.5 Treatment effect heterogeneity . . . 99
2.5.6 Consumer price indices and welfare . . . 101
2.6 Discussion and Conclusion . . . 102
2.6.1 Limitations . . . 104
2.7 Acknowledgements . . . 105
2.8 Appendix . . . 124

Chapter 3 The Effect of the 2008 Economic Stimulus Payments on Nutrient Demand . . . 139
3.1 Introduction . . . 140
3.2 Background and Related Literature . . . 141
3.2.1 Economic Stimulus Act of 2008 . . . 141
3.2.2 Unconditional cash transfers in the United States . . . 142
3.2.3 Demand for nutrients . . . 143
3.3 Data . . . 144
3.4 Methods . . . 148
3.4.1 Effect on payment on nutrient demand . . . 148
3.4.2 Heterogeneity . . . 149
3.5 Results . . . 150
3.5.1 Effect on panelist behaviors . . . 150
3.5.2 Effect on nutrient demand . . . 151
3.6 Discussion and Conclusion . . . 151
3.6.1 Limitations . . . 152
3.7 Acknowledgements . . . 153
3.8 Appendix . . . 166

LIST OF FIGURES

Figure 1.1: Attrition by treatment arm 25
Figure 1.2: Utility maximization problem for model students given the incentive structure in the experiment 34
Figure 1.3: Effect of grade incentive on videos watched 56
Figure 1.4: Weekly video watching by treatment arm 57
Figure 1.5: Video watch rates by video and treatment arm, grouped by incentive 58
Figure 1.6: Video watch rates by video and treatment arm, grouped by week 59
Figure 1.7: Distribution of videos counted towards incentive 72
Figure 1.8: Weekly video watching by exam topic 73
Figure 1.9: Effects of treatment along first midterm score, by midterm 2 74
Figure 1.10: Effects of treatment along first midterm score, by final 75
Figure 1.11: Distribution of max videos watched in one week 76
Figure 1.12: Video watch rates by video and treatment arm, grouped by incentive, ordered by video duration 77

Figure 2.1: Rollout of beverage taxes in the United States 82
Figure 2.2: Effects of Cook County tax on nutrients purchased 120
Figure 2.3: Cook County zip code layers 121
Figure 2.4: Soda budget share and price 122
Figure 2.5: Price indices 123
Figure 2.6: Cook County DMA: nutrient composition over time 129
Figure 2.7: Cook County DMA: panelist behaviors over time 130
Figure 2.8: Cook County DMA: beverage purchases over time 131
Figure 2.9: Cook County DMA: sugar by source over time 132
Figure 2.10: Cook County DMA: variation by store location 133
Figure 2.11: Cook County DMA: nutrient composition over time, ten largest counties as counterfactual 134
Figure 2.12: Cook County DMA: nutrient composition over time, ten nearest population density counties as counterfactual 135
Figure 2.13: Example price tag from Cook County grocery store 136
Figure 2.14: Distribution of differences in price per ounce during pretax versus taxed periods for carbonated beverages 137
Figure 2.15: Relative weekly search volume for “soda tax” 138

Figure 3.1: Dates ESP reported received 162
Figure 3.2: Effects of ESP receipt on shopping behaviors 163
Figure 3.3: Effects of ESP receipt on nutrient purchases 164
Figure 3.4: Difference in effects of ESP receipt for low liquidity households 165
Figure 3.5: Nutrient purchases over time by liquidity constraints 167
Figure 3.6: Nutrient purchases over time by liquidity constraints 168

LIST OF TABLES

Table 1.1: Comparison of information transmission formats 12
Table 1.2: Grade scheme by treatment arm. *Control* represents same grade scheme as *Above median*. Differences between the two grade schemes in bold. 13
Table 1.3: Predictions Across Models for Compliers, those induced by incentive to consume more videos. 36
Table 1.4: Effects of Grade Incentive on Video Watching 50
Table 1.5: Effects of Videos on Grades 51
Table 1.6: Spillover Effects of Incentive on Other Course Grades 52
Table 1.7: Spillover Effects of Incentive on Other Studying 53
Table 1.8: Spillover Effects during Subsequent Quarter 54
Table 1.9: Heterogeneous Effects of Treatment 55
Table 1.10: Baseline balance test, Midterm 2 sample 64
Table 1.11: Baseline balance test, Final Exam sample 65
Table 1.12: Anderson-Rubin confidence sets 66
Table 1.13: Candidate control variables for post-double-selection 67
Table 1.14: ITT model controls selected via post-double-selection 68
Table 1.15: LATE model controls selected via post-double-selection 71

Table 2.1: Summary statistics for Chicago DMA sample: demographics 110
Table 2.2: Effects of Cook County Beverage Tax on nutrients 111
Table 2.3: Effects of Cook County Beverage Tax on select beverage volume 112
Table 2.4: Effects of Cook County Beverage Tax on probability of purchasing select beverages 113
Table 2.5: Effects of Cook County Beverage Tax on sugar from select sources 114
Table 2.6: Effects of Cook County Beverage Tax on panelist behaviors 115
Table 2.7: Heterogeneous treatment effects of tax by household characteristics 116
Table 2.8: Effects of Cook County Beverage Tax by zip code layer 117
Table 2.9: Effects of Cook County Beverage Tax on cross-border All Trips 118
Table 2.10: Effects of tax on sugar by pretreatment soda decile, Cook County 119
Table 2.11: Summary statistics by soda decile, Cook County 124
Table 2.12: Effects of tax on sugar by pretreatment soda quintile, Philadelphia 125
Table 2.13: Effects of tax on sugar by pretreatment soda quintile, Seattle and San Francisco 126
Table 2.14: Effects of Cook County Beverage Tax on nutrients, 10 largest counties by population as counterfactual 127
Table 2.15: Effects of Cook County Beverage Tax on nutrients, 10 nearest population density counties as counterfactual 128

Table 3.1: Summary statistics, categorical variables 156
Table 3.2: Summary statistics, numeric variables 157
Table 3.3: Effects of ESP receipt on behaviors 158
Table 3.4: Effects of ESP receipt on nutrients 159
Table 3.5: Effects of ESP receipt on behaviors, low liquidity households 160
Table 3.6: Effects of ESP receipt on nutrients, low liquidity households 161

ACKNOWLEDGEMENTS

I would like to thank the Chair of my committee, Professor Gordon Dahl, for his encouragement and guidance throughout my graduate education. Besides providing feedback that has vastly improved my research, he has instilled in me an appreciation for rigorous causal inference and the tools applied economists use to better understand how the world works. I am grateful for Gordon's support of my academic, professional, and personal goals and his desire to see his students achieve what they want to achieve.

I would also like to recognize Professors Jeffrey Clemens, Craig McIntosh, and Katherine Meckel, who first taught me as their student before serving on my committee. I thank them for the many hours they have spent answering my questions and providing insights that have undoubtedly made me a stronger researcher and more learned economist. I want to additionally thank my collaborator, mentor, and committee member Professor Melissa Famulari. As her Teaching Assistant for three consecutive years, I learned most of my teaching skills shadowing one of the best.

I too would like to recognize those at UCSD not on my committee who have impacted my development as a researcher and person. I express my gratitude to the Teaching + Learning Commons for making possible the analysis in Chapter 1. I would like to thank Erilynn Heinrichsen for championing improved learning outcomes at UCSD and more than doubling the tools in my teaching toolbox. I would like to thank Molly Roberts for her guidance as I developed and instructed my own course for the first time. Others at UCSD I would like to thank include Kate Antonovics, Eli Berman, Julian Betts, Prashant Bharadwaj, Julie Cullen, Itzik Fadlon, Mark Jacobsen, Susan King, Karthik Muralidharan, Giacomo Rondina, and Jessica Williams.

I would be remiss if I did not acknowledge those at NC State who have tremendously influenced my career. I thank Professors Umut Dur and Doug Pearce for their mentorship and exposing me to economic research, Professor Walter Wessels for uncovering my love of microeconomics, and Professor David Hyman for convincing me to pursue economics beyond the general education requirement that was his class. Thanks also to the Park Scholarships staff and scholars for their support and fostering an undergraduate experience that produces not just good thinkers, but good people. I would like to recognize Larry “Coach” Brown who, beyond teaching me to dive, taught me to do everything “slow, relaxed, and with good technique.” I will surely “smile and be happy” for years to come thanks to his wisdom.

I would like to thank my peers in the UC San Diego Economics graduate program, both past and present. Never have I been surrounded by such an intelligent, supportive bunch so eager to create and share knowledge. Thank you Adam, Alex, Alyssa, Arman, Arushi, Bruno, Camila, Daniel, Evgenii, Greg, Jackson, Jason, Jianan, Jonathan, Mitch, Pablo, Pietro, Sasha, Yu-Chang, and others I have missed for your help and friendship. I especially want to thank Jacob Orchard for serving as a model co-author and friend.

I want to thank my friends outside economics for keeping me sane, grounded, and the best version of myself. I would like to thank my triathlon and cycling training partners and coaches for helping me maintain a fit mind and body. Thanks to Amy, Kyle, and the other members of the paratriathlon community for helping me rethink my own limits. I also want to thank my east coast friends for not abandoning me after I moved 3,000 miles away.

Lastly, I want to thank my family, to whom I owe so much for their love and unwavering belief in me. Thank you, my dear Beril, for serving as my guiding light in a chaotic world, for pushing me to be my best, and for being my partner through the good times and the bad. Thank you, my brother Justin, for the memes. Finally, thank you, Mom and Dad, for my existence, but moreover an existence worth living.

Chapter 1 is currently being prepared for submission for publication of the material. Famulari, Melissa; Goodman, Zachary A. The dissertation/thesis author was a primary investigator and author of this material.

Chapter 2 is currently being prepared for submission for publication of the material. Goodman, Zachary A.; Orchard, Jacob. The dissertation/thesis author was the primary investigator and author of this material.

Chapter 3 is currently being prepared for submission for publication of the material. Goodman, Zachary A. The dissertation/thesis author was the primary investigator and author of this material.

VITA

2016	Bachelor of Science in Economics, North Carolina State University
2016	Bachelor of Science in Mechanical Engineering, North Carolina State University
2018	Master of Science in Economics, University of California San Diego
2019	Candidate of Philosophy in Economics, University of California San Diego
2021	Doctor of Philosophy in Economics, University of California San Diego

ABSTRACT OF THE DISSERTATION

Essays in Applied Microeconomics

by

Zachary A. Goodman

Doctor of Philosophy in Economics

University of California San Diego, 2021

Professor Gordon Dahl, Chair

This dissertation contains three essays on topics in applied microeconomics. The first essay addresses effective pedagogical tools, and the latter two essays estimate the effects of two distinct tax policies on nutrient consumption.

In Chapter 1, we study a novel video-based textbook for intermediate microeconomics. Using a field experiment involving about 400 undergraduates, we estimate the effectiveness of watching videos on exam scores. We find that students experimentally induced to watch more videos perform significantly better on the midterm and final exams. We find no negative spillovers to other courses within the quarter of the experiment, and we find sustained takeup of the videos in the following quarter.

In Chapter 2, we study the 1-cent-per-ounce sweetened beverage tax in Cook County, the largest tax (in terms of population affected) of its kind in the United States and the only tax revoked to date. We find that the tax significantly decreases sugar purchases while active and has no lasting effects after the tax is revoked. We find that the tax has the largest sugar-reducing effects for high consumers of regular soda and those who live far from the border of the taxed jurisdiction. We weigh the welfare consequences of the tax by estimating the cost of living increase and find that each gram of sugar reduced cost 3.5 to 6.6 cents.

In Chapter 3, I examine the effects of the 2008 Economic Stimulus Act payments on nutrient purchases. I find that households with less than two months of income in savings increase total calories purchased in the month following receipt of the stimulus payment. Interestingly, the composition of the increased calories is not representative of the pre-stimulus nutrient bundle. Households increase carbohydrate and sugar purchases more than they increase fiber or protein purchases. I do not find evidence of sustained changes in nutrient purchases or changes in purchase behaviors by households with access to liquid savings.

Chapter 1

The effect of a supplementary video book on learning in intermediate microeconomics

You expect me to read the textbook? Ha!

— Anonymous student

1.1 Introduction

Every year, university students spend tens of thousands of dollars on tuition and course materials and hundreds of hours studying, in large part, to learn. Instructors can help their students learn more efficiently by providing and recommending pedagogical tools that have high returns per unit time and financial cost. Despite the value to students and instructors, little empirical work exists that estimates the effectiveness of different learning technologies (Allgood, Walstad, & Siegfried, 2015).

In this paper, we measure the impacts of one such technology, the Intermediate Microeconomic Video Handbook (IMVH), on outcomes in an intermediate microeconomics course. The IMVH was designed to *supplement* lecture as an audiovisual version of a conventional course textbook. Part of the impetus for creating the IMVH was a discussion with a student who described her inability to read the course text, not because of poor reading skills, but because she did not find the text engaging enough to command her attention. We hypothesized that modern students, who have had unprecedented exposure to electronic media, may find videos more engaging and, perhaps, more effective at building human capital.

The IMVH is comprised of 220 short videos, organized into a book, and cover the topics for a year-long interemediate microeconomics sequence with each topic covered by two videos: one with the verbal and graphical intuition and one with more formal definitions, identified by (Calculus) in the title. The IMVH was created by six UC San Diego faculty members, including one of the authors. Besides higher engagement than with conventional texts, the IMVH and video-based learning tools more broadly are of value to university educators and their students for four additional reasons. First, videos carry near-zero marginal cost and are accessible anywhere via internet, helping reduce financial and geographic barriers to high-quality educational materials. Second, video platforms can help students track what content they have already studied and what content they have yet to cover. Third, features embedded into videos, such as searchable captions and timestamps, can help students locate the information they are seeking faster. Finally, the perceived low cost of watching a brief video may be easier to overcome than perceived higher costs of other studying methods, potentially leading to more frequent studying, which decades of psychological research has demonstrated leads to more long-term learning than does cramming (Cepeda et al., 2006; Kornell, 2009).

Ultimately, the beneficial features of video-based learning tools are of value only if they can improve student learning outcomes, an empirical question we seek to answer in this paper. To estimate the effects of the IMVH, we administered a field experiment involving nearly 400 undergraduates enrolled in the same one-quarter-long microeconomics course over two years. Of note, these students all scored below the median on the first midterm exam, thereby making manifest a need to adjust studying habits, and perhaps standing to gain the most from an intervention that targets studying. We randomly assigned a grade-based incentive to half of these lower-performing students to encourage take-up of the IMVH, which was made available to all students in the class, allowing identification of intent to treat (ITT) effects and local average treatment effects (LATEs) while maintaining equitable access to learning resources. We tracked video watching at the student level using the software platform that hosts the IMVH. We observe grades, GPA, and video watching in both the term of the experiment and the subsequent term when students typically take the second intermediate microeconomics class in the sequence.

The first-stage impact of the exogenous encouragement on video watching is significant and substantial. Students who receive the grade-based incentive watch over 28% more unique videos by the second midterm and 63% more unique videos by the final exam, or about 1.1 and 3.4 hours of content, respectively, than did their control peers. We find large reduced-form effects of treatment on exam scores: for students in the bottom half of the class as of the first midterm, being assigned treatment (ITT) increases midterm and final exam scores by about 0.18 standard deviations. Our estimates imply that the marginal hour of videos watched increases exam scores (LATE) by between 0.05 and 0.16 standard deviations.

We interpret our results through a theoretical framework in which students, who value grades and leisure, have potentially incorrect priors about the returns of different studying methods. We do so not to test theory, but rather to help educators understand the potential welfare implications of providing students with a learning technology and paternalistic incentive structure like the one studied. Although we observe that treatment students performed better on course assessments, for welfare analysis one must consider where the time watching videos came from: leisure, work, student organizations, studying for other classes, studying for present class using other methods, etc. If students must reduce time allocated towards leisure or studying for other classes so they can watch more videos, then the welfare impacts of incentivizing video use could be negative depending on the students' preferences. On the other hand, if the videos are more productive than students realize compared to the next best studying technology, then incentivizing video watching could be utility enhancing.

To better understand the spillover effects of treatment, we examine other forms of studying including class attendance, visits to a tutoring center (specific to this course), and interacting with the class discussion board. We do not find any statistically significant changes in any observed studying method, and we can rule out large changes. In nearly all cases, treatment students used other studying methods at directionally *greater* rates than did their control peers. We also investigate spillovers to other courses taken during the term of the experiment and similarly find that treatment students perform directionally *better* than their control peers. Though not statistically significant, we can rule out large negative effects, suggesting that treatment did not cause students to substitute away from studying for other courses.

An important piece to the welfare puzzle is whether treatment students continue to use the IMVH at higher rates after exogenous incentives are removed. Persistent take-up in the absence of external prodding provides some confidence that students, now with updated priors, value the technology. Fortunately, we can observe video watching in the subsequent microeconomics course in the term following the experiment. Despite there being no direct incentives to watch videos in the subsequent course, treatment students persistently watched more videos than did control students, about 8 - 10 more unique videos, or 1.2 - 1.5 more hours of unique content. Our sample in the subsequent term is nearly half the original size, so we lack power to precisely estimate effects on exam scores; however, our confidence intervals include effect sizes consistent with those observed in the experiment term.

Collectively, we interpret our findings as evidence that requiring the IMVH is a net positive on underperforming students' academic achievement, both in the quarter of the experiment and beyond. Though formal welfare analysis is beyond the scope of this paper, we present suggestive evidence that requiring the IMVH is unlikely to be substantially utility harming, if not utility enhancing, as our results are consistent with a poor-information model of student learning. Our findings justify paternalistic incentive structures in settings where a large portion of the class is at risk of failing and the instructor has more information about the usefulness of a novel teaching technology than do her students.

The rest of the paper is organized as follows. Section 1.2 provides background on existing related literature. Section 1.3 describes the study design. Section 1.4 presents the results of the experiment, and Section 1.5 presents competing models of studying behavior that may explain the observed phenomena. Section 1.6 discusses the contributions and limitations of our study, and Section 1.7 concludes.

1.2 Related Literature

Students have many time-consuming activities to help them learn including attending class, watching recorded lectures, reading the textbook, doing homework, completing practice exams, attending tutoring labs, and more. While there is tremendous value in understanding how effectively each activity contributes to student learning, several empirical challenges make it difficult to estimate the causal effects of such activities. First, a student's decision to use a study method is likely influenced by unobservable student characteristics, such as motivation or ability, that also likely affect student exam performance. To estimate causal effects, researchers must address nonrandom selection into using the study method. Second, motivated students often visit instructors seeking to improve their study strategies after a negative exam shock, which suggests "dynamic selection" into the use of a study method and has been found empirically by Oettinger (2002), Krohn and O'Connor (2005), Stinebrickner and Stinebrickner (2008), Bonesrønning and Opstad (2012) and Bonesrønning and Opstad (2015).[1] Dynamic selection means including student fixed effects in class performance regressions will not uncover the causal effect of a study method. Third, study methods may be substitutes or complements in student learning, and experimental inducements to use one study strategy may affect takeup of another. In these cases, even randomized experiments will not identify the causal effects of a particular study method but will instead identify the causal

[1] Oettinger (2002) finds that students close to a grade threshold before the final exam perform better on the final. Krohn and O'Connor (2005) show that students reduce the number of hours they study after getting higher midterm scores. Stinebrickner and Stinebrickner (2008) find that IV estimates of studying on grades are much larger than OLS and provide suggestive evidence that students increase effort in semesters when semester-specific elements of grades are low. In two papers, Bonesrønning and Opstad (2012) and Bonesrønning and Opstad (2015) find that the difference between a student's expected and actual grade on an early assessment is positively correlated with a change in their study hours (after-assessment study hours minus before-assessment study hours). This suggests that students with a negative exam shock (actual grade worse than expected) may respond by increasing their study hours.

effects of a study policy and all of the changes in student behavior caused by that policy.[2] Finally, experimental inducements to use a study method may change the total time devoted the course. In this case, experiments jointly test the effectiveness of a particular learning method and devoting more (or less) time to the course.

We focus our review on research that uses experiments or quasi-experiments to explore the effects of learning acquisition that take students' time.[3] We organize these studies into two broad groups: *guided study*, such as attending lecture, tutoring labs, and discussion sections/recitations, or *self study*, which includes doing homework, practice exams, and watching recorded lectures.

First we examine guided study. Kirby and McElroy (2003) use student-reported travel time to campus as an instrument for lecture attendance and find a positive causal relationship between lecture attendance and exam grades. Dobkin, Gil, and Marion (2010) analyze a policy where lecture attendance became compulsary for students who scored below the median on the midterm exam and remained optional for those who scored above the median. Using a regression discontinuity approach, they find that a 10 percentage point increase in student attendance led to a 0.17 standard deviation increase in final exam score. Joyce et al. (2015) randomly assign 725 students taking introductory microeconomics students to twice-per-week and once-per-week lecture formats to identify the effects of classroom time in classes that are well-supported with online content (videos, quizzes, lectures slides, etc.) Students in the twice-per-week format scored 0.21 standard deviations higher on the midterm and 0.14 standard deviations higher on the final exam. Finally, Tang et al. (2020) analyze an experiment in which students were randomly assigned to either weekly or bi-weekly grading of in-class clicker questions. Weekly grading increased student lecture attendance by 11 percent, had no effect on self-study hours, and raised course grades by 6.31

[2] While the causal effects of an educational policy are useful for educators considering how to design their classes, they are less useful for students wanting to know the most productive use of their study time. We should also point out that instructors may find learning transmission methods substitutable or complementary. As an example, Morris, Swinnerton, and Coop (2019) find that many instructors report that lecture capture, where lectures are recorded and made available to students, changed the way they lectured in the classroom. Experiments randomly assigning students to classes taught one way versus another will not identify the causal effect of a study method if other aspects of learning transmission are simultaneously changed.

[3] We do not include research that examines how to make studying methods more productive for a fixed quantity of time, though the intensive margin of pedagogical tools is certainly an important area of study with potentially clearer welfare implications.

percent. Notably, the effects of weekly grading were strongest for students who preferred bi-weekly grading, had lower prior GPAs, and had lower self-control scores.

Arulampalam, Naylor, and Smith (2012) study discussion section attendance. The authors use time of a student's randomly assigned section as an instrument for attendance. They find no effects of section attendance for most students, but for students in the top quantiles, missing 10 percent of sections caused a 1 percentage point performance loss. Kapoor, Oosterveen, and Webbink (2020) examine the effects of section for second-year students whose first year GPA fell below a threshold. They find students just below the threshold, who were required to attend at least 70 percent of sections, attended 50 percent more sections and lectures but reported no significant difference in total study hours (lectures, section, and self study). Interestingly, the authors report zero overall effect on grades and a substantial *negative* effect in classes where attendance was optional for students above the threshold. Bratti and Staffolani (2013) examine attendance for lecture and discussion sections combined. Using student residence as an instrument and controlling for self-study hours, the authors find no significant effect of attendance on grades for most courses but significantly positive effects for quantitative courses.

At many universities, students can visit tutoring labs for help learning the material covered in lecture or discussion sections. Munley, Garvey, and McConnell (2010) find athletes are significantly more likely to attend peer tutoring labs, which the authors attribute to frequent reminders from coaches. Using one's status as an athlete as an instrument for tutoring hours, the authors find significantly positive effects of peer tutoring: attending tutoring one hour per week over a 14-week semester increases a student's final grade by one letter. Collectively, the aforementioned studies on the effects of guided study, across all types, consistently find that student performance is improved, but only if students do not substitute self-study time for guided-study time.

Turning to the effectiveness of self study, Stinebrickner and Stinebrickner (2008) examine 210 Berea College students who were randomly assigned a roommate. Students whose roommate brought a video game to college earn lower grades and spend less time studying. They authors instrument for study time using presence of a roommate with a video game and find that an additional hour of studying per day increases GPA by 0.36 points. On the other hand, Oreopoulos et al. (2019) randomly assign first year economics students at three different institutions to a variety of low-touch interventions to increase study hours (planning studying schedules, information about the value of studying, and weekly reminders about study plans) and find no effect on grades, credit accumulation, or retention despite increasing student's self-reported study hours. Clark et al. (2020) also explore a low touch intervention, having students set goals on the number of practice exams they will complete, and find those randomly assigned to set task-based goals completed 0.102 standard deviations more practice exams and increased total course points by .068 standard deviations.

Trost and Salehi-Isfahani (2012) examine the causal effects of homework by randomly assigning whether or not homework would contribute to students' grades in a Principles of Economics class. The authors find significant positive effects of homework on the first midterm but no effects on the final exam. Grodner and Rupp (2013) conduct a similar experiment, randomly assigning students to a "homework required" group, for whom homework was worth 10% of the final grade, and "homework not required" group, for whom exams were correspondingly upweighted. The authors find that 90% of treatment students completed 7 or more homework assignments compared to 6% of control students and on average scored higher on exams. As a whole, the research on the causal effects of self study is mixed and ranges from no significant effect on exams to large positive effects on GPA. While there is some evidence that self study may matter for early assessments, more research is warranted.

Finally, we review the research on lecture capture and flipped classrooms, both of which have aspects similar to the IMVH. Lecture capture provides students with recordings of lectures, which students may use as a substitute for lecture, to rewatch for enhancing understanding, or as review before exams. By necessity, recorded lectures cannot be made available to students before lecture and hence cannot help students prepare for lecture. To the best of our knowledge, there is no experimental or quasi-experimental research estimating the causal effects of lecture capture. In flipped classrooms, students watch material in advance of class and work on problems during class where instructional staff are available to answer questions. Two RCTs involving students at U.S.

military academies shed some light on the effectiveness of flipped classrooms in university settings. In a small RCT (137 students in 7 sections of an introduction to economics course), Wozny, Balser, and Ives (2018) assign lecture type (flipped vs traditional) to randomly chosen topics and randomly chosen instructors. They find that the flipped classroom increases exam scores by 0.16 standard deviations for medium-term assessments but no effect on short-term exams or the comprehensive final. In a larger RCT (1,328 students in 80 course sections and 29 instructors) across two courses, Principles of Economics and Introductory Calculus, Setren et al. (2021) randomly assign half of sections to cover a particular course topic using the flipped model and half to cover the topic using a traditional model. The authors find significant positive effects on a low-stakes quiz but no effects on the final exam.

1.3 Study Design

1.3.1 Description of the sample and institution

We conducted the field experiment in an undergraduates intermediate microeconomics course taught during fall 2018 and fall 2019 by one of the authors. The university is a large, diverse and selective public research university in the United States.[4] At this institution, intermediate microeconomics is a three-quarter sequence required for students majoring in Economics. The experiment was conducted in the first course of the sequence, *Micro A*. We also observe grades and video watching in the second course of the sequence, *Micro B* during the winter 2019 and winter 2020 quarters. The same instructor taught Micro B in both years (and was a different instructor than the one who taught Micro A). Both Micro A and B instructors created half of the videos relevant to their course in the IMVH.

[4] The Carnegie Classification of Institutions of Higher Education classifies the university as an R1 (very high research activity) university. For the 2017-2018 academic year, the undergraduate student body shared the following demographics: 49.1% female, 50.6% male; 75.0% in-state, 5.5% out-of-state, and 19.5% international; 59% students of color; 28.6% majoring in the social sciences, 26% of which major in Economics. Among newly admitted students, about one-third were transfer students, and average SAT scores were 652 and 605 for math and critical reading, respectively. About 34% of students are the first in their family to attend a four-year university.

The structure is similar across the three courses in the Micro sequence. Students have the option to attend one of two lectures offered back to back twice per week, each lasting about 90 minutes. Lectures are not recorded. Two midterm and final exams are held at a common time outside of lecture. In addition to lecture, students have access to weekly one-hour discussion sections run by graduate teaching assistants (TAs) who are all Economics PhD candidates, including, at the time, one of the authors. In lieu of office hours, the graduate TAs and Undergraduate Instructional Assistants staff a tutoring lab open between three and four hours per day, six days per week. Students may also attend weekly Supplemental Instruction (SI) sessions offered by undergraduates majoring in Economics and trained by the university in SI. Besides the IMVH, students have access to a variety of online learning resources including a discussion board moderated by the instructional team, four years of previous exam questions, weekly ungraded problem sets, and semi-weekly graded online quizzes.

Students were told about the experiment during the first lecture, given a printed copy of the consent form in the second lecture, and provided a virtual copy of the consent form in the syllabus on the course webpage. At any time during the quarter, students could opt out of having their data included in the analysis.[5] Students below the age of 18 at the start of the course as well as students enrolled via the university's extension program were removed from the analysis dataset.[6] Ultimately, four students under 18, five extension students, and seven students who opted-out were removed from the analysis dataset, leaving a sample of 850 students.

There are two unique demographic features of the class worth noting. First, many non-econ majors take the class to either satisfy general education requirements or to explore majoring in economics. As there are many students in the experiment on the margin of majoring in economics, an important outcome is the likelihood the student takes Micro B. Second, about 37% of the class is transfer students, for whom the class is not only their first experience with upper division

[5] Students could opt out via an online form visible to a third party university organization so that neither the instructor nor research team could observe which students elected to opt out.

[6] Students under the age of 18 were excluded per IRB protocol. We exclude extension students because of their potentially very different preparation for the course and our inability to observe pretreatment covariates and outcomes outside of Micro A.

coursework at a four-year research university, but also typically their first time taking classes under the faster-paced quarter system.[7] We examine treatment effect heterogeneity to understand whether transfer students might differentially benefit from the IMVH.

1.3.2 Description of the IMVH

The Intermediate Microeconomics Video Handbook (IMVH) is a collection of 220 short videos that cover the material in a year-long intermediate microeconomics course sequence.[8] The videos were designed as a complement to both lectures and the course textbook. Each topic typically has two videos, one with the graphical and verbal intuition and the other with the formal algebraic definitions and proofs, identified by having (Calculus) at the end of the video title.

The videos were created by six UC San Diego faculty members with professional videographer and production support. Many videos utilize the "learning glass," an innovative presentation technology where instructors write with neon markers on a large sheet of glass that has lights embedded along the glass edge to make the colors pop. The remaining videos feature faculty superimposed in front of slides that are often written on during the presentation. Videos are closed captioned and were checked by graduate students for accuracy.

Given the complexity of the material, the IMVH was developed with several key objectives in mind. First, the web interface is clean and simple to not distract from the content. Second, to help students find material quickly, videos can be accessed via either the table of contents or the index, each video contains time stamps of the concepts therein and captions are searchable, which allow the student to jump to the part of a video containing the searched-for word. Finally, the videos are organized by content area (e.g., consumer theory, producer theory, etc.) that help students understand where various topics "live" in intermediate microeconomics.

While we do not know of another textbook completely comprised of videos, the IMVH is similar to the Khan Academy website, lecture podcasts, and textbook websites that incorporate

[7] Community colleges, the most common previous institution for transfer students, are on the semester system in the state of the university.

[8] A preview of the IMVH can be found at https://iti.ucsd.edu/IMVH_Misc/Promo/IMVHPromo.html.

Table 1.1: Comparison of information transmission formats

Feature	Lecture	eText-book	Lecture Capture	IMVH
Instructor's time used	✓			
Instructor-learner interaction	✓			
Learner-learner interaction	✓			
Readable		✓	?	✓
Scalable	?	✓	✓	✓
Searchable		✓		✓
Skimmable		✓		✓
Stoppable	?	✓	✓	✓
Watchable	✓		✓	✓
Consumed on Demand		✓	Only After Lecture	✓

instructional videos. Table 1.1 presents a classification of some options to present course material to students.[9] Besides the engaging viewable nature, the IMVH differs from a traditional textbook in that the instructors explain, graph, and derive mathematical results in much the same way one would in a conventional lecture. The IMVH differs from a lecture in that students control the pace: they can rewatch, speed up, or slow down the videos. Other differences of the IMVH from lecture include the ability to read captions, clarity and ease of visibility (unlike in a large lecture hall), option to watch *before* lecture to prepare, and no recurring demand on the instructor's time. The primary benefit of lecture over the IMVH is that students can receive immediate help as soon as they have questions. Further, student questions may have import externalities for the learning of other students in the class. There is also an important social aspect of lectures as students can interact with each other before, during, and after lecture. Students cannot ask questions or interact with other students during an IMVH video. The IMVH differs from recorded lectures because the IMVH videos are much shorter, averaging under ten minutes. The videos are typically much more focused on one topic than are lecture recordings. Finally, lecture recordings typically include components that do not work well when recorded, such as group work or class discussion.

[9] This table is a slight modification of the classification table Martin Osborne proposed to one of the authors in an e-mail correspondence.

Table 1.2: Grade scheme by treatment arm. *Control* represents same grade scheme as *Above median*. Differences between the two grade schemes in bold.

Assessment	Incentive	Control
>40 videos	**4%**	**0%**
Midterm 1	**18%**	**22%**
Midterm 2	22%	22%
Final Exam	50%	50%
Math Quiz	1%	1%
Best 5 of 6 Quizzes	5%	5%
Total	100%	100%

1.3.3 Experiment Design

The experiment began four weeks into the ten week term following grading the first midterm exam. All students who scored above the median on the first midterm, the *Above median* arm, and half of students who scored below the median, the *Control* arm, were assigned a conventional grading scheme that places weight only on exams and quizzes. We assigned the remaining half of students below the median to the *Incentive* arm, whose grading scheme allots four percentage points conditional on watching at least 40 of 48 eligible videos in the IMVH.[10] These 48 videos cover new class content since the first midterm that would be assessed in the second midterm and final exam. All students could still view the 26 videos relevant to the first midterm and, as they could help students on the cumulative final exam, we include them in our measures of video watching despite not counting towards the grade incentive.

The two different grading schemes are outlined in Table 1.2. Notably, the four percentage points come at the expense of reduced weight placed on the first midterm score, which had already occurred at the time of treatment assignment. Hence, at the time of treatment assignment, the video incentive is the sole forward-looking difference between treatment arms.

To improve balance between *Incentive* and *Control* arms and increase statistical power, we assigned students to treatment arms using paired randomization (Athey & Imbens, 2017), matching

[10] Watched in standard speed, 40 videos would require students to spend between 5.5 and 7.1 hours, depending on the length of videos chosen (on average 9.7 minutes in length each). Watching all 48 incentivized videos in standard speed would require just shy of eight hours.

students by their first midterm scores before randomly assigning one member of each pair to *Incentive* and the other to *Control* (further details on treatment assignment can be found in Appendix 1.9.1). We emailed each student letting them know their assignment and grading scheme. Students could also find their assignment listed in the online gradebook. To confirm that students knew their assignment, we surveyed students using an in-class attendance quiz, and 94% of students correctly identified their grading scheme. We emailed the students who responded incorrectly to clarify their assignments.[11]

We informed *Incentive* students that they must watch the entire video and only one video at a time to get credit towards their 40 required videos. It is not possible to observe "watching" as students could, for example, minimize their browser, walk away from their computer, or otherwise play a video without actively watching it. As a proxy for watching, we use data recorded by the IMVH software that captures the video ID, student ID, and the date and time when a student opens a video link. We define the following measures:

1. *Videos*: Number of links opened, including duplicates
2. *Unique videos*: Number of unique video links opened
3. *Hours of videos*: Total runtime of video links opened
4. *Hours of unique videos*: Total runtime of video links opened with duplicates removed

For expositional ease, we use "watching" to refer to the link-opening behavior as defined above. Although video watching in our data is a binary measure, watching behavior can vary greatly in intensity. Some students take notes, pausing and rewatching portions of the video as needed. Other students, we suspect, play videos in the background without absorbing much material. Exploring the intensive margin of video watching remains an area for future research that will benefit from new technologies that can quantify video engagement including interactive content embedded in videos, eye-tracking devices, and more.

[11] 11 of 164 *Incentive*, 23 of 167 *Control*, and 10 of 373 *Above median* students did not identify their grading schemes correctly. 146 students did not answer the quiz, several of whom had dropped the course following the first midterm.

We helped students keep track of their progress towards 40 videos by periodically updating the online gradebook with counts determined from the IMVH data. Although nearly all students followed our instructions to watch videos completely and sequentially,[12] in a few exceptional cases, students opened 40 or more video links within a matter of a few minutes. We manually adjusted their video counts in the gradebook and emailed them a reminder of the requirements for videos to count towards the grade incentive.[13] Though it is doubtful these strategic students gained much from opening so many videos so quickly, to maintain interpretability of our results, we do *not* remove these clicks from our video count measures.[14] Our *unique* video measures, however, are less sensitive to this behavior.

To ensure fairness, we informed students that final letter grades would *not* be affected by being in the experiment. We accomplished parity between *Control* and *Incentive* arms through curving final grades. First, we applied a curve to the *Control* and *Above median* arms as one group to achieve a grade distribution in line with that of previous cohorts. Second, we curved the *Incentive* students' course grades to match the average course grade among *Control* students after curving in the previous step. Since course grades are by construction *ex post* equal between the *Control* and *Incentive* arms, we use exam scores as our primary outcomes of interest. Our secondary outcomes of interest include term GPA and number of courses passed, which help us understand how treatment may have affected other courses. We examine separately economics and non-economics courses in case effects differ by course content. To better understand mechanisms, we estimate effects of treatment on take-up of other studying tools within Micro A. Finally, we examine the likelihood a student continues to Micro B, the subsequent course in the intermediate microeconomics sequence, and explore video watching and grade outcomes in Micro B to see if treatment effects persist beyond one term.

[12] The time between timestamps for video links opened in succession was almost always longer than the runtime of the video.

[13] Although additional email communication as a result of treatment could violate the exclusion restriction, the small number of affected students is unlikely to have much (if any) influence on our results.

[14] Our causal effects are per "links opened" rather than per "links opened subject to certain qualifying conditions". The cost of this interpretability is likely downward bias on our causal effect estimates.

1.3.4 Empirical strategies

In this paper, we estimate the effect of being assigned to the *Incentive* arm on our outcome variables of interest, Intent To Treat (ITT) effects, as well as the effect of watching videos for those induced by the incentive to watch more videos, a Local Average Treatment Effect (LATE), also referred to as a Complier Average Treatment Effect (CATE) (Imbens & Rubin, 2015). We explore whether there is treatment effect heterogeneity across key demographic variables. Finally, we estimate treatment effects at the median midterm 1 score (the cutoff score for inclusion in the experiment) using a regression discontinuity approach. Below we outline the empirical strategies for estimating each of these effects.

Intention To Treat (ITT)

In this section, we examine the empirical strategy for estimating the causal effect of being assigned to the *Incentive* arm on outcomes of interest, such as exam scores. These are ITT estimates and not average treatment effect estimates because of two-sided non-compliance: some students in the treatment arm do not watch videos and some students in the control arm do watch videos. Since the incentive itself in our setting is representative of how future instructors may induce their students to watch videos, the ITT estimates are policy-relevant for instructors considering a similar grade incentive to encourage video-delivered learning methods in their courses.

Our baseline ITT specification is the partially linear model:

$$Y_i = \beta Z_i + f(X_i) + \varepsilon_i \tag{1.1}$$

where Y_i is an outcome of interest (e.g. videos watched or test scores) for student i, $Z_i \in \{0,1\}$ is a treatment indicator with those in the *Control* arm having $Z_i = 0$ and those in the *Incentive* arm having $Z_i = 1$, $f()$ is a generic function through which X_i, a vector of controls, affects Y_i, and ε_i is an unobserved residual. β, our parameter of interest, is the causal effect of being assigned to the

Incentive arm on the outcome of interest Y, assumed to be constant across class cohorts.[15] Under unconfoundedness, $\widehat{\beta}$ is an unbiased estimate of the ITT effect (Imbens & Rubin, 2015).[16]

In our baseline estimation of Equation 1.1, we include in X_i a year indicator and first midterm score, following the advice of Bruhn and McKenzie (2009) to control for all covariates used in seeking balance. In a second model, we include additional controls chosen using the Post-Double-Selection (PDS) procedure of Belloni, Chernozhukov, and Hansen (2014b), explained in detail in Appendix 1.9.1. In a third model, to check that our results are robust to potentially nonrandom attrition by treatment arm, we fit Equation 1.1 including pair fixed effects. These fixed effects subsume the year indicator (since pairs were assigned separately across years), so we drop the year indicator but keep midterm 1 score to control for small differences within pairs along that dimension. As entity fixed effects require at least two observations within the entity to be estimable, we drop any students whose matched pair attrited.

In our results, we present an additional nonparametric estimate using Neyman's (1923) repeated sampling approach, considering each pair (block) an independent, completely randomized experiment and averaging the results. We estimate the point estimate of the ITT as the mean difference in outcomes across pairs:

$$\widehat{\tau} = \frac{1}{J}\sum_{j=1}^{J}\widehat{\tau_j} = \frac{1}{J}\sum_{j=1}^{J} y_{j,I}^{\text{obs}} - y_{j,C}^{\text{obs}} \tag{1.2}$$

where $\widehat{\tau}$ is the point estimate of the ITT, J is the number of pairs in the sample, and $\widehat{\tau_j} = y_{j,I}^{obs} - y_{j,C}^{obs}$ is the observed difference in outcome for pair j. The estimated standard error of $\widehat{\tau}$

[15] Our experiment takes place over two years, and we pool the sample across both years. Out of the 850 student-years, one student repeated the course in both years, and hence there are 849 unique students. For simplicity, we drop the subscript t from our specifications, treating the one repeating student as independent across years. Dropping this student from the sample leaves the results virtually unchanged.

[16] Though we cannot test whether Z_i is confounded by unobservable covariates, we have confidence unconfoundedness holds given the random assignment of Z_i and the balance across observable covariates as demonstrated in Table 1.10 and 1.11.

(Athey & Imbens, 2017; Imai, 2008; Imbens & Rubin, 2015) is:

$$\widehat{SE}(\widehat{\tau}) = \Big(\frac{1}{J}\sum_{j=1}^{J}\widehat{V}(\widehat{\tau}_j)\Big)^{\frac{1}{2}} \tag{1.3}$$

where $\widehat{V}(\widehat{\tau}_j)$ is the estimated variance within block (pair) $j \in \{1,...,J\}$. This within-block variance given one control and one treated unit per block is (Imbens and Rubin, 2015, Athey and Imbens, 2017):

$$\widehat{V}(\widehat{\tau}_j) = s^2_{j,I} + s^2_{j,C} \tag{1.4}$$

where $s_{j,I}$ and $s_{j,C}$ are the *Incentive* and *Control* sample variances within block j, respectively. Unfortunately, these sample variances are not estimable in a matched-pair setting as there is only one unit in each arm per block. As such, we use the following estimator, which is conservative (confidence intervals wider) if there is heterogeneity in the treatment effect (Imai, 2008, Imbens and Rubin, 2015, Athey and Imbens, 2017):

$$\widehat{SE}(\widehat{\tau}_j) = \Big(\frac{1}{J(J-1)}\sum_{j=1}^{J}(\widehat{\tau}_j - \widehat{\tau})^2\Big)^{\frac{1}{2}} \tag{1.5}$$

Similar to the fixed effect model, the Neyman repeated sampling approach is only estimable if we drop all students whose matched pair attrited. This drop in observations increases the width of our confidence intervals, albeit modestly since including only matched pairs reduces unexplained variance in the outcome variables of interest. We present estimates from all four models to demonstrate that the results are generally similar and not sensitive to model specification or choice of control variables.

Local Average Treatment Effect (LATE)

Here we present the empirical strategies for estimating the causal effect of watching videos on outcomes of interest, exam scores. The average causal effect of watching videos can be modeled as:

$$Y_i = \gamma v_i + g(X_i) + u_i \tag{1.6}$$

where γ is the average causal effect of watching an additional video, Y_i is an outcome of interest (e.g., exam scores) for a student i, $g()$ is a generic function through which X_i, a vector of pretreatment covariates, affects Y_i, and u_i in an unobserved model residual. Because a student's decision to watch videos is likely correlated with unobservable factors (for example, motivation) that are also correlated with outcomes, regressing Y_i on endogenous videos v_i will provide biased estimates of γ. To solve this problem, we rely on variation in v_i induced by an exogenous instrument, Z_i. In our setting, Z_i is assignment to the *Incentive* grade scheme. If Z_i is a valid instrument for v_i, then we can estimate γ using two-stage least squares (2SLS):

$$v_i = \alpha Z_i + f(X_i) + e_i \tag{1.7}$$

$$Y_i = \gamma \widehat{v}_i + g(X_i) + u_i \tag{1.8}$$

where $f()$ and $g()$ are generic functions through which X_i affects v_i and Y_i, respectively, e_i and u_i are unobserved model residuals, and $\widehat{v}_i$ is instrumented videos estimated by Equation 1.7. We assume the influence of Z_i on v_i is monotonically increasing, that is, $v_I = \mathrm{E}(v_i|Z_i = 1) \geq \mathrm{E}(v_i|Z_i = 0) = v_C$. Hence, γ is the per-video average treatment effect, local to students induced by the incentive to watch on average $v_I - v_C$ additional videos.

Under the assumptions of unconfoundedness, excludability, monotonicity, and non-interference, $\widehat{\gamma}$ is an unbiased estimate of the LATE (J. D. Angrist & Imbens, 1995). Unconfoundedness requires that Z_i be independent of potential outcomes, a reasonable assumption given random assignment of students to the *Incentive* arm. Excludability assumes that outcomes (grades) are only affected by the instrument (incentive) through watching videos. This assumption could be violated if, for example, telling a student she is treated were to give her more confidence on subsequent exams during the quarter. Monotonicity, sometimes referred to as the "no defiers" assumption, is necessary because of two-sided noncompliance and requires that students assigned treatment watch weakly more videos than they would if they were assigned control. A violation of this assumption could occur if students get utility from rebelling against their assigned grade scheme. Non-interference, also known as the Stable Unit Treatment Value Assumption (SUTVA), assumes that each student's outcome depends only on their own treatment status and not the treatment status of their peers. Violations of SUTVA may include control students benefiting from having treatment students in the same class and, perhaps, studying together.

Although we believe unconfoundedness,[17] excludability,[18] and monotonicity[19] are reasonable assumptions, we have more concern about non-interference because of the potential for spillovers between students in the same class. If we had unlimited resources, a robust experimental design would assign treatment at the class (or coarser) level, reducing the chance for interactions between treated and control students. However, given our resource constraints, assigning treatment at coarser levels would have resulted in insufficient statistical power to detect reasonable effect sizes. Hence, we proceed acknowledging the potential for spillovers between students. We hypothesize that spillovers likely bias our estimates of the treatment effect *downwards* as we believe control students are more likely to benefit from having well-studied peers than they are to lose from, for example, having peers too busy watching videos to join a study group.[20]

Similar to our estimates of Equation 1.1, we estimate Equation 1.8 with three sets of controls:

[17] Although we randomly assigned treatment, one concern is nonrandom attrition. In the results section, we show that the *Incentive* and *Control* arms remain balanced across observables by the end of the experiment. Additionally, we find our results are similar when restricting the sample to students whose matched pair did not attrite.

[18] While this assumption is not testable, we took care in the experimental design to make the treatment and control arms as similar as possible except for the grading schemes. Of course, watching videos inherently requires time that takes away from some other activity. Hence, the results should be interpreted as the causal effects of more videos and less of whatever else they would have been doing. This subtle point could matter for external validity as a different population of students with zero leisure time may respond differently to the incentive.

[19] Though not testable directly, one testable implication of monotonicity is that the cumulative distribution function of videos watched for each treatment arm should not cross. Indeed, Figure **??** shows that the two CDFs do not cross.

