ESSAYS ON ADVERTISING, SODIUM INTAKE, AND SODA TAXES

ESSAYS ON ADVERTISING, SODIUM INTAKE, AND SODA TAXES

A Dissertation Presented

by

EZGI CENGIZ

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

September 2022

Department of Resource Economics

ESSAYS ON ADVERTISING, SODIUM INTAKE, AND SODA TAXES

A Dissertation Presented

by

EZGI CENGIZ

Approved as to style and content by:

____

Christian Rojas, Chair

____

Emily Yucai Wang, Member

____

Arindrajit Dube, Member

____

John Spraggon, Department Chair
Department of Resource Economics

DEDICATION

To Doruk, Leyla, and Ege.

ACKNOWLEDGMENTS

First, and foremost, I would like to thank my doctoral dissertation committee chair and mentor, Dr. Christian Rojas. That I have arrived here—as a junior academic about to begin a tenure-track position at an R1 university, a position I could not even have dared to dream about couple of years ago—is because Christian believed in me. When I started seven years ago, I had little but a keen desire to learn—and Christian valued that. Christian's constant advice and guidance helped me grow to be the person I am now. My goal is to emulate him, to inspire those around me and display the kind encouragement, dedication, and support for my students that he has shown me.

I am grateful for the advice and feedback I received from my committee members, Drs. Emily Yucai Wang and Arindrajit Dube. Throughout my years in the PhD program, I have admired Emily as a talented teacher and a bright scholar. I am indebted to Arin, one of the greatest minds of our time in the econometrics community, for his feedback.

I am lucky to have spent my years as a doctoral student in the Department of Resource Economics at the University of Massachusetts Amherst. The Resec community made me feel seen and supported. I am especially thankful to Dr. John Spraggon for seeing beyond an unimpressive bachelor GPA and giving me a second chance in academic life by accepting me to the program. I am also thankful to Drs. Debi Mohapatra, Christoph Bauner, Nathalie Lavoie, Daniel Lass, Nathan Chan, and David Keiser for their encouragement and support.

My experience at the University of Massachusetts Amherst wouldn't have been the same without the friendship from Drs. (and Drs.-to-be) Ceren Goral, Osman Can

Icoz, Eren Odabasi, Emrah Ilker Ozay, Jayash Paudel, Palista Kharel, Hongli Wei, Abdul Kidwai, Yasemin Dildar, Cem Oyvat, Serkan Demirkilic, Anders Fremstad, Alper Yagci, Baris Guven, Hasan Comert, Hasan Tekguc, Ebru Dogan, Elvan Cavac, Jung Hwan Koh, Deniz Besik, Rodrigo Mercado Fernandez, Apollo Demirel, Ekin Koker, Selin Secil Akin, Juan Montecino, Jessica Carrick-Hagenbarth, Joao Paulo De Souza, Gamze Onut, Jose David Garces, Isabel Montes, Jeronimo Callejas, Felipe Quezada, Manuel Enrique Estay Montecinos, Emma Grazier, and Onupurba Das. My friends on the other side of the Atlantic helped me survive many hardships. Special thanks to Elif Duyulan, Nilay Celik, Toprak Toprak, Cemre Cakmakli, Damla Surmen, Pelin Incel, Nazli Kemik, Sena Abay, Ozlem Ozbas, Gamze Okutucu, Sinem Suslu, Sait Cakir, Serdar Cagnis, Berkin Ipek, and Canberk Kocak for their cheerful texts, emotional emails, and spontaneous calls.

I owe a great deal to a very special set of people who supported me so that I could stand on my feet. I thank Drs. Onur and Aysen Karayal for moral support (and durable kitchen utensils), Amy Bekkerman for excellent copy-editing of my doctoral dissertation, Sophie Williamson for clear guidance on numerous administrative matters, Dr. Gabriella Siegel for delivering my babies, Frankie Bryant for nannying my babies, Dr. Bige Deniz for being my babies' Aunt Bige, Ozlem Ataman for online kundalini yoga lessons, and Drs. Turkan Turan and Nihal Tuncer for writing recommendation letters for my acceptance to the program.

My parents, Aysun Guventepe and Erhan Aksar, have always showed me smiling and proud faces every week from miles away. I am fortunate to have Ozan Aksar as my brother and a role model for dedication to his work. I thank my in-laws, Drs. Melahat Cengiz and Celalettin Cengiz, and my sister-in-law, Deniz Cengiz, for years-long support.

Last, but not least, I thank my better half, Doruk Cengiz, who always lifts me up so I feel really tall. None of my papers in this dissertation would have come to

fruition without Doruk's intellectual curiosity and willingness to help. Doruk, Leyla, and Ege always knew how to put a smile on my face, even when I was extra grumpy with loads of work (or laundry). My heart is so full.

ABSTRACT

ESSAYS ON ADVERTISING, SODIUM INTAKE, AND SODA TAXES

SEPTEMBER 2022

EZGI CENGIZ

B.A., ISTANBUL UNIVERSITY

M.Sc., UNIVERSITY OF MASSACHUSETTS AMHERST

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Christian Rojas

This dissertation empirically explores the economics of advertising, sodium intake, and sweetened beverage taxes in food and beverages industries in the United States. I use recent advances in econometrics and statistics and provide causally interpretable results to deepen our understanding of how individuals respond to changes in advertising and public health policies in consumer packaged good industries.[1]

In the first chapter, titled "**Is traditional advertising effective? New evidence from mass-produced lager beer in the United States**," I present new evidence regarding the effectiveness of traditional advertising on product demand for mass-produced lager beer in the United States using data on weekly sales

[1]These analyses are 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 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.

and advertising for 122 brands in $7,906$ stores from 2011 to 2015. While the existing literature has examined between-industry heterogeneity or the effect of TV advertisements alone, I report within-industry heterogeneity in own advertising effectiveness and explore heterogeneity in all traditional advertising channels and among demographic groups (as opposed to pooled adults). The empirical exercises are based on a comprehensive sample of products and a credible research design. The type of within-industry heterogeneity I focus on is whether advertising effectiveness differs across light and non-light beer segments. In the baseline specification, the long-run elasticity of demand with respect to own traditional advertising is 0.030 for light and 0.015 for non-light mass-produced lager beer. This finding implies that doubling advertising expenditures would result in a 3.42 percent average revenue in the light segment but only a 1.66 percent increase in the non-light segment. I also find that print and outdoor advertisements are important contributors to light beer sales and to the differential in advertising effectiveness across the two segments. Surprisingly, TV advertising accounts for only 33 percent of the effectiveness differential between light and non-light beer. Finally, I find that light beer advertisements are more effective among a specific set of adults: non-ethnic, white, millennial, and urban dwellers.

In the second chapter, titled "**Sodium intake and reformulation in the United States: Evidence from scanner data**," co-authored with Christian Rojas, we construct detailed, barcode-level information on the near-universe of packaged food products to assess the impacts of the 2009 National Salt Reduction Initiative (NSRI), a voluntary pledge to reduce sodium content in food products by 2014. First, we isolate and quantify the role that product reformulation, vis-à-vis consumer purchasing behavior, played in recently documented declines in sodium intake in the United States. Second, we evaluate whether the NSRI played a causal role in reformulation efforts to reduce sodium content. We find that reformulation efforts alone would have contributed to a 53 percent reduction in sodium intake between

2007 and 2015; however, changes in consumer shopping behavior negated almost the entire improvement, permitting only a 4.73 percent net decrease in sodium intake. Using a kink regression discontinuity design, we show that the NSRI played a causal role in the observed reformulation efforts: On average, products with sodium content above NSRI targets decreased excess sodium by approximately 40 percent. We also document nutritional inequality (as it pertains to sodium intake) across socioeconomic and demographic groups and discuss the implications of our findings for effective diet improvement policies.

In the third chapter, titled **"How much do people consume after a soda tax? Evidence from the Philadelphia, PA, tax event,"** I compile a comprehensive household-level sample and show empirical evidence of changes in price and barcoded beverage purchases as a result of the Philadelphia, PA, sweetened beverage tax. The longitudinal panel includes 24,793 households and their purchases of 3,729 barcoded beverages across 11,994 stores in the City of Philadelphia and neighboring localities from 2006 to 2017. To construct a counterfactual Philadelphia, I use both traditional difference-in-differences models and frontier causal machine learning methods, as in Athey et al. (2021). Results show that 64 percent of the tax is passed on to consumers (as opposed to the 97 percent pass-through rate reported in Seiler, Tuchman and Yao, 2021) and that households reduced their consumption by 21 percent–39 percent, depending on the estimator. I also find evidence that consumers avoid the tax via cross-border shopping, although the magnitude is one-third of that reported by Seiler, Tuchman and Yao (2021): 8 vs. 24 percentage points.

TABLE OF CONTENTS

---

[2]This chapter is a joint work with Christian Rojas.

LIST OF TABLES

LIST OF FIGURES

Figure **Page**

CHAPTER 1

IS TRADITIONAL ADVERTISING EFFECTIVE? NEW EVIDENCE FROM MASS-PRODUCED LAGER BEER IN THE UNITED STATES

1.1 Introduction

In 2020, firms in the United States spent $242 billion on advertising (1.2 percent of GDP, 2.7 percent of sales) and $90 billion on traditional advertising (37.2 percent of total media spending) despite the recession caused by the COVID-19 pandemic.[1] The main justification for these substantial expenses is that firms see advertising as a strategic tool to increase demand for their products. However, the impact of advertising on demand is contentious, and empirical evidence for advertising effectiveness is mixed. Some researchers have reported average own advertising elasticities of 0.10–0.20 (Henningsen, Heuke and Clement, 2011; Sethuraman, Tellis and Briesch, 2011), while others report more limited effects, only 0.01–0.02 (Shapiro, 2018; Shapiro, Hitsch and Tuchman, 2021; Tuchman, 2019).

The order of magnitude of this difference is intriguing. Earlier studies generally assumed that advertising *must* be effective—since firms spend large sums of money each year on advertisements—and so shifted their attention to the psychological aspects of advertising, examining *how* rather than *whether* advertising affects sales. However, the recent surge in data availability and advances in identification techniques have made it possible to quantify advertising effectiveness: *how much* of advertising

[1] See eMarketer (2021) for ad spending, Bureau of Economic Analysis (2021) for GDP, and Schonfeld and Associates Inc (2021) for advertising sales ratio.

expenses in fact results in larger revenues. As a result, recent literature has returned to this original question.

I use data on weekly sales and advertising expenditures for 122 brands in 7,906 stores from 2011 to 2015 to assess the differential impacts of advertising on the distribution of sales in the mass-produced lager beer (MPLB) industry in the United States.[2] I make three contributions to the existing literature: First, I present new empirical evidence on within-industry heterogeneity in traditional advertising effectiveness. Specifically, I show that a large elasticity differential exists between the light and non-light beer segments. Second, I decompose the average traditional advertising effects by individual media channels (TV, print, radio, and outdoor) and show that the elasticity differential would be substantially underestimated if effectiveness were measured based on TV advertisements alone. Third, I explore treatment effect heterogeneity in advertising among demographic groups. I find that the elasticity differential is more salient for a particular subset of beer consumers (non-ethnic, white, millennial, urban dwellers). For these empirical exercises, I construct comprehensive product-level data for frequently sold brands in up to 304 quasi-experiments in the borders of the top advertising markets over time. The richness of the sample allows for a fuller assessment of these advertising effects on product demand in each beer segment than has been available to date. I also use a credible research design that matches populations residing at the borders of neighboring advertising markets as good counterfactuals for one another. This design purges time-varying spatial confounders in an enriched triple difference model, controlling for additional time-varying heterogeneity on each side of an advertising market border,

[2] As detailed in Section 1.3.1, I chose MPLB because it represents nearly all sales and traditional advertising. This means that the changes in advertising spending will likely have the largest bite in this segment.

and therefore allows for more credibly causal estimates of advertising on sales than those found in previous literature.

Overall, I find substantial within-industry heterogeneity in traditional advertising in the MPLB sector. The long-run demand elasticity with respect to own traditional advertising in the light beer category (0.030) is twice as large as that in the non-light beer category (0.015) in the baseline specification. More specifically, I apply eight major specifications borrowed from the literature and find that the long-run own advertising elasticity for light beer (0.020–0.030) is always higher than that for non-light beer (0.015–0.019). The elasticity differential in the baseline specification (0.015) implies that firms could increase revenues by 3.42 percent in the light segment relative to a 1.66 percent increase in the non-light segment if advertising expenditures were doubled. All of these findings are statistically distinguishable from zero at the 1 percent significance level.[3]

To explore potential explanations for the effectiveness differential between the two beer types, I decompose own traditional advertising effectiveness into TV, print, radio, and outdoor (e.g., billboards, movie theatre, buses) media. I find that TV, print, and outdoor advertisements contribute to light beer sales significantly more than they contribute to non-light beer sales. Print advertising effectiveness for light beer is 0.006 compared to zero effectiveness for non-light beer. Outdoor advertisements are nearly three times more effective for light (0.0058) than for non-light beer (0.002). Interestingly, radio is the only traditional media channel that appears to be more effective for non-light beer (0.008); however, it is difficult to make a conclusive comparison as the average point estimate in the light beer segment is very noisy. On the other hand, TV advertising appears to explain only 33 percent of the total

[3] I also quantify rival advertising effects across brands within a beer type (light and non-light) and find significant positive spillover effects within each category.

elasticity differential between light and non-light beer. Nevertheless, TV is still the most effective medium for both light (0.0146) and non-light MPLB (0.0102).

I further explore the sources of this heterogeneity by investigating whether advertising effects, for each beer type, differ across populations. Relative to the findings for pooled adults (0.030 for light and 0.015 for non-light beer), regressions by demographic group show that advertisements for different types of beer have different levels of effectiveness for certain demographics, where the reported coefficients in the parentheses correspond to the effect of increasing the percentage share of a demographic by one standard deviation. In particular, light beer advertisements are more effective in areas with higher shares of millennials (0.039) and urban dwellers (0.037) and with smaller shares of African American (0.026), Asian (0.018), and Hispanic (0.022) populations. In contrast, non-light beer advertisements are less effective for millennials (0.011) but more effective for older generations (0.014 for baby boomers and 0.021 for Generation X) and women (0.016). These findings can prove useful as firms seek more effective ways to allocate advertising expenditures across locations.

My empirical approach for identification is to exploit the discontinuity in advertising exposure that exists across two bordering geographic regions (called designated market areas, or DMAs). More specifically, I specify and estimate an enriched triple difference model that takes the difference between the two difference-in-differences estimates through the inclusion of a battery of fixed effects that purge unobserved spatial confounders in the bordering markets. The identifying variation in advertising effects therefore comes from deviations from the average sales and advertising over weeks, stores, and brands. Intuitively, the research design relies on the assumption that differential trends in beer sales between the two opposite sides of a DMA border are the result of differences in advertising exposure.[4]

I estimate advertising effects across eight specifications that are frequently used in the empirical literature. Each specification can be categorized by three aspects. The first aspect pertains to how dynamics enter the model: To capture the long-run effects of advertising, I use either advertising stock or lagged sales. The second aspect considers how the direct and general (company-specific) advertisements are treated; one variant lumps both types of advertisements whereas another separately estimates their effects as in Shapiro, Hitsch and Tuchman (2021). The third aspect considers how bordering markets are defined. In the border region (BR) approach, I match all the stores that are at the opposite sides of an advertising market border; in the adjacent border county-pairs (ABCP) approach, I match only the stores that further share a county border.

I divide the MPLB industry into light and non-light beer segments rather than popular-priced, premium, and super-premium beer segments, a common practice adopted in previous beer studies. One reason for this segmentation is that light beer has been the most successful category introduction in the history of the industry. Light beer swiftly grew in popularity in the 1970s, capturing a large market share and expanding the market. The literature typically attributes the success of light beer to the spread of TV as the new medium to reach consumers and aggressive TV advertising by breweries, making the segmentation proposed in this research particularly relevant to study traditional advertising effects. Second, as a growing

[4] I build upon the empirical literature that relies on exogenous variation in the choice variable across neighboring border markets (Black, 1999, on the economic value of education; Card and Krueger, 1994, and Dube, Lester and Reich, 2010, on minimum wage effects; and Shapiro, 2018, and Tuchman, 2019 on advertising effects). Specifically, in addition to the identifying variation that comes from the common trend in the opposite sides of the same border, I also control for the unobserved demand shocks that are specific to the each side of the border with the inclusion of store-border-month fixed effects. See Section 1.4.2 for details.

proportion of the population seeks more nutritious diets, attributes such as low-calorie or light are now commonly used to segment products in other consumer packaged good (CPG) industries (e.g., carbonated beverages) thereby making my results more relevant. Last, the chosen segmentation is motivated by the divergent market structure that prevails in this industry. In particular, over the last two decades, the light beer segment has been characterized by higher (and persistent) market shares, concentration, and advertising outlays, while the non-light beer segment has become more fragmented and less advertising intensive. These divergent facts raise interesting questions for the role that advertising outlays and responsiveness can have on market structure, as in the endogenous sunk cost mechanism put forth by Sutton's (1991).

This study is related to a growing body of research that examines the effects of advertising on sales (see, e.g., Dubois, Griffith and O'Connell, 2017*a*; Shapiro, 2018; Shapiro, Hitsch and Tuchman, 2021; Tuchman, 2019). However, my approach to studying within-industry heterogeneity in advertising effectiveness is an important addition. Industry practitioners can craft more effective advertising strategies if credible identification of advertising effects is available for different segments within a broad product category. Second, the type of outlet matters. As I show in this study, print, radio, and outdoor ads (not included in prior studies) might be important determinants of sales. Last, compared to earlier literature (Shapiro, Hitsch and Tuchman, 2021), I examine the advertising effects for a wider range of brands and geographic markets; the advantage of this approach is that it alleviates potential concerns regarding the statistical power of the results while also providing a comprehensive assessment for a given industry.

Two caveats are in place. First, this paper does not measure or address digital advertising because of data limitations. Despite this limitation, traditional ads still constituted nearly 40 percent of total media advertising in 2020. Further, ad effectiveness in traditional media is expected to be particularly salient for popular products of everyday use (e.g., soft drinks, cereal, beer, yogurt), whereas digital media has an edge for its potential ability to target specific demographics, introduce new products, or grow niche markets. Second, this paper does not estimate a full-fledged structural model. Since the paper's focus is the causal measurement of a rich set of advertising effects, I emphasize a clear, convincing design-based identification strategy. As the empirical industrial organization literature acknowledges, a limitation of this approach is the inability to evaluate counterfactual scenarios in a broader set of environments (see, e.g., Nevo and Whinston, 2010).

The remainder of the paper is organized as follows: Section 1.2 reviews the literature on the advertising effects. Section 1.3, describes the data and sample construction and explains why the industry context of the MPLB segment provides an ideal setting to study the heterogeneous impacts of ad effectiveness on sales. Section 1.4 lays out the research design and presents a comprehensive list of potential threats to identification. In Section 1.5, I report the main results, examine their robustness, and discuss their implications. Section 2.7 concludes.

1.2 Literature review

Research on the economics of advertising expanded considerably in the second half of the twentieth century with the establishment of theoretical foundations for advertising's direct effects on sales and indirect effects on market structure and profits (see Bagwell, 2007, for a comprehensive survey). One group of studies primarily considered *how* advertising affects demand and laid out the psychological aspects of advertising (persuasive, informative, or complementary) on consumption. Another set of studies, motivated by the role of advertising on market structure à la Sutton (1991), argued that incumbent firms can escalate advertising and deter entry if advertising creates vertical differentiation. However, a relatively limited set of empirical studies has *quantified* advertising effectiveness or explored potential heterogeneity in effectiveness among product types in the same industry or among media types or demographic groups.

There are two main reasons for this gap in the empirical literature. First, the increased availability of detailed, brand-level datasets now allows researchers to measure ad effectiveness by matching advertisements to actual purchases in a finer way than was previously possible. For instance, the UPC-level scanner data used in this study for beer sales contain information on weekly store sales for millions of CPG products. Second, this increased data availability provides sufficient variation in variables of interest, which is crucial to credibly estimating the causal impact of advertising. More credible research designs, therefore, allow economists to more fully assess the policy effects of advertising using observational data and quasi-experiments (Angrist and Pischke, 2010).

Firms increase their advertising spending at times when and in places where sales are also likely to be higher. Indeed, to address the problem of unadjusted confounders in advertising effects, most of the marketing and economics literature to date has adopted an experimental approach. Earlier attempts in the marketing literature include split-run experiments in print media (Hovde, 1936; Waldman, 1956; West, 1939; Zubin and Peatman, 1945) and on television (Ackoff and Emshoff, 1975). Similarly, later literature (Aaker and Carman, 1982; Lodish et al., 1995) reported results from a large number of split-run advertising TV experiments. A common feature of this literature is that a large share of ads do not have a detectable effect.[5] The lack of an advertising effect in the Ackoff and Emshoff (1975) study, for example, led Anheuser-Busch InBev to drastically reduce its advertising. Importantly, these results should be interpreted with caution since there is insufficient detail to determine

[5]The only exception is Hu, Lodish and Krieger (2007), who, on average, detect a positive effect of advertising on sales.

(i) whether control and treatment groups were properly randomized or (ii) the quality of the statistical tests.

Recent literature in economics has also relied on the experimental method. Simester et al. (2009) randomly vary the number of catalogs that different consumers get and find a positive effect of advertising on sales. Lewis and Reiley (2014) run online advertising experiments using over 1,500,000 users at a major retailer and find modest (5 percent) but statistically significant effects of advertising exposure on profitability.

Even though the randomized controlled experiments are acknowledged to be the gold standard for causal inference in social sciences (Imbens and Rubin, 2015), the high cost associated with conducting these experiments has often led researchers to exploit naturally occurring variation in advertising when assessing the effects of advertising on sales. Bagwell (2007) reviews the literature that has looked at these types of correlations. A common finding in these studies is that advertising expenditures increase sales, although this result should be interpreted with caution given the aforementioned concerns about advertising endogeneity. Some studies using observational data have made important progress in identifying advertising effects. Tellis (1988), Tellis, Chandy and Thaivanich (2000), and Ackerberg (2001, 2003) exploit the panel structure of the data (advertising messages are displayed at high frequencies to the same individuals) for a more credible identification strategy. Ackerberg (2001, 2003) shows that the effect of advertising on sales is particularly salient for consumers who had not previously consumed the product. In the marketing literature, several meta-analyses (Assmus, Farley and Lehmann, 1984; Henningsen, Heuke and Clement, 2011; Sethuraman, Tellis and Briesch, 2011) report a distribution of short- and long-run elasticity estimates with averages as high as 0.2–0.3. Overall, the findings in these studies are consistent: Advertising is effective in increasing sales.

Using disaggregated and high-frequency data, four empirical questions for the advertising effects on sales have been posed in the literature. First, advertising might persuade current consumers of the brand or consumers who do not purchase any brand to increase their purchases of the advertised brand. Second, advertising might steal away consumers who are currently purchasing a rival brand. Third, advertising can have spillover effects when, say, consumers watch an advertisement from a brand that they do not favor, but the advertisement prompts them to purchase more of their desired brand. Fourth, there might be a carryover effect of advertising over time: Consumer might remember an advertisement that they have seen in the past when making a purchase today.

While the industrial organization literature has engaged in a theoretical debate regarding these effects, empirical work has been relatively scarce (Bagwell, 2007). A demand model that computes own and cross advertising elasticities and that allows for dynamic advertising can shed light on these important questions. Attempts to model cross advertising elasticities include Dubois, Griffith and O'Connell (2017*a*), Rojas and Peterson (2008), Shapiro (2018), Shapiro, Hitsch and Tuchman (2021), and Tuchman (2019). Dubé, Hitsch and Manchanda (2005), Dubois, Griffith and O'Connell (2017*a*), Erdem, Keane and Sun (2008), Shapiro (2018), and Shapiro, Hitsch and Tuchman (2021) consider models that include advertising stock in the demand estimation.

Sutton (1991) establishes a theoretical relationship between responsiveness to advertising (a sunk cost) and market structure. If demand is sufficiently responsive to advertising (or to other sunk costs such as R&D), then firms escalate their advertising expenditures, thereby generating an endogenous entry cost and a lower bound to concentration regardless of how large the market grows. Alternatively, if advertising is not very effective in increasing demand, larger markets can accommodate more firms and concentration decreases with market size. Several empirical studies

(Ellickson, 2007, in supermarkets; Gavazza, 2011, in mutual funds investment banking; and Bronnenberg, Dhar and Dubé, 2011, in a number of CPGs) have tested Sutton's predictions, and this paper adds to that literature, showing that advertising effectiveness differs by beer type in the off-premise retail beer industry.

This study is closest to a growing set of studies that examine the effects of advertising on sales using a similar identification strategy and similar data sources. Using Kantar Media Ad$pender and an identification strategy at the neighboring border markets, Shapiro (2018) reports advertising spillovers across rival brands in the antidepressants market. Tuchman (2019) measures the effect of e-cigarette advertising on combustible cigarettes using Nielsen scanner data and an identification strategy at the bordering advertising markets. Shapiro, Hitsch and Tuchman (2021) reports a distribution of own advertising effects from 288 brands across 70 CPG industries with the purpose of providing generalizable estimates. This paper adds to these studies by (i) demonstrating the existence of important within-industry heterogeneity in advertising; (ii) including print, outdoor, and radio in addition to TV ads; and (iii) considering a significantly more comprehensive range of brands and geographic markets.

1.3 Data and sample construction

In this section, I first explain why the U.S. mass-produced lager beer (MPLB) industry represents an ideal empirical setting in which to study heterogeneous advertising effectiveness. Next, I provide some context for the MPLB industry and a more contemporaneous industry profile. Last, I describe the data and document how I constructed the distinct samples used in the subsequent analyses.

1.3.1 Choice of industry

Beer has been the primary alcoholic drink of choice in the United States since the decades following the Civil War, with a 44 percent market share in 2020 (Stack, 2003; Statista, 2021). Despite significant losses in total beer sales in 2020 due to the COVID-19 pandemic, off-premise retail sales grew to $43.8 billion as consumers shifted their on-premise consumption to consumption at home (Beer Institute, 2021; Daily Beast, 2020; Brewbound, 2021). High and persistent levels of traditional advertising make the beer industry particularly suited to study advertising effects. In particular, changes in ad spending will have a larger bite in the MPLB industry than in any other beer segment because the MPLB industry is the *incumbent* segment of the beer industry, making up nearly 95 percent of all beer sales and traditional advertising (my analysis of the Nielsen RMS and Kantar Media Ad$pender from 2011–2015).

The recent surge in local and craft beer sales in the non-light beer segment has received attention because it sheds light on the relationship between market structure and a firm's choice variables, including price and advertising. However, local breweries often compete in product offerings, while mass-producing firms compete in price and advertising. Further, even though the combined volume of beer sales from craft breweries is reported to have been around 25.9 million barrels in 2018, nearly three times what it was a decade ago, Bud Light alone shipped more than 27 million barrels, making the success of craft beer seem less impressive (USA Today, 2020). Last, a majority of small-scale local breweries make most of their sales from on-site consumption. The pandemic therefore hit the craft sector particularly hard. This reversal of fortunes has helped cement the position of MPLB as the single largest segment in the U.S. retail beer market.[6] Quantifying advertising effects in this sector is therefore particularly relevant for firms and policy makers.

Studying the beer industry as a whole and the MPLB segment in particular provides greater opportunities for empirical research as these products are available in a large set of stores across the country. In this study, the causal identification of advertising effects depends on making comparisons across bordering markets, as I explain in detail in Section 1.4. A larger sample of stores ensures more useful variation in both sales and advertising in neighboring areas, which allows me to more credibly estimate impact and maintain statistical power. Similarly, industry assessment is most accurate if the major firms and their brands are represented in full. As noted in the literature, empirical results from a comprehensive sample are less likely to suffer from measurement error issues compared to those of limited scope.

Last, studying advertising effectiveness across beer types (e.g., light and non-light) makes it possible to isolate demand-side factors from potential supply-side confounders. For instance, Table 1.1 reports that one-third of macrobreweries produce both light and non-light beer, and only one company in the light segment covers just light beer (although it still produces both types of beer).[7] This suggests that both types of beer may share common supply-side factors. Indeed, both types of beer implement similar production technologies and similar distribution and marketing institutions. For example, to produce light beer, firms do not necessarily use different tanks or hops, but they often dilute non-light beer with water until the desired alcohol and calorie content are achieved. Similarly, it is hard to imagine why firms would

[6] Tremblay and Tremblay (2005) review the waves in the United States brewing industry in which the local brewers capture market shares from the mass-producing industry leaders. Over the past century, local breweries underwent a series of reversals of fortune, first with Prohibition, followed by the Great Depression and the rise of TV in the aftermath of the Second World War.

[7] The MPLB sample covers I.C. Light of Pittsburgh Brewing Company but excludes Iron City Beer, a non-light lager beer brand from the same company. See Section 1.3.4 for details on the criteria for selecting the MPLB brands.

transport products in different delivery vehicles and retail them in different stores. Ad buying practices over traditional media outlets also do not change depending on whether firms advertise for light or non-light beer. Rather, it is fairly common to see company or combo ads (what I refer to in this paper as general ads) in which both types of beer brands are advertised in a single advertisement.[8]

[8]This feature creates data challenges as the effect of these general ads on brand sales can be ambiguous, unlike the effect of focal brand ads on sales, which is assumed to be nonnegative. In Section 1.3.4, I detail the two alternative methods I used to account for the effect of general ads.

Table 1.1. Number of brands by firm and beer type, 2011–2015: light vs. non-light mass-produced lager beer

	Light	Non-light
ANHEUSER-BUSCH	12	23
MILLER COORS	6	20
HEINEKEN NV	3	5
PABST BREWING	2	6
CONSTELLATION BRANDS	1	5
KPS CAPITAL PARTNERS	1	2
BOSTON BEER	1	2
GAMBRINUS COMPANY	1	1
DG YUENGLING & SON BREWING	1	1
WILLIAM K BUSCH BREWING	1	1
PITTSBURGH BREWING COMPANY	1	0
ABITA BREWING COMPANY	0	4
CAPITAL BREWERY	0	3
NEW BELGIUM BREWING COMPANY	0	2
AUGUST SCHELL BREWING COMPANY	0	2
DIAGEO	0	1
NARRAGANSETT BEER	0	1
GREAT LAKES BREWING COMPANY	0	1
KONA BREWING COMPANY	0	1
STORE BRANDS	0	1
MULTIPLE VALUE	0	1
NEW GLARUS BREWING COMPANY	0	1
SESSION	0	1
LUCKY BUCKET BREWING COMPANY	0	1
ANCHOR BREWERS & DISTILLERS	0	1
BOULEVARD BREWING	0	1
FLORIDA ICE & FARM	0	1
AC GOLDEN BREWING	0	1
CRAFT BREWERS ALLIANCE	0	1
DEVILS BACKBONE BREWING COMPANY	0	1

Notes. Mass-produced lager beer (MPLB) brands are those with the top 95 percent cumulative shares and that have sold at least once in the top 90 percent stores in each DMA-year, where DMA denotes designated market areas or local advertising markets. Store brands, session (non-light lager beer with no more than 5 percent ABV), and multiple value are never-advertised, combined non-light beer subsegments. The sample covers 11 breweries with light beer and 29 breweries (including store brands, session, and multiple value) producing non-light beer and 10 breweries producing both light and non-light beer, for a total of 122 MPLB brands with 30 light and 92 non-light beer. 2011–2015 Nielsen RMS and Kantar Media Ad$pender.

Since the same scale advantages in production, distribution, and marketing are at play for both types of beer, any differential in advertising effectiveness across light and non-light segments would likely be attributed to demand-side factors.

1.3.2 Industry context

Five important, interrelated, developments have shaped the current state of the MPLB industry in the United States.[9] First, the influx of German immigrants in the nineteenth century helped German-style lager beer surpass British-style ale, porter, and stout by a significant margin by 1900. Despite the fact that lager requires longer storage time for fermentation (hence the German meaning of its name: *storehouse*) compared to British-style beer (6–10 days vs. 5–7 days) and that fermentation occurs at lower temperatures, suggesting that more care needs to be taken during production, beer and lager have been used synonymously ever since.

Second, new technologies—pasteurization, refrigeration, and bottling—increased shelf life, while the growing railroad system and, later, motorized delivery improved transportation, allowing beer to be distributed to distant markets. These technological advancements elevated beer from its small-scale, local origins and transformed it into a mass-produced, mass-consumed product.

Third, only a subset of breweries were able to survive Prohibition following the passage of the Volstead Act in 1919. By producing near-beer products and investing in delivery vehicles, these few breweries were able to stay in business until the Twenty-First Amendment repealed Prohibition in 1933. Further, a shift toward product packaging, in particular to canning, increased market concentration, edging out many small-scale regional breweries in the post-Prohibition era.

[9] See Stack (2000, 2003) and Tremblay and Tremblay (2005) for thorough historical reviews of the U.S. beer industry.

Fourth, the cost efficiencies gained from new packaging practices were accompanied by hundreds of mergers and acquisitions, further increasing the consolidation.

Last, and important for this study's empirical design, a new type of beer—light beer—started capturing and expanding the market. It is common in the literature to attribute the successful entry of light beer, followed swiftly by its dominance, to the spread of TV as the new medium to reach consumers and aggressive TV advertising by breweries.

Although the empirical literature has examined advertising effects in alternative beer segments—including ale, porters, stout, imported, and, more recently, craft beer—most industry studies have focused on mass-producers and their major products, where lager is the clear leader. It is common to divide lager into popular-priced, premium, and super-premium segments in order to tease out differential consumer willingness to pay a premium or to signal quality. Another possible categorization is regular, ice, dry, pilsner, light, and low-alcohol beer, depending on the alcohol and calorie content. This study has more in common with the second group, as I combine light and low-alcohol into a "light beer" category and regular, pilsner, and ice beer into a "non-light beer" category. Hence, the main difference between light and non-light lager beer is the alcohol, and to some extent carbohydrate, content.

After water and tea, beer is the most-consumed beverage worldwide (Nelson, 2005). In 2020, beer commands $94 billion in sales revenue ($116 billion in 2019) in the United States and $189 billion around the globe (The Beer Global Market Report, 2021; Brewers Association, 2021). About 88 percent of the U.S. beer volume sales belong to the mass-produced segment or *macrobreweries* where a majority of products are lagers. The five leading brands in the industry are Bud Light (13.24 percent), Coors Light (6.82 percent), Miller Lite (6.02 percent), Budweiser (5.14 percent), and

Michelob Ultra (4.97 percent), all of which are lagers with three being light lagers (USA Today, 2020). The average adult consumer in the United States drinks about one six-pack every week, but there is considerable regional variation in per capita consumption, from 20.2 gallons per person per year in Utah to 45.8 gallons per person per year in North Dakota (Beer Info, 2021).

1.3.3 Data

Whether there is differential effectiveness in traditional advertising by beer type is an empirical question. To assess a differential, the variation in sales volume in each beer segment needs to be captured and linked to the variation in advertising spending in that segment. I draw beer sales data from Nielsen Retail Measurement Systems (RMS) and beer advertising expenditure data from Kantar Media Ad$pender databases. I match the two datasets at brand-DMA-week level, where DMAs (designated market areas) are local advertising markets that are centered around a major city. Both databases provide detailed product-level data that can be aggregated up to the brand level for hundreds of CPG industries over many years at weekly granularity across the United States.[10] While the RMS data include scanner sales information from retail stores, Ad$pender is available at the DMA level. In addition, I use yearly county-level demographics and geography data from the Census Bureau and data on employment, wage, and CPI collected by the Bureau of Labor Statistics.

1.3.3.1 Nielsen RMS

The RMS data provide information on weekly store sales, measured in units and dollars, and detailed product attributes at the UPC (universal product code) level. Compared to brands, UPCs contain extra product information such as package size or product type. For example, both 6-pack and 12-pack Bud Light belong to a single

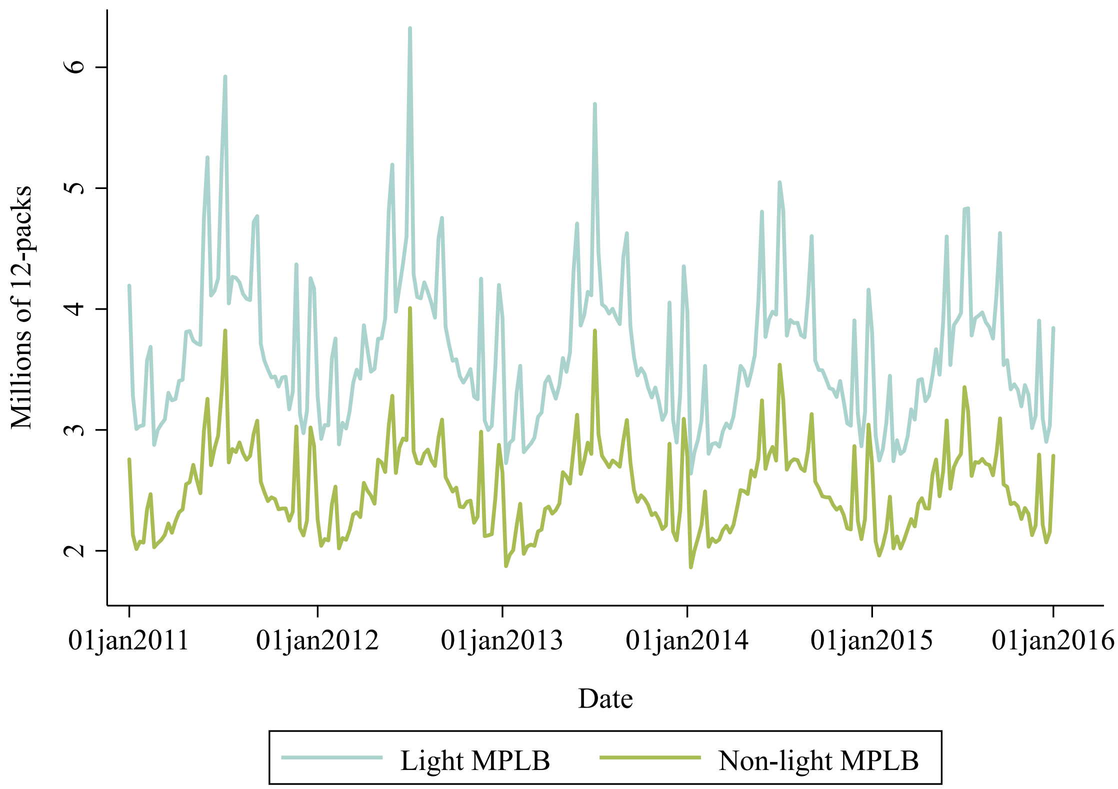
[10]This study uses data from 2011–2015 due to the limited availability of Kantar Media Ad$pender through the University of Massachusetts Amherst library databases.

brand (Bud Light), but they are represented by different UPCs. The 2011–2015 RMS beer sample records sales of 5,445 brands (276 light and 5,169 non-light) and 13,229 unique UPCs (1,991 light and 11,238 non-light). From this sample, I identify 3,142 UPCs (1,111 light and 2,031 non-light) as mass-produced lager beer and aggregate them into 122 brands (30 light and 92 non-light).

Each year, RMS tracks beer sales from more than 20,000 stores across multiple retail channels. In 2011, an estimated 35.1 percent of beer sales by volume were made at convenience stores, 32.3 percent were made at food, drug, and mass merchandise outlets, 5.5 percent were made at liquor stores, and 0.6 percent were made at other locations (2019 Nielsen RMS data manual). A total of 26,136 stores reported beer sales to Nielsen during the 2011–2015 period. Of these, the final sample I describe in Section 1.3.4 includes 7,906 stores.

Figure 1.1 plots the total number of 12-packs in light and non-light MPLB purchased over time for the 7,906 stores that continuously reported MPLB sales each year between 2011 and 2015. Light beer volume sales are persistently higher than non-light beer sales. However, both segments illustrate a fairly consistent and parallel trend throughout the sample period, with a slight decrease in sales during the second half of the period. Further, the plot depicts a clear seasonality pattern with lower sales during winter and higher sales during summer in both beer segments.

Figure 1.1. Aggregate weekly volume sales (millions of 12-packs), 2011–2015: light vs. non-light mass-produced lager beer



Notes. The figure shows the aggregate weekly sales in light and non-light mass-produced lager beer (MPLB) in the United States over 2011–2015. The y-axis denotes the aggregate volume sales in millions of 12-packs (144-ounce equivalent units) across all the MPLB brands available in the balanced all markets (AM) panel each week between 2011 and 2015. The x-axis denotes the date (a total of 261 weeks). The panel contains volume sales for 122 MPLB brands (30 light and 92 non-light) aggregated across 3,142 UPCs (1,111 light and 2,031 non-light) from 7,906 stores in the top 98 designated market areas (DMAs) that continuously report MPLB sales each year between 2011 and 2015. 2011–2015 Nielsen RMS and Kantar Media Ad$pender.

1.3.3.2 Kantar Media Ad$pender

Data on weekly advertising expenditures and occurrences across 17 traditional media types—including TV, newspapers, magazines, radio, and outdoor—come from Kantar Media Ad$pender. Information on ten of these media is available at the national scale, while the remaining seven are available at the local level across the top 101 designated market areas (DMAs). Even though DMAs were originally defined by Nielsen AC to be used by TV advertisers, each DMA has unique local ad content for all the remaining traditional ad channels reported by Kantar Media. Ads for the outdoor channels (e.g., billboards, movie theatres, buses) are originally collected from Standard Metropolitan Statistical Areas (SMSAs), but Ad$pender maps these SMSAs onto DMAs. Table 1.2 reports how I categorized each media type into one of four traditional advertising channels (TV, print, radio, and outdoor). In all subsequent empirical analyses, I either use these traditional channels separately or sum them and use total traditional media to estimate differential advertising effectiveness.

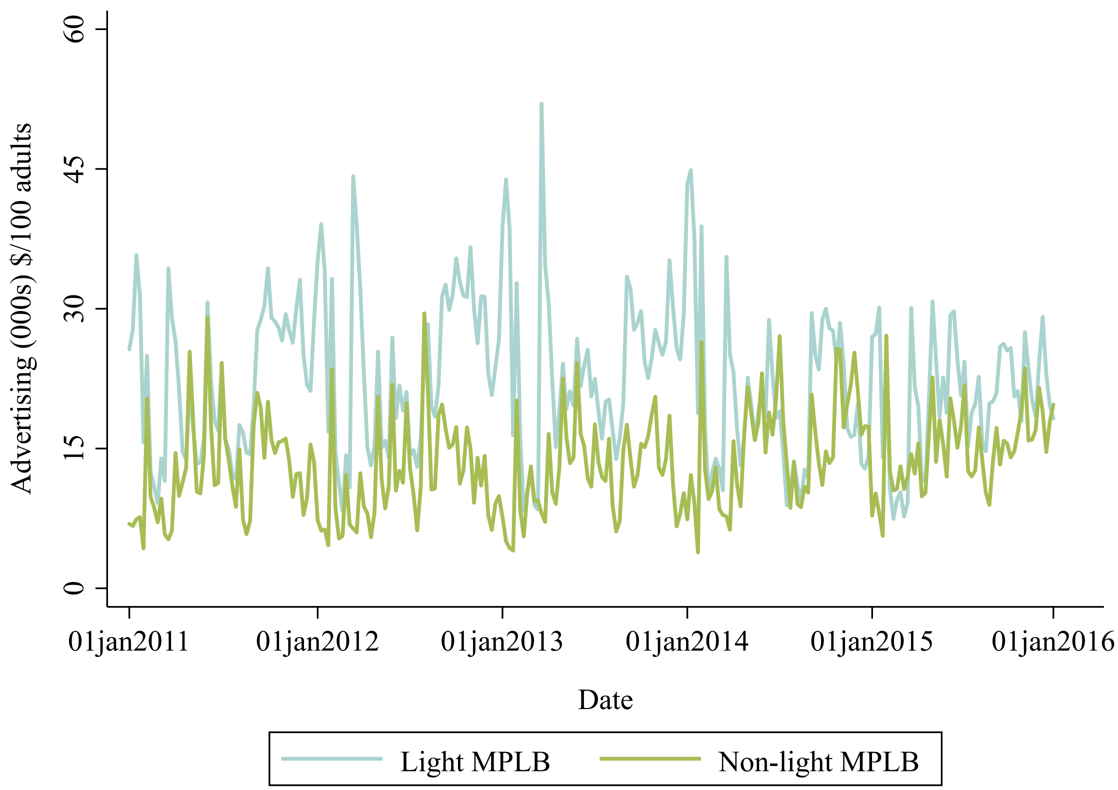
Table 1.2. Traditional advertising channels by media, 2011–2015 Ad$pender

Traditional ad channels	National	Local
TV	Network TV, Spanish Language Network TV, Cable TV, Syndication	Spot TV
Print	Magazines, Sunday Magazines, Hispanic Magazines, B-to-B Magazines, National Newspapers	Local Magazines, Newspapers, Hispanic Newspapers
Radio	Network Radio	National Spot Radio, Local Radio,
Outdoor	N/A	Outdoor

Notes. The table reports the distribution of 17 media (10 national and 7 local) into 4 traditional advertising channels. The raw Ad$pender data contain 19 media (11 national and 8 local) including U.S. Internet and Local History Radio. The internet data are limited and contain only display advertising from banners and some rich media types excluding sponsored social media ads and ads viewed on a mobile device. Local History Radio data are not available after 2013. Outdoor advertising data are available on a monthly basis and contain expenditures from billboards (poster and paint), painted walls, transit/bus shelters, in-store displays, convenience stores, shopping malls, airport, and taxi displays among other out-of-home advertising (Ad$pender Methodology). 2011–2015 Kantar Media Ad$pender.

Advertising data are typically available at the brand level. In some instances, however, an advertising expenditure observation may cover more than one brand. For example, “Amstel Light & Heineken Beer: Combo (F619.0.0)” is an entry that combines two brands. The expenditures in this type of entry are assigned to each of the brands based on the assignment criteria described in Section 1.3.4. Figure 1.2 plots the trend in total advertising over time by beer type.

Figure 1.2. Aggregate weekly traditional advertising spending (thousands of U.S. dollars per 100 adults), 2011–2015: light vs. non-light mass-produced lager beer



Notes. The figure shows the aggregate weekly advertising trend in the light and non-light MPLB in the United States over 2011–2015. The variable on the y-axis denotes the total traditional advertising across all the mass-produced lager beer brands available in each week and Designated Market Area (DMA). The unit of advertising is U.S. dollars per 100 legal drinking-age adults (21+). More specifically, I calculate total traditional advertising by summing up (i) all traditional media from TV, radio, print, and outdoor channels; (ii) all legal drinking-age population normalized national and local advertisements; and (iii) direct and general advertisements. I proportionally allocate the general advertisements among the affiliated brands: I first assign the corresponding total volume sales aggregated from all stores in that DMA and week, and then calculate the volume shares for each brand. The variable on the x-axis denotes the date (a total of 261 weeks). 2011--2015 Nielsen RMS and Kantar Media Ad$pender.

1.3.3.3 Census

Data on county-level demographics—including age, gender, race, and ethnicity—come from the annual estimates of the 2010 U.S. decennial census. I calculate shares of each adult generation (i.e., millennials, Generation X, baby boomers, and the silent generation) between 2011 and 2015 in the drinking-age adult population (i.e., 21 years of age and older). I further use county-level population to calculate DMA population, as DMAs are collections of non-overlapping counties. Urban density information is also available from the U.S. Census Bureau at the county level once every 10 years. Last, I use information on neighboring counties to create bordering markets in the form of either bordering regions (a collection of counties on each side of a border that divides two DMAs) or bordering county-pairs (all county-pairs that share the same DMA border).

1.3.3.4 Bureau of Labor Statistics

I use county-level employment and wage data from the Quarterly Census of Employment and Wages (QCEW) and Consumer Price Index (CPI) data from the U.S. Bureau of Labor Statistics between 2011 and 2015. The QCEW data provide the most accurate total employment and average earnings data each quarter for a large number of industries, coded by North American Industry Classification System (NAICS). I include all industries as well as both private and public sectors. Total employment is calculated as a simple average of monthly employment per each county-quarter pair. Average wage earnings denote total wage income divided by total employment in the area. I use CPI data to calculate real expenditure on sales, advertising, and wage earnings.

1.3.4 Sample construction

I construct three distinct samples: a sample of all markets (AM) and two samples of bordering markets (BM) where the markets are defined either as adjacent border county-pairs (ABCP) as in Dube, Lester and Reich (2010) or as adjacent border regions (BR) as in Shapiro (2018) and Tuchman (2019). In all samples, weekly brand sales and advertising are at the store level and all active stores are represented in a given market. Specifically, the AM sample contains all stores in both the border and interior counties across the DMAs, while the BM samples contain stores from only those counties that straddle a DMA border.

1.3.4.1 All markets (AM)

The raw RMS data contain volume sales and prices for all retail beer at the UPC-store-week level, including 26,136 stores, 206 DMAs, and 13,229 UPCs between 2011 and 2015. Data on sales and price are provided only if a UPC is purchased in a particular store and week. However, treating zero sales as if they were missing data might bias the estimates since, for example, a consumer might see an ad but decide not to make a purchase. To overcome this issue and allow for consumers' no-purchase decisions, I built a balanced sample of UPCs across stores and weeks and imputed prices for zero sales using a simple algorithm that I describe in Section 1.4.3.[11] Overall, 26.03 percent of all UPC-store-week combinations in the MPLB sample have zero sales and imputed prices.

To create the subset of MPLB products, I manipulate both sales and advertising databases (often simultaneously, to minimize manual work) on four main dimensions by preparing the sets of relevant markets, products, time periods, and variables.[12] The matching between the RMS and Ad$pender restricts the sample of markets to 100

[11] UPCs and brands appear to have zero sales for two reasons. First, brands are identified as MPLB in each DMA-category pair, separately. Hence, a MPLB brand in one local market might not rate as a MPLB in another, even if it has some sales in this second market. Second, I code zero sales and subsequently impute prices only for UPCs with at least one recorded purchase in a year, thereby distinguishing between availability and consumers' no-purchase decision.