[20] Although spillovers are possible, we believe the magnitude of the spillovers are likely small given that students have for the most part not yet formed strong social networks. 47% of students in the *Incentive* or *Control* arms are transfer students in their first term at the university. The remaining students are predominantly sophomores taking their first upper division course. Social dynamics at the university facilitate networks within "colleges" more than majors for the very reason of encouraging academic diversity among peer groups. One example of a possible positive spillover is the online discussion board where students could ask questions about content covered in the IMVH.

only year and first midterm score, controls chosen using PDS, and a fixed-effect model with controls chosen using PDS. We additionally estimate the LATE using Neyman's repeated sampling approach whose estimators we derive in Appendix 1.9.2.

Treatment Effect Heterogeneity

So far we have discussed estimating average treatment effects across all students below the median score on the first midterm. Now we investigate the extent to which treatment effects vary along key demographic variables. In particular, we add an interaction term to Equation 1.1:

$$Y_i = \beta_1 Z_i + \beta_2 Z_i \mathbb{1}_{x_i=d} + f(X_i) + \varepsilon_i \tag{1.9}$$

where $Z_i \mathbb{1}_{x_i=x}$ takes a value of 1 when treatment students have the value d for demographic variable $x \in X$. β_2 represents the difference in treatment effects for those with demographic d relative to those without. In practice, we estimate Equation 1.9 including in X_i dummies for year and the demographic characteristic of interest as well as first midterm score.

While we observe many demographic variables, we are underpowered to detect reasonable effect sizes after adjusting for multiplicity, given the small sample sizes of our subgroups. As such, we focus on heterogeneity along our blocking variables and covariates we hypothesized *ex ante* may have treatment heterogeneity. We estimate heterogeneity by levels of videos watched pretreatment since it is plausible that those with greater experience watching videos may have greater treatment effects. On the other hand, those who watched many videos pretreatment may have watched many videos during the experiment even if not treated, and hence the incentive may have little effect. As mentioned in Section 1.3, nearly all transfer students in the experiment are taking their first term at a four-year university and may not yet have optimized their studying practices. The university achievement gap between underrepresented and majority groups is well documented, including lower GPAs, longer time to graduation, lower graduation rates on average (Bowen, Chingos, & McPherson, 2009). It is important to know how treatment may affect this gap, potentially through mechanisms including reduced disparities in guidance on studying methods and technologies that may help those for whom English is not a native language, such as captions and the ability to reply videos.[21]

1.4 Results

In this section, we first examine attrition and establish that the *Incentive* and *Control* arms in our analysis sample are balanced on observable characteristics. Second, we show that the grade encouragement worked: students in the *Incentive* arm watched significantly more videos than did their *Control* peers. Third, we estimate the effects of being assigned to the *Incentive* arm (ITT) and the effects of videos (LATE) on grade outcomes. Fourth, to better understand mechanisms, we examine spillovers to other studying methods for Micro A as well as grades in other courses taken during the experiment term. Finally, we see whether behavior change persists after exogenous incentives are removed by estimating treatment effects on video watching and grades in the subsequent microeconomics course, Micro B.

1.4.1 Attrition and balance

At the university where the experiment took place, Micro A is the first upper-division economics course, often taken by students who have not yet declared a major. As such, it has higher withdrawal rates than most other economics courses, a product of both challenging course material and updated priors on interest in the field. In Micro A one year before the experiment, 8.5% and 13.4% of students who took the first midterm did not take the second midterm and final exam, respectively. Unsurprisingly, the withdrawal rates were greater for students who scored below the median on the first midterm: 14.7% and 24.2% of these students did not take the second midterm and final exam, respectively.[22] In the present study, high attrition is not problematic, other than

[21] Unfortunately, we are unable to observe native language directly or better proxies such as country of home address or visa status.

[22] The 2017 statistics are calculated from a sample that differs somewhat in inclusion critefria relative to the 2018 and 2019 samples. We provide these statistics to highlight the historically high rates of attrition and not to make comparisons with the experiment sample.

reducing statistical power, as long as attrition is independent of potential outcomes. If attrition is influenced by treatment status, which could occur, for example, if treatment students believed they were more likely to pass the course,[23] then the resulting nonrandom selection into our analysis could bias the results.

We assess the presence of nonrandom attrition by comparing rates of attrition by treatment arm and examining balance across observable demographic variables in the analysis sample. Students could have attrited in three ways. First, students under the age of 18 at the start of the experiment were removed from the analysis sample. Because of student privacy considerations, demographic variables (including age) were not observable until the conclusion of the experiment. Second, students who opted out of having their data included in the experiment analysis, an option available to students at any time during the experiment, were removed. The analysis sample was prepared and anonymized by a campus-based independent education research organization, per IRB requirement, which removed four minors and seven opt-outs from the sample and merged demographic variables before returning the anonymized data to the research team.[24]

The final and largest cause of attrition is withdrawing from the course. At the present university, students may formally drop a course without penalty up to the fourth week of the quarter. Between the fourth and sixth weeks, students may withdraw from a course, but a "W" grade is assigned in lieu of a letter grade, which does not affect GPA. From the seventh week onwards, students may no longer formally withdraw, but some may choose not to take the second midterm or final exams, which almost assuredly results in a failing grade in the course and does factor into the student's GPA. Because the first midterm took place in week four, the same week as the penalty-free drop deadline, many students took the exam and withdrew before finding out their grades, which were posted in week five. However, because of the lag between when a student drops a course and when the instructor is notified, we assigned treatment to several students not knowing they had already dropped the course. In total, 30 students - two *Above median*, 19 *Control* and nine *Incentive*

[23] As noted earlier, we informed students that final grades would be curved separately between *Control* and *Incentive* arms to remove any advantage treatment may carry towards passing the course.

[24] In total, five *Above median*, three *Control*, and three *Incentive* students were removed for age or opting-out.

- dropped the course before finding out their first midterm scores and treatment assignments. Among students who waited to find out their first midterm scores and treatment assignments, three *Control* and three *Incentive* students did not take the second midterm, and an additional nine *Control* and 13 *Incentive* students did not take the final exam.

We show attrition at each stage of the experiment in Figure **??**. The p-values in the figure are calculated using two-sample t-tests of the equality of attrition rates between the *Control* and *Incentive* arms at each stage. We do not find any statistically significant difference in attrition before the second midterm and final exams that can be attributed to treatment status. However, we do find a significant difference in enrollment rates in Micro B: treatment students were 9.7 percentage points *less* likely to enroll in Micro B than were control students. Conditioning on midterm 1 score and year reduces the gap slightly to 8.9 percentage points ($p = 0.06$). As mentioned in Section 1.3.4, as a robustness check against bias from potentially nonrandom attrition, we estimate treatment effects using models that only include matched-pairs.

Since all students below the median on the first midterm had equal probabilities of being assigned to the *Control* and *Incentive* arms, treatment arms are balanced on covariates in expectation. In practice, due to chance and nonrandom attrition, treatment arms can be unbalanced on covariates, which can bias estimates if not addressed in the analysis, particularly in small samples (Athey & Imbens, 2017). We check balance on observable characteristics after attrition for both the second midterm and final exam samples. As can be seen in Appendix Tables 1.10 and 1.11, we find no statistically significant difference between the *Control* and *Incentive* arms in observable covariates including first midterm score, year, previous term's cumulative GPA, videos watched before the first midterm, ethnicity, gender, and transfer status. However, as discussed in Section 1.3.4, to correct for potential imbalance and to improve precision, we estimate models that include controls chosen via the post-double-selection method of Belloni, Chernozhukov, and Hansen (2014b).

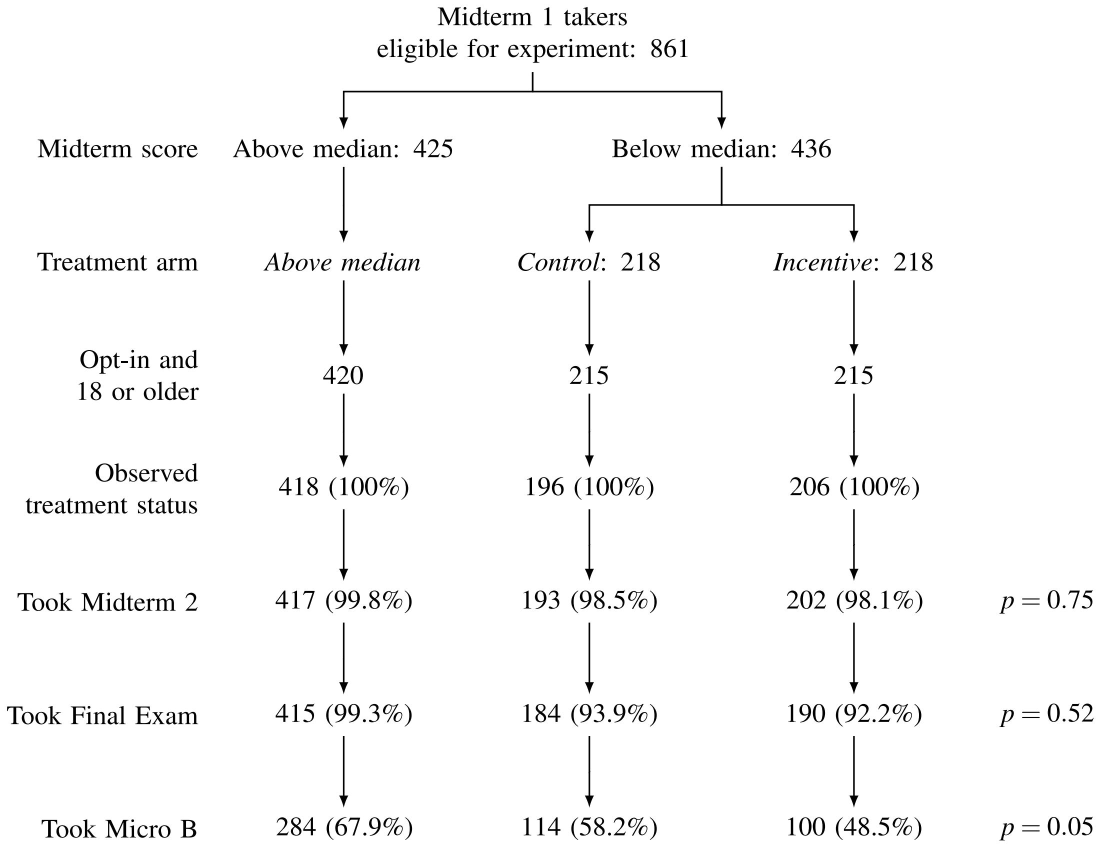

Figure 1.1: Attrition by treatment arm

Percentages in parentheses are the portion of students who observe their treatment status and took the second midterm, final exam, and enrolled in Micro B, calculated separately for each treatment arm. The p-values are from a two-sample t-test of the equality of attrition rates between the *Control* and *Incentive* arms at each stage.

1.4.2 Relevancy of the encouragement instrument

We use a Two-Stage Least Squares approach to estimate the LATE of watching videos on exam performance, as detailed in the Section 1.3.4. We must check that our instrument is both valid and relevant to ensure this method will produce an unbiased estimate of the LATE (Imbens & Rubin, 2015). The validity condition is met by assigning treatment at random conditional on midterm exam score and year of instruction. Balance across pretreatment observables, as demonstrated in Appendix Tables 1.10 and 1.11, give us further confidence that treatment status is uncorrelated with demographics.

Next we check instrument relevancy, that is, whether treatment status generates significantly more video watching. In Table 1.4 we present estimates from Equation 1.1. We find that by the second midterm exam, being assigned to the *Incentive* arm induces students to watch 9.1 - 10.5 videos and 6.0 - 6.8 unique videos more than being assigned to the *Control* arm. The gap between treatment and control grows by the final exam to 38.4 - 39.2 videos and 20.5 - 21.6 unique videos. The larger gap by the final is unsurprising given that the deadline to earn the grade incentive was the day before the final exam. Following the recommendations of Andrews, Stock, and Sun (2019), we assess the strength of our instrument using the effective F-statistic of Montiel Olea and Pflueger (2013) which, in our just-identified setting, coincides with the Wald statistic of Kleibergen and Paap (2006). The effective F-statistic for the second midterm and final exam first-stage specifications are 18.6 and 194.6, respectively, both of which are greater than the Stock and Yogo (2005) critical value of 16.4 and the rule-of-thumb cutoff of 10.

Graphically, we depict the distribution of videos watched as a function of treatment in Figure **??**. Notably, the gap between treatment and control distributions remains significantly positive at every level of video watching by the final exam. The difference is most pronounced near the required number of videos to earn the grade incentive, after which the difference diminishes towards zero. For the second midterm sample, the difference is smaller but significantly different from zero between zero and 62 videos watched. We also show time series plots of video watching in Figure **??**, which show no differences in video watching before the first midterm and marked increases before the second midterm and final exam. Collectively, given the highly significant first-stage regression results, large first-stage F-statistics, and monotonic increase in video watching across the sample, we have high confidence that our instrument meets the relevancy criterion.

1.4.3 Effects on exam scores

In this section, we estimate the causal effects of being assigned to the *Incentive* arm on exam scores (ITT). This estimate is relevant for educators interested in predicting how requiring videos will change exam scores in their classes using the same grade-based incentive implemented in our experiment. Additionally, we estimate the causal effect of watching videos on exam scores (LATE), which is of interest to educators deciding which teaching technologies to provide for their classes as well as to students choosing among different studying tools.

For both the ITTs and LATEs, we examine effects on the second midterm and final exams using both parametric methods (i.e. Equations 1.1 and 1.8) and nonparametric methods a la the repeated sampling framework of Neyman (1923). We check that our parametric results are robust to model specification by estimating Equations 1.1 and 1.8 with and without $f(X_i)$ as a vector of linear control variables chosen via PDS (Belloni, Chernozhukov, & Hansen, 2014b). To address bias from potentially nonrandom attrition across treatment arms, we fit a fixed effect model that drops any student whose matched pair attrited. These specifications and identification strategies are described in detail in Section 1.3.4.

Table 1.5 presents estimates of the effects of treatment on second-midterm and final exam scores. Across our four specifications, we estimate reduced-form (RF) impacts of being assigned to the *Incentive* arm of 0.17 - 0.18 standard deviations on the second midterm. These estimates, along with our first-stage estimates (Table 1.4), imply LATEs of 0.26 - 0.30 standard deviations per 10 unique videos, or 0.16 - 0.18 standard deviations per hour of unique content. For the final exam, we estimate similar ITT effects: being assigned to the *Incentive* arm raises scores by 0.14 - 0.18 standard deviations. However, given the larger first stage effects for the final exam, we estimate smaller LATEs: 0.08 - 0.09 standard deviations per 10 unique videos, or 0.04 - 0.05 standard deviations per hour of unique content.

Given the large F-statistics when estimating first-stage effects of our incentive instrument on videos watched, we are not particularly concerned about bias from weak instruments. However, following the advice of Andrews, Stock, and Sun (2019), we report Anderson-Rubin confidence sets, which are efficient regardless of the strength of our instrument. As can be seen in Appendix Table 1.12, we find that the weak-instrument-robust confidence intervals are very similar to those presented in Table 1.5.

1.4.4 Spillovers during concurrent term

We next estimate spillover effects to other courses taken concurrently during the term of the experiment. Although we find positive effects on exam scores in Micro A, it is important to examine spillover effects to understand the complete picture of how treatment affects student achievement. If the gains in Micro A come at the cost of slowed progress in other courses, then treatment may be counterproductive for some students. On the other hand, since the videos cover concepts and methods that could be covered in other Economics courses, treatment may help students outside Micro A. Although we cannot observe student time use, an important proxy is whether grade outcomes declined in courses unrelated in Micro A, which would suggest that students substitute time away from other courses to watch videos. On the other hand, if grades in these courses remain constant, it is more plausible that students either increased total studying time or substituted from alternative studying methods within Micro A.

Table 1.6 presents our estimates of Equation 1.1 where Y_i is GPA, number of classes passed, or portion of classes taken for a letter grade. We estimate the effects of treatment on term GPA calculated separately for all classes, all classes excluding Micro A, all classes outside of economics, and all economics classes excluding Micro A. In general, we find marginally significant or insignificant but directionally positive spillover effects on GPA.[25] We can rule out large negative spillover effects: in our worst-case specification for term GPA, our 95% confidence interval rules out negative spillover effects larger than -0.02 (on a 4.0 scale), or less than 1% of the mean term GPA among control students. There is no statistically significant difference between our estimates of spillover effects on GPA when restricting to only economics or non-economics courses.

We additionally estimate the effects of treatment on the number of classes passed and find small, insignificant, but directionally positive effects. We find that treatment caused students to pass 0.02 - 0.09 more classes, or about 1% - 3% more classes than the control mean. We find no effect of treatment on number of classes not passed or withdrawn. Interestingly, treatment students

[25] At this university, term GPA is affected only by classes taken for a letter grade. Hence, students may not have attrited from the sample but may have taken all courses Pass/No Pass and thus have no term GPA. As such, we report sample sizes for each GPA specification.

were somewhat less likely to take Micro A for a letter grade than were control students, but this difference is only marginally significant for one of the four specifications and insignificant for the rest. We find no relationship between treatment and fraction of classes taken for a letter grade versus Pass/No Pass. Across all grade spillovers examined, we find mostly small, directionally positive effects, which gives us confidence that treatment is likely not harmful to academic success outside of Micro A.

Besides estimating spillover effects on other course grades, we also examine spillovers to other studying methods within Micro A. Doing so helps us better understand mechanisms: do students substitute away from other studying when encouraged to watch more videos, or are they more likely to complement their video-watching with other unincentivized studying? In Table 1.7, we display the results of estimating equation 1.1 where Y_i is an alternative form of studying. We find that *Incentive* students are directionally less likely to attend class, though point estimates are near zero and not statistically significant. On the other hand, treatment students interacted with the online discussion board more than did control students, but again these estimates are not statistically distinguishable from zero. We do not find any significant relationship between treatment and tutoring attendance. Unfortunately, we do not observe the complete picture of student time use, but our evidence is consistent with students increasing self-study time while holding guided-study time nearly constant.

1.4.5 Spillovers to subsequent term

While we offered treatment students a grade incentive to watch videos during Micro A, students were not offered a grade incentive during the subsequent course in the intermediate microeconomics sequence, Micro B. However, all students in Micro B maintained access to the IMVH, were given direction on which videos to watch each week in the syllabus, and were verbally encouraged to watch videos as a study method. Fortunately, we are able to observe video watching and grade outcomes in Micro B.

We present our estimates of spillover effects during the subsequent term in Table 1.8. In

Panel A, we estimate the effect of treatment on videos watched during the subsequent term, for those who took Micro B. We find large and statistically significant effects: treatment caused students to watch 8.1 - 9.9 more unique videos and 1.2 - 1.5 hours of unique content compared to control students. In Panel B, we estimate equation 1.1 where Y_i is the first midterm, second midterm, or final exam score. Unfortunately given the small subsample of students who took Micro B, we are underpowered to detect effect sizes consistent with those observed during Micro A. Finally, in Panel C, we estimate the effect of treatment on taking Micro B and the number of classes passed and withdrawn. We find no effects statistically distinguishable from zero, though, as mentioned in section 1.4.1, treatment students were directionally less likely to take Micro B than were control students.

1.4.6 Treatment Effect Heterogeneity

Here we estimate Equation 1.9 to examine whether treatment effects vary along key demographic variables. We present our results in Table 1.9, which displays the coefficient estimates of β_2 from Equation 1.9. We find no evidence of treatment effect heterogeneity across our blocking variables, year and first midterm score. We are hesitant to make strong conclusions given the width of our confidence intervals, but this finding is consistent with stable treatment effects across the distribution of student abilities and years of the experiment. In Appendix Figures **??** and **??**, we fit local linear regressions of videos watched, midterm 2 scores, and final exam scores as functions of midterm 1 score. We do not find any statistically significant differences along the midterm 1 score distribution; however, for the final exam, the point estimates are largest for the bottom quarter of midterm 1 scorers. Next, we examine heterogeneity by levels of videos watched pretreatment and find no significant differences in three of four specifications. We find marginally significant positive effects on the second midterm for those with higher levels of pretreatment video watching.

Moving to student demographics, we do *not* find statistically significant heterogeneous treatment effects for transfer students, and the point estimates for the two exams are similar in magnitude and opposite in direction. Interestingly, we *do* find marginally significant negative heterogeneous effects for female students on the final exam, but no significant effect on the second midterm, though the point estimate remains negative. This observation that male students may have benefited more from the intervention is consistent with the findings of Clark et al. (2020) as well as the literature on self-control more broadly,[26] which suggests male students, who tend to have less self-control than female students, may benefit from interventions that address self-control problems. We find significant negative effects for Asian students and positive effects for White students on the second midterm, but no effect on the final exam. However, after adjusting for multiplicity, either using a Bonferroni correction or the less conservative methods proposed by List, Shaikh, and Xu (2019), none of our heterogeneity results remain significant. Our results motivate future work investigating differences along gender and racial dimensions, which has implications for instructor recommendations and personalized education more generally.

1.5 Models of Studying Behavior

To understand our empirical results in the context of economic theory, we discuss three models of student studying behavior: a neoclassical model, an imperfect information model, and a behavioral/procrastination model. For each model we consider the testable implications of a grade inducement to encourage adoption of a study method.[27] Neoclassical models of studying behavior assume that rational agents know their returns to studying using the methods available to them and allocate the optimal study time to each method given their utility function, which is increasing in leisure and grades and decreasing in time spent studying. College instructors have little knowledge about student utility functions and do not know student preferences over performance in other classes or other uses of the student's time which may also have large payoffs in the labor or marriage markets. In this model, there is no room for an instructor to increase student well-being by

[26] See, for example, Duckworth and Seligman (2006), Duckworth et al. (2015), and the works cited therein.

[27] For all three models, we do not address the issue that the IMVH is a relatively unique study tool in that, to our knowledge, it is the first instructional book to be created entirely of videos. However, given the availability of close substitutes to the IMVH (lecture capture, for example) we do not explore the added issues of inducing students to use a study tool whose usefulness is not known to the instructor.

intervening in their study decisions. Both Oettinger (2002) and Kapoor, Oosterveen, and Webbink (2020) find empirical support for the neoclassical model. Oettinger (2002) finds that student effort responds rationally to nonlinear grade incentives: across 1200 students in a principles of economics class with absolute grading standards, he finds evidence of bunching just above letter grade cutoffs and student performance on the final exam is higher if the student is just below a grade threshold. Kapoor, Oosterveen, and Webbink (2020) explore the effects of a university policy that required students who performed poorly in their first year to attend a large fraction of the tutorials for each class in their second year. At the poor performance threshold, both tutorial and lecture attendance increased by over 50 percent with a concomitant decline in self-study hours. Grades for students at the policy threshold were *lower* by 0.16-0.26 standard deviations. At least for students at the margin of poor performance, the requirement to attend tutorials appears to have hurt students by not allowing them to pick their optimal study strategy across self study, tutorials, and lectures.

A key assumption of the neoclassical model is that students possess complete information about the returns across studying methods. However, there is evidence from psychology that college students do not know the return to various study options[28] and many universities fund "Teaching and Learning Centers" or "Academic Skills Centers," part of whose mission is to help undergraduates learn to study more productively.[29] Further, the "raison d'etre" of higher education is not only to teach students specific skills but to teach students how to learn. As an alternative to the neoclassical model, we hypothesize that students supply a quantity of study time that is optimal given their information constraints. In this 'imperfect information' model, students choose study methods and quantities that are suboptimal relative to those they would have picked in a full information setting. Hence, an intervention by an entity that has more information about returns to studying across various methods (i.e. an instructor) can enhance student utility.

A third model is a behavioral one in which students plan to study more than they end up studying when the time comes. This phenomenon is consistent with two-self models in which a

[28] See, for example, McCabe (2011), Pashler et al. (2007), and Dunlosky et al. (2013)

[29] All nine University of California campuses have such a center. Examples outside the UC include Dartmouth's Academic Skills Center, Michigan's Center for Research on Teaching and Learning, UNC's Learning Center, and Yale's Teaching and Learning Center.

person's "planner" self, the one who desires high grades at the expense of leisure, is at odds with her "doer" self who must choose between immediately gratifying leisure and delayed gratification from higher grades. Indeed, survey and experimental data suggest that many students study less than they report they "should" and finish the term with grades lower than what they had anticipated they would earn at the start of the term.[30] Clark et al. (2020) provide empirical support for this model by finding that setting tasked-based goals helps improve college student performance. As descriptive evidence in support of this model, Beattie et al. (2019) find that students that do much worse than expected in college are those who say they have poor time management or procrastination issues, including a tendency to cram and spending very little time studying.

We consider the testable implications of the three models applied to a setting where students are incentivized to use a time-consuming educational input, say, a set of instructional videos (or attending class, reading the textbook, working on homework, etc.). The incentive is structured such that students who consume the educational input receive a higher grade in the course by consuming a set level of the input. In this simple setting, students gain utility only from leisure and grades. We assume grades, a function of time spent studying, and utility are both continuous, smooth, and increasing and concave in their inputs. Students can choose to study using the incentivized educational input, v, or some outside option that is not directly incentivized, o, or a combination thereof.

Across all three models, before the first educational input is incentivized, students allocate time to the two studying methods until the marginal benefit of each (through higher grades) is equal to the marginal cost of forgone leisure. Consider the population of students initially consuming below the requisite level to earn the grade incentive. These students must decide if earning the grade incentive is worth forgone leisure and less time allocated to their outside studying option. Next we explore the differences in predictions across the three models for *compliers*, those for whom the incentive induces greater take-up of the incentivized input.

In the neoclassical model, as long as o and v are not perfect substitutes in the grades

[30] See, for example, Ferrari (1992), Chen et al. (2017) and Lavecchia, Liu, and Oreopoulos (2016).

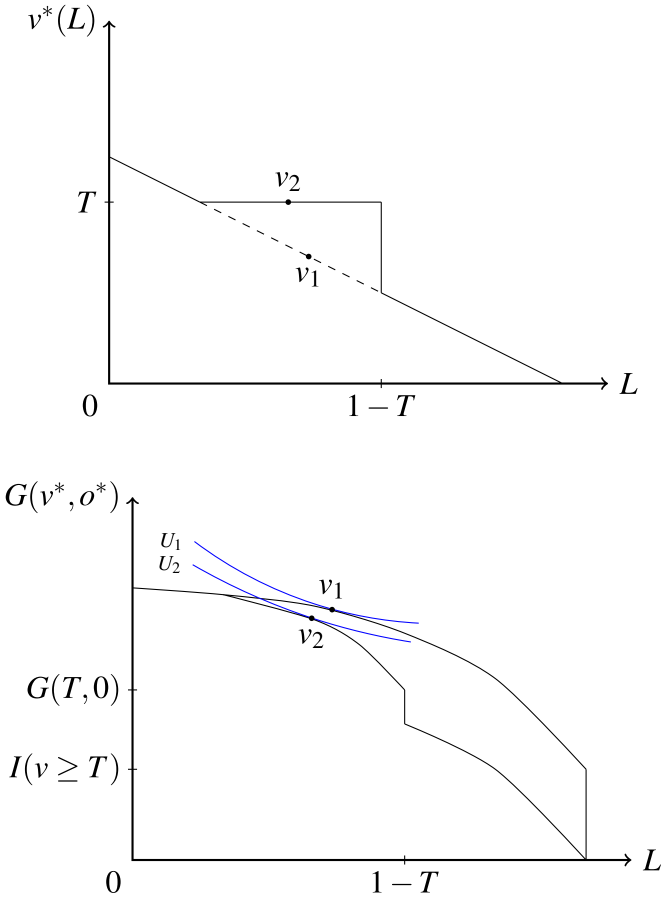

Figure 1.2: Utility maximization problem for model students given the incentive structure in the experiment

A student maximizes her utility U, a function of leisure hours, L, and grades, G. Grades are a function of video watching hours, v, and hours spent on the next best studying option, o. *Above*: Let $v^*(L)$ be the demand curve for video watching as a function of leisure $L \in [0,1]$. Assume the student maximizes her utility by watching v_1 videos and spending o_1 hours on her other study method. Suppose an instructor offers a grade incentive worth I units, which a student can earn by watching $v \geq T$ hours of videos. Note that at $L = 1-T$ hours of leisure, the student maximizes grades by spending all study time watching videos, that is, $v^* = T$. *Below*: Student's utility maximization problem for the neoclassical model. The grade incentive, I, is given to the student conditional on watching T hours of videos (inner budget constraint) or, in the unincentivized case, given regardless of video watching (outer budget constraint). Time bundles along the inner budget constraint are weakly less preferred since the incentive draws the student away from otherwise more efficient (v,o) combinations.

production function, the marginal return to grades of the incentivized input is less than that of the outside option for compliers. This model predicts bunching at the incentivized level cutoff since compliers would prefer to spend their marginal hours on leisure or studying with their other method. This model predicts a weak increase in video watching and weak decrease in other studying and leisure consumption. It is ambiguous whether cumulative study time increases or decreases as this depends on relative utility benefits of leisure and grades and the returns to studying by each method. However, if cumulative study time remains constant or decreases, then exam performance should strictly decrease since students are now suboptimally allocating study time versus their first-best allocation when considering only marginal returns to studying. On the other hand, if cumulative study time increases, students may earn greater exam grades but achieve lower utility compared to baseline. Importantly, this model predicts that in subsequent quarters students return to their pre-incentive levels of studying.

In the imperfect information model, students' *ex ante* allocations to each studying method are not necessarily first-best. The effect of the incentive on video-watching depends on whether students update their priors about the returns to watching videos as they work towards hitting the minimum required level for the grade incentive. At this cutoff, they make a decision whether to continue watching videos depending on their updated perceptions of the marginal benefit. We do not expect bunching in this model unless students believe the cutoff for the grade incentive is the optimal level or the updated marginal benefit at the cutoff is lower than the marginal benefit of the next best studying option or the marginal utility of leisure. A sharp prediction is that video watching will continue at the incentivized level in the absence of the grade incentive as students have learned an effective study tool. We also expect the treatment effect to be greater for students with more information problems, perhaps students in their first semester/quarter at university.

Finally, in the behavioral/procrastination model, the instructor's inducement helps students stick to study plans up to the minimum required level for the grade incentive. As long as total study time does not fall, the inducement will increase exam performance (see Ariely and Wertenbroch (2002) where students with externally set deadlines had higher grades relative to students who

Table 1.3: Predictions Across Models for Compliers, those induced by incentive to consume more videos.

Outcome	Neoclassical	Imperfect Information	Behavioral/ Procrastination
Number of Videos, v	up	up	up
Other study method, o	?	?	?
Total studying, $v+o$	?	?	?
Bunching at 40 videos, T	Yes	No	Yes
Exam Performance	Down*	Up	Up
Video use, future classes**	return to baseline	remain at incentivised level	return to baseline

Notes: *As long as total studying stays the same or falls. **We assume future classes do not incentivize video watching.

choose their own deadlines). This model also predicts bunching at the incentivized level cutoff as long as using the incentivized input does not change the student's "planner" and "doer" selves. In the absence of the inducement, i.e., in future classes, a sharp prediction is that video watching will revert to pre-inducement levels.

We summarize the predictions across models in Table 1.3. One key empirical difference is whether or not students return to their pre-incentive level of video watching in the absence of the incentive. We find that treatment students continued watching videos in the absence of any grade incentive in Micro B, which is inconsistent with the predictions from both the neoclassical and behavioral models of learning in our setting. Another key empirical difference across the models is whether students watch exactly 40 videos, the number of videos required to earn the grade incentive. In Figure **??** we plot the distribution of incentive-eligible videos watched, and one can see that treatment students typically watched more 40 videos. Since students did not receive immediate feedback about the number of videos they had watched, the lack of bunching may be influenced by student uncertainty about whether they had met the 40-video threshold.[31] Nevertheless, we view our results as most consistent with the imperfect information model of learning.

[31] Annecdotally, many students reported keeping track of which videos they had watched by checking-off the list of incentive-eligible videos and/or taking notes on each video watched.

1.6 Discussion

1.6.1 Contributions

Our results add to the literature on what motivates students to increase their performance in college. Previous research finds financial incentives have little effect (see papers cited in Gneezy, Meier, and Rey-Biel (2011)) but tend to work better if educational inputs as opposed to outputs are incentivized (Fryer Jr (2011), Gneezy, Meier, and Rey-Biel (2011)). There is mixed evidence on the effects of having students set goals on the use of educational inputs with no penalty for missing the goal (see Clark et al. (2020) who find positive grade effects of setting goals on number of practice quizzes to complete while Oreopoulos et al. (2019) find that setting a weekly schedule ahead of time and weekly reminders via a text message had only a small effect on study time and no effect on output as measured by grades, retention or credit accumulation.). Interestingly, Clark et al. (2020) find no effect of having students set goals on class outcomes, such as course grade or exam scores. We find that a small grade incentive is effective in motivating poorly performing college students to significantly take-up video watching, an educational input.[32] We also reduced the weight on an early assessment and allowed students to earn back the lost points fully by meeting the video watching requirement, which may also be an important motivator. Grade incentives have the unappealing feature that grades are directly a function of input use. The grade incentive used in this study was small, at most four percent of the student's grade, which may help mitigate this concern.

Second, we find that inducing students who performed poorly on an early assessment to increase the amount of time they spend watching instructional videos increased their exam performance. Since there was no drop in either grades for other courses taken in the same quarter or a drop in the use of many other educational for the class, this suggests that total study time for the course increased for treatment students. Unfortunately, we were not able to determine where the

[32] It is possible that the grade incentive is not an important motivator as Dobkin, Gil, and Marion (2010)) required students to attend class if they performed poorly on an early assessment and, though TAs carefully recorded attendance, there was no penalty for not attending class. Nevertheless, poorly performing students significantly increased class attendance.

added study time for the course did come from, which is important for student welfare calculations. Our results are consistent with other experimental and quasi-experimental studies that find positive effects of educational inputs on college student performance.

Universities often have policies that emphasize improving the performance poorly performing students, such as academic probation policies and the policy studied by Kapoor, Oosterveen, and Webbink (2020). We focus on students who perform poorly on the first midterm, a population similar to that of Dobkin, Gil, and Marion (2010). However, some researchers may ponder why we elected to include only the bottom half of the first midterm distribution in the experiment rather than the entire class. While including the entire class would have increased statistical power, we believe the additional precision might have come at the expense of welfare losses by high performing students. The first midterm provides a signal of which students likely know for themselves how to study, both methods and duration. Coercing these high-type students to spend time with a potentially different studying method runs a greater risk of harming utility. On the other hand, students who, through a low midterm score, make manifest a need for alternative studying practices stand to benefit the most from instructor-provided guidance.

Third, while not statistically significant, we found directionally positive point estimates on treated student's use of other educational inputs (attending class, downloading material from the course web page, and attending a tutoring lab) and grades in other courses taken in the same quarter compared to the control group. This surprising result was also found by Dobkin, Gil, and Marion (2010) who required poorly performing students to attend class. The authors posit that the statistically insignificant but positive spillovers they find may be due to fixed costs of coming to campus. A fixed costs argument is less compelling in our context.

Finally, we find support for an imperfect information model of learning because treatment students frequently watched more than the 40 videos they were required to watch to earn the grade incentive and because treatment students continued watching videos at a significantly higher rate in the following class when watching videos was not exogenously incentivized. J. Angrist, Lang, and Oreopoulos (2009) also find continued higher use of academic support services after the incentivized year for women. An imperfect information model of learning can also account for why incentivizing educational inputs has been found to be more effective than incentivizing grades or exam performance directly (see studies cited in Gneezy, Meier, and Rey-Biel (2011))

Allgood, Walstad, and Siegfried (2015) review the literature on using technology to provide supplemental aids for students in traditional classrooms and conclude that there is little causal evidence that student achievement is improved. The current study stands in contrast to this prior literature. In particular, they conclude that online instruction appears to have a negative effect on course grades and persistence. A more recent paper by Bettinger et al. (2017) confirms these results, particularly for low performing students. The IMVH is clearly the backbone for an online class as it includes all the videos one would require students to watch as part of an online course. Perhaps the main effect of the video-watching requirement in this study was to increase student study time. Would encouraging at-risk students to watch instructional videos in an online class be as effective as we find for a in-person class? We consider this an important area for future research.

1.6.2 Experiment design choices and *local* treatment effects

Our experiment contains several cutoffs: a midterm score cutoff below which students are eligible for the experiment, a video cutoff above which treatment students earn the grade incentive, and a date after which videos no longer count towards the incentive. All of these cutoffs influence the treatment effect estimates, which are *local* to compliers. Specifically, we estimate the effect of videos for those induced by the incentive to watch, so changing who is induced and the videos they are induced to watch will affect estimates of the LATE. Though the purpose of this study was not to identify optimal thresholds, we discuss some observations that may motivate future work and explain our intuition for the size of the *population* average treatment effect (ATE).

We required treatment students to watch 40 of 48 eligible videos to receive the incentive, giving students some agency in which videos they chose to watch. To better understand how the composition of watched videos differs between treatment arms, we plot video watch rates by video in Figure **??**. This plot reveals that the videos students chose to watch are not random. Watch rates in the control group vary across videos from nearly 0% up to over 70%, demonstrating that the perceived intrinsic value of videos varies considerably. For treatment students, watch rates vary less across videos but are positively correlated with control watch rates. Notably, treatment watch rates are 70% or greater for videos whose control watch rates are between 5% and 20%. This gap has implications for the estimated local average treatment effect: if treatment encourages students to watch low value videos, then the LATE estimate will be diluted relative to the population ATE.

We investigate this gap first by considering whether treatment encouraged students to watch shorter videos. One may hypothesize that treatment students would choose to skip the longest videos since earning the incentive is not dependent on length of videos. If students picked the shortest 40 videos, they would watch 1.6 fewer hours of content versus watching the longest 40 videos. We fit a linear model that predicts a video's treatment watch rate using the duration of the video and its control watch rate.[33] Using this model, we find that each additional minute in duration is associated with a 0.3 percentage point decrease in watch rate by the treatment group. For the longest incentivized video, this corresponds with a predicted 3.7 percentage point lower watch rate compared to the average length video. We interpret this finding, given the strong correlation between control and treatment watch rates, as evidence that *content* is a more important driver than is minimizing time cost when choosing which videos to watch. However, it is plausible that some students prioritized shorter-length videos.

Next, we examine how watch rates varied over time. In Figure **??** we plot watch rates by video organized by week each video's content was covered in class. Interestingly, control group watch rates taper off towards the end of the term while treatment watch rates remain much higher. This observation is consistent with the video incentive serving as a commitment mechanism, perhaps encouraging students to spend more time studying for the final exam. Alternatively, perhaps control students watch fewer videos in weeks nine and ten as they shift their study time to other methods while preparing for the final. It is theoretically ambiguous whether the gap in watch rates towards the end of the term suggests a larger or smaller LATE relative to the population ATE.

[33] It is important to control for the control group's watch rate since the educational value of a video may be endogenously related to its length (e.g., harder concepts take longer to explain).

Finally, we consider the influence of the due date on our treatment effect estimates. To earn the grade incentive, students had to watch 40 videos by the last day of instruction, which is the day before the final exam. This is an intuitive deadline, but it allows procrastination-prone students to delay watching many hours of videos until the final week of the term. It is plausible that these students could be harmed by treatment, or at the very least have very small treatment effects, which could explain part of the reason why our LATE estimates are smaller for the final exam than the second midterm. To get a sense of whether treatment students were more likely to "binge watch" videos, we plot the max number of videos watched in one week per student in Figure **??**. Though some control students watch 40 or more videos in one week, this behavior seems more prevalent in the treatment group. While this may very well be a useful, rational studying strategy for some, research suggests that spreading learning over time may be more effective (Kornell, 2009). Ariely and Wertenbroch (2002) find that offering deadlines, though costly from a rational agent perspective, results in better grades. As an alternative to offering one deadline, offering multiple deadlines, perhaps weekly or shortly before both exams, may improve consistency between treatment effect estimates for the two exams.

Collectively, it is not obvious how the population ATE compares with the LATEs we estimate. Neoclassical economic theory suggests that those who select into watching videos regardless of exogenous incentives likely do so because they benefit more than those who only select into watching videos only in the presence of exogenous incentives. This view would suggest that the population ATE is greater than the LATE. However, this view is, perhaps, inconsistent with the alternative that students may not know how much studying is optimal, or what they should be studying. While it is not possible to estimate the population ATE given our experimental design, our intuition is that the ATE is likely higher than the LATE estimated for the final exam and closer to that estimated for the second midterm, which is less likely to be diluted from deadline-induced binge watching and selecting the shortest videos.

1.6.3 Limitations

The present study has several limitations that should be considered before, for example, creating one's own video handbook and incentivizing students to use it. First, the population studied is students who score below the median on the first midterm of an intermediate microeconomics course at a large, highly-selective public research university. The extent to which treatment effects vary by course, instructor, university, or along the top half of the midterm score distribution is important but beyond the scope of this paper.[34] Additionally, the causal effects of watching videos that we estimate are *local* to compliers, i.e. students induced by the grade incentive to watch additional videos. As discussed in Section 1.6.2, we cannot recover the *population* average treatment effect.

The positive effects we estimate are attributable to watching IMVH videos and spending more time studying for the course. Our experimental design does not allow us to separately identify the effects of these two mechanisms for improving exam performance. Would a similar incentive structure that induces greater takeup of an alternative studying method have similar effects? We view this as an important question for future research. While all studying methods require time, some are more efficient than others in terms of effects per unit time. Besides time, learning tools often carry additional costs that should be taken into consideration. For example, students are often required to purchase textbooks for university courses, which may disadvantage students from low-resourced families. Some tools can add costs to the instructional team who, for example, must create and grade problem sets. Internet-delivered videos have the advantage of near-zero marginal financial cost to students and zero marginal time or financial cost to the instructional team.

Another consideration is the time frame during which the experiment took place, 2018 to 2019. About three months after the conclusion of our experiment, most students in the United States and all students at the studied university began remote learning as the COVID-19 pandemic

[34] As an example, Carrell and Kurlaender (2020) found that an intervention for poorly performing students (personalized e-mails from the professor with useul information about where to get help) that was effective in raising exams scores in introductory microeconomics classes at a large public research university did not raise exam scores across several different types of classes in a broader-access institution.

prompted stay-at-home orders. With increased experience learning via electronic media, it is possible that treatment effects will be higher in the future than we estimate in our paper. On the other hand, if students find online learning materials increasingly *less* engaging, we may find the opposite.

In addition to estimating the effects of video handbooks in other educational settings, future research should examine treatment effects in the presence of weekly deadlines instead of one final deadline at the end of the term. Given our observation of greater "binge watching" by treatment students and the smaller effect sizes before the final exam compared to the second midterm exam, we suspect weekly deadlines may reduce the deleterious effects of procrastination. Despite the rich literature on the advantages of spread-out studying (Cepeda et al., 2006; Kornell, 2009), we note that "binge watching" was not unique among treatment students. Indeed, most students within each treatment arm watched more videos the last week of the term than any other week.

1.7 Conclusion

We examine the effectiveness of an innovative educational technology, a video handbook composed of 220 brief instructional videos on intermediate microeconomic theory. We used random assignment of a grade-based incentive to experimentally vary takeup of the video handbook, and we found that greater takeup caused students to score significantly higher on exams. Specifically, we estimate that treatment causes students to score about 0.18 standard deviations higher on midterm and final exams. For students on the margin of watching videos, watching an additional hour of unique content causes students to score between 0.05 to 0.15 standard deviations higher on exams.

Instructors may have concerns about making a resource such as the IMVH available if they believe students may substitute away from lectures or other more productive studying methods (Kay, 2012). Another concern is that forcing students to spend more time studying in one's class may worsen performance in other classes. Our analysis provides some confidence that neither of these fears are first-order concerns. We do not find evidence that students decrease their consumption of other forms of studying, nor do we find that students perform worse in other courses during the same quarter. Our point estimates of the effect of treatment on takeup of other studying methods, though not statistically significantly different from zero, are *positive* for most alternatives, suggesting that if any, students consider the videos complements to other forms of studying. A potential mechanism might be that the videos help students realize what they *don't* know, increasing the marginal benefit of subsequent studying.

A final concern is one of welfare. In a neoclassical model, instructors cannot make their students better off by exogenously incentivizing quantities of studying they would not otherwise have chosen for themselves. In an imperfect information model, which we think is more appropriate in our university classroom setting, instructors *can* improve student welfare through intervention when information barriers lead to suboptimal time allocation decisions. We observe two phenomena that support this model. First, treatment students do not bunch at the cutoff for the grade incentive. Second, video consumption remains much higher among treatment students in the term following conclusion of the experiment.

While there are many educational interventions that instructors could offer their students, the research on causal effects of educational interventions remains limited. Our study serves as an example of a feasible research design that runs a lower risk of generating welfare losses for high performing students than does a class-wide experiment. It is our hope, as educators ourselves, that more research be conducted on the effectiveness of pedagogical technologies.

1.8 Acknowledgements

We thank the students who took intermediate microeconomics in the fall of 2018 and 2019 for participating in this study. We also thank UC San Diego's Teaching + Learning Commons for helping provide and anonymize the data used for analysis. We gratefully recognize Jessica Morales, Yingjia Zhang, and Shuli Zhu for outstanding research assistance. Finally, we thank the applied microeconomics group at UC San Diego for their helpful feedback on the experimental design and

analysis. This research was approved under UC San Diego's Human Research Protections Program (IRB approval 170886 in fall 2018 and 2019).

Chapter 1 is currently being prepared for submission for publication of the material. Famulari, Melissa; Goodman, Zachary A. The dissertation/thesis author was a primary investigator and author of this material.

Bibliography

Allgood, S., Walstad, W. B., & Siegfried, J. J. (2015). Research on teaching economics to undergraduates. *Journal of Economic Literature*, *53*(2), 285–325.

Andrews, I., Stock, J. H., & Sun, L. (2019). Weak instruments in instrumental variables regression: Theory and practice. *Annual Review of Economics*, *11*, 727–753.

Angrist, J., Lang, D., & Oreopoulos, P. (2009). Incentives and services for college achievement: Evidence from a randomized trial. *American Economic Journal: Applied Economics*, *1*(1), 136–63.

Angrist, J. D., & Imbens, G. W. (1995). Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American statistical Association*, *90*(430), 431–442.

Ariely, D., & Wertenbroch, K. (2002). Procrastination, deadlines, and performance: Self-control by precommitment. *Psychological Science*, *13*(3), 219–224.

Arulampalam, W., Naylor, R. A., & Smith, J. (2012). Am i missing something? the effects of absence from class on student performance. *Economics of Education Review*, *31*(4), 363–375.

Athey, S., & Imbens, G. W. (2017). The econometrics of randomized experiments. *Handbook of economic field experiments* (pp. 73–140). Elsevier.

Beattie, G., Laliberté, J.-W. P., Michaud-Leclerc, C., & Oreopoulos, P. (2019). What sets college thrivers and divers apart? a contrast in study habits, attitudes, and mental health. *Economics letters*, *178*, 50–53.

Belloni, A., Chernozhukov, V., & Hansen, C. (2014a). High-dimensional methods and inference on structural and treatment effects. *Journal of Economic Perspectives*, *28*(2), 29–50.

Belloni, A., Chernozhukov, V., & Hansen, C. (2014b). Inference on treatment effects after selection among high-dimensional controls. *The Review of Economic Studies*, *81*(2), 608–650.

Bettinger, E. P., Fox, L., Loeb, S., & Taylor, E. S. (2017). Virtual classrooms: How online college courses affect student success. *American Economic Review*, *107*(9), 2855–75.

Bonesrønning, H., & Opstad, L. (2012). How much is students' college performance affected by quantity of study? *International Review of Economics Education*, *11*(2), 46–63.

Bonesrønning, H., & Opstad, L. (2015). Can student effort be manipulated? does it matter? *Applied Economics*, *47*(15), 1511–1524.

Bowen, W. G., Chingos, M. M., & McPherson, M. S. (2009). *Crossing the finish line: Completing college at america's public universities*. Princeton University Press.

Bratti, M., & Staffolani, S. (2013). Student time allocation and educational production functions. *Annals of Economics and Statistics/ANNALES D'ÉCONOMIE ET DE STATISTIQUE*, 103–140.

Bruhn, M., & McKenzie, D. (2009). In pursuit of balance: Randomization in practice in development field experiments. *American economic journal: applied economics*, *1*(4), 200–232.

Carrell, S. E., & Kurlaender, M. (2020). *My professor cares: Experimental evidence on the role of faculty engagement* (Working Paper No. 27312). National Bureau of Economic Research.

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological bulletin*, *132*(3), 354.

Chen, P., Chavez, O., Ong, D. C., & Gunderson, B. (2017). Strategic resource use for learning: A self-administered intervention that guides self-reflection on effective resource use enhances academic performance. *Psychological Science*, *28*(6), 774–785.

Clark, D., Gill, D., Prowse, V., & Rush, M. (2020). Using goals to motivate college students: Theory and evidence from field experiments. *The Review of Economics and Statistics*, *102*(4), 648–663.

Dobkin, C., Gil, R., & Marion, J. (2010). Skipping class in college and exam performance: Evidence from a regression discontinuity classroom experiment. *Economics of Education Review*, *29*(4), 566–575.

Duckworth, A. L., Shulman, E. P., Mastronarde, A. J., Patrick, S. D., Zhang, J., & Druckman, J. (2015). Will not want: Self-control rather than motivation explains the female advantage in report card grades. *Learning and individual differences*, *39*, 13–23.

Duckworth, A. L., & Seligman, M. E. (2006). Self-discipline gives girls the edge: Gender in self-discipline, grades, and achievement test scores. *Journal of educational psychology*, *98*(1), 198.

Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, *14*(1), 4–58.

Ferrari, J. R. (1992). Psychometric validation of two procrastination inventories for adults: Arousal and avoidance measures. *Journal of Psychopathology and Behavioral Assessment*, *14*(2), 97–110.

Fryer Jr, R. G. (2011). Financial incentives and student achievement: Evidence from randomized trials. *The Quarterly Journal of Economics*, *126*(4), 1755–1798.

Gneezy, U., Meier, S., & Rey-Biel, P. (2011). When and why incentives (don't) work to modify behavior. *Journal of economic perspectives*, *25*(4), 191–210.

Grodner, A., & Rupp, N. (2013). The role of homework in student learning outcomes: Evidence from a field experiment. *The Journal of Economic Education*, *44*(2), 93–109.

Imai, K. (2008). Variance identification and efficiency analysis in randomized experiments under the matched-pair design. *Statistics in medicine*, *27*(24), 4857–4873.

Imbens, G. W., & Rubin, D. B. (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.

Joyce, T., Crockett, S., Jaeger, D. A., Altindag, O., & O'Connell, S. D. (2015). Does classroom time matter? *Economics of Education Review*, *46*(100), 64–77.

Kapoor, S., Oosterveen, M., & Webbink, D. (2020). The price of forced attendance. *Journal of Applied Econometrics*.

Kay, R. H. (2012). Review: Exploring the use of video podcasts in education: A comprehensive review of the literature. *28*(3).

Kirby, A., & McElroy, B. (2003). The effect of attendance on grade for first year economics students in university college cork. *The Economic and Social Review*, *34*, 311–326.

Kleibergen, F., & Paap, R. (2006). Generalized reduced rank tests using the singular value decomposition. *Journal of econometrics*, *133*(1), 97–126.

Kornell, N. (2009). Optimising learning using flashcards: Spacing is more effective than cramming. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *23*(9), 1297–1317.

Krohn, G. A., & O'Connor, C. M. (2005). Student effort and performance over the semester. *The Journal of Economic Education*, *36*(1), 3–28.

Lavecchia, A. M., Liu, H., & Oreopoulos, P. (2016). Behavioral economics of education: Progress and possibilities. *Handbook of the economics of education* (pp. 1–74). Elsevier.

List, J. A., Shaikh, A. M., & Xu, Y. (2019). Multiple hypothesis testing in experimental economics. *Experimental Economics*, *22*(4), 773–793.

McCabe, J. (2011). Metacognitive awareness of learning strategies in undergraduates. *Memory & Cognition*, *3*(39), 462–476.

Montiel Olea, J. L., & Pflueger, C. (2013). A robust test for weak instruments. *Journal of Business & Economic Statistics*, *31*(3).

Morris, N. P., Swinnerton, B., & Coop, T. (2019). Lecture recordings to support learning: A contested space between students and teachers. *Computers & Education*, *140*, 103604.

Munley, V. G., Garvey, E., & McConnell, M. J. (2010). The effectiveness of peer tutoring on student achievement at the university level. *American Economic Review*, *100*(2), 277–82.

Neyman, J. (1923). On the application of probability theory to agricultural experiments: Essay on principles, section 9. *Statistical Science*, *5*(4), 465–472.

Oettinger, G. S. (2002). The effect of nonlinear incentives on performance: Evidence from "econ 101". *The Review of Economics and Statistics*, *84*(3), 509–517.

Oreopoulos, P., Patterson, R. W., Petronijevic, U., & Pope, N. G. (2019). Low-touch attempts to improve time management among traditional and online college students. *Journal of Human Resources*, 0919–10426R1.

Pashler, H., Rohrer, D., Cepeda, N. J., & Carpenter, S. K. (2007). Enhancing learning and retarding forgetting: Choices and consequences. *Psychonomic Bulletin and Review*, *14*(2), 187–193.

Setren, E., Greenberg, K., Moore, O., & Yankovich, M. (2021). Effects of flipped classroom instruction: Evidence from a randomized trial. *Education Finance and Policy*, Forthcoming.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological science*, *22*(11), 1359–1366.

Stinebrickner, R., & Stinebrickner, T. R. (2008). The causal effect of studying on academic performance. *The B.E. Journal of Economic Analysis & Policy*, *8*(1), 1–55.

Stock, J. H., & Yogo, M. (2005). Testing for weak instruments in linear iv regression. *Identification and inference for econometric models: Essays in honor of Thomas Rothenberg*, *80*(4.2), 1.

Tang, L., Li, S., Auden, E., & Dhuey, E. (2020). Who benefits from regular class participation? *The Journal of Economic Education*, *51*(3-4), 243–256.

Trost, S., & Salehi-Isfahani, D. (2012). The effect of homework on exam performance: Experimental results from principles of economics. *Southern Economic Journal*, *79*(1), 224–242.

Wozny, N., Balser, C., & Ives, D. (2018). Evaluating the flipped classroom: A randomized controlled trial. *The Journal of Economic Education*, *49*(2), 115–129.

Table 1.4: Effects of Grade Incentive on Video Watching

	Control Mean	(1)	(2)	(3)	(4)
Panel A: By Midterm 2					
Videos	33.91	10.19***	10.54***	9.08***	9.58***
		(2.85)	(3.12)	(2.03)	(2.19)
Unique videos	23.13	6.63***	6.79***	5.97***	6.11***
		(1.54)	(1.70)	(0.98)	(1.11)
Hours of videos	5.88	1.68***	1.72***	1.48***	1.55***
		(0.50)	(0.55)	(0.35)	(0.38)
Hours of unique videos	3.85	1.10***	1.13***	0.99***	1.02***
		(0.25)	(0.28)	(0.16)	(0.18)
Observations		395	362	395	362
Panel B: By Final Exam					
Videos	53.09	39.25***	39.07***	38.57***	37.99***
		(4.06)	(4.37)	(3.40)	(3.69)
Unique videos	33.95	21.55***	21.08***	21.28***	20.49***
		(1.55)	(1.66)	(1.22)	(1.27)
Hours of videos	8.93	6.30***	6.26***	6.18***	6.05***
		(0.69)	(0.75)	(0.57)	(0.62)
Hours of unique videos	5.54	3.43***	3.36***	3.38***	3.26***
		(0.25)	(0.27)	(0.20)	(0.21)
Observations		374	332	374	332
Treatment assignment controls		Yes	No	Yes	Yes
Demographic controls		No	No	Yes	Yes
Pair Fixed Effects		No	No	No	Yes

Note: Model (1) contains linear controls for midterm 1 score and year; (2) is the difference in means and standard errors calculated using the repeated sampling framework of Neyman (1923); (3) and (4) use the post-double-selection (PDS) procedure of Belloni, Chernozhukov, and Hansen (2014b) to select control variables then estimate treatment effects and standard errors. The control variables selected using PDS are listed in Table 1.14. Models (2) and (4) include only students whose matched-pair did not attrite from the experiment. *Control Mean* is the mean for the Control students included in models (1) and (3), which is nearly identical to the mean for the Control students included in models (2) and (4). Standard errors in parentheses are robust to heteroskedasticity. ***, **, and * indicate significance at the 1, 5, and 10 percent critical levels, respectively.

Table 1.5: Effects of Videos on Grades

	(1)	(2)	(3)	(4)
Panel A: Midterm 2 score				
RF: Incentive	0.176*	0.183*	0.176*	0.174*
	(0.090)	(0.094)	(0.090)	(0.096)
2SLS: 10 videos	0.266*	0.270*	0.300**	0.293**
	(0.146)	(0.150)	(0.151)	(0.145)
2SLS: 1 hour of videos	0.160*	0.163*	0.181**	0.170*
	(0.087)	(0.090)	(0.090)	(0.095)
Observations	395	362	395	362
Panel B: Final exam score				
RF: Incentive	0.175**	0.174*	0.175**	0.138
	(0.089)	(0.103)	(0.088)	(0.103)
2SLS: 10 videos	0.081**	0.082*	0.082**	0.087*
	(0.041)	(0.049)	(0.041)	(0.046)
2SLS: 1 hour of videos	0.051**	0.052*	0.052**	0.043
	(0.026)	(0.031)	(0.026)	(0.031)
Observations	374	332	374	332
Treatment assignment controls	Yes	No	Yes	Yes
Demographic controls	No	No	Yes	Yes
Pair Fixed Effects	No	No	No	Yes

Note: This table reports coefficients on $Incentive_i$ from Equation 1.1 (Reduced Form, *RF*) and $\hat{Video}_i$ from Equation 1.8 (Two-Stage Least Squares, *2SLS*). Test scores are measured in standard deviation units. Model (1) contains linear controls for midterm 1 score and year; (2) is the difference in means and standard errors calculated using the repeated sampling framework of Neyman (1923); (3) and (4) use the post-double-selection (PDS) procedure of Belloni, Chernozhukov, and Hansen (2014b) to select control variables then estimate treatment effects and standard errors. The control variables selected using PDS are listed in Table 1.14. Models (2) and (4) include only students whose matched-pair did not attrite from the experiment. Standard errors in parentheses are robust to heteroskedasticity. ***, **, and * indicate significance at the 1, 5, and 10 percent critical levels, respectively.

Table 1.6: Spillover Effects of Incentive on Other Course Grades

	Control Mean	(1)	(2)	(3)	(4)
Panel A: Effects on Term GPA					
All classes	2.59	0.13**	0.13*	0.11*	0.10
		(0.06)	(0.07)	(0.06)	(0.06)
		373	332	373	332
Excluding Micro A	2.75	0.10	0.11	0.09	0.10
		(0.07)	(0.08)	(0.07)	(0.08)
		370	329	370	329
Excluding econ classes	2.99	0.06	0.09	0.06	0.08
		(0.10)	(0.09)	(0.09)	(0.12)
		315	278	315	278
Econ classes ex. Micro A	2.44	0.07	0.02	0.07	-0.03
		(0.09)	(0.08)	(0.09)	(0.12)
		258	228	258	228
Panel B: Effects on classes passed					
Num. classes passed	3.28	0.08	0.09	0.05	0.02
		(0.09)	(0.10)	(0.09)	(0.09)
Num. classes not passed	0.31	0.01	-0.01	0.01	-0.01
		(0.06)	(0.06)	(0.06)	(0.06)
Num. classes withdrawn	0.05	0.01	0.01	0.01	0.01
		(0.03)	(0.02)	(0.03)	(0.02)
Panel C: Effects on class grade type					
Letter grade in Micro A	0.95	-0.04	-0.05*	-0.03	-0.04
		(0.03)	(0.03)	(0.02)	(0.03)
% classes taken for letter	0.93	-0.01	-0.01	-0.01	-0.01
		(0.01)	(0.02)	(0.01)	(0.02)
% classes taken P/NP	0.07	0.01	0.01	0.01	0.01
		(0.01)	(0.02)	(0.01)	(0.02)
Observations		374	332	374	332
Treatment assignment controls		Yes	No	Yes	Yes
Demographic controls		No	No	Yes	Yes
Pair Fixed Effects		No	No	No	Yes

Note: This table reports coefficients on $Incentive_i$ from Equations 1.1. GPA is measured on a 4.0 scale and is only affected by courses taken for a letter grade. Courses taken for Pass/No Pass (P/NP) have no bearing on GPA, nor do withdrawn courses. Model (1) contains linear controls for midterm 1 score and year; (2) is the difference in means and standard errors calculated using the repeated sampling framework of Neyman (1923); (3) and (4) use the post-double-selection (PDS) procedure of Belloni, Chernozhukov, and Hansen (2014b) to select control variables then estimate treatment effects and standard errors. The control variables selected using PDS are listed in Table 1.14. Models (2) and (4) include only students whose matched-pair did not attrite from the experiment. *Control Mean* is the mean for the Control students included in models (1) and (3), which is nearly identical to the mean for the Control students included in models (2) and (4). Standard errors in parentheses are robust to heteroskedasticity. ***, **, and * indicate significance at the 1, 5, and 10 percent critical levels, respectively.

Table 1.7: Spillover Effects of Incentive on Other Studying

	Control Mean	(1)	(2)	(3)	(4)
Attendance checks	5.91	-0.08	-0.09	-0.16	-0.10
		(0.18)	(0.17)	(0.17)	(0.18)
Discussion board views	49.81	10.64	8.51	10.64	3.69
		(7.64)	(8.25)	(7.60)	(8.05)
Discussion board days online	10.40	1.43	1.89	1.43	1.67
		(1.55)	(1.59)	(1.54)	(1.65)
Discussion board questions asked	0.53	0.32	0.30	0.32	0.30
		(0.25)	(0.30)	(0.25)	(0.31)
Discussion board answers	0.47	0.08	0.01	0.08	-0.02
		(0.26)	(0.28)	(0.26)	(0.28)
Tutoring visits	0.41	0.05	-0.01	0.07	0.00
		(0.13)	(0.14)	(0.12)	(0.12)
Observations		374	332	374	332
Treatment assignment controls		Yes	No	Yes	Yes
Demographic controls		No	No	Yes	Yes
Pair Fixed Effects		No	No	No	Yes

Note: This table reports coefficients on $Incentive_i$ from Equations 1.1. There were seven *Attendance checks* during the quarter. *Tutoring visits* includes those after the first midterm. Model (1) contains linear controls for midterm 1 score and year; (2) is the difference in means and standard errors calculated using the repeated sampling framework of Neyman (1923); (3) and (4) use the post-double-selection (PDS) procedure of Belloni, Chernozhukov, and Hansen (2014b) to select control variables then estimate treatment effects and standard errors. The control variables selected using PDS are listed in Table 1.14. Models (2) and (4) include only students whose matched-pair did not attrite from the experiment. *Control Mean* is the mean for the Control students included in models (1) and (3), which is nearly identical to the mean for the Control students included in models (2) and (4). Standard errors in parentheses are robust to heteroskedasticity. ***, **, and * indicate significance at the 1, 5, and 10 percent critical levels, respectively.

Table 1.8: Spillover Effects during Subsequent Quarter

	Control Mean	(1)	(2)	(3)	(4)
Panel A: Videos during subsequent quarter					
Num. of videos	25.46	14.00***	12.78*	11.70***	11.35
		(4.45)	(6.74)	(4.24)	(7.08)
Num. unique videos	19.77	9.87***	8.85**	8.25***	8.07**
		(3.03)	(4.04)	(2.92)	(4.12)
Hours of videos	3.82	2.14***	1.88*	1.79***	1.70
		(0.68)	(1.03)	(0.64)	(1.08)
Hours unique videos	2.90	1.51***	1.33**	1.27***	1.22**
		(0.45)	(0.60)	(0.44)	(0.61)
Observations		211	108	211	108
Panel B: Effects on classes passed					
Midterm 1 score		-0.04	-0.24	-0.04	-0.30
		(0.13)	(0.18)	(0.13)	(0.19)
		213	112	213	112
Midterm 2 score		0.00	-0.04	0.00	0.03
		(0.13)	(0.20)	(0.13)	(0.21)
		214	112	214	112
Final exam score		0.12	0.00	0.12	0.23
		(0.14)	(0.18)	(0.14)	(0.23)
		211	108	211	108
Panel C: Effects on class grade type					
Took Micro B	0.61	-0.07	-0.07	-0.07	-0.08
		(0.05)	(0.05)	(0.05)	(0.06)
Num. classes passed	3.46	-0.07	-0.05	-0.07	-0.04
		(0.11)	(0.12)	(0.11)	(0.12)
Num. classes not passed	0.23	0.07	0.08	0.07	0.07
		(0.06)	(0.06)	(0.06)	(0.06)
Num. classes withdrawn	0.06	0.04	0.04	0.04	0.03
		(0.03)	(0.03)	(0.03)	(0.03)
Observations		374	332	374	332
Treatment assignment controls		Yes	No	Yes	Yes
Demographic controls		No	No	Yes	Yes
Pair Fixed Effects		No	No	No	Yes

Note: This table reports coefficients on $Incentive_i$ from Equations 1.1. Panel A restricts the sample to those who completed both the first and second microeconomics courses (Micro A and B). Panel C includes those who completed the first microeconomics course (Micro A). Test scores are measured in standard deviation units. Model (1) contains linear controls for midterm 1 score and year; (2) is the difference in means and standard errors calculated using the repeated sampling framework of Neyman (1923); (3) and (4) use the post-double-selection (PDS) procedure of Belloni, Chernozhukov, and Hansen (2014b) to select control variables then estimate treatment effects and standard errors. The control variables selected using PDS are listed in Table 1.14. Models (2) and (4) include only students whose matched-pair did not attrite from the experiment. *Control Mean* is the mean for the Control students included in models (1) and (3), which is nearly identical to the mean for the Control students included in models (2) and (4). Standard errors in parentheses are robust to heteroskedasticity. ***, **, and * indicate significance at the 1, 5, and 10 percent critical levels, respectively.

Table 1.9: Heterogeneous Effects of Treatment

Interaction Variable	Midterm 2	Final Exam
Midterm 1 score	0.009	0.002
	(0.086)	(0.095)
Year = 2019	0.068	0.017
	(0.181)	(0.179)
Pretreatment videos	0.011	-0.004
	(0.007)	(0.007)
Pretreatment videos, unique	0.020*	0.001
	(0.011)	(0.011)
Transfer	-0.139	0.130
	(0.183)	(0.178)
Female	-0.208	-0.322*
	(0.185)	(0.188)
Asian	-0.402**	-0.215
	(0.189)	(0.180)
Latinx	0.237	0.022
	(0.254)	(0.227)
White	0.638**	0.271
	(0.260)	(0.250)
Other ethnicity	-0.113	0.276
	(0.307)	(0.347)
Observations	395	374

Note: This table reports estimates for β_2 from Equation 1.9. Test scores are measured in standard deviation units. Standard errors in parentheses are robust to heteroskedasticity. ***, **, and * indicate significance at the 1, 5, and 10 percent critical levels, respectively.

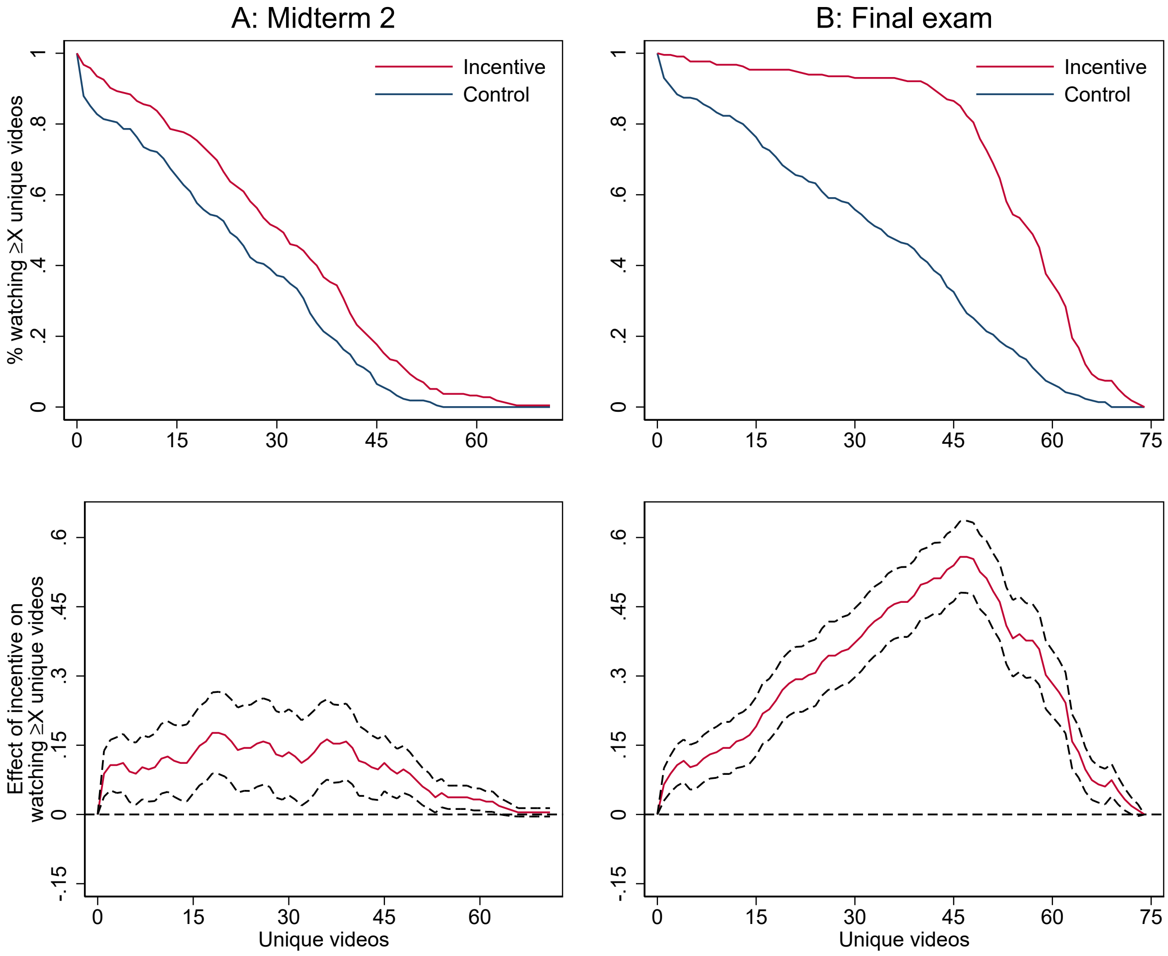

Figure 1.3: Effect of grade incentive on videos watched
Top panels display the percent students in the *Control* and *Incentive* arms that watched at least X unique videos (left) or hours of unique videos (right). Bottom panels display the differences between the two arms in the top panels with 95% confidence intervals estimated by regressing an indicator for whether on the student watched at least $X \in \{0, ..., X_{max}\}$ unique videos (or hours of unique videos) on the student's treatment status.

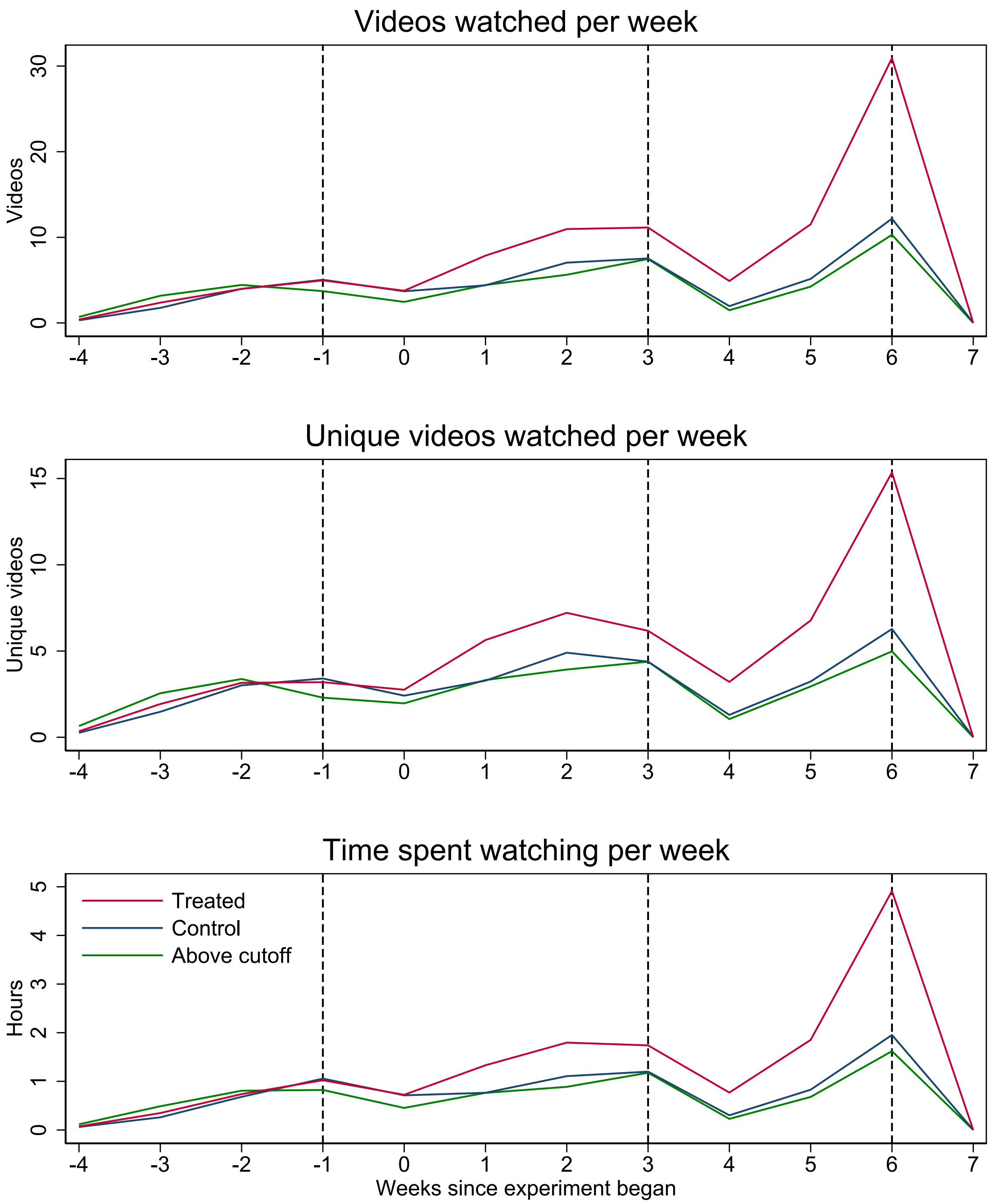

Figure 1.4: Weekly video watching by treatment arm
Dashed lines represent Midterm 1, Midterm 2, and Final exams

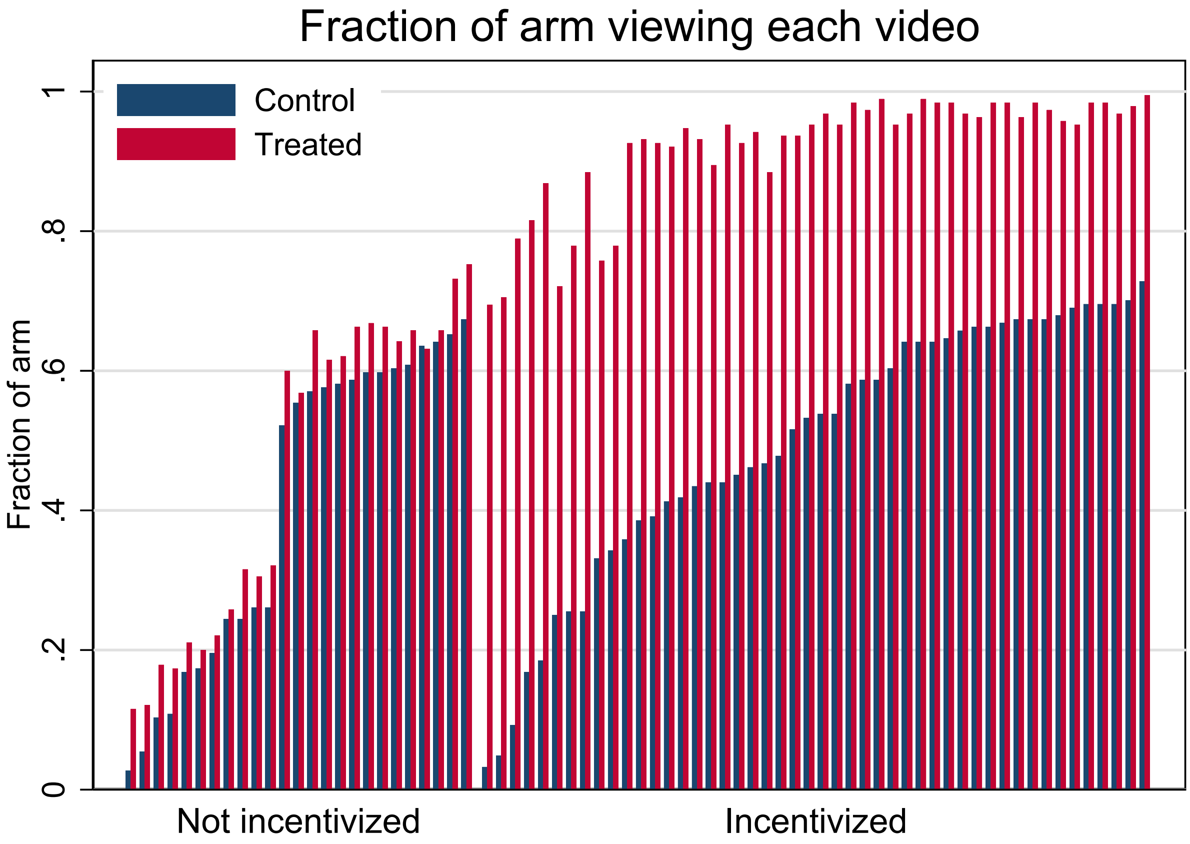

Figure 1.5: Video watch rates by video and treatment arm, grouped by incentive
Each bar represents the fraction of the treatment arm that watched a particular video. Bars are in order of control group watch rates separately for incentivized and non-incentivized videos.

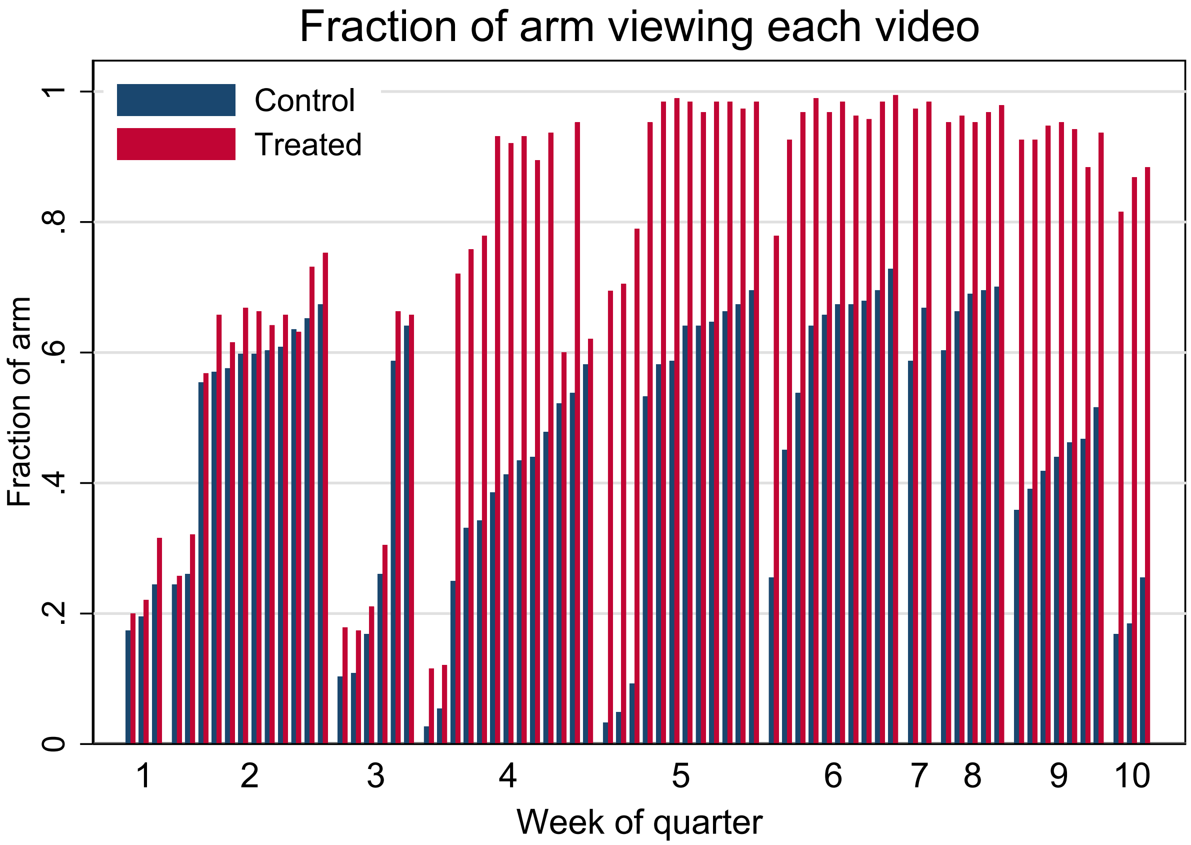

Figure 1.6: Video watch rates by video and treatment arm, grouped by week
Each bar represents the fraction of the treatment arm that watched a particular video. Bars are in order of control group watch rates separately for each week the corresponding content was covered in lecture, as listed in the syllabus.

1.9 Appendix

1.9.1 Additional experiment details

In this section we outline additional experiment details that could prove useful for replication or understanding our analysis choices.

Randomization

Students were assigned to treatment arms using a matched pairs design, a special case of blocked randomization in which each block contains exactly two units, one treated and one control. Several authors detail how matched pair designs can improve the *ex ante* precision of treatment effect estimates (versus complete randomization) by matching treatment units whose potential outcomes are similar (e.g. Imbens and Rubin, 2015, Athey and Imbens, 2017).

Additionally, we were unable to observe most pretreatment covariates until after the experiment had concluded because of student privacy considerations, thereby making it impossible to block on these variables. We learned from the previous cohorts' data that between the first midterm score and math quiz score, both observable at the time of randomization, the midterm score predicted significantly more variation in the final exam score. Hence, we stratified on midterm score when assigning treatment. While we could have used an alternative method (e.g. matching methods) that take into consideration multiple covariates when assigning treatment, we opted for a simpler design given the high correlation between midterm and math quiz score and the comparatively high number of missing observations for the latter assessment (the math quiz was given on the second class day and so before some students enrolled in the class).

We assigned treatment shortly after issuing the first midterm exam grades, which occurred during the fourth week of the quarter. To assign treatment, we ordered the students by exam score, then paired students along this ordering for students below the median. Within pairs, we randomly assigned one student to *Incentive*, the other to *Control*. By construction, these two arms were *ex ante* balanced on midterm exam score, and we verified at time of treatment that the arms were also balanced on math quiz score. Since this randomization was performed independently across year cohorts, by construction, the samples were also balanced on year.

Although our treatment assignment method provides a better chance of balance than does simple random sampling, by random chance and through non-random attrition, it is possible that the two treatment arms vary on *ex post* observable and unobservable covariates that are correlated with the outcomes of interest, thereby confounding our treatment effect estimates. The primary cause of attrition was withdrawing from the course, which reduced our experiment sample by 35 students before the second midterm and an additional 21 students before the final exam. A 13% withdrawal rate is in line with those observed in previous quarters. Another cause of attrition, albeit not from the course, is age: four students under the age of 18 during the experiment were removed from the analysis dataset. Additionally, seven students opted out of having their data included in the experiment analysis.

Since neither the students' intent to withdraw, age, nor opt-out preferences were observable at the time of treatment assignment, we could not *ex ante* balance this attrition across treatment arms. If students attrited non-randomly, that is, decided to attrite depending on their treatment status, then our treatment effect estimates would be biased. Fortunately, despite 8% attrition before the second midterm and 13% before the final exam, the two treatment arms below the median are balanced on nearly all observable pretreatment covariates, as shown in Tables 1.11 and 1.10, which gives us confidence that the *Control* arm is a good counterfactual for the *Incentive* arm.

Selection of control variables

In this section we discuss how we select control variables included in our linear models.

Equation 1.1 includes a vector of control variables related linearly to the outcomes of interest. Although d_i, the treatment indicator is randomly assigned and in expectation d_i is orthogonal to all observed and unobserved pretreatment covariates, in small samples stochastic imbalances can occur, which if controlled for can reduce bias of the treatment effect estimator (Athey & Imbens, 2017). Even if perfect balance is achieved, controlling for orthogonal covariates can improve precision of the treatment effect estimator if the covariates can predict unexplained variance in the outcome.

By definition it is not possible to guarantee balance on unobserved covariates. As discussed in Appendix 1.9.1, we mechanically balanced the treatment arms on first midterm score, one of the few observables at the time of treatment assignment, with our knowledge from previous cohorts' data that the first midterm score explains a significant amount of variance in final exam score. Hence, in our estimation strategies including controls, we always include the first midterm score and year, following the recommendations of Bruhn and McKenzie (2009) to control for all covariates used to seek balance when assigning treatment.

For variables unobservable at time of randomization but observable at time of analysis, we lack the luxury of guaranteed balance by construction, nor is it clear *ex ante*, beyond our intuition, which will predict variation in the outcome variables of interest. On one hand, failing to control for valid predictors reduces statistical power. On the other hand, hand-picking control variables increases researcher degrees of freedom, risking increasing the prevalence of Type I errors (Simmons, Nelson, & Simonsohn, 2011). As such, in addition to a model without controls beyond the ones used for treatment assignment (year and midterm score), we fit a second model that includes a vector of linear controls chosen using the post-double-selection (PDS) procedure introduced by Belloni, Chernozhukov, and Hansen (2014b).

PDS is a two step process in which first, model covariates are selected in an automated, principled fashion, and second, the model coefficients of interest are estimated while controlling for those selected covariates. The first step involves predicting, separately, both the outcome of interest (e.g., videos watched) and treatment status using lasso regression, which shrinks coefficient estimates towards zero. Note that since treatment is randomly assigned, the lasso should shrink most, if not all, of the coefficients towards zero when predicting treatment status. Next, the researcher takes the union of all covariates with non-zero coefficients and includes these covariates as controls in her model. With her control variables selected, she can now estimate treatment effects with reduced bias relative to including controls with less empirical rationale.

In Table 1.13, we describe all covariates observable in our study. In Table 1.14, we describe the covariates selected as controls for estimating the effect of treatment on each outcome variable of interest. All models include either pair fixed effects or year and first midterm score as controls. To ensure these controls are "selected" by the PDS procedure, we partialed out these controls from the first step prediction models by residualizing both sides of the equation as described in Belloni, Chernozhukov, and Hansen (2014a).

1.9.2 LATE estimators using Neyman's repeated sampling approach

In this section we derive LATE estimators using the repeated sampling approach of Neyman (1923), which considers each pair as an independent, completely randomized experiment.

Similar to a Wald estimator, the point estimate of the LATE is the mean within-pair difference in outcome divided by the mean within-pair difference in videos:

$$\widehat{\gamma} = \frac{\overline{\Delta y}}{\overline{\Delta v}} = \frac{\frac{1}{J}\sum_{j=1}^{J} \Delta y_j}{\frac{1}{J}\sum_{j=1}^{J} \Delta v_j} = \frac{\overline{y_I} - \overline{y_C}}{\overline{v_I} - \overline{v_C}} \tag{1.10}$$

where y is the outcome of interest (grades) and v is the number of videos, both indexed by pair $j \in J$ and treatment status C or I for *Control* or *Incentive*, respectively.

We use the delta method to calculate the approximate standard error of $\widehat{\gamma}$. First, we define the following normally-distributed random variables:

$$\begin{aligned} Y &= \overline{y_I} - \overline{y_C} \sim \mathcal{N}(\mu_Y, \sigma_Y^2) \\ V &= \overline{v_I} - \overline{v_C} \sim \mathcal{N}(\mu_V, \sigma_V^2) \end{aligned} \tag{1.11}$$

Using a first-order Taylor expansion and letting $g() = \frac{Y}{V}$, we have:

$$\begin{aligned} \mathrm{Var}(g) &= \mathrm{E}[(g - \mathrm{E}(g))^2] \\ &\approx \mathrm{E}[(g(\theta) + (Y - \theta_Y)g'_Y(\theta) + (V - \theta_V)g'_V(\theta) - g(\theta))^2] \\ &= \mathrm{E}[(Y - \theta_Y)^2(g'_Y(\theta))^2 + (V - \theta_V)^2(g'_V(\theta))^2 + 2(Y - \theta_Y)(V - \theta_V)g'_Y(\theta)g'_V(\theta)] \\ &= \mathrm{Var}(Y)(g'_Y(\theta))^2 + \mathrm{Var}(V)(g'_V(\theta))^2 + 2\mathrm{Cov}(Y,V)g'_Y(\theta)g'_V(\theta) \end{aligned} \tag{1.12}$$

Expanding about $\theta = (\theta_Y, \theta_V) = (\mu_Y, \mu_V)$ and letting $g'_Y(\theta) = \mu_V^{-1}$ and $g'_V(\theta) = \frac{-\mu_Y}{\mu_V^{-2}}$:

$$\begin{aligned} \mathrm{Var}(g) &\approx \frac{1}{\mu_V^2}\mathrm{Var}(Y) + \frac{\mu_Y^2}{\mu_V^4}\mathrm{Var}(V) + 2\frac{-\mu_Y}{\mu_V^{-2}}\mathrm{Cov}(Y,V) \\ &= \frac{\mu_Y^2}{\mu_V^2}\left(\frac{\sigma_Y^2}{\mu_Y^2} + \frac{\sigma_V^2}{\mu_V^2} - 2\frac{\mathrm{Cov}(Y,V)}{\mu_Y \mu_V}\right) \end{aligned} \tag{1.13}$$

We use the following variance estimators of Y and V from Equation 1.5:

$$\begin{aligned} \widehat{\mathrm{Var}}(Y) &= \widehat{\sigma_Y^2} = \frac{1}{J(J-1)}\sum_{j=1}^{J}(\Delta y_j - \overline{\Delta y})^2 \\ \widehat{\mathrm{Var}}(V) &= \widehat{\sigma_V^2} = \frac{1}{J(J-1)}\sum_{j=1}^{J}(\Delta v_j - \overline{\Delta v})^2 \\ \widehat{\mathrm{Cov}}(Y,V) &= \widehat{\sigma_{YV}} = \frac{1}{J(J-1)}\sum_{j=1}^{J}(\Delta y_j - \overline{\Delta y})(\Delta v_j - \overline{\Delta v}) \end{aligned} \tag{1.14}$$

and the following estimators for the population means of Y and V:

$$\begin{aligned}\widehat{\mu_Y} &= \mathrm{E}(\mu_Y) = \overline{\Delta y} \\ \widehat{\mu_V} &= \mathrm{E}(\mu_V) = \overline{\Delta v}\end{aligned} \tag{1.15}$$

Substituting these variance and means estimators into the final step of 1.13, we arrive at the standard error estimator for $\widehat{\gamma}$:

$$\widehat{\sigma_\gamma} = \frac{\overline{\Delta y}}{\overline{\Delta v}} \sqrt{\frac{\widehat{\sigma_Y^2}}{\overline{\Delta y}^2} + \frac{\widehat{\sigma_V^2}}{\overline{\Delta v}^2} - 2\frac{\widehat{\sigma_{YV}}}{\overline{\Delta y}\,\overline{\Delta v}}} \tag{1.16}$$

Table 1.10: Baseline balance test, Midterm 2 sample

	All students			P-values	Matched pairs		P-values
Variable	Above Median	Control	Incentive	(3) - (2)	Control	Incentive	(5) - (4)
Midterm 1 score	2.048	0.116	0.037	0.398	0.139	0.131	0.933
	(0.025)	(0.063)	(0.068)		(0.065)	(0.066)	
Year = 2019	0.492	0.513	0.500	0.797	0.514	0.514	1.000
	(0.025)	(0.036)	(0.035)		(0.037)	(0.037)	
Cumulative GPA	3.445	2.944	2.948	0.965	2.942	2.992	0.487
	(0.029)	(0.043)	(0.058)		(0.045)	(0.056)	
No cum. GPA	0.230	0.368	0.332	0.452	0.365	0.320	0.377
	(0.021)	(0.035)	(0.033)		(0.036)	(0.035)	
Math quiz score	0.592	0.037	0.106	0.471	0.054	0.137	0.396
	(0.044)	(0.070)	(0.065)		(0.071)	(0.068)	
Tutoring visits	0.269	0.259	0.223	0.655	0.276	0.232	0.612
	(0.042)	(0.059)	(0.056)		(0.062)	(0.061)	
Videos watched	13.228	13.368	13.777	0.750	13.663	13.729	0.961
	(0.681)	(0.886)	(0.931)		(0.929)	(0.986)	
Videos, unique	9.746	9.689	10.188	0.554	9.845	10.116	0.760
	(0.431)	(0.580)	(0.611)		(0.606)	(0.644)	
Hours videos	1.690	1.782	1.825	0.818	1.827	1.804	0.906
	(0.093)	(0.127)	(0.135)		(0.133)	(0.142)	
Hours videos, unique	1.291	1.355	1.387	0.802	1.382	1.364	0.897
	(0.062)	(0.090)	(0.092)		(0.095)	(0.096)	
Asian	0.700	0.694	0.668	0.581	0.713	0.652	0.215
	(0.022)	(0.033)	(0.033)		(0.034)	(0.036)	
Latinx	0.060	0.135	0.158	0.506	0.133	0.166	0.377
	(0.012)	(0.025)	(0.026)		(0.025)	(0.028)	
White	0.151	0.114	0.124	0.765	0.105	0.138	0.336
	(0.018)	(0.023)	(0.023)		(0.023)	(0.026)	
Other ethnicity	0.089	0.057	0.050	0.741	0.050	0.044	0.804
	(0.014)	(0.017)	(0.015)		(0.016)	(0.015)	
Female	0.393	0.342	0.391	0.312	0.343	0.392	0.328
	(0.024)	(0.034)	(0.034)		(0.035)	(0.036)	
Male	0.592	0.653	0.604	0.316	0.652	0.602	0.329
	(0.024)	(0.034)	(0.034)		(0.036)	(0.036)	
Transfer	0.271	0.477	0.455	0.673	0.470	0.436	0.528
	(0.022)	(0.036)	(0.035)		(0.037)	(0.037)	
Observations	417	193	202		181	181	

Note: This table includes all students who completed the second midterm. Descriptions of each variable can be found in Table 1.13. *Male* and *Female* are coded zero for nine students who do not report a gender. *P-values* are reported for the Welch's t-test of equal means between the *Control* and *Incentive* arms. Standard errors in parentheses are robust to heteroskedasticity. ***, **, and * indicate significance at the 1, 5, and 10 percent critical levels, respectively.