[12] See Appendix A for a more detailed description of the preparation of each dataset and the matching between the two.

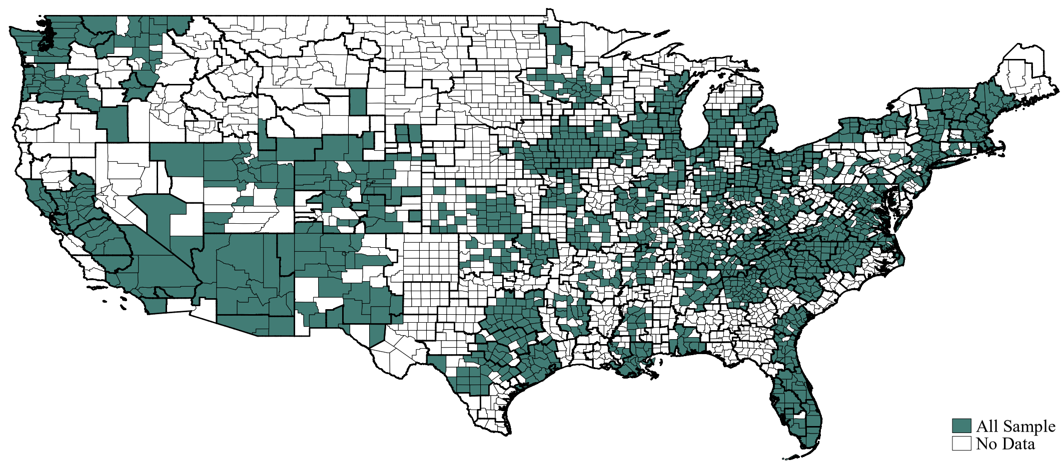
DMAs (Kantar Media monitors the top 101 DMAs, but Nielsen does not report sales for Honolulu, HI).[13] Next, I keep a balanced panel of stores (those that continuously report sales to Nielsen during the sample period) and remove split counties (those that change DMAs over time due to redrawn borders), split stores (those that change counties over time), and smaller stores (those at the bottom 10 percent of aggregate beer revenue) from the sample. The final sample covers 7,906 stores in 963 counties and 98 DMAs.[14] Figure 1.3 shows the map of all counties and their DMAs in the all markets (AM) sample.

---

[13] Ad$pender codes a designated market, All Other, to denote all the remaining DMAs. However it is not feasible to separate out individual markets and match with those in Nielsen RMS.

[14] The two other DMAs, Harrisburg, PA, and Johnstown, PA, are dropped when I remove smaller RMS stores.

Figure 1.3. All markets (AM) sample: 963 counties across the top 98 designated market areas (DMAs) with mass-produced lager beer sales and advertising



Notes. All markets sample comes from the matched sample between Nielsen RMS and Kantar Media Ad$pender. The figure shows the location of all counties from the matched sample on a map of the United States. The sample contains 963 counties across the top 98 designated market areas (DMAs). Thicker lines represent DMA borders, thinner lines represent county borders. 2011–2015 Nielsen RMS and Kantar Media Ad$pender.

Nielsen reports a rich set of product attributes including brand, size, category, container (can, bottle, keg), style (imported, domestic), and type (e.g., lager, ale) for each UPC. To identify whether a beer is a lager, I use descriptions provided by UPC (upc_descr), brand (brand_descr), and type (type_descr). Of 13,229 total beer UPCs, 6,326 (47.8 percent) are identified as lager, 113 (0.9 percent) as not lager, and the remaining 6,790 (51.3 percent) and 3,738 brands are left with unknown lager status, requiring a manual search. However, to make the manual search more manageable, I first keep only the most popular brands in a given DMA and beer category (i.e., those brands that make up the top 95 percent of volume sales in a DMA-category pair). I then search for lager status in this much smaller sample of brands. I then remove brands with a share of less than 1 percent in this smaller pool.[15] Selecting brands using a local market share threshold allows me to include a regional dominance criterion, an acknowledged characteristic of the beer industry due to the high cost of transporting beer, and removing brands with shares of less than 1 percent assures that brands that are not mass-produced are not included. A total of 122 brands, 32 light and 92 non-light, are left in the final MPLB sample.

Kantar Media Ad$pender records advertisements every week beginning on Monday, while Nielsen RMS records sales every week ending on Saturday; this ensures a 6-day overlap between the two databases. Consumers start seeing advertisements on Monday of the week, and they make purchases until Saturday of that week. 2011–2015 Ad$pender reports weekly advertising information on Mondays from December 27, 2010, through December 21, 2015, while Nielsen Retail Scanner covers the period between January 1, 2011, and December 26, 2015 (Saturdays). Therefore, the matched sample spans 261 weeks between 2011 and 2015.

[15] Mass-produced beer companies are either large breweries that operate nationally, with annual sales over 6 million barrels, or exceptionally large regional breweries with annual sales of 4–6 million barrels; microbreweries (craft beer) sell fewer than 15,000 barrels a year. See the market segment statistics in Brewers Association (2018) for further details.

Last, I identify and categorize all potential ads from Ad$pender for the selected 122 MPLB brands in the matched 98 DMAs over 2011–2015. Ad spending is provided at product level, which is a combined description of brands, categories, and type of ads as described in Section 1.3.3. I aggregate product-level ads into brands and categorize each brand-level ad as either direct advertisement (i.e., exclusively pertaining to one focal brand) or general advertisement (i.e., those that can be assigned to multiple brands) per each ad channel (TV, print, radio, and outdoor). To categorize, I clean the data on parent companies, considering several mergers and acquisitions that were approved before and during the sample period.[16]

1.3.4.2 Bordering markets (BM)

I build two distinct samples for bordering markets (BM) depending on the number of stores and counties considered in each local neighboring market: adjacent border county-pairs (ABCP) and border regions (BR).[17] For both samples, I first create a list of all neighboring counties in the AM sample using the 2010 County Adjacency File from the Census Bureau. Next, I keep the neighboring counties on different sides of a DMA border. This restricts the sample to 343 counties and 82 DMAs. For the ABCP sample, I observe 304 distinct county-pairs, or 608 bordering markets, that are continuously active over 261 weeks. In the BR sample, I observe 96 distinct local markets with two neighboring DMAs (or, equivalently, 192 bordering markets) over the same 261 weeks.

[16]Frequent merger activity is a stylized fact in the beer industry, which is why many papers in the industrial organization literature use a case study of beer when estimating merger effects on price and welfare. For example, Ashenfelter, Hosken and Weinberg (2015) and Miller and Weinberg (2017) analyze the price effects of a merger between Molson Coors and SAB Miller, the two largest beer companies, which were ranked second and third by the time the merger was proposed.

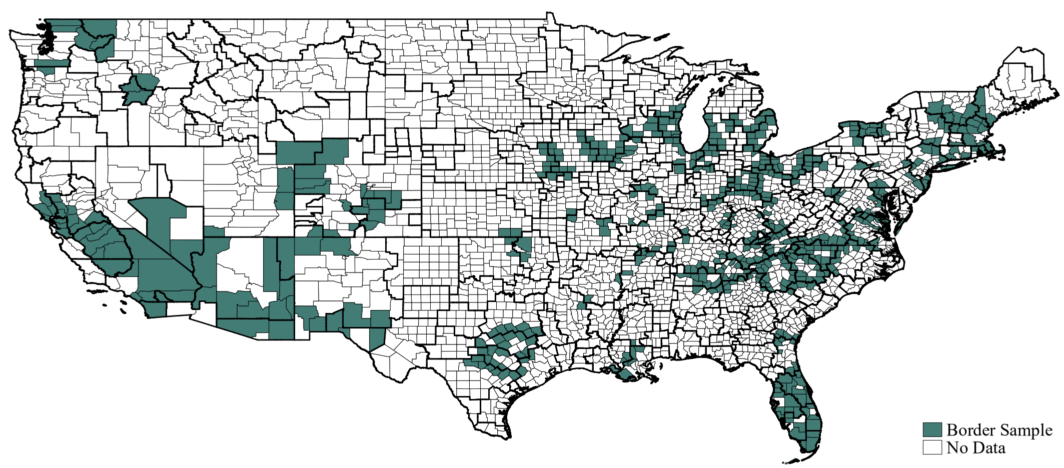
[17]Dube, Lester and Reich (2010) use the ABCP approach to identify employment effects of minimum wage policies using contiguous county-pairs that share a state border, while Shapiro (2018) and Tuchman (2019) use the BR approach and neighboring markets across DMA borders to measure the impact of TV advertising on antidepressants and cigarettes, respectively.

BM samples would include data from stores repetitively if some counties neighbored multiple counties. As noted by Dube, Lester and Reich (2010), these repetitive observations would downsize standard errors for estimates and inflate precision. Following Dube, Lester and Reich (2010), I overcome this bias in standard errors by two-way clustering them on the DMA and border, separately, for each sample.

The level of observation is brand-border-store-week, where border denotes either ABCP or BR depending on the sample. Each border-store-week combination constitutes a local market and will be considered as a local experiment, as I explain in greater detail in Section 1.4. Figure 1.4 shows a map of counties in the BM sample with their corresponding DMA borders.[18]

[18] Even though both samples are at the store level, the exact location of these stores remains anonymous. Therefore, the maps are provided at the county level.

Figure 1.4. Border markets (BM) sample: 343 bordering counties across 82 designated market areas (DMAs) with mass-produced lager beer sales and advertising



Notes. In the adjacent border county-pairs (ABCP) sample there are 304 borders; in the border region (BR) sample there are 96 borders. The border markets sample comes from the matched Nielsen RMS and Kantar Media Ad$pender sample. The figure shows the location of all counties that share a designated market area (DMA) border in the matched sample on a map of the United States. The sample contains 343 counties across 82 DMAs. Thicker lines represent DMA borders, thinner lines represent county borders. 2011–2015 Nielsen RMS and Kantar Media Ad$pender.

1.4 Research design

Firms do not make advertising decisions randomly; rather, they likely spend more on advertising during times and in places where consumption is also likely to be higher. This co-movement poses a challenge for the identification of causal effect. A naïve analysis might attribute all the increases in beer sales to the advertisements, in which case the estimates would be biased upward.[19]

To address this endogeneity issue, I use a credible research design along the lines of Dube, Lester and Reich (2010) and Shapiro (2018) that allows me to identify the causal impact of advertising on sales. This exercise provides credible counterfactuals for the sales as if the advertisements had never taken place. Firms make different advertising decisions for different DMAs, which creates useful variation in advertising between the opposite sides of bordering DMAs. More specifically, I trace out the variation in sales generated by consumers who reside on opposite sides of a DMA border over time and measure how that variation relates to the variation in advertisements to which these people are exposed. Importantly, these consumers are not only similar in terms of observables such as demographics, but they are also exposed to the same business cycles, economic conditions, and unobserved local demand shocks that are correlated with firms' advertising decisions and sales. As a result, my research design leaves only the level of advertising exposure to explain the differences in beer sales between neighboring DMA populations.

1.4.1 Identification using bordering markets

1.4.1.1 Choice of research design for traditional advertising effects

I use two alternative approaches from the empirical economics literature to extract exogenous variation in traditional ad spending: border regions (BR) and adjacent

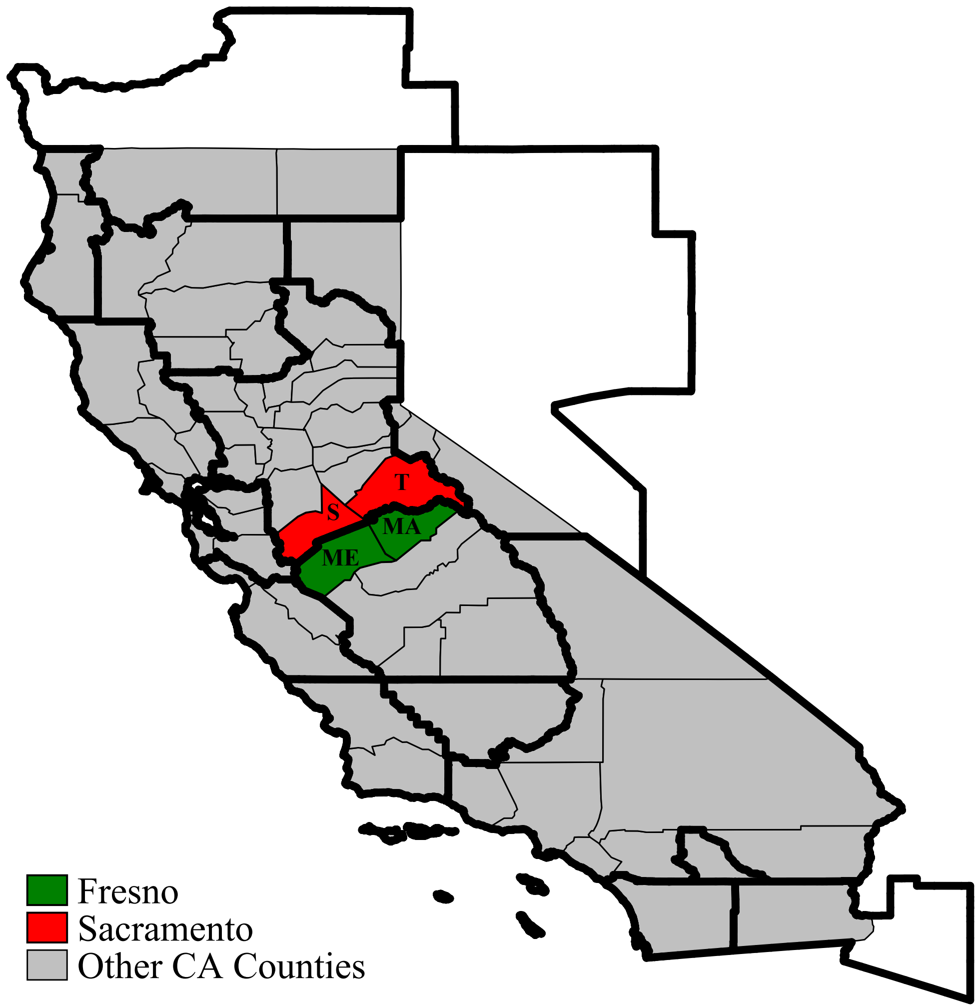
[19] For example, both beer sales and ad expenditures are high during the Super Bowl.

border county-pairs (ABCP).[20] Specifically, the BR approach compares all stores in all counties on one side of a DMA border with stores on the other side of the border without requiring each county on one side of the border to be adjacent to every other county on the other side. The ABCP approach compares stores only if two counties in different DMAs share a border.

To see how each approach works, Figure 1.5 illustrates the border between the Sacramento, CA, and Fresno, CA, DMAs. Each DMA has two counties adjacent to their shared border: Stanislaus (S) and Tuolumne (T) in the Sacramento DMA, and Merced (ME) and Mariposa (MA) in the Fresno DMA. Among these counties, two county-pairs are adjacent to one another across the DMA border: (i) S and ME and (ii) T and MA. The BR approach groups the counties on either side of the DMA border and runs a single experiment with observations from the two bordering regions in each week. More precisely, I group all the stores in S and T and all the stores in ME and MA. I then trace out the differences in sales and advertisements over time using all the stores in the S/T pair and comparing them with all the stores in the ME/MA pair. On the other hand, the ABCP approach matches only the adjacent county-pairs, yielding two county-pairs in the Sacramento-Fresno border: (i) S and ME and (ii) T and MA. In this approach, therefore, each local experiment contains the stores only in adjacent county-pairs.

[20]The identification strategy based on bordering markets has been used in various contexts, including borders that divide electricity markets, school districts, congressional districts, metropolitan statistical areas (MSAs) and states.

Figure 1.5. California and its DMAs



Notes. The figure shows the bordering counties in the Sacramento, CA, (in red) and Fresno, CA, (in green) designated market areas (DMAs). Each DMA has two counties adjacent to the shared border: Stanislaus (S) and Tuolumne (T) in the Sacramento DMA, and Merced (ME) and Mariposa (MA) in the Fresno DMA. Among these counties, two county-pairs are adjacent to one another across the DMA border: (i) S and ME and (ii) T and MA. Thicker lines represent DMA borders, thinner lines represent county borders. 2011–2015 Nielsen RMS and Kantar Media Ad$pender.

Importantly, the BR approach has the advantage of including more observations in each local experiment, leading to a smaller sampling error when compared to the ABCP approach. That is, each BR experiment is less likely to represent a treatment effect on the tails. On the other hand, the ABCP approach has the advantage of providing more localized experiments, which are useful to account for unobserved spatial heterogeneity in the presence of more localized confounders.[21] Furthermore, the number of experiments in the ABCP approach increases threefold compared to the BR approach (304 vs. 96). Even if the ABCP approach is more vulnerable to experiment-specific errors due to smaller sample size in each experiment, these errors on average would cancel each other out since the weight of each experiment would be less in a larger number of experiments.[22]

The research design controls for unobserved spatial heterogeneity using a battery of fixed effects and ensures matching and comparisons between treatment and control groups using only the neighboring border markets. In all model specifications, I use three sets of fixed effects: brand-border-store, brand-border-month, and store-border-month. Brand-border-store fixed effects sweep out all the persistent regional differences in demand for a specific brand in a given store. For instance, historically higher demand for Anheuser-Busch brands (Bud Light, Budweiser) around St. Louis, MO, where the company headquarters are located, will be accounted for with these fixed effects. Brand-border-month fixed effects allow me to partial out all of the time-specific variation between the two sides of the border each month. These fixed effects will capture whether, for example, a brand suddenly loses its popularity across a region of the United States for a reason unrelated to own or rival advertising

[21]Unobserved demand shocks are spatial in nature; hence, they do not jump up or down discontinuously when crossing the border.

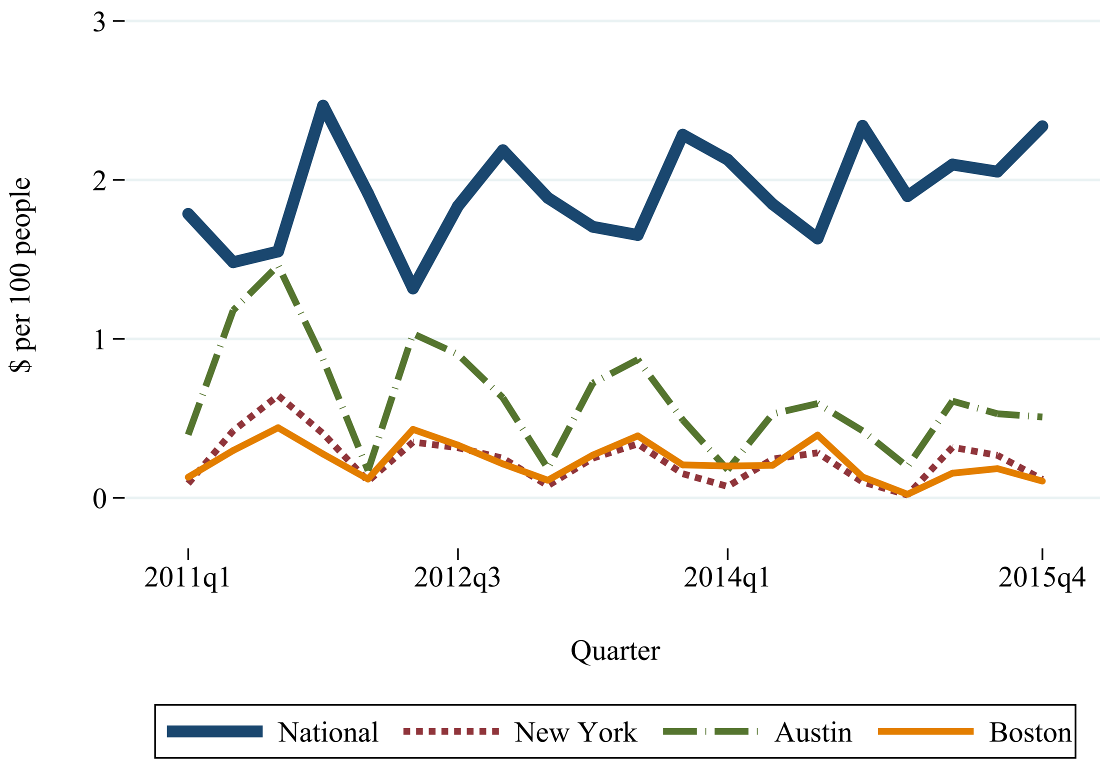
[22]This advantage will be smaller if advertising effects are compromised due to regional spillovers that might create attenuation bias, which would occur, for example, when adults have seen an ad on one side of the border but made the purchase on the other side. I discuss this issue in more detail in Section 1.4.2, and show that the main results are not affected by this attenuation bias.

expenditures. Store-border-month fixed effects control for a store-specific time trend of any functional form on either side of the border. They purge any store-specific overall beer demand changes that are not related to advertising.

1.4.1.2 Regional time-varying variation in beer advertisements and sales

The identification depends on whether there is useful variation both across DMAs and over time in local advertising and sales. Otherwise, the regressions fail to produce results with sufficient statistical power. As an example, Figure 1.6 plots the quarterly local and national traditional advertising ($ per 100 adults) in three cities (New York, NY; Austin, TX; and Boston, MA) during 2011–2015. The figure confirms that there is significant variation both across bordering markets and over time, even though a majority of advertising spending is carried out at a national scale.

Figure 1.6. Regional variation across three markets in advertising, 2011–2015: light vs. non-light mass-produced lager beer

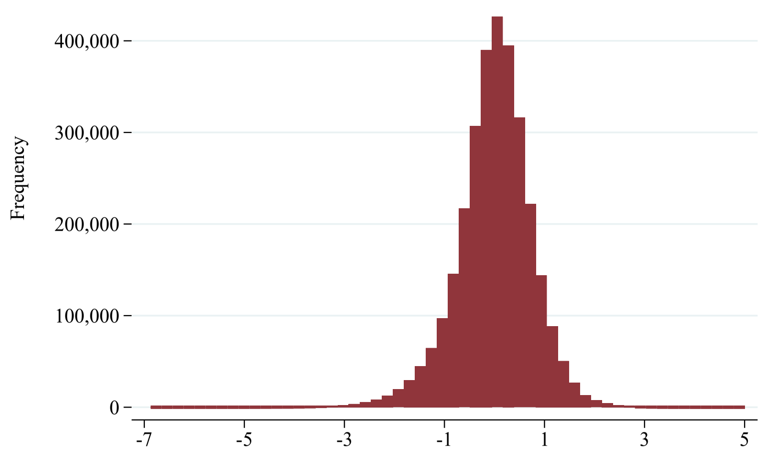


Notes. The figure shows the time-varying regional variation in local traditional advertising spending across three local designated market areas (DMAs) as well as the time-varying variation in national advertising spending. The y-axis denotes mean advertising across all the mass-produced lager beer (MPLB) brands available in each quarter and DMA. The figure also reports mean advertising ($/100 adults) for the national media. More specifically, national advertising refers to brand advertising aired or distributed throughout the United States, while local advertising refers to DMA-exclusive advertising. For each brand, I calculate the national advertising in a DMA by dividing national advertising expenditures by the U.S. population of legal drinking age (21+). For the local advertising, I divide the DMA-specific advertising dollars by the population of legal drinking age in that DMA, where I calculate the DMA population by summing the population from all counties in that DMA. 2011–2015 Nielsen RMS and Kantar Media Ad$pender.

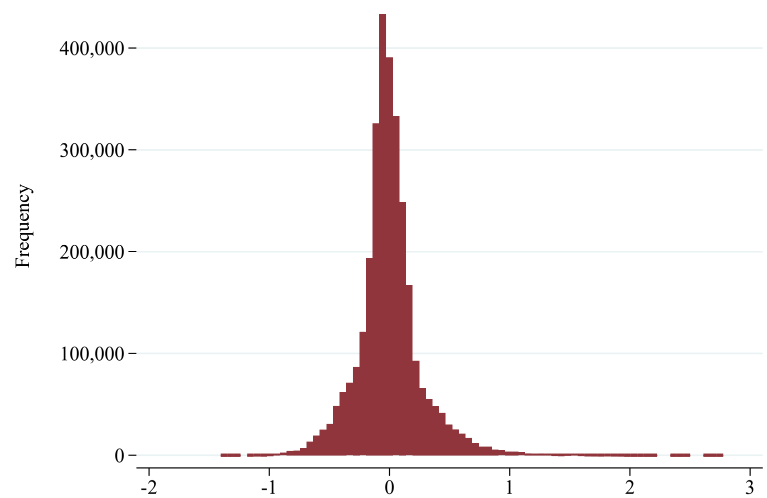
Figure 1.7 shows the distribution of residuals net of fixed effects and controls, providing further support that both sales and advertising exhibit important variation with a standard deviations of 0.27 for advertising (Panel A) and 0.75 for sales (Panel B).

Figure 1.7. Variation in sales and advertising net of controls: border region sample

Panel A. Volume sales



Panel B. Advertising



Notes. Panel A shows the distribution of residual volume sales in baseline specification (Equation 1.1) net of fixed effects and controls (own advertising, rival advertising, own price, and rival price) with a standard deviation of 0.75. Panel B shows the distribution of residual advertising net of fixed effects and controls (own price and rival price) with a standard deviation of 0.27. 2011–2015 Nielsen RMS and Kantar Media Ad$pender.

1.4.2 Potential threats to the research design

1.4.2.1 Remaining confounders on one side of a border

The identification would fail if there were remaining unobserved confounders that are correlated with sales and advertising on one side of a DMA border and not the other. The neighboring populations would no longer be similar and would therefore be unsuitable controls for each other. More broadly, the identifying assumption would be violated if a DMA border coincides with, for example, a state border; in this case, a state-level policy on one side of the DMA border would appear to correlate with advertisement spending (and therefore demand) while, in reality, they are independent events. The existing literature deals with this issue by removing the borders that overlap with other borders.[23] However, in my empirical setting, I additionally control for store-border-week fixed effects, which would deal with the concern of remaining confounders as long as the unobserved policy on one side of the border does not affect a single brand or a subset of brands. Therefore, the distribution laws noted in Ashenfelter, Hosken and Weinberg (2015) and other alcohol-related policies, such as the local excise taxes discussed in Tuchman (2019), should not threaten my preferred research design.

1.4.2.2 Dissimilar adjacent DMAs

One of the main assumptions of my research design is that the bordering DMAs are quite similar to each other and that, therefore, discrepancies in beer sales are, on average, largely driven by firms' advertisement decisions. If this is not the case and the traditional advertising effectiveness has a differential impact on certain demographics, then my findings could be biased.[24] To assess the validity of this

[23]For example, both Shapiro (2018) and Tuchman (2019) drop DMA borders that overlap with state borders.

[24]In Section 1.5.1, I provide results for the differential effectiveness of ads between the light and non-light beer by demographic group; in Section 1.5.3, I discuss their implications.

assumption, I compare demographic data across the two sides of every border included in the estimation: 96 borders in the BR approach and 304 county-pairs in the ABCP approach.

For ease of interpretation, I calculate a standardized absolute distance metric that measures the difference in standard deviations between the two sides of a bordering market for each demographic characteristic, as in Tuchman (2019). Specifically, for each demographic characteristic d and bordering markets b_1 and b_2, the standardized absolute distance is calculated as $\left|\frac{d_{b_1} - d_{b2}}{\sigma_d}\right|$, where σ_d denotes the standard deviation of demographic d across all neighboring border markets (192 in the BR approach and 608 in the ABCP approach). Table 1.3 shows the results using the BR approach. I find that the bordering markets are quite similar to each other. The median bordering market pair is within less than half of a standard deviation of each other for 8 out of 11 demographic characteristics.

Table 1.3. Descriptive statistics of the standardized absolute distance in demographics across neighboring border markets using BR approach

Demographic	median	mean	N
Percent Millennial	0.141	0.187	96
Percent Generation X	0.449	0.756	96
Percent Baby Boomers	0.307	0.474	96
Percent Silent	0.615	0.814	96
Percent Women	0.398	0.425	96
Percent White	0.474	0.555	96
Percent African American	0.793	0.896	96
Percent Asian	0.304	0.540	96
Percent Other	0.621	0.783	96
Percent Hispanic	0.247	0.343	96
Percent Urban	0.169	0.307	96

Notes. BR is border region. BR approach contains 96 border regions and thus 192 neighboring border markets. Percentage urban is defined at the county level only and available every 10 years, using data from the 2010 decennial census. All other demographics are available at the county-year level. Descriptive results based on the ABCP (adjacent border county-pairs) approach are similar and available upon request.

1.4.2.3 Attenuation bias

There could be a scenario in which a consumer is exposed to an advertisement in a bordering county but purchases the product on the other side of the border in a different DMA (e.g., when returning from work). Another scenario could be when a consumer sees an advertisement in a bordering county but commutes to an interior county that is not included in the estimation to make the purchase. One final—but less likely—scenario would be if a consumer is exposed to an ad, either TV or radio, from a neighboring DMA because of an overlapping signal (which can occur if the consumer is located very close to the border).[25]

To explore whether attenuation bias may exist in my application, I use Nielsen Homescan data and assess how often households located near a border make purchases at stores located in the neighboring DMA. Table 1.4 reports the results. I find that, only 1 percent (2 percent) are made in a neighboring DMA when households purchased a light (non-light) beer. In addition to cross-border shopping behavior, I repeated the same analysis to measure the prevalence of shopping trips to the interiors of the same DMAs and found that only 2 percent (3 percent) are made in the interior of DMAs when households purchased a light (non-light) beer. These findings suggest that attenuation bias would most likely be negligible.

Table 1.4. Household transactions (shopping trips), 2011–2015 light and non-light mass-produced lager beer

	Light	Non-light
Across border	0.0089	0.0016
To interior	0.0018	0.0028

Notes. I focus on the set of Nielsen panelists who ever buy light or non-light mass-produced beer in the household scanner database (Homescan) and who reside in DMA-bordering counties. First row in the table reports the fractions of light or non-light mass-produced beer purchases that were made by these households in a DMA other than their DMA of residence. Of the 96,129 (92,563) light (non-light) beer transactions made by these panelists, only 1 percent (2 percent) of transactions were made outside of the household's DMA of residence. The second row reports the fractions of beer purchases by these households in the interiors of the DMAs they reside in. 2011–2015 Nielsen Homescan.

[25] To reduce the likelihood of this possibility, I drop stores in markets where DMA borders have been redrawn due to competing signals over the air.

While attenuation bias is not of substantial concern in my application, to the extent that it remains in the data it would render my estimates as a lower bound for the causal advertising effects.

1.4.2.4 Nickell bias

The specifications that include the lagged dependent variable to capture the dynamic effects of advertising could suffer from Nickell (1981) bias. The differencing causes a mechanical correlation between the lagged dependent variable and the error term. This can be problematic for small T, the number of time periods in the sample, but it diminishes to zero as T goes to infinity. I apply the analytical formula provided by Nickell (1981) to solve for the bias as a function of T.[26] In the current empirical analysis, T is *a reasonably large number* (261). As a result, the carryover coefficient estimates register a near-negligible bias for both light (−0.004) and non-light (−0.005) segments.[27]

1.4.2.5 External validity

When estimation is performed based on a subsample, a valid concern arises as to whether the results are generalizable to a larger population. Because firms and policy makers make decisions considering entire populations, the current analysis would be limited if the causal impact of advertising were different in interior and bordering markets. My results can be cautiously generalized to the United States because the number of experiments in my analysis is relatively large (96 in the BR and 304 in the ABCP approach).

[26]The analytical formula for (Nickell, 1981) is $plim_{N\to\infty}(\hat{\lambda} - \lambda) = \left\{\frac{2\lambda}{1-\lambda^2} - \left[\frac{1+\lambda}{T-1}\left(1 - \frac{1}{T}\frac{1-\lambda^T}{1-\lambda}\right)\right]^{-1}\right\}^{-1} \approx \frac{-(1+\lambda)}{T-1}$, where $\hat{\lambda}$ denotes the coefficient estimate of the lagged dependent variable.

[27]See Section 1.5.1 for the reported estimates of the carryover coefficient.

1.4.3 Estimation

1.4.3.1 Main specification

I estimate the impact of advertising on brand sales of light and non-light MPLB. I specify an enriched triple difference (TD) model that matches contiguous DMAs. The identifying variation comes from the three dimensions in my data: store, brand, and time. The baseline regression equation is as follows:

$$Q_{jbsw} = \boldsymbol{\gamma} \boldsymbol{A}_{jbd(s)w} - \boldsymbol{\beta} \boldsymbol{p}_{jbsw} + \boldsymbol{\gamma}_l l_j \boldsymbol{A}_{jbd(s)w} - \boldsymbol{\beta}_l l_j \boldsymbol{p}_{jbsw} + \theta_{jbs} + \theta_{jbm} + \theta_{sbm} + \varepsilon_{jbsw}, \quad (1.1)$$

where Q_{jbsw} is the number of 12-packs (144-ounce equivalent units) of brand j sold in border region b, at store s, in week w; $\boldsymbol{A}_{jbd(s)w}$ is the vector of own and rival advertising stocks in U.S. dollars per 100 adults (21+) for brand j in border region b, DMA $d(s)$, and week w. The ad stock measure is the sum of total direct and proportional general ad stock based on the volume sales shares of the affected brands in a given market and time. $\boldsymbol{p}_{jbsw}$ is the vector of own and rival prices in U.S. dollars per 12-pack. The model controls for unobserved spatial heterogeneity through the use of three sets of fixed effects: (i) θ_{jbs} represents the brand-border-store fixed effects that control for time-invariant regional differences across markets; (ii) θ_{jbm} represents the brand-border-month fixed effects that absorb common trends on opposite sides of a DMA border; and (iii) θ_{sbm} represents the store-border-month fixed effects that attribute a common trend to all beer at the either side of a DMA border.[28] The equation is an interacted model, where l_j indicates whether j is a light beer product, and the parameters $\boldsymbol{\gamma}_l$ and $\boldsymbol{\beta}_l$ refer to the differential advertising and price effects of light beer brands compared to non-light products. In this and all subsequent

[28] Despite the weekly granularity of the data, the triple difference regressions use monthly fixed effects. Estimating the model with weekly fixed effects reduces statistical power.

specifications, I use this interacted model design by interacting the covariates with a light indicator to clearly lay out the sign, magnitude, and statistical significance of the differential effects between light and non-light MPLB.

I specify the advertising stock as $\boldsymbol{A}_{jbd(s)w}$, as in Shapiro, Hitsch and Tuchman (2021):

$$\boldsymbol{A}_{jbd(s)w} = \sum_{W=w-L}^{w} \delta^{w-W} \boldsymbol{a}_{jbd(s)W}, \tag{1.2}$$

where $\boldsymbol{a}_{jbd(s)w}$ is a vector of contemporaneous own and rival advertising spending in U.S. dollars per 100 adults for brand j in border region b, DMA $d(s)$, and week w. δ is the depreciation parameter that discounts the effect of past advertising in a cumulative lagged structure over time. In subsequent analyses, δ is equal to 0.3. As detailed in Section 1.5.2, I estimate δ using a grid search procedure and show that 0.3 is the optimum value in my application. L is the number of lags, and I set it equal to 52. As noted in the literature, the long-run effects of advertising are often no longer than a year (Bagwell, 2007). In fact, it takes about 5–6 weeks for the advertising effect to fully dissipate when δ is equal to 0.3.[29] Intuitively, the advertising stock accounts for the effect of lagged advertising effects on contemporaneous sales. In addition to this simple carryover effect mechanism, advertising stock also captures the state dependence in which past sales (driven by past advertising) are positively related to contemporaneous sales.

For specification 2, I modify the baseline specification (Equation 1.1) by separately including direct and general advertisements in the regression equation as follows:

[29] $0.3^5 = 0.002$ and $0.3^6 = 0.001$.

$$Q_{jbsw} = \boldsymbol{\gamma}^D \boldsymbol{A}^D_{jbd(s)w} + \boldsymbol{\gamma}^G \boldsymbol{A}^G_{jbd(s)w} - \boldsymbol{\beta} \boldsymbol{p}_{jbsw} + \boldsymbol{\gamma}^D_l l_j \boldsymbol{A}^D_{jbd(s)w} + \boldsymbol{\gamma}^G_l l_j \boldsymbol{A}^G_{jbd(s)w} - \boldsymbol{\beta}_l l_j \boldsymbol{p}_{jbsw}$$
$$+ \theta_{jbs} + \theta_{jbm} + \theta_{sbm} + \varepsilon_{jbsw}, \tag{1.3}$$

where the superscript D denotes direct advertisements and G denotes general advertisements. Similar to the baseline specification, Q_{jbsw} is the number of 12-packs sold, while $\boldsymbol{A}^D_{jbd(s)w}$, $\boldsymbol{A}^G_{jbd(s)w}$, and $\boldsymbol{p}_{jbsw}$ are vectors of own and rival direct advertising stocks, own and rival general advertising stocks, and own and rival prices, respectively. Economic theory suggests a nonnegative impact of direct own brand advertisements on own brand sales ($\gamma^D + \gamma^D_l$ for light and γ^D for non-light MPLB). However, the impact of the general advertisements on own brand sales ($\gamma^G + \gamma^G_l$ for light and γ^G for non-light MPLB) might be positive or negative depending on how a general ad impacts the different affected brands. For example, a general Anheuser-Busch advertisement might increase Bud Light sales at the expense of Budweiser sales, suggesting a business-stealing channel. Therefore, in Equation 1.3, I let the data directly test the ambiguous effect of general advertisements on the focal brand. The baseline specification in 1.1, however, lumps direct and general advertising; in this specification general ads are distributed across brands using a proportional allocation rule (see Appendix A).

In specification 3, I modify the baseline specification by including contemporaneous and long-run effects of advertising separately, as follows:

$$Q_{jbsw} = \lambda Q_{jbs,w-1} + \boldsymbol{\gamma} \boldsymbol{a}_{jbd(s)w} - \boldsymbol{\beta} \boldsymbol{p}_{jbsw} + \lambda_l l_j Q_{jbs,w-1} + \boldsymbol{\gamma}_l l_j \boldsymbol{a}_{jbd(s)w} - \boldsymbol{\beta}_l l_j \boldsymbol{p}_{jbsw}$$
$$+ \theta_{jbs} + \theta_{jbm} + \theta_{sbm} + \varepsilon_{jbsw}. \tag{1.4}$$

The dynamics enter the model through ($Q_{jbs,w-1}$), the 1-week-lagged number of 12-packs of brand j sold in border b in store s. $\boldsymbol{a}_{jbd(s)w}$ is a vector of contemporaneous own and rival advertising for brand j in border region b, DMA $d(s)$, and week w. The only difference between this specification and the baseline specification (Equation 1.1) is the way in which dynamics enter the model.[30] The carryover coefficient λ in Equation 1.4 has a similar interpretation as the depreciation parameter, δ, in the advertising stock formulation; both terms capture the amount of it takes advertising effects to dissipate. The carryover coefficient also captures the state dependence on sales (as predicted in models of addiction, for example), since consumers are more likely to make a brand purchase today if they bought it in the past. This could be due to brand loyalty (see Allcott, Gentzkow and Song, 2021; Tuchman, 2019) or habit formation (see DeCicca et al., 2021; Tremblay and Tremblay, 2005).

Specification 4 modifies specification 3 by identifying the direct and general advertisements separately:

$$
\begin{aligned}
Q_{jbsw} = {} & \lambda Q_{jbs,w-1} + \boldsymbol{\gamma}^D \boldsymbol{a}^D_{jbd(s)w} + \boldsymbol{\gamma}^G \boldsymbol{a}^G_{jbd(s)w} - \boldsymbol{\beta} \boldsymbol{p}_{jbsw} \\
& + \lambda_l l_j Q_{jbs,w-1} + \boldsymbol{\gamma}^D_l l_j \boldsymbol{a}^D_{jbd(s)w} + \boldsymbol{\gamma}^G_l l_j \boldsymbol{a}^G_{jbd(s)w} - \boldsymbol{\beta}_l l_j \boldsymbol{p}_{jbsw} \qquad (1.5) \\
& + \theta_{jbs} + \theta_{jbm} + \theta_{sbm} + \varepsilon_{jbsw}.
\end{aligned}
$$

As in specification 3, the long-run effects of advertising are captured through the lagged volume sales. The contemporaneous advertising effects are incorporated by vectors of own and rival direct advertisements, $\boldsymbol{a}^D_{jbd(s)w}$, and general advertisements, $\boldsymbol{a}^G_{jbd(s)w}$. Specifications 1–4 use the BR approach. I estimate these four specifications separately for the BR approach and the ABCP approach (designated as specifications 5–8).

[30] Another method to capture the dynamic advertising effects is to include contemporaneous and lagged advertising separately in the model. However, the data do not allow these effects to be identified separately because the contemporaneous and 1-week-lagged advertising are highly correlated, with the correlation coefficient $\rho = 0.64$.

1.4.3.2 Decomposition by traditional advertising channels

These specifications consider the totality of traditional advertising channels. However, advertising effectiveness may vary across individual advertising channels. The advertising literature typically studies the effects of TV advertising in CPG industries. While about 80 percent of all media spending is concentrated in TV advertising in the average market in the MPLB sector, this might not be representative of the whole picture. To investigate this issue, I estimate the effect of advertising for each channel separately. This decomposition is carried out by modifying the baseline model (Equation 1.1) to

$$
\begin{aligned}
Q_{jbsw} = {} & \boldsymbol{\gamma}^{TV} \boldsymbol{A}^{TV}_{jbd(s)w} + \boldsymbol{\gamma}^{outdoor} \boldsymbol{A}^{outdoor}_{jbd(s)w} + \boldsymbol{\gamma}^{print} \boldsymbol{A}^{print}_{jbd(s)w} + \boldsymbol{\gamma}^{radio} \boldsymbol{A}^{radio}_{jbd(s)w} - \boldsymbol{\beta} \boldsymbol{p}_{jbsw} \\
& + l_j(\boldsymbol{\gamma}^{TV}_l \boldsymbol{A}^{TV}_{jbd(s)w} + \boldsymbol{\gamma}^{outdoor}_l \boldsymbol{A}^{outdoor}_{jbd(s)w} + \boldsymbol{\gamma}^{print}_l \boldsymbol{A}^{print}_{jbd(s)w} + \boldsymbol{\gamma}^{radio}_l \boldsymbol{A}^{radio}_{jbd(s)w} - \boldsymbol{\beta}_l \boldsymbol{p}_{jbsw}) \\
& + \theta_{jbs} + \theta_{jbm} + \theta_{sbm} + \varepsilon_{jbsw}, \qquad (1.6)
\end{aligned}
$$

where $\boldsymbol{A}^{TV}_{jbd(s)w}$, $\boldsymbol{A}^{outdoor}_{jbd(s)w}$, $\boldsymbol{A}^{print}_{jbd(s)w}$, and $\boldsymbol{A}^{radio}_{jbd(s)w}$ are, respectively, the vectors of own and rival TV, outdoor, print, and radio advertising stocks of brand j in the DMA d side of the border region b in week w. As in the baseline specification, the model controls for own and rival prices, $\boldsymbol{p}_{jbsw}$, and for time-invariant, θ_{jbs}, and time-varying confounders, θ_{jbm}, θ_{sbm}. The differential effects of the light beer brands are captured via light interactions, l_j.

1.4.3.3 Heterogeneity by demographic group

The advertising estimates from the previous specifications provide the average advertising effects across U.S. adults. To explore whether traditional advertising works better for one type of beer than for the other among certain demographic groups, I specify and estimate the following TD model:

$$
\begin{aligned}
Q_{jbsw} = &\boldsymbol{\gamma} \boldsymbol{A}_{jbd(s)w} + \sum_d \gamma^d A^{own}_{jbd(s)w} \times d_{c(s)y(w)} \\
&+ l_j \left[\boldsymbol{\gamma}_l \boldsymbol{A}_{jbd(s)w} + \left(\sum_d \gamma^d_l A^{own}_{jbd(s)w} \times d_{c(s)y(w)} \right) - \boldsymbol{\beta}_l \boldsymbol{p}_{jbsw} \right] \\
&+ \theta_{jbs} + \theta_{jbm} + \theta_{sbm} + \varepsilon_{jbsw}.
\end{aligned} \tag{1.7}
$$

Equation 1.7 contains interactions between a vector of different demographics d in county $c(s)$ and year $y(w)$ and the own traditional advertising $Adv_{jbd(s)w}$ for brand j in the DMA d side of the border region b in week w. More specifically, I interact own traditional advertising with the shares of millennials, Generation X, women, African Americans, Asians, Hispanics, and urban populations as well as with average wage earnings and total employment. The reference group is the combined share of baby boomers and the silent generation for age, the share of white for race, and the share of non-Hispanic for ethnicity. Descriptive statistics for the list of demographics from the 343 counties in the BR sample are available in Table 1.5. All demographics are standardized to mean 0 and standard deviation 1 to make the interpretation more intuitive.[31] As in the baseline specification, the model further includes an interaction with a light-brand indicator, l_j, and controls for rival advertising, own price, rival price, and the usual set of fixed effects (θ_{jbs}, θ_{jbm}, θ_{sbm}).

[31] The standardization allows me to evaluate advertising effectiveness with respect to a 1-standard-deviation increase in the demographic variable.

Table 1.5. Demographic descriptive statistics in 2011–2015

Demographic	5 percent	25percent	75percent	95 percent	mean	s.d.	N
Percent Millennial	0.172	0.211	0.262	0.333	0.241	0.05	1,715
Percent Generation X	0.119	0.148	0.173	0.191	0.159	0.022	1,715
Percent Baby Boomers	0.382	0.43	0.466	0.496	0.445	0.034	1,715
Percent Silent	0.103	0.13	0.17	0.23	0.155	0.041	1,715
Percent Women	0.472	0.505	0.52	0.529	0.509	0.019	1,715
Percent White	0.635	0.818	0.958	0.981	0.87	0.12	1,715
Percent African American	0.005	0.015	0.124	0.341	0.091	0.114	1,715
Percent Asian	0.003	0.005	0.017	0.06	0.019	0.034	1,715
Percent Other	0.008	0.012	0.02	0.058	0.021	0.021	1,715
Percent Hispanic	0.009	0.019	0.094	0.324	0.085	0.118	1,715
Percent Urban	0	0.32	0.761	0.97	0.53	0.23	343
Wages (000s)	7.303	8.262	10.345	13.595	9.632	2.19	6,860
Employment (000s)	3.557	9.74	56.542	380.413	84.807	196.517	6,860

Notes. Demographic information is available for county-level wage and employment on a quarterly basis. Percentage urban is defined at the county level only and available every 10 years, using data from the 2010 decennial census. All other demographics are available at the county-year level. The balanced border market sample includes 343 counties.

1.5 Empirical findings

1.5.1 Main results

1.5.1.1 Within-industry heterogeneity in traditional advertising

Table 1.6 reports the differential own advertising effects in the light and non-light segments of the mass-produced lager beer industry for all eight specifications. As detailed in Section 1.4.3, unobserved spatial confounders are accounted for by using brand-border-store, brand-border-month, and store-border-month fixed effects in all regressions. Following Dube, Lester and Reich (2010), I compute standard errors using a two-way (DMA and border) clustering procedure. This procedure controls for mechanical correlation across the bordering pairs from the presence of a single store in multiple pairs (see Appendix A for details).

To facilitate the interpretation of advertising effects, Panel A in Table 1.7 and Figure 1.8 show long-run elasticities of demand (including standard errors) with respect to own traditional advertising spending across the eight specifications. I calculate the elasticities by multiplying the advertising coefficients from the TD regressions reported in Table 1.6 with the ratio of the mean traditional advertising spending (Panel C in Table 1.7) to the mean number of 12-packs (Panel B in Table 1.7). This procedure is applied to both long-run estimates from the specifications with advertising stocks (specifications 1, 2, 5, and 6) and the short-run elasticity estimates from the specifications with contemporaneous advertising and lagged sales (specifications 3, 4, 7, and 8). Following Clarke (1976), I calculate the long-run elasticities by dividing the short-run estimates by $1 - \lambda$, where λ is the estimated carryover coefficient for lagged volume sales.

Table 1.6. Own traditional advertising effects across eight specifications: light vs. non-light mass-produced lager beer

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Direct + general advertising	0.978***		0.953***		0.990***		0.967***	
	(0.141)		(0.155)		(0.125)		(0.137)	
Direct + general advertising × light	0.753***		0.925***		0.721***		0.880***	
	(0.256)		(0.284)		(0.223)		(0.249)	
Direct advertising		0.995***		0.994***		0.997***		0.999***
		(0.147)		(0.167)		(0.132)		(0.147)
Direct advertising × light		0.530**		0.689**		0.532**		0.677**
		(0.253)		(0.292)		(0.225)		(0.257)
General advertising		0.258***		0.179***		0.265***		0.183***
		(0.065)		(0.054)		(0.056)		(0.045)
General advertising × light		-0.347***		-0.356***		-0.304***		-0.311***
		(0.110)		(0.097)		(0.095)		(0.093)
Lagged volume sales			0.326***	0.326***			0.314***	0.314***
			(0.027)	(0.027)			(0.0236)	(0.0236)
Lagged volume sales × light			-0.195**	-0.195**			-0.181**	-0.181**
			(0.0831)	(0.0832)			(0.0752)	(0.0752)
Observations	21,964,423	21,964,423	21,878,427	21,878,427	31,620,468	31,620,468	31,496,824	31,496,824
Brand-border-store FE	Y	Y	Y	Y	Y	Y	Y	Y
Brand-border-month FE	Y	Y	Y	Y	Y	Y	Y	Y
Store-border-month FE	Y	Y	Y	Y	Y	Y	Y	Y
Rival advertising	Y	Y	Y	Y	Y	Y	Y	Y
Own price	Y	Y	Y	Y	Y	Y	Y	Y
Rival price	Y	Y	Y	Y	Y	Y	Y	Y

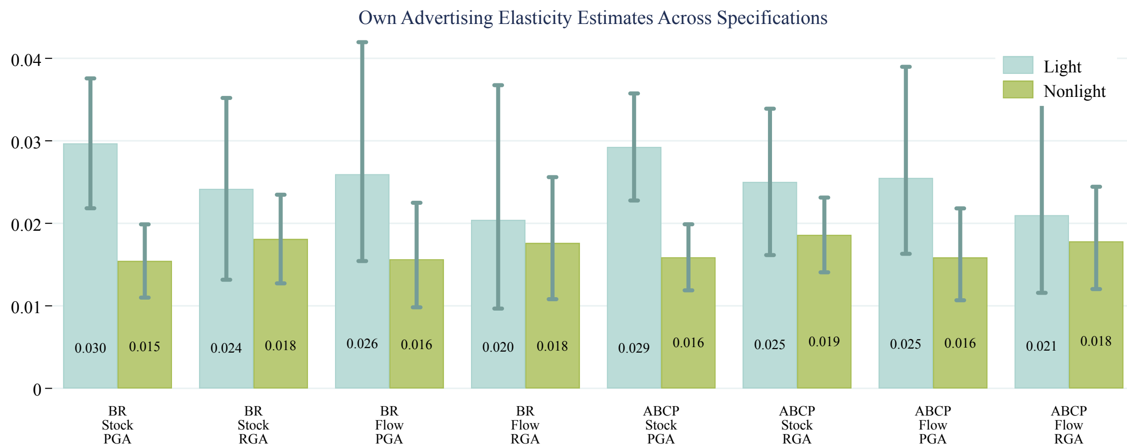
Notes. Robust standard errors in parentheses are two-way clustered at the DMA and border segment level for all specifications. Significance levels: *10 percent, **5 percent, ***1 percent. 2011–2015 Nielsen RMS and Kantar Media Ad$pender. See the text for further details.