Table 1.11: Baseline balance test, Final Exam sample

	All students			P-values	Matched pairs		P-values
Variable	Above Median	Control	Incentive	(3) - (2)	Control	Incentive	(5) - (4)
Midterm 1 score	2.049	0.153	0.057	0.291	0.177	0.170	0.938
	(0.025)	(0.061)	(0.069)		(0.064)	(0.065)	
Year = 2019	0.489	0.516	0.500	0.753	0.518	0.518	1.000
	(0.025)	(0.037)	(0.036)		(0.039)	(0.039)	
Cumulative GPA	3.445	2.946	2.959	0.864	2.929	3.001	0.346
	(0.029)	(0.044)	(0.060)		(0.047)	(0.059)	
No cum. GPA	0.231	0.359	0.332	0.583	0.367	0.313	0.299
	(0.021)	(0.035)	(0.034)		(0.038)	(0.036)	
Math quiz score	0.599	0.071	0.152	0.396	0.061	0.157	0.338
	(0.043)	(0.068)	(0.066)		(0.071)	(0.071)	
Tutoring visits	0.270	0.272	0.237	0.684	0.283	0.253	0.746
	(0.043)	(0.061)	(0.060)		(0.066)	(0.066)	
Videos watched	13.292	13.418	13.658	0.856	13.729	13.789	0.966
	(0.682)	(0.909)	(0.953)		(0.978)	(1.023)	
Videos, unique	9.793	9.783	10.111	0.704	9.795	10.181	0.674
	(0.432)	(0.598)	(0.622)		(0.630)	(0.665)	
Hours videos	1.698	1.788	1.805	0.929	1.812	1.803	0.967
	(0.094)	(0.130)	(0.138)		(0.138)	(0.148)	
Hours videos, unique	1.297	1.369	1.372	0.985	1.363	1.366	0.985
	(0.062)	(0.093)	(0.094)		(0.098)	(0.100)	
Asian	0.701	0.696	0.653	0.376	0.711	0.633	0.129
	(0.022)	(0.034)	(0.035)		(0.035)	(0.038)	
Latinx	0.060	0.141	0.158	0.654	0.139	0.169	0.448
	(0.012)	(0.026)	(0.027)		(0.027)	(0.029)	
White	0.149	0.109	0.132	0.497	0.102	0.145	0.244
	(0.018)	(0.023)	(0.025)		(0.024)	(0.027)	
Other ethnicity	0.089	0.054	0.058	0.882	0.048	0.054	0.804
	(0.014)	(0.017)	(0.017)		(0.017)	(0.018)	
Female	0.393	0.348	0.405	0.253	0.337	0.404	0.212
	(0.024)	(0.035)	(0.036)		(0.037)	(0.038)	
Male	0.593	0.647	0.584	0.215	0.657	0.584	0.176
	(0.024)	(0.035)	(0.036)		(0.037)	(0.038)	
Transfer	0.272	0.462	0.447	0.778	0.470	0.416	0.321
	(0.022)	(0.037)	(0.036)		(0.039)	(0.038)	
Observations	415	184	190		166	166	

Note: This table includes all students who completed the final exam. Descriptions of each variable can be found in Table 1.13. *Male* and *Female* are coded zero for nine students who do not report a gender. *P-values* are reported for the Welch's t-test of equal means between the *Control* and *Incentive* arms. Standard errors in parentheses are robust to heteroskedasticity. ***, **, and * indicate significance at the 1, 5, and 10 percent critical levels, respectively.

Table 1.12: Anderson-Rubin confidence sets

Outcome variable	Endogenous variable	Anderson-Rubin CI
Midterm 2 score	10 unique videos	[-0.001, 0.653]
Midterm 2 score	1 hour videos	[-0.001, 0.392]
Final exam score	10 unique videos	[0.000, 0.163]
Final exam score	1 hour videos	[0.000, 0.102]

Note: This table displays Anderson-Rubin confidence sets at the 95% confidence level for the 2SLS estimator $\hat{\gamma}$ from Equation 1.8 including year dummies and first midterm score as controls. Outcomes are measured in standard deviations. Instrumented endogenous variables are measured in 10s of unique videos or hours of unique content.

Table 1.13: Candidate control variables for post-double-selection

Variable	Description
Midterm 1 score	Score on the first midterm
Year = 2019	1 if course taken in 2019, 0 otherwise
Cumulative GPA	Cumulative GPA from prior term, 0 if not observed
No cum. GPA	1 if Cumulative GPA unobserved, 0 otherwise
Math quiz score	Score on a quiz assessing prerequisite math skills
Tutoring visits	Number of group tutoring lab visits as of the first midterm
Videos watched	Number unique videos watched as of the first midterm
Hours videos	Hours of unique videos watched as of the first midterm
Asian	1 if ethnicity is Asian, 0 otherwise
Latinx	1 if ethnicity is Latinx, 0 otherwise
White	1 if ethnicity is White, 0 otherwise
Female	1 if female, 0 otherwise
Transfer	1 if transfer student, 0 otherwise

Note: *Midterm 1 score* and *Math quiz score* are measured in control standard deviations. *Cumulative GPA* is measured on a 4.0 scale. Videos included in *Videos watched* and *Hours videos* are unique course-relevant videos. The ethnicity variables are coded by university records: *Asian* includes ”Chinese/Chinese American”, ”Vietnamese”, ”East Indian/Pakistani”, ”Japanese/Japanese American”, ”Korean/Korean American”, and ”All other Asian/Asian American”; *Latinx* includes ”Mexican/Mexican American”, ”Chicano”, and ”All other Spanish-American/Latino”; *White* includes ”White/Caucasian”; and the omitted category inludes ”African American/Black”, ”Pacific Islander”, and ”Not give/declined to state”.

Table 1.14: ITT model controls selected via post-double-selection

Table	Dependent Variable	Controls, All Observations	Controls, Fixed Effects
Table 1	Hours unique videos by Final	Hours videos Videos	Hours videos Videos
	Hours unique videos by Mid. 2	Hours videos	Hours videos Videos
	Hours videos by Final	Hours videos	Hours videos
	Hours videos by Mid. 2	Hours videos	Hours videos Tutoring visits Videos
	Num. unique videos before Final	Hours videos Videos	Videos
	Num. unique videos before Mid. 2	Hours videos Videos	Videos
	Num. videos before Final	Hours videos Videos	Hours videos Videos
	Num. videos before Mid. 2	Hours videos Videos	Hours videos Tutoring visits Videos
Table 2	Final exam score	None	Math quiz score Transfer
	Midterm 2 score	None	Math quiz score
Table 3	All classes	Cumulative GPA	Cumulative GPA Math quiz score Transfer
	Econ classes ex. Micro A	None	Cumulative GPA Transfer
	Excluding Micro A	Cumulative GPA	Transfer
	Excluding econ classes	None	None
	Letter grade in Micro A	Cumulative GPA Latinx Transfer	Cumulative GPA
	Num. classes not passed	None	None
	Num. classes passed	Cumulative GPA Transfer	Cumulative GPA Transfer
	Num. classes taken P/NP	Latinx	Latinx
	Num. classes taken for letter	Cumulative GPA No cum. GPA	Cumulative GPA
	Num. classes withdrawn	None	None
	Num. units taken P/NP	Latinx	Latinx
	Num. units taken for letter grade	Cumulative GPA No cum. GPA	Cumulative GPA
	Num. units withdrawn	None	None
	% classes taken P/NP	None	Latinx
	% classes taken for letter	None	Latinx
Table 4	Attendance checks	Female Math quiz score Tutoring visits	Tutoring visits

Continued on next page

Table 1.14 (continued)

	Discussion board answers	None	None
	Discussion board days online	None	None
	Discussion board questions asked	None	None
	Discussion board views	None	Asian
	Tutoring visits	Tutoring visits	Tutoring visits
Table 5	Hours of videos	Hours videos	Hours videos Latinx Math quiz score Tutoring visits Videos
	Midterm 1 score	None	Latinx Math quiz score Videos
	Midterm 2 score	None	Asian Latinx Math quiz score Videos
	Num. classes not passed	None	None
	Num. classes passed	None	None
	Num. classes taken P/NP	None	Transfer
	Num. classes taken for letter	None	No cum. GPA
	Num. classes withdrawn	None	None
	Num. of videos	Hours videos	Hours videos Latinx Math quiz score Tutoring visits Videos
	Num. units taken P/NP	None	Transfer
	Num. units taken for letter grade	None	None
	Num. units withdrawn	None	None
	Term GPA	Cumulative GPA	Cumulative GPA Tutoring visits
	Term GPA, econ courses ex. Micro B, winter	None	Math quiz score
	Term GPA, ex. Micro B	Cumulative GPA	Cumulative GPA Tutoring visits
	Term GPA, ex. econ courses	None	Tutoring visits
	Took Micro B	None	Math quiz score
	% classes taken P/NP	None	No cum. GPA Transfer
	% classes taken for letter	None	No cum. GPA Transfer
Table None	Final exam score	None	Latinx Math quiz score Videos
	Hours unique videos	Hours videos	Hours videos Latinx Math quiz score Tutoring visits Videos

Continued on next page

Table 1.14 (continued)

Num. unique videos	Hours videos	Hours videos Latinx Math quiz score Tutoring visits Videos
Pass Micro B	None	Latinx Math quiz score Videos

Note: Controls chosen via the PDS procedure of Belloni, Chernozhukov, and Hansen (2014b). In the *All Observations* model, *Midterm 1 score* and *Year = 2019* are additionally included as controls. In the *Fixed Effects* model, pair fixed effects and *Midterm 1 score* are included. All control variables are measured before the start of the experiment, e.g. *Hours videos* is the hours of videos watched as of the first midterm.

Table 1.15: LATE model controls selected via post-double-selection

Dependent Variable	Instrumented	Controls, All Observations	Controls, Fixed Effects
Final exam score	Hours videos, unique	Hours videos Math quiz score Transfer Videos	Hours videos Videos
Final exam score	Videos, unique	Hours videos Videos	Hours videos Videos
Midterm 2 score	Hours videos, unique	Hours videos Math quiz score Tutoring visits Videos	Hours videos
Midterm 2 score	Videos, unique	Hours videos Videos	Hours videos Videos

Note: Controls chosen via the PDS procedure of Belloni, Chernozhukov, and Hansen (2014b). In the *All Observations* model, *Midterm 1 score* and *Year = 2019* are additionally included as controls. In the *Fixed Effects* model, pair fixed effects and *Midterm 1 score* are included. All control variables are measured before the start of the experiment, e.g. *Hours videos* is the hours of videos watched as of the first midterm.

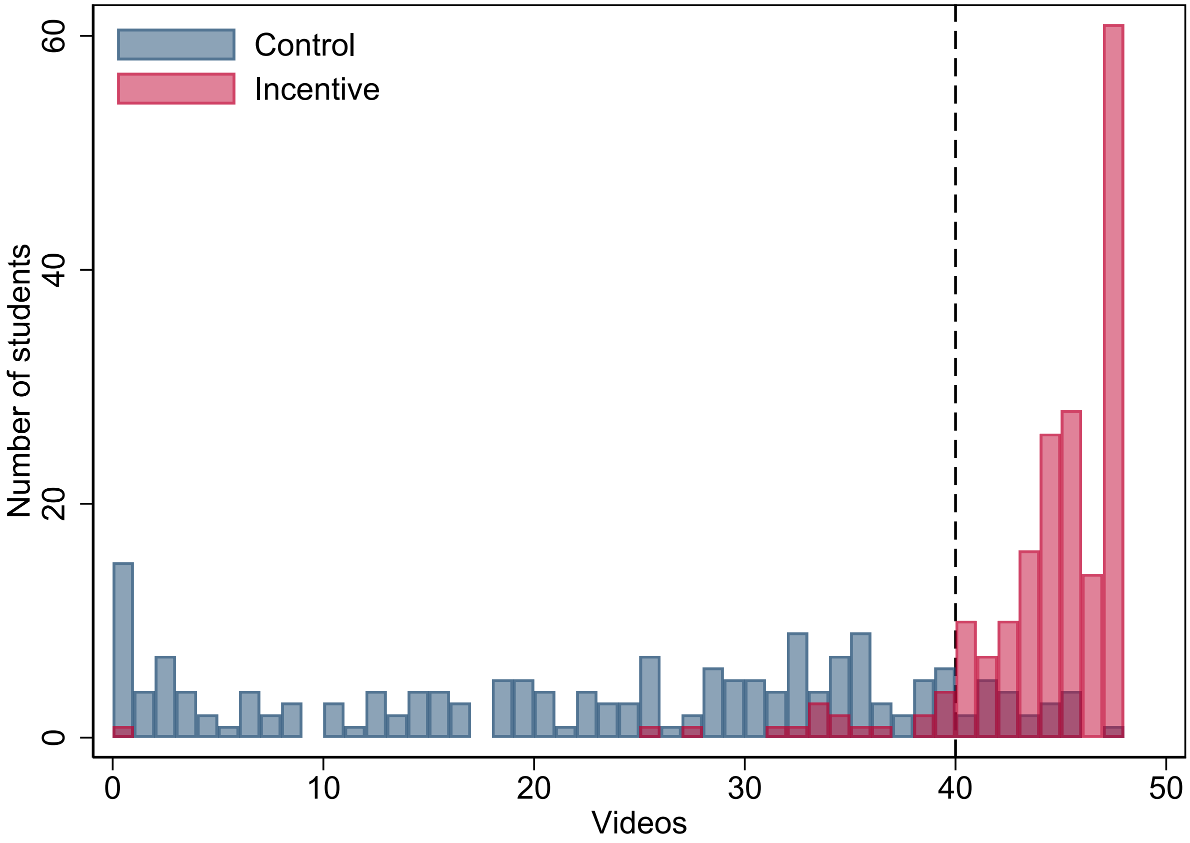

Figure 1.7: Distribution of videos counted towards incentive
This plot includes only videos that would have counted towards the earning the grade incentive. Students were required to watch 40 unique of 48 eligible videos between the first midterm and final exam to earn the grade incentive. 91% of *Incentive* students met the requirements for the grade incentive versus 11% of *Control* students.

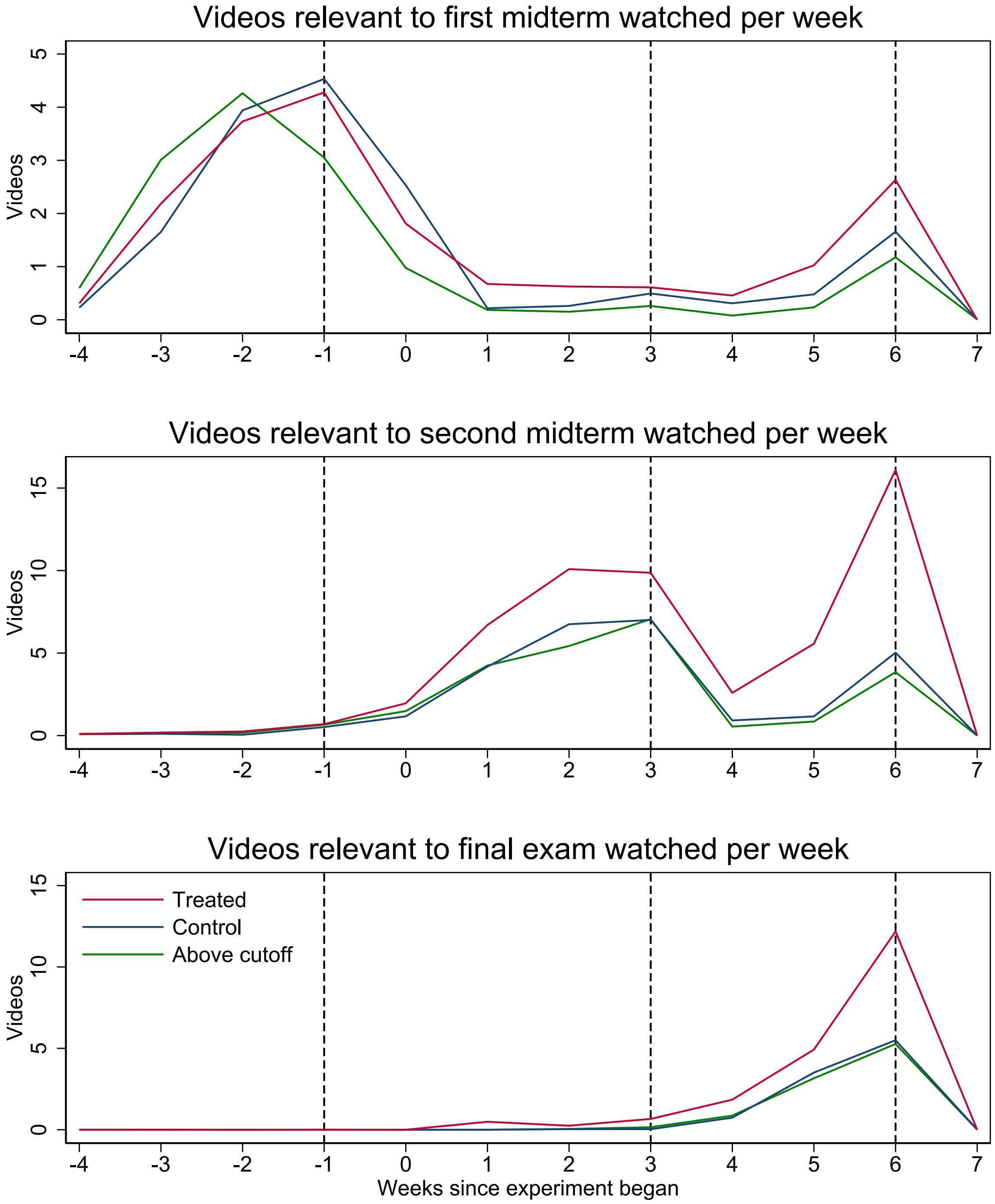

Figure 1.8: Weekly video watching by exam topic
Dashed lines represent Midterm 1, Midterm 2, and Final exams.

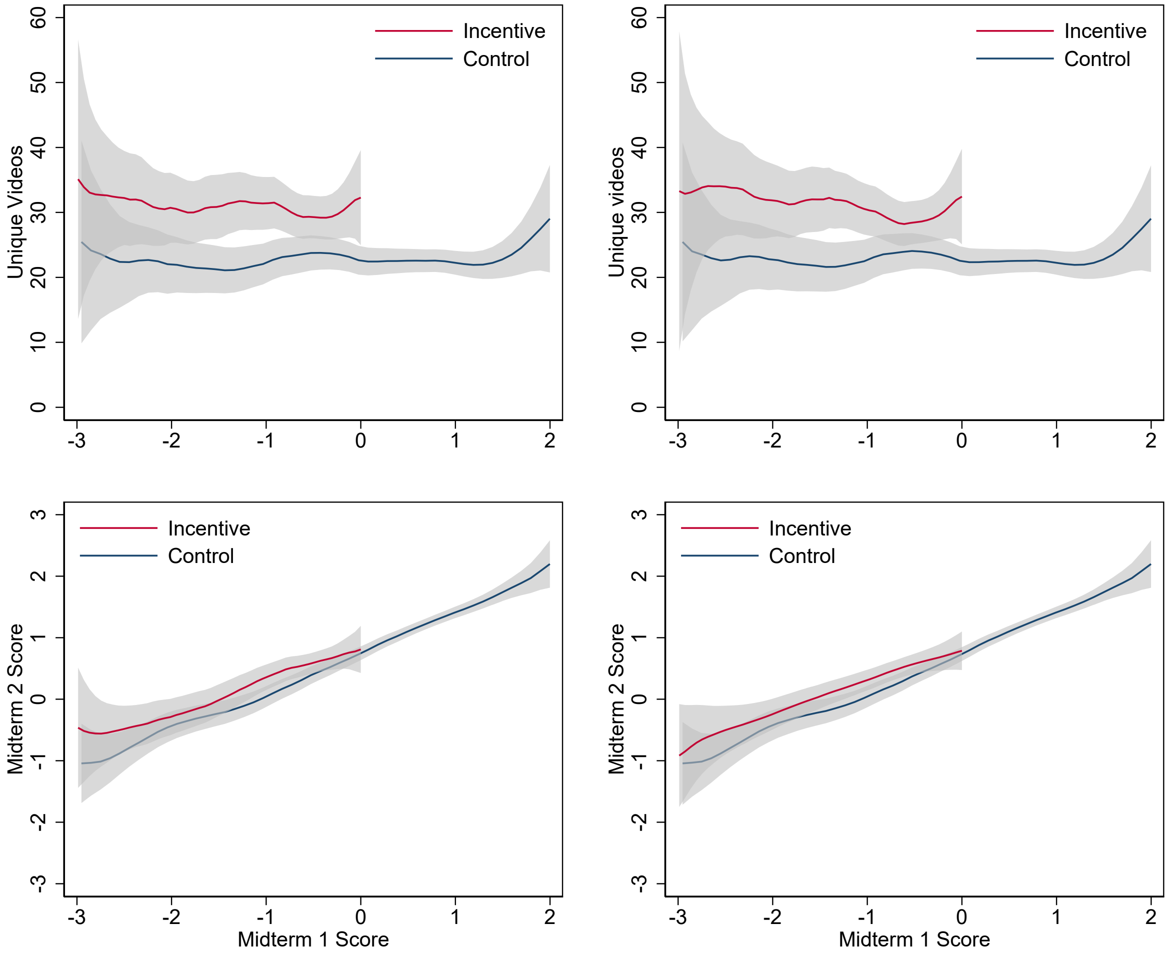

Figure 1.9: Effects of treatment along first midterm score, by midterm 2
Videos (top) includes unique videos watched before the second midterm exam. Exam scores (bottom) are measured in control standard deviations. Confidence bands represent 95% confidence intervals of the conditional mean outcome. The left plots includes all students who took the second midterm while the right plots exclude any students whose matched pair attrited.

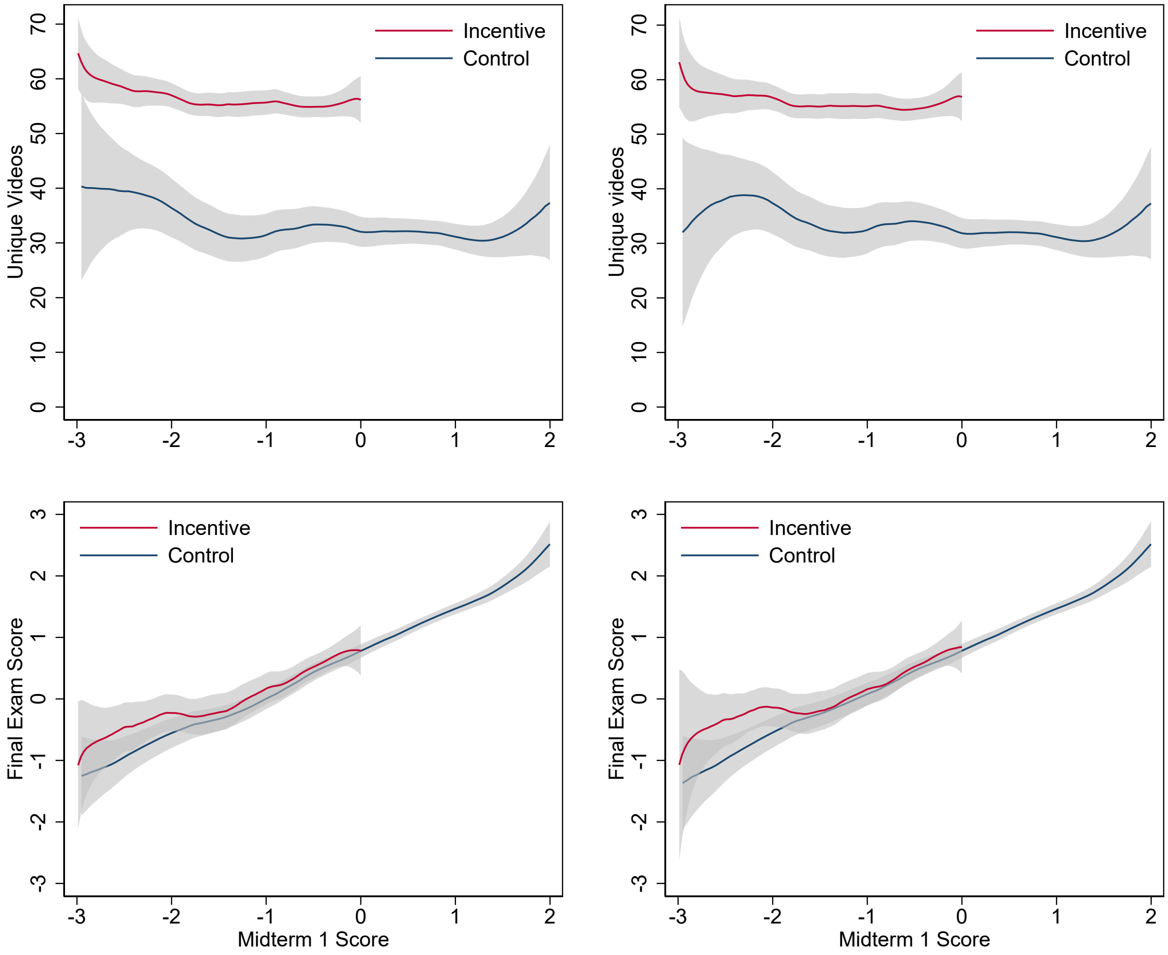

Figure 1.10: Effects of treatment along first midterm score, by final
Videos (top) includes unique videos watched before the final exam. Exam scores (bottom) are measured in control standard deviations. Confidence bands represent 95% confidence intervals of the conditional mean outcome. The left plots includes all students who took the final exam while the right plots exclude any students whose matched pair attrited.

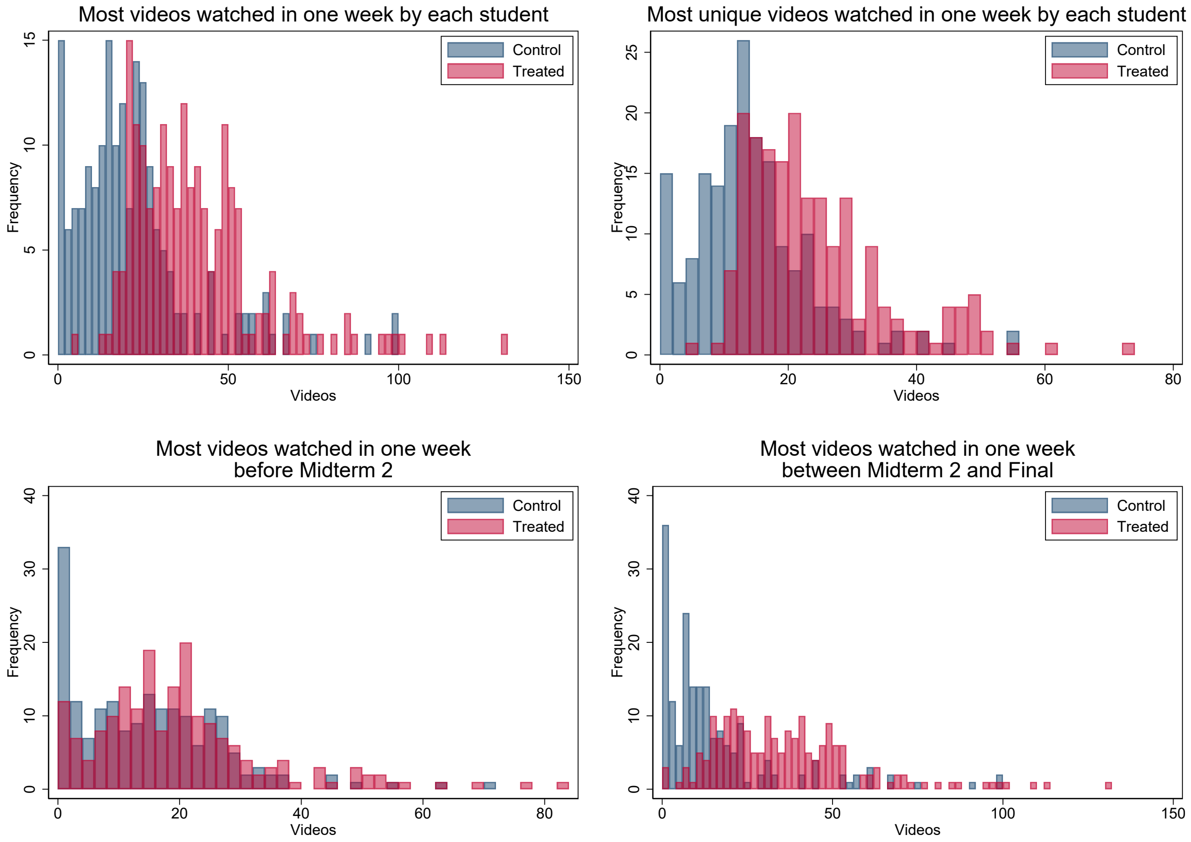

Figure 1.11: Distribution of max videos watched in one week
These plots help illustrate potential "binge watching" behavior. Compared to the *Control* students, *Incentive* students are more likely to watch 40 or more unique videos in a week, which occurs in the weeks preceding the final and not the second midterm.

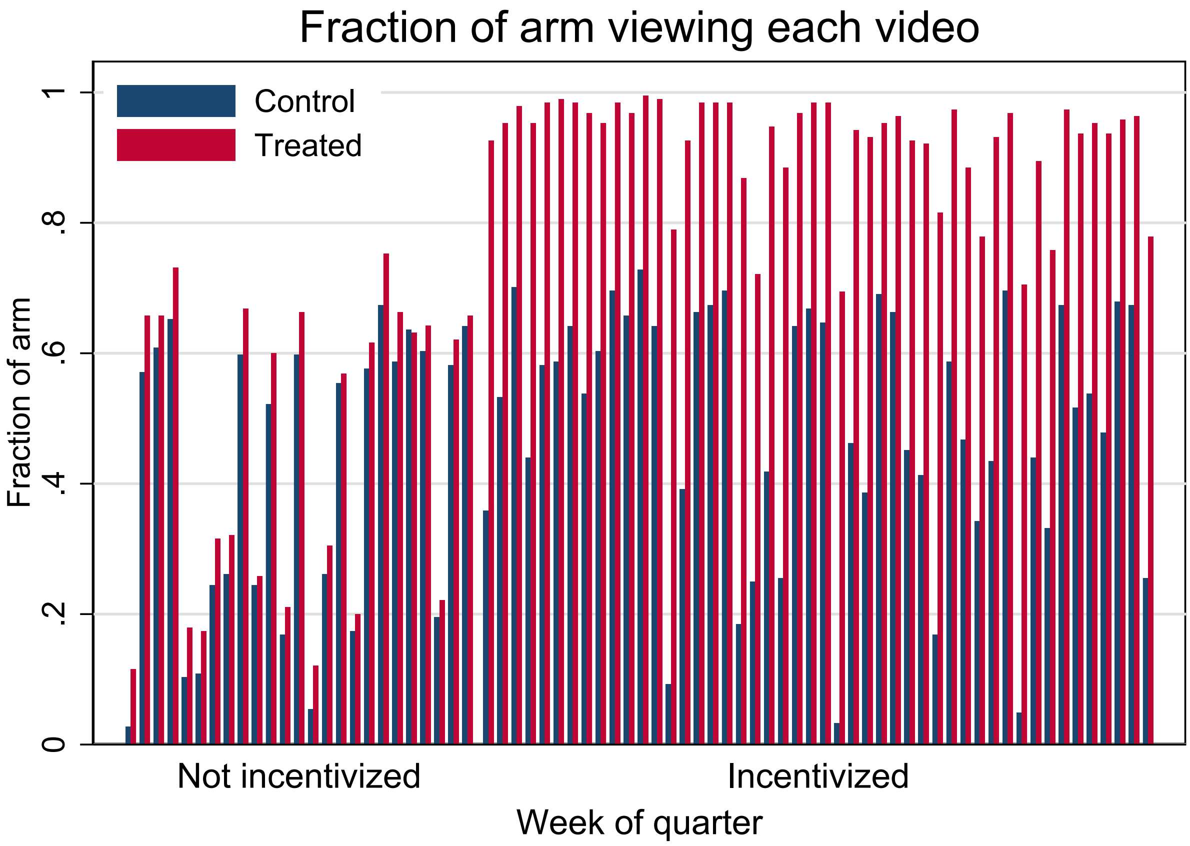

Figure 1.12: Video watch rates by video and treatment arm, grouped by incentive, ordered by video duration

Each bar represents the fraction of the treatment arm that watched a particular video. Bars are in order of video duration separately for incentivized and non-incentivized videos.

Chapter 2

The Effect of the Cook County, Illinois Sweetened Beverage Tax on Sugar Demand

2.1 Introduction

Obesity remains among the costliest health issues in the United States today. Finkelstein et al. (2009) find that obesity-related health expenses cost the U.S. between $107 billion to $184 billion in 2008 in 2018 dollars. Cawley and Meyerhoefer (2012) estimate significantly larger aggregate expenses totaling $261 billion using an IV approach. As a potential solution to reducing the high medical costs of obesity, government agencies have relatively recently suggested reducing added sugar consumption (USDA, 2015). Sugar sweetened beverages (SSBs) are the source of nearly 40% of the average American's consumption of added sugar (USDA, 2015) and about 7% of caloric intake (Allcott, Lockwood, & Taubinsky, 2019b). Several policymakers have recommended imposing taxes on sweetened beverages to reduce sugar consumption and associated health costs, a strategy that has worked for other harmful-to-health products like tobacco (Chaloupka, Yurekli, & Fong, 2012).

While excise taxes on soft drink taxes have existed for many years in the U.S. and other countries, beverage taxes levied at the *ounce* level have been passed in eight U.S. localities starting with Berekely, California in 2015.[1] In this paper we focus on the largest of these localities, Cook County, Illinois, which enacted and revoked a one-cent-per-ounce tax on sweetened beverages in 2017. As the only sweetened beverage tax revoked in the United States as of this writing, Cook County's tax is uniquely interesting for studying the effects of short-term taxes on long-term behavioral change. Additionally, Cook County's tax affected the largest population among all beverage taxes in the U.S., allowing for increased precision of effect estimators.

While other researchers have estimated the effects of beverage taxes on beverage demand and sugar from beverages, few have examined the impact on the policy-relevant metric of total sugar consumption. It is important to consider the complete basket of foods consumed when estimating the effects of taxes on nutrients: examining only the taxed product class in isolation may overstate the effects of the tax if consumers substitute towards other food product classes (for example if

[1] For expositional ease, *beverage tax* hereafter refers to the class of beverage taxes levied at the ounce level.

consumers substitute from taxed soda to untaxed cookies). On the other hand, if the tax carries a signal that causes consumers to update their priors about the healthfulness of a particular product attribute, then the estimated effects could be *understated*.

Unfortunately, it is challenging to reliably observe nutrient consumption across all foods over an extended period of time. Common methods, such as surveys and food diaries, suffer from small sample sizes, limited time frames, and high measurement error from poor recall and incorrect labeling. To overcome this limitation, we use a large representative panel dataset with barcode-level information about the products *purchased* in a household. We then match these data with barcode-level nutrition panel data to observe nutrients purchased over time within a household. While nutrient consumption and nutrient purchases are not equivalent, we treat purchases as a proxy for consumption.

Our primary outcome measure of interest is grams of sugar purchased, both during the tax and shortly after the tax is removed. We exploit variation across time and geography in a differences-in-differences specification to conduct our analysis. Specifically, we use households residing in the Chicago Designated Market Area (DMA)[2] as our primary counterfactual for households living inside Cook County. We construct a second counterfactual composed of the top ten largest counties by population (other than Cook County). While the first counterfactual shares region-specific shocks with Cook County, the second counterfactual is less likely to suffer from spillover effects and other violations of the Stable Unit Treatment Value Assumption (SUTVA).

We find that on average, the Cook County beverage tax reduced sugar purchases from all foods by 14.9% compared to those households outside the border. The results are similar when using large counties as the counterfactual, albeit somewhat smaller magnitudes. We find no evidence that the tax caused reductions in sugar purchases that persisted after the tax was removed, nor do we find an increase beyond pretax levels of sugar. Similar to other authors, we find that the taxes decreased purchases of taxed beverages across both extensive and intensive margins. However, unlike prior work, we confirm that the reduction in sugar purchases is explained almost entirely

[2] Nielsen's DMAs are similar to the Census Bureau's Metropolitan Statistical Areas.

by taxed products, suggesting that substitution towards untaxed products is likely modest for most households.

We examine heterogeneity across several dimensions. First, we find that reductions in sugar are largest for the poorest households in our sample, and we find no differences in reductions across race or education dimensions. Second, we estimate the effects of the tax as a function of distance from a household's zip code to the Cook County border, and we find smaller and statistically insignificant effects for households on the border and larger and significant effects for households residing further inside the border. Finally, we estimate the effects of the tax as a function of pretax levels of soda purchases. While the tax had no statistically significant effect on sugar purchases for low-soda households, the tax had large and significant effects on the top three deciles of households by pretax soda purchases.

Finally, we compare the SSB taxes possible benefits of reduced sugar consumption with its main cost: an increase in consumer prices. Specifically, we develop a simple household model where the household has CES preferences over a range of products, some of which are taxed. We estimate the households implicit price index before and after the tax in Cook County and our counterfactual (Chicago DMA sans Cook County) and find that the tax cost the representative household between \$7.63 and \$24.21 during the four month period of the tax. While the tax decreased sugar consumption by a significant amount, the reduction came at the cost of approximately 2.1 to 6.6 cents per gram.

The remainder of the paper is organized as follows. Section 2.2 provides a background of the policies enacted and related literature. Sections 2.4 and 2.3 introduce the methods and data used, respectively. Section 2.5 summarizes our results. Section 2.6 concludes with a discussion of those results.

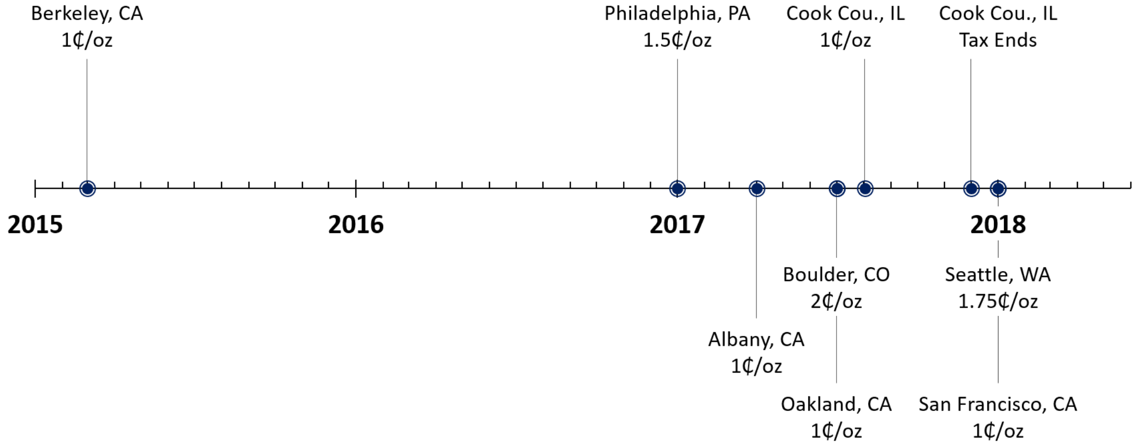

Figure 2.1: Rollout of beverage taxes in the United States

2.2 Background and Related Literature

In this section, we provide background on beverage taxes and summarize existing related literature.

2.2.1 Beverage taxes in the United States

While dozens of cities have considered taxing beverages, the first city to enact a beverage tax was Berkeley, California in March 2015. Berkeley's Measure D levied a one-cent-per-ounce excise tax on the distribution of beverages with added sugar. Alcohol, milk products, beverages for medical use, 100% fruit and vegetable juices, water, and artificially sweetened beverages are all exempt from the tax. Nearly two years after Berkeley, other cities began enacting similar SSB taxes including Philadelphia, PA, (2017); Oakland, CA, (2017); Albany, CA (2017); Boulder, CO, (2017); San Francisco, CA, (2018); Seattle, WA, (2018). Cook County became the first county to enact a county-wide beverage tax on August 2, 2017. Two months later on December 1, the county became the first locality in the United States to remove an SSB tax. One can observe a timeline of SSB tax events in Figure 2.1.

The Berkeley tax is the longest standing SSB tax in the U.S. and the most studied for its impact on consumption. Falbe et al. (2015) randomly surveyed retail establishments in San Francisco, Oakland, and Berkeley both before and after the tax change and found that the one-cent-per-ounce tax led to 0.46-0.67 cent increase in the price per ounce, depending on the type of beverage. In a follow-up study, Falbe et al. (2016) used a repeated cross-sectional survey of low-income and minority households before and after the tax. Participants were asked how often they consume different types of beverages each week or month, and total SSB consumption was inferred from their responses. The authors found that SSB consumption decreased by 21% in Berkeley while it increased by 4% in comparison cities. L. D. Silver et al. (2017) used store surveys and point-of-sale scanner data for three Berkeley and six non-Berkeley supermarkets to assess the tax pass through. They found that the tax pass-through was near 100% for large chain supermarkets and smaller for others. They also found that sales of SSBs decreased in Berkeley relative to other cities. Cawley and Frisvold (2017) add to the literature that stores in Berkeley pass through more of the tax to consumers the further away they are from borders with neighboring untaxed localities, ceteris paribus.

Each of the aforementioned studies find evidence that the tax in Berkeley resulted in high, but not 100%, pass-through of the tax to consumers, which reduced SSB consumption. From these findings we learn that consumers are not perfectly elastic in their demand for SSBs, but we cannot say whether the taxes reduced sugar consumption because the authors did not examine demand interrelationships with other goods that contain sugar. While about 39% of the average American's daily sugar consumption (USDA, 2015) and about 7% of daily calories (Allcott, Lockwood, & Taubinsky, 2019b) come from beverages that are taxable under the examined policies, it would be a problem for policymakers if, for example, taxed consumers substituted away from soda towards chocolate milk or candy. Most of these studies also suffer from only sampling consumers in one region and failing to track where consumers purchase SSBs, which may have changed following the tax. Additionally, these studies rely on survey data, which likely contain significant response bias. Given the increased attention of the potential negative health effects of SSBs and respondents' tendencies to provide prosocial answers to surveyors[3], it would not be surprising to find that respondents underreport their SSB consumption to a larger degree following implementation of the tax.

More recently, other authors have examined the SSB taxes in Boulder (Cawley et al., 2021), Oakland (Cawley et al., 2020b), and Philadelphia (Cawley et al., 2020a). In Boulder, Cawley and coauthors find that nearly 80% of the two-cents-per-ounce tax in Boulder was passed on to the consumer. Importantly, they note that about one-fifth of the retailers surveyed added the SSB tax at the register rather than into the shelf-price, which may have reduced saliency of the tax, perhaps lowering its effectiveness (Chetty, Looney, & Kroft, 2009). In Oakland, the authors find that the one-cent-per-ounce tax was passed through at a rate of about 60% and did not statistically significantly decrease sweetened beverages purchased per trip or added sugar consumed from beverages. In Philadelphia, the authors find that consumers purchased 8.9 ounces of SSBs per shopping trip less from stores within the city relative to consumers who purchased SSBs from stores outside the city. They also find evidence that city residents increased their purchases of SSBs outside the city. Cawley, Frisvold, and Jones (2020) estimate the effects of SSB taxes in four cities (Oakland, Philadelphia, San Francisco, and Seattle) on consumer purchases using household receipt data from InfoScout. They find that a one-cent-per-ounce increase in tax rate reduces purchases of taxed products by 12.2%, driven primarily by panelists in Philadelphia.

2.2.2 Cook County's beverage tax

Here we review the history of the Cook County beverage tax and provide specific details about timing and implementation.

The Cook County Board of Commissioners passed the Cook County Sweetened Beverage Tax Ordinance with a 9-8 vote on November 10, 2016. Initially, the one-cent-per-ounce tax was slated to go into effect on July 1, 2017; however, four days earlier on June 27, the Illinois Retail

[3] Researchers have observed that people tend to overreport socially positive behaviors like voting (B. D. Silver, Anderson, & Abramson, 1986) and attending church (Hadaway, Marler, & Chaves, 1993) but underreport socially negative behaviors like declaring bankruptcy (Locander, Sudman, & Bradburn, 1976) and using illegal drugs (Mensch & Kandel, 1988). See Krumpal (2013) for a review of the literature on "Social Desirability Bias."

Merchants Association filed suit, questioning the constitutionality of the tax. A state judge issued a temporary restraining order on the new tax until the case could be reviewed. On July 28, the same judge dismissed the lawsuit, and shortly thereafter on August 2, 2017, the beverage tax went into effect. As Cook County is home to 5.2 million residents, this tax became the most widespread beverage tax in the United States.

The Cook County beverage tax is unique among the other beverage taxes in the U.S. Besides having impacted the largest U.S. population to date, the law required that the tax be "borne by the purchaser of the sweetened beverage," ("Sweetened Beverage Tax Ordinance 16-5931", 2016). In practice, this meant that although the tax was collected by distributors, the law required that the tax be added to the price faced by consumers, usually in the form of a line-item on the receipt. This element of the policy has important implications for pass through and salience. Several large chain retailers added to the price tags of beverages a disclaimer that the tax would be added at the register (see Appendix 2.13 for an example.) However, it is possible that many retailers added the tax at checkout without including the tax in the displayed price, which may have reduced salience of the tax and possibly the impact on consumption (Chetty, Looney, & Kroft, 2009).

Another difference between the Cook County tax and beverage taxes in other localities is the taxation of both regular and diet beverages. Besides Philadelphia, Pennsylvania, Cook County is the only other locality in the US to have included diet beverages in its beverage tax. Since diet beverages are close substitutes for regular beverages but contain no sugar, taxing them may be counterproductive if the goal is reduced sugar consumption. On the other hand, if increasing tax revenue is part of the objective function, then taxing diet beverages may be a dominant policy strategy.

To grasp the magnitude of the tax in Cook County, consider some popular beverages from our data. The most popular (in gross sales) carbonated beverage in our data for Cook County consumers is a 12-pack of 12-ounce cans that retails for $3.57 on average. The Cook County beverage tax adds $1.44, or 40%, to the price of this 12-pack on average. The most popular two-liter carbonated beverage retails for $1.42 on average, and the tax adds 48% to the price of this 67.6-ounce product.

Across all carbonated beverages purchased in our dataset, we estimate the average percent tax to be about 23.5%.[4]

It is worth noting that Chicago levies a 3% tax on non-fountain soft drinks (added at the register) as well as a 9% tax on wholesale fountain drink syrups (paid by distributor) since 1994. Beverages in Chicago are also subject to an Illinois state sales tax of 6.25%, Cook County sales tax of 1.75%, Chicago sales tax of 1.25%, and a Regional Transit Authority tax of 1%. The beverage tax studied in this paper is in addition to these aforementioned taxes, which remain constant during the period of study.

Importantly, sweetened beverages purchased using SNAP benefits were exempt from the Cook County beverage tax since federal law prohibits taxation of products purchased using SNAP. All other localities in the US that have enacted beverage taxes levied them on distributors and did not require that the tax be added at the register. Hence, the passed-through portion of the taxes in these localities could be paid for using SNAP benefits. Given this difference, it is plausible that effects of the beverage tax in Cook County could have been smaller for the poorest households relative to effects of similar taxes in other localities. Another element to consider is that poor households most likely to receive SNAP benefits are also those most likely to face the nutrition informational barriers that drive internalities from SSB consumption (Allcott, Lockwood, & Taubinsky, 2019b). If so, then exempting SNAP benefits from the tax may actually hurt the poorest consumers.

On October 11, 2017, the Cook County Board of Commissioners voted 15-2 to repeal the ordinance, becoming the first governing body in U.S. history to revoke a beverage tax. The one-cent-per-ounce tax expired on December 1, 2017.

2.2.3 Theoretical rational for beverage taxes

Medical researchers have presented evidence that excess sugar consumption is strongly associated with metabolic-related diseases such as obesity and type 2 diabetes, dental carries, and

[4] We only include carbonated beverages in this calculation since nearly all carbonated beverages are taxed whereas other beverage categories include both taxed and untaxed products (e.g. 100% fruit juices are exempt but share a category with fruit-flavored drinks).

cognitive decline (Imamura et al., 2015; Malik & Hu, 2019; Malik et al., 2010). Additionally, several randomized trials have demonstrated a causal link between reduced SSB consumption and improved health (de Ruyter et al., 2012; Ebbeling et al., 2012; Ebbeling et al., 2006; Vartanian, Schwartz, & Brownell, 2007). Wang et al. (2012) and Long et al. (2015) both estimate that each ounce of SSB consumed carries with it approximately one cent of health system cost. If the health system collects these costs from consumers of health care, then there would not be externalities from harmful effects of SSBs on health. On the other hand, if these costs are not covered by consumers themselves, then taxes on beverages could help pass the social costs of SSB consumption to the consumer.

Cawley and Meyerhoefer (2012) find that about 88% of obesity-related medical expenses are paid for by parties other than the consumer of the health care. Taking this estimate along with those of Wang et al. (2012) and Long et al. (2015), Allcott, Lockwood, and Taubinsky (2019a) estimate that the external cost of SSB consumption is in the range of 0.8 to 0.9 cents per ounce.

If one takes the perspective of a social planner seeking to maximize long-term welfare, as Allcott, Lockwood, and Taubinsky (2019a), then consumers may face *internalities* from their own myopia and overconsumption of SSBs, which carry costs in the hyperbolically-discounted future. Additionally, information frictions may impose internalities; that is, consumers might consume fewer SSBs if they have the same knowledge of the health effects as do dietitians and nutritionists. Allcott and coauthors estimate that, in dollar terms, the marginal internality from SSB consumption is 0.9 to 2.1 cents per ounce. In sum, the authors suggest that the optimal federal beverage tax is between 1 and 2.1 cents per ounce, or around 0.4 cents per ounce if only considering externalities. Note that this “optimal tax” is reduced from the sum of the externalities and internalities by taking into account the regressive financial cost of the tax, which reduces the optimal tax by about 0.5 cents per ounce. For city-level taxes, the authors suggest that the optimal tax may range 0.5 to 1 cent per ounce to account for 50% and 25% cross-border shopping, respectively.

2.3 Data

We identify the effects of SSB taxes by linking several data sources. First, we observe household-level food purchase and demographic data from the Nielsen Homescan Survey, a nationally representative panel of 40-60,000 households from 2004 to 2016. These data have been used to asses the effects of other taxes. Harding, Leibtag, and Lovenheim (2012) use the Homescan data to asses the impact of cigarette taxes, especially for panelists living close to state borders. Cotti, Nesson, and Tefft (2016) look at how tobacco taxes affect smoking cessation product purchases. Dharmasena and Capps Jr (2012) assess the effects of a hypothetical tax on sugar. Colchero et al. (2016) and Colchero et al. (2017) use Nielsen panel data in Mexico to evaluate the impact of a one-peso per liter tax on SSBs.

Nielsen provides recruited households with a handheld barcode scanner and asks them to record every item they purchase. By scanning their purchased products, households earn points that can be redeemed in a rewards catalog much like credit card reward programs. The households can earn higher value rewards the longer they remain in the panel, so households are additionally incentivized to participate in the panel for multiple years (the median household participation duration is about 7 years). Nielsen collects household characteristics including zip code level geographic identifiers, household size, age of household members, race, income range, and more. In Table 2.1, we present descriptive statistics of households in the Chicago DMA sample by treatment status.

Participants are asked to scan the UPC codes of *all* the products they buy following shopping trips and provide information on the location of the purchase. If the purchase was completed at a Nielsen partner retail outlet, the price of the purchased item is automatically recorded as the average price of that product during the week that the customer recorded the purchase. If the purchase was made at a non-participating retail outlet, the participant is asked to manually input the price. If the product does not have a barcode as is the case with raw fruits and vegetables, households can record their purchase by scanning a barcode for the corresponding item printed in a provided booklet.

Given that households may forget or choose to omit recording some of their purchases, the Nielsen Homescan Survey data likely contain measurement error. However, provided the measurement error is uncorrelated with introduction of SSB taxes, our findings, presented as percent changes, should remain unbiased. Levels of consumption presented in the data should be considered underestimates given the high likelihood of under-reporting by households (Einav, Leibtag, & Nevo, 2010).

We match the UPC codes from purchases in the Nielsen data with product-level nutrition data, thereby permitting us to observe total quantities of nutrients purchased by each household per month. Our nutrition data come from three sources. First, we use proprietary nutrition data from Syndigo, who provide UPC-level product information including the Nutrition Facts (amount of each micro and macronutrient listed), ingredients, product size, description, and brand for over 220,000 branded packages.[5] We use an imputation procedure, described next, to match similar products that do not have a direct UPC match. Second, we use the US Department of Agriculture's FoodData Central to gather nutrition information for as many of the remaining products as possible. Finally, for all remaining products, we use the search features of major online retailers to find images of the Nutrition Facts panel and manually record the information.

To match nutrition data to the Nielsen product data, we follow a procedure similar to that of Dubois, Griffith, and Nevo (2014). We first drop non-food products, alcohol, tobacco, weight-loss/diet aids, and reference card goods.[6] We then merge on UPC code, which matches about 52.8% of the purchased UPCs. Next, for products without a direct match, we match within product (string description of the product), brand, product module,[7], size type (measured in counts versus milliliters versus grams), flavor, variety, type, formula, and style. This step adds 23.8% to the matched data. Next, we relax the brand requirement, which permits matching of store-brand goods

[5] The data are licensed by over 2,000 consumer applications and have been used to estimate nutrients purchased by Nielsen panelists by other authors including Dubois, Griffith, and Nevo (2014).

[6] Reference card goods are products without a barcode that panelists can report they purchased by scanning barcodes listed in a barcode reference booklet. These reference barcodes are associated with pictures and standardized item descriptions (such as "bakery item"). We drop these products since we cannot infer accurate nutrition information, and reference card goods are vastly underreported relative to products that have barcodes. Nielsen provides alternative panelist weights for analyses that include reference card goods.

[7] over 1,000 product categories defined by Nielsen

and adds 11.2% of purchased products.[8] Next we relax the flavor, variety, type, formula, and style requirements, which allow for another 7.6% of products to be matched. For 4.0% of products, we impute within product module for those product modules that have sufficient matched observations. For the remaining product modules, or 0.5% of purchased products mostly comprised of fresh produce, we manually look up and enter nutrition information.

To reduce the influence of outliers, we topcode our aggregated nutrient measures at the 99th percentile for all values greater than the 99th percentile.

2.4 Methods

In this section, we describe the empirical strategies used in this paper.

2.4.1 Effect on food and nutrient purchases

We use a differences-in-differences specification to examine the impact of the beverage tax on food and nutrient purchases. We compare the change in monthly purchases during the taxed period by households living in Cook County, Illinois to that among counterfactual households. Our specification is as follows:

$$y_{it} = \beta_1 \tau_{s(i)t} + \gamma_i + \delta_t + \varepsilon_{it} \tag{2.1}$$

where y_{it} is the total purchases of a food or nutrient y by household i during month t, γ_i and δ_t are household and time fixed effects, respectively, and ε_{it} is an unobserved model residual. $\tau_{s(i)t}$ is the beverage tax in a locality-month with units cents-per-ounce where $s(i)$ is the locality where household i resides. Under the parallel trends assumption, unconfoundedness, and SUTVA, the coefficient of interest, β_1, is interpreted as the average causal effect of the tax on monthly purchases

[8] The Syndigo data do not include store-brand goods.

of y during the observed taxed period. In practice, we will take the inverse hyperbolic sine of y such that our results can be interpreted as percent changes.

Choosing a counterfactual that satisfies all of the assumptions aforementioned is challenging. In our preferred specification, we use Chicago DMA households that reside outside of the Cook County border as the counterfactual. While it is reasonable that households in the same DMA likely experience similar region-specific shocks thereby increasing the plausibility of the parallel trends and unconfoundedness assumptions, SUTVA is potentially less reliable. For example, it is conceivable that control households may shop at grocery stores inside Cook County, or stores outside the border offer promotions or increase marketing to attract additional business during the taxed period. Hence, we additionally estimate our specifications using households who reside in the top ten largest U.S. counties by population, other than Cook County and other counties that pass or enact a beverage tax during the period studied. While it is less plausible that households in distant counties experience similar regional shocks, it is reasonable to expect that these households are insulated from spillovers from the Cook County beverage tax.

To examine heterogeneous treatment effects, we add an interaction term to Equation 2.1:

$$y_{it} = \beta_1 \tau_{s(i)t} + \beta_2 \mathrm{I}_i \tau_{s(i)t} + \gamma_i + \delta_t + \varepsilon_{it} \tag{2.2}$$

where I_i is a household-specific, time-invariant covariate of interest. Specifically, we add interaction terms with key demographic variables, such as household income, to identify if the tax may have had differential effects across these demographic variables. Household income is of particular interest, besides for examining the potential regressivity of the tax, since beverages purchased with SNAP benefits were exempted from the tax.[9]

We also estimate Equation 2.2 with I_i as percentiles of pretax levels of regular soda purchases,

[9] We cannot directly observe whether SNAP benefits were used to pay for products, but we can observe household income ranges as well as if an alternative to cash and credit was used. Unfortunately, fewer than 100 households in our treatment sample can be classified as SNAP-eligible households, and an even smaller subset of these use cash and credit alternatives. Hence, we focus our analysis examining heterogeneity along income ranges.

scaled by the inverse of total monthly expenditures. In this specification, β_2 identifies the difference in treatment effects for high soda purchasing households to those that do not purchase soda. Low soda-purchasing households in Cook County can be thought of as an additional control group since they face less exposure to the tax yet experience the same county-specific shocks that high soda-consuming households experience.[10]

Finally, given the potential for Cook County residents to purchase beverages across the border, we estimate Equation 2.2 setting I_i as indicators for the household's "layer" within the county, thereby helping us understand whether the tax had larger effects for those further from the border. We code all households in our sample using their residential zip code to be one of five categories: outside border, cross border, on border, and two inside-border categories depending on linear distance to the nearest untaxed zip code. We depict these layers in Figure 2.3. In our analysis, we drop the 50 cross-border households given their small numbers and unclear taxed status.[11] We omit an indicator for outside-border households, and hence all coefficients are interpreted as the average change in purchases during the tax for a given layer relative to outside-border households.

When estimating Equations 2.1 and 2.2, we cluster our standard errors at the zip code level to allow for arbitrary serial correlation of residuals among households in the same zip code over time. While the Cook County beverage tax was levied at the county level, we are unable to reliably cluster our errors at such fine a level since our sample is comprised of too few counties. Additionally, we weight households in our sample using a Nielsen-provided projection factor, which helps make the panel more representative of the general U.S. population.[12] Finally, for months where a household has zero store trips and purchases, we code observations as missing. To check that our results are not substantially biased by potentially nonrandom missing purchase data, we estimate our specifications restricted to a subsample of panelists that have at least one purchase per month.

[10] Regular soda was not the only beverage taxed, so these households may have still experienced some exposure to the tax.

[11] Including an indicator variable for this group in our specifications has little influence on the results, and we do not have ample precision to draw any conclusions from this group.

[12] Uniformly weighting each household has a small but not substantive impact on the results.

2.4.2 Welfare

While the above methods help us understand whether the tax may reduce sugar consumption and improve health, they do not capture the implicit costs to households in terms of an increase in their price index. In order to examine the welfare costs of the tax, we develop a simple household model where households derive utility from grocery consumption and from their health. Some of the health effects of sugar are non-internalized, which induces the policy maker to impose a tax. The household's utility is given as:

$$U(C_t, H(S_t)) = \log(C_t) + \alpha H(S_t) \tag{2.3}$$

where H($\cdot$) is a function of sugar on the health of the consumer that is not internalized. C is a nested CES aggregation of N_m grocery goods within M different categories:

$$C_t = \left(\sum_m^M \left(\sum_k^N (\varphi_{k,t} c_{k,t})^{\frac{\sigma_m - 1}{\sigma_m}} \right)^{\frac{\sigma_m}{\sigma_m - 1}} \right)^{\frac{\sigma}{\sigma - 1}}. \tag{2.4}$$

The household derives utility from consumption of each product at time t $c_{k,t}$. This per product utility varies with time specific taste shocks for each product $\varphi_{k,t}$. This taste parameter is meant to represent all of the internalized utility that the household derives from the product (including its health benefits). So if the sugar tax increases the household's knowledge about the negative health effects of sugar than we might expect φ to decrease for products with high amounts of sugar. Each product group has it's own specific elasticity of substitution σ_m and there is another elasticity of substitution across product categories σ.

Following standard CES optimization, each product category m has price index at time t of:

$$P_{tm} = \left[\sum_k \left(\frac{p_{k,t}}{\varphi_{k,t}} \right)^{1-\sigma_m} \right]^{\frac{1}{1-\sigma_m}}. \tag{2.5}$$

What matters for the household's price index is the taste adjusted price of the good $\frac{p_k}{\varphi_k}$ rather than simply the price. This is important for our analysis as the tax increases the price of SSB, but can also affect the "taste" for sugar given that the tax may signal to consumers that sugar is a bad. It should be noted that both the taxes affect on increasing the price of SSB as well as decreasing the taste for sugar will both act to increase the consumer's price index. So the price index could increase even more than the simple case depending on the consumer's ability to substitute to other goods.

We use the reverse weighting method from Redding and Weinstein (2020) to calculate the elasticity of substitution for each product module σ_m and the elasticity of substitution across product modules σ using all of the data in our sample (both inside and outside Cook County). We also follow Redding and Weinstein (2020) and invert the price index in (2.5) so that we do not need to estimate the time specific good taste parameter $\varphi_{k,t}$. The change in the price index in a product category m between any two periods then becomes:

$$\log \frac{P_t}{P_{t-1}} = \frac{1}{\sigma_m - 1} \log \frac{\lambda_{t,t-1}}{\lambda t-1,t} + \frac{1}{N_{t,t-1}} \sum_k \log \left(\frac{p_{kt}}{p_{k,t-1}} \right) + \frac{1}{\sigma_m - 1} \frac{1}{N_{t,t-1}} \sum_k \log \left(\frac{s^*_{kt}}{s^*_{k,t-1}} \right). \tag{2.6}$$

The first term $\frac{1}{\sigma_m - 1} \log \frac{\lambda_{t,t-1}}{\lambda t-1,t}$ is the variety adjustment term first seen in Feenstra (1994) that accounts for increases in welfare due to new products ($\lambda_{t,t-1}$ is the budget share spent on products common to both periods). The number of products common to both periods is denoted by $N_{t,t-1}$ and the budget share of a product as a percentage of the total spent on common products is denoted by s^*. We estimate this price index change for every product category (module described in data section) for households within and outside of Cook County. We then treat the category level P_m as prices and estimate (2.6) using category prices and shares to get aggregate price indexes for Cook County and the comparison group outside of Cook County.

After determining the four price levels (Cook vs. counterfactual, pretax vs. post-tax), we calculate the average compensating variation (in dollars) with:

$$CV = \bar{C}\left(\frac{P_{1,\text{Cook}}}{P_{0,\text{Cook}}} - \frac{P_{1,\text{counterfactual}}}{P_{0,\text{counterfactual}}}\right) \tag{2.7}$$

where $\bar{C}$ is the average consumption (in dollars) among Cook residents in the pretax period. The intuition is as follows: suppose the price level increases by 10% in Cook County during the taxed period. Over the same time period, the price level increased in counterfactual counties by 6%. Hence, in the absence of treatment, (we assume) the price level in Cook County would have increased by only 6%, so the effective price increase caused by the tax is 4%. To "compensate" the consumer for the increased prices (and return her to her baseline level of consumption), we must provide her with 4% of the money she spent in the pretax period.

As a base month we use July 2017, the month prior to the implementation of the tax. We also compute a simple Laspeyres index in Cook County and the outside counterfactual as a simple robustness check.

We use this compensating variation to estimate the dollar cost per-gram of reducing sugar consumption, which is useful for conducting cost-benefit analyses of beverage taxes. These analyses must take into account 1) grams of sugar reduced per unit tax, 2) health cost savings per gram of sugar reduction, 3) compensating variation (social cost in terms of an increase in the price index), and 4) tax revenue. This paper provides estimates of 1) and 3). We leave it to other researchers to determine 2) and 4).

2.5 Results

In this section we present results using the methodology discussed in Section 2.4 and data introduced in Section 2.3.

2.5.1 Effect on nutrients

Here we estimate the effect of the Cook County beverage tax on nutrients purchased, from all scanned food products, using Equation 2.1. To instill some confidence that our comparison group - Chicago DMA panelists who live outside Cook County - is a credible counterfactual, we plot the time series of average nutrients purchased per month by treatment group in Figure **??**. We similarly examine panelist behaviors, such as number of trips, total expenditures, and items scanned per month across treatment arms in Figure **??**. In both sets of time series plots, one can observe that pretreatment trends are remarkably similar with shocks occurring in both groups in the same months.

We present our estimates of Equation 2.1 for Nutrition Facts nutrients in Table 2.2. We find that during the four months of the tax, Cook County households purchased 14.4 to 16.3 percent fewer grams of sugar than did control households during the same time period. We find significant decreases in carbohydrates purchased, but no significant decrease in non-sugar carbohydrates purchased. We find no change in fat or calories from sources other than sugar, and we find marginally significant decreases in calories, fiber, protein, and sodium, but these decreases become insignificant after adding sampling weights and balancing the panel. Interestingly, in the four months following removal of the tax, treatment households did *not* increase purchases of any nutrient relative to their pretax baseline. In particular, we can reject that households increased their sugar purchases in the four months following the tax to make up for the reduction in sugar purchased during the tax.

2.5.2 Effect on beverage purchases

The Cook County beverage tax targeted beverages that contain added caloric sweeteners, with some exceptions such as beverages whose first ingredient is milk, 100% fruit juices, and baby formula. We investigate how much the tax reduced purchases of some categories of beverages by estimating Equation 2.1 using the inverse hyperbolic sine of volume purchased of each beverage category as the outcome variable. Table 2.3 shows our findings. We find that regular soda volume decreased by 33.2 - 39.1%, diet soda volume decreased 14.3 - 27.0%, fruit drinks[13] by 30.5 - 35.5%, and liquid tea by 18.4 - 24.8%. Among these product categories that contain taxed products, we find no compensating increase among in the four months following the tax. However, the point estimates for regular soda are positive and one-sixth to one-third the magnitude of the decrease during the tax, suggesting that consumers may have partially offset the temporary reductions in regular soda purchased. Notably, we can reject that households maintained reduced levels of SSBs after the tax was removed.

Moving to product categories that contain mostly untaxed beverages, we find that bottled water demand increased 8.5% to 14.8%, but we cannot reject a 0% increase. We find weak evidence of increased demand for solid coffee but no change for liquid coffee.[14] We find no change in demand for milk and no significant increase in alcohol purchases, but we cannot reject large increases.

In Table 2.4, we estimate effects of the tax on the intensive margin of purchasing select beverage categories. That is, we estimate Equation 2.1 setting the outcome variable to 1 if the household purchased any product in that category that month, 0 otherwise. We find that the tax decreased the probability of purchasing regular soda by 5.1 to 6.6 percentage points, diet soda by 2.0 to 3.8 percentage points, fruit drinks by 4.8 to 5.5 percentage points, and liquid tea by 3.4 to 4.1 percentage points. We find no significant change in the probability of purchasing beverages of other categories.

[13] 100% fruit juice was not eligible for the tax, so some products in this category were not taxed.

[14] Liquid coffee products with added sweeteners were taxed, except those whose first ingredient is milk.

2.5.3 Effect on sugar from select sources

Here we characterize the change in sugar during the tax by the source of that sugar. We generate the results in Table 2.5 by fitting Equation 2.1 with levels of sugar from a given source as the outcome variable. During the tax, treatment households reduced their sugar purchases by 81 to 93 grams per month. We find that sugar reductions from regular soda can explain 61% to 81% of this total reduction, and fruit drinks can explain 17% to 25%. To see if households substituted towards sugary non-beverage products, we estimated the effects of the tax on grams of sugar from the ”Candy” and ”Cookies” product groups. We find directionally positive but not statistically significant changes in sugar from sweets, which we interpret as suggestive evidence that households may have substituted soda for other sweet foods, albeit not one-for-one in terms of grams of sugar.

After removal of the tax, we find that treatment households increased their purchases of sugar by 31 to 43 grams per month relative to their pretax baseline. Though this change is not statistically significant, it is between 33% and 47% of the taxed-period decrease. Interestingly, a large portion of this increase can be attributed to increases in sugar from fruit drinks. We do *not* find large increases in sugar from regular soda following removal of the tax.

2.5.4 Effect on shopping and scanning behaviors

Next we examine whether panelists changed their shopping or scanning behaviors during the tax. We again estimate Equation 2.1 using the inverse hyperbolic sine of each panelist behavior as the outcome variable of interest, and we show the results in Table 2.6. We find no evidence of changes in trips, total amount spent (as shown on the receipt), total amount spent on scanned items (as input by panelists), or number of scanned items with coupons or deals. We find marginally significant reductions in food expenditures scanned, number of items scanned, and number of food items scanned; however, these reductions become insignificant after including sampling weights. We find a marginally significant reduction in number of storebrand items scanned, which remains significant after adding sampling weights and balancing the panel. In sum, we interpret these results as suggestive evidence that the tax did not reduce trips or amount spent at stores but may have shifted some spending from food to nonfood.

2.5.5 Treatment effect heterogeneity

In this section we look into heterogeneous treatment effects by household characteristics, geography, and pretreatment levels of soda purchases.

We begin by estimating Equation 2.2 using as outcomes the inverse hyperbolic sine of grams of sugar and ounces of regular soda purchased. For the interaction terms, we use dummies for household income range, race, and education.[15] Relative to the baseline category of households with income under $35 thousand, higher income households reduced their sugar less during the tax on average, though no difference is statistically significant. Compared to White households, Black and Asian households changed their sugar about the same amount, but households that identify as Other had much larger decreases, albeit not statistically significant. The differences by education are the smallest, showing little difference in treatment effects by education level.

Next, we examine the importance of distance from untaxed stores with respect to the effectiveness of the tax at reducing sugar consumption. Theoretically, households near the border face lower costs to avoid the tax than do households in the center of the taxed county since the cost to travel to an untaxed store is lower. We estimate Equation 2.2 after coding each household with an indicator variable for their zip-code layer as depicted in Figure 2.3. We present the coefficient estimates on the interaction term between the tax and the zip layer dummy in Table 2.8. While the coefficients are indistinguishable from zero for households who live in a border zip code, the coefficients are negative and significant for the two inner zip code layers.

To further examine whether cross-border shopping may explain the differences in treatment effects by geography, we estimate Equation 2.1 using the inverse hyperbolic sine of trips, expenditures, and grams of sugar separately for stores located on the same side of the Cook County border as one's residence ("inside border") and on the opposite side of the border ("cross border"). As

[15] Race and education is that of the female head of household, or the male head of household if there is no female head.

can be seen in Table 2.9, we find that trips to inside-border stores decreased 2.2% to 4.5%, offset by a nearly equal increase in trips to cross-border stores. Expenditures decreased somewhat in inside-border stores while expenditures did not change in cross-border stores. Interestingly, sugar purchases decreased by 21% to 22% in inside-border stores while they *increased* 6.9% to 10.3% in cross-border stores. This observation highlights the importance of including products purchased at stores outside the taxed jurisdiction when estimating effects of county-level taxes. In our example, failing to account for cross-border purchases would overstate the effect of the tax on sugar purchased by more than 50%.

Finally, we examine effects of the beverage tax on sugar by pretreatment levels of soda purchases. We construct deciles by first calculating each households average monthly soda volume purchases during the year prior to the tax, excluding one month prior to the tax (to reduce bias from anticipatory effects). We divide this amount by monthly expenditures to so that the ranking does not simply capture the size of the household's budget. Summary statistics for each decile can be found in Table 2.11. While deciles 1 and 2 purchase no soda before the tax, the top decile purchases about 730 ounces of regular soda per month and gets nearly 80% of its sugar from regular soda.

Ex ante, one would expect to find smaller treatment effects for low soda purchasing households relative to high soda purchasing households. Hence, estimating effects by soda decile serves as a placebo check, though it should be noted that low soda purchasing households could still have negative treatment effects since soda was not the only taxed beverage category. We present our heterogeneous treatment effect estimates by pretreatment soda decile in Table 2.10. We find no significant effects of the tax on sugar purchases for deciles 1 through 6, marginally significant negative effects for decile 7, and large, significant negative effects for deciles 8 through 10. The coefficients for the top deciles indicate that the tax reduced sugar purchases by 29% to 44%. Additionally, we do not find compensating increases for any decile following removal of the tax. We interpret these findings as evidence that the tax reduced sugar purchases for high soda purchasing households while the tax was in effect but had no lasting impacts.

The decreases in sugar observed for high soda purchasing households are large enough that we are powered to detect effects with relatively small sample sizes. As a replication exercise, we estimate treatment effects along the pretreatment soda dimension in two other sets of localities that passed beverage taxes: Philadelphia, and San Francisco and Seattle.[16] In both sets of results, as seen in Tables 2.12 and 2.13, the tax reduced sugar purchases substantially (roughly 33% in Philadelphia and 60% in San Francisco/Seattle) during the first four months of the tax for the highest soda purchasing quintiles. During the subsequent four months, the point estimates stay negative but shrink towards zero, becoming statistically insignificant. This finding suggests that beverage taxes may lose some efficacy over time in their ability to reduce sugar consumption.

2.5.6 Consumer price indices and welfare

As described in Section 2.4, we estimate consumer price indices in the Chicago area for households that reside both within and outside Cook County. We find that Cook County had a relatively significant increase in the price indices of the soda product module and a minor increase in the overall retail price index during the period of the tax.

Figure 2.4 shows the change in the aggregate household budget share spent on soda before and after the tax. The budget shares in Cook County and the rest of Chicago track each other closely except in the months directly prior to the implementation of the tax, perhaps evidence of stockpiling behavior. The similar trajectory of the share of soda products inside and outside Cook County indicate that a simple Laspeyres index may be an appropriate estimate for the change in the cost of living due to the tax. The price of soda in Cook County increases by around 1 cent per ounce, which indicates complete pass through.

Figure 2.5 shows our estimated results for the price index inside and outside Cook County. The top row shows the price index estimated for only carbonated beverages. The left hand side is the simple Laspeyres price index, while the right hand side is the CES Unified Price Index (CUPI) where the aggregate household is allowed to have time varying taste shocks and CES utility. Both methods yield a roughly 60 percent increase in the price index on soda for households in Cook County. If we

[16] We group San Francisco and Seattle since both cities enacted beverage taxes concurrently, and grouping the two provides sufficient power. We are underpowered in the remaining cities.

combine this with information from Figure 2.4, we get that the representative household's increase in the cost of living during the period of the tax was \$7.63 via the CUPI measure and \$8.43 via the simple Laspeyres price index. When we combine that with the total estimated decline in sugar purchases during the same period (368 grams) we estimate that each gram of sugar mitigation cost approximately 2.1-2.3 cents. Since carbonated drinks are only one of the taxed beverage categories, this estimate is a lower bound for the cost of sugar reduction.

The bottom row of Figure 2.5 shows the change in the overall retail price index in Cook County and outside, of which soda is only a small part. Again, the left hand side shows the results from the Laspeyres price index, while the right hand side shows the full CUPI retail price index. The Laspeyres price index in Cook County is higher than outside Cook County for every month of the tax, with the exception of September. The Retail CUPI index paints a clearer picture: the price index of households within Cook County increased by about 4% in August (the first month of the tax) and a further 1.25% in October before falling in December to 2% over July prices. This is compared to outside Cook County that only had a 2.5% increase in retail prices during this period. Given that households in Cook County on average spent \$285 on scanned products each month in 2017, the cumulative cost per household during the four month tax was \$11.25 using the difference in Laspeyeres price indices and \$24.21 using the difference in CUPI price indices. The estimated decline in sugar purchases during the same period is 368 grams, indicating that each gram of sugar reduction cost households 3.1 to 6.6 cents. However, unlike the increase in the soda-only price index, the pretreatment trends in price indices are not as obviously parallel, which does not lend credence to the parallel trends assumption. As such, we leave this as suggestive evidence that the cost per gram of sugar reduction is likely larger than two 2 cents.

2.6 Discussion and Conclusion

In this paper, we use Nielsen household scanner data and UPC-level nutrition data to examine the effect of the Cook County, Illinois beverage tax of one cent per ounce on sugar consumption.

Using a differences-in-differences specification, we find that during the taxed period, residents of Cook County reduced total sugar purchased by 14.9%, which could be explained almost entirely by reduced purchases of regular soda and fruit drinks. The tax did *not* cause sustained reduction in sugar purchases after the tax was removed. We find that the tax primarily affected households living far from the county border, and we present suggestive evidence that households may have increased trips and sugar purchases at cross-border stores by a modest amount. We show that the tax reduced sugar consumption the most for high regular-soda consuming households.

Finally, we estimate that the reduction in sugar cost households, using compensating variation, approximately 3.1 - 6.6 cents per gram. This cost is considerably larger than the 0.4 cents per gram increase in prices expected for a one-cent-per-ounce tax estimated by Grummon et al. (2019). Part of the difference can be explained by the fact that Cook County's tax included beverages sweetened with non-nutritive sweeteners, which raised prices without impacting sugar consumption. Other explanations may include that consumers are more inelastic in their demand for high sugar-density beverages, consumers substitute towards higher sugar-density beverages when prices increase, and that consumers shop outside of the taxed jurisdiction or substitute to untaxed sugar-sweetened products. As previously mentioned, policymakers should consider the benefits of the observed sugar reduction and decide whether these costs are justified.

Our findings are robust to several empirical checks. We verify that our results are not sensitive to whether we include sampling weights or restrict the sample only to panelists who record purchases every month during the period of study. We check our results using an alternative counterfactual constructed of urban households that live far from Cook County and find substantively similar results. We find that the net decrease in sugar purchases is driven by high soda-purchasing households, which gives us confidence that the decrease observed is driven through the studied tax and not general sugar purchase trends. Finally, we repeat our empirical exercise in Philadelphia, Seattle, and San Francisco and find large, significant decreases in sugar purchases by high pretax consumers of regular soda during the first four months of the tax. We also find negative but insignificant decreases during the second four months of the tax.

2.6.1 Limitations

Our approach has a few limitations. We observe sugar purchases, not sugar consumption, which is the policy-relevant metric. However, if we assume that the fraction of sugar purchased to sugar consumed does not change following the tax, our estimate of the effect of the tax on sugar purchases would additionally estimate the effect on sugar consumption. This assumption could be violated if, for example, consumers buy less sugar after the tax is enacted but consume the same amount by consuming a greater portion of their purchased sugar. It is not possible to test this assumption using the Nielsen data, though violations of this assumption likely bias our estimates *upwards* since consumers presumably start consuming a larger portion of the purchased sugar as it becomes relatively more expensive. Additionally, we cannot observe some intensive margin changes to soda consumption like consuming more soda from a container before disposal.

Additionally, we observe grocery store purchases that households scan, which do not include unscanned items such as food consumed away from home. Since the tax studied in this paper are levied based on volume, taxes on beverages sold at grocery stores are generally larger on a percentage basis than are taxes on beverages sold at restaurants, which could result in relatively larger effect sizes at grocery stores relative to restaurants. For example, Marinello et al. (2020) find that Oakland's one-cent-per-ounce tax raised prices on bottled regular soda sold in restaurants by 8%. The pass-through of the tax may also be greater at grocery stores compared to restaurants, as Cawley et al. (2021) found to be true in Boulder.

Finally, we do not examine health directly. While sugar consumption decreases during the taxed period, agents may simultaneously change other behaviors, such as exercise. Consider the agent who rationally reduces time spent exercising as her total caloric intake decreases. If large portions of the public behave in such a manner, the health care cost savings from reduced sugar consumption may be partially erased by other behavioral changes. Future research that examines the effect of SSB taxes on health expenditures directly is warranted.

2.7 Acknowledgements

We gratefully recognize Prashant Bharadwaj, Jeffrey Clemens, Gordon Dahl, Itzik Fadlon, Mark Jacobsen, Katherine Meckel, Kaspar Wuthrich, and numerous seminar participants for their helpful comments and suggestions. We thank Eric Youn for outstanding research assistance. Our analyses are calculated based in part on data from Nielsen Consumer LLC and marketing databases provided through the NielsenIQ Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the NielsenIQ data are those of our own and do not reflect the views of NielsenIQ. NielsenIQ is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein.

Chapter 2 is currently being prepared for submission for publication of the material. Goodman, Zachary A.; Orchard, Jacob. The dissertation/thesis author was the primary investigator and author of this material.

Bibliography

Allcott, H., Lockwood, B. B., & Taubinsky, D. (2019a). Regressive sin taxes, with an application to the optimal soda tax. *The Quarterly Journal of Economics*, *134*(3), 1557–1626.

Allcott, H., Lockwood, B. B., & Taubinsky, D. (2019b). Should we tax sugar-sweetened beverages? an overview of theory and evidence. *Journal of Economic Perspectives*, *33*(3), 202–27.

Cawley, J., Frisvold, D., Hill, A., & Jones, D. (2020a). The impact of the philadelphia beverage tax on prices and product availability. *Journal of Policy Analysis and Management*, *39*(3), 605–628.

Cawley, J., Frisvold, D., Hill, A., & Jones, D. (2020b). Oakland's sugar-sweetened beverage tax: Impacts on prices, purchases and consumption by adults and children. *Economics & Human Biology*, *37*, 100865.

Cawley, J., Frisvold, D., & Jones, D. (2020). The impact of sugar-sweetened beverage taxes on purchases: Evidence from four city-level taxes in the united states. *Health Economics*, *29*(10), 1289–1306.

Cawley, J., Frisvold, D., Jones, D., & Lensing, C. (2021). The pass-through of a tax on sugar-sweetened beverages in boulder, colorado. *American Journal of Agricultural Economics*, *103*(3), 987–1005.

Cawley, J., & Frisvold, D. E. (2017). The pass-through of taxes on sugar-sweetened beverages to retail prices: The case of berkeley, california. *Journal of Policy Analysis and Management*, *36*(2), 303–326.

Cawley, J., & Meyerhoefer, C. (2012). The medical care costs of obesity: An instrumental variables approach. *Journal of health economics*, *31*(1), 219–230.

Chaloupka, F. J., Yurekli, A., & Fong, G. T. (2012). Tobacco taxes as a tobacco control strategy. *Tobacco control*, *21*(2), 172–180.

Chetty, R., Looney, A., & Kroft, K. (2009). Salience and taxation: Theory and evidence. *American economic review*, *99*(4), 1145–77.

Colchero, M. A., Popkin, B. M., Rivera, J. A., & Ng, S. W. (2016). Beverage purchases from stores in mexico under the excise tax on sugar sweetened beverages: Observational study. *bmj*, *352*.

Colchero, M. A., Rivera-Dommarco, J., Popkin, B. M., & Ng, S. W. (2017). In mexico, evidence of sustained consumer response two years after implementing a sugar-sweetened beverage tax. *Health Affairs*, *36*(3), 564–571.

Cotti, C., Nesson, E., & Tefft, N. (2016). The effects of tobacco control policies on tobacco products, tar, and nicotine purchases among adults: Evidence from household panel data. *American Economic Journal: Economic Policy*, *8*(4), 103–23.

de Ruyter, J. C., Olthof, M. R., Seidell, J. C., & Katan, M. B. (2012). A trial of sugar-free or sugar-sweetened beverages and body weight in children. *New England Journal of Medicine*, *367*(15), 1397–1406.

Dharmasena, S., & Capps Jr, O. (2012). Intended and unintended consequences of a proposed national tax on sugar-sweetened beverages to combat the us obesity problem. *Health economics*, *21*(6), 669–694.

Dubois, P., Griffith, R., & Nevo, A. (2014). Do prices and attributes explain international differences in food purchases? *American Economic Review*, *104*(3), 832–67.

Ebbeling, C. B., Feldman, H. A., Chomitz, V. R., Antonelli, T. A., Gortmaker, S. L., Osganian, S. K., & Ludwig, D. S. (2012). A randomized trial of sugar-sweetened beverages and adolescent body weight. *N Engl J Med*, *367*, 1407–1416.

Ebbeling, C. B., Feldman, H. A., Osganian, S. K., Chomitz, V. R., Ellenbogen, S. J., & Ludwig, D. S. (2006). Effects of decreasing sugar-sweetened beverage consumption on body weight in adolescents: A randomized, controlled pilot study. *Pediatrics*, *117*(3), 673–680.

Einav, L., Leibtag, E., & Nevo, A. (2010). Recording discrepancies in nielsen homescan data: Are they present and do they matter? *QME*, *8*(2), 207–239.

Falbe, J., Rojas, N., Grummon, A. H., & Madsen, K. A. (2015). Higher retail prices of sugar-sweetened beverages 3 months after implementation of an excise tax in berkeley, california. *American journal of public health*, *105*(11), 2194–2201.

Falbe, J., Thompson, H. R., Becker, C. M., Rojas, N., McCulloch, C. E., & Madsen, K. A. (2016). Impact of the berkeley excise tax on sugar-sweetened beverage consumption. *American journal of public health*, *106*(10), 1865–1871.

Feenstra, R. C. (1994). New product varieties and the measurement of international prices. *The American Economic Review*, 157–177.

Finkelstein, E. A., Trogdon, J. G., Cohen, J. W., & Dietz, W. (2009). Annual medical spending attributable to obesity: Payer-and service-specific estimates: Amid calls for health reform, real cost savings are more likely to be achieved through reducing obesity and related risk factors. *Health affairs*, *28*(Suppl1), w822–w831.

Grummon, A. H., Lockwood, B. B., Taubinsky, D., & Allcott, H. (2019). Designing better sugary drink taxes. *Science*, *365*(6457), 989–990.

Hadaway, C. K., Marler, P. L., & Chaves, M. (1993). What the polls don't show: A closer look at us church attendance. *American Sociological Review*, 741–752.

Harding, M., Leibtag, E., & Lovenheim, M. F. (2012). The heterogeneous geographic and socioeconomic incidence of cigarette taxes: Evidence from nielsen homescan data. *American Economic Journal: Economic Policy*, *4*(4), 169–98.

Imamura, F., O'Connor, L., Ye, Z., Mursu, J., Hayashino, Y., Bhupathiraju, S. N., & Forouhi, N. G. (2015). Consumption of sugar sweetened beverages, artificially sweetened beverages, and fruit juice and incidence of type 2 diabetes: Systematic review, meta-analysis, and estimation of population attributable fraction. *Bmj*, *351*.

Krumpal, I. (2013). Determinants of social desirability bias in sensitive surveys: A literature review. *Quality & Quantity*, *47*(4), 2025–2047.

Locander, W., Sudman, S., & Bradburn, N. (1976). An investigation of interview method, threat and response distortion. *Journal of the American Statistical Association*, *71*(354), 269–275.

Long, M. W., Gortmaker, S. L., Ward, Z. J., Resch, S. C., Moodie, M. L., Sacks, G., Swinburn, B. A., Carter, R. C., & Wang, Y. C. (2015). Cost effectiveness of a sugar-sweetened beverage excise tax in the us. *American journal of preventive medicine*, *49*(1), 112–123.

Malik, V. S., & Hu, F. B. (2019). Sugar-sweetened beverages and cardiometabolic health: An update of the evidence. *Nutrients*, *11*(8), 1840.

Malik, V. S., Popkin, B. M., Bray, G. A., Després, J.-P., Willett, W. C., & Hu, F. B. (2010). Sugar-sweetened beverages and risk of metabolic syndrome and type 2 diabetes: A meta-analysis. *Diabetes care*, *33*(11), 2477–2483.

Marinello, S., Pipito, A. A., Leider, J., Pugach, O., & Powell, L. M. (2020). The impact of the oakland sugar-sweetened beverage tax on bottled soda and fountain drink prices in fast-food restaurants. *Preventive medicine reports*, *17*, 101034.

Mensch, B. S., & Kandel, D. B. (1988). Underreporting of substance use in a national longitudinal youth cohort: Individual and interviewer effects. *Public Opinion Quarterly*, *52*(1), 100–124.

Redding, S. J., & Weinstein, D. E. (2020). Measuring aggregate price indices with taste shocks: Theory and evidence for ces preferences. *The Quarterly Journal of Economics*, *135*(1), 503–560.

Silver, B. D., Anderson, B. A., & Abramson, P. R. (1986). Who overreports voting? *The American Political Science Review*, 613–624.

Silver, L. D., Ng, S. W., Ryan-Ibarra, S., Taillie, L. S., Induni, M., Miles, D. R., Poti, J. M., & Popkin, B. M. (2017). Changes in prices, sales, consumer spending, and beverage consumption one year after a tax on sugar-sweetened beverages in berkeley, california, us: A before-and-after study. *PLoS medicine*, *14*(4), e1002283.

Sweetened beverage tax ordinance 16-5931. (2016). *Cook County Board of Commissioners*.

USDA. (2015). *Dietary guidelines for americans 2015-2020*. Government Printing Office.

Vartanian, L. R., Schwartz, M. B., & Brownell, K. D. (2007). Effects of soft drink consumption on nutrition and health: A systematic review and meta-analysis. *American journal of public health*, *97*(4), 667–675.

Wang, Y. C., Coxson, P., Shen, Y.-M., Goldman, L., & Bibbins-Domingo, K. (2012). A penny-per-ounce tax on sugar-sweetened beverages would cut health and cost burdens of diabetes. *Health Affairs*, *31*(1), 199–207.

Table 2.1: Summary statistics for Chicago DMA sample: demographics

	Control	Cook	Total
	%	%	%
Hours employment/week of male head of HH			
No male head	25.14	36.34	31.18
< 30	2.61	3.86	3.28
30 - 34	1.97	1.72	1.83
≥ 35	49.88	35.66	42.22
Not employed	20.40	22.42	21.49
Hours employment/week of female head of HH			
No female head	20.35	23.23	21.90
< 30	10.52	10.40	10.46
30 - 34	5.85	3.77	4.73
≥ 35	32.89	32.80	32.84
Not employed	30.39	29.79	30.07
Racial identity of the household			
White	78.78	60.08	68.70
Black	7.87	23.63	16.36
Asian	3.91	5.72	4.89
Other	9.44	10.57	10.05
Household income			
<$35K	21.46	31.46	26.85
$35K - $59,999	18.33	21.11	19.83
$60K - $99,999	25.61	21.29	23.28
>$100K	34.60	26.14	30.04
Age of the (female) head of household			
<35	14.86	16.36	15.67
35 - 49	31.55	27.58	29.41
50-64	35.90	31.37	33.46
65+	17.69	24.70	21.47
Education of the (female) head of household			
< HS	2.99	2.12	2.52
HS Grad	27.69	25.35	26.43
Some College	28.28	31.77	30.16
Bachelor's+	41.04	40.76	40.89
Household size, top-coded at 6			
Mean	2.58	2.27	2.44
Any children < 18	39.94	26.67	32.79
Households	1,076	965	2,041
Household-months	12,795	11,432	24,227

All person-level variables (race, age, education) take the value of the female head of household unless there is no female head, in which case the variables take the value of the male head of household. Control includes households that reside in the Chicago DMA outside of Cook County limits. Summary statistics take into account sampling weights.

Table 2.2: Effects of Cook County Beverage Tax on nutrients

	During tax			4 months post tax		
Dependent Variable	(1)	(2)	(3)	(4)	(5)	(6)
All sugar	-0.144***	-0.149**	-0.163**	-0.009	0.006	0.013
	(0.036)	(0.052)	(0.059)	(0.036)	(0.047)	(0.050)
Carbohydrates	-0.104**	-0.111*	-0.119*	-0.020	-0.010	-0.004
	(0.034)	(0.049)	(0.055)	(0.034)	(0.044)	(0.045)
Carb., non-sugar	-0.059	-0.077	-0.064	-0.026	-0.017	-0.014
	(0.037)	(0.047)	(0.051)	(0.038)	(0.051)	(0.050)
Calories	-0.081*	-0.092	-0.096	-0.022	-0.013	-0.004
	(0.036)	(0.053)	(0.060)	(0.035)	(0.048)	(0.047)
Calories, non-sugar	-0.057	-0.087	-0.056	-0.047	-0.058	0.021
	(0.043)	(0.059)	(0.059)	(0.040)	(0.066)	(0.066)
Fat	-0.045	-0.056	-0.035	-0.025	-0.035	-0.012
	(0.030)	(0.041)	(0.043)	(0.031)	(0.043)	(0.045)
Fiber	-0.063*	-0.060	-0.060	-0.021	-0.017	-0.014
	(0.027)	(0.036)	(0.040)	(0.029)	(0.038)	(0.039)
Protein	-0.070*	-0.088*	-0.078	-0.029	-0.034	-0.010
	(0.031)	(0.041)	(0.044)	(0.032)	(0.045)	(0.045)
Sodium	-0.060*	-0.054	-0.046	-0.005	-0.004	-0.005
	(0.024)	(0.030)	(0.032)	(0.024)	(0.032)	(0.035)
Treated Households	1142	1142	719	1220	1220	624
Households	2400	2400	1530	2575	2575	1302
Household-months	30272	30272	22950	29831	29831	19530
Balanced Panel	No	No	Yes	No	No	Yes
Sampling Weights	No	Yes	Yes	No	Yes	Yes

Note: Coefficients represent the percent change in the dependent variable during the four months that the Cook County beverage tax was active (1 - 3) and the four months after the tax was no longer active (4 - 6) relative to the 12 months preceding the tax, omitting the month prior to the tax. All specifications include household and month fixed effects. Standard errors in parentheses are clustered at the zip code level. ***, **, and * indicate significance at the 1, 5, and 10 percent critical levels, respectively.