Table 1.7. Long-run own traditional advertising elasticities and sample means across eight specifications: light vs. non-light mass-produced lager beer

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Long-run own advertising elasticity								
Light	0.030***	0.024***	0.026***	0.020***	0.029***	0.025***	0.025***	0.021***
	(0.004)	(0.0055)	(0.0035)	(0.0042)	(0.0033)	(0.0045)	(0.0029)	(0.0034)
Non-light	0.015***	0.018***	0.016***	0.018***	0.016***	0.019***	0.016***	0.018***
	(0.0022)	(0.0027)	(0.0018)	(0.0021)	(0.002)	(0.0023)	(0.0016)	(0.0017)
Panel B: Mean own advertising ($/100 adults)								
Light	0.85	0.83	0.59	0.58	0.83	0.82	0.58	0.57
Non-light	0.31	0.28	0.22	0.20	0.30	0.28	0.21	0.19
Observations	21,964,423	21,964,423	21,878,427	21,878,427	31,620,468	31,620,468	31,496,824	31,496,824
Brand-border-store FE	Y	Y	Y	Y	Y	Y	Y	Y
Brand-border-month FE	Y	Y	Y	Y	Y	Y	Y	Y
Store-border-month FE	Y	Y	Y	Y	Y	Y	Y	Y
Rival advertising	Y	Y	Y	Y	Y	Y	Y	Y
Own price	Y	Y	Y	Y	Y	Y	Y	Y
Rival price	Y	Y	Y	Y	Y	Y	Y	Y

Notes. Mean volume sales (number of 12–packs) is 49.46 in light and 19.43 in non-light beer. Robust standard errors in parentheses are two-way clustered at the DMA and border segment level for all specifications. Significance levels: *10 percent, **5 percent, ***1 percent. 2011–2015 Nielsen RMS and Kantar Media Ad$pender. See the text for further details.

Figure 1.8. Effectiveness of own traditional advertising on light and non-light mass-produced lager beer sales: alternative specifications



Notes. The figure shows long-run elasticities of demand (point estimates and error bars) with respect to own traditional advertising spending across the eight specifications. Blue bars represent light beer, green bars represent non-light beer. BR is border region sample; ABCP is adjacent border county-pairs sample; Stock is advertising stock; and Flow is contemporaneous advertising. PGA refers to the specifications that lumps direct and general advertisements following a proportional allocation rule; RGA refers to the specifications where direct and general advertisements enter the model separately and additively. 2011–2015 Nielsen RMS and Kantar Media Ad$pender.

The results show that there is sizable within-industry heterogeneity in own traditional advertising effectiveness across all eight specifications. The elasticities range between 0.020 and 0.030 for light and 0.015 and 0.019 for non-light MPLB. All of these coefficients are statistically significant at the 1 percent level. The differential effectiveness between the light and non-light beer is robust across all eight specifications and ranges from 0.003 to 0.015.

More specifically, the differential effectiveness between light and non-light beer ranges between 0.006 and 0.015 where advertising stocks are modeled (specifications

1, 2, 5, and 6) and between 0.003 and 0.010 in models where dynamics are captured by lagged volume sales (specifications 3, 4, 7, and 8). The estimated differential is larger for odd-numbered specifications—which incorporate general advertisements proportional to the volume shares of the affected brands (0.010–0.015)—relative to the differential observed in even-numbered specifications, in which general advertisements enter the model additively (0.003–0.007). Finally, the differential is virtually the same, regardless of which bordering market approach is adopted: the BR approach (specifications 1–4, with a differential of 0.003–0.015) or the ABCP approach (specifications 5–8, with differential of 0.003–0.014).

I note that the main finding of the within-industry heterogeneity in own traditional advertising effectiveness is not statistically significantly different across the eight specifications. I choose specification 1 as the baseline model for two reasons: First, it is the most parsimonious specification as it (i) captures direct and general advertising jointly and (ii) captures the long-run effects in a more comprehensive way—the stock formulation accounts for both the carryover effect (the impact of past advertising on current sales) as well as state dependence (e.g., the addictive nature of beer). Second, the BR sample has a smaller sample size than the ABCP sample (21,964,423 vs. 31,620,468 observations), which alleviates the computational burden in subsequent analyses.

1.5.1.2 Advertising effects by media channel

Table 1.8 provides the estimates from Equation **??**. I further report the estimates as elasticities in Table 1.9 and Figure 1.9. The decomposition results show that nearly all traditional media channels are more effective in driving the beer sales in the light compared to the non-light segment. The only exception is radio, where the effectiveness (0.008) is larger in the non-light segment and not statistically distinguishable from zero for the light beer (0.003).

Table 1.8. Long-run own traditional advertising effects by traditional advertising channels: light vs. non-light mass-produced lager beer

TV advertising	0.808***
	(0.116)
TV advertising × light	0.171
	(0.176)
Print advertising	0.338
	(0.308)
Print advertising × light	6.732***
	(0.517)
Radio advertising	9.678***
	(1.972)
Radio advertising × light	-3.846
	(3.015)
Outdoor advertising	0.994***
	(0.207)
Outdoor advertising × light	3.917***
	(1.350)
Observations	21,964,423
Brand-border-store FE	Y
Brand-border-month FE	Y
Store-border-month FE	Y
Rival advertising	Y
Own price	Y
Rival price	Y

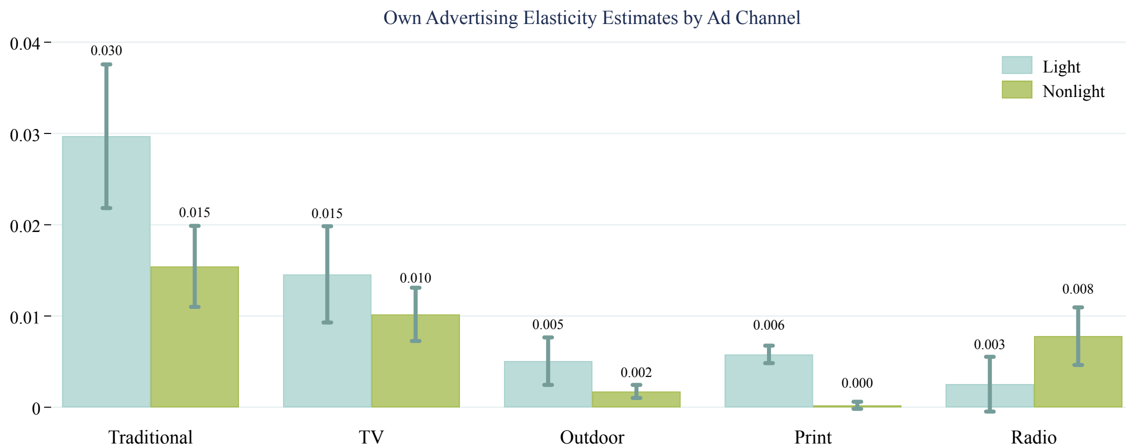
Notes. The reported estimates are long-run own advertising effects from different traditional outlets: TV, print, radio and outdoor. These estimates are from an enriched triple difference model that are estimated regressing volume sales of mass-produced lager beer brands for each beer type (light and non-light) on own advertising stocks per each outlet (TV, print, radio, outdoor) and covariates (rival advertising, own price, rival price). The model includes brand-border-store fixed effects, brand-border-month fixed effects, and store border-month fixed effects. See Equation **??** for the regression equation. Robust standard errors in parentheses are two-way clustered at the DMA and border segment level. Significance levels: *10 percent, **5 percent, ***1 percent. 2011–2015 Nielsen RMS and Kantar Media Ad$pender.

Table 1.9. Decomposition of long-run own traditional advertising elasticities and sample means by traditional advertising channels: light vs. non-light mass-produced lager beer

	Traditional	TV	Print	Radio	Outdoor
Panel A: Long-run own advertising elasticity					
Light	0.030***	0.0146***	0.0058***	0.0025	0.0051***
	(0.004)	(0.0027)	(0.0005)	(0.0015)	(0.0013)
Non-light	0.015***	0.0102***	0.0002	0.0078***	0.0017***
	(0.002)	(0.0015)	(0.0002)	(0.0016)	(0.0004)
Panel B: Mean advertising ($/100 *adults*)					
Light	0.85	0.74	0.04	0.02	0.05
Non-light	0.31	0.24	0.01	0.02	0.03
Brand-border-store FE	Y	Y	Y	Y	Y
Brand-border-month FE	Y	Y	Y	Y	Y
Store-border-month FE	Y	Y	Y	Y	Y
Rival advertising	Y	Y	Y	Y	Y
Own price	Y	Y	Y	Y	Y
Rival price	Y	Y	Y	Y	Y

Notes. Mean volume sales (number of 12–packs) is 49.46 in light and 19.43 in non-light beer. Robust standard errors in parentheses are two-way clustered at the DMA and border segment level for all specifications. Sample size equals 21,964,423 for both specifications that are estimated using Equation 1.1 and Equation **??**. Panel A reports the long-run own traditional advertising elasticity (column labeled Traditional) and elasticities for four traditional outlets (columns labeled TV, Print, Radio, and Outlet). Traditional advertising elasticity is calculated by multiplying the coefficients on own advertising stocks (reported in Column 1 of Table 1.6) by the mean advertising sales ratio for each beer type. The elasticities for advertising outlets are calculated by multiplying the coefficients on own advertising stock channels (reported in Table 1.8) by the mean advertising sales ratio for each channel and beer type. Panel B reports mean advertising ($/100 drinking-age adults, 21+) for total traditional ad spending and individual advertising outlets. Because mean values for each advertising channel are different, the individual channel elasticities in Panel A do not add up to total traditional elasticity. Robust standard errors in parentheses are two-way clustered at the DMA and border segment level. Significance levels: *10 percent, **5 percent, ***1 percent. 2011–2015 Nielsen RMS and Kantar Media Ad$pender.

Figure 1.9. Heterogeneity in traditional own advertising effectiveness by channel: light vs. non-light mass-produced lager beer



Notes. The figure shows long-run elasticities of demand (point estimates and error bars) with respect to own traditional advertising spending per each channel in baseline specification (i.e., I use border region sample and advertising stocks with δ=0.3 and lump direct and general advertising). Traditional is total traditional advertising (i.e., the sum of TV, outdoor, print, and radio advertising) and included in the figure for comparison even though it is not included in regression (Equation **??**). Note that decomposition elements do not necessarily add up to the total traditional effectiveness (denoted by Traditional), as the total is a weighted average of the individual decomposition components in the current figure (i.e., comes from the regression in Equation 1.1). Blue bars represent light beer, green bars represent non-light beer. 2011–2015 Nielsen RMS and Kantar Media Ad$pender.

Figure 1.9 shows that about half of the effectiveness of light beer own advertising in the baseline model (0.030) comes from TV advertisements (0.015). Notably, print and outdoor advertisements are also effective in increasing sales (0.006 and 0.005, respectively). For non-light beer, TV advertisements explain two thirds of the effectiveness measured in the baseline model (i.e., 0.010 out of 0.015), with radio being almost as important (0.008).

Decomposition findings provide some empirical support for the general tendency to use TV advertisements when measuring traditional media effectiveness. Furthermore, TV advertising spending accounts for more than 85 percent of all traditional media advertising for light beer and nearly 80 percent for non-light beer, as shown in Panel C of Table 1.9. The remainder is shared by outdoor (6 percent for light vs. 10 percent for non-light), print (5 percent for light vs. 3 percent for non-light), and radio (2 percent for light vs. 7 percent for non-light). This is why the magnitudes of the own elasticities in the decomposition elements do not necessarily add up to the total traditional effectiveness, as the total is a weighted average of the individual decomposition components.

1.5.1.3 Within-industry heterogeneity by demographic group

Table 1.10 reports the results of the model set forth in Equation 1.7. As before, for ease of interpretation, I further report elasticities in Table 1.11 and Figure 1.10.

Table 1.10. Heterogeneity in own traditional advertising effects by demographic group: light vs. non-light mass-produced lager

Advertising	0.901***	Asian	-0.144
	(0.135)		(0.157)
Advertising × light	0.804***	Asian × light	-0.518**
	(0.191)		(0.206)
Millennials	-0.217***	Hispanic	0.183
	(0.068)		(0.172)
Millennials × light	0.824***	Hispanic × light	-0.617***
	(0.190)		(0.220)
Generation X	0.474***	Urban	-0.050
	(0.114)		(0.146)
Generation X × light	-0.152	Urban × light	0.502**
	(0.217)		(0.216)
Women	0.152**	Wages	0.323**
	(0.075)		(0.144)
Women × light	-0.010	Wages × light	-0.055
	(0.101)		(0.173)
African American	-0.105	Employment	-0.153
	(0.137)		(0.102)
African American × light	-0.095	Employment × light	0.094
	(0.103)		(0.156)

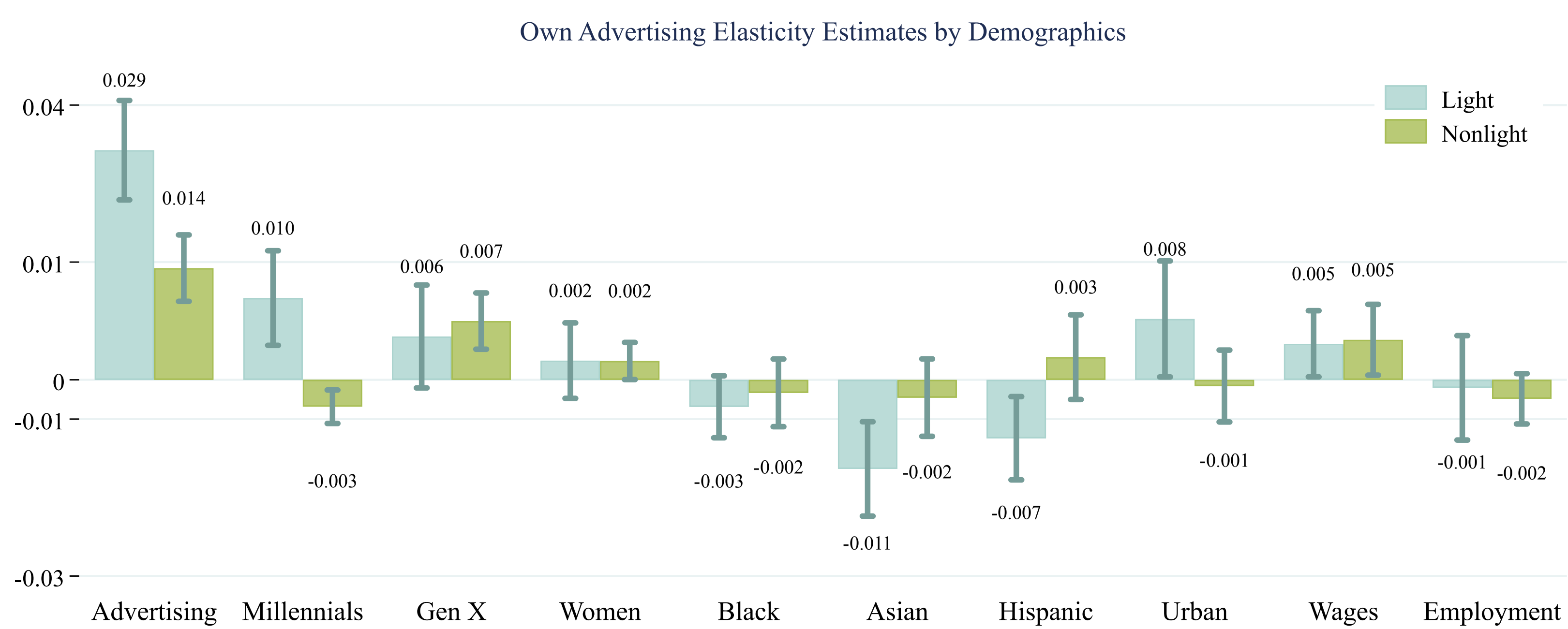
Notes. Sample size equals to 21,964,423. Robust standard errors in parentheses are two-way clustered at the DMA and border segment level for all specifications. Significance levels: *10 percent, **5 percent, ***1 percent. 2011–2015 Nielsen RMS and Kantar Media Ad$pender.

Table 1.11. Heterogeneity in long-run own traditional advertising elasticities by demographic group: light vs. non-light mass-produced lager beer

	Advertising	Millennials	Generation X	Women	African American
Light	0.029***	0.010***	0.006*	0.002	-0.003*
	(0.003)	(0.003)	(0.003)	(0.002)	(0.002)
Non-light	0.014***	-0.003***	0.007***	0.002**	-0.002
	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)
	Asian	Hispanic	Urban	Wages	Employment
Light	-0.011***	-0.007***	0.008**	0.005**	-0.001
	(0.003)	(0.003)	(0.004)	(0.002)	(0.003)
Non-light	-0.002	0.003	-0.001	0.005**	-0.002
	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)
Brand-border-store FE	Y	Y	Y	Y	Y
Brand-border-month FE	Y	Y	Y	Y	Y
Store-border-month FE	Y	Y	Y	Y	Y
Rival advertising	Y	Y	Y	Y	Y
Own price	Y	Y	Y	Y	Y
Rival price	Y	Y	Y	Y	Y

Notes. Sample size equals to 21,964,423. Robust standard errors in parentheses are two-way clustered at the DMA and border segment level for all specifications. Significance levels: *10 percent, **5 percent, ***1 percent. 2011–2015 Nielsen RMS and Kantar Media Ad$pender.

Figure 1.10. Heterogeneity in traditional own advertising effectiveness by demographic group: light vs. non-light mass-produced lager beer



Notes. The figure shows long-run elasticities of demand (point estimates and error bars) with respect to own traditional advertising spending among demographics in baseline specification (i.e., I use border region sample and advertising stocks with δ=0.3 and lump direct and general advertising). The bars (except Advertising) represent interactions of own traditional advertising with the shares of millennials, Generation X, women, African Americans, Asians, Hispanics, and urban populations as well as with average wage earnings and total employment. Advertising is the mean traditional advertising for the remaining demographics among U.S. adults (i.e., baby boomers, silent generation, men, white, non-Hispanic, and rural). All demographics are standardized to mean 0 and standard deviation 1 to make the interpretation more intuitive (i.e., I evaluate advertising effectiveness with respect to a 1-standard-deviation increase in each demographic variable). Blue bars represent light beer, green bars represent non-light beer. 2011–2015 Nielsen RMS and Kantar Media Ad$pender.

The results show that light beer advertising is more effective in areas that have a greater share of millennials (0.039) compared to areas with a greater share of older generations (0.035 for Generation X and 0.029 for the combined reference group of baby boomers and the silent generation). Furthermore, light beer advertisements appear to be more effective in areas with greater percentages of white population (around 0.029) compared to African American (0.026) and Asian (0.018) populations, less effective for a larger share of Hispanic populations (0.022), and more effective in urban areas (0.037).

On the other hand, traditional non-light beer advertisements seem to be less effective for millennials (0.011) but more effective for women (0.016). The results also show that traditional advertisements from both types of beer are more effective among wealthier individuals. Overall, these results suggest that there is a demographic component to within-industry heterogeneity in the MPLB sector. Light beer advertisements appear to work better in areas with greater populations of non-ethnic, white, and millennial urban dwellers.

1.5.1.4 Rival advertising, own price, and lagged sales

In Table 1.12, I report the long-run rival traditional advertising elasticities in the two beer segments across the eight specifications. The results show brand spillovers from the same beer segment in both light (0.017–0.054) and non-light beer (0.049–0.060) across all eight specifications. These spillovers are statistically significant at the 1 percent level in both beer types, and the magnitudes appear to be larger for non-light beer. However, the difference in spillover effects between the two types is not statistically significant at the conventional levels in any of these specifications. Therefore, these results do not show within-industry heterogeneity in rival advertising effectiveness, where the rivals only come from the same beer category (e.g., other light beer brand rivals for a light beer brand). As I explain in Appendix

A, my data do not provide sufficient variation to directly identify different-category rival advertising effects.

Table 1.12. Long-run rival traditional advertising elasticities and sample means across eight specifications: light vs. non-light mass-produced lager beer

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: Long-run rival advertising elasticity								
Light	0.023***	0.054***	0.020***	0.040***	0.020***	0.049***	0.017***	0.036***
	(0.0044)	(0.0084)	(0.0041)	(0.0066)	(0.0045)	(0.0082)	(0.0038)	(0.0062)
Non-light	0.053***	0.059***	0.049***	0.057***	0.055***	0.060***	0.049***	0.056***
	(0.0065)	(0.0073)	(0.0048)	(0.0057)	(0.0062)	(0.0068)	(0.0045)	(0.0052)
Panel B: Mean rival advertising ($/100 adults)								
Light	7.12	6.99	4.99	4.9	7.06	6.94	4.94	4.86
Non-light	4.64	4.19	3.26	2.94	4.57	4.14	3.2	2.9
Observations	21,964,423	21,964,423	21,878,427	21,878,427	31,620,468	31,620,468	31,496,824	31,496,824
Brand-border-store FE	Y	Y	Y	Y	Y	Y	Y	Y
Brand-border-month FE	Y	Y	Y	Y	Y	Y	Y	Y
Store-border-month FE	Y	Y	Y	Y	Y	Y	Y	Y
Own advertising	Y	Y	Y	Y	Y	Y	Y	Y
Own price	Y	Y	Y	Y	Y	Y	Y	Y
Rival price	Y	Y	Y	Y	Y	Y	Y	Y

Notes. Mean volume sales (number of 12–packs) is 49.46 in light and 19.43 in non-light beer. Robust standard errors in parentheses are two-way clustered at the DMA and border segment level for all specifications. Significance levels: *10 percent, **5 percent, ***1 percent. 2011–2015 Nielsen RMS and Kantar Media Ad$pender.

Even though the main focus of this study is to measure long-run advertising effects, I note that within-industry heterogeneity in own price effectiveness exists and is statistically significant at the 1 percent level. Figure 1.11 illustrates the own price elasticities in the MPLB sector. Across all eight specifications, demand is more price elastic in light beer (ranging between -1.687 and -1.775) than in non-light beer (ranging between -1.290 and -1.515). Light beer is cheaper ($9.46/12-pack vs. $10.76/12-pack for non-light) and therefore consumed more intensely by lower income consumers (Schiff et al., 2021), a feature that is consistent with the greater price sensitivity observed in the light segment. By contrast, the relatively less price-sensitive consumers in the non-light beer segment might suggest that consumers are more willing to pay a premium, say, for the wider spectrum of taste and variety that typically characterize this beer segment.

Figure 1.11. Effectiveness of own price on light and non-light mass-produced lager beer sales: alternative specifications

Notes. The figure shows elasticities of demand (point estimates and error bars) with respect to own price across the eight specifications. Blue bars represent light beer, green bars represent non-light beer. BR is border region sample; ABCP is adjacent border county-pairs sample; Stock is advertising stock; and Flow is contemporaneous advertising. PGA refers to the specifications that lumps direct and general advertisements following a proportional allocation rule; RGA refers to the specifications where direct and general advertisements enter the model separately and additively. 2011–2015 Nielsen RMS and Kantar Media Ad$pender.

Last, in addition to own advertising effects, Table 1.6 further reports the carryover coefficients for specifications 3, 4, 7, and 8. This coefficient has a similar interpretation as the depreciation parameter, δ, in the ad stock formulation provided in Equation 1.2, and it implies the persistence of sales, as in Shapiro (2018). The results show that the carryover effects are smaller in the light compared to the non-light segment, despite the finding that light beer advertisements are more effective in the short run. A potential explanation is that the carryover coefficient might be picking up habit formation, since non-light beer products contain more alcohol than light beer products and can be more addictive (see Tremblay and Tremblay, 2005).

1.5.2 Robustness checks

1.5.2.1 Identification of price effects using modified Hausman instruments

I include price in my model because pricing strategies are important determinants of demand. Similar to the nature of advertising decisions, firms do not make their pricing decisions randomly but choose prices based on anticipated demand. The common period fixed effects at the bordering markets (brand-border-month and store-border-month) would also deal with bias due to confounders as long as these confounders are correlated with both price and demand.

However, as an additional sensitivity check, I use a modified version of Hausman price instruments used in Nevo (2000). Specifically, I compute price instrument for a brand-store-week observation by taking the average price of that brand in other local markets (but preferably in the same retail chain) in a given week. While restricting the sample to distant markets ensures that unobserved shocks will be uncorrelated between the markets, the prices would be more similar when brand prices are averaged out using stores from the same retail chain in the same local market, as noted by Hitsch, Hortacsu and Lin (2019). Furthermore, I create different tiers of pairs of distant local markets and retail chains to compute the best possible instruments with available data. Whenever possible, I first use data from different DMAs, then from different counties in the same DMA, and finally from different stores in the same county. In each case, I prioritize the set of stores from the same retail chain.

Table 1.13 describes the distant markets used for the price instruments by each tier (Column 1) and reports the share of price data instrumented by each tier in the Nielsen RMS sample (Column 2). Nearly 91 percent of all prices for a brand-store-week combination are instrumented by using average price data from stores in a different DMA and from the same retail chain (Tier 1), while 8.32 percent are from stores in a different DMA and from a different retail chain (Tier 2). All remaining tiers represent less than 1 percent of the price instruments. The table further reports that only Tier 1 (96 percent) and Tier 2 (4 percent) price instruments are left in the border region sample (Column 3).

Table 1.13. Modified Hausman instruments in 2011–2015 mass-produced lager beer

Distant market tier	Distant market description for the instruments	N observations and percent share in the Nielsen RMS	N observations and percent share in the Border Region (BR) sample
Tier 1	Same retailer, different DMA	54,197,629 (90.99 percent)	21,073,707 (95.94)
Tier 2	Different retailer, different DMA	4,955,414 (8.32 percent)	890,716 (4.06)
Tier 3	Same retailer, different county	389,285 (0.65 percent)	N/A
Tier 4	Different retailer, different county	22,556 (0.04 percent)	N/A
Tier 5	Same retailer, different store	1,252 (0.00 percent)	N/A
Tier 6	Different retailer, different store	105 (0.00 percent)	N/A
Tier 7	Same store	417 (0.00 percent)	N/A

Notes. The table describes the distant markets used for the price instruments by each tier (Column 1) and reports the share of price data instrumented by each tier in the Nielsen RMS sample (Column 2). 2011–2015 Nielsen RMS.

Table 1.14 reports the estimates of own price and own advertising effects, comparing the baseline specification (Column 1) with a version in which Hausman price instruments are used (Column 2). The increase in the absolute magnitude of the price coefficient is consistent with the reduction in bias as noted in the literature, while the findings of the own advertising effects in each beer segment as well as the within-industry heterogeneity in own advertising effects are hold.

Table 1.14. Own advertising and own price effects with Hausman price instruments: light vs. non-light mass-produced lager beer

	No Price Instruments (Baseline Model)	Hausman Price Instruments
Advertising	0.978***	0.851***
	(0.141)	(0.124)
Advertising × light	0.753***	0.740***
	(0.256)	(0.239)
Price	-2.736***	-5.493***
	(0.371)	(0.571)
Price × light	-6.318***	-11.51***
	(1.413)	(2.213)
Observations	21,964,423	21,964,423
Brand-border-store FE	Y	Y
Brand-border-month FE	Y	Y
Store-border-month FE	Y	Y
Rival advertising	Y	Y
Rival price	Y	Y

Notes. Robust standard errors in parentheses are two-way clustered at the DMA and border segment level for all specifications. Significance levels: *10 percent, **5 percent, ***1 percent. 2011–2015 Nielsen RMS and Kantar Media Ad$pender.

1.5.2.2 Robustness to the selection of the cutoff brands

All analyses in this paper use brands that make up the top 95 percent market share in each beer type-market pair.. To check the sensitivity of results to the selection of this MPLB sample, I re-estimate the baseline model varying the cumulative brand cutoff from 0.75 to 0.95 in increments of 0.05. Table 1.15 presents the advertising estimates from these alternative brand cutoffs. Column 5 represents the baseline specification for comparison. All own and rival advertising estimates are statistically significant at the 1 percent level. The within-industry heterogeneity in own advertising effects is statistically significant at the 10 percent level or better.

Table 1.15. Own and rival advertising effects from the regressions with alternative brand cutoffs: light vs. non-light mass-produced lager beer

	Top 75 percent brands	Top 80 percent brands	Top 85 percent brands	Top 90 percent brands	Top 95 percent brands
Own advertising	0.928***	0.966***	0.979***	0.977***	0.978***
	(0.146)	(0.143)	(0.142)	(0.142)	(0.141)
Own advertising × light	0.411*	0.540**	0.611**	0.696***	0.753***
	(0.211)	(0.217)	(0.233)	(0.249)	(0.256)
Rival advertising	0.256***	0.227***	0.219***	0.219***	0.220***
	(0.030)	(0.029)	(0.028)	(0.028)	(0.027)
Rival advertising × light	0.156	0.0821	0.0279	-0.0178	-0.0575
	(0.107)	(0.078)	(0.058)	(0.043)	(0.037)
Observations	14,374,910	16,972,179	18,587,261	20,210,623	21,964,423
rMSE	34.23	31.96	31.12	30.1	29.03
Brand-border-store FE	Y	Y	Y	Y	Y
Brand-border-month FE	Y	Y	Y	Y	Y
Store-border-month FE	Y	Y	Y	Y	Y
Own price	Y	Y	Y	Y	Y
Rival price	Y	Y	Y	Y	Y

Notes. Robust standard errors in parentheses are clustered at the DMA and border segment level for all specifications. Significance levels: *10 percent, **5 percent, ***1 percent. 2011–2015 Nielsen RMS and Kantar Media Ad$pender. See the text for further details.

1.5.2.3 Robustness to never-sold

As noted in Section 1.3.4, Nielsen records sales only when a purchase is made, resulting in 26.03 percent of zero sales in the data. I have imputed the UPC-level prices provided that at least one purchase is made in a given year. This helps to distinguish between product availability and consumers' no-purchase decisions. As shown in Table 1.16, about 5 percent of observations (1,132,541 out of 21,964,423) in the baseline estimation sample have zero sales.

To see the direction and magnitude of a potential bias that might result from treating zero sales as if they were missing, I re-estimated the baseline specification after removing observations with zero sales. Table 1.16 reports the results. Overall, I found that the elasticity differential between the light and non-light segments gets slightly larger if zero sales are dropped. Specifically, the light beer advertising becomes slightly more effective than the baseline (0.0299 vs. 0.0297), while the non-light advertising becomes slightly less effective (0.0149 vs. 0.0154). This suggests that consumers' no-purchase decision might vary between the two types of beer: Light beer consumers might be opting out of their no-purchase decisions more often than the non-light beer consumers, consistent with the greater effectiveness of advertisements in the light beer segment.

Table 1.16. The long-run own traditional advertising elasticities and sample means with and without zero sales: light vs. non-light mass-produced lager

	With Zero Sales (Baseline Model)	Without Zero Sales
Panel A: Long-run own advertising elasticity		
Light	0.0297***	0.0299***
	(0.00396)	(0.00393)
Non-light	0.0154***	0.0149***
	(0.00223)	(0.0022)
Panel B: Mean sales (N 12-packs)		
Light	49.46	51.33
Non-light	19.43	20.69
Panel C: Mean advertising ($/100 *adults*)		
Light	0.85	0.87
Non-light	0.31	0.32
Observations	21,964,423	20,831,882
Brand-border-store FE	Y	Y
Brand-border-month FE	Y	Y
Store-border-month FE	Y	Y
Rival advertising	Y	Y
Own price	Y	Y
Rival price	Y	Y

Notes. Robust standard errors in parentheses are clustered at the DMA and border segment level for all specifications. Significance levels: *10 percent, **5 percent, ***1 percent. 2011–2015 Nielsen RMS and Kantar Media Ad$pender. See the text for further details.

1.5.2.4 Robustness to never-advertisers

The estimation samples are constructed to represent a greater population of advertisers, including those that heavily and frequently advertise as well as those that never advertise throughout the sample period or never advertise for at least one full year. Of 122 brands (representing less than 0.4 percent of the sales) in the MPLB sample, 6 do not have any advertising throughout the sample period. However, 15 out of 122 brands never advertised for at least one year over 2011–2015, representing less than 0.6 percent of sales in the period they never advertised and less than 0.9 percent of sales throughout 2011–2015.[32]

Although they represent only a fraction of MPLB sales, the inclusion of never-advertisers among established beer brands is important to provide a more comprehensive picture of advertising effects in the industry. Throughout brewing history, many large breweries—including Anheuser-Busch InBev and Miller Coors—have purposefully cut their entire advertising expenses on various traditional advertising channels in a given time and market (Tremblay and Tremblay, 2005). However, the empirical evidence would be less reliable if the relatively small set of never-advertisers were driving the results.

To address this, I re-estimate the baseline model after removing the never-advertisers. Table 1.17 reports the results. Compared to the baseline specification that considers never-advertisers (Column 1), excluding the brands that are never advertised throughout the sample period (Column 2) or that are never advertised in at least one year during the sample period (Column 3) produces slightly larger within-industry heterogeneity between light and non-light beer, mainly as a result of smaller own advertising effects in the non-light beer sector.

[32] I note that these brands are categorized as never-advertisers in the MPLB sample only; they might be advertising on some traditional media in the top DMAs in a given year without *mass-producing* in that market. This implies that they are at the bottom 5 percent of the distribution of lager beer sales in that DMA-by-year pair, as detailed in Section 1.3.4.

Table 1.17. Own and rival traditional advertising effects with and without never-advertisers: light vs. non-light mass-produced lager beer

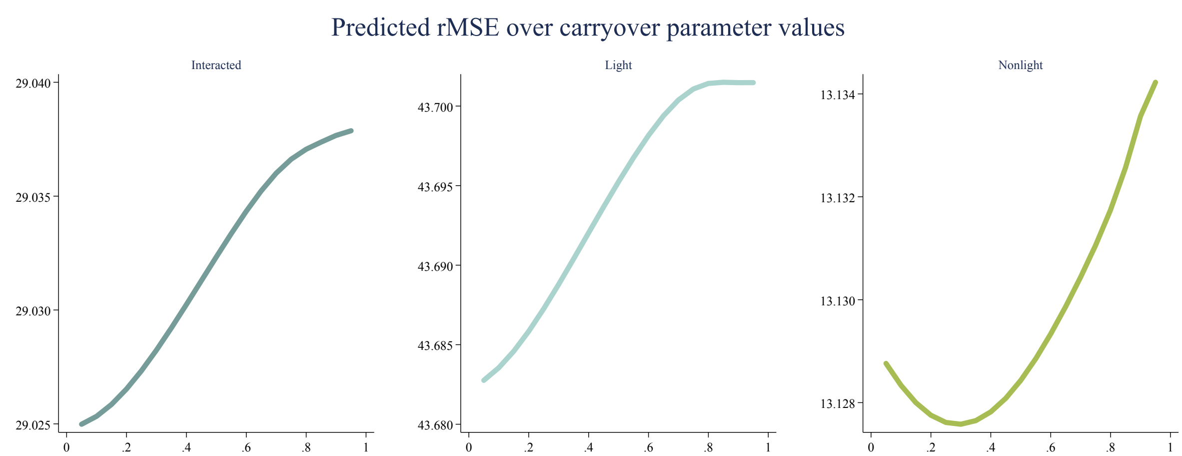
	With never-advertisers (Baseline Model)	Without those never advertised throughout the sample	Without those never advertised in at least 1-year
Own advertising	0.978***	0.974***	0.966***
	(0.141)	(0.142)	(0.141)
Own advertising × light	0.753***	0.756***	0.764***
	(0.256)	(0.257)	(0.256)
Rival advertising	0.220***	0.222***	0.229***
	(0.0271)	(0.0268)	(0.0277)
Rival advertising × light	-0.0575	-0.0597	-0.0665*
	(0.0366)	(0.0363)	(0.0368)
Observations	21,964,423	21,808,563	21,580,539
Brand-border-store FE	Y	Y	Y
Brand-border-month FE	Y	Y	Y
Store-border-month FE	Y	Y	Y
Own price	Y	Y	Y
Rival price	Y	Y	Y

Notes. Robust standard errors in parentheses are clustered at the DMA and border segment level for all specifications. Significance levels: *10 percent, **5 percent, ***1 percent. 2011–2015 Nielsen RMS and Kantar Media Ad$pender. See the text for further details.

1.5.2.5 Estimation of the depreciation parameter

I set δ, the depreciation parameter, to be 0.3 in the baseline model. To see how the model fits the data when δ changes, I estimate δ using a grid search similar to that in Shapiro, Hitsch and Tuchman (2021). Specifically, I first use a grid from 0 to 1 in increments of 0.05, excluding the boundaries of 0 and 1, for δ and re-calculate the advertising stock as in Equation 1.2. Next, I estimate the interacted baseline model and the split models of the baseline specification per industry segment for each δ, for a total of 19 fixed-effects regressions. Figure 1.12 and Table 1.18 report that the best fit as provided by the predicted minimum root mean squared error (rMSE) occurs at smaller levels of the depreciation parameter in both the interacted and split models, while $\delta = 0.3$ is the optimum value in the non-light MPLB segment. Consistent with this finding, Table 1.6 shows that the carryover coefficient estimate in front of the lagged volume sales is 0.13 in the light beer segment (the 95 percent confidence interval is between 0 and 0.30) and 0.31–0.33 in the non-light segment across the specifications in which the dynamics are captured through the lagged volume sales.

Figure 1.12. Grid search estimates of the advertising depreciation parameter: interacted vs. split models in the mass-produced lager beer industry



Notes. The figure shows predicted root mean squared error (rMSE) from estimations of the interacted baseline model (the graph on the far left) and the split models of the baseline specification per industry segment (the middle graph is for light and the graph on the far right is for non-light) for each δ (i.e., depreciation parameter), for a total of 19 fixed-effects regressions. The variable on the y-axis is rMSE; the variable on the x-axis is δ. 2011–2015 Nielsen RMS and Kantar Media Ad$pender.

Table 1.18. Predicted rMSE from grid search regressions: interacted, light, and non-light mass-produced lager beer

Depreciation Parameter δ	Interacted Model	Split Model: Light MPLB	Split Model: Non-light MPLB
0.05	29.02498	43.68276	13.12876
0.10	29.02533	43.68355	13.12834
0.15	29.02585	43.68459	13.12800
0.20	29.02653	43.68584	13.12776
0.25	29.02733	43.68726	13.12762
0.30	29.02824	43.68879	13.12758
0.35	29.02922	43.6904	13.12765
0.40	29.03024	43.69203	13.12782
0.45	29.03128	43.69366	13.12809
0.50	29.03233	43.69524	13.12843
0.55	29.03336	43.69675	13.12885
0.60	29.03434	43.69815	13.12933
0.65	29.03523	43.69938	13.12987
0.70	29.03601	43.70038	13.13044
0.75	29.03662	43.70107	13.13106
0.80	29.03706	43.70142	13.13175
0.85	29.03737	43.70150	13.13257
0.90	29.03767	43.70147	13.13357
0.95	29.03787	43.70147	13.13423

Notes. The table reports predicted root mean squared errors (rMSE) that result from estimation of the baseline model (second column) for different values of the depreciation parameter δ. The table also reports predicted rMSE from the split regressions of the baseline model separately for light (third column) and non-light (fourth column) mass-produced lager beer. See the text for further details.

A smaller δ than that estimated in Shapiro, Hitsch and Tuchman (2021) (0.3 vs. 0.9) makes intuitive sense since a slower dissipation of past advertising effects on current demand is expected when a smaller set of top brands with the highest ranking is considered (their sample contains the top 11 beer brands compared to the top 122 brands in my analysis). Furthermore, Shapiro, Hitsch and Tuchman (2021) estimate regression equations separately for each CPG brand and report statistics from the distribution of own advertising elasticities which, on average, decreases the weight coming from the carryover effects that are observed in the beer alone. This is consistent with a lower δ reported in Tuchman (2019), who found that smaller values of δ predict a better fit when estimation is done at the category level for cigarettes and cessation products after aggregating a large set of brands in each category.

1.5.3 Discussion

1.5.3.1 The need for the identification on bordering markets

In the presence of unobserved spatial heterogeneity, advertising estimates would suffer from an endogeneity issue, as explained in Section 1.4. I check whether unobserved spatial heterogeneity exists and creates bias for the advertising effects by comparing the estimates from three specifications: a least saturated naïve model, a traditional fixed-effects model, and a most saturated model with border fixed effects.

Table 1.19 reports the results. Column 1 shows the own and rival traditional advertising estimates from the naïve model; Column 2 shows the estimates from a traditional fixed effects model in which I control for brand-store, brand-month, and store-month fixed effects; and Column 3 shows the estimates from the baseline model with border fixed effects.

Table 1.19. Evolution of the own and rival advertising effects from naïve to fixed effects specifications: light vs. non-light mass-produced lager beer

	Naïve	Fixed Effects	Baseline
Own advertising	18.51***	1.305***	0.978***
	(1.758)	(0.194)	(0.141)
Own advertising × light	6.984**	1.002**	0.753***
	(3.188)	(0.431)	(0.256)
Rival advertising	-0.798***	0.172***	0.220***
	(0.244)	(0.033)	(0.027)
Rival advertising × light	-1.431***	-0.044	-0.058
	(0.402)	(0.044)	(0.037)
Observations	21,964,423	21,964,423	21,964,423
rMSE	57.64	29.76	29.03
Fixed effects	None	brand-store, brand-month, store-month	brand-border-store, brand-border-month, store-border-month
Own price	Y	Y	Y
Rival price	Y	Y	Y

Notes. Robust standard errors in parentheses are clustered at the DMA and border segment level for all specifications. Significance levels: *10 percent, **5 percent, ***1 percent. 2011–2015 Nielsen RMS and Kantar Media Ad$pender. See the text for further details.

Overall, the own advertising effects decrease while the rival effects increase as I move toward the more sophisticated alternatives. The move from the naïve to the traditional fixed effects model provides the most drastic results. In both beer segments, own advertising effects are reduced by an order of magnitude, while rival effects change direction. These results are consistent with the presence of important confounders. In the case of rival effects, the positive effects support the implication that a market expansion mechanism is at play (with advertising spillovers) as opposed to the business-stealing mechanism provided in the naïve model. However, as the switch from the fixed-effects model to the baseline shows, there are still remaining confounders with unique spatial characteristics. The magnitudes of both the own and rival effects are reduced in the baseline, though the advertising spillovers still prevail in both beer segments. Together, these results provide evidence for the presence of spatial confounders and therefore justify the use of bordering markets.

1.5.3.2 Implications of the within-industry heterogeneity in advertising

What do these differential long-run traditional advertising estimates in the two different beer segments imply? Why does within-industry heterogeneity in advertising effectiveness matter? To tackle these questions, I measure the impact of advertising effectiveness on sales revenue in each beer segment. An intuitive way to do this is to calculate the percentage change in sales revenue if advertising were to increase by 100 percent in each segment. I note that this would boost the light beer segment more than twice as much as the non-light segment since the mean advertising is \$0.85 per 100 adults in the light segment relative to \$0.31 per 100 adults in the non-light segment. To check the following calculation procedure, I refer the reader to Column 1 in Table 1.7 for the own traditional advertising elasticities and to Table 1.20 for the descriptive statistics from the baseline specification.

Table 1.20. Descriptive statistics in the baseline model: light vs. non-light mass-produced lager beer

	Light		Non-light	
	mean	s.d.	mean	s.d.
Sales (N 12-packs)	49.46	92.91	19.43	36.04
Revenue ($)	414.88	683.11	188.58	350.34
Own advertising stock ($/100 *adults*)	0.85	1.34	0.31	0.59
Rival advertising stock ($\sum_{J^-}$ $/100 *adults*)	7.12	3.05	4.64	2.25
Own price ($/12-packs)	9.46	2.4	10.76	3.34
Rival price ($/12-packs)	9.53	1.05	10.77	1.03

Notes. Sales volume is the number of 12-packs, revenue is the total dollar sales, and own advertising stock is the total direct and proportionally added general advertisements, total local and national advertisements ($/100 adults age 21 and over) with 0.3 depreciation over time. Rival advertising is the total advertising stocks of rival brands. Price is $/12-packs. Rival price is the average shelf price of rival brands. Descriptive statistics come from the border region sample, with 21,964,423 observations at brand-border-store-week level. 2011–2015 Nielsen RMS and Kantar Media Ad$pender.

More specifically, the elasticity findings imply that the firm could make an extra 1.5 light beer sales (49.45×0.03) in an average store and on a typical week by spending $1.70 per 100 adults. Similarly, by spending $0.62 per 100 adults, the firm could make an extra 0.3 non-light beer sales (19.43×0.015). This suggests $14.18 more revenue ($\$9.46 \times 1.5$) for the light beer segment relative to $3.13 more revenue ($\$10.76 \times 0.3$) for the non-light beer segment. Overall, doubling the advertisements in each beer segment implies a 3.42 percent increase in the light beer segment ($14.18/$414.88) compared to a 1.66 percent increase in the non-light segment ($3.13/$188.57).

The difference of 1.76 percentage points in extra revenue that could be generated by doubling the ads in the light beer segment relative to the non-light beer segment suggests that firms could do better if they considered the heterogeneity in advertising effectiveness between the two types of beer when making marketing decisions. Even a redistribution of allotted traditional advertising expenditures from the non-light to the light beer category, rather than increasing the marketing budget, might generate more profits in the long run.

1.6 Conclusion

Many U.S. firms did not cut spending on traditional advertising in 2020, despite the fact that many sectors were hit hard by the COVID-19 pandemic. Despite significant delays in the supply chain of consumer packaged goods industries, the demand for most products—beer among them—did not fall. As consumer demand switched from on-premise to off-premise beer sales, the recent surge in craft beer came to a sudden halt. By contrast, sales of the mass-produced lager beer (MPLB) products that populate retail store shelves experienced growth.

In this study, I provide robust empirical evidence of within-industry heterogeneity in long-run traditional advertising effectiveness in the U.S. MPLB market. Specifically, I show that own light beer advertisements are twice as effective (0.030) in generating sales as their non-light counterparts (0.015). This finding is robust across eight alternative specifications and a substantive list of sensitivity tests borrowed from the economics literature. The research design credibly demonstrates how the time-varying regional unobservables that confound both sales and advertising are swept out, clearing the way for causal estimates.

To explore the sources of heterogeneity in the effects, I first decompose traditional advertising effectiveness into TV, print, radio, and outdoor media channels. I find that print and outdoor advertising are effective in generating light beer sales but less so for non-light beer. Conversely, radio advertising appears to work only for non-light beer. Furthermore, TV advertising—though still the most effective medium for beer advertisers—explains only 33 percent of the observed differential ad effectiveness between light and non-light beer. This suggests that the common practice in the advertising literature of studying TV advertisements alone might fail to fully quantify both the magnitude and the heterogeneous character of traditional advertising effectiveness in each segment.

Second, using an enriched triple difference model, I interact advertising effects with demographic groups and examine whether the differential in advertising effectiveness between the two beer types comes from areas with greater shares of certain populations. The evidence shows that light beer advertisements are more effective for millennials than for older generations, particularly baby boomers, and more effective in areas with greater shares of non-ethnic populations, white, and urban dwellers. By contrast, non-light beer advertisements are more effective in areas with greater shares of older and female populations.

In addition to advertising effects, the industry also illustrates differential price effectiveness between the light and non-light beer segments. Across the eight specifications, the elasticity of demand with respect to own brand price ranges from -1.687 to -1.775 in the light beer segment and from -1.290 to -1.515 in the non-light segment. The presence of more price-sensitive consumers in the light beer segment makes economic sense and suggests that lower-income households consume light beer more intensively, as light beer products are less expensive ($9.46 per 12-pack) relative to non-light products ($10.76 per 12-pack). On the other hand, the presence of less price-sensitive consumers in the non-light beer segment is consistent with consumers' greater willingness to pay a price premium for, say, the wider spectrum of taste and variety that typically characterizes this beer segment.

The finding of within-industry heterogeneity in traditional advertising effectiveness implies a 1.75-percentage-point difference in extra revenue that could have been generated if firms had chosen to double their advertising spending in the light rather than the non-light beer category. Shifting a portion of the marketing budget to the light beer segment from non-light beer, or to other traditional outlets from TV, might secure more profits for the firm in the long run. To the extent that wasteful advertisements create a cost for society as a whole, this type of optimal advertising reallocation might be beneficial for the consumer as well. A fuller assessment of these implications is left to future research.

CHAPTER 2

SODIUM INTAKE AND REFORMULATION IN THE UNITED STATES: EVIDENCE FROM SCANNER DATA

2.1 Introduction[1]

Responding to ample evidence and worldwide consensus regarding the importance of a healthy diet, this paper focuses on sodium, a key factor in improving diet quality. The scientific evidence has shown a consistent and strong positive association between sodium intake and noncommunicable diseases, including high blood pressure (i.e., hypertension) and cardiovascular disease.[2] It has been estimated that reducing sodium intake could prevent 100,000 deaths per year (Danaei et al., 2009) and decrease medical expenses by $32 billion in the United States (Polar and Sturm, 2009; Smith-Spangler et al., 2010). Concerns about sodium intake in the United States have been voiced for over 50 years.[3] Similar concerns have been voiced worldwide. Most recently, the World Health Organization has targeted a 30 percent relative reduction in average sodium intake by 2025 (WHO et al., 2013).

---

[1] This chapter is a joint work with Christian Rojas.

[2] The incidence of high blood pressure in the United States is alarming: About 45 percent of the adult population suffers from hypertension (Heron, 2019), which is, in turn, associated with a significantly higher risk of cardiovascular disease (see Meyers et al., 2006; Boon et al., 2010, and references therein). Heart disease is the most common cause of death in the United States (23 percent of deaths) and stroke is the fifth most common (5.2 percent of deaths; Heron, 2019).

[3] The first recorded effort to reduce sodium content was the 1969 White House Conference on Food, Nutrition, and Health. Besides recommending reducing individual sodium intake, the conference encouraged food processors to reduce the amount of salt in their products and to include sodium content in product labels (see Boon et al., 2010, for details).

Despite numerous initiatives in the United States to curb sodium intake, the use of this dietary mineral remains largely unregulated. While there is a consensus that a healthy adult's sodium consumption should not exceed 2,300 mg/day as set forth in the 2015–2020 Dietary Guidelines for Americans, this is only a recommendation for a tolerable upper limit for consumption.[4] Furthermore, this reference value applies to an individual's total daily intake; as such, it has no direct implications for appropriate sodium content for individual products. The FDA and Administration (2021) only recently released federal guidance for recommended sodium content in products.

The most substantive attempt to curb sodium intake was achieved in 2009 by the National Salt Reduction Initiative (NSRI).[5] The NSRI, led by the New York City Health Department (NYCHD), was a broad partnership that included the American Medical Association, the American Heart Association, the American Public Health Association, and more than 40 other entities (e.g., health organizations, cities, and states). The NSRI convened many of the leading U.S. food manufacturers with the objective of promoting gradual and substantive reductions in the sodium content of packaged and restaurant foods. The overarching objective of the NSRI was to reduce population-level sodium intake by 20 percent in 5 years, to bring adults' average sodium intake (approximately 3,400 mg/day at the time) closer to the 2,300 mg/day upper limit set by the 2015–2020 Dietary Guidelines for Americans. To this end, the NSRI set targets for several groups of sodium-intensive food products (e.g., cereals, snacks, meats) for the years 2012 and 2014, to which several large food manufacturers voluntarily adhered.[6]

---

[4] This upper limit was set in accordance with the scientific standards established by the Academy of Sciences (Meyers et al., 2006).

[5] Initial work on the initiative began in 2008. The initiative was later expanded to include sugar and is now called the National Salt and Sugar Reduction Initiative (see NSSRI).

[6] We present the sodium-intensive food groups in Section 2.6.1 and provide definitions of the NSRI food groups in Table B.1. Table B.2 reports the names of the NSRI signatories, the manufacturers who voluntarily adhere to the NSRI.

Sodium, sugar, and fat are key product attributes in manufacturers' efforts to produce tastier food. Thus, voluntary agreements to reduce the amounts of these valuable attributes may be difficult to rationalize. A growing consensus for the economic rationale behind these efforts appears to be the industry's recognition of a credible threat of future government intervention (see Moss, 2013). One example is the persistent efforts by the government and consumer advocacy groups to regulate trans fat content in food. These efforts initiated in the 1990s and culminated with a ban, effective in 2015.[7]

Prior research and our results in this paper show that sodium intake has gradually decreased over the last 15 years or so (see Curtis et al., 2016; Rehm et al., 2016; Clapp et al., 2018), a period of time that coincides with the NSRI. The drivers of this decline, however, are not well understood. To fill this gap, we make two major advances to the extant literature. First, we study the extent to which the noted decline in sodium intake is due to manufacturers' reformulation efforts vis-à-vis changes in consumers' purchase habits. Further, we explore the role that demographics and socioeconomic variables play in this decomposition. Understanding the importance of each factor for different populations is particularly valuable as policy makers weigh the efficacy of reformulation interventions relative to policies aimed at altering consumer behavior. Second, we study whether the NSRI played a causal role in manufacturers' product reformulation efforts and whether the effect of the NSRI extended beyond signatory firms. We explore the impact of the NSRI policy on manufacturers' reformulations (as opposed to consumer response) for two reasons. First, the NSRI was a concerted effort between industry and policy-makers. Second, as we elaborate later, the sodium reduction targets established by the NSRI generated a particular kind of sodium-reduction incentive that, if effective, should only be observed on the supply side.

[7]More recent examples include successful efforts to tax sugary drinks and the inclusion of added sugar content in the nutrition facts label.

For our empirical exercises, we construct a comprehensive sample on nutrition and household purchases for 589,377 barcoded products in the years 2007 and 2015. As far as we are aware, no other dataset of this type has been constructed at this level of detail, breadth, and rigor. The dataset allows us to track sodium content at the finest level of product granularity (i.e., barcode level) rather than at a broader product category level; this feature allows us to fully quantify manufacturers' reformulation efforts, both through temporal changes in sodium content within a product that remains on the shelves and through product turnover (i.e., the entry and exit of products). In addition, this type of data allows us to isolate, via a decomposition analysis, the role that product reformulation, vis-à-vis changes in consumers' shopping behavior, has played in the temporal evolution of sodium intake.[8]

Our identification approach for the NSRI effects exploits the fact that the NSRI established sodium content targets that varied widely across 57 food categories. Importantly, as we explain in Section 2.3, these targets were set by policy makers with little or no influence from food manufacturers. If manufacturers responded to these exogenously set sodium targets, we would expect to see reduced sodium content in products that registered above-target sodium content in 2007 (i.e., *noncompliant* Universal Product Codes [UPCs]) and no change in products that registered below-target sodium content in 2007 (i.e., *compliant* UPCs). Further, if the agreement had its intended effect, the observed sodium reduction should be gradual for noncompliant products and more pronounced for products with a larger sodium excess.

We use a kink regression discontinuity (KRD) design to evaluate whether the predicted result is borne out by the data. This approach captures a discontinuity in the *slope* of the relation between the running variable and the outcome variable; this

[8]The decomposition method is borrowed from the productivity literature. In particular, we use consumption-weighted average measures of sodium intake to isolate the role of reformulation from the role of consumer switching. See Section 2.5.1.2 for details.

contrasts with the traditional regression discontinuity (RD) design, which is aimed at capturing a discontinuity in the *level*. KRD design is thus well-suited for our purposes since the NSRI generated gradually larger incentives to reduce sodium (the outcome variable) for products with larger amounts of excess sodium (the running variable). As far as we are aware, ours is the first study pursuing causal identification of the effect of voluntary agreements to reformulate food products in the United States.

Our main findings are as follows: First, we find that sodium intake declined by 4.73 percent over the 8–year period we study. Second, we find that reformulation and consumer shopping behavior exerted strikingly opposing effects: Had consumers maintained their spending habits at 2007 levels, manufacturers' reformulation efforts would have resulted in a 53 percent sodium intake reduction by 2015. In other words, changes in consumer shopping habits contributed to a deterioration in sodium intake that wiped out more than 90 percent of the improvement as a result of reformulation. Third, we document important sodium intake disparities across different populations and show that these disparities have worsened over time along income and racial/ethnic minority dimensions. In particular, we find that Black, Hispanic, and lower-income households experienced less pronounced improvements, despite the availability of healthier alternatives to these populations. This suggests the need for policies that are better designed to incentivize a healthier diet for more vulnerable segments of society.

Fourth, we find evidence that supports the notion that the NSRI was an important causal factor in the substantial reformulation efforts that we quantify. In particular, we find evidence of a discontinuity in slope at the threshold, where the threshold is given by the sodium targets set by the NSRI. Our regression discontinuity results indicate that the NSRI was approximately 40 percent effective in generating perfect compliance with the voluntary agreement. Further, we find evidence that the effect was present in products across all food manufacturers including both NSRI signatories and nonsignatories, indicating that the NSRI had a wider reach.[9]

The rest of the paper is organized as follows: Section 2.2 surveys the relevant literature. Section 2.3 provides institutional details on the NSRI and the role of sodium in food with a focus for motivating our identification strategy and empirical approach. Section 2.4 describes the data used in the analysis. Section 2.5 describes the methodology and Section 3.5 presents the main results of the analysis. Section 2.7 concludes.

2.2 Contribution to existing work

The number of countries with a sodium reduction initiative has grown rapidly, from 32 in 2010 to 75 in 2015 (Trieu et al., 2015). Only a few policies are mandatory, however. Mandatory measures include explicit limits on sodium content or direct taxation of sodium. A common nonmandatory policy is to set suggested sodium content targets, which manufacturers may or may not pledge to (or abide by). Other nonmandatory measures seek to change consumer behavior by raising awareness of the risks associated with high levels of sodium consumption. These policies include education campaigns as well as mandates to display sodium content on product packaging labels.

Although the evidence suggests that policies worldwide may have been effective, this assessment is not comprehensive. Of the 75 initiatives, only 13 have been empirically evaluated, registering reductions in sodium intake that range between 4.9 percent and 36 percent. However, only six country-level policies have been studied using actual measures of sodium intake, including urinary sodium excretion and, in two cases, actual food purchases. Evaluations in the remaining seven countries

[9] This finding is consistent with the informal accounts provided by NSRI officials. We discuss these accounts in Sections 2.3.2–2.3.4).

have been based on approximations of the intake using data from self-reported food consumption surveys. Our work adds significantly to this literature.

Two studies are closely connected to our work. Curtis et al. (2016) employ a sample of the most representative barcoded packaged food products to describe the sales-weighted sodium content before and after NSRI implementation. Using sales data from retailers, the authors find a 6.8 percent overall reduction in sales-weighted sodium content, falling short of the targeted 20 percent reduction. Using a similar dataset, Clapp et al. (2018) find an average reduction of 6 percent in sodium content across 2,979 UPCs between 2009 and 2015.

Our work differs from (and adds substantially to) this prior work. First, by covering the near-universe of UPCs (our sample is about 200 times larger than in earlier work) we capture households' food baskets more precisely and reduce sampling error. Second, we construct a comprehensive sample of household-level food purchases and nutrition. This allows us to speak about per capita consumption, carry out decomposition analysis, and study the role of socioeconomic and demographic factors. Finally, we go beyond descriptive analyses and pursue causal identification of the NSRI's effects on product reformulation, a question that has not been addressed previously.

A third related study is Griffith, O'Connell and Smith (2017), which reports that the 5.1 percent reduction in sodium intake observed in the United Kingdom between 2005 and 2011 was largely driven by reformulation efforts from manufacturers. Using a decomposition method similar to ours, the authors find that reformulation alone would have resulted in a 6.5 percent reduction in sodium consumption; in other words, consumers switched to saltier products over time, thereby negating 1.4 percent percentage points of the improvement coming from reformulation efforts. We find a similar overall improvement in the United States, but the components are of a much larger magnitude: Reformulation alone has contributed to a 53 percent decline in sodium intake, whereas consumer switching toward saltier products wiped out all but 4.73 percentage points.\footnote{One reason why our estimates differ from those in the United Kingdom could be due to U.S. diet quality being inferior to that in the United Kingdom. Dubois, Griffith and Nevo (2014), for example, show that U.S. households consume a substantially higher amount of prepared foods (where sodium content is high) than U.K. households. This difference implies, for instance, that for a given reformulation amount (i.e., reduction in sodium across products), a larger sodium reduction would be observed in the United States relative to the United Kingdom.} An important aspect in which our work complements the Griffith, O'Connell and Smith (2017) study is that we also quantify whether a voluntary initiative to reformulate had an effect on product-level sodium content. More generally, as far as we are aware, ours is the first study pursuing causal identification of the effect of voluntary agreements to reformulate food products in the United States.

More broadly, our study is related to work aimed at explaining when manufacturers' self-regulation (or voluntary regulation) could be effective. Empirical evidence suggests that voluntary regulation can be effective in some cases (Sharma, Teret and Brownell, 2010; Lim et al., 2020). An important determinant of success appears to be the existence of credible external threats. In the food industry, these could take the form of negative public attitudes or likely government action (e.g., taxation or reformulation mandates). An example is trans fat regulation. Regulatory action, including a mandate to disclose trans fat content, and media attention surrounding the health risks of trans fat consumption originated in the 1990s; the increasing scrutiny of trans fat culminated in 2015 with a government decision to ban it from packaged foods. Yet food manufacturers engaged in active reformulation (trans fat removal) years before the ban was implemented (Unnevehr and Jagmanaite, 2008). An interpretation of our findings in light of this literature is that the NSRI, together with worldwide support for reducing sodium intake, was likely perceived as a credible external threat of stringent future regulation (e.g., mandatory upper limits on sodium content).

Another strand of relevant literature includes the effectiveness of government interventions to improve consumer choice. Some of these interventions aim to equip consumers with general nutrition knowledge. An example is product labeling, which helps consumers obtain standardized nutrition information across products. So far, the evidence on whether this regulation has been successful in improving consumers' dietary choices appears to be mixed. While there is some evidence that dietary choices may not improve as a result of nutrition labeling (Bollinger, Leslie and Sorensen, 2011; Moorman, Ferraro and Huber, 2012), labeling may be effective when conveyed in a simpler fashion or in the form of a warning (e.g., Kiesel and Villas-Boas, 2013; Zhu, Lopez and Liu, 2016; Alé-Chilet and Moshary, 2020).

Finally, our paper is related to a handful of studies that have used barcode-level data to assess the healthfulness of consumer food choices. Dubois, Griffith and Nevo (2014) examine how household food baskets differ across developed nations (the United States, the United Kingdom, and France), whereas Allcott et al. (2019) study healthfulness disparities across populations in the United States. Other related papers in this vein are Hut and Oster (2018), who analyze the effects of diet-related news on food choices, and Cengiz, Rojas and Wang (2020), who address how the quality of food choices by American households has evolved over time.

2.3 The role of sodium in food manufacturing and the NSRI

In this section, we provide institutional details regarding the role of sodium in food manufacturing and the implementation of the NSRI. These details provide context for our empirical exercises and help explain our identification strategy for the role of NSRI in product reformulation.

2.3.1 The role of sodium in food manufacturing[10]

Sodium is popular as a food additive for a variety of reasons. First, sodium is a taste enhancer. For example, sodium can mask undesirable flavors (e.g., bitterness) or enhance sweetness (e.g., in ready-to-eat cereals). Second, sodium can improve products' texture (e.g., stickiness in dough, tenderness in meat, meltability in cheese, thickening in sauces). Third, sodium can be a key factor for food safety: It is an inhibitor of water activity, thereby aiding in food preservation.[11]

The role of sodium varies widely across categories. For some products, sodium's primary role is as a tissue binder, providing the distinctive texture of cured meats, for example. Sodium is also an effective preservative, inhibiting bacterial growth, as in hard cheeses. There are some products (e.g., butter, rice, pasta, vegetable juices) in which sodium's unique purpose is to enhance flavor. In some products, such as salad dressing, sodium performs all three roles. As a result, sodium content varies widely across food categories: Table 2.1 displays the distribution of sodium content (in mg/100 g) across UPCs in selected food categories.

---

[10] Part of this section draws from scientific evidence reviewed in the National Academies report on sodium (Boon et al., 2010, Chapters 3 and 4).

[11] For instance, sodium benzoate limits the growth of bacteria, mold, and other microbes. It is commonly used in acidic foods (e.g., soda, pickled foods, salad dressings, and soy sauce).

Table 2.1. UPC-level sodium content (mg/100 g) in 2007 and the 2014 NSRI Target

Description	Percentile (2007)					St. Dev. (2007)	Mean (2007)	2014 NSRI target (w.r.t. 2007 mean)
	10%	25%	50%	75%	90%			
2.2 Breakfast cereals	403	536	536	592	760	189	545	-39%
3.1 Cold cuts	776	970	1,502	1,658	1,658	416	1,319	-25%
4.1 Grated hard cheese	1,786	1,980	1,980	1,980	2,027	183	1,975	-15%
7.1 Flavored chips	335	467	635	700	811	196	605	-40%
11.2 Canned vegetables	129	192	221	279	412	198	261	-41%

Notes: The table reports the distribution of UPC-level sodium content in 2007 and the 2014 NSRI target for selected NSRI food categories. Coefficients of variation are 0.35 for breakfast cereals, 0.32 for for cold cuts, 0.09 for grated hard cheese, 0.32 for flavored chips, and 0.76 for canned vegetables. Nutrition data from Gladson and NSRI targets from NSRI.

While the food-specific role of sodium may explain its varying levels across food categories, it falls short of explaining the large variance in sodium content that is frequently observed *within* narrowly defined food categories. To illustrate this point, Table 2.1 reports descriptive statistics (across UPCs) for a sample of food categories targeted by the NSRI in 2007.[12] Except for grated hard cheese, all categories exhibit a large standard deviation with respect to the mean and median. One takeaway of these patterns is that sodium content reduction should be feasible as products located on the right-hand side of the distribution could technically reduce sodium content without compromising food safety. Further, greater reductions could be achieved in food categories that exhibit a distribution with a longer and fatter right tail. This observation is in line with the logic used by the NSRI to set sodium reduction targets, a point to which we turn next.

2.3.2 The NSRI: target setting

The implementation of the NSRI in 2009 was preceded by a year of work led by the New York City Health Department (NYCHD).[13] The process began with a proposed set of food categories that would be subject to the commitment as well as a proposed target for products in that category set in terms of milligrams of sodium per 100 g of product. This initial proposal was crafted by staff at the NYCHD based on data similar to those employed in this paper. First, products with high sodium content were identified and lumped into categories based on product similarity. Then, a sodium content target (discussed below) was proposed for each category. The set of food categories and targets was disseminated to industry participants who then had the opportunity to comment in a series of technical meetings. There appears

[12] Appendix Table B.1 lists the full set of food categories targeted by the NSRI.

[13] We thank Elizabeth Solomon, Director of Nutrition and Policy Programs, and Andrea Sharkey, NSRI Program Director, for their valuable time in providing this information.

to have been extensive discussion with industry but limited disagreement. To the extent that industry objections arose, they focused on category definitions rather than on actual target levels.[14] For example, industry requested that hard cheese be categorized separately since it requires substantially higher amounts of sodium to ensure preservation compared with other cheeses.

Table 2.2. National Salt Reduction Initiative food groups and targets

Food Group	2009 Baseline	2012 Target	2014 Target
1. Bakery	530	366.9	385.7
2. Cereal	585.5	485	380
3. Meats	1,069.8	982.5	873.8
4. Dairy	869.2	810	716
5. Fats and Oils	764.5	697.5	610
6. Sauces	698.2	636	538
7. Snacks	852.3	720	592.5
8. Soups	339	300	245
9. Potatoes	1037	885	725
10. Mixed Dishes	521.4	467.5	398.8
11. Vegetables	300.2	258	206
12. Legumes	372.5	330	275
13. Canned Fish	374	350	330
14. Seasonings	415	350	290
15. Nut Butters	457	410	340

Notes. The table shows average sodium content in mg/100 g across the 15 NSRI food groups for 2009 baseline, 2012 target, and 2014 target. The values are the averages for the corresponding 57 NSRI categories. Except *meats*, *dairy*, and *canned fish*, all groups are targeted to reduce sodium content at least by 20 percent by 2014. Table B.1 reports the baseline and target sodium content in mg/100 g for the 57 NSRI food categories and their correspondence with NSRI food groups. Data come from NSRI.

[14] This is corroborated by the National Academies report on sodium (Boon et al., 2010, Appendix G).

In the end, NYCHD identified 57 sodium-intensive food categories and cataloged them in 15 broad groups. Table 2.2 shows the average sodium content in 2009 and associated 2012 and 2014 targets for these 15 NSRI food groups. *Cereal* was targeted for the greatest reduction in sodium content (35.1 percent) by 2014, while *canned fish* had the smallest target (11.7 percent). Most food groups (12 of 15) were targeted to reduce sodium content by more than 20 percent by 2014, with a 26 percent median reduction target across all food groups. The intermediate NSRI targets were less aggressive, ranging from 6 percent for *canned fish* to 30 percent for *bakery products*, with an 11 percent median reduction target by 2012.

Determining the NSRI targets incorporated and leveraged the institutional facts described in the previous subsection. First, the targets recognized that the importance of sodium as a food additive varies widely depending on the type of product. Second, greater reductions were pursued in categories where sodium reductions were less likely to interfere with product viability or safety (i.e., where dispersion was greater). Another way to see this logic is that if many products existed on the lower end of the distribution, it should be possible for products on the higher end of the distribution to significantly lower their sodium content and still make competition viable. The last three columns of Table 2.1 illustrate this relationship: Food categories with greater dispersion in sodium content in 2007 were subject to greater target reductions in sodium content. We argue that the way in which the NSRI set targets creates an opportunity to establish credible causal inference. Specifically, if the NSRI had no bite whatsoever, one would expect changes in sodium content over time to be relatively similar on both sides of the threshold (i.e., the NSRI targets), particularly for products nearest to the threshold, where they are most comparable. This setting allows us to test whether the NSRI perturbed the pattern of sodium reductions that would have been observed absent the NSRI. Section 2.5 provides more details on this identification strategy.

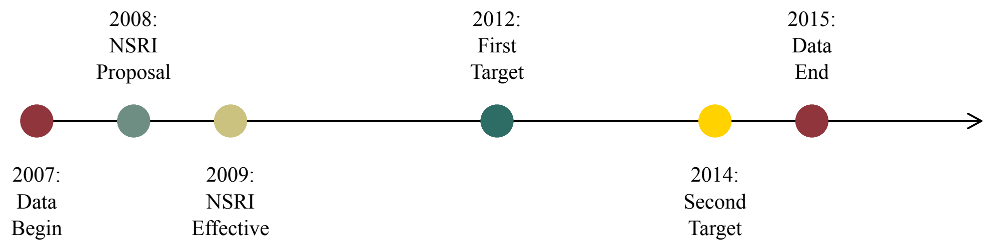
2.3.3 The NSRI: implementation

It is well known that the human preference for food components such as salt, fat, and sugar is adaptable: Small reductions in sodium content, for example, go largely unnoticed. As a result, recent literature has suggested that in order for food reformulation to effectively improve diets, it needs to be carried out gradually and silently (see Boon et al., 2010, Chapter 1, and Gressier et al., 2021). Similarly, highly publicized sodium reductions may signal reduced taste and result in demand reductions. These arguments have been advocated by both policy makers (who are interested in improving consumers' diets; e.g., Jacobs, 2021) as well as by industry (whose interest is to preserve consumer loyalty, e.g., Brat and Tamman, 2010). The NSRI was consistent with this approach.

First, the NSRI was implemented gradually: It set out to achieve its goal over a 5-year period with an intermediate (less aggressive) target 3 years out (2012) and a final target in 2014 (due to data availability, our focus is the 2014 target only).[15] Second, companies did not advertise their participation in the agreement nor was there large or noticeable media coverage of the agreement. This behavior is consistent with media reports documenting firms' aversion to publicizing improvements in sodium content (see, e.g., Brat and Tamman, 2010). Figure 2.1 shows a timeline of the NSRI implementation and our sample window. Our sample of nutrition and food purchases begins in 2007, a year before the initial NYCHD discussions for the NSRI proposal began in 2008. The NSRI became effective and signatory firms declared their voluntary participation in the program in 2009. The sample ends in 2015, a year after the 2014 final NSRI target, providing data for the year following the treatment.

[15] Although the spirit of the initiative contemplated a gradual reduction, the widespread consumer switching (away from low-sodium products and toward high-sodium products) that we document below (see Section 2.6.3) suggests that sodium reductions carried out by manufacturers might have been too aggressive (see Section 2.7 for further discussion).

Figure 2.1. Timeline of the National Salt Reduction Initiative and our sample



Notes. The figure shows the gradual process in the NSRI implementation. Our sample of nutrition and food purchases begins in 2007, a year before the initial work on the NSRI proposal started in 2008. The NSRI became effective in 2009. The first NSRI target in 2012 was designed to be a less aggressive, intermediate target. Our sample ends in 2015, a year after the final NSRI target in 2014.

Industry participation was completely voluntary and involved no concessions or special treatment from local or national regulators. The NSRI targeted two segments: packaged foods and restaurant foods. Our data allow us to study behavior in the packaged foods segment. In our KRD analysis below, we study the effects of the NSRI on all products in the market as well as for the subset of products manufactured by NSRI signatories (listed in Table B.2) and non-NSRI signatories.[16]

We asked NYCHD staff whether the NSRI targets were being exclusively used by signatories, and they indicated that other large companies had used the NSRI targets as their sodium content reference guide, especially when introducing new products. To investigate whether any detectable reduction in sodium content is observed in a wider pool of firms (beyond NSRI signatories), we also consider an expanded list of firms and include those that participated in two other voluntary initiatives. The first is the 2007 Healthy Weight Commitment Foundation (HWCF) pledge for firms to collectively sell 1 trillion fewer calories by 2012 and 1.5 trillion fewer calories by

[16] We rely on "root" UPC digits (typically the first 8 UPC digits) that are publicly registered to a specific manufacturer to identify all UPCs belonging to a particular firm.

2015. The second is the International Food and Beverage Alliance (IFBA), which in 2008 committed to improving the nutrition profiles of their products on various dimensions. Table B.2 lists the firms that signed these two agreements and the NSRI participants. Of 38 firms listed, 22 are NSRI signatories. Some firms participated in more than one initiative. For example, Campbell Soup Company participated in both NSRI and HWCF, while Kraft Foods, Mars Food US, and Unilever participated in all three initiatives.

2.3.4 Other details

Sodium is an inexpensive food additive for which there is no perfect substitute. However, manufacturers can decrease sodium intensity in an increasing number of ways. Simply reducing sodium content is a viable option for some products. Other possibilities include using finer salt particles or increasing the concentration of sodium toward a product's surface; both techniques can produce the impression of more saltiness while using less sodium (Brat and Tamman, 2010). Alternatively, some manufacturers can use a substitute; potassium chloride (available to and used by some firms prior to the NSRI) is considered to be the closest substitute.

Further, extensive research in the food industry over the last two decades has made finding and using sodium substitutes increasingly more viable. In other words, the NSRI did not take the industry by surprise; firms already had substantial ability and knowledge about how to transition away from sodium. It was thus unlikely that reformulation significantly altered firms' production costs (sunk, fixed, or marginal). Salt is tasty and inexpensive; thus, the main drawback of reformulation for firms is not increased costs of production but the the risk of losing demand as products become less palatable (e.g., potassium chloride, while also inexpensive, is not as tasty as sodium). Thus, to the extent that price changes accompanied reformulation, they were likely driven by changes in consumer demand rather than by changes in production costs. We investigate this channel in Section 2.6.5.2 by checking whether products with reduced sodium became pricier.

There is some evidence for complementarities between sodium and other food components (i.e., fat and sugar). For example, it is known that food manufacturers often simultaneously manipulate the levels of sodium, sugar, and fat to achieve a target taste (see Moss, 2013). In addition, for some products the level of sodium required to preserve food texture or safety may depend to some degree on the amount of sugar or fat. In Section 2.6.5.2, we investigate whether the effects of the NSRI on product-level sodium content were also accompanied by recalibrations of sugar and fat content.

Given the thousands of products involved, evaluating compliance with the commitment on a product-by-product basis was not practical. Instead, the NSRI established a weighted average measure. Manufacturers' compliance with sodium targets was evaluated based on sales-weighted sodium content (mg/100 g) across all UPCs sold by the company. We carry out two types of empirical analyses. The first decomposes changes in sodium intake and relies on consumption-weighted average measures of sodium content across products; we explore several weights for this measure but use NSRI's sales-weighted version in our baseline specification. The second set of exercises focuses on the NSRI's effect on manufacturers' product reformulation efforts. Because the object of analysis in these causal exercises is product-level reformulation, the outcome of interest is UPC-level sodium content.

2.4 Data

We employ two types of datasets. The first comes from Nielsen, which collects data obtained from scanning devices. Scanner datasets contain barcode-level information on purchases made across thousands of locations in the United States from 2006 to 2019. The second dataset comes from Gladson and contains barcode-level nutrition on all food products sold in the United States at two points in time: 2007 and 2015. Our analysis focuses on 2007 and 2015, the years for which we have detailed nutrition data. The next sections provide a brief overview of the datasets and the matching procedure we used. The online appendix provides extensive details on the datasets, their preparation for analysis, and the matching procedure.

2.4.1 Scanner datasets

Nielsen's Home Measurement Service (HMS) dataset tracks weekly retail purchases made by a panel of households; we refer to this dataset as "household" or HMS. Participant households scan their retail purchases at home upon returning from shopping trips made to retail outlets; then, a hand-operated scanning device transmits the information to Nielsen for processing. The recorded data include price and volume information at the UPC (also known as "barcode") level, as well as an anonymized code that identifies the visited retailer. In addition, Nielsen collects demographic information on each household member (e.g., age and gender) as well as education and employment for the household heads. Race, ethnicity, income, marital status, and presence of children are recorded for each household. Over 60,000 households are represented in each of the 2 years we study (2007 and 2015), 23,556 of which are present in both years. We aggregate weekly purchases up to household-year-UPC level. That is, an observation comprises the number of ounces purchased by the household for a particular UPC in a given year as well as the corresponding average per unit price. The resulting dataset contains 21,058,401 household-year-UPC observations.[17] We use survey weights whenever feasible to scale the results of our household sample to national representativeness.

[17]Recording errors are possible in the HMS data. However, Einav, Leibtag and Nevo (2010) show, these errors are not unusual in many datasets commonly used by researchers.

Another Nielsen scanner dataset (Retail Measurement Systems, or RMS) contains information on weekly store-level sales (see the online appendix for details). We do not use this dataset for our main analyses but employ it for robustness checks in Section 2.6.5. The reason for this choice is that our decomposition analyses requires that each unit of analysis remain stable over time, otherwise quantified effects may be confounded with changes in the characteristics of the observed units. Even though households also change over time (e.g., as they might get older, wealthier, more educated), these changes are likely to be less pronounced than changes that occur for a given store over time. For example, frequent changes in management and product availability would pose a threat to stable unit treatment. Furthermore, in Section 2.6.3.2 we conduct household-level demographic analyses, which allows us to assess the drivers of the sodium decline (product reformulation vis-à-vis consumer shopping behavior) among different populations at a finer level of granularity than would be possible from examining area-level demographics (e.g., census data).

2.4.2 Nutrition dataset

UPC-specific nutrition information comes from Gladson, the leading provider of syndicated consumer packaged goods (CPG) on product images and information.[18] Gladson collects a large number of product attributes, including nutrient and non-nutrient components for all CPG products at the barcode level. Unlike other databases, Gladson collects information on several hundred product attributes; for the

[18]To our knowledge, Gladson has the largest, most complete, and most accurate database of product attributes for consumer packaged goods (CPG). Leading retailers and CPG manufacturers (e.g., Albertson's, CVS, Dean Dairy, General Mills, Kellogg's, Kraft, Kroger, Meijer, Safeway, Supervalu, Unilever) periodically send their products to Gladson for product information recording. The two other sources for this type of data of which we are aware are USDA (obtained through IRI and Gladson) and the Mintel database; both have limitations for our research question of the policy effects on product reformulation and consumer switching. We learned from USDA personnel that their UPC-level nutrition information is only available from 2012 onward. The Mintel database only provides nutrition information for newly launched products, and it does not track reformulation of existing products.

purposes of this paper, we focus our attention on sodium content but also use calories, carbohydrates, fat, protein, and sugar for auxiliary analyses (see Section 2.6.5).[19] Of importance to our research is Gladson's practice of continuously and periodically (weekly) updating their database; this provides precise information on products that are subject to reformulation as well as on product turnover (entering and exiting products). Gladson provided the data based on the list of food UPCs identified in Nielsen scanner datasets. Not all food UPCs in the scanner datasets, however, are exactly matched in Gladson's database; to address this issue, we employed a matching procedure described in Section 2.4.3.

We focus our analysis on two years: 2007 and 2015. One reason for our focus is that producing a merged dataset at this level of granularity is extremely time consuming.[20] These two years were chosen to provide the most relevant snapshots of the packed food market a year before the start of the NSRI efforts in 2008 and right after the final NSRI target in 2014. Of the 589,377 UPCs in our sample, 27 percent exited the market prior to 2015, 38 percent did not exist in 2007 and only 35 percent were present in both years. We limit our study to food and non-alcoholic beverage categories.

2.4.3 Matching scanner and nutrition datasets

Gladson nutrition data have an exact match for 12 percent (of 262,087 UPCs) and 25 percent (of 327,290 UPCs) of Nielsen UPCs in 2007 and 2015, respectively. Although this may seem like a small coverage, the nutrition information can be reliably proxied for a significant portion of the unmatched UPCs, many of which are identical to a matched UPC in terms of nutrients. Since a UPC is a product-

[19] We also utilize Gladson information on calories and servings for some of our calculations. In separate work, we utilize several product attributes to analyze overall product healthfulness (Cengiz, Rojas and Wang, 2020).

[20] Another reason is the prohibitive cost of the data.

package combination, the content (brand and, therefore, nutrients) is identical across several UPCs except for their package size (i.e., both 8-ounce and 2-liter Coca-Cola have exactly the same nutrient profile on a per ounce basis). Using the nutrition information across all UPCs of identical characteristics raises the fraction of matched UPCs to 56 percent and 60 percent, respectively, for each of the sample years.

We use nutrition of the most similar UPC in the pool of matched UPCs to proxy nutrition for the remaining unmatched UPCs. Similarity between matched and unmatched UPCs is based on several criteria, including core product characteristics (year, brand, product category) and detailed product attributes (flavor, formula, form, product, type, variety, style). This procedure allows us to match a large number of UPCs that are cataloged as store brands (also known as private label brands), which often mimic the attributes of the leading branded products. A detailed documentation of the proxying procedure (including the number of UPCs matched at each step) is available in the online appendix.

We faced several data challenges in constructing the matched sample. Since sodium content is defined on a weight basis (i.e., milligrams of sodium per 100 grams of product), it is essential to have the total weight of the UPC in at least one database (Nielsen or Gladson). However, some UPCs are defined in "counts" (e.g., slices of bread) rather than by weight; we circumvent this problem by using the weight of the most similar UPCs as a proxy. Further, we detected several outliers (observations with unusually large nutritional values) in the Gladson database, likely caused by erroneous recording of the appropriate weight unit (e.g., grams instead of milligrams) or an incorrect coding of the nutrient value. We applied an interquartile range outlier detection method to deal with these cases. The online appendix documents an extensive list of details of how we overcame these and other issues in the matching procedure. Finally, to further ensure the quality of our matching procedure, we scraped nutrition information from a variety of websites to conduct random checks for a subset of several thousands of UPCs in our final matched data.[21]

As documented in the online appendix, we invested a considerable amount of time and resources ensuring that the matched dataset had the most precise, transparent, and reliable barcode-level nutritional content.[22] We thus favored a high-quality and reliable dataset for 2 years over a longer but less reliable panel. Prior work that has carried out a similar matching procedure (Dubois, Griffith and Nevo, 2014; Allcott et al., 2019; Hut and Oster, 2018; Hut, 2018) does not provide sufficient documentation to allow us to compare the quality of our matching procedure; however, from the information available, we conjecture that our sample is superior to earlier efforts as it tracks nutrient information precisely over time. Hut (2018) and Hut and Oster (2018), for example, use a single cross-section of nutrition information (2017) to impute nutrition information for the 2008--2015 period. This imputation is not innocuous since product reformulation and turnover are not tail events in food manufacturing.

2.5 Methodology

Our methodology is divided in two portions, both aimed at understanding the temporal change in sodium levels. First, we study changes in *sodium intake* and the role that product reformulation has played vis-à-vis consumer shopping behavior. Since sodium intake is a function of how much of each product consumers purchase, the methodology uses both household consumption data (HMS) as well as product

[21] Originally we observe 329,628 and 350,714 UPCs in 2007 and 2015, respectively. Although we apply proxying procedures to this larger set of UPCs, for the analysis we carry out in this paper, we kept the UPCs for which we found our matching procedure to be most reliable. See Section 3 of the online appendix for details.

[22] Further, we wrote the online appendix to allow future researchers to readily replicate the database and reproduce our results.

nutrition data (Gladson). Second, we assess product reformulation in isolation using nutrition data only. Here, our main goal is to investigate whether the NSRI had a causal impact on the *sodium content* of targeted products.

We focus our efforts on studying sodium before and after NSRI targets (focusing only on the final NSRI target in 2014). We use two sufficiently distant periods (2007 and 2015) to increase the odds of capturing changes in firm reformulation and consumer preferences, which likely evolve over longer horizons. Further, our discussions with industry experts suggest that even a small change in a single ingredient (e.g., a decrease in sodium) could take significant time to be incorporated in production. For example, long-term contracts with suppliers and distributors may significantly hinder firms' ability to reduce sodium content via product turnover (entry and exit). Conversely, there can be constraints within the firm (e.g., resistance from managers or divisions to adopt or introduce a reformulated product).

A limitation in our analysis is that we use food purchases to proxy actual food intake. This creates two measurement errors that work in opposite directions. While we overestimate sodium intake because food waste is not considered, we underestimate intake because our data do not capture food away from home (e.g., restaurants, fast food chains, school cafeterias). Despite this limitation, some evidence suggests that households' retail purchases provide a good approximation for individuals' overall diets; as we report in Section 2.6.2, our data produce similar per calorie sodium intake measures as those obtained in studies that use data on individuals' full diet. In addition, the COVID-19 pandemic has, perhaps permanently, decreased households' consumption of food away from home.[23] Finally, we provide robustness tests that use comprehensive sales data from retail stores (instead of purchases registered by the sampled households) and verify that conclusions are qualitatively similar (see

[23]See, e.g., https://www.ers.usda.gov/covid-19/food-and-consumers/.

Section 2.6.5.1); this robustness test is valuable as food sales across all retailers are arguably superior measures of food intake across outlets (i.e., many restaurants and other food-away-from-home outlets procure food at retail stores).

2.5.1 Sodium intake and its decomposition

We investigate the evolution of sodium intake and the role of reformulation (vis-à-vis consumer shopping behavior) in that evolution. We study sodium intake across four measures, described in Section 2.5.1.1, that are adopted in the empirical literature, including the sales-weighted sodium intake measure adopted by the NSRI. In Section 2.5.1.2, we detail the decomposition method used to isolate the role of product reformulation from that of changing consumer shopping habits.

2.5.1.1 Sodium intake measures

Strictly speaking, sodium intake is daily individual sodium consumption, typically measured in milligrams of sodium per day (mg/day). To compute daily individual intake, we sum up the sodium content in milligrams across the purchased UPCs in a year by a particular household, and divide the result by 365 times the adult-equivalent (AE) number of household members.[24] We call this measure per capita sodium intake (PC). Adult equivalence is computed using the tolerable upper sodium intake levels set by the 2015–2020 Dietary Guidelines for Americans. These upper limits vary by age: 1,500 mg/day for ages 1–3, 1,900 mg/day for ages 4–8, 2,200 mg/day for ages 9–13, and 2,300 mg/day for ages 14 and above. For example, a household that has one adult and one 2-year old would have 1.65 AE members.[25]

[24]This measure is not possible with retail data since we do not have the number of consumers who purchased the products at the sampled retail stores.

[25]We calculate AE by summing the tolerable upper sodium intake levels by the household members and dividing the sum by 2,300 mg, the tolerable upper sodium intake for the American adult. In this example, AE is equal to $\frac{1{,}500\ mg+2{,}300\ mg}{2{,}300\ mg} = 1.6522$.

Since the PC measure is based on aggregate sodium consumption across all products, it cannot be used to carry out the decomposition analysis, which requires the evaluation of individual products per person. For the decomposition analysis we therefore employ three alternative measures proposed in the literature. These measures are constructed as weighted averages of the sodium content across UPCs, where the weights are given by a measure of consumption.

First, if one is interested in product weight in grams as a measure of food consumption, the weighted average would use each UPC's contribution to the entire weight of the food basket (Griffith, O'Connell and Smith, 2017); we refer to this measure as "gram-weighted" sodium intake (GW). Second, if one is interested in food expenditures as a measure of consumption, the weighted average uses the sales share of a UPC to measure each UPC's contribution to the average (see, e.g., Curtis et al., 2016); we refer to this measure as "sales-weighted" sodium intake (SW), and this is the measure used by NSRI officials to evaluate firms' compliance with NSRI targets. Third, we report a calorie-based measure of the intake as set forth by the Institute of Medicine (Boon et al., 2010). Specifically, a UPC's sodium content is defined on the basis of milligrams per 1,000 kilocalories (kcal), and the weights are given by each UPC's contribution to the caloric content of the entire food basket. We call this measure "calorie-weighted" sodium intake (CW). Table 2.3 summarizes these alternative measures of sodium intake.

To compute the weighted average measures, we first standardize the UPC-specific sodium content and label it as d (see third column, Table 2.3). Next, we apply the weights given by the corresponding consumption measure (product weight share, expenditure share, or calorie share). Finally, the weighted average measure of sodium intake for each household in our data, $(D_{h,\,t})$, is computed as

$$D_{h,\,t} = \sum_{k \in K_{h,\,t}} q_{k,\,h,\,t}\; d_{k,\,t} \tag{2.1}$$

where h denotes a household, $t \in \{2007,\ 2015\}$, and $q_{k,h,t}$ is the share of UPC k in the basket of all UPCs purchased by the household h in year t (i.e., $K_{h,t}$).

Table 2.3. Summary of sodium intake measures

Sodium Intake Measure (D)	Description	UPC Sodium Content (d)	UPC Weight for Weighted Average (q)
Per Capita (PC)	mg of sodium per adult equivalent (AE) per day (pooled across all UPCs)	N/A	N/A
Gram-Weighted (GW)	Sodium intake weighted by weight (in grams)	mg of sodium per 100 g	% product weight (in grams) with respect to combined weight across all UPCs
Sales-Weighted (SW)	Sodium intake weighted by expenditures	mg of sodium per 100 g	% expenditures with respect to combined expenditure across all UPCs
Calorie-Weighted (CW)	Sodium intake weighted by calories (per 1,000 kcal)	mg of sodium per 1,000 kcal	% kcal with respect to combined kcal across all UPCs

Notes. Per capita sodium intake (PC) is our measure of daily individual sodium consumption. It is not suitable for decomposition analysis because of its aggregate nature (see text). GW intake is used in (Griffith, O'Connell and Smith, 2017), SW intake is the preferred measure by the NSRI, while the CW intake is a measure suggested by the Institute of Medicine (Boon et al., 2010).

Since the alternative weighted average intake measures are not direct measures of the intake (unlike the per capita measure) but rather consumption-weighted averages of sodium content across products, they are often referred as measures of "sodium intensity" (Griffith, O'Connell and Smith, 2017) or "sodium density" (Curtis et al., 2016). We note this difference, but for consistency and expositional simplicity we use the term "sodium intake" across alternative measures.[26]

[26]Further, note that the change in the GW measure is arithmetically identical to the change in the PC measure. Further, as it is later reported, all measures we consider yield qualitatively similar conclusions.

2.5.1.2 Decomposition of the change in sodium intake

To analyze the drivers of the improvement in sodium intake that we observe in the data (reported in Section 2.6.2), we apply a decomposition method pioneered by and used extensively in the productivity literature (Baily et al., 1992; Foster, Haltiwanger, and Krizan, 2001; Foster, Haltiwanger and Syverson, 2008; see Melitz and Polanec, 2015, for a comparison of methodologies). The decomposition method is applicable to the study of changes in weighted average measures such as those we consider here.[27] First, UPCs are grouped into three mutually exclusive sets: continuing (UPCs are present in both years, labeled C), entering (UPCs are present only in 2015, labeled N), and exiting (UPCs are present only in 2007, labeled X). The change in sodium intake measure (D) for each household can then be expressed as

$$\begin{aligned}\Delta D_h &= D_{h,\,t} - D_{h,\,t-1} \\ &= \underbrace{\sum_{k\in C} q_{k,\,t-1,\,h}\,\Delta d_k}_{\textit{Reformulation}} + \underbrace{\sum_{k\in N} q_{k,\,t,\,h}\,(d_{k,\,t} - D_{h,\,t-1})}_{\textit{Entry}} - \underbrace{\sum_{k\in X} q_{k,\,t-1,\,h}\,(d_{k,\,t-1} - D_{h,\,t-1})}_{\textit{Exit}} \\ &\quad + \underbrace{\sum_{k\in C}(d_{k,\,t-1} - D_{h,\,t-1})\Delta q_{k,\,h} + \sum_{k\in C}\Delta d_k\,\Delta q_{k,\,h}}_{\textit{Switching}}\,.\end{aligned} \tag{2.2}$$

The Reformulation term isolates the effect of the reformulation of products that remain in the market: It weighs the change in the sodium content of UPC k (Δd_k), keeping weights fixed at 2007 (subscript $t-1$) levels. If the sodium content across UPCs, on average, decreases over time, this term would be negative. The Entry term captures the contribution of entering products: It considers the weighted difference between the sodium content of each entering UPC in 2015 and the overall sodium

[27] In the productivity literature, the decomposition method is used to study how different types of firms (exiting, entering, continuing) contribute to average productivity changes over time. See, e.g., (Foster, Haltiwanger and Krizan, 2001; Foster, Haltiwanger and Syverson, 2008).

intake of the household in 2007. If an entering product has a sodium content that is lower than the overall 2007 sodium intake, then its contribution to the change in weighted average is negative, (i.e., a healthier addition to the household's grocery basket). Similar mechanics apply to the Exit term, but the set of products considered here is exiting UPCs. If exiting products are generally worse (i.e., higher in sodium content) than overall sodium intake in 2007, this term would (because of the preceding negative sign) contribute to a decrease in sodium intake by household h (ΔD_h).

The Switching term combines two terms that capture the effect of consumer switching across continuing products; we label these as $S_A = \sum_{k \in C}(d_{k,\,t-1} - D_{h,\,t-1})\Delta q_{k,\,h}$ and $S_B = \sum_{k \in C} \Delta d_k \, \Delta q_{k,\,h}$. Since S_A holds sodium content fixed at 2007 levels, this term can be thought of as the component of switching that is driven by consumers' evolving shopping behavior alone. If consumers' shopping behavior was such that it gravitated toward (away from) products that are saltier (less salty) than the mean intake in 2007 (i.e., $d_{k,\,t-1} > D_{h,\,t-1}$ and $\Delta q_{k,\,h} > 0$ or $d_{k,\,t-1} < D_{h,\,t-1}$ and $\Delta q_{k,\,h} < 0$), then S_A would be positive. On the other hand, S_B can be thought of as consumers' *reaction* to reformulation. If S_B is positive, then consumers are, on average, switching away from (toward) products that reduced (increased) sodium content over time (i.e., $\Delta d_k < 0$ and $\Delta q_{k,\,h} < 0$ or $\Delta d_k > 0$ and $\Delta q_{k,\,h} > 0$).

Firms that opt to reformulate products (in this case, reduce sodium content) may pursue this strategy by either keeping their products in the market and changing their nutritional profile or by replacing some of their existing products with newer, less salty, ones; hence, product turnover (entry and exit) can be cast as a reformulation mechanism. Thus, our main conclusion on product reformulation refers to the *combined effect* of the first three terms in equation (2.2). Conversely, we interpret switching (the last two terms in equation 2.2) to capture changes in consumer shopping behavior.[28]

Unlike the productivity literature, which computes a scalar for each of the decomposition terms at a point in time, we compute decomposition components for each household in the data and hence generate a distribution for each component.[29] This feature allows us to study how socioeconomic and demographic factors are related to decomposition components (see Section 2.6.3.2). Finally, the decomposition analysis is feasible only for the subset of households that are observed in both 2007 and 2015. One advantage of using a balanced sample of households is that results are more reliable because units are comparable over time, in line with the stable unit treatment variable assumption (SUTVA) needed in the potential outcomes framework. While our decomposition analysis is not strictly a causal exercise, it can be cast as a counterfactual exercise as it simulates consumption in 2015 using 2007 shopping behavior (shares). In this sense, ensuring SUTVA increases the reliability of the decomposition analysis.

2.5.2 Causal identification of the NSRI effects

We pursue causal identification of the NSRI's effects on product-level sodium content. We focus on sodium content rather than sodium intake for four reasons: First, the NSRI was a concerted effort between policy makers and industry that focused on product reformulation; participation by consumer advocacy groups was all but absent. Second, the NSRI pledge was not a highly publicized event. In fact, as

---

[28]We could also argue that the entry and exit terms capture some of the consumer switching since they capture consumer shopping behavior away from some products (exiting) and into others (entering). However, entry and exit are natural "supply-side" decisions. Further, as we show in Section 3.5, to the extent that switching occurs through entry and exit, the effect is very limited; the main results are largely driven by the reformulation and switching components in equation (2.2).