Table 2.3: Effects of Cook County Beverage Tax on select beverage volume

	During tax			4 months post tax		
Dependent Variable	(1)	(2)	(3)	(4)	(5)	(6)
Reg. soda	-0.332***	-0.376***	-0.391***	0.062	0.057	0.106
	(0.083)	(0.101)	(0.114)	(0.083)	(0.103)	(0.121)
Diet soda	-0.270***	-0.226*	-0.143	0.031	0.016	0.084
	(0.078)	(0.096)	(0.108)	(0.079)	(0.099)	(0.117)
Fruit drinks	-0.348***	-0.305**	-0.355***	-0.054	-0.042	-0.014
	(0.071)	(0.093)	(0.107)	(0.080)	(0.102)	(0.127)
Milk	-0.044	0.040	0.050	-0.024	-0.086	-0.081
	(0.067)	(0.095)	(0.110)	(0.066)	(0.098)	(0.112)
Bottled water	0.085	0.124	0.148	-0.033	-0.132	-0.153
	(0.080)	(0.109)	(0.134)	(0.086)	(0.122)	(0.145)
Alcohol	0.080	0.051	-0.001	0.075	0.137	0.147
	(0.055)	(0.075)	(0.089)	(0.061)	(0.077)	(0.094)
Liquid tea	-0.202***	-0.184*	-0.248**	-0.025	-0.055	-0.065
	(0.057)	(0.082)	(0.087)	(0.061)	(0.084)	(0.102)
Liquid coffee	-0.020	0.002	-0.024	-0.020	-0.021	-0.041
	(0.031)	(0.038)	(0.041)	(0.027)	(0.036)	(0.041)
Solid coffee	0.031	0.084*	0.055	-0.000	0.046	0.047
	(0.035)	(0.040)	(0.047)	(0.042)	(0.050)	(0.059)
Treated Households	1142	1142	719	1220	1220	624
Households	2400	2400	1530	2575	2575	1302
Household-months	30272	30272	22950	29831	29831	19530
Balanced Panel	No	No	Yes	No	No	Yes
Sampling Weights	No	Yes	Yes	No	Yes	Yes

Note: Coefficients represent the percent change in the dependent variable during the four months that the Cook County beverage tax was active (1 - 3) and the four months after the tax was no longer active (4 - 6) relative to the 12 months preceding the tax, omitting the month prior to the tax. All specifications include household and month fixed effects. Standard errors in parentheses are clustered at the zip code level. ***, **, and * indicate significance at the 1, 5, and 10 percent critical levels, respectively.

Table 2.4: Effects of Cook County Beverage Tax on probability of purchasing select beverages

	During tax			4 months post tax		
Dependent Variable	(1)	(2)	(3)	(4)	(5)	(6)
Reg. soda	-0.051***	-0.059***	-0.066***	0.009	0.011	0.015
	(0.014)	(0.017)	(0.019)	(0.014)	(0.017)	(0.021)
Diet soda	-0.038**	-0.034*	-0.020	0.003	-0.000	0.006
	(0.012)	(0.015)	(0.017)	(0.013)	(0.017)	(0.020)
Fruit drinks	-0.055***	-0.048**	-0.055**	-0.017	-0.025	-0.023
	(0.013)	(0.016)	(0.018)	(0.014)	(0.018)	(0.022)
Milk	-0.008	0.008	0.009	-0.007	-0.018	-0.018
	(0.011)	(0.015)	(0.018)	(0.011)	(0.016)	(0.018)
Bottled water	0.012	0.019	0.023	-0.002	-0.014	-0.020
	(0.012)	(0.017)	(0.020)	(0.013)	(0.019)	(0.022)
Alcohol	0.013	0.009	0.002	0.014	0.027*	0.035*
	(0.010)	(0.013)	(0.016)	(0.011)	(0.014)	(0.017)
Liquid tea	-0.040***	-0.034*	-0.041**	-0.011	-0.016	-0.016
	(0.010)	(0.015)	(0.016)	(0.012)	(0.016)	(0.020)
Liquid coffee	-0.005	-0.001	-0.010	-0.006	-0.006	-0.010
	(0.008)	(0.009)	(0.010)	(0.006)	(0.009)	(0.010)
Solid coffee	0.007	0.019	0.013	-0.004	0.010	0.012
	(0.009)	(0.010)	(0.012)	(0.010)	(0.012)	(0.014)
Treated Households	1142	1142	719	1220	1220	624
Households	2400	2400	1530	2575	2575	1302
Household-months	30272	30272	22950	29831	29831	19530
Balanced Panel	No	No	Yes	No	No	Yes
Sampling Weights	No	Yes	Yes	No	Yes	Yes

Note: Coefficients represent the change in probability of purchasing the dependent variable during the four months that the Cook County beverage tax was active (1 - 3) and the four months after the tax was no longer active (4 - 6) relative to the 12 months preceding the tax, omitting the month prior to the tax. All specifications include household and month fixed effects. Standard errors in parentheses are clustered at the zip code level. ***, **, and * indicate significance at the 1, 5, and 10 percent critical levels, respectively.

Table 2.5: Effects of Cook County Beverage Tax on sugar from select sources

	During tax			4 months post tax		
Dependent Variable	(1)	(2)	(3)	(4)	(5)	(6)
All sugar	-80.972**	-92.053*	-92.691*	31.785	42.950	30.561
	(28.789)	(37.169)	(40.755)	(30.574)	(40.479)	(44.589)
Sugar from reg. soda	-55.588*	-74.164**	-56.092	11.969	-1.145	-4.064
	(22.220)	(28.337)	(31.843)	(22.699)	(29.410)	(32.745)
Sugar from fruit drinks	-16.486**	-15.938*	-23.496**	9.198	18.647*	16.209
	(5.026)	(6.652)	(7.710)	(5.454)	(7.847)	(9.614)
Sugar from milk	-0.506	-0.696	-0.375	0.860	0.989	0.582
	(0.723)	(0.965)	(1.096)	(0.783)	(1.027)	(1.180)
Sugar from liq. tea	-4.131	-3.826	-6.714	2.034	-0.273	-0.952
	(2.420)	(3.429)	(3.752)	(2.584)	(3.025)	(3.887)
Sugar from liq. coffee	-0.230	0.280	0.381	-0.092	-0.276	-0.588
	(0.525)	(0.664)	(0.676)	(0.528)	(0.719)	(0.735)
Sugar from sweets	1.342	3.318	5.109	-4.873	4.417	2.734
	(3.973)	(5.906)	(6.394)	(4.615)	(6.849)	(8.123)
Treated Households	1142	1142	719	1220	1220	624
Households	2400	2400	1530	2575	2575	1302
Household-months	30272	30272	22950	29831	29831	19530
Balanced Panel	No	No	Yes	No	No	Yes
Sampling Weights	No	Yes	Yes	No	Yes	Yes

Note: Coefficients represent the absolute change in the dependent variable during the four months that the Cook County beverage tax was active (1 - 3) and the four months after the tax was no longer active (4 - 6) relative to the 12 months preceding the tax, omitting the month prior to the tax. All specifications include household and month fixed effects. Standard errors in parentheses are clustered at the zip code level. ***, **, and * indicate significance at the 1, 5, and 10 percent critical levels, respectively.

Table 2.6: Effects of Cook County Beverage Tax on panelist behaviors

	During tax			4 months post tax		
Dependent Variable	(1)	(2)	(3)	(4)	(5)	(6)
All Trips	-0.012	-0.004	0.004	0.002	0.000	0.003
	(0.015)	(0.022)	(0.024)	(0.017)	(0.022)	(0.024)
All expenditures	-0.027	-0.021	-0.008	-0.019	-0.010	0.004
	(0.017)	(0.023)	(0.025)	(0.018)	(0.023)	(0.024)
Expenditures scanned	-0.026	-0.017	-0.011	-0.012	-0.013	-0.001
	(0.017)	(0.024)	(0.027)	(0.018)	(0.023)	(0.025)
Food expenditures scanned	-0.053*	-0.041	-0.040	-0.001	-0.006	0.012
	(0.026)	(0.035)	(0.039)	(0.026)	(0.034)	(0.034)
Items scanned	-0.040*	-0.024	-0.012	-0.011	-0.001	0.001
	(0.018)	(0.024)	(0.028)	(0.019)	(0.026)	(0.027)
Food items scanned	-0.064**	-0.050	-0.038	-0.013	-0.007	0.004
	(0.024)	(0.032)	(0.036)	(0.025)	(0.033)	(0.034)
Storebrand items scanned	-0.082*	-0.100*	-0.098*	-0.064	-0.077	-0.085
	(0.032)	(0.041)	(0.047)	(0.034)	(0.044)	(0.050)
Items with coupon scanned	-0.013	-0.001	0.023	-0.009	0.014	0.053
	(0.024)	(0.031)	(0.033)	(0.029)	(0.038)	(0.045)
Items with deals scanned	-0.018	0.009	0.062	0.024	0.043	0.053
	(0.031)	(0.043)	(0.047)	(0.036)	(0.050)	(0.052)
Treated Households	1142	1142	719	1220	1220	624
Households	2400	2400	1530	2575	2575	1302
Household-months	30272	30272	22950	29831	29831	19530
Balanced Panel	No	No	Yes	No	No	Yes
Sampling Weights	No	Yes	Yes	No	Yes	Yes

Note: Coefficients represent the percent change in the dependent variable during the four months that the Cook County beverage tax was active (1 - 3) and the four months after the tax was no longer active (4 - 6) relative to the 12 months preceding the tax, omitting the month prior to the tax. All specifications include household and month fixed effects. Standard errors in parentheses are clustered at the zip code level. ***, **, and * indicate significance at the 1, 5, and 10 percent critical levels, respectively.

Table 2.7: Heterogeneous treatment effects of tax by household characteristics

	Sugar (g)	Reg. soda (oz)
Panel A: Household income		
$35K - $59,999	0.062	-0.332
	(0.102)	(0.183)
$60K - $99,999	0.007	-0.086
	(0.090)	(0.131)
>$100K	0.078	0.115
	(0.089)	(0.182)
Panel B: Household race		
Black	0.024	-0.135
	(0.075)	(0.171)
Asian	0.021	-0.089
	(0.130)	(0.232)
Other	-0.213	-0.418
	(0.169)	(0.219)
Panel C: Household education		
Some College	-0.025	0.220
	(0.097)	(0.164)
BA+	0.043	0.506**
	(0.074)	(0.155)
Treated Households	1142	1142
Households	2400	2400
Household-months	30272	30272
Balanced Panel	No	No
Sampling Weights	Yes	Yes

Note: Coefficients represent the percent change in the dependent variable for those with the given household characteristic during the four months that the Cook County beverage tax was active relative to the 12 months preceding the tax, omitting the month prior to the tax. Omitted categories are income under $35K, race = White, and education = HS grad or less. All specifications include household and month fixed effects. Standard errors in parentheses are clustered at the zip code level. ***, **, and * indicate significance at the 1, 5, and 10 percent critical levels, respectively.

Table 2.8: Effects of Cook County Beverage Tax by zip code layer

	During tax		
Zip layer	(1)	(2)	(3)
Border zips	-0.086	-0.070	-0.109
	(0.054)	(0.070)	(0.079)
Middle layer	-0.156**	-0.140*	-0.187*
	(0.050)	(0.066)	(0.072)
Innermost layer	-0.188*	-0.187	-0.140
	(0.078)	(0.098)	(0.106)
Treated Households	884	884	587
Households	1862	1862	1253
Household-months	24108	24108	18795
Balanced Panel	No	No	Yes
Sampling Weights	No	Yes	Yes

Note: Coefficients represent the percent change in sugar purchased during the four months that the Cook County beverage tax was active, separated by zip code layer within Cook County. The omitted group is zip codes outside of Cook County, and the specification does not include zip codes that cross the border. The sample includes only households that have positive purchases of regular soda during the year prior to the tax, omitting the month prior to the tax. All specifications include household and month fixed effects. Standard errors in parentheses are clustered at the zip code level. ***, **, and * indicate significance at the 1, 5, and 10 percent critical levels, respectively.

Table 2.9: Effects of Cook County Beverage Tax on cross-border All Trips

	During tax			4 months post tax		
Dependent Variable	(1)	(2)	(3)	(4)	(5)	(6)
IB trips	-0.045**	-0.022	-0.025	0.004	0.014	0.009
	(0.017)	(0.022)	(0.025)	(0.017)	(0.022)	(0.025)
CB trips	0.028	0.026	0.036	0.010	-0.008	-0.013
	(0.020)	(0.028)	(0.033)	(0.021)	(0.026)	(0.030)
IB expenditures	-0.092**	-0.057	-0.049	-0.023	-0.005	-0.005
	(0.029)	(0.031)	(0.032)	(0.024)	(0.030)	(0.032)
CB expenditures	-0.019	-0.014	0.021	-0.035	-0.067	-0.036
	(0.041)	(0.063)	(0.072)	(0.042)	(0.055)	(0.064)
IB sugar	-0.217***	-0.208***	-0.205**	-0.023	-0.003	-0.003
	(0.045)	(0.057)	(0.064)	(0.042)	(0.055)	(0.059)
CB sugar	0.088	0.103	0.069	0.100	0.102	0.061
	(0.064)	(0.088)	(0.100)	(0.053)	(0.069)	(0.087)
Treated Households	1142	1142	719	1220	1220	624
Households	2400	2400	1530	2575	2575	1302
Household-months	30272	30272	22950	29831	29831	19530
Balanced Panel	No	No	Yes	No	No	Yes
Sampling Weights	No	Yes	Yes	No	Yes	Yes

Note: IB = Inside Border, CB = Cross Border. Coefficients represent the percent change in the dependent variable during the four months that the Cook County beverage tax was active (1 - 3) and the four months after the tax was no longer active (4 - 6) relative to the 12 months preceding the tax, omitting the month prior to the tax. All specifications include household and month fixed effects. Standard errors in parentheses are clustered at the zip code level. ***, **, and * indicate significance at the 1, 5, and 10 percent critical levels, respectively.

Table 2.10: Effects of tax on sugar by pretreatment soda decile, Cook County

	Sugar during tax			Sugar 4 months post tax		
Pre-tax soda decile	(1)	(2)	(3)	(4)	(5)	(6)
Deciles 1 - 2	-0.051	-0.106	-0.108	0.129	0.131	0.110
	(0.061)	(0.086)	(0.097)	(0.070)	(0.092)	(0.087)
Decile 3	0.065	0.098	0.003	0.046	0.145	0.114
	(0.072)	(0.082)	(0.092)	(0.101)	(0.138)	(0.168)
Decile 4	-0.082	-0.013	-0.040	0.093	0.190*	0.059
	(0.066)	(0.084)	(0.098)	(0.066)	(0.089)	(0.087)
Decile 5	-0.156	-0.088	-0.119	-0.042	0.043	0.039
	(0.089)	(0.110)	(0.140)	(0.072)	(0.080)	(0.067)
Decile 6	-0.022	-0.079	-0.198	0.048	0.092	0.114
	(0.087)	(0.174)	(0.215)	(0.067)	(0.085)	(0.101)
Decile 7	-0.185**	-0.141	-0.168	0.050	0.056	-0.005
	(0.070)	(0.096)	(0.098)	(0.077)	(0.107)	(0.089)
Decile 8	-0.291**	-0.269**	-0.083	-0.089	-0.138	-0.121
	(0.109)	(0.096)	(0.076)	(0.088)	(0.091)	(0.099)
Decile 9	-0.326***	-0.345***	-0.439***	-0.016	0.016	0.084
	(0.084)	(0.100)	(0.104)	(0.072)	(0.097)	(0.126)
Decile 10	-0.340***	-0.311***	-0.290**	-0.034	-0.044	-0.029
	(0.074)	(0.084)	(0.089)	(0.082)	(0.122)	(0.107)
Treated Households	1142	1142	720	1220	1220	625
Households	2400	2400	1535	2575	2575	1304
Household-months	30272	30272	23025	37909	37909	24776
Balanced Panel	No	No	Yes	No	No	Yes
Sampling Weights	No	Yes	Yes	No	Yes	Yes

Note: Coefficients represent the percent change in sugar purchased during the four months that the Cook County beverage tax was active (1 - 3) and the four months after the tax was no longer active (4 - 6), separated by expenditure-weighted pretreatment regular soda volume purchase decile. We omit the month prior to the tax to reduce bias from anticipatory effects. Deciles were calculated based on the treated households' average monthly volume of soda purchased during the previous year omitting the month prior to the tax. All of deciles 1 and 2 report zero regular soda purchases before the tax. All specifications include household and month fixed effects. Standard errors in parentheses are clustered at the zip code level. ***, **, and * indicate significance at the 1, 5, and 10 percent critical levels, respectively.

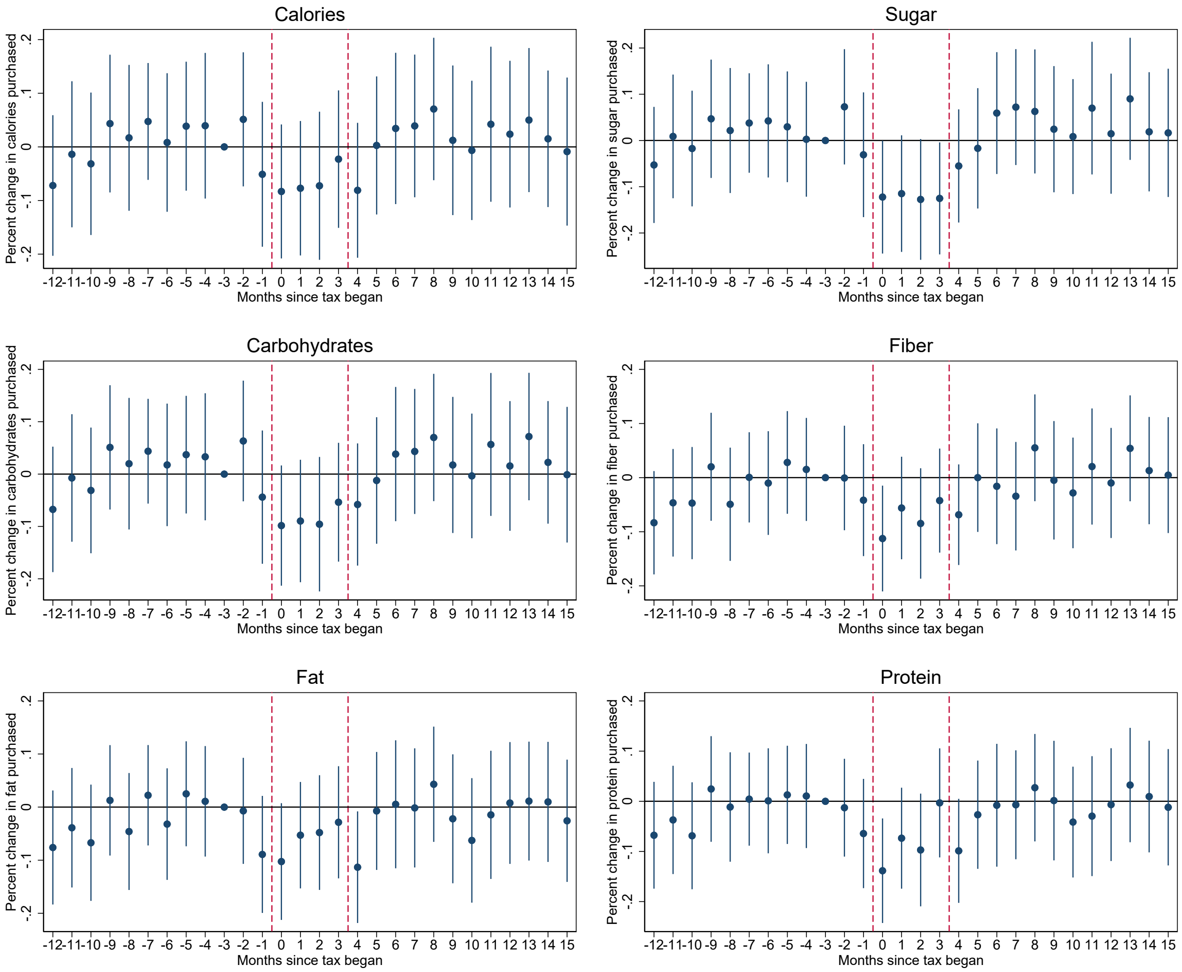

Figure 2.2: Effects of Cook County tax on nutrients purchased
Each panel displays the monthly differences in log quantities purchased by residents of Cook County, Illinois and those in the same DMA outside of the county limits. The regressions used include household and month fixed-effects, and errors are clustered and the zip-code level. Vertical dashed lines represent the start and end of the tax in Cook County.

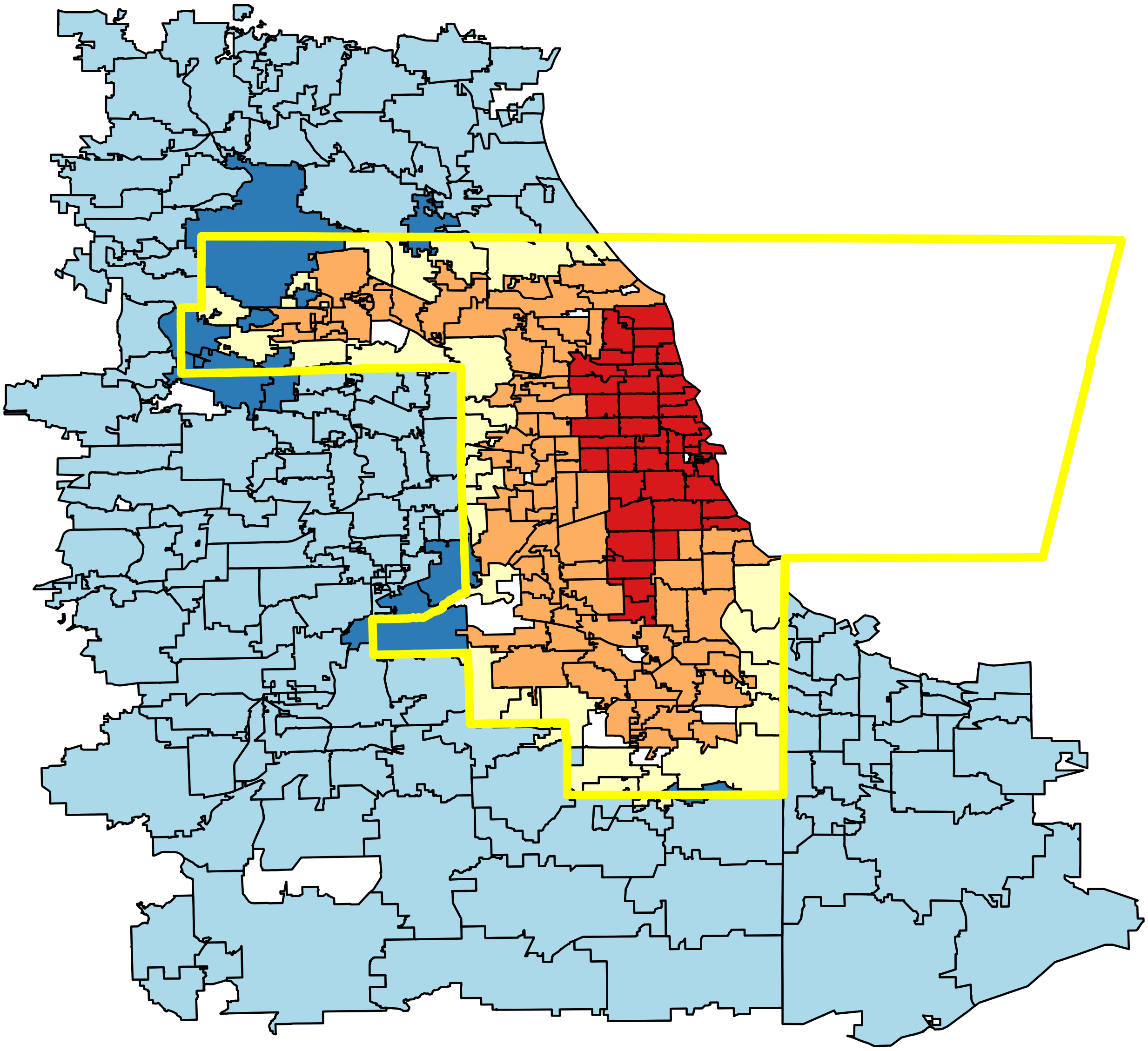

Figure 2.3: Cook County zip code layers
Zip code layers used for examining heterogeneous treatment effects by distance from the nearest border. White regions in the map have no corresponding zip code information (lakes, forest, and other land without mailing addresses) and hence are excluded from the analysis. The darkest blue zip codes contain households both inside and outside Cook County's border. The lighter-yellow zips lie on the border entirely within Cook County. The remaining layers are constructed using linear distance to the nearest untaxed zip.

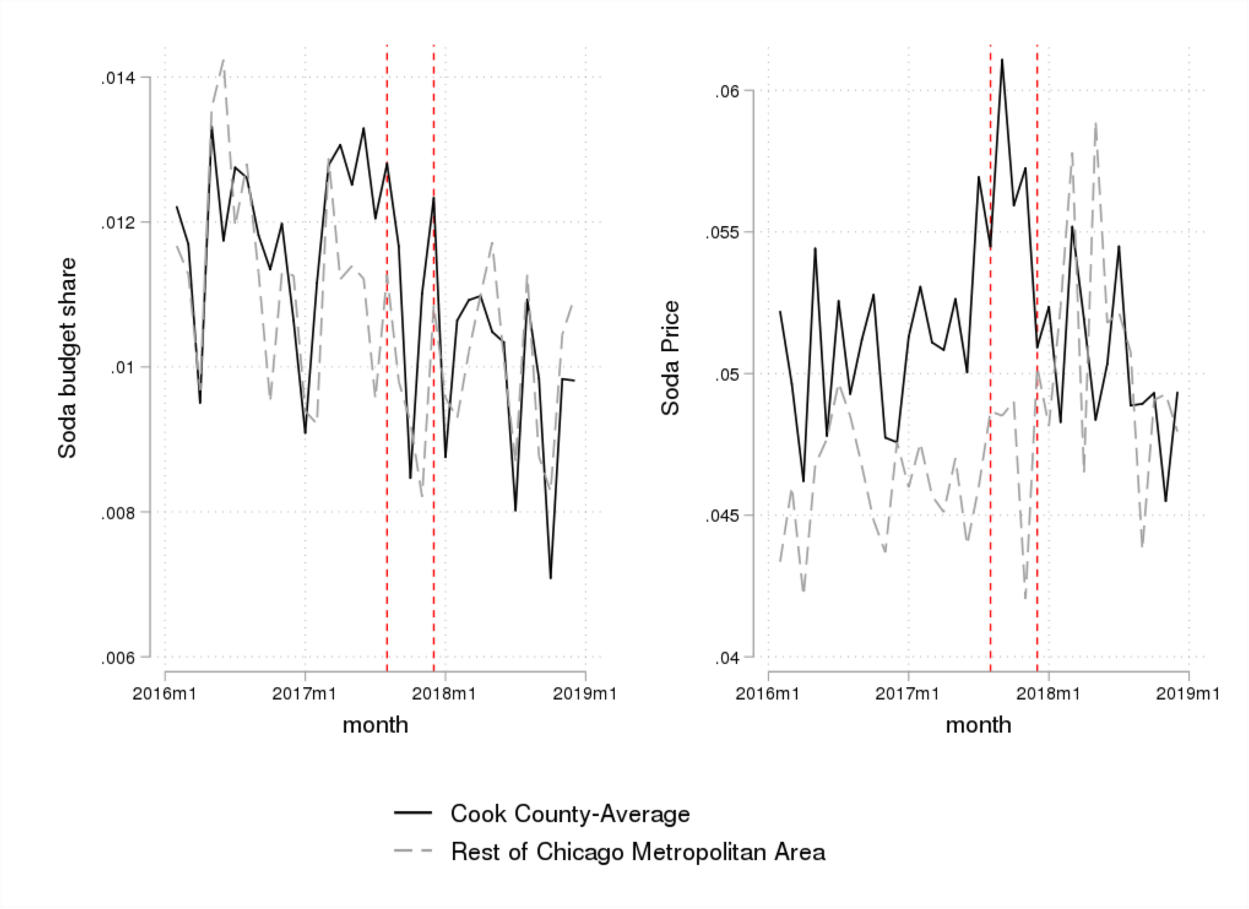

Figure 2.4: Soda budget share and price

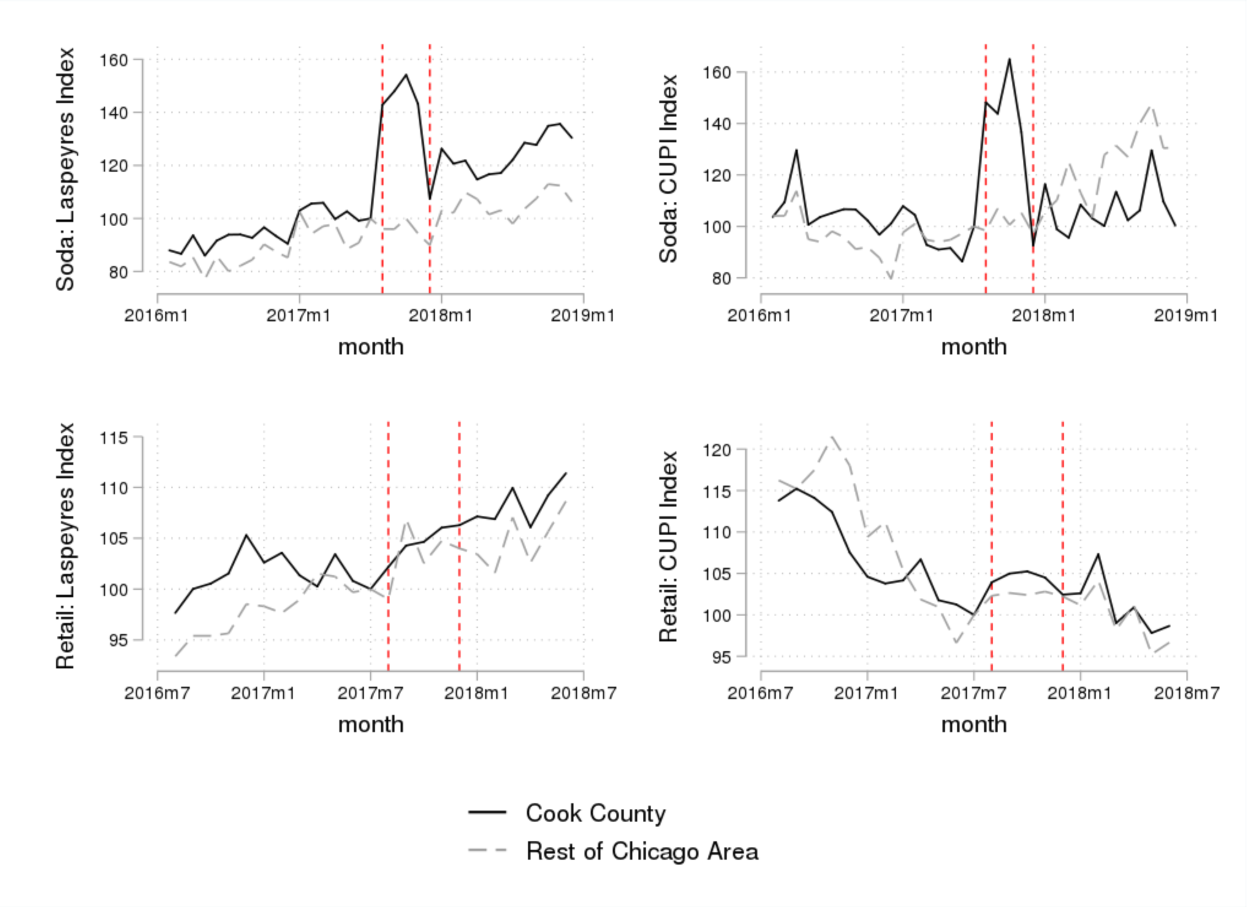

Figure 2.5: Price indices

Note: The top row shows the change in the soda only price index in Cook County and the rest of the Chicago Area. The bottom row shows the full retail price index. The right side assumes that households have nested CES preferences with time varying taste parameters, while the left hand-side is a simple Laspeyres weighted index. All price indices are set at 100 in July 2017.

2.8 Appendix

Table 2.11: Summary statistics by soda decile, Cook County

	Reg. soda (oz)	Sugar (g)	Sugar from soda (g)	Trips	Items scanned	Total spent ($)	Hh. size
Deciles 1 - 2	0.00	442.04	0.00	11.45	81.75	456.29	2.00
	(0.00)	(448.05)	(0.00)	(9.51)	(53.41)	(362.63)	(1.18)
Decile 3	7.38	560.01	11.62	11.54	99.98	505.12	2.12
	(23.59)	(576.91)	(48.87)	(6.93)	(62.91)	(325.10)	(1.15)
Decile 4	25.07	674.40	46.28	14.25	108.58	567.62	2.04
	(62.62)	(603.01)	(144.72)	(9.97)	(70.41)	(392.79)	(1.02)
Decile 5	52.11	793.11	110.68	13.79	117.54	601.94	2.31
	(120.54)	(779.81)	(350.65)	(8.88)	(75.19)	(584.85)	(1.30)
Decile 6	85.38	861.31	186.44	15.49	108.42	584.54	2.21
	(159.70)	(902.21)	(430.67)	(14.58)	(72.25)	(426.79)	(1.21)
Decile 7	133.10	983.27	293.43	12.36	114.14	556.78	2.64
	(215.90)	(991.17)	(628.80)	(8.46)	(74.32)	(420.58)	(1.34)
Decile 8	224.79	1300.58	575.75	13.83	116.80	564.61	2.39
	(290.68)	(1274.11)	(923.06)	(9.86)	(75.06)	(401.53)	(1.23)
Decile 9	371.25	1743.45	904.20	14.15	134.41	558.40	2.55
	(446.85)	(1622.65)	(1287.80)	(11.24)	(99.54)	(421.21)	(1.37)
Decile 10	729.56	2490.04	1959.81	12.17	114.15	454.44	2.28
	(704.40)	(2035.11)	(2259.00)	(8.53)	(73.93)	(339.24)	(1.26)

Note: Deciles are calculated using each household's average monthly soda volume purchased per dollar spent during the year prior to the tax, omitting the month prior to the tax. Standard deviations in parentheses.

Table 2.12: Effects of tax on sugar by pretreatment soda quintile, Philadelphia

	Months 1 - 4, sugar			Months 5 - 8, sugar		
Pre-tax soda quintile	(1)	(2)	(3)	(4)	(5)	(6)
Quintile 1	0.093	0.054	0.105	0.204	0.019	0.202
	(0.078)	(0.094)	(0.090)	(0.139)	(0.170)	(0.114)
Quintile 2	0.218	0.313*	0.355**	0.212**	0.212	0.160
	(0.120)	(0.125)	(0.111)	(0.074)	(0.126)	(0.139)
Quintile 3	0.091	0.056	-0.006	-0.012	-0.011	0.046
	(0.081)	(0.107)	(0.114)	(0.099)	(0.109)	(0.134)
Quintile 4	-0.252*	-0.277	-0.228	-0.132	-0.044	-0.191
	(0.123)	(0.154)	(0.147)	(0.094)	(0.084)	(0.105)
Quintile 5	-0.292*	-0.330*	-0.319*	-0.247	-0.193	-0.200
	(0.140)	(0.149)	(0.153)	(0.156)	(0.176)	(0.226)
Treated Households	262	262	149	262	262	146
Households	2012	2012	1228	2012	2012	1198
Household-months	25319	25319	18420	31937	31937	22762
Balanced Panel	No	No	Yes	No	No	Yes
Sampling Weights	No	Yes	Yes	No	Yes	Yes

Note: Coefficients represent the percent change in sugar purchased during the first four months (1 - 3) and second four months (4 - 6) after the Philadelphia beverage tax became active, separated by expenditure-weighted pretreatment regular soda volume purchase decile. Quintiles were calculated based on the treated households' average monthly volume of soda purchased during the previous year omitting the month prior to the tax. Nearly all of quintile 1 report zero regular soda purchases before the tax. Standard errors in parentheses are clustered at the zip code level. ***, **, and * indicate significance at the 1, 5, and 10 percent critical levels, respectively.

Table 2.13: Effects of tax on sugar by pretreatment soda quintile, Seattle and San Francisco

	Months 1 - 4, sugar			Months 5 - 8, sugar		
Pre-tax soda quintile	(1)	(2)	(3)	(4)	(5)	(6)
Quintiles 1 - 2	0.053	-0.144	-0.172	0.065	-0.001	0.093
	(0.085)	(0.142)	(0.142)	(0.097)	(0.135)	(0.138)
Quintile 3	0.150	0.099	0.070	0.061	0.258	0.187
	(0.129)	(0.148)	(0.164)	(0.144)	(0.193)	(0.211)
Quintile 4	-0.021	-0.116	-0.172	-0.180	-0.336	-0.337
	(0.105)	(0.139)	(0.148)	(0.156)	(0.218)	(0.238)
Quintile 5	-0.473**	-0.596**	-0.613**	-0.098	-0.128	-0.161
	(0.148)	(0.229)	(0.237)	(0.102)	(0.189)	(0.201)
Treated Households	229	229	153	229	229	151
Households	2141	2141	1402	2141	2141	1375
Household-months	27663	27663	21030	34932	34932	26125
Balanced Panel	No	No	Yes	No	No	Yes
Sampling Weights	No	Yes	Yes	No	Yes	Yes

Note: Coefficients represent the percent change in sugar purchased during the first four months (1 - 3) and second four months (4 - 6) after the Seattle and San Francisco beverage taxes became active, separated by expenditure-weighted pretreatment regular soda volume purchase decile. We omit the month prior to the tax to reduce bias from anticipatory effects. Quintiles were calculated based on the treated households' average monthly volume of soda purchased during the previous year omitting the month prior to the tax. All of quintile 1 and nearly all of quintile 2 report zero regular soda purchases before the tax. Standard errors in parentheses are clustered at the zip code level. ***, **, and * indicate significance at the 1, 5, and 10 percent critical levels, respectively.

Table 2.14: Effects of Cook County Beverage Tax on nutrients, 10 largest counties by population as counterfactual

	During tax			4 months post tax		
Dependent Variable	(1)	(2)	(3)	(4)	(5)	(6)
All sugar	-0.129***	-0.087*	-0.095*	0.020	0.056	0.072
	(0.031)	(0.040)	(0.043)	(0.030)	(0.036)	(0.039)
Carbohydrates	-0.069*	-0.027	-0.031	0.020	0.048	0.052
	(0.029)	(0.037)	(0.039)	(0.028)	(0.034)	(0.035)
Carb., non-sugar	0.033	0.061	0.056	0.042	0.062	0.053
	(0.032)	(0.038)	(0.043)	(0.031)	(0.039)	(0.040)
Calories	-0.035	0.003	-0.003	0.018	0.044	0.043
	(0.031)	(0.039)	(0.041)	(0.029)	(0.036)	(0.035)
Calories, non-sugar	0.028	0.056	0.059	0.028	0.042	0.047
	(0.037)	(0.046)	(0.045)	(0.033)	(0.047)	(0.041)
Fat	0.007	0.053	0.052	0.027	0.052	0.044
	(0.026)	(0.033)	(0.033)	(0.024)	(0.034)	(0.034)
Fiber	-0.011	0.037	0.034	0.041	0.069*	0.062*
	(0.023)	(0.028)	(0.030)	(0.024)	(0.030)	(0.031)
Protein	-0.010	0.030	0.031	0.030	0.059	0.061
	(0.027)	(0.032)	(0.034)	(0.026)	(0.035)	(0.034)
Sodium	-0.031	0.000	0.013	0.040*	0.053*	0.055*
	(0.021)	(0.024)	(0.025)	(0.020)	(0.025)	(0.026)
Treated Households	1142	1142	719	1220	1220	624
Households	7464	7464	4453	8250	8250	3804
Household-months	92291	92291	66795	91334	91334	57060
Balanced Panel	No	No	Yes	No	No	Yes
Sampling Weights	No	Yes	Yes	No	Yes	Yes

Note: Coefficients represent the percent change in the dependent variable during the four months that the Cook County beverage tax was active (1 - 3) and the four months after the tax was no longer active (4 - 6) relative to the 12 months preceding the tax, omitting the month prior to the tax. All specifications include household and month fixed effects. Standard errors in parentheses are clustered at the zip code level. ***, **, and * indicate significance at the 1, 5, and 10 percent critical levels, respectively.

Table 2.15: Effects of Cook County Beverage Tax on nutrients, 10 nearest population density counties as counterfactual

	During tax			4 months post tax		
Dependent Variable	(1)	(2)	(3)	(4)	(5)	(6)
All sugar	-0.117**	-0.138**	-0.104*	-0.016	0.029	0.083
	(0.039)	(0.050)	(0.052)	(0.039)	(0.050)	(0.058)
Carbohydrates	-0.089*	-0.113*	-0.072	-0.032	0.004	0.047
	(0.037)	(0.047)	(0.048)	(0.037)	(0.048)	(0.056)
Carb., non-sugar	-0.022	-0.047	-0.008	-0.038	-0.002	0.013
	(0.049)	(0.054)	(0.057)	(0.050)	(0.056)	(0.065)
Calories	-0.062	-0.085	-0.046	-0.045	-0.002	0.037
	(0.038)	(0.048)	(0.049)	(0.039)	(0.051)	(0.057)
Calories, non-sugar	-0.001	-0.033	0.028	-0.029	0.030	0.053
	(0.051)	(0.056)	(0.056)	(0.049)	(0.066)	(0.066)
Fat	0.005	0.002	0.044	-0.028	0.016	0.048
	(0.033)	(0.043)	(0.045)	(0.034)	(0.046)	(0.053)
Fiber	-0.014	-0.021	-0.003	-0.024	0.007	0.032
	(0.033)	(0.040)	(0.043)	(0.033)	(0.043)	(0.048)
Protein	-0.029	-0.034	0.005	-0.033	0.023	0.065
	(0.035)	(0.042)	(0.044)	(0.038)	(0.051)	(0.058)
Sodium	-0.013	-0.010	0.025	-0.003	0.027	0.059
	(0.026)	(0.032)	(0.034)	(0.027)	(0.036)	(0.043)
Treated Households	1142	1142	719	1220	1220	624
Households	1907	1907	1202	2059	2059	1029
Household-months	24004	24004	18030	23630	23630	15435
Balanced Panel	No	No	Yes	No	No	Yes
Sampling Weights	No	Yes	Yes	No	Yes	Yes

Note: Coefficients represent the percent change in the dependent variable during the four months that the Cook County beverage tax was active (1 - 3) and the four months after the tax was no longer active (4 - 6) relative to the 12 months preceding the tax, omitting the month prior to the tax. All specifications include household and month fixed effects. Standard errors in parentheses are clustered at the zip code level. ***, **, and * indicate significance at the 1, 5, and 10 percent critical levels, respectively.

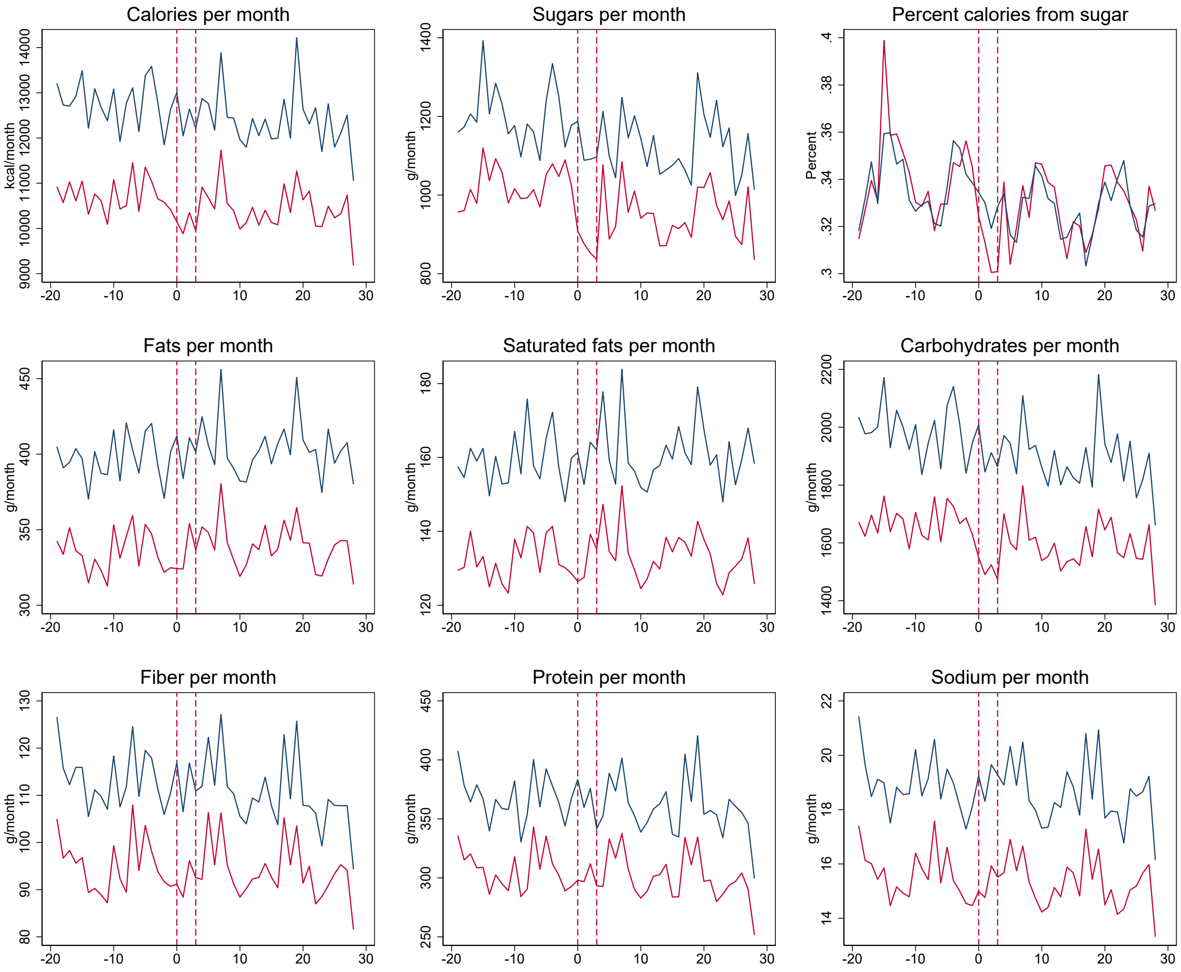

Figure 2.6: Cook County DMA: nutrient composition over time
Each panel displays monthly sums or percentages of different nutrients for residents of Cook County, Illinois (red) and those in the same DMA outside of the county limits (blue). Vertical dashed lines represent the start and end of the tax in Cook County.