[29]In the productivity literature, q represents the volume share of each plant's production and d (D) represents a plant's productivity measure (across-plant weighted average).

described in Section 2.3.3, firms appear to have purposefully adopted a silent strategy as it pertained to their reformulation efforts. As a result of these two factors, consumer awareness of firms' reformulation efforts stemming directly from the NSRI is likely to have been limited. Third, our decomposition results indicate that, to the extent that the NSRI caused its desired effect (i.e., a reduction in sodium intake), it was driven by sodium content reductions (and not by consumers switching to healthier options). Last, as we argue below, the institutional details of how the NSRI was implemented provide an opportunity for reliable causal identification of the NSRI effects on product reformulation. Given this backdrop, the most meaningful and reliable place to look for a causal effect of the NSRI is on suppliers' behavior to reduce sodium content, as this should be the primary effect of the agreement.[30]

Our approach relies on a kink regression discontinuity (KRD) design, as in Card et al. (2015). We argue that the NSRI targets create a discontinuity in firms' incentives to reduce sodium content. The NSRI did not institute a target that was equally stringent across all products in a category: Manufacturers of products with sodium content in excess of the target had more incentive to reformulate. Thus, the discontinuity produced by the NSRI is gradual (as opposed to a discrete jump). This gradual discontinuity allows us to more precisely discern sodium reductions resulting from the NSRI from those that would have occurred otherwise (e.g., a 10 percent sodium reduction across all products). Further, as we argue below, this discontinuity is likely to be exogenous as the observed kink is not present in other product attributes.

[30]We are not suggesting that the NSRI had no effect on demand. Indeed, our decomposition results suggest that consumers perceived and disliked sodium reductions, some of which may have been driven indirectly by the NSRI. Similarly, our descriptive results indicate that firms reduced sodium across the board. What we argue here is that the institutional details of how the NSRI was implemented provide an opportunity to reliably *isolate* the causal effect of the NSRI on the supply side (reformulation).

Specifically, we look at whether the *change* in sodium content differs across sets of products whose pre-NSRI sodium content were on the opposing sides of the NSRI target. Figure 2.2 depicts perfect compliance with the NSRI. The running variable is the distance between a product's pre-NSRI sodium content and the NSRI target (i.e., excess sodium relative to the NSRI target). The y-axis represents a *change* in UPC sodium content pre- and post-NSRI. Both axes are measured in milligrams of sodium per 100 grams. The theoretical prediction under perfect compliance serves as the basis for our empirical approach: If the NSRI had its expected effect, one would observe no (or insignificant) change in sodium content for compliant products (i.e., those with 2007 sodium content below the NSRI target) but an increasing reduction the farther away a noncompliant product is from the target (i.e., a 45° line under perfect compliance).

Figure 2.2. Theoretical prediction of perfect compliance under the NSRI

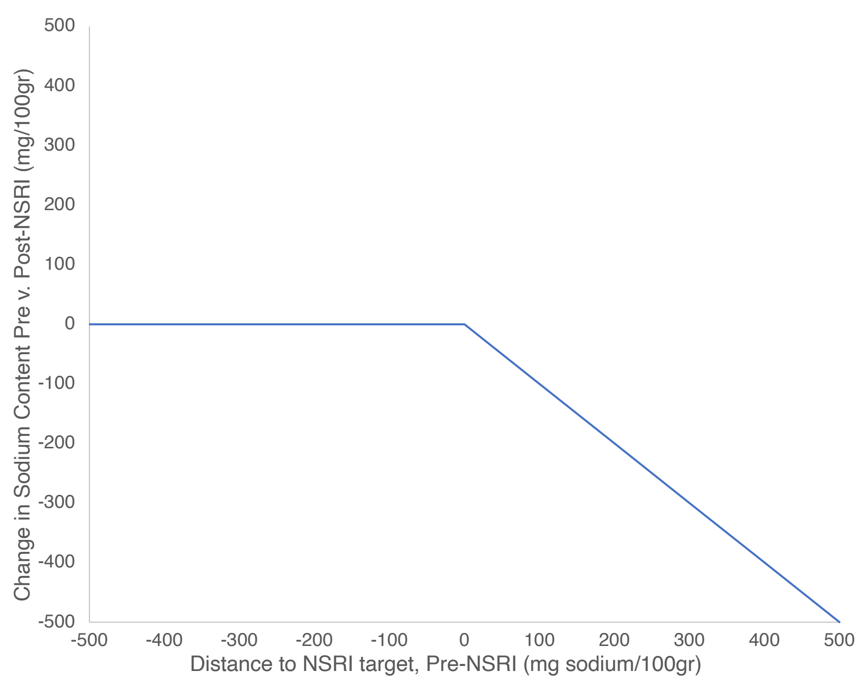
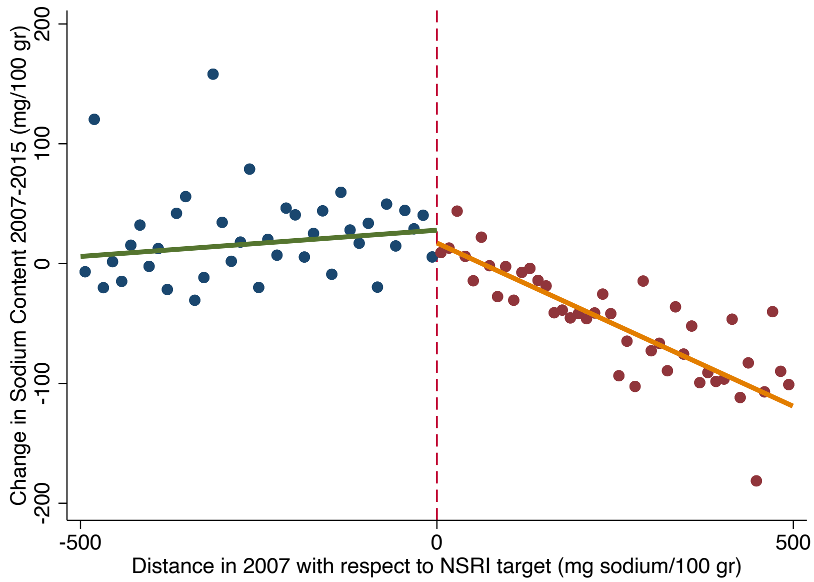


Figure 2.3. Kink regression discontinuity plot, products in NSRI categories that appear in both 2007 and 2015



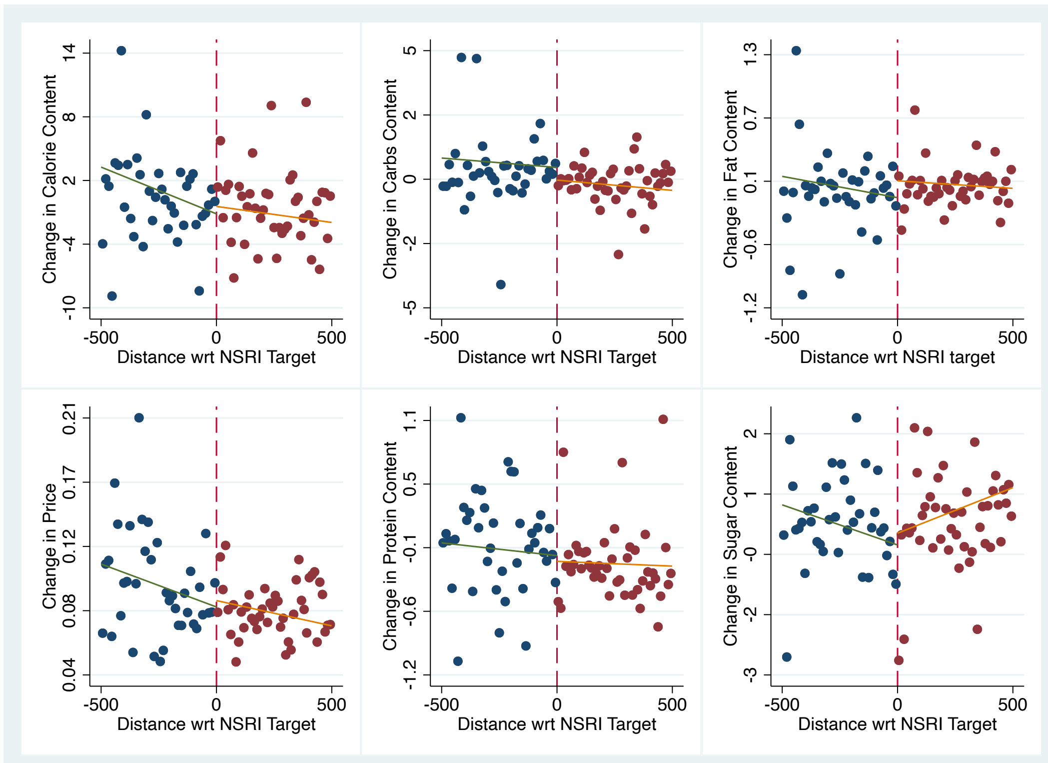
Notes. The RD plot is generated by applying 45 bins on each side of the threshold and fitting a linear trend on each side. Each observation corresponds to the average across UPCs within a bin.

Figure 2.3 uses our data to depict the empirical counterpart of Figure 2.2.[31] As predicted, a) the change in sodium content for products on the left of the threshold appears to have revolved around 0 and with a relatively flat slope and b) the change in sodium content for products on the right of the threshold are negative, much more sizable, and become larger as the distance from the threshold increases. Visually, the NSRI appears to have had a sizable effect, albeit below that predicted by perfect compliance. We interpret this evidence as strongly suggestive of the NSRI (as opposed

[31] We use 2007 as a proxy of sodium content in 2008. Sodium content in 2007 was not (yet) affected by the policy and is the best approximation for 2008 sodium levels.

to other sodium reduction efforts that might have occurred in tandem)[32] having caused a decline in product-level sodium content.

Figure 2.4. Kink regression discontinuity plots, other product attributes



Notes. The RD plot is generated by applying 45 bins on each side of the threshold and fitting a linear trend on each side. Each observation corresponds to the average across UPCs within a bin. The x axis in all plots is expressed in mg/100 g; the y axis in all plots except calories (which is expressed in kcal/100 g) and price (which is expressed in \$/100 g).

As we argued in Section 2.3.2, target setting in the NSRI was likely exogenous, which increases the reliability of our causal interpretation of the observed discontinuity in slope. To add credibility to this argument, we run visual placebo tests and inspect whether the identified kink is present in other product attributes. As seen in Figure 2.4, the kink in slope is not evident (or as salient) as in Figure

[32]The timing of the HWCF and IFBA initiatives (see Section 2.3.3) roughly coincides with that of the NSRI. However, neither initiative targeted sodium directly. The HWCFA focused exclusively on reducing calories. The IFBA sought to reduce sodium content, among other things, but it left it up to companies to decide how much to reformulate

2.3. There appears to be a kink in slope for sugar. However, we rule this out via formal placebo tests for all product attributes (see Section 2.6.5.2; last column in Table 2.10). Section 2.6.5.2 discusses other robustness checks and formal tests of the assumptions of the KRD design.

Estimation details

We estimate a standard kink regression discontinuity model. In our setting, the KRD design tests for a discontinuity in the *slope* of the change in sodium content for products located at either side of (and close to) the kink threshold (in our case the NSRI target). The KRD estimand is given by as in Card et al. (2015):

$$\hat{\tau}_{KRD} = \frac{\hat{\mu}^+ - \hat{\mu}^-}{\kappa^+ - \kappa^-}, \tag{2.3}$$

where $\hat{\mu}^+$ and $\hat{\mu}^-$ are estimators of the slope on either side of the threshold. The values κ^+ and κ^- are deterministic values that depend on the policy being evaluated. In our case, $\kappa^- = 0$ and $\kappa^+ = -1$ since the predicted slopes at the left and right of the threshold, respectively, are predicted to be equal to 0 and -1 (i.e., perfect compliance, see Figure 2.2). To estimate $\hat{\mu}^+ - \hat{\mu}^-$, we use standard procedures and specify a local polynomial model (of order $\overline{p} > 0$):

$$E[z_i \mid x_i] = \sum_{p=0}^{\bar{p}} \left[\gamma_p x_i^p + \nu_p x_i^p \times D\right], \tag{2.4}$$

where z_i denotes the change in sodium content for UPC i on either side of the threshold, x_i measures the distance of UPC i's sodium content in 2007 with respect to the 2014 NSRI target (i.e., the kink threshold in our setting is equal to 0), and $D = \mathbb{I}[x_i \geq 0]$ is an indicator for being above the kink threshold. Both z_i and x_i are measured in milligrams of sodium per 100 g. Note that this specification allows direct estimation of the numerator in equation (2.3) via $\hat{\nu}_1 = \hat{\mu}^+ - \hat{\mu}^-$.

In terms of implementation, we use observations within a bandwidth b from the threshold (i.e., $|x_i| < b$) and determine the optimal threshold (b) using a minimum-square-error (MSE) optimization procedure (Figure B.5 also reports robustness results with varying bandwidth sizes). Our preferred specification is linear (see Section 2.6.5.2 for sensitivity to $\overline{p} = 2$). We estimate equation (2.4) via weighted least squares using a triangular kernel function to weigh observations.[33] Standard errors are clustered at the most conservative level; specifically, following Dubois, Griffith and Nevo (2014), we divide UPCs into nine broad food departments and allow for correlation across all UPCs within a food type.[34]

Since the outcome of interest is the change in sodium content, we study the subset of products that appear in both 2007 and 2015. Focusing on continuing products increases the credibility of our identification strategy since product-specific unobservables (e.g., brand reputation or consumer loyalty) can be accounted for. Further, this design increases the similarity of products on either side of the threshold: Continuing products are arguably more comparable to one other (e.g., they may be well-known products with stable demand) than to entering or exiting products. Beyond reporting whether a change in slope exists across all continuing products, our results below separately investigate which types of products are responsible for the effect (i.e., products manufactured by NSRI signatories vis-à-vis products belonging to NSRI categories but manufactured by nonsignatories).[35]

---

[33]Our results and conclusions are not sensitive to the type of kernel used.

[34]The nine broad food departments are dairy, drinks, fruits, grains, meats, oils, prepared, sweets, and vegetables.

[35]We also look at whether the result is driven by a wider pool of firms: NSRI signatories *plus* other large firms that committed to other voluntary reformulation initiatives around this period; see Section 2.3.3 for details on these firms and Section 2.6.4 for additional discussion.

2.6 Results

We divide our results into five parts. Section 2.6.1, provides a descriptive analysis of sodium content and product turnover at the UPC level. Section 2.6.2 presents sodium intake measures. Section 2.6.3 reports decomposition results, Section 2.6.4 reports NSRI effects results, and Section 2.6.5 discusses robustness checks and additional results.

2.6.1 UPC-level sodium content and product turnover

Table 2.4 reports the median sodium content across all UPCs in each of the two sample years.[36] The table shows the sodium content using two alternative measures: by weight (milligrams of sodium per 100 g of product) and by calories (milligrams of sodium per 1,000 kcal). In Panel A, we show the median for all UPCs as well as for two relevant subsets of UPCs: NSRI and non-NSRI products. We define a product as a UPC. We use NSRI's detailed descriptions in each of the 15 broad food groups and 57 finer food categories to find all UPCs in the scanner datasets that match this description; we call these matched products *NSRI* UPCs. (The remaining UPCs are labeled *non-NSRI*.)[37] Panel B breaks down the median sodium content across UPCs in each of the 15 NSRI food groups.

As expected, non-NSRI UPCs register far lower sodium content than NSRI UPCs. While the median sodium content across the entire pool of UPCs (NSRI and non-NSRI combined) shows mixed results (a 4 percent increase in the weight measure and a 10 percent decrease in the calorie measure), we observe a significant decrease in NSRI UPCs for both measures (-10 percent and -9 percent, respectively). Furthermore,

[36] The means produce qualitatively similar results as the medians. We report the median because it is not sensitive to outliers In our data, there are certain UPCs (e.g., table salt) with large amounts of sodium.

[37] Specifically, we create a correspondence between Nielsen's finest level product categories (PMCs) and NSRI food categories. See online appendix Section 3.1.4 for details.

the measure of sodium content by weight registers a decrease in all but *dairy* and *potatoes*, which exhibit virtually no change.

Something similar occurs in the calorie measure: two NSRI categories register an increase (10 percent in *potatoes* and 232 percent in *seasonings*). The large increase in *seasonings* requires some explanation. Seasonings are high-sodium products with very few calories (e.g., chili seasonings), making the calorie measure highly sensitive to changes in sodium content. Furthermore, this sensitivity is particularly salient in the seasoning category because a large number of UPCs (many of them of reduced caloric content) were introduced in the market after 2007 (there are 53 percent more UPCs in this category in 2015 than in 2007, as shown in the last column of Table 2.4).

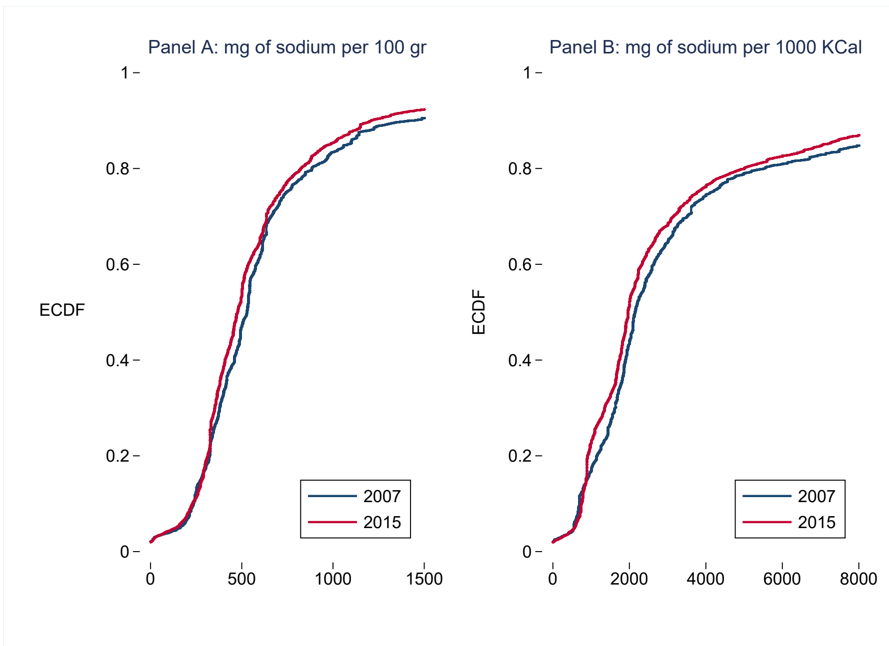
Table 2.4. UPC-by-year median sodium content [# UPCs]

	Median Sodium Content (mg of sodium per)				% Change 2007–2015		
	2007		2015				
	100 grams	1,000 kcal	100 grams	1,000 kcal	100 grams	1,000 kcal	#UPCs
Panel A: Overall							
All	287.23	1,147	299.83	1,028	4%	-10%	25%
	[262,087]		[327,290]				
Non-NSRI	48.99	335	50	346	2%	3%	22%
	[138,391]		[169,093]				
NSRI:	528.03	2,147	476.2	1,960	-10%	-9%	28%
	[123,696]		[158,197]				
Panel B: By-NSRI food group							
1. Bakery	455.03	1,538	377.94	967	-17%	-37%	39%
	[34,258]		[47,476]				
2. Cereal	536.12	1,438	492.02	1,333	-8%	-8%	4%
	[5,357]		[5,576]				
3. Meats	970.03	3,615	778.54	3,556	-20%	-2%	27%
	[7,986]		[10,127]				
4. Dairy	634.93	1,936	634.93	1,802	0%	-7%	24%
	[10,428]		[12,938]				
5. Fats and Oils	547	2,098	493.84	1,972	-10%	-6%	11%
	[8,035]		[8,922]				
6. Sauces & Dips	645	11,827	648.97	10,720	1%	-9%	26%
	[10,196]		[12,847]				
7. Snacks	769.08	1,699	761.92	1,590	-1%	-6%	29%
	[10,722]		[13,788]				
8. Soups	447.61	8,970	379.2	7,982	-15%	-11%	16%
	[4,450]		[5,175]				
9. Potatoes	331.38	1,846	331.22	2,044	0%	10%	0%
	[2,420]		[2,431]				
10. Mixed Dishes	468.77	2,289	451.93	2,211	-4%	-3%	41%
	[13,715]		[19,279]				
11. Vegetables	221.14	6,905	212.15	6,612	-4%	-4%	8%
	[6,904]		[7,444]				
12. Legumes	275.73	4,222	270.34	3,726	-2%	-12%	11%
	[2,783]		[3,084]				
13. Canned Fish	367.44	3,447	291.36	3,426	-21%	-1%	3%
	[613]		[634]				
14. Seasonings	9813.13	38,262	7170.55	127,137	-27%	232%	53%
	[4,814]		[7,375]				
15. Nut Butters	411.62	663	392.77	654	-5%	-2%	8%
	[1,015]		[1,101]				

Notes. NSRI = National Salt Reduction Initiative. NSRI UPCs are those in food categories targeted by the NSRI. Appendix Table B.1 provides details on the NSRI categories and corresponding subcategories. The online appendix (Section 3.1.4) contains a detailed description of the NSRI categories and corresponding product module codes (PMC) in Nielsen.

To provide initial evidence of how the NSRI policy might have influenced the sodium content at the product level over the 8-year period, Figure 2.5 plots the empirical cumulative distribution function (ECDF) of sodium content registered by NSRI UPCs. We find strongly suggestive evidence for a sodium reduction: The 2015 ECDF stochastically dominates the 2007 ECDF, suggesting a reduction in sodium content at almost every quantile of the distribution (not just at the median, as reported in Table 2.4).

Figure 2.5. Empirical cumulative distribution of sodium content, NSRI UPCs



Notes. ECDF = Empirical Cumulative Distribution Function.

There is a caveat with UPC-level descriptive statistics. Sodium intake (which we discuss below) is a function of UPC-level content *and* the amount of each UPC purchased. Since consumption (volume purchases) for a given household varies significantly across UPCs, we expect sodium intake measures to differ significantly from the patterns reported here. For instance, if households' consumption is disproportionately concentrated among UPCs that decreased sodium content over time, one would observe a decrease in sodium intake that is more pronounced than

Table 2.5. UPC count by food group and presence

	2007: Exiting, X		2015: Entering, N		2007 & 2015: Continuing, C		Exit Rate, $\frac{X}{C+X}$	Entry Rate, $\frac{N}{C+N}$
Panel A: Overall								
ALL	158,908	[33%]	224,111	[46%]	103,179	[21%]	61%	68%
NON-NSRI	83,340	[33%]	114,042	[45%]	55,051	[22%]	60%	67%
NSRI	75,568	[32%]	110,069	[47%]	48,128	[21%]	61%	70%
Panel B: By NSRI group								
1. Bakery	22,283	[32%]	35,501	[51%]	11,975	[17%]	65%	75%
2. Cereal	3,767	[40%]	3,986	[43%]	1,590	[17%]	70%	71%
3. Meats	5,078	[33%]	7,219	[47%]	2,908	[19%]	64%	71%
4. Dairy	4,995	[28%]	7,505	[42%]	5,433	[30%]	48%	58%
5. Fats and Oils	4,625	[34%]	5,512	[41%]	3,410	[25%]	58%	62%
6. Sauces	5,810	[31%]	8,461	[45%]	4,386	[24%]	57%	66%
7. Snacks	7,558	[35%]	10,624	[50%]	3,164	[15%]	70%	77%
8. Soups	2,422	[32%]	3,147	[41%]	2,028	[27%]	54%	61%
9. Potatoes	1,530	[39%]	1,541	[39%]	890	[22%]	63%	63%
10. Mixed Dishes	9,839	[34%]	15,403	[53%]	3,876	[13%]	72%	80%
11. Vegetables	3,031	[29%]	3,571	[34%]	3,873	[37%]	44%	48%
12. Legumes	1,114	[27%]	1,415	[34%]	1,669	[40%]	40%	46%
13. Canned Fish	441	[41%]	462	[43%]	172	[16%]	72%	73%
14. Seasonings	2,604	[26%]	5,165	[52%]	2,210	[22%]	54%	70%
15. Nut Butters	471	[30%]	557	[35%]	544	[35%]	46%	51%

Notes. This table reflects the number of *unique* UPCs in the data (i.e., it counts continuing UPCs, denoted by set C, only once).

what the UPC-level descriptive analysis suggests. Put differently, changes in the median UPC-level sodium content (Table 2.4) would be identical to changes in median sodium intake only if consumers purchased all UPCs with equal intensity.

We finalize the description of UPC-level data by looking at product turnover. Since reformulation (as well as total sodium intake) is partially driven by product entry and exit, it is important to describe the extent to which product turnover occurs in the data. Further, product turnover is integral to the decomposition analysis (see Section 2.5.1.2). Table 2.5 shows UPC counts by product presence: 2007 only (X), 2015 only (N), and both years products (C). Table 2.5 also shows the corresponding entry and exit rates. All NSRI food groups have roughly similar entry and exit rates with *dairy*, *vegetables*, *legumes*, and *nut butters* registering somewhat smaller entry and exit rates than the rest of the groups. Product turnover, as measured by the entry and exit rates, is very large for all groups. Almost two-thirds of 2007 UPCs exited by 2015; similarly, two-thirds of 2015 UPCs did not exist in 2007. Entry and exit rates vary somewhat across the NSRI groups but the values are not too different than those observed in non-NSRI products, suggesting that the NSRI did not alter firms' rate of product turnover.

2.6.2 Sodium intake measures

This section provides a descriptive analysis of the temporal change in sodium intake across the alternative measures described in Section 2.5.1.1. Table 2.6 reports the median sodium intake values in each of the four measures using two distinct household samples. The first sample includes all households observed in each year regardless of whether the household is present in either or both years, while the balanced panel contains the subset of households that appear in both years.[38] The improvement (reduction) in sodium intake is consistently observed and ranges from 4.02 percent to 8.00 percent across the four measures and two samples.[39]

[38]Decomposition analysis is only possible when the unit of analysis (household or county, depending on the dataset) is present in both years (i.e., a balanced panel). For comparability, we present results of balanced and unbalanced panels whenever possible.

[39]Similarity of results across balanced and unbalanced panel suggests that attrition bias would be minimal (or unimportant).

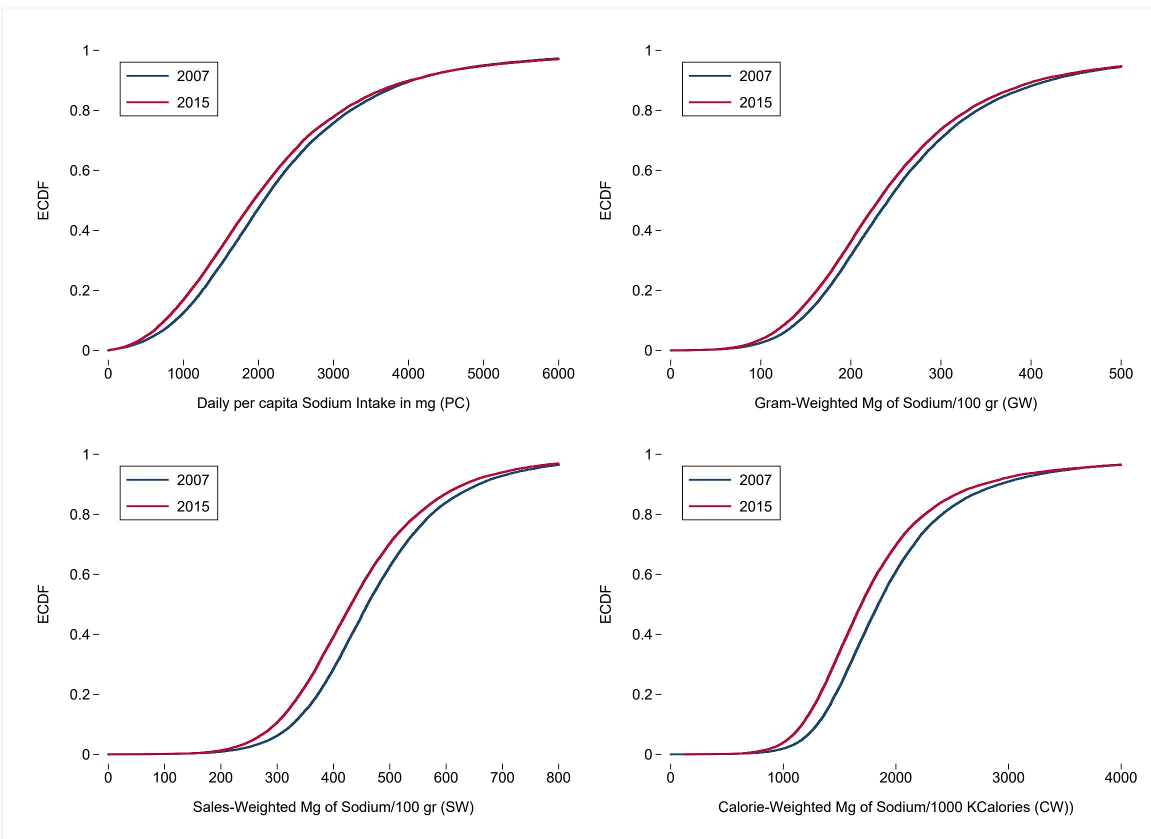
Table 2.6. Median sodium intake by measure type and year

	All Households			Balanced Panel of Households		
	2007	2015	% Change	2007	2015	% Change
Per Capita (PC)	1,941.45	1,797.75	-7.40	2,016.99	1,875.22	-7.03
Gram-Weighted (GW)	242.3	232.56	-4.02	240.37	229.42	-4.56
Sales-Weighted (SW)	465.33	442.17	-4.98	463.26	434.88	-6.13
Calorie-Weighted (CW)	1,846.56	1,719.63	-6.87	1,854.74	1,706.27	-8.00
# Households	63,325	61,351		23,556	23,556	

Notes. Balanced panel of households represents the households that are sampled in both 2007 and 2015. To ensure national representativeness, we use population weights to compute medians in the unbalanced panel of households (i.e., all household sample); weights are not feasible for the balanced panel. 2007 and 2015 Nielsen Home Measurement System (HMS) and Gladson.

Figure 2.6 provides visual evidence for the way in which sodium intake evolved for the entire distribution of the balanced panel of households. The figure reports the empirical cumulative distribution function (ECDF) of households in each year and for each of the four measures. All measures show a discernible improvement, as captured by the horizontal distance between 2007 and 2015 ECDFs; the improvement is relatively more pronounced for the sales- and calorie-weighted measures. These results suggest that the reduction of sodium intake was widespread among households: Nearly every quantile in the distribution experienced an improvement, a result that is consistent with evidence on the median sodium intake levels reported in Table 2.6.

Figure 2.6. Empirical cumulative distribution of sodium intake measures (balanced panel)



Notes. Figures created using 23,556 households that appear in both 2007 and 2015 (balanced panel of households). Figures with unbalanced panel (the sample of all households) show almost identical patterns (see Appendix Figure B.1. 2007 and 2015 Nielsen Home Measurement System (HMS) and Gladson.

While our data do not capture food consumption away from home, our analysis appears to generate reasonable approximations of individuals' sodium intake in their overall diet. The calorie-weighted (CW) measure of sodium intake that we report in Tables 2.6 and B.4 aligns well with estimates of individuals' entire diet. For example, the Institute of Medicine reports a CW intake measure between 1,500 and 1,600 mg of sodium per 1,000 kcal (see Boon et al., 2010, Figures 2–14), while our estimates range from (roughly) 1,700 to 1,850 mg of sodium per 1,000 kcal. Our estimates of sodium intake decline are also consistent with prior findings (Curtis et al., 2016, report a 6.8 percent decline in the SW measure).

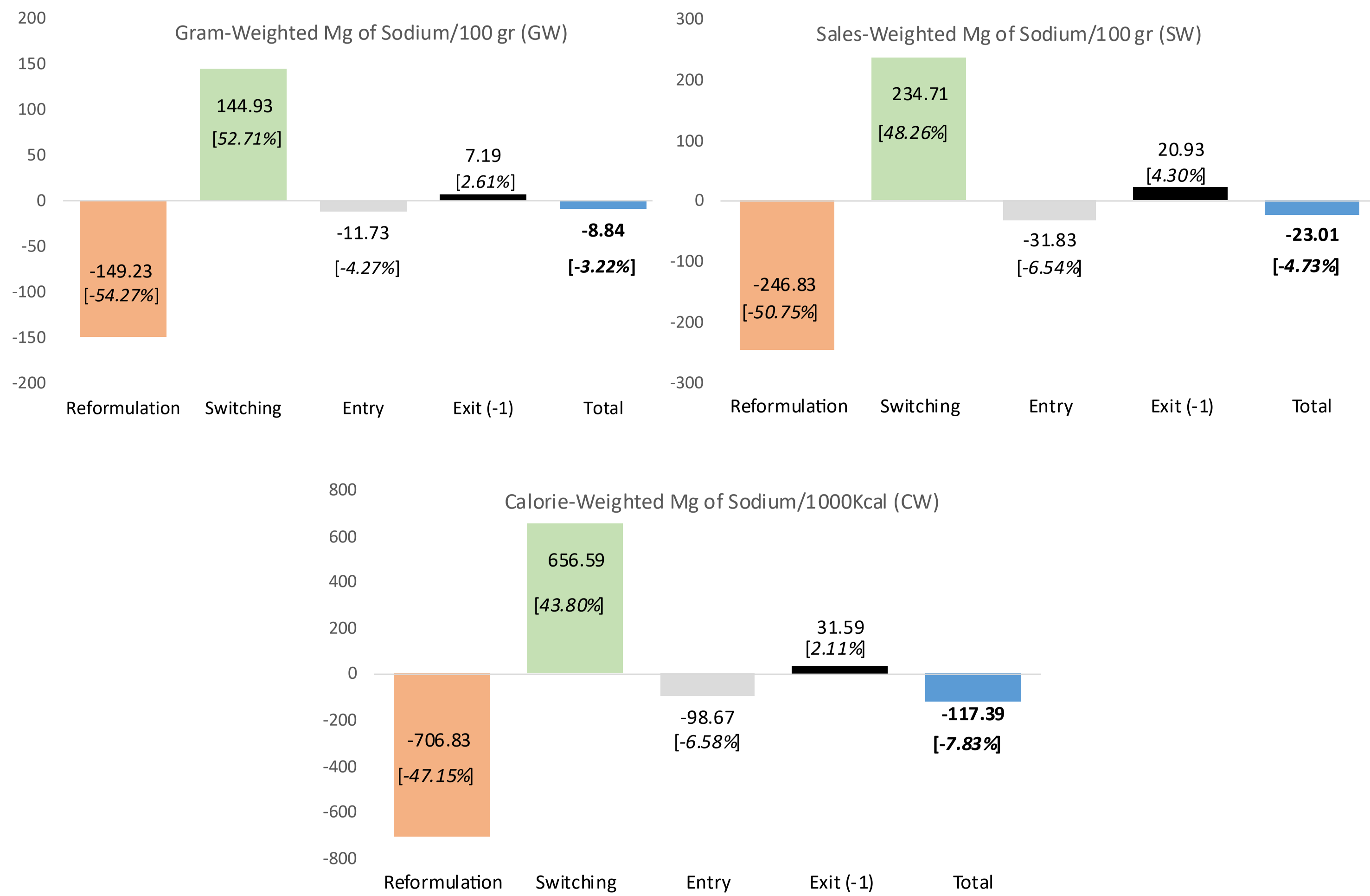
2.6.3 Decomposition of the temporal change in sodium intake

We carry out the decomposition exercise (equation 2.2) for each household in the balanced panel sample. Figure 2.7 shows a decomposition summary that reports the mean (across all households) of each of the four decomposition elements (reformulation, switching, entry, and exit) as well as the mean of the total temporal change. The figures show results for the three intake measures for which decomposition analysis is feasible.

Results are robust across measures. First, reformulation and switching have the largest effects, while entry and exit (product turnover) play a more limited role in their contribution toward the temporal change in sodium intake. Second, reformulation and switching display opposing forces as do entry and exit (Figure 2.7 accounts for the negative effect of exit; see equation 2.2). Third, the magnitudes of entry and exit are much smaller than those seen in reformulation and switching; in all cases, however, the net effect of entry is to improve (decrease) sodium intake. The fourth, and most salient, result is that reformulation alone produced a reduction of approximately 50 percent in sodium intake, a decline well above the 20 percent NSRI target and the 30 percent target established by the WHO for 2030. However, consumer switching toward saltier products has offset nearly all the reduction in sodium intake due to product reformulation efforts.

The net result is that total reduction in sodium intake (which ranges from 3.22 percent to 7.83 percent across the three measures) is small with respect to what could have been achieved given the observed reformulation efforts. As we argued earlier, product turnover can be interpreted as a reformulation mechanism. Using this broader interpretation of reformulation and focusing on the sales-weighted (SW) intake measure (the evaluation measure adopted by the NSRI), we observe that the 4.73 percent decline in sodium intake over 2007–2015 results from a 52.99 percent reduction generated by reformulation efforts (50.75 percent from pure reformulation and 2.24 percent from product turnover) and a 48.26 percent increase due to consumer switching.

Figure 2.7. Decomposition of the temporal change in sodium intake measures



Notes. The figure illustrates decomposition analysis across three sodium intake measures: gram-weighted (GW), sales-weighted (SW), and calorie-weighted (CW). Every bar depicts the mean of the decomposition component (reformulation, switching, entry, and exit) across all units (households). The decomposition analysis uses a balanced panel of 23,556 households. Values in the "Total" bars also refer to means and thus do not coincide with the percentage change in (median) values reported in the right panel of Table 2.6. Percentage changes in decomposition components are provided in brackets. The mean sodium intake in 2007 is 274.96 mg/100 g for GW, 486.36 mg/100 g for SW, and 1,499.20 mg/1,000 kcal for CW. 2007 and 2015 Nielsen Home Measurement System (HMS) and Gladson.

Focusing on the SW measure results (top right panel in Figure 2.7), we find that the components of switching (see Section 2.5.1.2) are $S_A = -171.79$ and $S_B = 406.51$. These magnitudes suggest two things. First, absent reformulation (e.g., no reductions in sodium, or $\Delta d_{k,\,h} = 0 = S_B$), consumers would have gravitated toward healthier alternatives (since S_A is negative). Second, consumers significantly disliked reduced-sodium products (i.e., S_B is positive). One likely reason for why reformulation efforts backfired in this particular way is that, despite the attempt at a gradual approach, reformulation might have occurred too rapidly to go unnoticed by consumers (see Section 2.3.3 and our discussion in Section 2.7). It is possible that consumers shied away from reformulated products as reduced-sodium products became pricier, but this explanation is not supported by our RD results (see robustness results, Section 2.6.5.2).

2.6.3.1 Decomposition by food group

Figure 2.8 shows the contribution of each food group to reformulation (Panel A) and switching (Panel B) components. To simplify the analysis, we adopt the broader definition of reformulation and lump entry and exit into the reformulation component.[40] The figure shows the contributions of all non-NSRI products as well as the individual contributions of products in each NSRI food group. To conserve space, we report the results from the sales-weighted (SW) measure.[41] The figure shows that NSRI products make up 64 percent of the reformulation component despite representing 48 percent of all UPCs in the market (see Table 2.4). At the same time, NSRI products are also the main drivers of the switching component (59 percent). The four largest contributors to reformulation and switching (*seasonings*, *bakery*, *dairy*,

[40]Since entry and exit are marginal components, this aggregation is inconsequential for our conclusions.

[41]Results for gram- and calorie-weighted measures are similar and available upon request.

and *meats*) make up more than 50 percent of the total contribution by products in NSRI categories; the seasonings food group is particularly notable as it makes up 19 percent (0.12/0.64) of the NSRI contribution to reformulation (17 percent for switching; 0.10/0.59) despite only making up about 4 percent of all NSRI UPCs.

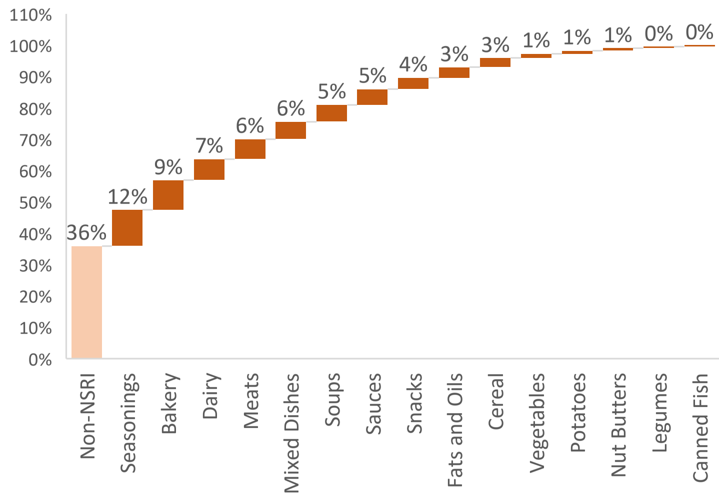
As stated in Section 2.6.1, reformulation patterns at the intake level need not reflect UPC-level results. In particular, we observe that the *median* UPC in non-NSRI food groups registered a slight increase in sodium content (see Table 2.4) while the reformulation component (Panel A in 2.8) registers a sizable improvement (i.e., decrease in sodium). This result is explained by the fact that a disproportionate portion of households' consumption in 2007 was concentrated in the *subset* of non-NSRI products that decreased their sodium content over the sample period. By the same token, the magnitude of switching for NSRI products (Panel B of Figure 2.8) is nearly 3 times larger than the reduction in median sodium content across NSRI UPCs reported in Table 2.4 (29 percent[42] vs. 10 percent). Put differently, firms disproportionately concentrated their sodium reduction efforts (for both NSRI and non-NSRI products) among popular products.

---

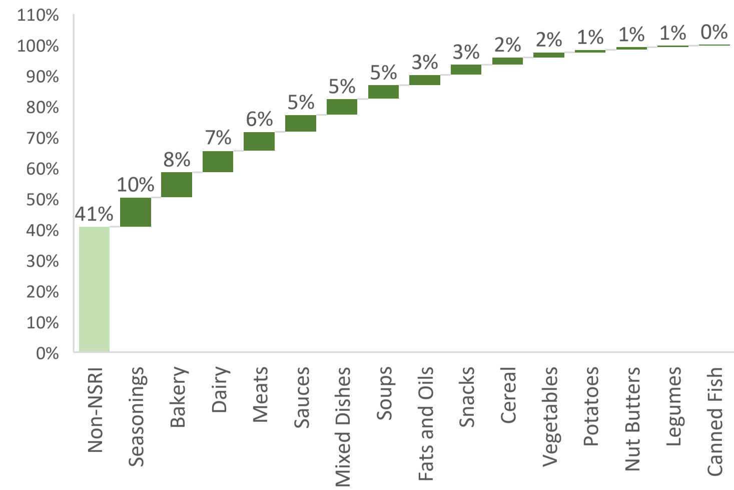
[42] $0.58 \times 0.4826 = 0.2847$.

Figure 2.8. Decomposition of reformulation and switching by food group

Panel A



Panel B

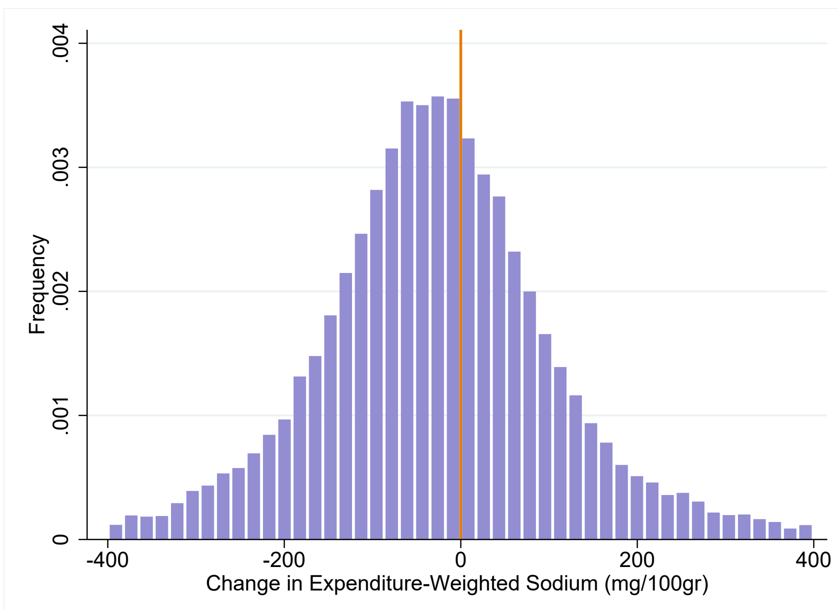


Notes. The figure shows the percentage contribution of the non-NSRI and individual NSRI food groups to the mean (across households) reformulation (Panel A) and mean switching (Panel B) reported in Figure 2.7. Reformulation denotes total reformulation, including pure reformulation, entry, and exit.

2.6.3.2 Socioeconomic and demographic factors

In this section, we examine whether socioeconomic and demographic factors are associated with the observed temporal changes in sodium intake. To conserve space, we report results using the sales-weighted (SW) measure (results using other measures are comparable and available upon request). While our prior results show that average sodium intake has decreased over time, there is substantial variation across households, as shown in Figure 2.9. Of the 23,556 households that we observe in both years, 60 percent (14,170 households) experience an improvement (i.e., a reduction) in sodium intake.

Figure 2.9. Distribution of the temporal change in the SW sodium intake measure, household data



Notes. Figure created with 23,556 households that appear in both 2007 and 2015 (balanced panel of households). The x-axis shows the change in the sales-weighted (SW) sodium intake measure over time and across households.2007 and 2015 Nielsen Home Measurement System (HMS) and Gladson.

To understand the association between sodium intake and socioeconomic and demographic factors, we run a battery of regressions designed to capture the relation at both the extensive margin (whether households cut their sodium intake) and the intensive margin (the extent to which households cut their sodium intake).

Explanatory variables capture income quartiles (as indicator variables), household characteristics (household size, presence of children, marital status), gender (share of women in the household), age (average age of household heads), education (average number of schooling years by household heads), employment (average number of weekly hours worked by household heads), ethnicity (Hispanic vs. non-Hispanic), and race (white, Black, Asian).[43]

Table 2.7 reports ordinary least squares (OLS) regressions on the intensive margin. The dependent variable is the change in sales-weighted sodium intake (ΔD_h) in the second column and the individual decomposition elements (reformulation, switching, and net entry) in the following columns. Reporting regressions on each decomposition element allows us to decompose the role of each socioeconomic or demographic factor: In a given row, the reformulation, switching, and net entry coefficients sum to the coefficient in the ΔD_h column. The main result from the ΔD_h regression is that income matters: Households in the second and third quartiles experience the greatest improvement, whereas households in the lowest quartile of the income distribution experience a less sizable (but significant) improvement in sodium intake. Also, households with children experience significantly smaller improvements.

[43]The data also contain another racial category labeled "other." The number of observations in this category is small. To simplify the analysis, we subsume these observations in the "white" category. (It is also difficult to interpret what a coefficient on this racial category might mean.) The online appendix contains further details on the explanatory variables used.

Table 2.7. Regressions of sodium intake changes on socioeconomic and demographic (SW measure, household data)

Demographics	ΔD_h	Reformulation	Switching	Net Entry
Income Q2	-16.48***	-8.890***	-0.0260	-7.566*
	(5.745)	(3.376)	(3.708)	(3.917)
Income Q3	-14.31**	-13.49***	5.732	-6.544
	(6.354)	(3.683)	(3.984)	(4.717)
Income Q4	-8.193	-13.11***	7.239*	-2.321
	(7.859)	(4.853)	(4.394)	(5.523)
Household size	-1.104	-1.449	1.601	-1.257
	(3.007)	(2.003)	(1.439)	(2.120)
Children	18.68**	13.13***	-4.936	10.49*
	(8.301)	(4.337)	(4.007)	(5.798)
Married	-4.517	-4.636	1.766	-1.648
	(5.426)	(3.289)	(3.272)	(3.946)
% Women	-7.241	-13.06***	8.658*	-2.836
	(7.339)	(4.260)	(4.424)	(5.341)
Age	0.0975	0.224	-0.194	0.0678
	(0.244)	(0.143)	(0.139)	(0.168)
Education	-1.771	2.174***	-2.216***	-1.729
	(1.252)	(0.735)	(0.734)	(1.171)
Employment	0.228	0.312***	-0.102	0.0181
	(0.164)	(0.0991)	(0.0950)	(0.116)
Hispanic	-6.131	-18.40***	17.23***	-4.965
	(7.810)	(4.741)	(4.902)	(5.102)
Black	13.61	-35.31***	39.20***	9.716
	(9.726)	(4.551)	(6.600)	(6.315)
Asian	16.06	0.923	9.489	5.651
	(12.76)	(8.518)	(5.943)	(12.05)
Constant	1.464	-273.1***	260.5***	14.10
	(26.50)	(17.93)	(15.43)	(18.13)

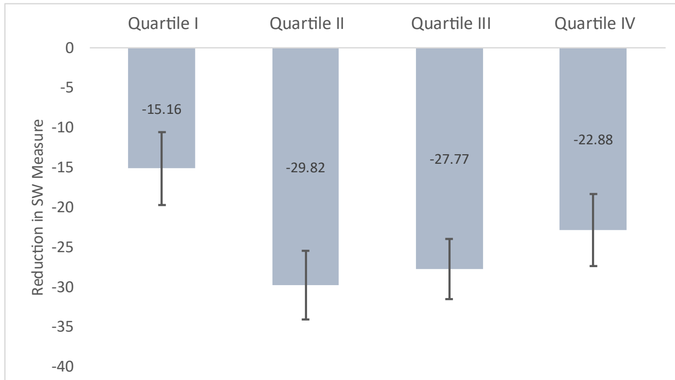
Notes. Sample size equals to 23,556. Robust standard errors in parentheses; *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Household size is the number of household members, children denotes the presence of children, % women is the share of women in the household; age, years of education, and employment (hours/week) are means of the household heads; Hispanic denotes ethnicity; reference group for race is white. Entry and exit are combined as net entry in the last column. Balanced panel of households. 2007 and 2015 Nielsen Home Measurement System (HMS) and Gladson.

Regressions on the decomposition elements confirm that a) the largest contributor for the observed overall sodium reduction is reformulation and b) switching behavior counteracts (and often neutralizes) this effect. This is particularly salient for gender (share of women in the household), ethnicity (Hispanic), and for households that identify as Black and less salient for households with more children and education. Consistent with results in Section 2.6.3, net entry plays a weakly significant role at best.

These results have at least two policy implications. First, Hispanic and Black populations appear to gravitate toward saltier products even though products redesigned to contain less sodium are available to these populations. This finding suggests that policies to incentivize healthier eating in these populations are not working very well (in their current format) or that (if they do have an effect) a heavier investment should be made in them. A second implication is drawn from the large and positive coefficient on children in the reformulation and net entry regressions, which suggests that manufacturers have not done as much in terms of reducing sodium content for products targeted to children (e.g., snacks, ketchup) or perhaps have worsened their nutritional profile. This observation is consistent with results in Table 2.4: Of all NSRI categories, *snacks* and *sauces* display the smallest improvements in sodium content.

To visualize how each income quartile stacks against each other in terms of their improvement (i.e., a negative ΔD_h), Figure 2.10 reports the sales-weighted (SW) sodium intake reduction in each income quartile that results from evaluating the regression coefficients for each of the four groups of households at their regressor means. The figure illustrates that all quartiles—including households at the bottom of the income distribution—exhibit a significant reduction in sodium intake; however, the reduction in the upper three quartiles is significantly larger than that in the bottom quartile (almost twice as large for the second and third quartiles).

Figure 2.10. Reduction in sodium intake by income quartile, SW measure (household data)



Notes. Based on coefficients from ΔD_h regression in Table 2.7. For each income quartile, $\Delta\widehat{D_h}$ is evaluated at the mean values of the regressors in that quartile. Balanced panel of households. 2007 and 2015 Nielsen Home Measurement System (HMS) and Gladson.

Table 2.8 further explores the heterogeneity in the temporal sodium intake change. First, we explore a possible nonlinearity in the association between sodium intake and socioeconomic and demographic factors. We run separate regressions for households that experience an improvement in and worsening of sodium intake ($\Delta D_h < 0$ and $\Delta D_h > 0$, respectively). Second, we consider an extensive-margin specification: a probit model on whether a household experiences an improvement. OLS results (the second and third column) reveal that the income effect observed in Table 2.7 is driven by households that experience a worsening in sodium intake over time. From a policy perspective, this result suggests that the disparities in dietary quality across income levels are becoming more pronounced in more vulnerable populations (i.e., those that are not reducing their sodium intake). A similar takeaway can be drawn from ethnicity and race: Hispanic, Black, and Asian coefficients in the worsening regression are positive (although only the coefficient on Black is statistically significant). Interestingly, among households that improved their sodium intake over time, the disparities across ethnicity (Hispanic vs. non-Hispanic) and race (white vs. non-white) decrease substantially and significantly.

Table 2.8. Improvement and worsening regressions (SW measure, household data)

Demographics	Improvement $\Delta D_h < 0$ [OLS]	Worsening $\Delta D_h > 0$ [OLS]	$I = \begin{cases} 1 & \Delta D_h < 0 \\ 0 & \Delta D_h > 0 \end{cases}$ [Probit]
Income Q2	1.552	-20.97**	0.0207**
	(5.040)	(8.561)	(0.00890)
Income Q3	7.192	-17.89*	0.0349***
	(5.822)	(9.695)	(0.00944)
Income Q4	5.805	-0.730	0.0337***
	(6.785)	(11.15)	(0.0108)
Household size	-4.209	4.417	-0.00583
	(3.440)	(3.855)	(0.00438)
Children	26.73***	-9.162	-0.0232*
	(7.046)	(11.46)	(0.0129)
Married	13.61***	-33.91***	0.0127
	(4.811)	(9.030)	(0.00897)
% Women	-12.78**	-3.090	-0.0113
	(5.691)	(10.76)	(0.0114)
Age	0.239	-0.0316	0.000582
	(0.187)	(0.339)	(0.000361)
Education	-1.462	0.212	0.00242
	(0.956)	(1.925)	(0.00168)
Employment	-0.00277	0.0907	-0.000721***
	(0.128)	(0.275)	(0.000257)
Hispanic	-24.79***	10.07	-0.0263*
	(8.126)	(10.13)	(0.0137)
Black	-41.74***	73.84***	-0.0320***
	(6.522)	(15.34)	(0.0116)
Asian	-27.09**	20.83	-0.0463**
	(11.21)	(14.82)	(0.0200)
Constant	-110.8***	129.8***	
	(22.93)	(41.44)	
Observations	14,170	9,386	23,556

Notes. Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1. Household size is the number of household members, children denotes the presence of children, % women is the share of women in the household; age, years of education, and employment (hours/week) are means of the household heads; Hispanic denotes ethnicity; reference group for race is white. Probit results correspond to the marginal effects. Balanced panel of households. 2007 and 2015 Nielsen Home Measurement System (HMS) and Gladson.

Some extensive margin results presented in the last column in Table 2.8 are in line with the overall intensive margin regressions reported in the second column of Table 2.7. Specifically, households in the lowest income quartile are less likely to improve their sodium intake over time, as are households with children. In contrast to the intensive margin results, however, Hispanic, Black, and Asian populations are less likely to experience an improvement, as are household heads that work more. Taken together, results from this section support the notion that diet quality disparities along the income and racial/ethnic minority dimensions exist and that they may have worsened over time.

2.6.4 NSRI effects

Table 2.9 displays RD results from four specifications. Specification 1 uses all continuing UPCs that fall within the selected bandwidth around the target threshold, and Specifications 2–4 focus on different subsets of UPCs. Specification 1 reveals that UPCs located near and on opposing sides of the threshold reformulated products differently over time: Products to the right of the threshold reduced sodium at an increasing rate ($\widehat{\nu}_1$ is negative and significant), whereas products to the left did not have a detectable change in sodium content ($\widehat{\gamma}_1$ is neither economically nor statistically significant). Note that, as in the visual evidence, these results are not only consistent with a kink in slope at the threshold but they confirm the theoretical prediction that the slope below the threshold is indistinguishable from 0. The magnitude of the estimated coefficient implies that noncompliant products reduced, on average, approximately 40 percent of the excess sodium (as defined by NSRI targets). Loosely speaking, compliance on the intensive margin was partially effective.

Table 2.9. KRD estimation results

D.V.: Δ Sodium (mg/100 g)	Spec. 1: *All Firms*		Spec. 2: *NSRI Signatories*		Spec. 3: *Other Signatories*		Spec. 4: *Non-Signatories*	
Panel A: Estimates								
KRD estimate: $\widehat{\nu}_1$	-0.40		-0.29		-0.33		-0.35	
p-value	<0.001		0.09		0.003		0.004	
LHS slope: $\widehat{\gamma}_1$	0.01		-0.08		0.02		0.00	
p-value	0.75		0.012		0.36		0.99	
Panel B: N and bandwidth (b) around the cutoff=0 (c)								
	L	R	L	R	L	R	L	R
Total N	9,701	37,360	714	2,864	453	2,145	8,534	32,351
Effective N	7,824	31,213	588	2309	444	1,889	7,235	27,972
MSE-Optimal Bandwidth (b)	-475	475	-430	430	-650	650	-551	551

Notes: D.V. is dependent variable. KRD is kink regression discontinuity. N is number of observations. Bandwidth (b) is in mg/100 g. P-values are based on robust inference with standard errors clustered at 9 broad food departments as described in Section 2.5.2. See Section 2.5.2 for further estimation details.

When applied to the subset of products manufactured by NSRI signatories (Specification 2 in Table 2.9), the KRD parameter provides similar evidence (although the magnitude is smaller); this specification, however, suggests that the slope below the threshold is slightly negative. Specifications 3 and 4 investigate whether the effect is present in products manufactured by non-NSRI signatories. Specification 3 (other signatories) looks at products manufactured by the set of firms that committed to other nutrition improvement initiatives around the time of the NSRI (the HWCF and IFBA initiatives; see Section 2.3.3), and Specification 4 looks at the remainder of the products (those manufactured by "non-signatories," firms that did not participate in any voluntary reformulation initiatives). Results in Specifications 3 and 4 are consistent with those in Specification 1 (for both $\widehat{\nu}_1$ and $\widehat{\gamma}_1$). More interestingly, these findings indicate that NSRI effects were not limited to NSRI participants but widespread across the entire food industry.

Our finding that the effect is also present in non-NSRI firms may seem surprising at the outset. We note, however, that this result is consistent with anecdotal evidence that other companies were using the NSRI targets as the sodium content reference in their manufacturing decisions (see Section 2.3.3). Further, and perhaps more importantly, we note that this result is consistent with the food industry view that reformulation should be carried out silently (see Section 2.3.3); an explicit commitment to the NSRI would have run counter to firms' objective of silent reformulation.

2.6.5 Robustness checks and additional Results

2.6.5.1 Decomposition

As explained in Section 2.4.1, we choose household data over retail data (RMS) for our main analysis because it is more suitable for the decomposition analysis, which relies on the inspected units being as comparable and consistent as possible over time.[44] Units of analysis in RMS are not households but stores; as such, these units are likely to experience important changes over time (e.g., changes in the store's product mix, the mix of consumers visiting the store, change of ownership/manager). In addition, the household data allow for a more detailed analysis of how socioeconomic and demographic factors affected changes in sodium intake (see Section 2.6.3.2).

In this section, we report robustness checks of our results in Sections 2.6.2 and 2.6.3 using Nielsen's retail dataset. To simplify the analysis for the RMS data, we aggregate store data up to the county level, which we then use in the decomposition analysis.[45] Results, reported in Appendix B, are qualitatively similar to those reported for the household data. Table B.4 shows that the decrease in the median sodium intake ranges from 2.66 percent to 7.02 percent across the alternative measures and county samples. Compared to the household data (−4.02 percent to −8 percent range), sodium reduction in RMS appears less sizable but consistent across intake measures and samples.[46] We observe a similarly consistent result for the distribution of sodium intake; Figures B.2 and B.3 show a generalized sodium decrease across most quantiles, although in a less pronounced magnitude for some quantiles than that registered in the household data. Figure B.4 and Tables B.5–B.7 present the decomposition results

---

[44] We do acknowledge, however, that households may also experience changes over time (e.g., income, number of members). We argue that units in RMS data are subject to substantially less consistency.

[45] RMS reports data at the UPC-week-store level (see the online appendix for details). One option is to carry out the analysis using store-level retail data, but this approach is not ideal since there is a nontrivial amount of store turnover in the panel and the decomposition analysis relies on observing data for the same unit across periods. Further, the basket of food products is both limited and unrepresentative in many of the stores (e.g., drug and convenience stores). Aggregating stores up to the county level results in a more meaningful analysis as it is possible to capture consumers' food purchases across stores in a geographic area.

[46] One explanation for the difference in magnitudes across household and retail datasets is that retail data include sales from outlets that are unlikely to be representative of the population's shopping basket. For instance, many customers of mass-merchandisers (included in the retail data) are small-business owners that buy in bulk. Also, convenience stores (also included in the retail data) represent purchases from a limited set of products that are not representative of the typical family diet.

for the retail data. While magnitudes are less sizable than those in the household data, the overall results are qualitatively similar across the alternative measures of sodium intake: a) reformulation and switching have the greatest effects, but they nearly cancel each other out, and b) net product entry plays a marginal role.

2.6.5.2 Causal inference

The key testable assumption for the validity of a KRD design is that predetermined covariates do not exhibit a kink (the smoothness condition, see Card et al., 2015). To probe this assumption, we utilize the same approach and mechanics that we applied to our outcome of interest (Section 2.6.4), except that here the dependent variable is replaced by a predetermined variable. We consider seven covariates (at their 2007 levels): calories, carbohydrates, saturated fat (fat), price, protein, and sugar. The first column of Table 2.10 reports the corresponding $\widehat{\nu}_1$ coefficient estimated in each separate regression. None of these results raise suspicion regarding a violation of the smoothness condition.

We also probe the results via the placebo tests visually presented in Section 2.5.2. The idea is similar to the smoothness tests, except that here the dependent variable is given by the temporal *change* in product characteristics (other than sodium content). The last column in Table 2.10 reports the corresponding $\widehat{\nu}_1$ coefficient for these regressions. The statistical evidence (or lack thereof) shown in these results confirms the visual evidence seen in Section 2.5.2. Further, these results suggest that manufacturers did not compensate for sodium reductions by making changes to other product attributes; this result is interesting since food manufacturers are known to simultaneously manipulate several product constituents to obtain a product's desired taste (see Section 2.3.4). Similarly, the lack of significance in the price regression indicates that sodium reductions were not accompanied by a change in the equilibrium price (see Section 2.3.4).[47]

Table 2.10. KRD estimate ($\widehat{\nu}_1$) for smoothness and placebo tests

Dependent Variable (D.V.)	Smoothness Tests: *D.V. at 2007 levels*	Placebo Test: Δ *in D.V. 2007–2015*
Calories (kcal/100 g)	0.17 [0.63]	0.04 [0.84]
Carbs (mg/100 g)	0.03 [0.52]	0.001 [0.79]
Fat (mg/100 g)	-0.001 [0.89]	-0.000 [0.90]
Price ($/100 g)	-0.001 [0.65]	0.000 [0.54]
Protein (mg/100 g)	0.01 [0.66]	0.000 [0.63]
Sugar (mg/100 g)	-0.02 [0.36]	0.01 [0.16]

Notes: [p-values] based on robust (clustered SE) inference (see Section 2.5.2 for details). All regressions employ the same bandwidth as the baseline regression (Specification 1, Table 2.9).

We also expose our specification to the addition of covariates. We re-estimate $\widehat{\nu}_1$ by including a rich set of product-category specific effects in equation (2.4). To maximize the explanatory power of these covariates, we use the finest product category definitions provided by Nielsen, called product module codes (PMCs); there are nearly 600 PMCs in our dataset.[48] An advantage of this rich set of fixed

[47]Since ours is a reduced-form approach, we can only measure the direction of change, not its sources. The finding would be consistent (but not conclusive) with firms not adopting a different pricing strategy for reformulated products. For instance, our finding is consistent with a simultaneous contraction of demand and supply.

[48]See the online appendix, Section 1.1; Section 3.1.4 in the online appendix displays the correspondence of PMCs to NSRI food categories. We refer to PMCs as product categories, which is different from the "food categories" term we use to refer to NSRI's targeted categories.

effects is that they absorb differences in (average) reformulation across different product categories, thereby increasing the comparability of the measured effect across categories. Our results, although less precise, remain largely consistent with our main results (p-value): -0.32 (0.03), -0.10 (0.37), -0.36 (<0.01), and -0.32 (0.14) (for Specifications 1–4, respectively).

We run a battery of additional tests, reported in Appendix C, including robustness to bandwidth choice and polynomial order. In terms of bandwidth, results remain robust for an ample range although (as usual with RD designs) preciseness declines with bandwidth size. In terms of polynomial, we find that our results in Table 2.9 remain consistent, although Specifications 2 and 3 lose significance due to the substantially lower number of observations with respect to Specifications 1 and 4. In addition, we conduct a data-driven exercise to verify that the location of the kink in slope at the NSRI thresholds is quantitatively supported vis-à-vis other potential kink points.

2.7 Conclusion

In this paper, we present, as far as we are aware, the most detailed and comprehensive documentation of the evolution of sodium content and intake in the United States in recent times. We focus on the periods before and after the implementation of the most important societal effort to reduce sodium intake, the National Salt Reduction Initiative (NSRI). Our results strongly confirm that product sodium content and consumer sodium intake have significantly declined over this period of time. Further, our empirical analysis indicates that the NSRI played a significant causal role: All else equal, sodium content fell sharply for products that exceeded the NSRI's sodium content targets but not for those that were below the targets. Interestingly, efforts to reduce sodium in products were evident in both signatory and nonsignatory firms. This suggests that the entire industry might have perceived future regulatory intervention (e.g., in the form of sodium content limits) to be a credible threat and that adhering to the NSRI was likely only symbolic. This finding is also consistent with the industry's preference for pursuing reformulation in a silent fashion (see Section 2.3.3).

While these results may seem encouraging, we find that consumers have shifted their purchases to saltier alternatives, thereby significantly (and almost entirely) limiting the effect that reformulation efforts have had on total intake. Had consumer shopping habits in 2015 remained similar to those seen in 2007, sodium intake reduction targets set forth by the NSRI and the World Health Organization would have been met. Our results suggest that one reason for this setback is that reformulation might have occurred too quickly. While manufacturers and policy makers agree sodium content reductions should be gradual so as to be effective in reducing intake (see Section 2.3), there is some anecdotal evidence suggesting that some manufacturers may sometimes miscalculate the speed of gradual reductions (see Brat and Tamman, 2010).

Taken together, our results suggest that while progress will likely continue on the supply side, changes in consumer behavior could achieve much more substantial gains in the efforts to curb sodium consumption. From a policy perspective, our research suggests that significant resources and study should be devoted to understanding the reasons behind consumers' aversion to reformulated products. While one explanation for this aversion is that reformulated (reduced-sodium) products were introduced at higher price point, our causal results are not consistent with this notion. Designing consumer-oriented strategies that focus on the most salient (non-price) aspects behind consumers' responsiveness (or lack thereof) is thus essential. Our results provide some light in this regard: Policy emphasis should be placed on lower income groups and minorities, where the disparities in diet quality are not only most evident but have also become more pronounced over time.

Our work is not without limitations. We only analyze food at home. However, while food away from home is an important component, we do not have reason to believe that firms have adopted a dramatically different strategy as it pertains to reducing sodium content in their products. For example, the NSRI had a separate set of targets (and signatories) for the food away from home segment. Also, many food manufacturing conglomerates sell their products through both channels (e.g., restaurants and supermarkets), making reformulation efforts likely across the board. While our work does look at effects that reformulation could have on other market outcomes (certain food components as well as price), we do not study other possible mechanisms (e.g., advertising). We conjecture, however, that advertising reactions were limited or nonexistent given the strong stance the industry has taken on making sure that reformulations go unnoticed by the consumer (Brat and Tamman, 2010).

Yet manufacturers can still have important strategic responses in tandem with reformulation, such as market segmentation strategies designed to offer different product lines catered to different populations (e.g., health vs. nonhealth conscious consumers) with likely different willingness to pay (or price sensitivity). These responses would be accompanied by more elaborate pricing and packaging changes (e.g., modification in serving size and/or total product size). Structural work, which would allow for more precise disentanglement of supply and demand forces, would be more appropriate to answer these types of questions. An additional question that we leave for future research is whether socioeconomic and demographic disparities in access to healthier (i.e., lower sodium) food are related to geographic disparities in retail configuration and availability; for example, households in food deserts may not be able to buy lower sodium products even if they want to.

CHAPTER 3

HOW MUCH DO PEOPLE CONSUME AFTER A SODA TAX? EVIDENCE FROM THE PHILADELPHIA, PA, TAX EVENT

3.1 Introduction

On average, Americans consume 13.8 ounces of soft drinks every day.[1] A single 12-ounce can of Coca-Cola contains 39 grams of sugar, a daily sugar intake already well above the tolerable upper limit recommended by the American Health Association for adult women (24 grams) and men (36 grams). Soda consumption is associated with obesity, diabetes, heart attack, and stroke—all of which the World Health Organization (2021*b*) classifies as preventable diseases and together represent 25 percent of healthcare costs in the United States.[2] One method proposed to combat these diseases is to impose sweetened beverage taxes (also known as soda taxes).

Taxing soda, if effective, could reduce sugary drink purchases that degrade people's health and raise public revenue for programs improving public health, a win-win situation for society that has already proven effective in the case of tobacco (World Health Organization, 2021*a*). As a result, seven major U.S. cities—Berkeley, CA (2015), Philadelphia, PA (2017), Boulder, CO (2017), Albany, CA (2017), Oakland, CA (2017), San Francisco, CA (2018), and Seattle, WA (2018)— and one county—

---

[1] https://www.statista.com/statistics/306836/us-per-capita-consumption-of-soft-drinks/.

[2] Finkelstein et al. (2009) find that annual obesity costs in 2008 represent 10 percent of all medical costs; the American Diabetes Association (2018) calculates economic costs of diabetes to be as much as $328 billion in 2017 (nearly 9 percent of all healthcare costs in 2018); and Benjamin et al. (2018) estimates $199 billion medical costs related to heart disease and stroke in 2017 (6 percent of all healthcare costs in 2018). See Fortune (2019) for total U.S. healthcare costs in 2018.

Cook County, IL (2016)—implemented sweetened beverages taxes between 2015 and 2018. These taxes are levied on manufacturers and reflected in shelf price with the dual purpose of discouraging soda consumption and raising public revenue.[3] Growing scrutiny of the beverage industry over the past decade has raised questions about whether soda will face a similar fate as tobacco and surprise future generations with its once-upon-a-time high-consumption prevalence.

An active literature explores whether soda taxes are effective. The primary policy effect of a tax is acknowledged to be how much (if any) people cut their soda consumption as a result of increased prices, which is typically measured by calculating the price elasticity of demand for taxed products.[4] The empirical evidence is mixed. One group of studies have reported little to no effect, while others report more sizable reductions in soda consumption after the tax. For example, a recent meta-analysis by Powell et al. (2021) reports price elasticity of demand for taxed beverages of 0.06 [−5.05; 6.15] in Rojas and Wang (2021) compared to −13.88 [−20.44; −7.57] in Lee et al. (2019) for the tax event in Berkeley, CA, where numbers in brackets represent 95 percent confidence intervals.[5] For the Philadelphia, PA, event, estimates range between −0.10 [−1.08; 0.88] in Lawman et al. (2020) and −2.52 [−4.94; −0.11] in Cawley et al. (2019*b*).

[3]Tax revenue is collected ideally to improve public health in a more comprehensive manner. For example, city authorities often declare that tax money will be spent on building or upgrading parks and recreation centers. For example, Philadelphia Mayor Jim Kenney designed and advocated a program called "Rebuild," a comprehensive project to show where tax dollars go. The program has funded the renovation of nine recreation centers and libraries as of 2022, with another nine under construction (Levy, 2022).

[4]Some studies report price elasticities with and without accounting for tax avoidance mechanisms. For example, Seiler, Tuchman and Yao (2021) finds elastic demand without accounting for cross-border shopping behavior (−1.35) but inelastic demand (−0.65) with cross-border shopping. As the authors note, the first measure is relevant for tax revenue assessment, while the second is relevant for assessing the impact of the tax on people's consumption and nutrition intake.

[5]Using 26 estimates from 19 peer-reviewed studies of five local soda tax events, the authors report an average 20 percent decline in soda consumption, with an estimated price elasticity of demand of −1.47. They conclude that taxing soda is an effective policy tool associated with significant reductions in soda consumption.

Empirical analysis of whether soda taxes work to curb consumption is still novel in this literature and could help account for discrepancies between the predicted effects of soda taxes and how consumption actually changes after their implementation. I use detailed, household-level scanner purchases data on 3,729 barcoded beverage products by 24,793 households in 11,994 stores from 2006 to 2017 to assess firm and consumer responses to the Philadelphia, PA, soda tax. My contributions are two-fold: First, I construct a comprehensive sample of household-level sugary drink purchases for a larger panel of households, stores, products, and time periods than has been previously reported in the extant literature. Earlier studies often used retail scanner sales data or self-reported survey data. Unlike retail sales data, household purchases have the advantage of providing a crisper picture of individual beverage consumption and sugar intake. As oppose to self-reported survey data, household scanner purchases are less vulnerable to measurement error and thus might provide greater accuracy and precision for the estimates.[6]

Second, I build upon the existing empirical literature by utilizing recent advances in causal machine learning tools that I borrow from the statistics and econometrics causal inference literatures. In particular, I report empirical evidence using matrix completion and transposed elastic-net methods[7] as in Athey et al. (2021) and provide evidence from traditional difference-in-differences approaches following Card and Krueger (1994), Cawley et al. (2019*b*), Rojas and Wang (2021), and Seiler, Tuchman and Yao (2021). Reporting evidence from alternative research techniques has an advantage of assessing how fragile the results are to different identifying assumptions made by these techniques, thereby providing more credibly causal (and robust)

[6]Nielsen Homescan allows households to record purchase prices and volumes at the barcode level with the use of scanning devices, while self-reported survey data are based on recall. However, an important limitation is that Homescan data do not record food away from home, which is estimated to make up as much as 28 percent of total U.S. food consumption (Allcott, Lockwood and Taubinsky, 2019*a*). I also consider purchases to be equivalent to consumption, assuming zero waste or gifting.

[7]Transposed elastic-net is also referred as “horizontal regression” in Athey et al. (2021).

estimates. More specifically, machine learning findings alleviate concerns of potential issues that might result from the violation of the parallel trends assumption between affected and unaffected regions.

I use three main databases and build a comprehensive sample of household purchases of taxed and untaxed products in affected and control regions. First, I use household-level scanner data on weekly barcoded purchases from Nielsen from 2006 to 2017. I use population weights for panelists in my data to better measure tax impacts on firm response in terms of pass-through rates (i.e., how much of the tax is passed on shelf price). Population weights are annually assigned by Nielsen to participating households using a projection methodology that employs demographic and socioeconomic data from the U.S. Census Bureau. As noted in the literature (e.g., Hitsch, Hortacsu and Lin, 2019), population-weighted household-level scanner data better represent the U.S. population for aggregate measures such as total household spending in retail stores compared to store-level scanner data (e.g., Nielsen RMS) that contain information only from a select sample of stores and products that are not necessarily representative of the entire U.S. population. Hence, the use of household scanner data with sampling weights provides a more accurate assessment of what portion of the tax is passed on to U.S. consumers.

Second, I complement the scanner data with barcode-level grams of sugar and ingredients list from Gladson and determine the tax status of products. The City of Philadelphia has provided (and periodically updated) detailed guidelines for which products are subject to the sweetened beverages tax.[8] Beverages that contain caloric sweeteners (e.g., high fructose corn syrup, cane sugar, dextrose, fructose, glucose), including juice concentrates that are in excess of "what would be expected from the same volume of 100 percent fruit or vegetable juice of the same type" (e.g., apple

[8]Detailed guidelines for the sweetened beverage tax are posted on the website of the City of Philadelphia (2017).

juice concentrate), and nonnutritive sweeteners (e.g., aspartame, sucralose) have been subjected to the tax. Gladson's product ingredients list contains useful information on whether the product contains caloric or artificial sweetener and whether it has more sugar than one would expect from 100 percent juice. I also utilize Gladson data when cataloging product categories for taxed (e.g., soda, sweetened juice) and untaxed (e.g., 100 percent juice, water) beverages.

Last, I complement these two databases with 2010 and 2014 MABLE GeoCorr geographic correspondence data. GeoCorr provides mapping between cities and 5-digit zip code areas as well as information on latitudes and longitudes of the 5-digit zip code areas. I use latitudes and longitudes to calculate the geodesic distance between a 5-digit zip code area outside Philadelphia, PA, and the closest 5-digit zip code area in the city. By calculating 3D geodesic distance rather than 2D planar distance, I account for the curved surface of the earth and calculate distance and direction more accurately in any direction between any locations.[9]

Despite mixed findings, the empirical literature on sweetened beverage taxes appears to have reached a consensus regarding the two major channels that need to be analyzed to assess effectiveness of a tax. These channels are (i) the pass-through rate that measures the impact on retail price of the taxed beverages and (ii) the change in consumption accounting for different tax avoidance mechanisms such as engaging in cross-border shopping behavior (shopping in nearby untaxed stores) and switching to untaxed products in the taxed jurisdiction. As excise soda taxes are levied on manufacturers, the proportion of the tax to be passed through to the final consumer in the form of increased retail prices becomes the firm's decision. At certain times and in certain places it might be more optimal for firms to share a portion of the tax burden with their consumers. This scenario would be more likely if consumers could

[9]Planar distance could be the preferred method of calculating distance between two location if speed is of more concern than accuracy.

not avoid the tax either in product (by switching to untaxed product alternatives) or geographic space (by switching to stores outside tax jurisdiction). In the case of Philadelphia, the taxed product categories are soda (both regular and diet), sports drinks, energy drinks, sweetened fruit juice, sweetened milk, and sweetened tea/coffee, while untaxed beverage categories are pure water, 100 percent juice, unsweetened milk, and unsweetened tea/coffee. Narrower sets of untaxed categories compared to those in other tax events (e.g., Berkeley, CA) might make it harder for Philadelphia households to switch to alternatives in product space. However, the city is in close proximity to several untaxed regions, making it likely that Philadelphia households will shop at stores in these regions.

I start with traditional difference-in-differences (DiD) methods. This baseline estimation model allows me to directly test the effect of the tax on two outcome variables: per ounce price in cents (where I can use the DiD coefficient to infer the pass-through rate by dividing the coefficient estimate by the tax rate of 1.5 cents per ounce) and log volume purchases (where the DiD coefficient refers to the percentage change in consumption). To address concerns that neighboring zip codes might be partially treated due to potential spatial spillovers of the policy, I experiment with alternative comparison groups that are not too close to the treated regions, namely creating an artificial "doughnut hole." At the same time, I utilize purchases made by the affected households in nearby "unaffected" regions (i.e., the hole in the doughnut) and account for the effect of cross-border shopping behavior on consumption, a noted tax avoidance mechanism in the literature.

The fact that I can clearly distinguish treated and control regions allows me to implement event-based approaches. Thus, I utilize two frontier causal machine learning tools, matrix completion and transposed elastic-net, as in Athey et al. (2021), in addition to the traditional DiD estimator. I compare results from these alternative estimators as further robustness checks of the DiD findings.

Overall, I find that 64 percent of the tax has passed to consumers, leading to a 12.3 percent price increase. The result is robust to alternative control samples that I construct using a criterion based on a distance threshold from the city border: Evidence shows an average 69 percent pass-though rate. Consumption results come from both benchmark DiD and alternative machine learning estimators and report approximately 39 percent reduction using both the DiD and matrix completion methods and 21 percent reduction from the transposed elastic-net estimator. I also find that the decline in consumption becomes slightly less pronounced (31 percent) after accounting for the cross-border shopping behavior of the affected households as a way of avoiding the tax.

When cross-border shopping behavior is not considered, the estimated price and quantity responses constitute a price elasticity of demand of -3.17, using the benchmark DiD estimator (it is -1.73 with transposed elastic-net and -3.22 with matrix completion). Once cross-border shopping behavior is accounted for, the estimated elasticity for the benchmark model becomes -2.49. As noted in the literature, the first set of elasticities is relevant when assessing tax revenue, while the second elasticity is relevant when assessing the tax impact on all purchases by Philadelphia households and thus more fully reflects the actual decline and can be used to assess the change in sugar intake.[10]

The remainder of the paper is organized as follows: Section 3.2 reviews the literature on the effects of local sweetened beverage taxes. Section 3.3 describes the data and sample construction. Section 3.4 presents the research design and alternative

[10]Relatively larger magnitudes of estimated elasticities compared to those in the extant literature in part come from a broader coverage of taxed beverages including barcoded products from sweetened milk and sweetened coffee/tea. The elasticities are -1.69 without cross-border shopping and -1.34 with cross-border shopping when I keep soda and diet soda categories, to which the majority of products belong. These magnitudes are more in line with what is reported in the extant literature for the Philadelphia tax event (e.g., -1.35 in Seiler, Tuchman and Yao (2021), -2.52 in Cawley et al. (2019*b*)).

estimators. Section 3.5 reports the main results of price and consumption effects of the Philadelphia soda tax and examines their sensitivity. Section 3.6 concludes.

3.2 Literature

The impacts of sweetened beverage taxes on price and consumption have been studied in detail in the existing literature (see Cawley et al., 2019*a*; Allcott, Lockwood and Taubinsky, 2019*b*, for reviews). Most studies report consumption effects using either store-level retail sales or self-reported survey data. This research contributes to this broad literature by reporting evidence on price and consumption using detailed, household-level purchases data at the barcode level for the Philadelphia, PA, tax event.[11]

Hypothetical event studies (Andreyeva, Long and Brownell, 2010; Dharmasena and Capps, 2012; Dubois, Griffith and O'Connell, 2017*b*; Wang, 2015; Wang, Rojas and Colantuoni, 2017) use the structural method and report a relatively high pass-through rate (e.g., full pass-through in Wang, 2015) and negligible consumption reduction. Event studies of soda taxes in U.S. localities, however, generally use traditional difference-in-differences (DiD) techniques and estimate the impact on price and/or consumption in the affected tax region using control regions that are typically chosen based on their proximity to the affected region. In my main estimations of changes in price and consumption as a result of the tax event in Philadelphia, PA, I too use a DiD estimator. Unlike the existing literature, however, I use this estimator on more disaggregate sample data (i.e., at the barcode level in price and household level in consumption regressions) while also utilizing two alternative estimators (matrix completion and transposed elastic-net) borrowed from the causal machine learning

[11]Some studies report detailed data at the barcode level, but these either use store-level scanner data, refer to different tax events, or both (e.g., Bollinger and Sexton, 2018; Aaker and Carman, 1982, for Berkeley, CA).

literature. While more granular data allow me to draw a crisper picture of causal tax effects by flexibly considering the effects of individual products, machine learning estimators allow me to build robust counterfactual tax regions that provide further credibility for the parallel trends assumption (i.e., the identifying assumption in the potential outcomes framework for the causal effect).

The first U.S. locality that implemented an excise soda tax was Berkeley, CA. Evidence for its effectiveness is mixed. Using manually collected survey data, Cawley and Frisvold (2017) and Falbe et al. (2016) compared prices in Berkeley, CA, before and after the implementation of the tax to nearby locations (Oakland or San Francisco) and reported 40 percent–50 percent pass-though rates, while Falbe et al. (2016) also reports a 21 percent decline in consumption. Silver et al. (2017) reports a higher pass-through rate of 67 percent but a reduction in consumption of just under 10 percent. By contrast, Rojas and Wang (2017) find no effect on sales (consumption) and a relatively moderate pass-through rate of 24 percent.[12]

Philadelphia, PA, was the second U.S. locality to implement an excise soda tax. With hand-collected data, Cawley, Willage and Frisvold (2018) exploit a natural experiment at the Philadelphia International Airport, where only half of stores were taxed, and report a passed-through rate of 61 percent. To account for network effects (i.e., the fact that the comparison stores might also respond to the tax by increasing price) they also estimate the effect with a time-series model using a before-after comparison and report a pass-through rate of 93 percent. Cawley et al. (2020*a*) report full pass-through and provide suggestive evidence on cross-border shopping behavior (higher pass-through rates from taxed stores that are farther away from untaxed nearby stores) and regressive tax structure (higher pass-through rates in poorer

[12]Most of these studies also break down the effects on pass-through rates and consumption by beverage category (regular soda vs. energy drinks), product size (12-ounce 12 packs vs. 2 liter bottles), and store type (convenience stores vs. grocery stores) as well as different demographic and income groups.

neighborhoods). Seiler, Tuchman and Yao (2021) analyze the impact using both store-level scanner data (price and quantities) and hand-coded nutritional information for stores in and outside of the City of Philadelphia at the level of 3-digit zip code areas. They report an average pass-through rate of 97 percent accompanied by a 46 percent decline in sales (consumption) of the taxed products without any significant substitution to untaxed products. However, after accounting for cross-shopping behavior, the net reduction in consumption is reported as 22 percent. The current study is most related to Seiler, Tuchman and Yao (2021), as I too report evidence on price and consumption using disaggregate data from the same population. My study differs from Seiler, Tuchman and Yao (2021) however by (i) using a long panel of household scanner data (as opposed to retail scanner data) and (ii) reporting robust findings from a larger pool of control groups (counterfactual Philadelphia's). More specifically, I construct alternative control samples both from alternative estimators (as opposed to DiD alone) and from a larger panel of stores (and households) outside the city (rather than stores in Pennsylvania and New Jersey alone) depending on their distance to the city border (e.g., 10 miles, 20 miles).

Two more studies exclusively analyze the effect of the Philadelphia tax on consumption using repeated cross sections of individual-level data. Zhong et al. (2018) collect phone survey data for a short time span before and after the implication of the tax and report 40 percent and 64 percent reductions in consumption of soda and energy drinks, respectively, while bottled water consumption rose by 58 percent. Cawley et al. (2019*b*) investigate the impact of the tax on consumption over a longer time span and report a 30 percent reduction in daily consumption of regular soda but no impact on bottled water consumption. Apart from the known issues inherent to self-reported survey data, both studies have information from a limited number of comparison groups that are too close to Philadelphia; hence, they are unable to account for the tax's possible spillover effects outside of the city. This study reports evidence on price and consumption from more comprehensive (barcode-level longitudinal as opposed to short panels of a couple of months) and reliable household-level data (collected from hand-held scanning devices rather than manually collected survey data).

An increasing number of studies add to the empirical literature providing evidence from more recent soda tax events in the United States. Using Nielsen retail scanner data, Powell and Leider (2020) report a 59 percent pass-through rate and a 22 percent decline in sales with no cross-border shopping in the Seattle, WA, event. Leider, Pipito and Powell (2018) and Powell, Leider and Léger (2020) report an overshifting of the tax, with 114 percent and 119 percent pass-through rates in Cook County, IL. Using hand-collected data, Cawley et al. (2021) report a 79 percent pass-through rate in stores and a 69 percent pass-through rate in restaurants from the Boulder, CO, soda tax event; Cawley et al. (2020*b*) report a 60 percent pass-through rate but no significant change in consumption from the Oakland, CA, soda tax event; and Falbe et al. (2020) report a 92 percent pass-through rate the from Oakland, CA, and full pass-through (100 percent) from the San Francisco, CA, soda tax events. Using secondary household-level receipt data and four event studies, Cawley, Frisvold and Jones (2020) report an average 12 percent decline in consumption where the majority of reduction comes from the Philadelphia, PA, soda tax event.

In addition to this purely applied work, two influential studies have tackled two important questions: (Allcott, Lockwood and Taubinsky, 2019*a*) investigate the optimal rate of soda taxes, while (Dubois, Griffith and O'Connell, 2020) assess who is most affected by these taxes. Allcott, Lockwood and Taubinsky (2019*a*) estimate an optimal nationwide tax on sugar-sweetened beverages using Nielsen household-level scanner data and report an optimal nationwide tax of 1 to 2.1 cents per ounce and optimal local city taxes of as low as 0.4 cents per ounce after considering cross-border shopping behavior. Dubois, Griffith and O'Connell (2020) specify a structural model of demand assuming a 100 percent pass-through rate to prices. They find that soda taxes are effective at targeting the young but less effective for those with sugar-intensive diets and those with lower incomes. The evidence on optimal tax rates and targeted populations provides a benchmark for effective tax designs and thus helps with the interpretation of empirical evidence from case studies of actual tax events.

Another group of papers studies the effects of soda taxes at the country level, including France (Berardi et al., 2016), Denmark (Bergman and Hansen, 2010), Mexico (Aguilar, Gutierrez and Seira, 2016; Colchero et al., 2017), and Portugal (Gonçalves and Dos Santos, 2020). These studies report pass-through rates of above 100 percent, accompanied by significant reductions in the consumption of the taxed products. However, country-level studies lack comparison groups with stores unaffected by the tax, and the evidence typically comes from a before-and-after analysis in the taxed country.[13]

3.3 Data and sample construction

3.3.1 Institutional context

Table 3.1 reports soda tax events in eight U.S. localities between 2015 and 2018. On January 1, 2017, a 1.5 cent per ounce beverage tax went into effect in the City of Philadelphia. This was the second city-level beverage tax in the United States, but it was the first excise soda tax implemented in a large U.S. city. Excise taxes are levied on manufacturers and expected to be passed through to retail prices. Some evidence shows that excise taxes are more effective than sales taxes e.g. Brownell et al., 2009, perhaps because consumers are more likely to react to shelf price than to a sales tax paid at the cashier, which is often a combined tax, making it hard to discern the portion that comes from soda alone.

[13]A smaller number of studies (e.g., Berardi et al., 2016, in France; Gonçalves and Dos Santos, 2020, in Portugal) use untaxed product categories such as water as the control.

Although the tax standard is 1 cent per ounce, Philadelphia, Seattle, and Boulder imposed taxes of 1.50 cents per ounce, 1.75 cents per ounce, and 2 cents per ounce, respectively. In the City of Philadelphia and Cook County, the tax is imposed not just on sugar-sweetened beverages (e.g., regular soda) but also on diet products (e.g., diet soda), but the county board in Cook County repealed the tax after only four months. This leaves the Philadelphia tax event alone with the largest universe of taxed products and the additional distinction of a tax rate 50 percent higher than the 1-cent-per-ounce standard.

Table 3.1. Local sweetened beverages taxes in the United States

Location	Tax (¢/*ounce*)	Date Approved	Date Imposed	Coverage
Berkeley, CA	1¢	November 2014	March 1, 2015	SSB
Philadelphia, PA	1.5¢	June 2016	January 1, 2017	SSB + Diet
Cook County, IL	1¢	November 2016	August 2, 2017	SSB + Diet
Albany, CA	1¢	November 2016	April 1, 2017	SSB
Boulder, CO	2¢	November 2016	July 1, 2017	SSB
Oakland, CA	1¢	November 2016	July 1, 2017	SSB
San Francisco, CA	1¢	November 2016	January 1, 2018	SSB
Seattle, WA	1.75¢	June 5, 2017	January 1, 2018	SSB

Notes. Sugar-sweetened beverages (SSB) include regular soda, sports drinks, energy drinks, sugary juice (less than 50 percent juice), presweetened coffee and tea, presweetened milk, flavored water, and mixers for alcoholic drinks. Diet refers to artificially sweetened, noncaloric beverages, such as diet soda. Cook County, IL, includes Chicago. The tax was repealed by the Cook County Board of Commissioners, effective December 1, 2017. The other seven taxes were still in effect when this paper was written.

An active debate on the implementation and effectiveness of local soda taxes is manifested in extensive media coverage and ongoing lawsuits between local authorities and beverage manufacturers and associations. Tax advocates unanimously argue that the sweetened beverages tax—just like any other "sin" tax (e.g., tobacco, alcohol, and fat (margarine) taxes)—would decrease consumption, improve diet quality, and decrease medical costs while generating tax revenue that might potentially be used to improve public health (e.g., by opening parks and recreation centers). By contrast, soda manufacturers have spent billions of dollars to fight against soda taxes through a series of lawsuits, funded research, and lobbying; they argue that there are no significant health gains from soda taxes, only losses in terms of job losses and price hikes. In the midst of the continuing debate, it is important to provide credible results and inform policy makers and consumer advocacy groups on whether and the extent to which these taxes work.

3.3.2 Data sources and sample construction

I use Nielsen Homescan (also referred as Home Measurement Service, or HMS) data to compose a detailed dataset of weekly barcoded (at the Universal Product Code [UPC] level) beverage purchases across all stores (in and outside the City of Philadelphia) at which households shop. The final sample of price and volume of beverages is kept at household-store-UPC-week level from 2006 to 2017. Next, I complement household-level scanner purchases with barcoded nutrition (calories, sugar, and ingredients list) collected by Gladson. Last, I utilize MABLE GeoCorr geographic correspondence data to identify affected and unaffected localities at five-digit zip code levels and calculate geodesic distances between any pair of two localities.

3.3.2.1 Household purchases

Nielsen Homescan tracks store purchases made by a panel of households. Section 2.4.1 describes this dataset in detail. For the benchmark DiD estimation, I use 2015–2017 purchases, covering the period from two years before to a year after the tax was implemented, and measure price and consumption responses based on this shorter panel. One reason for this time window is to minimize the attrition bias that might result from some panelists dropping out the sample. Another, related reason is to give more credibility to the parallel trends assumption between the affected and control regions. However, matrix completion and transpose elastic-net methods utilize the complete sample period from 2006 to 2017, as these methods let the data construct counterfactual Philadelphia and utilizing a larger pretreatment period for both treated and control regions is helpful for a better fit.

3.3.2.2 Barcoded ingredients

Gladson extracts barcode-level ingredients lists from the nutrition facts labels of food products in the United States. Along with the city guidelines for the taxed beverages, I utilize Gladson's ingredients lists to determine whether a beverage is taxed. When a UPC in the scanner data does not have a match in the Gladson data, I use the ingredients of the most similar UPC in the pool of matched UPCs to proxy ingredients for the remaining unmatched UPCs. I describe Gladson database and the proxying procedure for unmatched UPCs in more detail in 2.4.1.

3.3.2.3 Geographic mapping and geodesic distances

I use the GeoCorr database first to identify zip code areas in each tax locality and second to calculate geodesic distances between pairs of localities. The database provides detailed geographic correspondence information between U.S. cities (or counties) and three- or five-digit zip code areas. GeoCorr also provides latitudes and longitudes for the zip code areas, which makes it possible to calculate geodesic distances between zip code areas in and outside of the tax jurisdiction. I use STATA's vincenty package to calculate geodesic distances between any two localities (five-digit zip code areas between Philadelphia and outside Philadelphia) on the surface of the earth using an accurate ellipsoidal model of the earth.

Table 3.2 reports five-digit zip code areas (column labeled N Five-digit zip code areas) and their residents (column labeled N Households), both in and up to 350 miles outside the city border of the City of Philadelphia, between 2016 and 2017, a year before and a year after the soda tax in Philadelphia, PA, went into effect. I focus only on five-digit zip code areas for which I have household data.

Table 3.2. The number of five-digit zip code areas and their residents by geodesic distance and year

	N Five-digit zip code areas		*N* Households	
	2016	2017	2016	2017
Philadelphia	43	43	236	220
4 miles outside Philadelphia	42	41	172	165
6 miles outside Philadelphia	69	68	260	256
10 miles outside Philadelphia	109	106	478	468
20 miles outside Philadelphia	187	185	807	796
50 miles outside Philadelphia	447	444	1,988	1,986
100 miles outside Philadelphia	1,277	1,263	5,410	5,447
200 miles outside Philadelphia	1,870	1,851	7,253	7,273
350 miles outside Philadelphia	2,456	2,436	9,350	9,376

Notes. I focus on five-digit zip code areas in six U.S. localities that are geographically proximate to the City of Philadelphia: five U.S. states (the rest of Pennsylvania, Delaware, Maryland, New Jersey, New York) and Washington, DC. I use STATA's vincenty package to calculate geodesic distances between any pair of localities using an ellipsoidal model of the earth. For example, "4 miles" refers to the region between a five-digit zip code area in the city border and all other zip code areas outside the city border up to 4 miles. 2016–2017 Nielsen Homescan; 2014 MABLE GeoCorr.

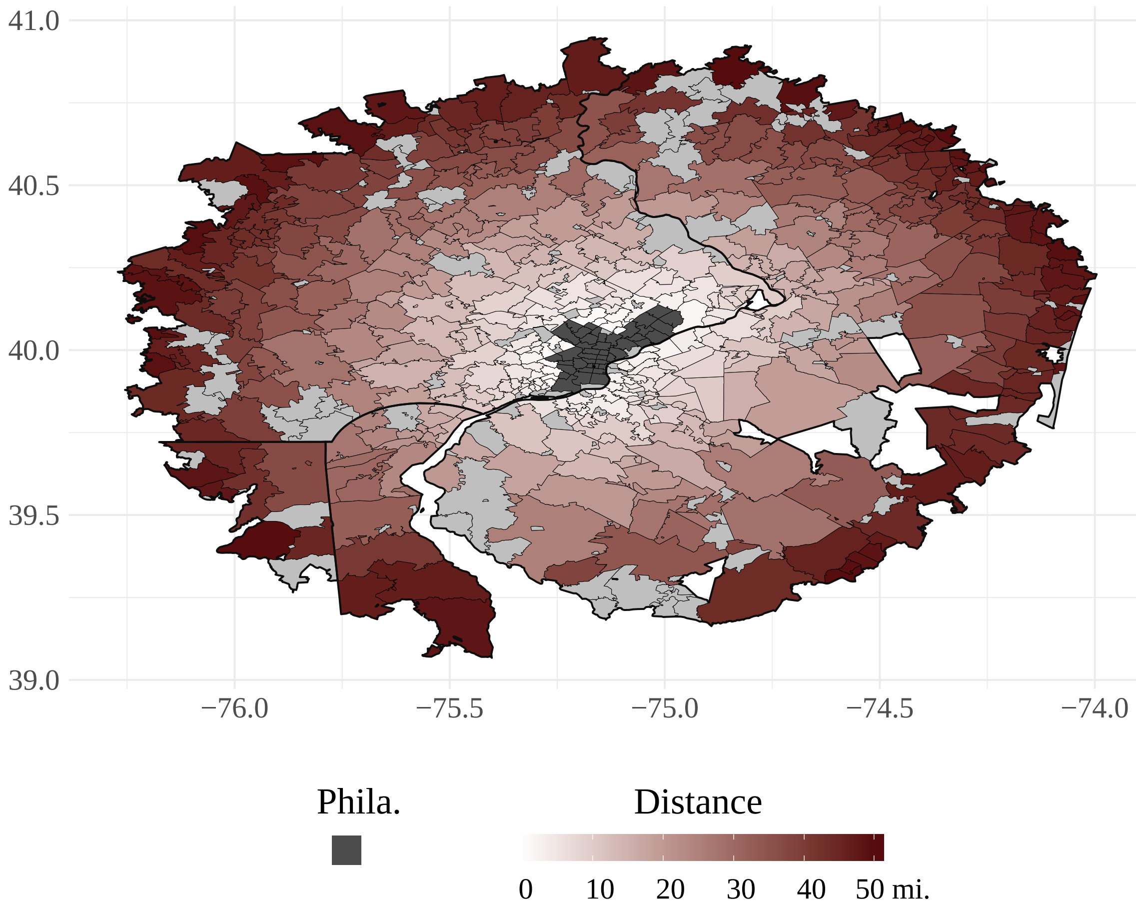
For control regions, I focus on five-digit zip code areas in six U.S. localities: five U.S. states (the rest of Pennsylvania, Delaware, Maryland, New Jersey, New York) and Washington, DC.[14] Figure 3.1 provides a map of five-digit zip codes in and outside of Philadelphia up to 50 miles, and Figure C.1 shows a similar map up to 350 miles.