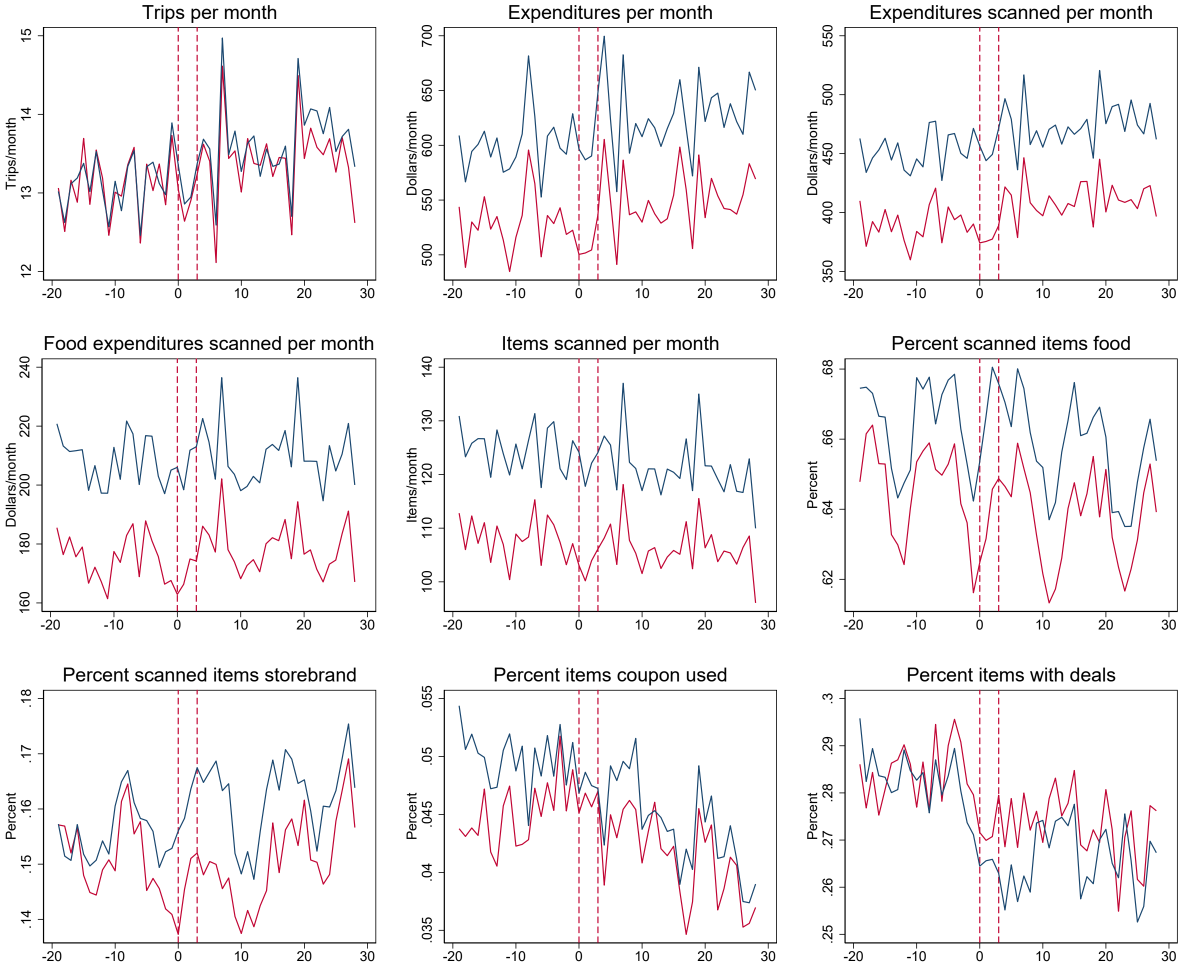

Figure 2.7: Cook County DMA: panelist behaviors over time
Each panel displays monthly sums or percentages of panelist behaviors for residents of Cook County, Illinois (red) and those in the same DMA outside of the county limits (blue). Vertical dashed lines represent the start and end of the tax in Cook County.

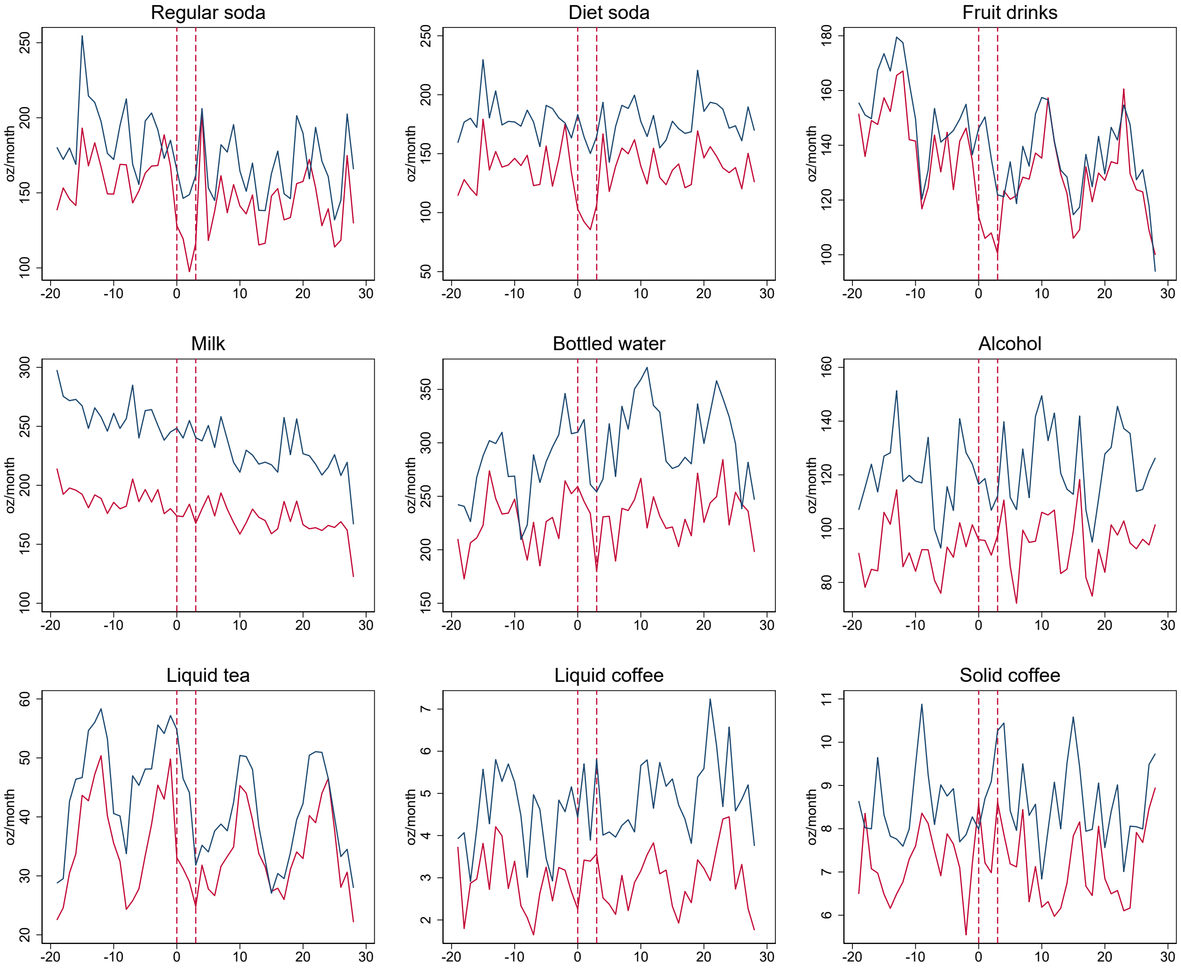

Figure 2.8: Cook County DMA: beverage purchases over time
Each panel displays monthly sums of purchased beverages (in ounces) for residents of Cook County, Illinois (red) and those in the same DMA outside of the county limits (blue). Vertical dashed lines represent the start and end of the tax in Cook County.

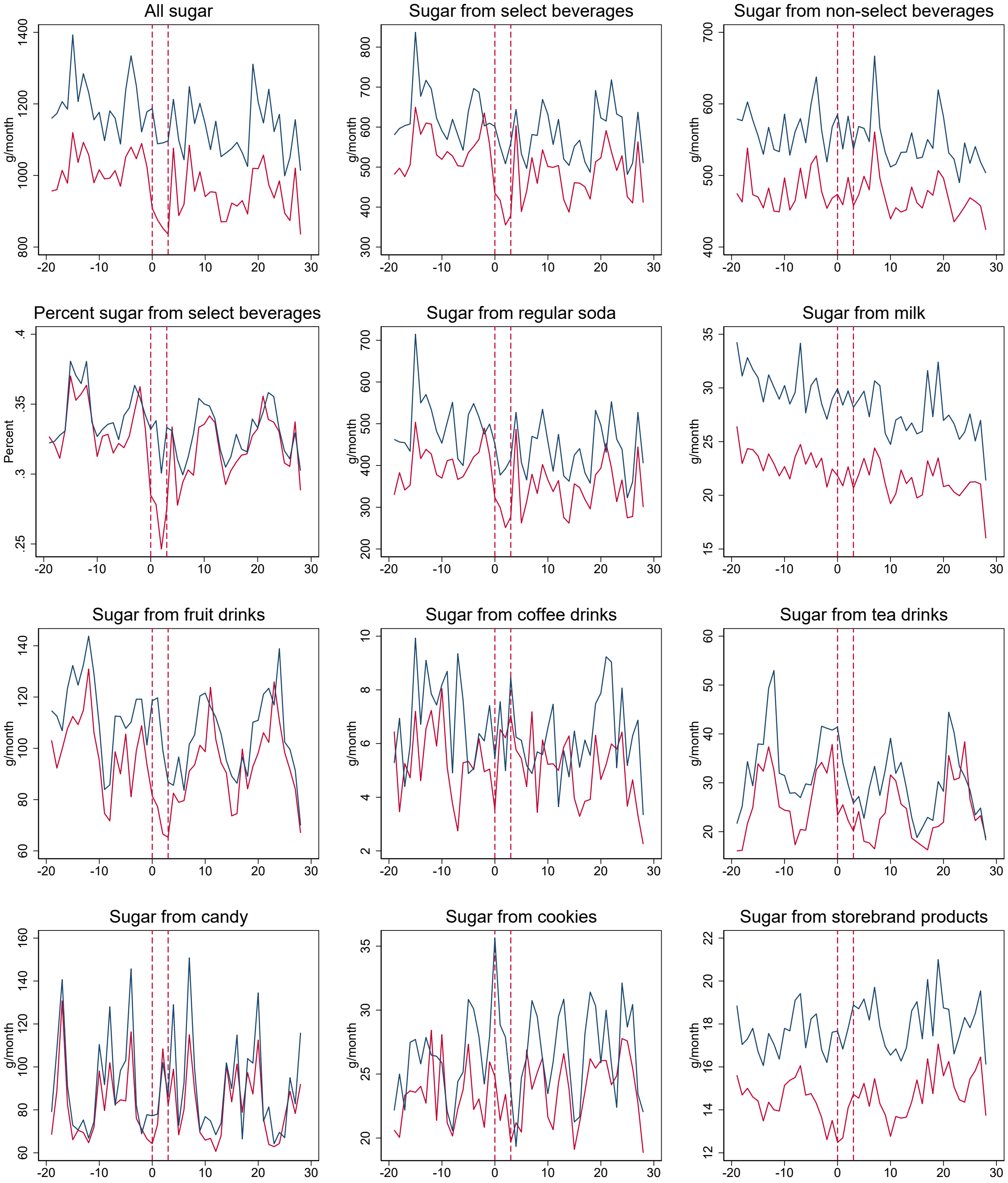

Figure 2.9: Cook County DMA: sugar by source over time
Each panel displays monthly sums of sugar (in grams) for residents of Cook County, Illinois (red) and those in the same DMA outside of the county limits (blue) from different product groups. Vertical dashed lines represent the start and end of the tax in Cook County.

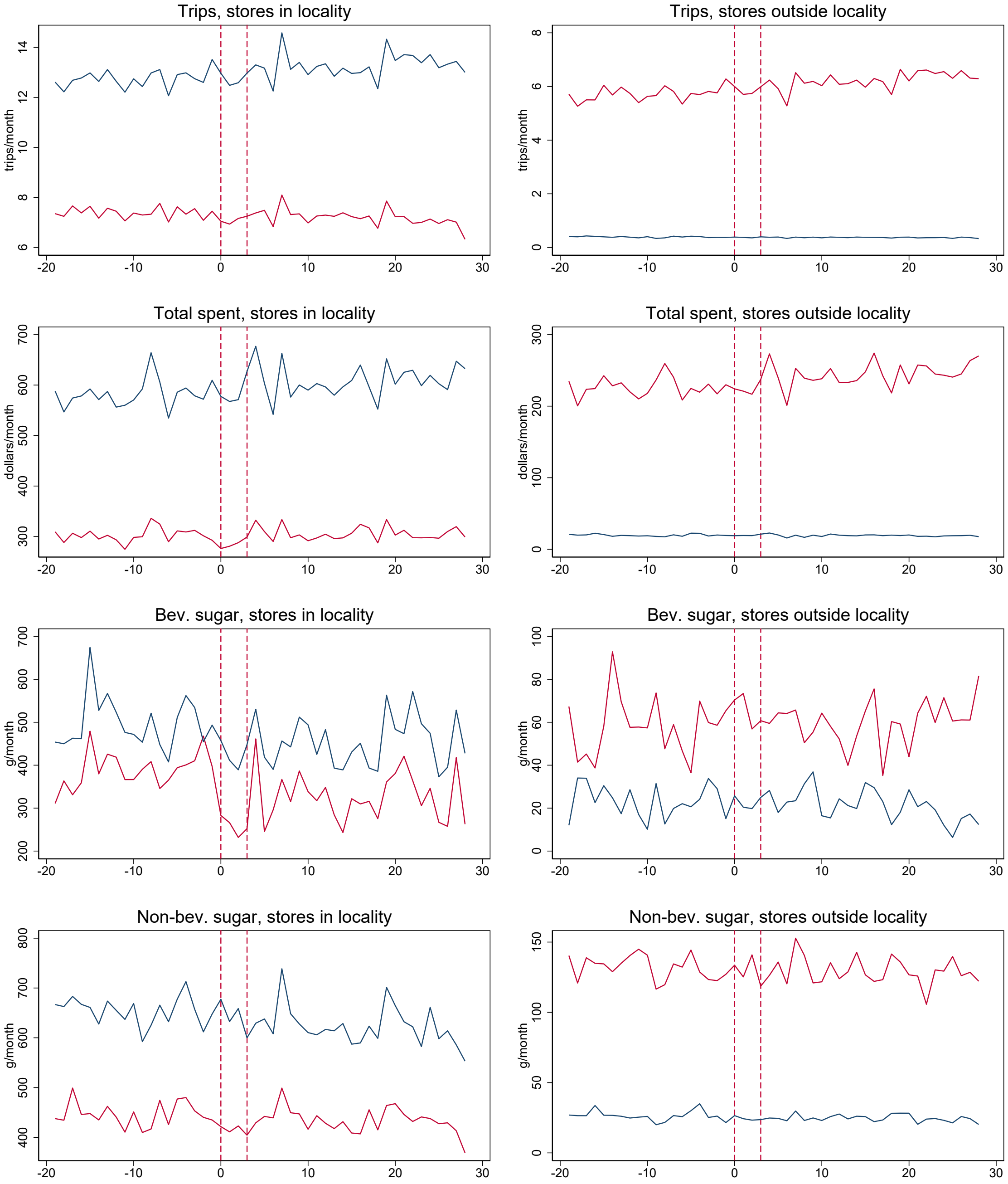

Figure 2.10: Cook County DMA: variation by store location
Each panel displays monthly sums of trips, total purchases, and grams of sugar (from beverages or non-beverages) for residents of Cook County, Illinois (red) and those in the same DMA outside of the county limits (blue). Left panels are sums within Cook County stores for Cook County residents and non-Cook County stores for control residents. Right panels show the reverse: sums within non-Cook County stores for Cook County residents and Cook County stores for control residents. Vertical dashed lines represent the start and end of the tax in Cook County.

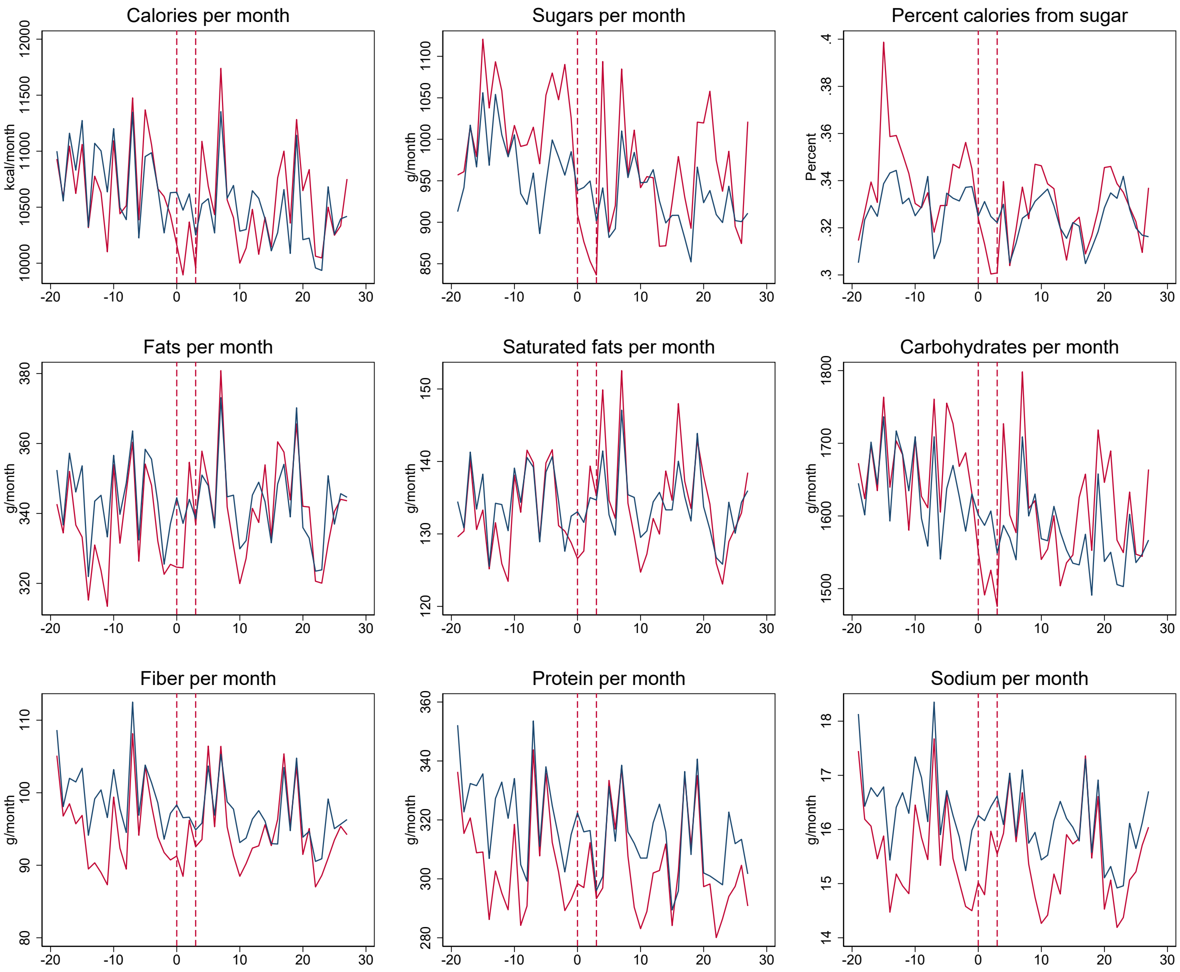

Figure 2.11: Cook County DMA: nutrient composition over time, ten largest counties as counterfactual

Each panel displays monthly sums or percentages of different nutrients for residents of Cook County, Illinois (red) and those in the ten largest counties by population, excluding Cook County (blue). Vertical dashed lines represent the start and end of the tax in Cook County.

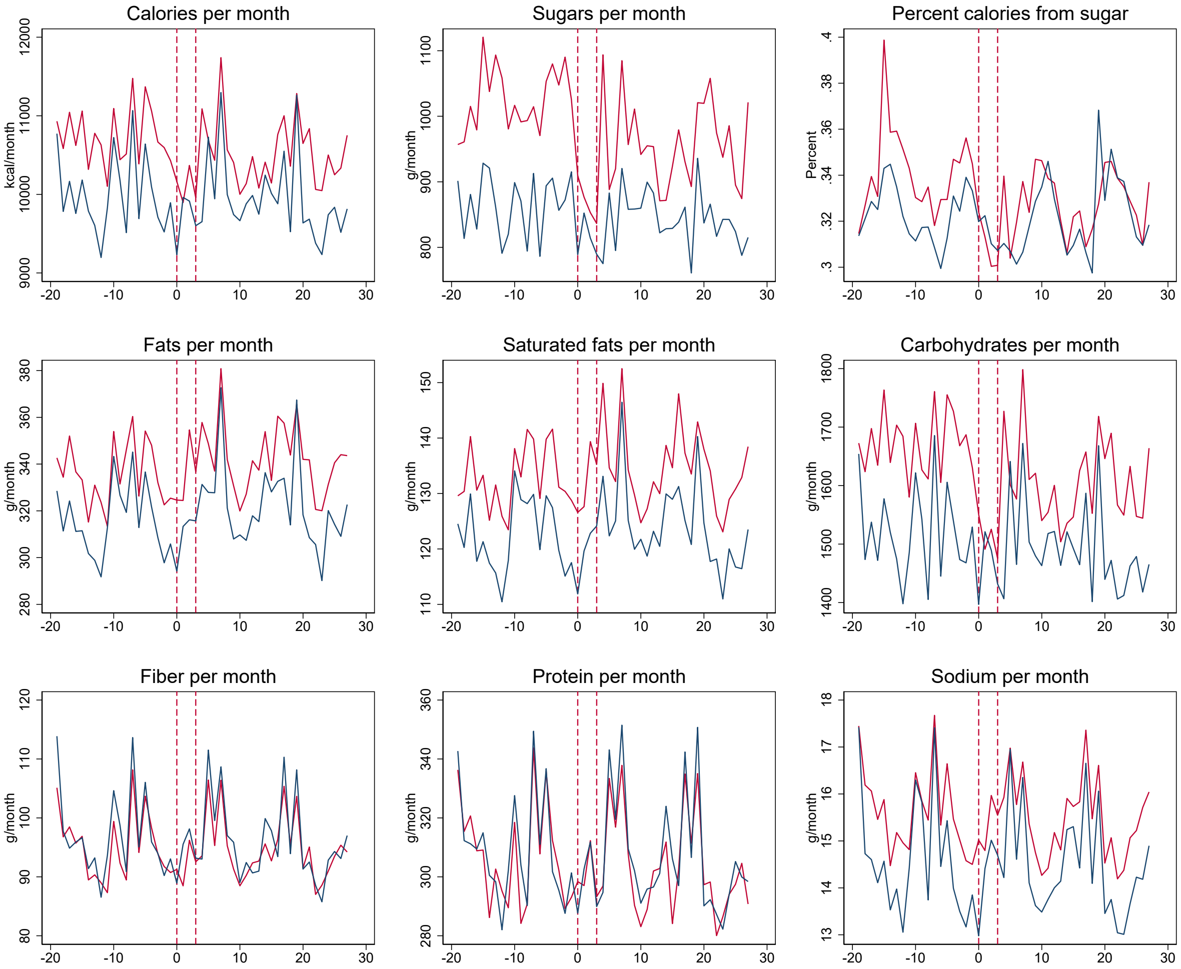

Figure 2.12: Cook County DMA: nutrient composition over time, ten nearest population density counties as counterfactual

Each panel displays monthly sums or percentages of different nutrients for residents of Cook County, Illinois (red) and those in the ten nearest counties by population density to Cook County (blue). Vertical dashed lines represent the start and end of the tax in Cook County.

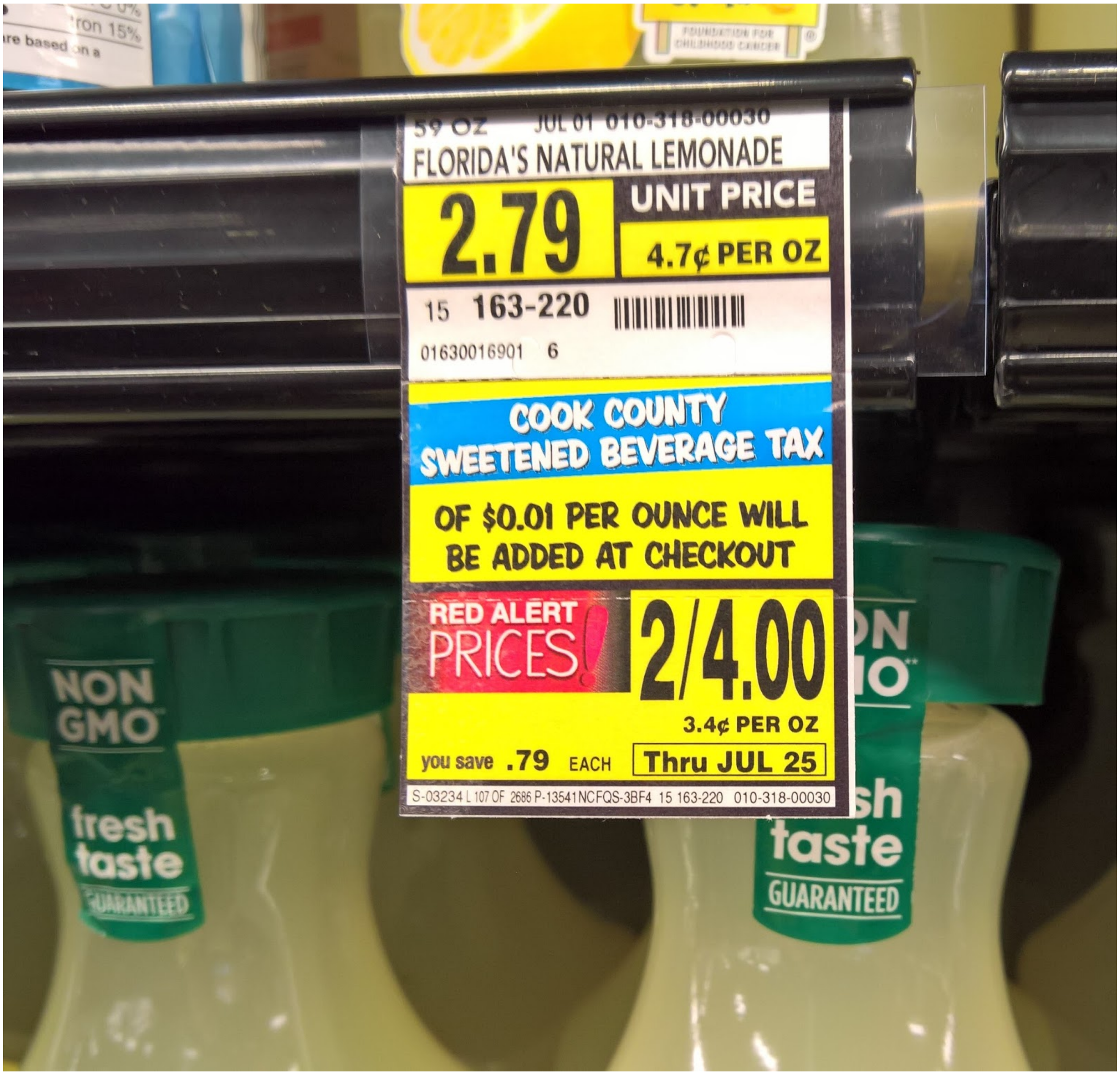

Figure 2.13: Example price tag from Cook County grocery store
This price tag includes a warning about the beverage tax in Cook County. Price tags like these may have increased salience of the tax whereas price tags in smaller, non-chain stores may not have.

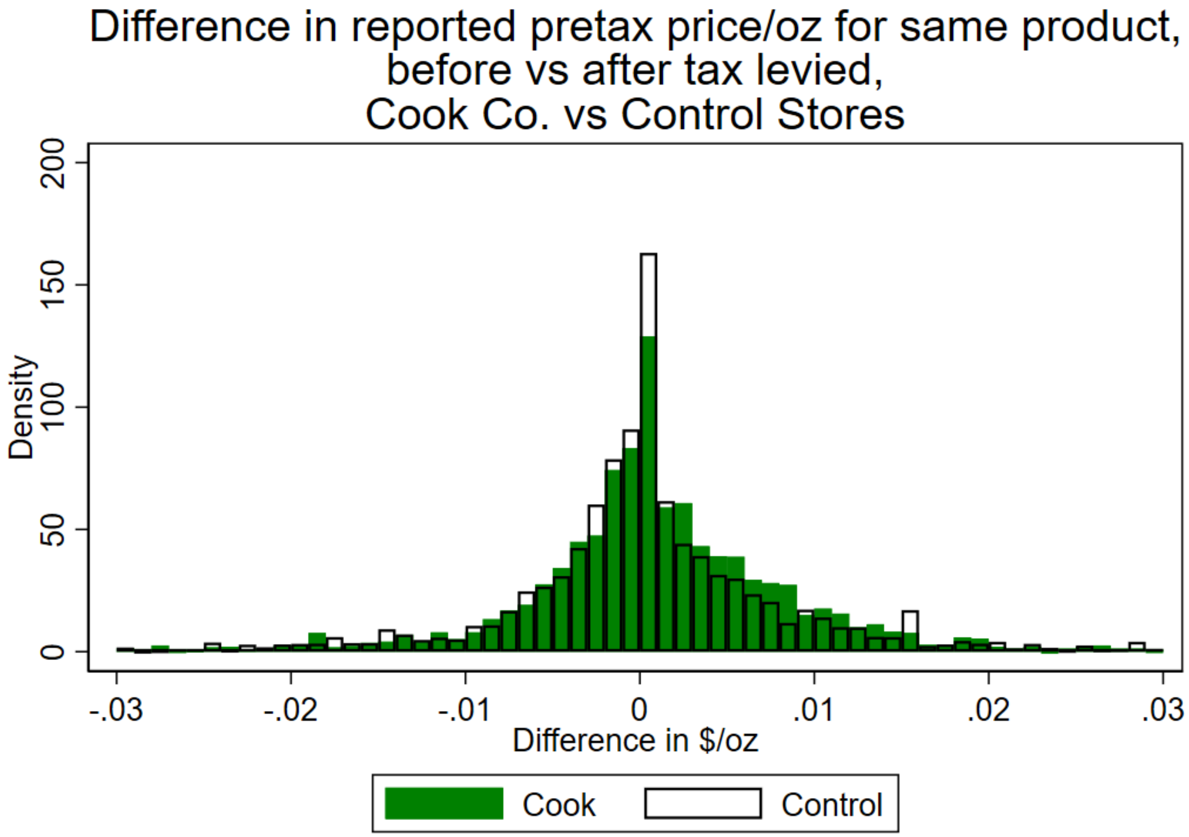

Figure 2.14: Distribution of differences in price per ounce during pretax versus taxed periods for carbonated beverages

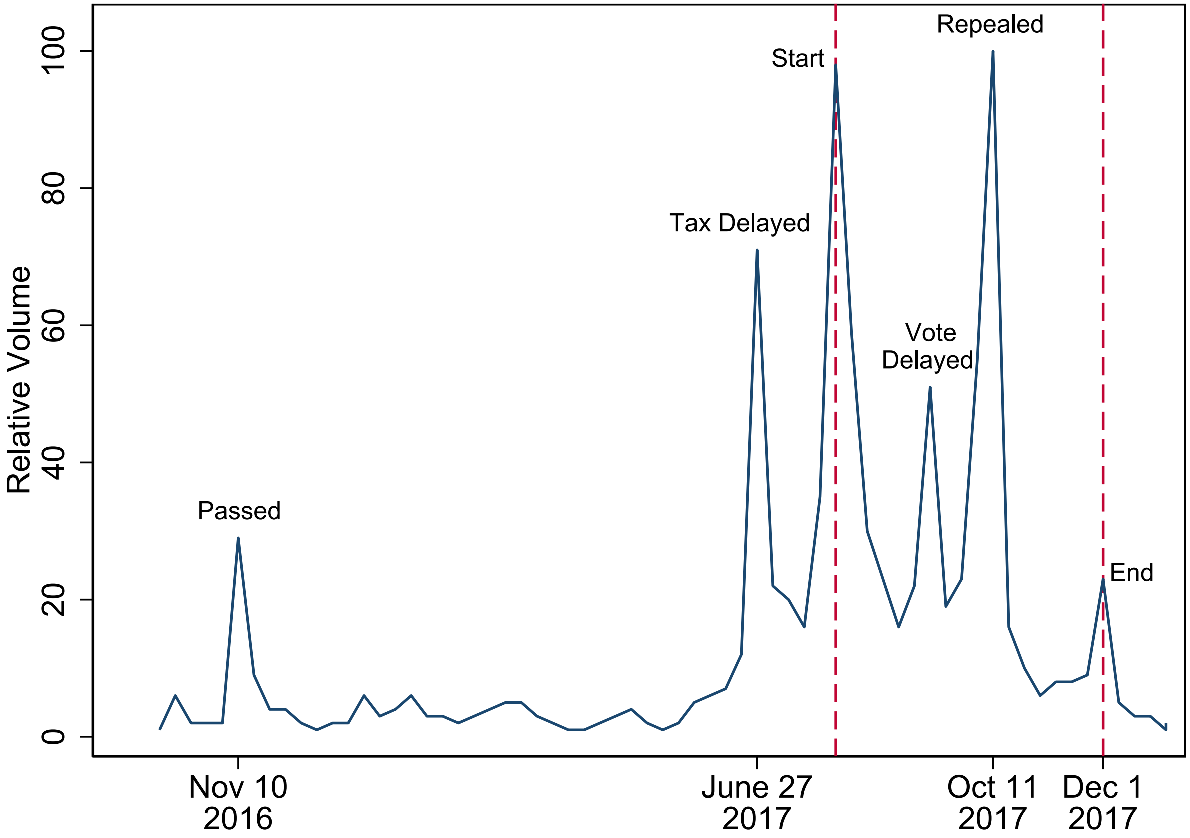

Figure 2.15: Relative weekly search volume for “soda tax”
This plot displays Google search volume in Illinois for the keywords “soda tax”. Each point is the volume in Illinois relative to the total search volume for that week in Illinois scaled so that the highest week’s search volume is 100.

Chapter 3

The Effect of the 2008 Economic Stimulus Payments on Nutrient Demand

3.1 Introduction

The US spends about $100 billion annually on programs designed to address food insecurity including the Supplemental Nutrition Assistance Program, Special Supplemental Nutrition Program for Women, Infants, and Children, and school meal programs. These programs serve tens of millions of Americans: one in four uses a program offered by the US Department of Agriculture (USDA) Food and Nutrition Service at least once in a year (of Agriculture Food & Service, n.d.). Additionally, food insecurity and poor nutrition are often cited by politicians as justification for transfer programs. Moreover, what we eat and drink plays a large role in health outcomes including obesity, which costs taxpayers north of $200 billion dollars annually (Cawley & Meyerhoefer, 2012).

Unconditional cash transfers (UCTs) have been proposed as ways to address poverty without spending resources on means testing (Hanna & Olken, 2018) or increase aggregate consumption thereby helping reducing the duration of a recession (Broda & Parker, 2014). However, little is known how UCTs affect nutritional choices in developed countries. In this paper, I examine one such UCT, economic stimulus payments (ESPs) dispersed as part of the Economic Stimulus Act of 2008 (hereafter ESA), on demand for nutrients. These ESPs were sent to households starting May 2008 in an effort to increase spending and reduce the duration of the recession caused by the 2007 financial crisis. In total, over $100 billion in ESPs were sent to 130 million taxpayers in 2008. Fortuitously, the IRS decided when each taxpayer would receive her ESP using the last two digits of her Social Security number, which is effectively random assigned.[1] The random timing of the ESP allows for identification of causal effects, which other authors such as Broda and Parker (2014) have exploited to estimate the impact of receiving an ESP on consumption. I use a similar approach to estimate the impact of receiving an ESP on demand for nutrients.

To estimate this effect, I calculate household-level nutrient purchases over time using data from the Nielsen Consumer Panel (NCP), which provides household-level purchase data at the

[1] Social Security applicants are assigned the last four digits of their Social Security numbers sequentially within their the geographic area and group, which determine the first seven digits of their number.

barcode level for about 60,000 households over several years. I merge these panel data with barcode-level Nutrition Facts data to observe quantities of each nutrient purchased per household over time. Finally, I observe when panelists receive their ESPs using a special survey constructed by Broda and Parker (2014) answered by NCP panelists before and during the period of time that ESPs were distributed.

I find that during the concurrent week the ESP is received, households increase total spending by 6% and only increase spending on food by a modest, statistically insignificant amount. However, households that do not have at least two months of income as liquid assets see large, significant effects on purchases of food. These households increase their purchases of all nutrients, but the effects are heterogeneous. Notably, carbohydrates and sugar increase more than other nutrients while protein and fiber increase the least. These findings are consistent with the Permanent Income Hypothesis and the ”wealthy hand-to-mouth” theory of Kaplan and Violante (2014a). Additionally, these findings suggest that one-time, large income shocks are unlikely to improve nutritional quality, at least for populations similar to the sample studied, or perhaps developed countries more broadly.

The remainder of the paper is organized as follows. Section 3.2 provides a background of the policies enacted and related literature. Sections 3.4 and 3.3 introduce the methods and data used, respectively. Section 3.5 summarizes our results. Section 3.6 discusses those results and concludes.

3.2 Background and Related Literature

In this section, I review the policy studied as well as the literature on the effects of ESPs in the United States and the demand for nutritious foods.

3.2.1 Economic Stimulus Act of 2008

First I provide background on the Economic Stimulus Act of 2008. The ESA was signed into law by President George W. Bush on February 13, 2008, which promised to provide ESPs to taxpayers who filed a 2007 tax return. Each eligible taxpayer received between $300 and $600 if filing single or twice that amount if filing jointly, depending on tax liability in 2007 (Congress, 2008). To be eligible for an ESP, one must have earned at least $3,000 in income in 2007 and have a sufficiently low income. Adjusted gross income beyond a threshold of $75,000 per qualifying adult phased out ESP benefits at a rate of 5%. Additionally, each dependent child increased one's ESP by $300. In total, the program disbursed about $100 billion in payments to 130 million taxpayers (Parker et al., 2013).

Although signed into law in February, taxpayers did not begin to receive ESPs until May. For those taxpayers who provided direct deposit information, the IRS distributed electronic transfers of ESPs during a three-week period starting in late April. For the remaining taxpayers, the IRS mailed paper checks during a nine-week period starting in early May through July. In advance of sending the ESP, the IRS mailed recipients a notice that an ESP would be arriving soon. The week a taxpayer received her ESP was determined by the last two digits of her Social Security number, which as mentioned earlier is assigned as good as randomly. Hence, the *timing* of receiving an ESP is random conditional on transfer method (electronic vs mail). It is important to note that the amount of the ESP is *not* random, nor is whether or not one receives an ESP. Hence, identification of causal effects in this paper comes from comparing those who receive their ESP early to those who receive their ESP later, regardless of amount, conditional on transfer method.

3.2.2 Unconditional cash transfers in the United States

Here I review papers that examine causal effects of UCTs in the US Johnson, Parker, and Souleles (2006) use the quasi-random timing of tax rebate checks sent to households as part of the Economic Growth and Tax Relief Reconciliation Act of 2001 to estimate the effects of income shocks on consumption. Using a unique set of questions added to the Consumer Expenditure Survey after passage of the 2001 Tax Act, they observe when households receive their payments. Similar to the ESPs distributed in 2008, the timing of payments in 2001 depended on one's SSN, and hence the timing of receipt is as good as randomly assigned. The authors find that households spent about two-thirds of their rebates over the six months following receipt on nondurable goods.

Broda and Parker (2014) conduct the survey of Nielsen Consumer Panel households that I use in this paper to observe if and when households receive their stimulus payments, how much they received, and some information about preferences and how the household intends to spend their payments. The authors find that households do not respond to news about the incoming payment but that aggregate spending rises by 10% the week after the payment is received. Additionally, spending remains higher for the three months following receipt of the payment, and the response is greatest for households that report low liquid wealth and low income. In another paper, Parker et al. (2013) add additional survey questions to the Consumer Expenditure Survey related to receipt of 2008 ESPs. The authors find that households spent between 50 - 90% of the payments, with about a third of that spent on nondurable goods.

Kaplan and Violante (2014a) construct a model that can explain the smaller consumption response in 2008 vs 2001, which they detail in Kaplan and Violante (2014b). The authors describe "wealthy hand-to-mouth" households who hold large amount of wealth mostly as illiquid assets. These households display large marginal propensities to consume upon receiving a positive income shock but not upon receiving news of the shock, which runs counter to conventional hypotheses that uses a one-asset framework and ignore the liquidity of such an asset. The key theoretical takeaway that I examine in this work is that liquidity constrained households, even those with significant savings, may experience consumption changes upon receiving an ESP.

3.2.3 Demand for nutrients

Here I review empirical work on demand for "healthy" (non-sugar carbohydrates, fiber, unsaturated fats) and "unhealthy" nutrients (saturated fats, sugar, sodium).[2]. Allcott et al. (2019) examine whether entrance of a supermarket to a "food desert" reduces nutritional inequality between wealthy and poorer households. The authors find that entrance of a new store does not significantly reduce nutritional inequality, which follows from the observation that households travel far to purchase groceries. A closer proximity store simply changes where the households make their

[2] *Healthy* in this context is guided by recommendations from US government agencies, such as the USDA's *Dietary Guidelines for Americans* 2020, which are guided by current average levels of nutrient consumption in the US.

purchases and modestly helps households through decreased transit costs and increased variety. The authors also find that moving to a healthier neighborhood does not make a large dent in nutritional inequality over several years following the move. Finally, the authors report that providing poorer households with the same prices available to higher income households would reduce nutritional inequality by at most 10% while the remaining 90% is driven by differences in demand. They suggest subsidizing healthy groceries, at a cost of 15% of the budget for SNAP, to eliminate nutritional inequality.

Other authors have provided alternative explanations for preferences for nutritious foods. Hut (2020) finds that migrants are mostly unaffected by local differences in demand for nutritious foods shortly after moving, but within three to four decades, about half of the difference in healthfulness is closed. Harding and Lovenheim (2017) use structural modeling to estimate the impact of nutrient-specific taxes on demand, which they estimate using data from the NCP. They find that a 20% tax on fat, sugar, and salt reduces demand for that nutrient by 30.25%, 16.41%, and 10.03%, respectively, all of which are more effective than are taxes on product classes (like sugar-sweetened beverages). Griffith, O'Connell, and Smith (2017) report that decreased salt intake observed in the UK is due to firms reformulating products and not because of consumers choosing products with less salt. In a recent paper, Harris-Lagoudakis (2020) finds that the introduction of an online shopping service modestly reduces the share of budget spent on sweets and candies but no evidence of improvement across other measures of healthfulness.

3.3 Data

I identify the causal effects of UCTs on nutrient demand using several data sources. I start with data from the Nielsen Homescan Survey, a nationally-representative panel that allows me to observe household-level food purchases and demographic data for about 60,000 households in 2008. Households in the panel are provided a barcode scanner by Nielsen, who asks households to scan all items purchased. To incentivize households to scan items, Nielsen offers a rewards catalog, which provides panelists higher value rewards the longer they remain in the sample.[3]

When scanning barcodes, Nielsen panelists are asked to provide information about the store where the product was purchased and the price of the item. Prices are automatically recorded for products purchased at a Nielsen partner retail outlet as the average price during the week the panelist purchased the product. Panelists are asked to manually input the price of products made at non-partner retail outlets. For barcodeless products like some produce and bakery items, Nielsen provides a reference booklet with barcodes associated with a photo and description of a product.[4] In practice, households may choose to omit reporting certain purchases or forget to scan and do not report all purchases (Einav, Leibtag, & Nevo, 2010). Hence I report coefficient estimates as percent changes. Provided the degree of underreporting in the homescan data is not correlated with ESP timing, the treatment effect estimates should not be biased by this measurement error of all purchases.

To observe nutrition information, I match the UPC codes associated with each Nielsen product with barcode-level nutrition data. I then collapse these data to observe total quantities of each nutrient purchased per household per week. The nutrition information come from three sources. The first source is from Syndigo, who license their nutrition dataset covering over 220,000 products containing information from the Nutrition Facts panel, ingredients list, and general product attributes like brand, size, and description.[5] The second source is the US Department of Agriculture's FoodData Central, which provides product-specific nutrient information. The final source is from images of products provided by major online retailers, from which I hand-record nutrition information from pictures of the Nutrition Facts panel.

I follow an imputation process similar to that of Dubois, Griffith, and Nevo (2014) to label each purchased product with nutrient information. I begin by dropping non-food products, alcohol,

[3] The median household stays in the sample for about seven years.

[4] "Reference card goods" are underreported in the Nielsen data, and Nielsen provides sampling weights for researchers who choose to include reference card goods in their analysis that upweight households that scan reference card items.

[5] Over 2,000 consumer applications have licensed these data, as have other researchers such as Dubois, Griffith, and Nevo (2014) who also match these data to Nielsen products.

tobacco, weight-loss/diet aids, and reference card goods.[6] I then match Nielsen products to Syndigo nutrients, covering about 62.7% of purchased products. I next match products without a direct match to those that share the same product description, brand, product module,[7], size type,[8] flavor, variety, type, formula, and style, which adds 22.2% to the matched data. After that, I allow matches across brands, including storebrand goods, which do not have matches in the Syndigo data. This step matches 8.5% of purchased products. The next step relaxes the flavor, variety, type, formula, and style restrictions, labeling 4.3% of products. I then impute with product module for ones that have sufficient labeled observations, which covers 2.1% of products. Finally, I manually label the remaining 0.3% of products using the USDA and online retailer data sources. After aggregating to the household-week level, I topcode each nutrient measure at the 99th percentile for all values greater than the 99th percentile to reduce the influence of outliers.

Finally, I use the survey constructed by Broda and Parker (2014) to observe when panelists received their ESPs. The survey was sent to all NCP households who met Nielsen's reporting requirement, who were offered rewards points to answer. The instructions asked that the survey be answered by "the adult most knowledgeable about your household's income tax returns". The first wave of the survey asked households for details about whether they had received a payment or if not, when they expected to receive a payment, if any. Subsequent waves followed up with households that had not yet received a payment but expected to in the future or were not sure if they would receive a payment. For households who reported receiving a payment, the survey asked for the amount and date the ESP was received as well as how the respondents' household anticipated spending or saving the money. The survey also asked all respondents a series of questions related to liquidity and savings. For a detailed description of the survey questions and timing, see the Appendix of Broda and Parker (2014).

Of the approximately 60,000 households receiving the survey, 80% responded, providing a

[6] Reference card goods do not have barcodes and are generally underreported by panelists. Additionally, they are reported as *counts* as opposed to weights, thereby making it difficult to label these products with correct nutrition information.

[7] Nielsen defines over 1,000 product categories called *product modules*.

[8] "Size type" is whether the product is measured in counts versus volume versus weight, which is generally (but not always) consistent within product module.

pre-trimmed sample of about 48,000. By necessity, I drop all households who report not receiving an ESP or omit a date of receipt, which removes about 20% of respondents. Additionally, I follow Broda and Parker (2014) in dropping households who report obviously wrong or inconsistent information. Specifically, I drop households who report receiving an ESP before ESPs of their type were distributed, those who report receiving an ESP at a date prior to an earlier survey wave in which they reported not receiving an ESP, and those who report receiving an ESP in a future date after the survey's timestamp. I also drop households who receive their ESPs later than the IRS-published disbursement schedules, which is possible but not random which households received their ESPs late. Finally, I drop households who do not make purchases both before and after receipt of their ESP, which is necessary to include household fixed effects. Very few households did not report purchases both before and after receipt, so this step likely has minimal impact on results. The resulting sample includes 19,961 households of which 9,190 received their ESP by direct deposit, 10,744 by mail, and the remaining 27 unsure.[9]

Clearly the sample remaining is not randomly selected. However, if the timing of receipt for this sample is random conditional on receipt type, estimated effects are unbiased, and the effects are representative of trimmed sample. There may be a few concerns that could lead to nonrandom timing of receipt. First, it is possible that households do not report their receipt dates correctly. Second, even if households report the dates correctly, we cannot observe when households were supposed to receive their payments. As such, it is possible that there is nonrandom selection where those reporting receiving their payments late, albeit still inside the permissible period, were supposed to receive their payments sooner. To give some confidence that households are reporting correctly, I plot the daily counts of reported receipt dates in Figure **??**. Interestingly, the spikes early in the distribution period line up with the distribution dates reported by the IRS for direct deposit. Hence, I proceed under the assumption that all households reported correct receipt dates, with the caveat that the reported dates are perhaps more accurate in the direct-deposit subsample. Furthermore, the inclusion window for mailed checks is much larger, which provides greater opportunity for

[9] For those households unsure of method of receipt, I used the union of eligibility dates across both methods of receipt when trimming the sample.

nonrandom selection of late recipients into the sample.