In Sections 3.5.1.2 and 3.5.2, I provide evidence on estimated price and consumption effects of the tax when results come from alternative control regions (i.e., five-digit zip code areas in unaffected tax regions) depending on how far they are outside the city border. Specifically, I assign control regions depending on their

[14] Appendix Table C.1 reports demographic in the City of Philadelphia and these six surrounding localities by their distance from the city border.

distance from the city border in increments of 10 miles, as shown in Figures 3.3–3.5 and Table C.4.[15]

Figure 3.1. Map of the City of Philadelphia and its surrounding localities by 50-miles distance



Notes. I focus on five-digit zip code areas in six U.S. localities that are geographically proximate to the City of Philadelphia: five U.S. states (the rest of Pennsylvania, New York, Delaware, Maryland, New Jersey) and Washington, DC. I use STATA's vincenty package to calculate geodesic distances between any pair of localities using an ellipsoidal model of the earth. There are no household scanner data for areas in light gray. 2006–2017 Nielsen Homescan; 2014 MABLE GeoCorr.

[15]In these regressions, I treat the region corresponding to 0–6 miles outside the city border in two alternative ways: I either exclude it from the sample and report estimates using Philadelphia households' Philadelphia-exclusive beverage purchases alone, as in Figure 3.3, (i.e., an artificial doughnut hole) or I include them as part of the affected region (i.e., Philadelphia households' all beverage purchases) and account for the effect of cross-border shopping by Philadelphia households when assessing consumption effects of the tax, as in Figure 3.5.

3.4 Research design

3.4.1 Difference-in-differences

I specify and estimate price and consumption regressions separately for taxed and untaxed products and for each month. Price regressions in the benchmark DiD model are of the following form:

$$p_{jst} = \alpha(TreatedLocality_s \times Effective_t) + \delta_{s(z)t} + \gamma_{jt} + \epsilon_{jst}, \tag{3.1}$$

where p_{jst} is the per ounce price of UPC j in store s and month t. The coefficient of interest, α, is the DiD coefficient and represents the price effect of the soda tax event in Philadelphia, PA, which is indicated by the interaction between binary $TreatedLocality_s$ (takes a value of 1 if a store is in the City of Philadelphia and a value of 0 otherwise) and binary $Effective_t$ (takes a value of 1 during the effective period of the tax, i.e., for all time periods after January 2017 for the Philadelphia tax). I include two sets of fixed effects: Store's 3-digit zip code-by-month fixed effects ($\delta_{s(z)t}$) purge time-varying regional confounders such as seasonal product availability shocks in a specific area. UPC-by-month fixed effects (γ_{jt}) purge time-varying confounders for each product (at the UPC level). The regression error is denoted by ϵ_{jst}.

I specify consumption regressions as in Cawley, Frisvold and Jones (2020):

$$q_{ht} = \alpha(TreatedLocality_h \times Effective_t) + \delta_h + \gamma_t + \epsilon_{ht}, \tag{3.2}$$

where q_{ht} is the log of all volume purchases by household h in month t. I calculate (approximate) percentage changes in volume purchases using inverse hyperbolic sine transformation, as in Clemens and Hunt (2017), rather than adding 1 to the variable prior to its logarithmic transformation (i.e., $\log[variable + 1]$). Both transformations retain zero purchases in the balanced panel, but differences between two logs of smaller and larger values are more comparable and hence the results are more accurate when inverse hyperbolic sine transformation is adopted. The coefficient of interest, α, is the DiD coefficient and represents the consumption effect of the soda tax event in Philadelphia, PA, which is indicated by the interaction between binary $TreatedLocality_h$ (takes a value of 1 if household h is a resident in the City of Philadelphia and a value of 0 otherwise) and binary $Effective_t$ with identical interpretation as in equation 3.1. I include household fixed effects (δ_h) to account for systematic differences across households, while month fixed effects (γ_t) control for macroeconomic shocks. The regression error is denoted by ϵ_{ht}.

To construct counterfactual Philadelphia, where the tax policy was never enacted, I initially focus my sample on six U.S. localities in close proximity to the City of Philadelphia (i.e., the rest of Pennsylvania, Delaware, Maryland, New Jersey, New York, and Washington, DC) and exclude all other U.S. localities. The underlying assumption is that geographically proximate areas are more likely to follow similar economic trends and business cycles. Thus, changes that might occur after the policy can be attributed to the policy rather than to some other shock.

Next, as robustness checks, I impose a more restrictive criteria in selecting control areas: I choose only those that are less than certain miles away from Philadelphia (e.g., 10 miles away, 20 miles away, and so on). To avoid spillover effects contaminating my estimates, I exclude all 5-digit zip code areas that are within 6 miles of the treated region, as in Seiler, Tuchman and Yao (2021), and create an artificial doughnut hole. Unlike Seiler, Tuchman and Yao (2021), however, I run regressions from alternative control samples: The doughnut gets bigger as control stores come from farther away surroundings (6–10 miles away, 6–20 miles away, etc.), though the doughnut hole remains the same (0–6 miles away from the city border).

3.4.2 Machine learning

In the literature, researchers often rely on DiD methods to estimate the effects of the tax. However, tax localities might be different from control regions in a systematic and time-varying way. Therefore, in addition to the benchmark models in equations 3.1 and 3.2, I employ more advanced machine learning methods and let the machine pick the trends in data and capture the causal effect of the tax policy.

In general, causal machine learning methods rely on data to construct counterfactual treated regions (or units) and impute missing potential control outcomes for treated regions to estimate the causal average treatment effect on the treated, or ATT. I use two of these methods—matrix completion and transposed elastic-net—as in Athey et al. (2021).[16] The primary advantage of this methodology over the DiD benchmark is that these methods rely on data to purge unobserved confounders. Thus, individually, each method promises to provide credible estimates of tax effects as they likely produce an accurate representation of the counterfactual tax locality.

For machine learning exercises, I use a larger sample in both the pretreatment period (2006–2016 vs. 2015–2016) and in the control regions (all U.S. vs. six U.S. localities). An advantage of using a larger sample is the possibility that data will capture a trend between treated and control regions that might not be as possible in a smaller sample. Yet another advantage would be more useful potential variation in data and hence more statistical power in findings. However, a larger sample may also be more likely to capture noise. In Section 3.5.1.2, I show that ATT findings across

[16] Other alternative causal machine learning tools include synthetic control by Abadie, Diamond and Hainmueller (2010) and generalized synthetic control by Xu (2017). While some studies in soda tax literature use synthetic control when constructing counterfactual treatment regions—such as the counterfactual Berkeley, CA, in Rojas and Wang (2021)—other strands of the applied microeconomics use generalized synthetic control methodology (e.g., Cengiz and Tekgüç, 2022, on the effects of Syrian immigration).

machine learning and DiD models are quite similar, which suggests that capturing noise is not likely in my analysis.

In particular, matrix completion uses singular value decomposition to construct counterfactuals using observed values (both pre- and posttreatment) in control regions and pretreatment values in the treated region. Intuitively, it tries to reconstruct a matrix where no region experienced a tax. Therefore, in the Philadelphia soda tax event, it considers Philadelphia from 2017 onward to be missing and tries to populate the matrix. By contrast, transposed elastic-net, or horizontal regression as it referred in Athey et al. (2021), uses pretreatment observations of the control regions to match posttreatment observations. In this sense, it is somewhat similar to standard time-series models, but it uses the penalty factor to determine which lags are important.

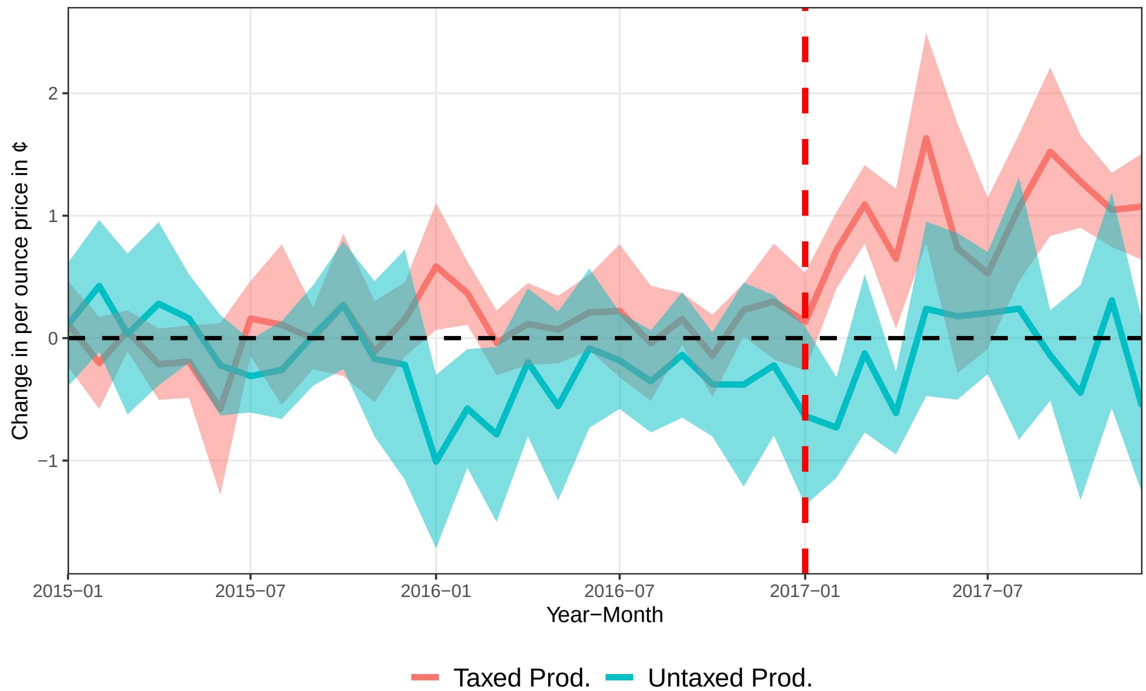
3.5 Results

3.5.1 Main results

3.5.1.1 Pass-through rate

Figure 3.2 plots the estimated monthly price effects of the tax between 2015 and 2017 for taxed and untaxed products separately using equation 3.1. The solid red line shows the change in price in taxed stores compared to the change in price in untaxed stores (i.e., those that are 6–346 miles away from the city border) for taxed products. The solid blue line reports the estimated change for untaxed products. Shaded areas correspond to the 95 percent confidence intervals: red for taxed, blue for untaxed products. The treatment is shown by the vertical dashed red line, and it is in effect since January 1, 2017, when the soda tax in Philadelphia (1.5 cents per ounce) was enacted.

Figure 3.2. Price effects of the tax on taxed and untaxed beverage products in Philadelphia compared to counterfactual Philadelphia over time



Notes. The figure shows the estimated monthly price effects of the tax using equation 3.1 separately on taxed (solid red line) and untaxed (solid blue line) products. Shaded areas correspond to the 95 percent confidence intervals. Taxed location is the 3-digit 191 zip code (City of Philadelphia). Untaxed comparison regions are other 3-digit zip codes in Pennsylvania, and those in New York, Delaware, Maryland, New Jersey, and Washington DC, except areas in 0–6 miles outside the city border (i.e., an artificial doughnut hole). Taxed products contain soda (regular and diet), energy drinks, sports drinks, sweetened fruit juice, sweetened water, sweetened coffee/tea, and sweetened milk. Untaxed products are pure water, 100 percent juice, unsweetened milk, and unsweetened coffee/tea. The Philadelphia soda tax went into effect on January 1, 2017. 2015–2017 Nielsen Homescan.

The figure shows that the solid red line of monthly estimates of taxed products always lies above 0 in the posttreatment period, while the solid blue line of untaxed products lies around 0, suggesting that taxed products alone experienced an increase in price. Tables C.3 and C.4 show regression results for the graphical illustration in Figure 3.2. I report monthly point estimates and 95 percent confidence intervals before and after the tax was enacted for the taxed (Table C.3) and untaxed products (Table C.4). All of the findings from monthly regressions in Table C.3 are statistically distinguishable from 0 at the 1 percent significance level, except January 2017, where the effect is insignificant.[17]

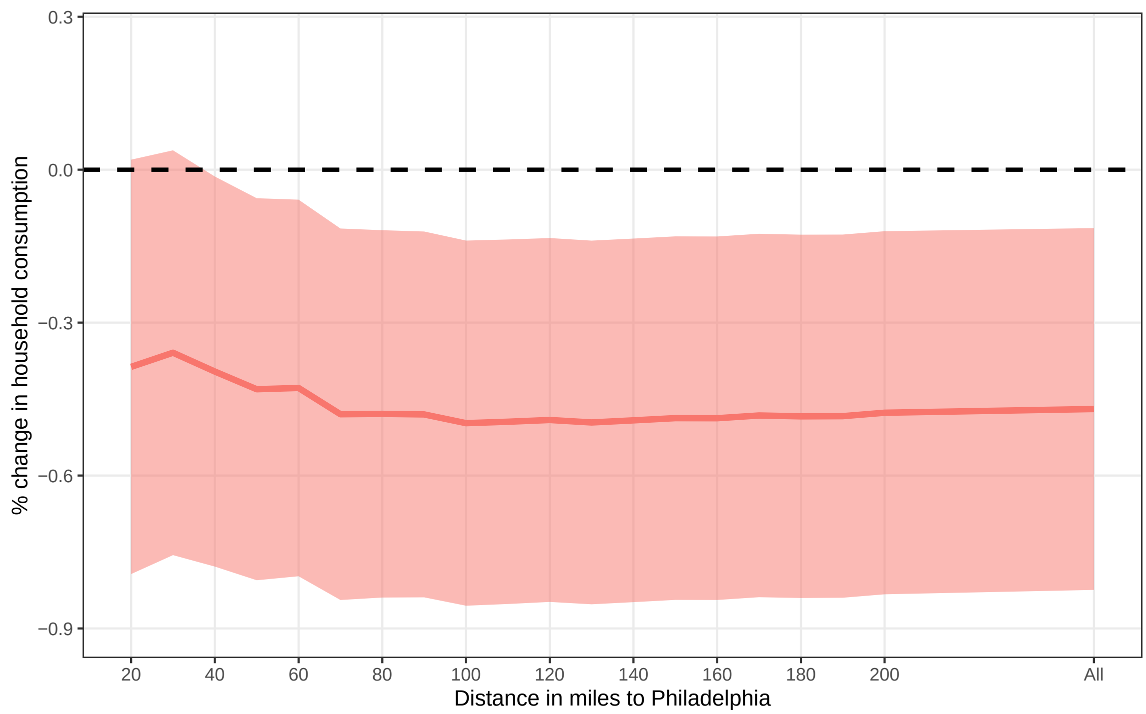
By taking the average of the monthly price estimates, I calculate the average price estimate for the posttreatment period as 0.96 cents per ounce. This translates into 64 percent pass-through rate (¢0.956/¢1.5 = 0.637) and a 12.3 percent increase in price (¢0.956/¢7.8 = 0.123), where 7.8 cents per ounce is the average pretax price for taxed products.

3.5.1.2 Consumption

Figure 3.3 plots estimated consumption effects of the soda tax on taxed products. In these regressions, I estimate equation 3.2 using alternative control region samples of a range of 6 to certain miles away from the city border. For example, the 20-mile distance threshold only includes stores from 6 miles to 20 miles away from the Philadelphia border as the control region (i.e., counterfactual Philadelphia). The red line shows the percentage change in consumption in the taxed regions compared to the untaxed regions for taxed products. The shaded area represents the 95 percent confidence interval.

[17] There is no statistically significant difference between Philadelphia and control regions in the first month after the tax was enacted. As noted in (Seiler, Tuchman and Yao, 2021), not all retail stores adjusted prices immediately after the tax went into effect but rather adjusted prices with a lag or gradually. The authors label the first four months after the tax was enacted as the adjustment period and omit this period from their main analysis. While I keep this period in my analysis and report regression estimates per each month, I note that the insignificant price difference between Philadelphia, PA, and its controls is suggestive of an adjustment period. One rationale for this adjustment period is the difficulty in following city guidelines for the taxed products: Some retail stores weren't immediately sure which products were subject to tax, and there have been cases in which they taxed products that shouldn't be taxed and failed to tax some products that should have been taxed (Orso, 2017).

Figure 3.3. Consumption effects on taxed beverages in Philadelphia compared to counterfactual Philadelphia by distance excluding cross-border shopping

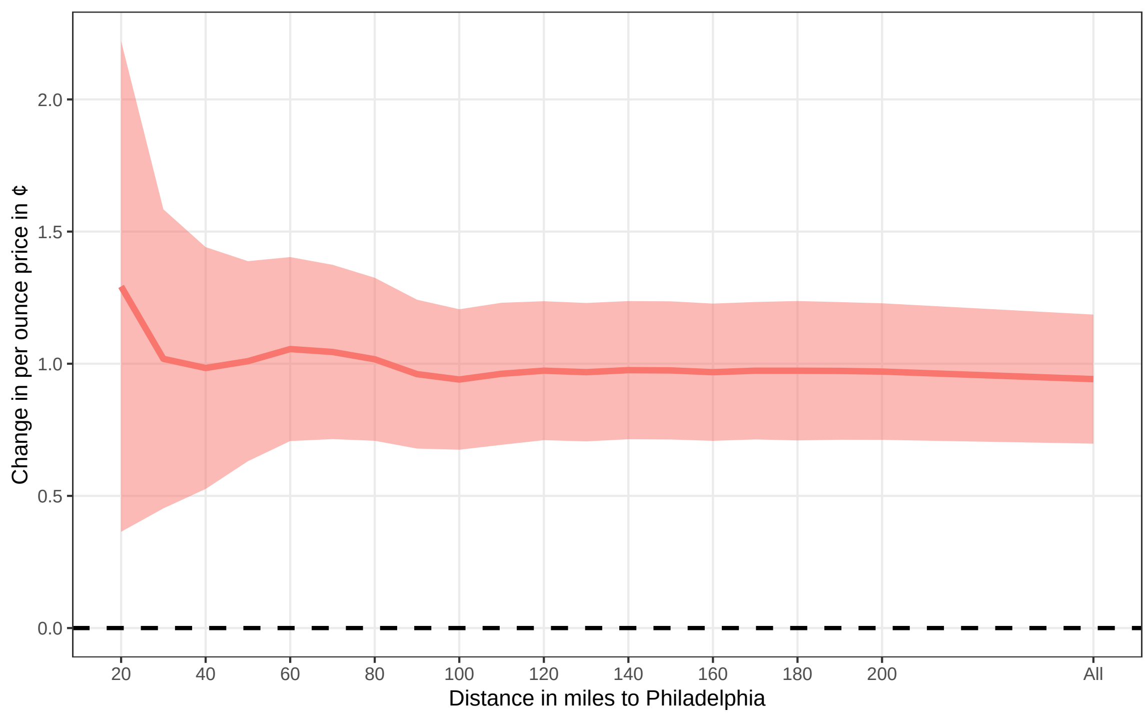


Notes. The figure shows the estimated consumption effects of the tax on taxed products using equation 3.2 for alternative control samples constructed using the distance in miles to Philadelphia criterion. The red line shows the percentage change in consumption in taxed regions compared to untaxed regions for taxed products. The x-axis shows the distance criterion in increments of 10 miles up to 200 miles and then up to 346 miles (labelled as "All"). I keep Philadelphia households' Philadelphia-exclusive purchases alone, therefore excluding cross-border shopping. The shaded area corresponds to the 95 percent confidence interval. The taxed location is the 3-digit 191 zip code (City of Philadelphia). Untaxed comparison regions are other 3-digit zip codes in Pennsylvania and those in New York, Delaware, Maryland, New Jersey, and Washington, DC, except areas 0–6 miles outside the city border (i.e., an artificial doughnut hole). Taxed products contain soda (regular and diet), energy drinks, sports drinks, sweetened fruit juice, sweetened water, sweetened coffee/tea, and sweetened milk. The Philadelphia soda tax went into effect on January 1, 2017. 2015–2017 Nielsen Homescan; 2014 MABLE GeoCorr.

3.5.2 Sensitivity analyses

Figure 3.4 shows the estimated price effects of the tax on taxed products using equation 3.1 for alternative control samples constructed using the geodesic distance in miles to Philadelphia criterion. The red line shows the change in price in the taxed regions compared to the untaxed regions for taxed products. I find that the change in price is 69 percent of the tax when I construct control groups using this distance criterion. Table C.5 reports the illustrated regression results in table format.

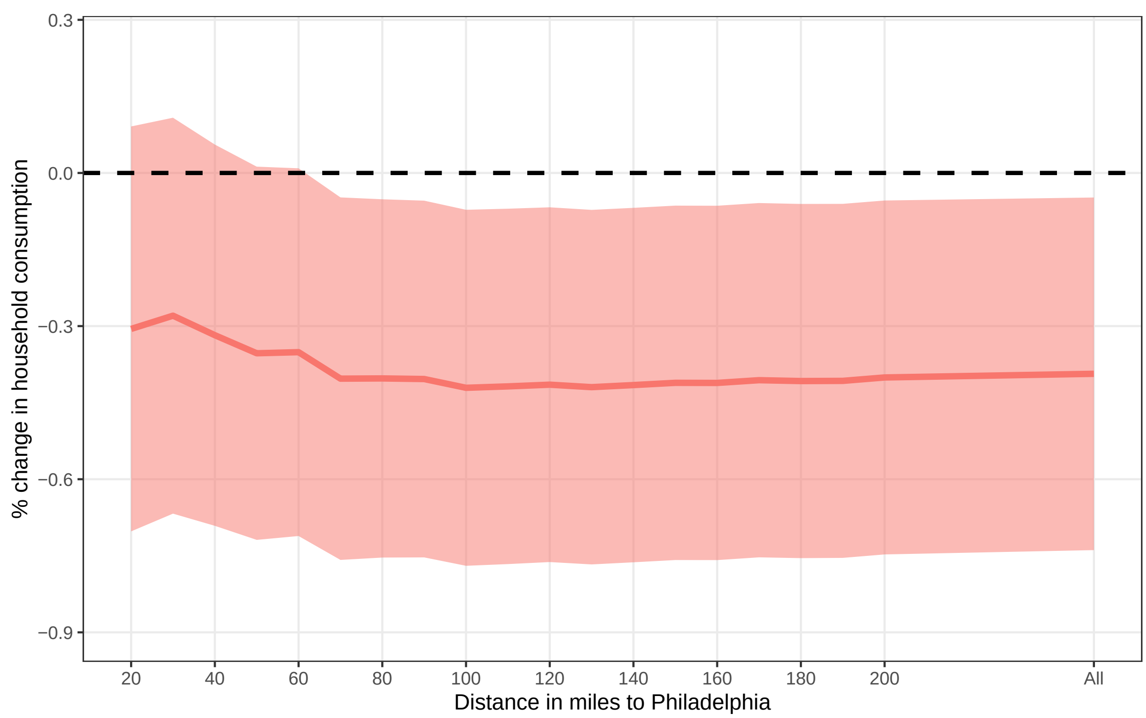
Figure 3.4. Price effects on taxed beverages in Philadelphia compared to counterfactual Philadelphia by distance



Notes. The figure shows the estimated price effects of the tax on taxed products using equation 3.1 for alternative control samples constructed using the distance in miles to Philadelphia criterion. The red line shows the change in price in taxed regions compared to untaxed regions for taxed products. The shaded area corresponds to the 95 percent confidence interval. The taxed location is the 3-digit 191 zip code (City of Philadelphia). Untaxed comparison regions are other 3-digit zip codes in Pennsylvania and those in New York, Delaware, Maryland, New Jersey, and Washington, DC. Taxed products contain soda (regular and diet), energy drinks, sports drinks, sweetened fruit juice, sweetened water, sweetened coffee/tea, and sweetened milk. Philadelphia soda tax went into effect on January 1, 2017. 2015–2017 Nielsen Homescan; 2014 MABLE GeoCorr.

In Figure 3.3, I reported consumption estimates coming from Philadelphia households' Philadelphia-exclusive purchases alone, excluding cross-border shopping in the neighboring untaxed regions. In contrast, Figure 3.5 reports consumption estimates from all purchases made by Philadelphia households, including their cross-border shopping. Following Seiler, Tuchman and Yao (2021), I define the cross-border shopping zone as up to 6 miles outside the City of Philadelphia. Once I account for households' cross-border shopping behavior, the estimated decline in consumption becomes less pronounced. This shows that affected households engage in cross-border shopping as a way of avoiding the tax.

Figure 3.5. Consumption effects on taxed beverages in Philadelphia compared to counterfactual Philadelphia including cross-border shopping



Notes. The figure shows the estimated consumption effects of the tax on taxed products using equation 3.2 for alternative control samples constructed using the distance in miles to Philadelphia criterion. The red line shows the percentage change in consumption in taxed regions compared to untaxed regions for taxed products. In these regressions, I keep all purchases made by Philadelphia households, including cross-border shopping. The taxed location is the 3-digit 191 zip code (City of Philadelphia). Untaxed comparison regions are other 3-digit zip codes in Pennsylvania and those in New York, Delaware, Maryland, New Jersey, and Washington, DC. Taxed products contain soda (regular and diet), energy drinks, sports drinks, sweetened fruit juice, sweetened water, sweetened coffee/tea, and sweetened milk. 2015–2017 Nielsen Homescan; 2014 MABLE GeoCorr.

Table 3.3 reports consumption estimates (in percentage changes) across alternative estimators. All three models point towards a decline of about 20 percent to 40 percent. While DiD and matrix completion models largely agree on the magnitude of the decline, the transposed elastic-net suggests a smaller decline. Taken together, the estimated price and quantity responses report a price elasticity of demand of -3.17 ($-0.389/0.123 = -3.17$) using the benchmark DiD estimator, where 0.123 is the price increase calculated in Section 3.5.1.1. The elasticity is -1.73 ($-0.212/0.123 = -1.73$) with transposed elastic-net and -3.22 ($-0.394/0.123 = -3.22$) with matrix completion. Once cross-border shopping behavior is accounted for, the estimated elasticity for the benchmark model becomes -2.49 ($-0.305/0.123 = -2.49$).

Table 3.3. Estimated log change in consumption

Difference-in-differences	Matrix completion	Transposed elastic-net
−0.388831	−0.39416	−0.2121545

Notes. The table reports the average estimated consumption effect of the tax on taxed products across alternative methods. The taxed location is the 3-digit 191 zip code area (City of Philadelphia). Untaxed comparison regions are other 3-digit zip codes in Pennsylvania and those in New York, Delaware, Maryland, New Jersey, and Washington, DC, for the difference-in-differences model and all U.S. for matrix completion and transposed elastic-net. Taxed products contain soda (regular and diet), energy drinks, sports drinks, sweetened fruit juice, sweetened water, sweetened coffee/tea, and sweetened milk. Difference-in-differences model uses 2015–2016 period as pretreatment sample, while machine learning models use 2006–2016 period. The Philadelphia soda tax went into effect on January 1, 2017. 2006–2017 Nielsen Homescan.

3.6 Conclusion

In 2020, the soft drinks industry was one of the leading consumer packaged goods industries in the United States, with $270 billion in sales; more than 40 percent of U.S. adults are obese.[18] Despite the high stakes for beverage manufacturers, mounting evidence empirically shows that taxing soda might be an effective method for combating obesity and other preventable diseases. This study adds to this literature by providing robust evidence from the Philadelphia soda tax event.

Using a highly granular sample and alternative estimators (difference-in-differences and advanced causal machine learning tools), I report that firms increased price by 12.3 percent (meaning that 64 percent of the tax was passed to consumers) and households cut soda consumption by 20 percent to 40 percent in response to the Philadelphia soda tax. Evidence also shows that consumers took their business to stores in nearby untaxed localities, but they didn't switch to untaxed beverages.

As recognized in the literature, cross-border shopping is one way to avoid the tax, and switching to untaxed but unhealthy products (e.g., artificially sweetened products) that are associated with adverse health outcomes is another (Dubois, Griffith and O'Connell, 2020; Powell et al., 2021; Seiler, Tuchman and Yao, 2021). The Philadelphia soda tax event is unique in taxing both diet and regular soda and hence limiting tax avoidance in product space. The literature has also recognized two other tax avoidance mechanisms: households' stockpiling behavior of taxed beverages during sale periods (Wang, 2015; Wang, Rojas and Colantuoni, 2017), and, at a more diet quality level, switching to sweetened nonbeverage products such as ice cream or candy (Allcott, Lockwood and Taubinsky, 2019*a*) or to other sin products like tobacco or alcoholic beverages. A fuller assessment of these channels is left to future research.

[18]See Globe Newswire for 2020 soft drinks revenue in the United States; see CDC (2020) for the U.S. obesity prevalence data.

APPENDIX A

APPENDIX OF CHAPTER 1

A.1 Choice of products

Nielsen records regular and light beer sales in two different product modules. I start cleaning by generating non-overlapping lists of light and non-light brands at the universal product code (UPC) level. Next, I convert all UPCs into ounces and manually fill in the missing data via online web search.

A.2 The matched sample between sales and advertising databases

The matching between the Nielsen Retail Scanner sales database and Ad$pender advertising database is not straightforward as the scales and descriptions of geographic units, time periods, and brands are different. Ad$pender spans the top 101 designated marketing areas (DMAs) while Nielsen covers as many as 206 DMAs. Discussions with Kantar Media representatives assured me that there is one-to-one DMA correspondence between the two databases, even though the descriptions of the same DMA are different. For example, Nielsen's "Boston and Manchester" DMA matches perfectly with the "Boston" DMA in Ad$pender.

A.3 Own advertising measures

The calculation of the total own advertising spending for each mass-produced lager beer (MPLB) brand in a given week, and DMA can be expressed in up to three aggregations. First, I take the sum of all traditional media across different traditional advertising channels (TV, radio, print, and outdoor) in order to represent the total traditional advertising spending as described in Section 1.3.3 and shown in Table 1.2.

The second summation is of national and local advertisements, which are distinguished from one another by the scale of viewership. More specifically, national advertising refers to brand advertising aired or distributed across the U.S., while local advertising refers to DMA-exclusive advertising. For example, the total Bud Light advertisements from traditional media in Atlanta in a given week is the sum of national Bud Light advertisements (those that are aired through national TV channels and radio stations as well as those that are distributed through national newspapers and magazines) and local Bud Light advertisements (those that are aired and distributed through Atlanta-specific channels, including outdoor displays). I calculate the contemporaneous advertising, $(\boldsymbol{a}_{jbd(s)w})$, for brand j in DMA $d(s)$ and week w as the drinking population normalized summation of total traditional advertising dollars from both national and local advertisements. More specifically, I calculate the national advertising in a DMA by dividing national advertising expenditures by the U.S. population of legal drinking age (age 21 and above). For the local advertising, I divide the DMA-specific advertising dollars by the legal drinking-age population in that DMA, where I calculate the DMA population by summing the population from the counties in that DMA. I then obtain the total by adding up the local and national and multiplying by 100 in order to represent the dollars spent for 100 people of drinking age: \$ per 100 adults.

A third summation occurs for specifications that proportionally add general advertisements onto direct advertisements. A general advertisement can be as broad as a parent company advertisement (e.g., an Anheuser-Busch advertisement that refers to all Anheuser-Busch brands regardless of beer type (light or non-light) in a given week and DMA). Conversely, a *combo* advertisement contains at least two brands, including the focal brand for which I estimate sales. An example would be a sponsored event advertisement for both Amstel Light and Heineken. To proportionally allocate these general advertisements among the included brands, I first assign the corresponding total volume sales aggregated from all stores in that DMA and week and then calculate the volume shares for each brand.

A.4 Rival advertising, own price, and rival price

In a given store and week, I calculate rival advertising for brand j as the summation of advertising by all the other brands within the same beer category as j. The rival advertising from a different category for a given brand-store-week combination (e.g., the summation of advertising from all present non-light brands for any light beer brand) repeats the same rival advertising value without producing any variation that would be required to identify the different-category rival advertising effect.

I calculate own price as the volume-weighted average price of all UPCs purchased in store-week pairs. Compared to revenue weighted average price, a volume-weighted average is a better measure of brand effects because it captures consumer preferences more fully where the frequency of purchases are accounted for. For zero-sales, however, I impute prices for UPCs in store-week pairs using the largest mode in the corresponding six-month period, consistent with the assumption that price would likely to be higher in the absence of sales, as noted by Hitsch, Hortacsu and Lin (2019). There are approximately 1,500,000 UPC-store pairs in the 2011\textendash2015 MPLB data, where 7.7 percent have sales throughout the 261 weeks. However, for 99.3 percent of the cases, the pair appears at least once in each of the five years over the 2011–2015 period. Importantly, 69 percent of the pairs have sales in at least half of the available period, which requires the other half to be imputed—this could be as low as 26 weeks or as high as 181 weeks depending on the timeline that the pair is available. Only 2.5 percent of pairs have the extreme case of not having enough observations for at least one six-month period (e.g., 1 observation per year, 2 or 3 observations for every 2 years, 3–5 observations for every 3 years). In such cases, the prices are imputed using the annual mode price in the corresponding year. Therefore, for the set of zero sales, the brand price is in fact a simple average of per ounce price, implying that each unit is sold once to make the brand-level UPC aggregation feasible.

I set the rival price for brand j as the average of shelf price of all the remaining brands in the same category in a given store and week pair. As for advertising, computing different-category rival price is not helpful for identification. Furthermore, I compute a simple average shelf price rather than a volume-weighted average price as in the own price calculation for two main reasons. First, the volume-weighted average rival price is the same for all zero-sales products, producing no useful variation in the data. Also, the rival price would be missing if a brand had rivals with no sales. Second, the shelf price is the price that people actually see when they shop in a store. Hence, beer consumers would consider the shelf price of rival brands when making a purchase rather than a weighted average price that accounts for how many times each brand is actually sold.

A.5 Choice of the functional form of variables that enter the regression

Variables enter the triple difference (TD) regression models linearly in order to account for the high prevalence of zero sales and zero advertisements appropriately. To the best of my knowledge, the empirical literature addresses the presence of zero sales and/or advertising in three ways. The first way by removing them from the regression sample, either explicitly or implicitly through the use of natural log transformations of the variables. A second popular way is to apply an approximate natural log transformation where the values of variables are increased by a constant, most commonly incased by 1, to make the log structure feasible without dropping the zero values. A third way takes the inverse hyperbolic sine of the variables. As noted in Clemens and Hunt (2017, p.16),

> regression coefficients on variables transformed with the inverse hyperbolic sine can be interpreted identically to those using the traditional log transformation (as approximating percentage changes). But unlike the log transformation, the inverse hyperbolic sine has desirable properties near zero and is defined at zero (asinh $0 = 0$).

I run simulations in which I calculate the own advertising coefficients from three approximate log transformations using a naïve model that regresses sales on advertising. First, I adopt an approximate log transformation where the values of variables (advertising and sales) are increased by 1 (e.g., $log(1 + advertising)$). Second, the values are rather increased by the sample minimum (e.g., $log(minimum\ advertising + advertising)$), and third I use an inverse hyperbolic sine transformation (e.g., $asinh(advertising)$). I find drastically different results depending on the which transformation is used.

The main purpose of this exercise is to show that the choice of an additive constant is arbitrary and therefore can bias the estimates in unforeseeable ways. Furthermore, the use of an inverse hyperbolic sine transformation is limited when there are many values, in absolute, smaller than 1, which poses an additional restriction in the current empirical setting with population normalized values of advertising. Therefore, I use levels of variables instead of using approximate log transformations to avoid the likely arbitrary treatments of values when addressing zero sales and advertisements.

A.6 Standard errors

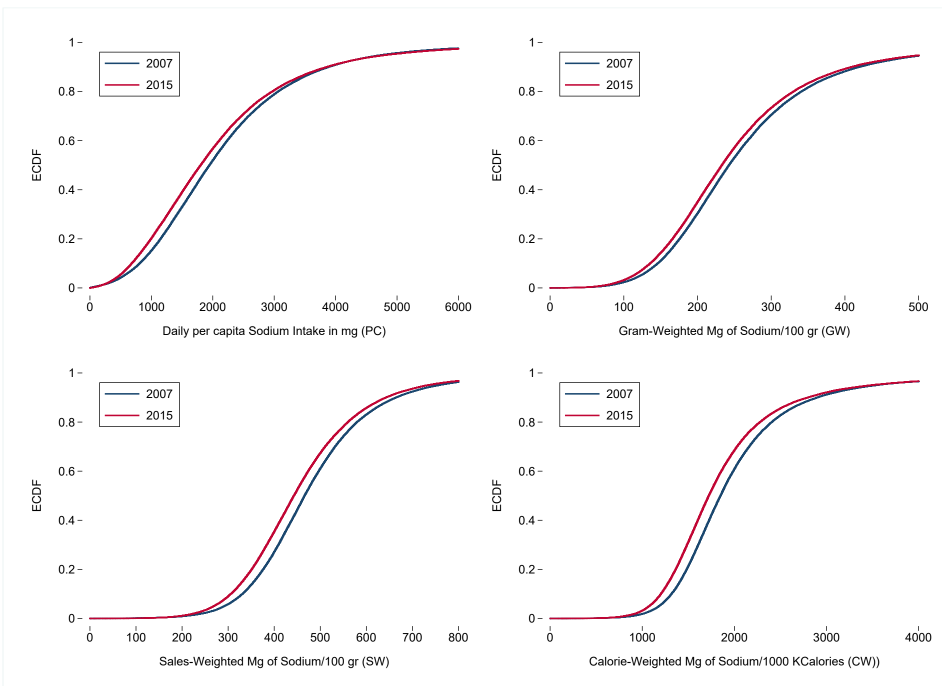
Following Dube, Lester and Reich (2010), I two-way cluster the standard errors on the DMA and border separately for each sample. The two-way clustering allows me to appropriately calculate the standard errors, which otherwise imply higher precision due to the repetitive observations in the bordering samples. Intuitively, the standard errors would already be biased downward with the constant treatment effect at each store within each DMA. To the extent that the set of stores included in a single county occurs repetitively in multiple pairs, there will be additional correlation across county-pairs (or grouped county-pairs as in the BR sample) mechanically induced by the research design, suggesting that residuals will no longer be independent. The two-way clustering of the standard errors accounts for these sources of bias while also correcting for heteroskedasticity.

APPENDIX B

APPENDIX OF CHAPTER 2

B.1 Additional tables and figures

Figure B.1. Empirical cumulative distribution of sodium intake measures (unbalanced panel)



Notes. Unbalanced panel of households (all households sample). 2007 and 2015 Nielsen Home Measurement System (HMS) and Gladson.

Table B.1. NSRI Food Groups, Categories, and Sodium Targets (mg/100 gr)

Food Group	Category	2009 Base	2012 Target	2014 Target
1. Bakery Products				
	1.1 Breads and rolls	485	440	360
	1.2 Sweet breads and rolls	295	270	220
	1.3 Tortillas and wraps	717	650	540
	1.4 Cakes, snack cakes, muffins, and toaster pastries	359	310	250
	1.5 Cookies	367	310	260
	1.6 Crackers	918	78	640
	1.7 French toast, pancakes, and waffles	569	510	430
2. Cereal and Other Grain Products				
	2.1 Instant hot cereal	562	480	390
	2.2 Breakfast cereals, light and medium weight	609	490	370
3. Meats				
	3.1 Cold cuts	1,085	980	810
	3.3 Cooked sausage	1,834	1,740	1,560
	3.4 Uncooked sausage	898	810	720
	3.5 Hot dogs	1,059	950	850
	3.6 Bacon	1,792	1,610	1,470
	3.7 Uncooked whole muscle meat and poultry	500*	450	400
	3.8 Canned meat and sausage	987	940	840
	3.9 Canned chicken and turkey	403	380	340

Food Group	Category	2009 Base	2012 Target	2014 Target
4. Dairy Products and Substitutes				
	4.1 Grated hard cheese	1,530	1,450	1,300
	4.2 Cheddar, Colby, Jack, Mozzarella, Muenster, provolone, and Swiss cheese	668	630	600
	4.3 Cream cheese	408	390	350
	4.4 Cottage cheese	347	330	290
	4.5 Processed cheese	1,393	1,250	1,040
5. Fats and Oils				
	5.1 Margarine and other spreads	718	650	570
	5.2 Salted butter	608	580	540
	5.3 Mayonnaise and mayonnaise-type dressing	713	640	570
	5.4 Salad dressing	1,019	920	760
6. Sauces, Dips, Gravies and Condiments				
	6.1 Major main entrée sauce	442	400	330
	6.2 Minor main entrée sauce	550	500	410
	6.3 Salsa, dips, and dipping sauce	712	640	530
	6.4 Barbecue sauce, ketchup, marinades, and steak sauce	1,081	1,000	860
	6.5 Asian-style condiments	706	640	560
7. Snacks				
	7.1 Flavored chips	711	570	430
	7.3 Puffed corn snacks	524	470	420
	7.4 Popcorn	969	820	680
	7.5 Pretzels and snack mixes	1,205	1,020	840

Food Group	Category	2009 Base	2012 Target	2014 Target
8. Soups				
	8.1 Canned soup	326	280	230
	8.2 Broth and stock	352	320	260
9. Potatoes				
	9.1 Frozen and refrigerated potatoes	349	300	240
	9.2 Seasoned processed potatoes	1,725	1,470	1,210
10. Mixed Dishes				
	10.1 Frozen entrees and sides $<$ 6 oz per serving	537	480	400
	10.2 Frozen entrees and sides 6-10 oz per serving	337	300	250
	10.3 Frozen entrees and sides $\geq$ 10 oz per serving	294	260	220
	10.4 Frozen and refrigerated pizza	558	500	450
	10.5 Refrigerated entrees and sides	531	480	400
	10.6 Canned chili, pasta, and hash	377	340	280
	10.7 Seasoned pasta and stuffing mixes	700	630	560
	10.8 Seasoned grain mixes	837	750	630
11. Vegetables				
	11.1 Frozen vegetables in sauce	343	310	260
	11.2 Canned vegetables	238	190	140
	11.3 Canned whole tomatoes	152	130	100
	11.4 Diced, crushed, and stewed tomatoes	289	250	190
	11.5 Vegetable juice	479	410	340

Food Group	Category	2009 Base	2012 Target	2014 Target
12. Legumes				
	12.1 Baked beans	408	370	310
	12.2 Canned beans	337	290	240
13. Canned Fish				
	13.1 Canned fish	374	350	330
14. Seasoning Mixes				
	14.1 Dry seasoning mixes	415	350	290
15. Nut Butters				
	15.1 Nut butters	457	410	340

Notes. NSRI = National Salt Reduction Initiative. Data come from NSRI. Base and target values are mg/100 g.

Table B.2. List of Signatory Firms for NSRI, HWCF and IFBA

Firm Name	**NSRI**	**HWCF**	**IFBA**
Black Bear European Style Deli	✓		
Boar's Head Provisions Co.	✓		
Bumble Bee Foods		✓	
Butterball	✓		
Campbell Soup Company	✓	✓	
ConAgra Foods		✓	
Danone			✓
Delhaize America (retailer)	✓		
Dietz & Watson	✓		
Ferrero,			✓
Furmano's	✓		
General Mills		✓	✓
Goya Foods	✓		
Grupo Bimbo			✓
Hain Celestial	✓		
Heinz	✓		
Hostess	✓		
Kellogg Company		✓	✓
Ken's Foods	✓		
Kraft Foods/Mondelez	✓	✓	✓
LiDestri Foods / Francesco Rinaldi	✓		
Mars Food US	✓	✓	✓
McCain Foods	✓		
McCormick & Company, Inc.		✓	

Firm Name	**NSRI**	**HWCF**	**IFBA**
McDonald's			✓
Nestle		✓	✓
PepsiCo		✓	✓
Post Foods/ Ralston Foods, LLC		✓	
Premio	✓		
Red Gold	✓		
Sara Lee Corporation		✓	
Snyder's-Lance, Inc.	✓		
Starbucks Coffee Company	✓		
The Coca-Cola Company		✓	✓
The Hershey Company		✓	
The J.M. Smucker Company		✓	
Unilever	✓	✓	✓
White Rose (distributor)	✓		

Notes. National Salt Reduction Initiative (NSRI), Healthy Weight Commitment Foundation (HWCF), and International Food and Beverage Alliance (IFBA).

Table B.3. Decomposition by UPC type and NSRI food group, CW measure

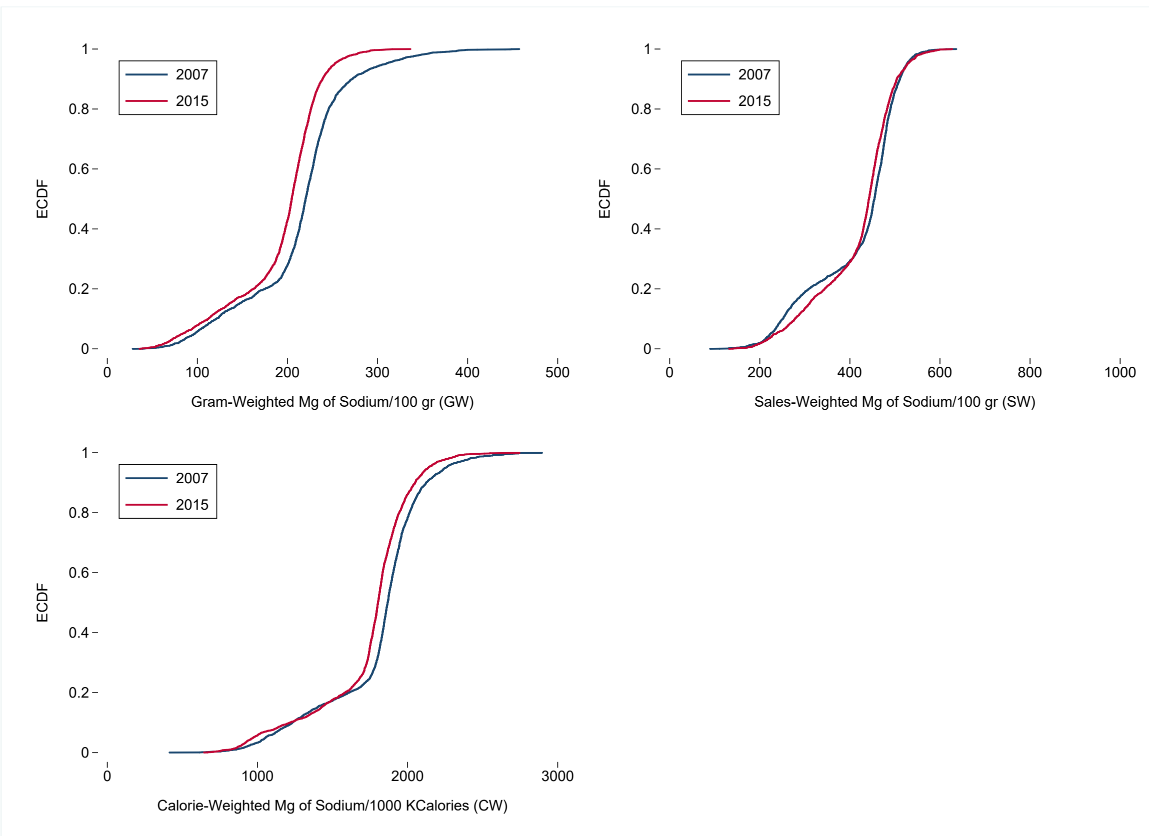
	Reformulation	Switching	Entry	Exit	Total
1. Bakery	-45%	42%	-9%	2%	-10%
2. Cereal	-39%	30%	-5%	-1%	-15%
3. Meats	-47%	44%	-3%	-1%	-7%
4. Dairy	-56%	50%	-4%	1%	-10%
5. Fats and Oils	-44%	41%	1%	1%	-1%
6. Sauces	-60%	56%	-8%	7%	-6%
7. Snacks	-29%	28%	-13%	2%	-13%
8. Soups	-70%	68%	-2%	0%	-4%
9. Potatoes	-55%	42%	3%	0%	-9%
10. Mixed Dishes	-39%	34%	-6%	1%	-11%
11. Vegetables	-70%	61%	-6%	1%	-15%
12. Legumes	-68%	50%	-7%	1%	-25%
13. Canned Fish	-26%	17%	-8%	0%	-17%
14. Seasonings	-73%	41%	13%	1%	-19%
15. Nut Butters	-63%	44%	-1%	0%	-20%

Notes. Balanced panel of households. 2007 and 2015 Nielsen Home Measurement System (HMS) and Gladson.