I present descriptive statistics of the selected sample in Tables 3.1 and 3.2, which may be helpful for deciding whether the results may generalize to other populations and grasping the magnitude of effects.

3.4 Methods

In this section, I describe the empirical strategies used in this paper.

3.4.1 Effect on payment on nutrient demand

I use a stacked event study design to estimate the impact of receiving an ESP on purchases. The primary specification is as follows:

$$y_{it} = \text{ESP}_{i(t)} + \alpha_i + \gamma_t + \varepsilon_{it} \tag{3.1}$$

where y_{it} is the total purchases of a food or nutrient y by household i during week t, α_i is a household fixed effect, γ_t are week fixed effects, and ε_{it} is an unobserved error term. $\text{ESP}_{i(t)}$ is a set of indicators for week since the household received its ESP, and I omit the week two weeks prior to when the household receives its check such that all coefficients are relative to that week.[10] I additionally drop a second relative-week indicator far in advance of treatment, which is required to avoid multicollinearity as discussed in Borusyak and Jaravel (2017).[11] Identification of the indicator dummies comes from timing of the ESPs, which is assigned as-good-as random conditional on ESP type (paper check or direct deposit) and receiving the payment when expected, as described in detail in Section 3.2.1. As such, I interact γ_t by method of payment indicators as well as estimate my primary specifications restricting the sample to only those who receive their ESP by the same

[10] I omit two weeks instead of the week immediately prior to allow for anticipatory effects.

[11] Doing so implicitly assumes the coefficient is zero, which is reasonable if there are no pretreatment trends. Another option is to not include household fixed effects.

method. Under unconfoundedness and SUTVA, the coefficients $\text{ESP}_{i(t)}$ for $i(t) \geq 0$ identify the (weighted) average treatment effect on the treated (ATT) of receiving an ESP on purchases of y after t weeks. I use the inverse hyperbolic sine of levels of y such that the coefficients can be interpreted as percent changes. When reporting average effects over multiple weeks, I average the relevant indicator coefficients.

In this setting, the *timing* of treatment is randomly assigned, and treated units remain "treated" for all following periods. Athey and Imbens (2021) show that causal effects in such a "staggered adoption design" can be estimated using the standard differences-in-differences (DID) estimator and interpreted as a weighted average of the average effect of changing the adoption date. The authors also prove that the clustered bootstrap provides a conservative estimate of the variance of the coefficients of interest. Sun and Abraham (2020) examine the event-study specification, which I use in this paper, and demonstrate that, even in settings where all units are treated, treatment effect estimates remain unbiased provided that the pattern of treatment effect does not change over time. I follow the guidance of that work as well as the principles shared in other recent papers (Callaway & Sant'Anna, 2020; De Chaisemartin & d'Haultfoeuille, 2020; Goodman-Bacon, 2018) when performing robustness checks. Of note, I do not include relative-time dummies beyond the last period for which I have at least one untreated group (set of households who have not yet received their impending ESPs). This limits my post-treatment observation period to 12 weeks with the full sample, 9 weeks with the paper-check sample, and 4 weeks with the direct deposit sample.

3.4.2 Heterogeneity

According to the Permanent Income Hypothesis, those with sufficient liquidity should not change their nutrient purchases dramatically upon receiving an ESP. To examine heterogeneous treatment effects of the ESP, I adjust 3.1 to include an interaction term:

$$y_{it} = \text{ESP}_{i(t)} + \beta_t \text{I}_i * \text{ESP}_{i(t)} + \alpha_i + \gamma_t + \varepsilon_{it} \tag{3.2}$$

where I_i is an indicator equal to 1 when a household has a time-invariant characteristic of interest, such as having at least two months of income saved in liquid assets. Note that time-invariant, in this setting, requires that the characteristic be constant during the periods studied, which is assumed to be true for all (annually-updated) Nielsen-provided demographic information. β_t, the parameter of interest in this specification, is the average difference in the outcome variable during week t for household who have the time-invariant characteristic of interest.

3.5 Results

In this section I present results using the methodology discussed in Section 3.4 estimated using data described in Section 3.3.

3.5.1 Effect on panelist behaviors

I first estimate the effects of receiving an ESP on panelist behaviors, such as number of trips taken, items scanned, and total amount spent using Equation 3.1. I present my findings in Table 3.3 and select event study coefficient plots in Figure **??**. In general, I find small, often insignificant effects in the first few weeks after receipt, and zero effect beyond that. Specifically, like Broda and Parker (2014), I find significant effects on total spending in the first week after receipt; however, I find more modest effects of about 6%. I find insignificant and close to zero effects on total amount spent on scanned items, scanned food, and quantities of both. Finally, I find a tight zero effect on number of trips, suggesting that on average, the ESP did not increase the number of trips taken by panelists.

I second estimate Equation 3.2 using liquidity constraints as the interaction variable of interest. Specifically, I can observe whether a household reports having at least two months of income accessible in liquid assets. I report results in Table 3.5 and event study relative-time coefficients in the left panel of Figure **??**. Interestingly, all outcomes of interest are statistically significantly significant and large. Compared to high liquidity households, low liquidity households increased their total expenditures by 17% on all purchases, 15% on scanned goods, and 12% on food items. These households purchased about 10% more items and went on 5% more trips in the first week. By the second week, the effects fade somewhat but remain significant for total expenditures and trips. The results are driven by direct deposit panelists, though the results are directionally positive for check-recipients as well.

3.5.2 Effect on nutrient demand

Next I examine the effects of receiving an ESP on nutrients purchased. The procedure is similar to that discussed in the previous section with panelist behaviors: I start by estimating the average effect across the sample and then estimate the difference in treatment effects for the low-liquidity subsample.

The story is similar for nutrients. When estimating treatment effects across the entire sample, as can be seen in Table 3.4 and Figure **??**, the average effects are small and insignificant. However, when including an interaction term for low liquidity households, the average effects are nearly all significant and large during the week of receipt, as can be seen in Table 3.6. Notably, calories increase by 19%, which can be explained by increases in fat, carbohydrates, and protein of 11%, 16%, and 13%, respectively. Alcohol (volume) purchases increase by about 10%, which may carry externalities that could be nontrivial in the aggregate. By the second week, the average effects have tapered off and are often insignificant, though still directionally positive and often large.

3.6 Discussion and Conclusion

While conditional cash transfers in the U.S. are well studied, there is limited research on the effects of unconditional cash transfers. Given that ESPs can be sent quickly to help reverse contractionary economic periods, political leaders are increasingly talking about Universal Basic Income, and ESPs have recently been deployed during the COVID-19 epidemic, it is important to understand how ESPs affect consumer decisions that have downstream health consequences.

Little is known about how ESPs affect nutritional choices, which is troubling given the large public investment in programs that assist nutritional choices and costs incurred by nutrition-related health complications.

In this paper I exploit the ESPs sent as part of the Economic Stimulus Act of 2008, which was designed to increase consumer spending and help reduce the downturn from the 2007 financial crisis. This policy delivered payments to US taxpayers at quasi-random delivery times, allowing for identification of causal effects using the exogenous timing. In the sample studied, these payments averaged $900 and arrived between late April and late July 2008, allowing me to identify the effect of receiving an ESP up to 12 weeks early. In general, I find evidence that the average effects on nutrient purchases were modest, which is consistent with the Permanent Income Hypothesis. However, households with low amounts of liquid assets had large and significant increases of purchases of all nutrients, but particularly those the USDA has encouraged reducing consumption. Collectively, I interpret this as evidence that one-time, large income shocks are unlikely to improve diet healthfulness in the short run and may even reduce healthfulness. However, one may view this optimistically as evidence that households splurge on special occasions (e.g. receiving a large payment from the government) and may otherwise follow a comparatively healthier diet.

3.6.1 Limitations

This paper is primarily limited by data concerns, particularly reliance on self-reporting of ESP receipt date. Besides containing measurement error, those who receive their payments later in the cycle (by factors other than random chance) may have unobservable factors that confound the treatment effect estimates. I present some evidence that these populations are likely small, at least among those reporting direct deposit payments; alas, this error is not observable. Second, the results are specific to the time period studied, namely the beginning of the worst economic recession in recent history. It is possible that the results may look very different if, for example, payments were made during an expansionary period. Third, the estimates are specific to a one-time transfer. Further research is warranted for estimating the impact of repeat transfers on household nutrition, which would help policymakers predict nutrition-related health impacts of transfer programs like universal basic income.

3.7 Acknowledgements

I thank Jeffrey Clemens, Gordon Dahl, Melissa Famulari, Craig McIntosh, Katherine Meckel, and numerous seminar participants for their comments and suggestions. My analyses are calculated based in part on data from Nielsen Consumer LLC and marketing databases provided through the NielsenIQ Datasets at the Kilts Center for Marketing Data Center at The University of Chicago Booth School of Business. The conclusions drawn from the NielsenIQ data are those of my own and do not reflect the views of NielsenIQ. NielsenIQ is not responsible for, had no role in, and was not involved in analyzing and preparing the results reported herein.

Chapter 3 is currently being prepared for submission for publication of the material. Goodman, Zachary A. The dissertation/thesis author was the primary investigator and author of this material.

Bibliography

Allcott, H., Diamond, R., Dubé, J.-P., Handbury, J., Rahkovsky, I., & Schnell, M. (2019). Food deserts and the causes of nutritional inequality. *The Quarterly Journal of Economics*, *134*(4), 1793–1844.

Athey, S., & Imbens, G. W. (2021). Design-based analysis in difference-in-differences settings with staggered adoption. *Journal of Econometrics*.

Borusyak, K., & Jaravel, X. (2017). Revisiting event study designs. *Available at SSRN 2826228*.

Broda, C., & Parker, J. A. (2014). The economic stimulus payments of 2008 and the aggregate demand for consumption. *Journal of Monetary Economics*, *68*, S20–S36.

Callaway, B., & Sant'Anna, P. H. (2020). Difference-in-differences with multiple time periods. *Journal of Econometrics*.

Cawley, J., & Meyerhoefer, C. (2012). The medical care costs of obesity: An instrumental variables approach. *Journal of health economics*, *31*(1), 219–230.

Congress, U. H. 1. (2008). H.r.5140 - economic stimulus act of 2008.

De Chaisemartin, C., & d'Haultfoeuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, *110*(9), 2964–96.

Dubois, P., Griffith, R., & Nevo, A. (2014). Do prices and attributes explain international differences in food purchases? *American Economic Review*, *104*(3), 832–67.

Einav, L., Leibtag, E., & Nevo, A. (2010). Recording discrepancies in nielsen homescan data: Are they present and do they matter? *QME*, *8*(2), 207–239.

Goodman-Bacon, A. (2018). *Difference-in-differences with variation in treatment timing* (tech. rep.). National Bureau of Economic Research.

Griffith, R., O'Connell, M., & Smith, K. (2017). The importance of product reformulation versus consumer choice in improving diet quality. *Economica*, *84*(333), 34–53.

Hanna, R., & Olken, B. A. (2018). Universal basic incomes versus targeted transfers: Anti-poverty programs in developing countries. *Journal of Economic Perspectives*, *32*(4), 201–26.

Harding, M., & Lovenheim, M. (2017). The effect of prices on nutrition: Comparing the impact of product-and nutrient-specific taxes. *Journal of Health Economics*, *53*, 53–71.

Harris-Lagoudakis, K. (2020). Online shopping and the healthfulness of grocery purchases. *Working paper*.

Hut, S. (2020). Determinants of dietary choice in the us: Evidence from consumer migration. *Journal of Health Economics*, *72*, 102327.

Johnson, D. S., Parker, J. A., & Souleles, N. S. (2006). Household expenditure and the income tax rebates of 2001. *American Economic Review*, *96*(5), 1589–1610.

Kaplan, G., & Violante, G. L. (2014a). A model of the consumption response to fiscal stimulus payments. *Econometrica*, *82*(4), 1199–1239.

Kaplan, G., & Violante, G. L. (2014b). A tale of two stimulus payments: 2001 versus 2008. *American Economic Review*, *104*(5), 116–21.

of Agriculture, U. D., of Health, U. D., & Services, H. (2020). Dietary guidelines for americans 2020-2025 [Accessed May 2021]. http://www.dietaryguidelines.gov/

of Agriculture Food, U. D., & Service, N. (n.d.). Fns nutrition programs [Accessed May 2021]. https://www.fns.usda.gov/programs

Parker, J. A., Souleles, N. S., Johnson, D. S., & McClelland, R. (2013). Consumer spending and the economic stimulus payments of 2008. *American Economic Review*, *103*(6), 2530–53.

Sun, L., & Abraham, S. (2020). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*.

Table 3.1: Summary statistics, categorical variables

	At least two months income in liquid assets		
	Illiquid	Liquid	Total
	%	%	%
Hours employment/week of male head of HH			
No head of this gender	28.05	26.76	27.22
< 30	2.45	2.92	2.76
30 - 34	1.89	1.58	1.69
≥ 35	49.37	37.96	42.04
Not employed	18.23	30.77	26.29
Hours employment/week of female head of HH			
No head of this gender	10.57	13.66	12.55
< 30	11.76	9.72	10.45
30 - 34	4.64	3.82	4.11
≥ 35	38.87	30.42	33.45
Not employed	34.16	42.37	39.44
Racial identity			
White	79.54	84.01	82.41
Black	12.18	7.64	9.26
Asian	1.64	3.59	2.89
Other	6.64	4.76	5.44
Household income			
<$35K	44.70	33.26	37.35
$35K - $59,999	28.87	29.27	29.13
$60K - $99,999	23.32	31.03	28.27
>$100K	3.11	6.44	5.25
Age of the (female) head of household			
<35	17.71	8.59	11.85
35 - 49	39.88	22.92	28.99
50-64	32.40	37.17	35.46
65+	10.01	31.31	23.70
Education of the (female) head of household			
< HS	4.50	3.27	3.71
HS Grad	34.64	33.01	33.59
Some College	33.28	29.22	30.67
Bachelor’s+	27.58	34.51	32.03
Any children < 18			
No	60.78	80.36	73.36
Yes	39.22	19.64	26.64
WIC indicator			
No	74.24	92.25	85.81
Current	3.34	0.80	1.70
Previously	22.42	6.96	12.48
Households	7,136	12,825	19,961

Table 3.2: Summary statistics, numeric variables

	Illiquid mean	sd	Liquid mean	sd	Total mean	sd
ESP by check	0.45	0.50	0.54	0.50	0.50	0.50
ESP by dir. dep.	0.54	0.50	0.46	0.50	0.49	0.50
ESP amount	906.76	555.96	864.12	516.83	881.88	533.89
Trips	3.12	1.98	2.95	1.72	3.02	1.83
Total spent	129.75	74.14	117.66	69.95	122.70	71.97
Total spent, scanned items	93.35	56.86	83.24	51.83	87.45	54.21
Total spent, scanned food	46.44	25.88	41.77	23.67	43.72	24.72
Scanned items	33.08	17.98	28.50	16.06	30.40	17.03
Storebrand items	5.67	4.87	4.65	3.98	5.08	4.40
Food items	23.64	13.13	20.50	11.62	21.80	12.37
Items with deals	6.82	9.39	8.06	9.71	7.55	9.60
Items with coupons	1.53	2.79	1.62	2.75	1.58	2.77
Coupon value	2.50	4.92	2.67	5.02	2.60	4.98
Calories	3920.57	2400.06	3146.84	2021.15	3469.10	2219.98
Calories from fat	996.14	582.96	816.46	506.29	891.30	546.77
Carbohydrates	652.20	448.04	510.45	365.33	569.49	407.88
Cholesterol	0.14	0.09	0.12	0.09	0.13	0.09
Fat	110.05	64.58	90.15	56.01	98.44	60.53
Fiber	31.34	17.80	28.47	16.54	29.66	17.13
Protein	98.46	55.49	86.62	50.39	91.55	52.90
Saturated fat	42.66	25.61	34.93	22.26	38.15	24.02
Sodium	6.14	3.55	5.09	3.10	5.53	3.33
Sugar	423.01	354.94	321.46	276.92	363.76	315.79
Transfat	1.44	1.18	1.04	0.94	1.21	1.07
Households	7,136		12,825		19,961	

Note: Variables are at the household-week level. Nutrient variables are all in grams, except for calories. Finanical variables are all in nominal dollars. Statistics are weighted using each household's projection factor as provided by Nielsen.

Table 3.3: Effects of ESP receipt on behaviors

	First week			Two weeks		
Outcome	All	Dir. Dep.	Check	All	Dir. Dep.	Check
Total spent	0.058*	0.074	0.044	0.067*	0.041	0.066
	(0.034)	(0.066)	(0.042)	(0.036)	(0.082)	(0.044)
Total spent, scanned items	0.024	0.018	0.024	0.036	-0.005	0.042
	(0.032)	(0.063)	(0.039)	(0.034)	(0.077)	(0.041)
Total spent, scanned food	0.007	0.022	0.010	0.015	0.006	0.014
	(0.030)	(0.058)	(0.038)	(0.032)	(0.071)	(0.040)
Trips	0.010	-0.003	0.009	0.012	-0.016	0.016
	(0.012)	(0.023)	(0.015)	(0.012)	(0.029)	(0.015)
Scanned items	0.016	0.024	0.014	0.021	-0.002	0.023
	(0.026)	(0.050)	(0.033)	(0.028)	(0.061)	(0.034)
Scanned food items	0.010	0.014	0.016	0.012	-0.010	0.017
	(0.026)	(0.049)	(0.033)	(0.027)	(0.060)	(0.034)
Households	19,961	9,190	10,744	19,961	9,190	10,744

Note: All specifications include household and calendar-week-by-method-of-payment fixed effects. Standard errors in parentheses are clustered at household level. ***, **, and * indicate significance at the 1, 5, and 10 percent critical levels, respectively.

Table 3.4: Effects of ESP receipt on nutrients

	First week			Two weeks		
Outcome	All	Dir. Dep.	Check	All	Dir. Dep.	Check
Calories	0.005	0.047	0.009	0.025	0.009	0.027
	(0.056)	(0.108)	(0.070)	(0.060)	(0.135)	(0.074)
Fats	0.001	0.033	-0.002	0.006	-0.009	0.002
	(0.036)	(0.068)	(0.045)	(0.038)	(0.084)	(0.047)
Saturated fat	0.019	0.050	0.015	0.020	0.014	0.014
	(0.030)	(0.057)	(0.038)	(0.032)	(0.071)	(0.040)
Trans fat	-0.001	-0.018	0.009	-0.005	-0.019	-0.000
	(0.014)	(0.027)	(0.018)	(0.015)	(0.031)	(0.018)
Carbohydrates	0.008	0.042	0.012	0.021	0.010	0.020
	(0.045)	(0.087)	(0.056)	(0.048)	(0.108)	(0.060)
Sugar	0.034	0.070	0.039	0.035	0.042	0.037
	(0.043)	(0.083)	(0.053)	(0.045)	(0.102)	(0.056)
Fiber	0.004	0.017	-0.004	0.009	-0.017	0.001
	(0.029)	(0.055)	(0.036)	(0.031)	(0.068)	(0.038)
Sodium	0.007	0.003	0.010	0.006	-0.020	0.010
	(0.020)	(0.037)	(0.025)	(0.021)	(0.045)	(0.026)
Protein	0.007	0.019	0.008	0.018	-0.014	0.020
	(0.036)	(0.068)	(0.044)	(0.038)	(0.083)	(0.047)
Alcohol	0.025	0.086*	0.002	0.018	0.057	0.020
	(0.025)	(0.048)	(0.032)	(0.026)	(0.056)	(0.033)
Cholesterol	0.003	0.005	0.001	0.002	0.002	0.002
	(0.003)	(0.005)	(0.003)	(0.003)	(0.006)	(0.003)
Households	19,961	9,190	10,744	19,961	9,190	10,744

Note: All specifications include household and calendar-week-by-method-of-payment fixed effects. Standard errors in parentheses are clustered at household level. ***, **, and * indicate significance at the 1, 5, and 10 percent critical levels, respectively.

Table 3.5: Effects of ESP receipt on behaviors, low liquidity households

	First week			Two weeks		
Outcome	All	Dir. Dep.	Check	All	Dir. Dep.	Check
Total spent	0.167***	0.254***	0.096	0.114***	0.194*	0.132
	(0.057)	(0.088)	(0.082)	(0.057)	(0.115)	(0.082)
Total spent, scanned items	0.151***	0.237***	0.089	0.106**	0.200*	0.115
	(0.054)	(0.084)	(0.077)	(0.054)	(0.108)	(0.077)
Total spent, scanned food	0.124***	0.193***	0.095	0.079	0.167	0.092
	(0.052)	(0.081)	(0.073)	(0.052)	(0.105)	(0.074)
Trips	0.051***	0.067***	0.039	0.035*	0.072*	0.047
	(0.020)	(0.031)	(0.029)	(0.020)	(0.041)	(0.029)
Scanned items	0.103***	0.160***	0.076	0.068	0.139	0.083
	(0.044)	(0.069)	(0.064)	(0.045)	(0.088)	(0.064)
Scanned food items	0.092***	0.158***	0.068	0.055	0.152*	0.065
	(0.045)	(0.070)	(0.063)	(0.045)	(0.089)	(0.064)
Households	19,961	9,190	10,744	19,961	9,190	10,744

Note: The first set of columns is β_1 from Equation 3.2 while the second set includes the average of β_1 and β_2. All specifications include household and calendar-week-by-method-of-payment fixed effects. Standard errors in parentheses are clustered at household level. ***, **, and * indicate significance at the 1, 5, and 10 percent critical levels, respectively.

Table 3.6: Effects of ESP receipt on nutrients, low liquidity households

	First week			Two weeks		
Outcome	All	Dir. Dep.	Check	All	Dir. Dep.	Check
Calories	0.186*	0.333***	0.125	0.123	0.296	0.137
	(0.097)	(0.152)	(0.136)	(0.097)	(0.197)	(0.137)
Fats	0.109*	0.206***	0.042	0.071	0.219*	0.054
	(0.062)	(0.096)	(0.088)	(0.062)	(0.124)	(0.089)
Saturated fat	0.103**	0.197***	0.043	0.066	0.203*	0.053
	(0.052)	(0.081)	(0.074)	(0.053)	(0.104)	(0.076)
Trans fat	0.051***	0.071*	0.063*	0.009	0.031	0.024
	(0.025)	(0.038)	(0.037)	(0.025)	(0.048)	(0.037)
Carbohydrates	0.163***	0.277***	0.136	0.112	0.249	0.142
	(0.078)	(0.123)	(0.110)	(0.079)	(0.159)	(0.112)
Sugar	0.160***	0.254***	0.152	0.116	0.252*	0.156
	(0.074)	(0.116)	(0.104)	(0.075)	(0.150)	(0.105)
Fiber	0.137***	0.196***	0.109	0.092*	0.202**	0.106
	(0.051)	(0.079)	(0.072)	(0.051)	(0.101)	(0.073)
Sodium	0.109***	0.156***	0.098**	0.060*	0.130*	0.068
	(0.035)	(0.054)	(0.049)	(0.035)	(0.068)	(0.050)
Protein	0.126***	0.215***	0.081	0.073	0.199	0.071
	(0.061)	(0.095)	(0.087)	(0.062)	(0.121)	(0.088)
Alcohol	0.096***	0.159***	0.087	0.080*	0.093	0.080
	(0.043)	(0.068)	(0.058)	(0.044)	(0.081)	(0.061)
Cholesterol	0.007	0.010	0.007	0.001	0.006	0.004
	(0.005)	(0.007)	(0.007)	(0.005)	(0.009)	(0.007)
Households	19,961	9,190	10,744	19,961	9,190	10,744

Note: The first set of columns is β_1 from Equation 3.2 while the second set includes the average of β_1 and β_2. All specifications include household and calendar-week-by-method-of-payment fixed effects. Standard errors in parentheses are clustered at household level. ***, **, and * indicate significance at the 1, 5, and 10 percent critical levels, respectively.

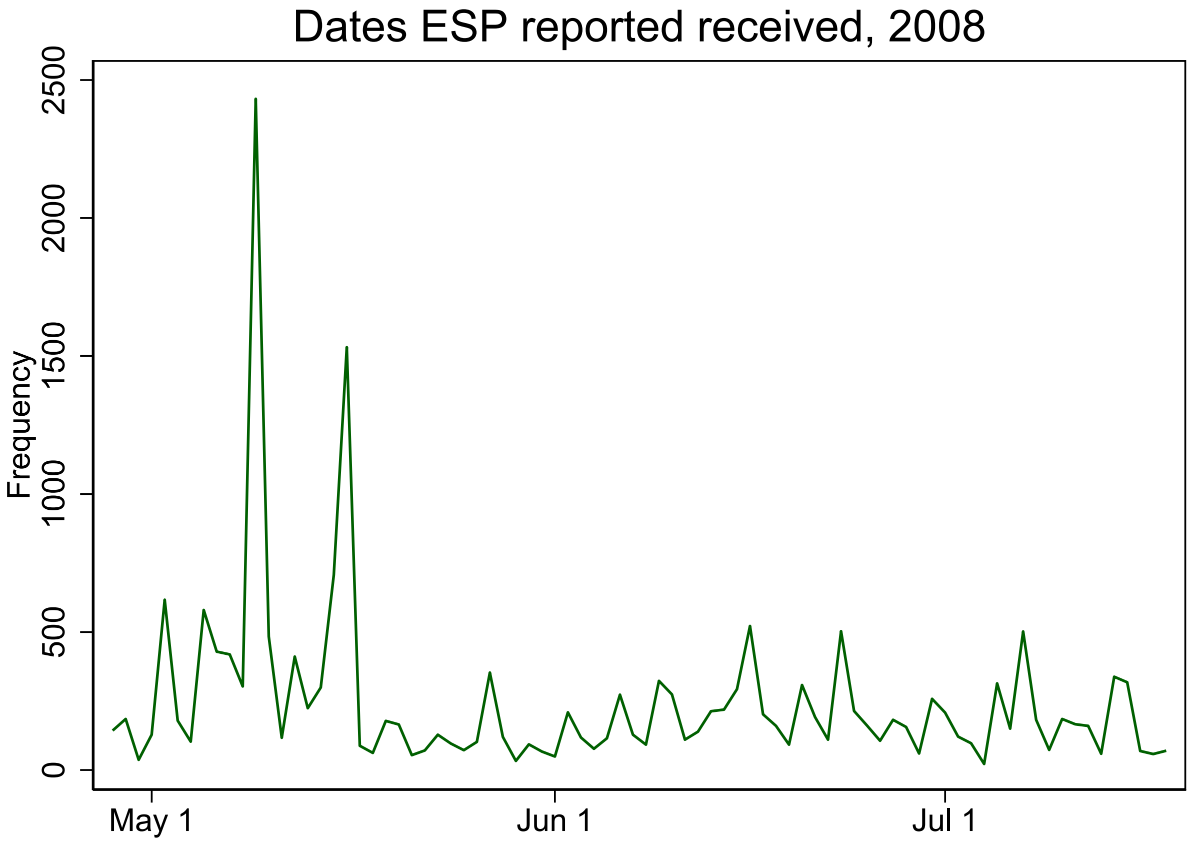

Figure 3.1: Dates ESP reported received

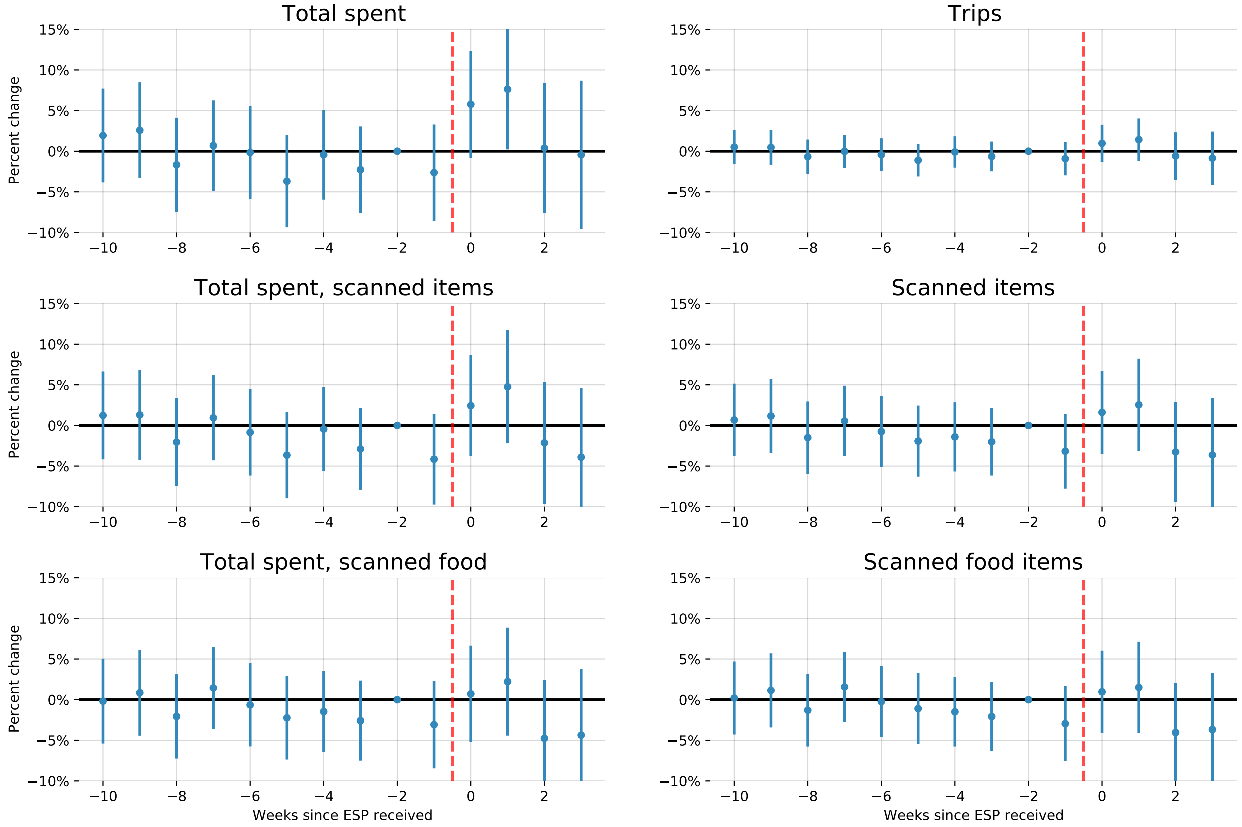

Figure 3.2: Effects of ESP receipt on shopping behaviors
Each panel displays the weekly percent differences for the given outcome variable fit using Equation 3.1. The regressions used include household and week-by-method-of-receipt fixed effects. Errors are clustered at the household level.

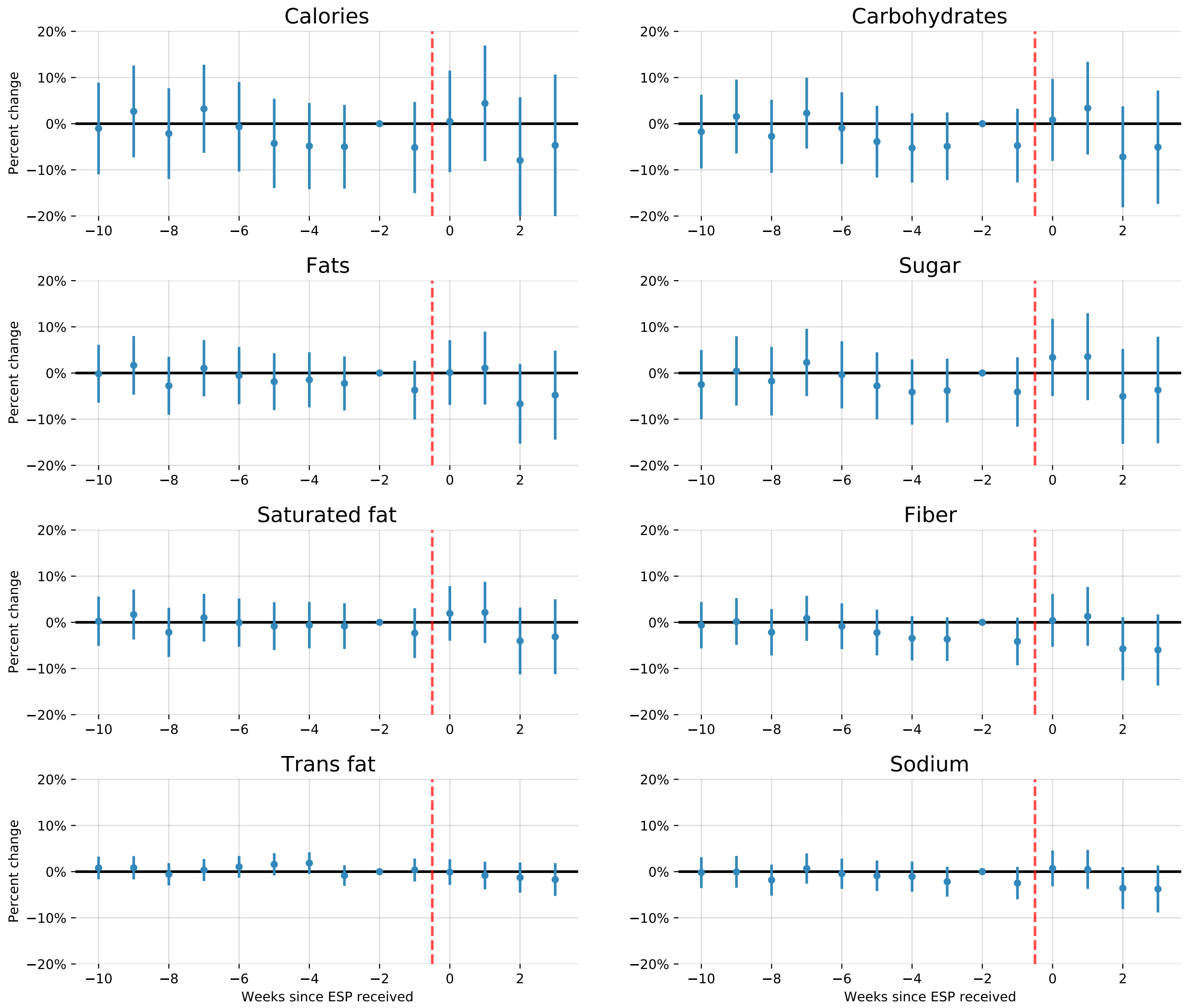

Figure 3.3: Effects of ESP receipt on nutrient purchases
Each panel displays the weekly percent differences for the given outcome variable fit using Equation 3.1. The regressions used include household and week-by-method-of-receipt fixed effects. Errors are clustered at the household level.

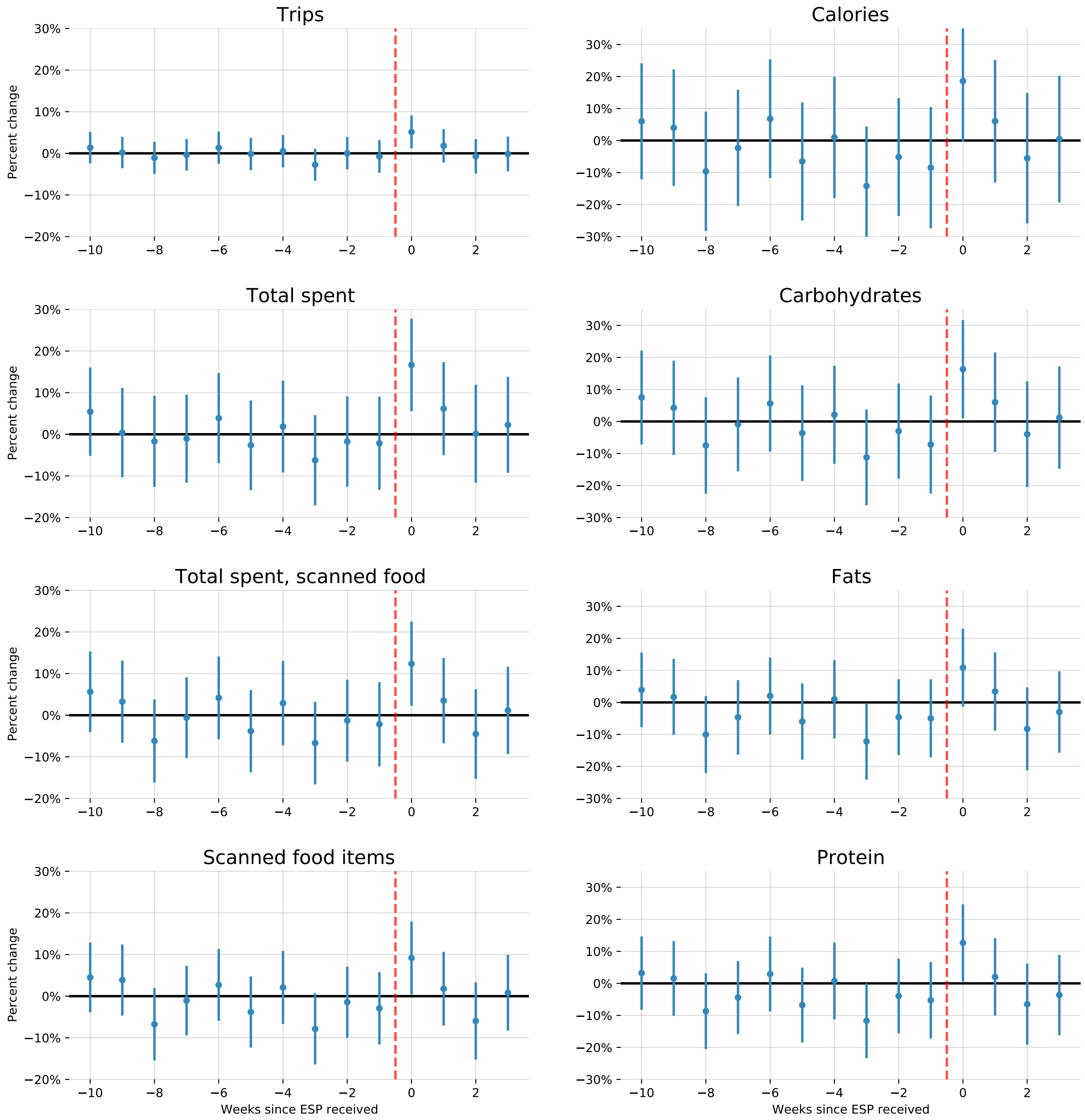

Figure 3.4: Difference in effects of ESP receipt for low liquidity households
Each panel displays β_t from Equation 3.2. The regressions used include household and week-by-method-of-receipt fixed effects. Errors are clustered at the household level.

3.8 Appendix

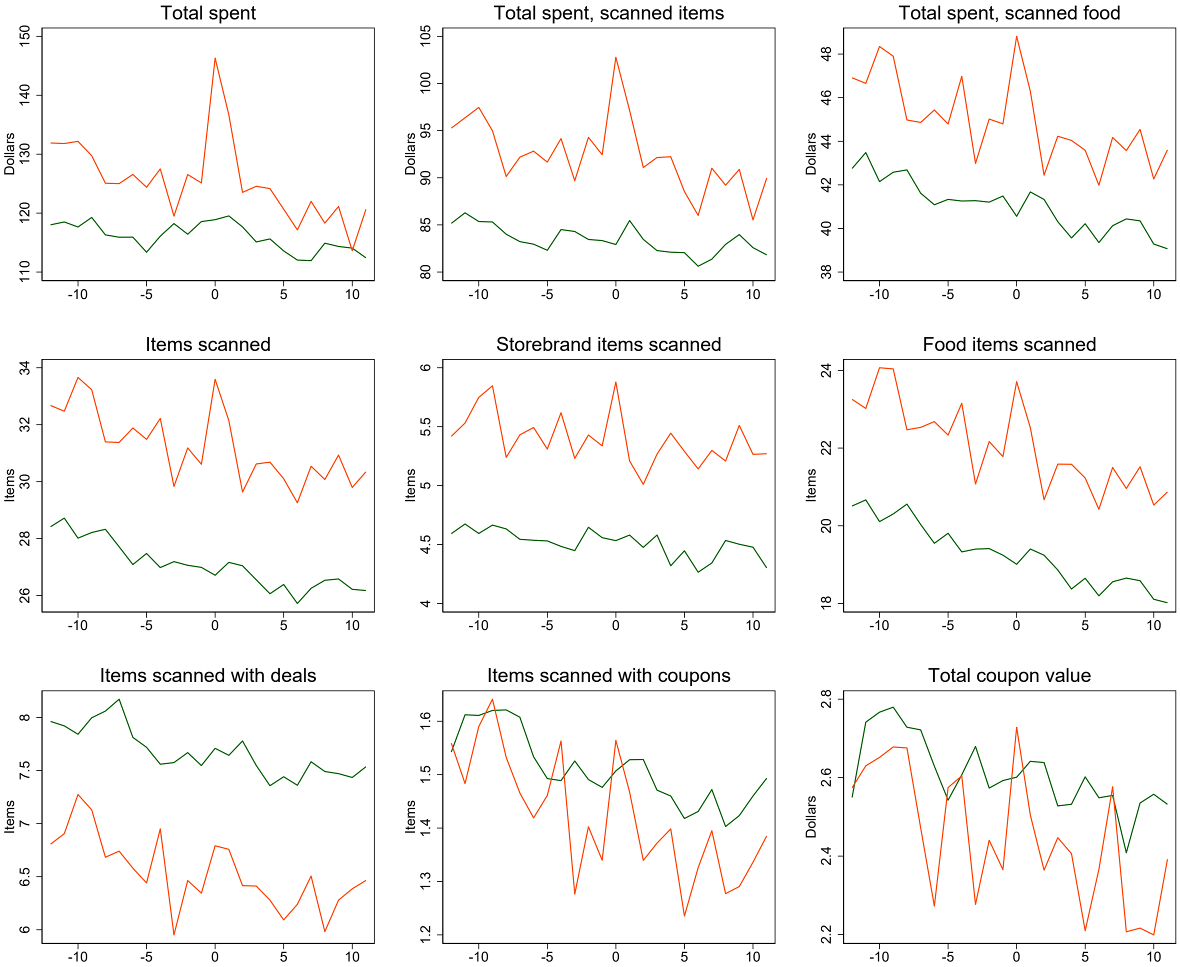

Figure 3.5: Nutrient purchases over time by liquidity constraints
Each panel displays weekly sums of different panelist behaviors averaged across panelists in the sample by whether they have at least two months of income in liquid assets. Orange (generally top) line is liquidity-constrained households, green (generally bottom) line is those with liquidity. Averages are weighted by panelist projection factors provided by Nielsen.

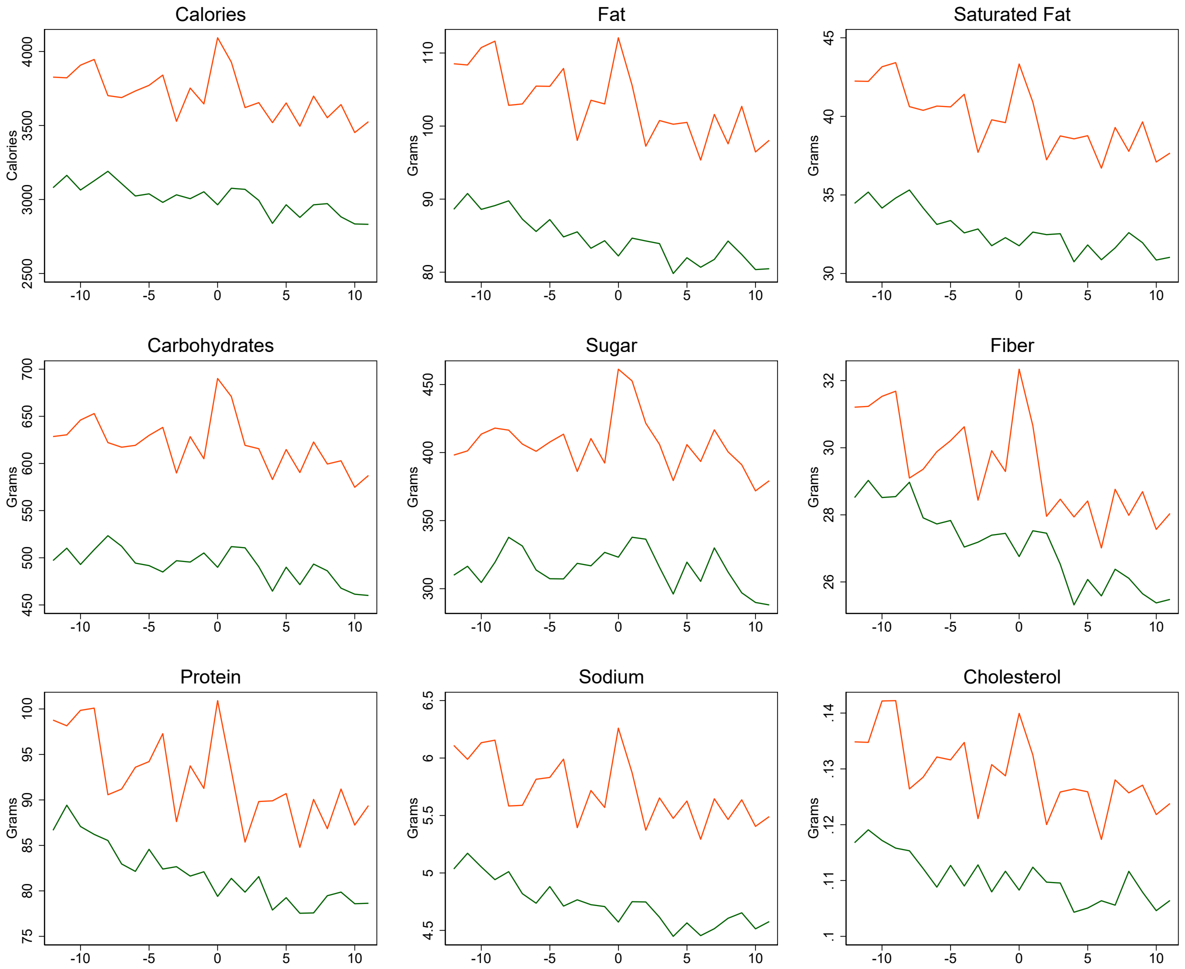

Figure 3.6: Nutrient purchases over time by liquidity constraints
Each panel displays weekly sums of different nutrients averaged across panelists in the sample by whether they have at least two months of income in liquid assets. Orange (top) line is liquidity-constrained households, green (bottom) line is those with liquidity. Averages are weighted by panelist projection factors provided by Nielsen.