B.2 Retail data results

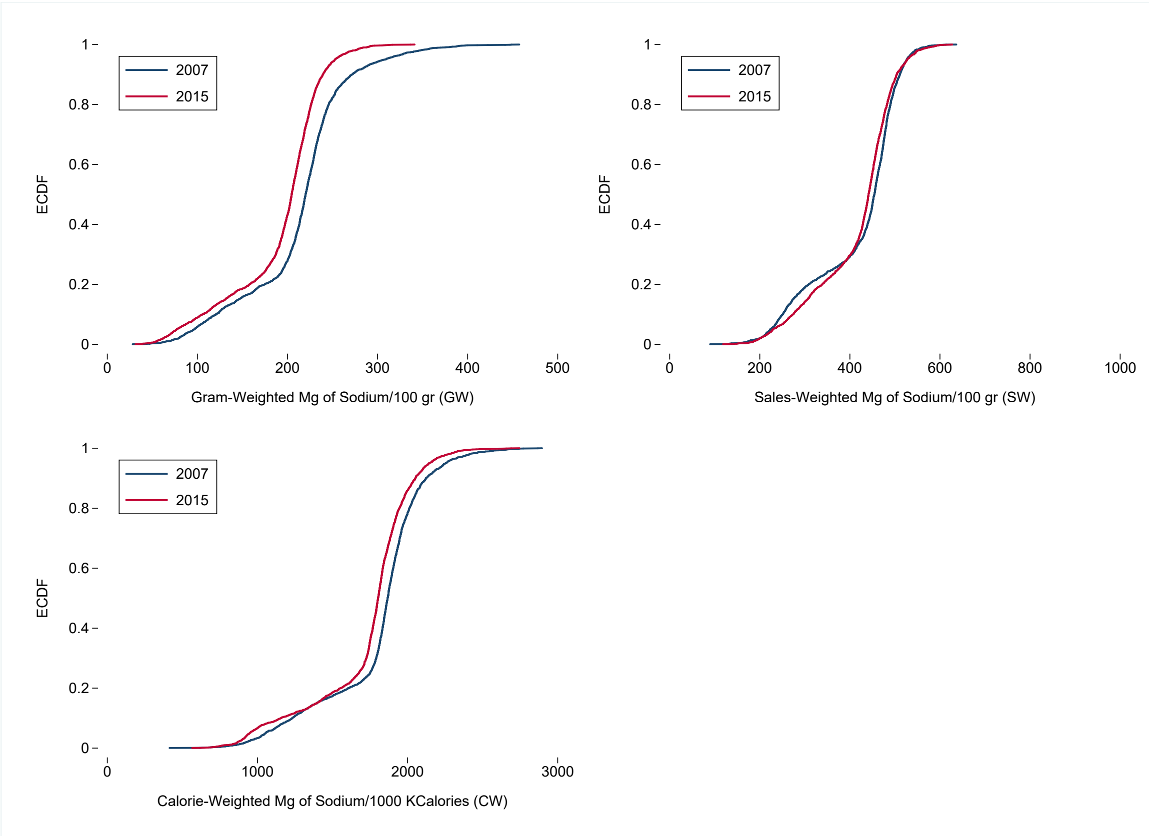
In this appendix we report sodium intake and decomposition results using retail (RMS) data. We report results that are analogous to the HMS analysis in the paper (various intake measures as well as for all vs. balanced panel of counties). Since we do not observe the number of customers associated with the purchases made in the retail dataset, it is not possible to compute the per capita (PC) intake measure; hence, the results focus on the three alternative measures (gram-weighted, sales-weighted, and calorie-weighted).

Figure B.2. Empirical cumulative distribution of sodium intake measures, retail data (balanced panel)



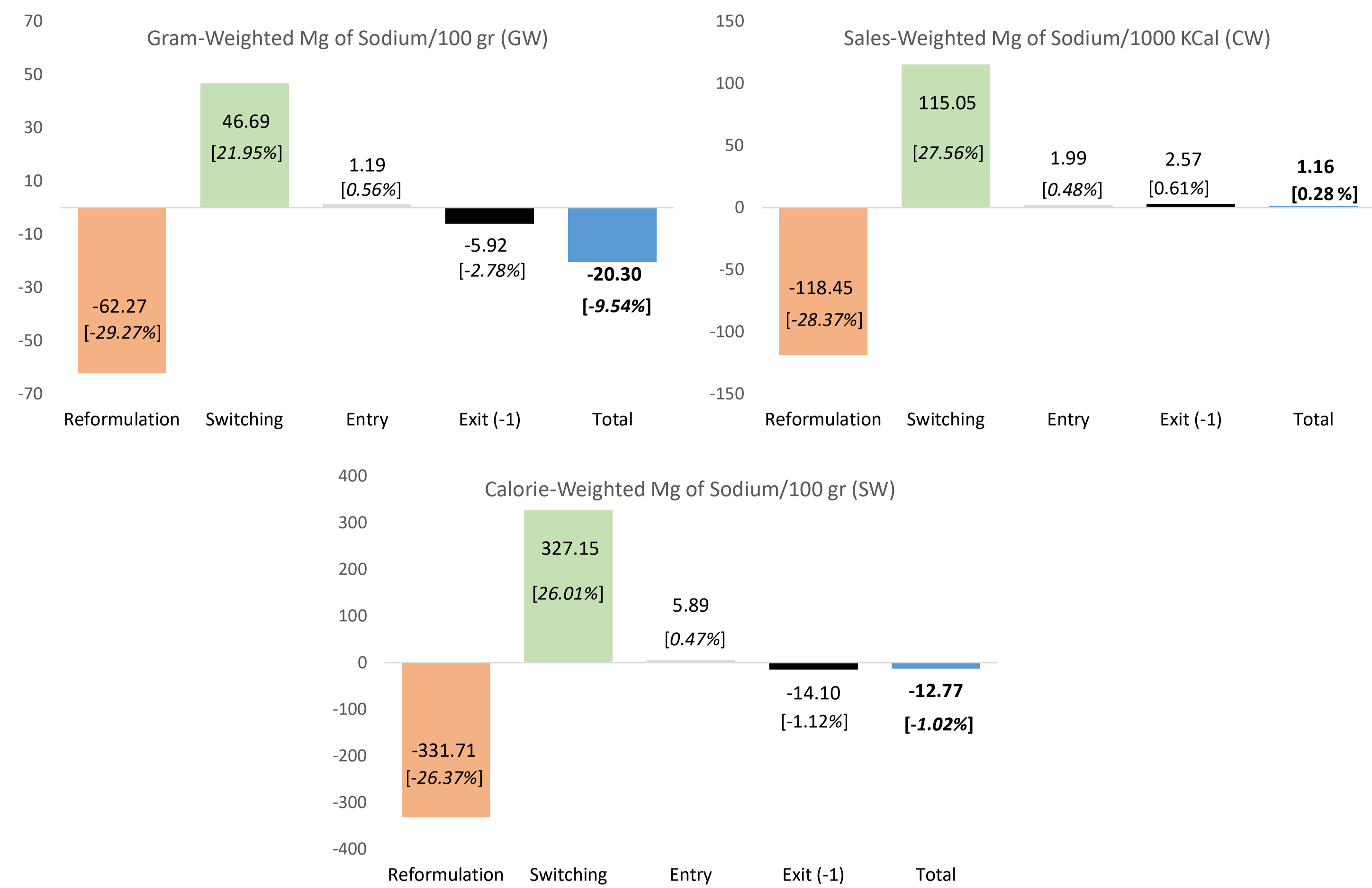
Notes. Figures created using 2,315 counties that appear in both 2007 and 2015 (balanced panel). Figures with unbalanced panel show almost identical patterns (see Appendix Figure B.3). 2007 and 2015 Nielsen Retail Measurement System (RMS) and Gladson.

Figure B.3. Empirical cumulative distribution of sodium intake measures, retail data (unbalanced panel)



Notes. Figures with balanced panel show almost identical patterns (see Appendix Figure **??**). 2007 and 2015 Nielsen Retail Measurement System (RMS) and Gladson.

Figure B.4. Decomposition of the temporal change in sodium intake measures, retail data



Notes. Every bar depicts the mean of the decomposition component across all units (counties). Figure uses 2,315 counties (balanced panel of counties). The "Total" bar in the decomposition figure does not coincide with the changes reported in Table B.4. 2007 and 2015 Nielsen Retail Measurement System (RMS) and Gladson.

Table B.4. Median sodium intake by measure type and year, retail data

	All Counties			Balanced Panel of Counties		
	2007	2015	% Change	2007	2015	% Change
Gram-Weighted (GW)	220.17	204.72	-7.02	220.18	204.75	-7.01
Sales-Weighted (SW)	454.51	441.43	-2.88	454.52	442.44	-2.66
Calorie-Weighted (CW)	1867.28	1803.51	-3.42	1,867.30	1,802.79	-3.45
# Counties	2,316	2,400		2,315	2,315	

Notes. Balanced panel of counties represents the counties that are sampled in both 2007 and 2015. 2007 and 2015 Nielsen Retail Measurement System (RMS) and Gladson.

Table B.5. Decomposition by UPC type and NSRI food group, SW measure (retail data)

	Reformulation	Switching	Entry	Exit	Total
Panel A: Overall					
Non-NSRI	-26%	24%	3%	5%	6%
NSRI	-30%	30%	-5%	1%	-4%
Panel B: By-NSRI group					
1. Bakery	-34%	30%	-6%	1%	-9%
2. Cereal	-50%	43%	-6%	2%	-11%
3. Meats	-26%	23%	-2%	-1%	-6%
4. Dairy	-20%	11%	-3%	0%	-12%
5. Fats and Oils	-21%	14%	-3%	1%	-8%
6. Sauces	-24%	29%	-3%	2%	4%
7. Snacks	-32%	34%	-2%	0%	0%
8. Soups	-31%	49%	-9%	1%	9%
9. Potatoes	-44%	33%	-4%	8%	-7%
10. Mixed Dishes	-24%	21%	-4%	3%	-5%
11. Vegetables	-35%	31%	-8%	0%	-12%
12. Legumes	-26%	21%	6%	0%	1%
13. Canned Fish	-56%	49%	-21%	0%	-28%
14. Seasonings	-34%	23%	-12%	4%	-19%
15. Nut Butters	-57%	53%	-4%	0%	-8%

Notes. Balanced panel of counties. SW is sales-weighted intake measure. 2007 and 2015 Nielsen Retail Measurement System (RMS) and Gladson.

Table B.6. Decomposition by UPC type and NSRI food group, GW measure (retail data)

	Reformulation	Switching	Entry	Exit	Total
1. Bakery	-33%	7%	40%	-26%	-12%
2. Cereal	-49%	6%	46%	-32%	-29%
3. Meats	-26%	4%	52%	-29%	1%
4. Dairy	-17%	14%	34%	-23%	8%
5. Fats and Oils	-20%	5%	34%	-23%	-4%
6. Sauces	-30%	12%	38%	-21%	-2%
7. Snacks	-34%	4%	76%	-52%	-6%
8. Soups	-27%	3%	20%	-8%	-13%
9. Potatoes	-25%	12%	38%	-22%	2%
10. Mixed Dishes	-22%	3%	55%	-38%	-3%
11. Vegetables	-25%	8%	13%	-9%	-13%
12. Legumes	-13%	13%	27%	-14%	14%
13. Canned Fish	-58%	8%	36%	-30%	-45%
14. Seasonings	-21%	15%	47%	-39%	2%
15. Nut Butters	-52%	-7%	47%	-5%	-18%

Notes. Balanced panel of counties. GW is gram-weighted intake measure. 2007 and 2015 Nielsen Retail Measurement System (RMS) and Gladson.

Table B.7. Decomposition by UPC type and NSRI food group, CW Measure (retail data)

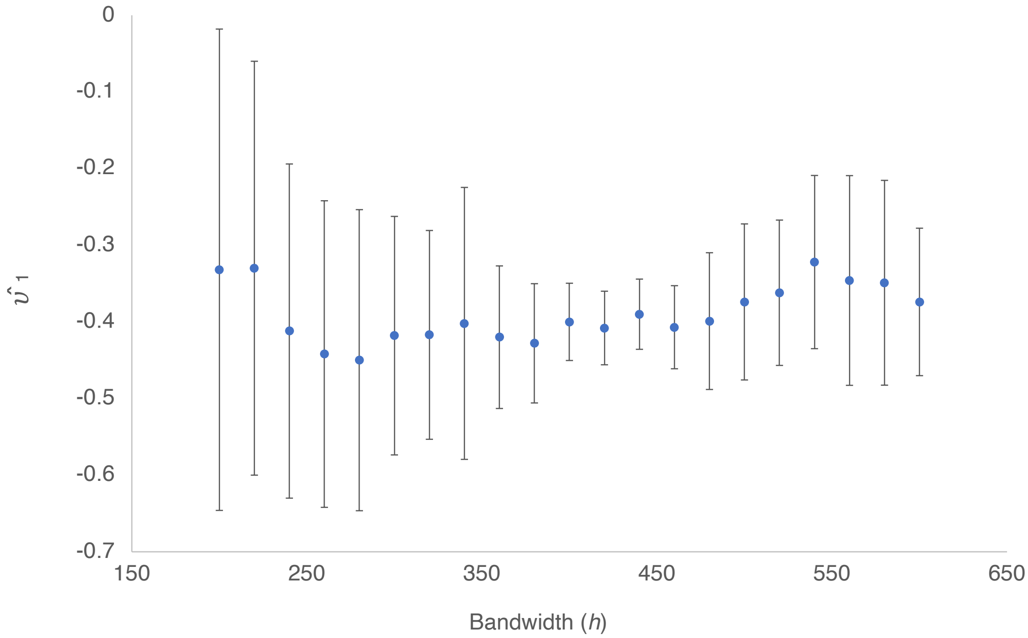
	Reformulation	Switching	Entry	Exit	Total
Panel A: Overall					
Non-NSRI	-24%	13%	23%	-9%	2%
NSRI	-27%	16%	24%	-16%	-2%
Panel B: By-NSRI group					
1. Bakery	-33%	12%	37%	-23%	-7%
2. Cereal	-49%	8%	47%	-31%	-24%
3. Meats	-21%	2%	54%	-28%	7%
4. Dairy	-17%	16%	35%	-22%	11%
5. Fats and Oils	-15%	3%	34%	-22%	1%
6. Sauces	-27%	10%	41%	-22%	2%
7. Snacks	-34%	7%	76%	-48%	1%
8. Soups	-25%	3%	21%	-8%	-9%
9. Potatoes	-22%	11%	40%	-22%	7%
10. Mixed Dishes	-21%	4%	56%	-37%	2%
11. Vegetables	-25%	12%	14%	-9%	-9%
12. Legumes	-12%	16%	28%	-14%	18%
13. Canned Fish	-57%	7%	38%	-29%	-41%
14. Seasonings	-6%	0%	41%	-52%	-17%
15. Nut Butters	-52%	-6%	48%	-5%	-14%

Notes. Balanced panel of counties. CW is calorie-weighted intake measure. 2007 and 2015 Nielsen Retail Measurement System (RMS) and Gladson.

B.3 Additional KRD Tests

Figure B.5 reports the estimated KRD estimate ($\widehat{\nu}_1$) for the main specification for bandwidths 200–600 (mg/100 g). Estimates fluctuate around the main specification's estimate (-0.4) and are always precisely estimated (95 percent confidence intervals). Similar graphs are obtained for Specifications 2–4 (available upon request).

Figure B.5. Sensitivity of $\widehat{\nu}_1$ to bandwidth



Notes. Figure reports KRD estimate for main specification (Specification 1 in Table 2.9). Vertical bars denote 95 percent confidence intervals.

Table B.8 reports results for a quadratic specification (bandwidth is fixed at the same length as in Table 2.9); for comparison purposes, Table B.8 replicates the linear specification results from Table 2.9 in the text. The main takeaway is that results are qualitatively consistent with the linear specification. Specifications 2 and 3 lose statistical significance as the power of these regressions is substantially lower (Specifications 1 and 4 have more than 10 times the number of observations).

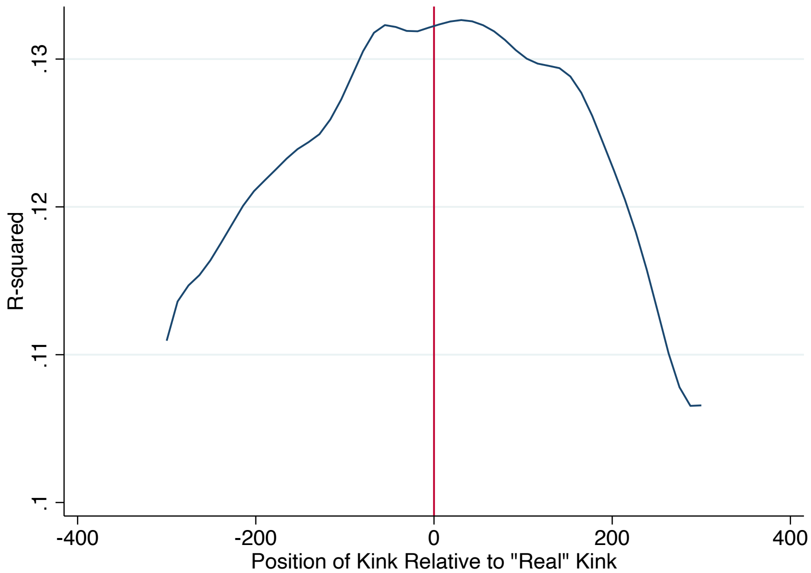
Table B.8. $\widehat{\nu}_1$ for varying order of polynomial, $\widehat{\nu}_1$

D.V.: Δ Sodium (mg/100g)	Spec. 1: *All Firms*	Spec. 2: *NSRI Signatories*	Spec. 3: *Other Signatories*	Spec. 4: *Non-Signatories*
$p = 1$ (Table 2.9)	-0.40	-0.29	-0.33	-0.35
	[<0.001]	[0.09]	[0.003]	[0.004]
$p = 2$	-0.41	-0.22	-0.16	-0.56
	[0.01]	[0.14]	[0.42]	[0.003]

Notes: D.V. is dependent variable. p is the order of polynomial. $\widehat{\nu}_1$ is kink regression discontinuity (KRD) estimate. P-values are in brackets and based on robust inference with standard errors clustered at 9 broad food types. See Section 2.5.2 for further estimation details.

One concern in the KRD design is that the estimation is picking up a spurious break point in the relationship between the assignment variable and the outcome of interest. One way to test for potential spuriousness is to search for a break point to check that our chosen threshold is consistent with the kink that would have been chosen if one had no prior information of where the kink could be. Following the literature in the detection of structural breaks, we implement a nonparametric test by running our main specification (first column, Table 2.9) for a large number of virtual kink points and then carrying out a grid search for the optimal (highest) R^2 value. The idea is that the kink that we use should produce a fitness measure (R-squared) that is at or around the highest observed level. Figure B.6 reports the results. We interpret this visual evidence as strongly supportive of the kink point being at 0, for two reasons. First, the R-squared measure increases sharply as one approaches 0 from either side, providing evidence for the validity of our KRD design. Second, the maximum R-squared value is located at 7.5, only slightly to the right of the kink point that we utilize. (Furthermore, we cannot reject the null hypothesis that the kink point is located at 0.)

Figure B.6. R^2 as a function of kink point location

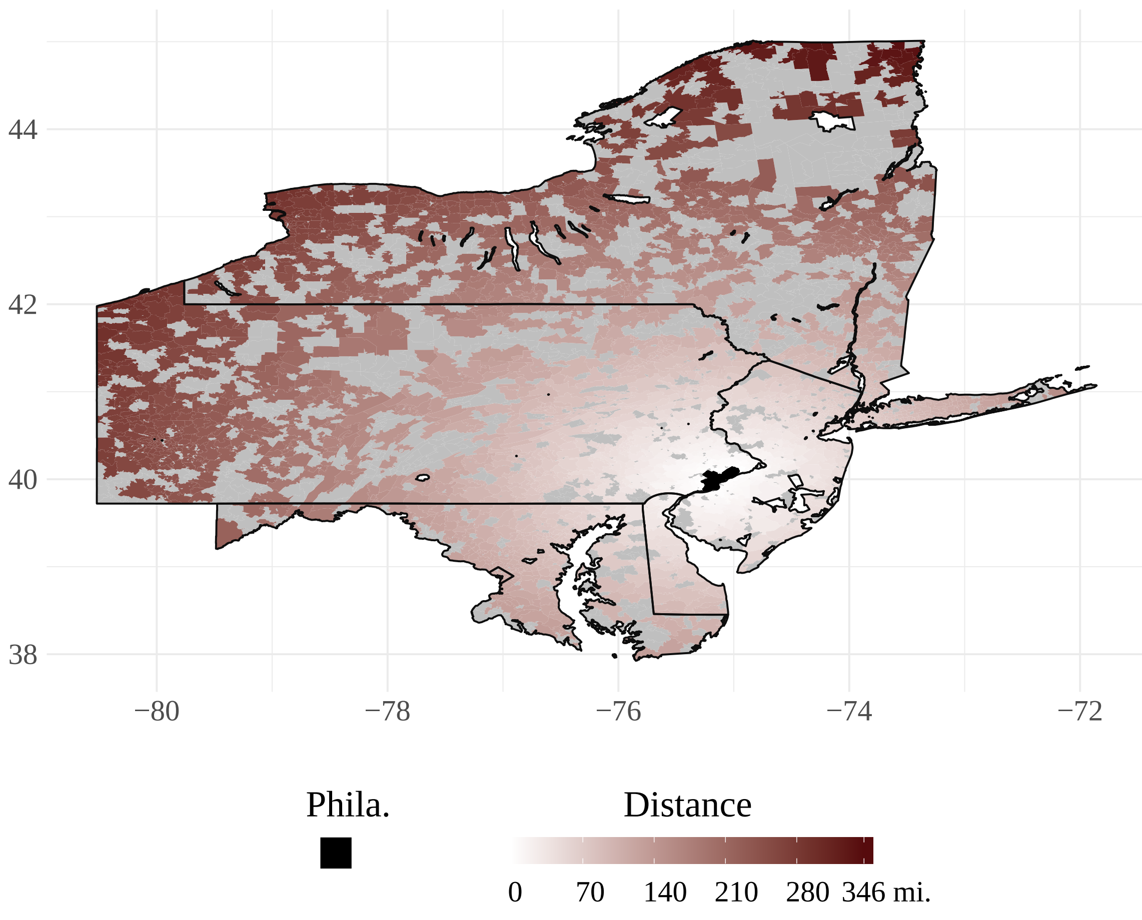


Notes. Figure reports the R-squared obtained in Specification 1 for varying kink point locations.

APPENDIX C

APPENDIX OF CHAPTER 3

Figure C.1. Map of the City of Philadelphia and its surrounding localities by 350-miles distance



Notes. I focus on five-digit zip code areas in six U.S. localities that are geographically proximate to the City of Philadelphia: five U.S. states (Pennsylvania, New York, Delaware, Maryland, New Jersey) and Washington, DC. I use STATA's vincenty package to calculate geodesic distances between any pair of localities using an ellipsoidal model of the earth. 2006–2017 Nielsen Homescan; 2014 MABLE GeoCorr.

Table C.1. Demographic breakdown in and outside Philadelphia, PA, by distance, 2016

	Share of White	Share of Hispanic	Mean Education	Mean Age	Income (US $)
Philadelphia	0.458	0.122	14.09	56.19	52,728.99
4 miles	0.665	0.102	14.52	52.72	67,151.27
6 miles	0.712	0.092	14.54	53.05	70,536.14
10 miles	0.747	0.074	14.22	54.00	72,307.30
20 miles	0.750	0.064	14.45	53.68	78,410.70
50 miles	0.784	0.100	14.30	52.88	77,991.96

Notes. I focus on five-digit zip code areas in six U.S. localities that are geographically proximate to the City of Philadelphia: five U.S. states (Pennsylvania, New York, Delaware, Maryland, New Jersey) and Washington, DC. I use STATA's vincenty package to calculate geodesic distances between any pair of localities using an ellipsoidal model of the earth. 2016 Nielsen Homescan; 2014 MABLE GeoCorr.

Table C.2. Demographic breakdown in and outside Philadelphia, PA, by distance, 2017

	Share of White	Share of Hispanic	Mean Education	Mean Age	Income (US $)
Philadelphia	0.483	0.102	13.92	55.92	51,024.40
4 miles	0.685	0.115	13.91	53.20	70,928.13
6 miles	0.729	0.102	14.03	53.42	72,614.29
10 miles	0.753	0.087	13.97	53.99	72,754.34
20 miles	0.754	0.084	14.33	53.54	77,209.58
50 miles	0.792	0.097	14.24	53.05	76,216.47

Notes. I focus on five-digit zip code areas in six U.S. localities that are geographically proximate to the City of Philadelphia: five U.S. states (Pennsylvania, New York, Delaware, Maryland, New Jersey) and Washington, DC. I use STATA's vincenty package to calculate geodesic distances between any pair of localities using an ellipsoidal model of the earth. 2017 Nielsen Homescan; 2014 MABLE GeoCorr.

Table C.3. Monthly price effects on taxed beverages in Philadelphia, PA, compared to counterfactual Philadelphia, between 2015–2017

Year	Month	Coefficient	SE	t value	pr(>|t|)	-95%	+95%
2015	1	0.0012	0.0014	0.8472	0.3974	-0.0011	0.0034
2015	2	-0.0020	0.0015	-1.3968	0.1633	-0.0044	0.0004
2015	3	0.0006	0.0005	1.1167	0.2648	-0.0003	0.0015
2015	4	-0.0021	0.0010	-2.2203	0.0270	-0.0037	-0.0005
2015	5	-0.0019	0.0011	-1.7515	0.0806	-0.0037	-0.0001
2015	6	-0.0058	0.0027	-2.1811	0.0298	-0.0102	-0.0014
2015	7	0.0016	0.0010	1.6340	0.1031	0.0000	0.0032
2015	8	0.0011	0.0027	0.4039	0.6865	-0.0034	0.0055
2015	9	0.0000	0.0010	-0.0300	0.9760	-0.0017	0.0016
2015	10	0.0027	0.0024	1.1559	0.2484	-0.0012	0.0066
2015	11	-0.0011	0.0016	-0.7147	0.4752	-0.0037	0.0015
2015	12	0.0015	0.0012	1.2529	0.2110	-0.0005	0.0035
2016	1	0.0059	0.0020	2.9162	0.0037	0.0025	0.0092
2016	2	0.0037	0.0009	4.2788	0.0000	0.0023	0.0051
2016	3	-0.0004	0.0008	-0.4968	0.6196	-0.0017	0.0009
2016	4	0.0012	0.0012	0.9659	0.3347	-0.0008	0.0031
2016	5	0.0007	0.0011	0.6402	0.5224	-0.0011	0.0025
2016	6	0.0021	0.0012	1.6905	0.0917	0.0001	0.0041
2016	7	0.0022	0.0019	1.1518	0.2501	-0.0010	0.0054
2016	8	-0.0004	0.0018	-0.2325	0.8162	-0.0034	0.0026
2016	9	0.0016	0.0007	2.3222	0.0207	0.0005	0.0027
2016	10	-0.0014	0.0013	-1.0848	0.2787	-0.0036	0.0007
2016	11	0.0023	0.0007	3.1409	0.0018	0.0011	0.0035
2016	12	0.0030	0.0018	1.7005	0.0898	0.0001	0.0059
2017	1	0.0013	0.0016	0.8146	0.4158	-0.0014	0.0040
2017	2	0.0072	0.0012	5.8022	0.0000	0.0051	0.0092
2017	3	0.0109	0.0011	9.7478	0.0000	0.0091	0.0128
2017	4	0.0065	0.0021	3.0612	0.0024	0.0030	0.0100
2017	5	0.0163	0.0036	4.5145	0.0000	0.0104	0.0223
2017	6	0.0073	0.0040	1.8437	0.0660	0.0008	0.0139
2017	7	0.0053	0.0026	2.0463	0.0414	0.0010	0.0095
2017	8	0.0106	0.0022	4.9350	0.0000	0.0071	0.0142
2017	9	0.0152	0.0023	6.7381	0.0000	0.0115	0.0190
2017	10	0.0128	0.0011	11.4629	0.0000	0.0109	0.0146
2017	11	0.0105	0.0010	10.3015	0.0000	0.0088	0.0122
2017	12	0.0107	0.0017	6.4988	0.0000	0.0080	0.0135

Notes. The table reports the regression results from equation 3.1 for taxed products. The taxed location is the 3-digit 191 zip code (City of Philadelphia). The untaxed comparison regions represent other 3-digit zip codes in Pennsylvania, New York, Delaware, Maryland, New Jersey, and Washington DC. Taxed products contain soda (regular and diet), energy drinks, sports drinks, sweetened fruit juice, sweetened water, sweetened coffee/tea, and sweetened milk. 2015–2017 Nielsen Homescan; 2014 MABLE GeoCorr.

Table C.4. Monthly price effects on untaxed beverages in Philadelphia, PA, compared to counterfactual Philadelphia, between 2015–2017

Year	Month	Coefficient	SE	t value	pr(>|t|)	-95%	+95%
2015	1	0.0011	0.0018	0.5992	0.5494	-0.0019	0.0041
2015	2	0.0043	0.0021	2.0727	0.0389	0.0009	0.0076
2015	3	0.0003	0.0027	0.1231	0.9021	-0.0041	0.0047
2015	4	0.0028	0.0021	1.3439	0.1798	-0.0006	0.0063
2015	5	0.0016	0.0012	1.2990	0.1947	-0.0004	0.0036
2015	6	-0.0022	0.0015	-1.5079	0.1324	-0.0046	0.0002
2015	7	-0.0031	0.0011	-2.8249	0.0050	-0.0049	-0.0013
2015	8	-0.0026	0.0013	-1.9351	0.0537	-0.0048	-0.0004
2015	9	0.0002	0.0017	0.1281	0.8981	-0.0025	0.0030
2015	10	0.0027	0.0020	1.3768	0.1694	-0.0005	0.0059
2015	11	-0.0017	0.0022	-0.7818	0.4348	-0.0053	0.0019
2015	12	-0.0022	0.0036	-0.5985	0.5499	-0.0081	0.0038
2016	1	-0.0101	0.0025	-3.9686	0.0001	-0.0143	-0.0059
2016	2	-0.0057	0.0018	-3.1584	0.0017	-0.0088	-0.0027
2016	3	-0.0079	0.0025	-3.1289	0.0019	-0.0120	-0.0037
2016	4	-0.0020	0.0024	-0.8266	0.4090	-0.0059	0.0020
2016	5	-0.0056	0.0028	-1.9531	0.0515	-0.0102	-0.0009
2016	6	-0.0008	0.0021	-0.4029	0.6872	-0.0042	0.0026
2016	7	-0.0018	0.0014	-1.3127	0.1901	-0.0041	0.0005
2016	8	-0.0035	0.0013	-2.7291	0.0066	-0.0057	-0.0014
2016	9	-0.0014	0.0019	-0.7266	0.4679	-0.0045	0.0017
2016	10	-0.0038	0.0016	-2.3689	0.0183	-0.0064	-0.0011
2016	11	-0.0038	0.0028	-1.3739	0.1703	-0.0084	0.0008
2016	12	-0.0022	0.0023	-0.9801	0.3276	-0.0060	0.0015
2017	1	-0.0064	0.0029	-2.2065	0.0279	-0.0111	-0.0016
2017	2	-0.0073	0.0013	-5.4380	0.0000	-0.0095	-0.0051
2017	3	-0.0012	0.0023	-0.5472	0.5845	-0.0050	0.0025
2017	4	-0.0061	0.0011	-5.3403	0.0000	-0.0080	-0.0042
2017	5	0.0024	0.0030	0.7896	0.4303	-0.0026	0.0074
2017	6	0.0018	0.0029	0.6067	0.5444	-0.0031	0.0066
2017	7	0.0020	0.0017	1.2214	0.2227	-0.0007	0.0048
2017	8	0.0024	0.0047	0.5161	0.6061	-0.0053	0.0101
2017	9	-0.0014	0.0012	-1.1579	0.2476	-0.0034	0.0006
2017	10	-0.0045	0.0031	-1.4195	0.1566	-0.0096	0.0007
2017	11	0.0031	0.0034	0.8972	0.3702	-0.0026	0.0088
2017	12	-0.0056	0.0026	-2.1506	0.0321	-0.0099	-0.0013

Notes. The table reports the regression results from equation 3.1 for untaxed products. The taxed location is the 3-digit 191 zip code (City of Philadelphia). The untaxed comparison regions represent other 3-digit zip codes in Pennsylvania, New York, Delaware, Maryland, New Jersey, and Washington, DC. Untaxed products are pure water, 100 percent juice, unsweetened milk, and unsweetened coffee/tea. 2015–2017 Nielsen Homescan; 2014 MABLE GeoCorr.

Table C.5. Price effects on taxed beverages in Philadelphia, PA, compared to counterfactual Philadelphia, by distance

Miles	Coefficient	SE	t-value	pr(>|t|)
10	0.0175	0.0086	2.043	0.041
20	0.0129	0.0024	5.365	0.000
30	0.0102	0.0017	5.941	0.000
40	0.0098	0.0015	6.389	0.000
50	0.0101	0.0014	7.283	0.000
60	0.0106	0.0014	7.769	0.000
70	0.0104	0.0013	7.937	0.000
80	0.0102	0.0013	7.982	0.000
90	0.0096	0.0012	8.099	0.000
100	0.0094	0.0011	8.318	0.000
110	0.0096	0.0011	8.366	0.000
120	0.0097	0.0011	8.614	0.000
130	0.0097	0.0011	8.585	0.000
140	0.0098	0.0011	8.658	0.000
150	0.0097	0.0011	8.639	0.000
160	0.0097	0.0011	8.611	0.000
170	0.0097	0.0011	8.638	0.000
180	0.0097	0.0011	8.480	0.000
190	0.0097	0.0011	8.560	0.000
200	0.0097	0.0011	8.590	0.000
346	0.0094	0.0011	8.625	0.000

Notes. The table shows the estimated price effects of the tax on taxed products using equation 3.1 for alternative control samples constructed using the distance in miles to Philadelphia criterion. The taxed location is the 3-digit 191 zip code (City of Philadelphia). The untaxed comparison regions represent other 3-digit zip codes in Pennsylvania, New York, Delaware, Maryland, New Jersey, and Washington, DC. Taxed products contain soda (regular and diet), energy drinks, sports drinks, sweetened fruit juice, sweetened water, sweetened coffee/tea, and sweetened milk. 2015–2017 Nielsen Homescan; 2014 MABLE GeoCorr.

...

BIBLIOGRAPHY

Aaker, David A, and James M Carman. 1982. “Are you over-advertizing.” *Journal of Advertising Research*, 22(4): 57–70.

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 2010. “Synthetic control methods for comparative case studies: Estimating the effect of California’s tobacco control program.” *Journal of the American statistical Association*, 105(490): 493–505.

Ackerberg, Daniel A. 2001. “Empirically distinguishing informative and prestige effects of advertising.” *RAND Journal of Economics*, 316–333.

Ackerberg, Daniel A. 2003. “Advertising, learning, and consumer choice in experience good markets: an empirical examination.” *International Economic Review*, 44(3): 1007–1040.

Ackoff, Russell L, and James R Emshoff. 1975. “Advertising research at anheuser-busch, inc.(1963-68).” *Sloan Management Review (pre-1986)*, 16(2): 1.

Aguilar, Arturo, Emilio Gutierrez, and Enrique Seira. 2016. “Taxing to reduce obesity.” *Wroking Paper*, 1–39.

Alé-Chilet, Jorge, and Sarah Moshary. 2020. “Beyond Consumer Switching: Supply Responses to Food Packaging and Advertising Regulations.” *SSRN Working Paper 3678744*.

Allcott, Hunt, Benjamin B Lockwood, and Dmitry Taubinsky. 2019*a*. “Regressive sin taxes, with an application to the optimal soda tax.” *The Quarterly Journal of Economics*, 134(3): 1557–1626.

Allcott, Hunt, Benjamin B Lockwood, and Dmitry Taubinsky. 2019*b*. “Should we tax sugar-sweetened beverages? An overview of theory and evidence.” *Journal of Economic Perspectives*, 33(3): 202–27.

Allcott, Hunt, Matthew Gentzkow, and Lena Song. 2021. “Digital addiction.” National Bureau of Economic Research.

Allcott, Hunt, Rebecca Diamond, Jean-Pierre Dubé, Jessie Handbury, Ilya Rahkovsky, and Molly Schnell. 2019. “Food Deserts and the Causes of Nutritional Inequality.” *Quarterly Journal of Economics*, Forthcoming.

American Diabetes Association, ADA. 2018. "Economic costs of diabetes in the US in 2017." *Diabetes care*, 41(5): 917–928.

Andreyeva, Tatiana, Michael W Long, and Kelly D Brownell. 2010. "The impact of food prices on consumption: a systematic review of research on the price elasticity of demand for food." *American journal of public health*, 100(2): 216–222.

Angrist, Joshua D, and Jörn-Steffen Pischke. 2010. "The credibility revolution in empirical economics: How better research design is taking the con out of econometrics." *Journal of economic perspectives*, 24(2): 3–30.

Ashenfelter, Orley C, Daniel S Hosken, and Matthew C Weinberg. 2015. "Efficiencies brewed: pricing and consolidation in the US beer industry." *The RAND Journal of Economics*, 46(2): 328–361.

Assmus, Gert, John U Farley, and Donald R Lehmann. 1984. "How advertising affects sales: meta-analysis of econometric results." *Journal of Marketing Research*, 21(1): 65–74.

Athey, Susan, Mohsen Bayati, Nikolay Doudchenko, Guido Imbens, and Khashayar Khosravi. 2021. "Matrix completion methods for causal panel data models." *Journal of the American Statistical Association*, 116(536): 1716–1730.

Bagwell, Kyle. 2007. "The economic analysis of advertising." *Handbook of industrial organization*, 3: 1701–1844.

Baily, Martin Neil, Charles Hulten, David Campbell, Timothy Bresnahan, and Richard E Caves. 1992. "Productivity dynamics in manufacturing plants." *Brookings papers on economic activity. Microeconomics*, 1992: 187–267.

Beer Info, BI. 2021. "Beer Info 2021." http://beerinfo.com/beer–consumption–by–state–per–capita/ (August 16, 2021).

Beer Institute, BI. 2021. "Annual Report 2020." https://www.beerinstitute2020annualreport.com/ (August 11, 2021).

Benjamin, Emelia J, Salim S Virani, Clifton W Callaway, Alanna M Chamberlain, Alexander R Chang, Susan Cheng, Stephanie E Chiuve, Mary Cushman, Francesca N Delling, Rajat Deo, et al. 2018. "Heart disease and stroke statistics-2018 update: a report from the American Heart Association." *Circulation*, 137(12): e67.

Berardi, Nicoletta, Patrick Sevestre, Marine Tepaut, and Alexandre Vigneron. 2016. "The impact of a ?soda tax?on prices: evidence from French micro data." *Applied Economics*, 48(41): 3976–3994.

Bergman, U Michael, and Niels Lynggård Hansen. 2010. "Are excise taxes on beverages fully passed through to prices? The Danish evidence."

Black, Sandra E. 1999. "Do better schools matter? Parental valuation of elementary education." *The quarterly journal of economics*, 114(2): 577–599.

Bollinger, Bryan, and Steven Sexton. 2018. "Local excise taxes, sticky prices, and spillovers: evidence from Berkeley's soda tax." *Sticky Prices, and Spillovers: Evidence from Berkeley's Soda Tax (January 12, 2018).*

Bollinger, Bryan, Phillip Leslie, and Alan Sorensen. 2011. "Calorie posting in chain restaurants." *American Economic Journal: Economic Policy*, 3(1): 91–128.

Boon, Caitlin S, Christine L Taylor, Jane E Henney, et al. 2010. *Strategies to reduce sodium intake in the United States.* National Academies Press.

Brat, Ilan, and Maurice Tamman. 2010. "Food Makers Quietly Cut Back on Salt."

Brewbound. 2021. "IRI: Off-Premise Beer Category Sales Reach Nearly 43.8 Billion Dollars in 2020." https://www.brewbound.com/news/iri–off–premise–beer–category–sales–reach–nearly–43–8–billion–in–2020 (January 8, 2021).

Brewers Association, BA. 2018. "Market Segments." https://www.brewersassociation.org/statistics/market–segments/. (January 8, 2019).

Brewers Association, BA. 2021. "National Beer Sales and Production Data." https://www.brewersassociation.org/statistics–and–data/national–beer–stats/ (August 16, 2021).

Bronnenberg, Bart J, Sanjay K Dhar, and Jean-Pierre H Dubé. 2011. "Endogenous sunk costs and the geographic differences in the market structures of CPG categories." *Quantitative Marketing and Economics*, 9(1): 1–23.

Brownell, Kelly D, Thomas Farley, Walter C Willett, Barry M Popkin, Frank J Chaloupka, Joseph W Thompson, and David S Ludwig. 2009. "The public health and economic benefits of taxing sugar-sweetened beverages."

Bureau of Economic Analysis, BEA. 2021. "Gross Domestic Product, Fourth Quarter and Year 2020 (Advanced Estimate)." *Bureau of Economic Analysis*, https://www.bea.gov/news/2021/gross–domestic–product–4th–quarter–and–year–2020–advance–estimate.

Card, David, and Alan B Krueger. 1994. "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania." *The American Economic Review*, 84(4): 772–793.

Card, David, David S. Lee, Zhuan Pei, and Andrea Weber. 2015. "Inference on Causal Effects in a Generalized Regression Kink Design." *Econometrica*, 83(6): 2453–2483.

Cawley, John, and David E Frisvold. 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, John, Anne Marie Thow, Katherine Wen, and David Frisvold. 2019*a*. "The economics of taxes on sugar-sweetened beverages: a review of the effects on prices, sales, cross-border shopping, and consumption." *Annual review of nutrition*, 39: 317–338.

Cawley, John, Barton Willage, and David Frisvold. 2018. "Pass-through of a tax on sugar-sweetened beverages at the Philadelphia International Airport." *Jama*, 319(3): 305–306.

Cawley, John, David Frisvold, and David Jones. 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, John, David Frisvold, Anna Hill, and David Jones. 2019*b*. "The impact of the Philadelphia beverage tax on purchases and consumption by adults and children." *Journal of Health Economics*, 67: 102225.

Cawley, John, David Frisvold, Anna Hill, and David Jones. 2020*a*. "The impact of the Philadelphia beverage tax on prices and product availability." *Journal of Policy Analysis and Management*, 39(3): 605–628.

Cawley, John, David Frisvold, Anna Hill, and David Jones. 2020*b*. "Oakland?s sugar-sweetened beverage tax: Impacts on prices, purchases and consumption by adults and children." *Economics & Human Biology*, 37: 100865.

Cawley, John, David Frisvold, David Jones, and Chelsea Lensing. 2021. "The Pass-Through of a Tax on Sugar-Sweetened Beverages in Boulder, Colorado." *American Journal of Agricultural Economics*, 103(3): 987–1005.

CDC. 2020. "Adult Obesity Facts." https://www.cdc.gov/obesity/data/adult.html.

Cengiz, Doruk, and Hasan Tekgüç. 2022. "Is it merely a labor supply shock? impacts of Syrian migrants on local economies in Turkey." *ILR Review*, 75(3): 741–768.

Cengiz, Ezgi, Christian Rojas, and Emily Wang. 2020. "The Evolution of Diet Quality in the US: Evidence from Scanner Data."

City of Philadelphia. 2017. "Sugar-sweetened beverage tax ("SBT") regulations." https://www.phila.gov/media/20170808092824/Sweetened-Bev-Tax-Regs-Comp-8-4-2017-final.pdf.

Clapp, Jenifer E, Sarah A Niederman, Elizabeth Leonard, and Christine J Curtis. 2018. "Changes in Serving Size, Calories, and Sodium Content in Processed Foods From 2009 to 2015." *Preventing chronic disease*, 15.

Clarke, Darral G. 1976. "Econometric measurement of the duration of advertising effect on sales." *Journal of Marketing Research*, 13(4): 345–357.

Clemens, Michael A, and Jennifer Hunt. 2017. "The labor market effects of refugee waves: reconciling conflicting results." *ILR Review*, 0019793918824597.

Colchero, M Arantxa, Juan Rivera-Dommarco, Barry M Popkin, and Shu Wen Ng. 2017. "In Mexico, evidence of sustained consumer response two years after implementing a sugar-sweetened beverage tax." *Health Affairs*, 36(3): 564–571.

Curtis, Christine J, Jenifer Clapp, Sarah A Niederman, Shu Wen Ng, and Sonia Y Angell. 2016. "US food industry progress during the National Salt Reduction Initiative: 2009–2014." *American journal of public health*, 106(10): 1815–1819.

Daily Beast, DB. 2020. "No, Beer Sales Aren't Booming." https://www.thedailybeast.com/no–beer–sales–arent–booming (September 22, 2020).

Danaei, Goodarz, Eric L Ding, Dariush Mozaffarian, Ben Taylor, Jürgen Rehm, Christopher JL Murray, and Majid Ezzati. 2009. "The preventable causes of death in the United States: comparative risk assessment of dietary, lifestyle, and metabolic risk factors." *PLoS Med*, 6(4): e1000058.

DeCicca, Philip, Donald Kenkel, Feng Liu, and Jason Somerville. 2021. "Quantifying Brand Loyalty: Evidence from the Cigarette Market." *Journal of Health Economics*, 102512.

Dharmasena, Senarath, and Oral Capps. 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.

Dube, Arindrajit, T William Lester, and Michael Reich. 2010. "Minimum wage effects across state borders: Estimates using contiguous counties." *The Review of Economics and Statistics*, 92(4): 945–964.

Dubé, Jean-Pierre, Günter J Hitsch, and Puneet Manchanda. 2005. "An empirical model of advertising dynamics." *Quantitative marketing and economics*, 3(2): 107–144.

Dubois, Pierre, Rachel Griffith, and Aviv Nevo. 2014. "Do prices and attributes explain international differences in food purchases?" *American Economic Review*, 104(3): 832–67.

Dubois, Pierre, Rachel Griffith, and Martin O'Connell. 2017*a*. "The effects of banning advertising in junk food markets." *The Review of Economic Studies*, 85(1): 396–436.

Dubois, Pierre, Rachel Griffith, and Martin O'Connell. 2017*b*. "How well targeted are soda taxes?"

Dubois, Pierre, Rachel Griffith, and Martin O'Connell. 2020. "How well targeted are soda taxes?" *American Economic Review*, 110(11): 3661–3704.

Einav, Liran, Ephraim Leibtag, and Aviv Nevo. 2010. "Recording discrepancies in Nielsen Homescan data: Are they present and do they matter?" *QME*, 8(2): 207–239.

Ellickson, Paul B. 2007. "Does Sutton apply to supermarkets?" *The RAND Journal of Economics*, 38(1): 43–59.

eMarketer. 2021. "Total Media Ad Spending in the U.S." https://www.emarketer.com/forecasts. (March 2021).

Erdem, Tülin, Michael P Keane, and Baohong Sun. 2008. "The impact of advertising on consumer price sensitivity in experience goods markets." *Quantitative Marketing and Economics*, 6(2): 139–176.

Falbe, Jennifer, Hannah R Thompson, Christina M Becker, Nadia Rojas, Charles E McCulloch, and Kristine A Madsen. 2016. "Impact of the Berkeley excise tax on sugar-sweetened beverage consumption." *American journal of public health*, 106(10): 1865–1871.

Falbe, Jennifer, Matthew M Lee, Scott Kaplan, Nadia A Rojas, Alberto M Ortega Hinojosa, and Kristine A Madsen. 2020. "Higher Sugar-Sweetened Beverage Retail Prices After Excise Taxes in Oakland and San Francisco." *American Journal of Public Health*, , (0): e1–e7.

FDA, Food, and Drug Administration. 2021. "Guidance for Industry: Voluntary Sodium Reduction Goals." *Federal Registry 86 FR 57156*, 57156–57159.

Finkelstein, Eric A, Justin G Trogdon, Joel W Cohen, and William Dietz. 2009. "Annual medical spending attributable to obesity: payer-and service-specific estimates." *Health affairs*, 28(5): w822–w831.

Fortune. 2019. "U.S. Health Care Costs Skyrocketed to 3.65 Trillion Dollars in 2018." http://fortune.com/2019/02/21/us–health–care–costs–2/. (February 21, 2019).

Foster, Lucia, John C Haltiwanger, and Cornell John Krizan. 2001. "Aggregate productivity growth: Lessons from microeconomic evidence." In *New developments in productivity analysis*. 303–372. University of Chicago Press.

Foster, Lucia, John Haltiwanger, and Chad Syverson. 2008. "Reallocation, firm turnover, and efficiency: selection on productivity or profitability?" *American Economic Review*, 98(1): 394–425.

Gavazza, Alessandro. 2011. "Demand spillovers and market outcomes in the mutual fund industry." *The RAND Journal of Economics*, 42(4): 776–804.

Gonçalves, Judite, and João Pereira Dos Santos. 2020. "Brown sugar, how come you taste so good? The impact of a soda tax on prices and consumption." *Social Science & Medicine*, 264: 113332.

Gressier, Mathilde, Boyd Swinburn, Gary Frost, Alexa B Segal, and Franco Sassi. 2021. "What is the impact of food reformulation on individuals' behaviour, nutrient intakes and health status? A systematic review of empirical evidence." *Obesity Reviews*, 22(2): e13139.

Griffith, Rachel, Martin O'Connell, and Kate Smith. 2017. "The importance of product reformulation versus consumer choice in improving diet quality." *Economica*, 84(333): 34–53.

Henningsen, Sina, Rebecca Heuke, and Michel Clement. 2011. "Determinants of advertising effectiveness: The development of an international advertising elasticity database and a meta-analysis." *Business Research*, 4(2): 193–239.

Heron, Melonie P. 2019. "Deaths: leading causes for 2017." *National Vital Statistics Reports*, 68(6).

Hitsch, Günter J, Ali Hortacsu, and Xiliang Lin. 2019. "Prices and promotions in us retail markets: Evidence from big data." National Bureau of Economic Research.

Hovde, Howard T. 1936. "Recent trends in the development of market research." *American Marketing Journal*, , (1): 3–19.

Hut, Stefan. 2018. "Determinants of Dietary Choice in the US: Evidence from Consumer Migration." Working Paper, Brown University.

Hut, Stefan, and Emily Oster. 2018. "Changes in Household Diet: Determinants and Predictability." National Bureau of Economic Research.

Hu, Ye, Leonard M Lodish, and Abba M Krieger. 2007. "An analysis of real world TV advertising tests: A 15-year update." *Journal of Advertising Research*, 47(3): 341–353.

Imbens, Guido W, and Donald B Rubin. 2015. *Causal inference in statistics, social, and biomedical sciences.* Cambridge University Press.

Jacobs, Andrew. 2021. "F.D.A. Issues Guidelines to Reduce Salt in Foods." *Wall Street Journal.*

Kiesel, Kristin, and Sofia B Villas-Boas. 2013. "Can information costs affect consumer choice? Nutritional labels in a supermarket experiment." *International Journal of Industrial Organization.*

Lawman, Hannah G, Sara N Bleich, Jiali Yan, Sophia V Hua, Caitlin M Lowery, Ana Peterhans, Michael T LeVasseur, Nandita Mitra, Laura A Gibson, and Christina A Roberto. 2020. "One-year changes in sugar-sweetened beverage consumers? purchases following implementation of a beverage tax: a longitudinal quasi-experiment." *The American journal of clinical nutrition*, 112(3): 644–651.

Lee, Matthew M, Jennifer Falbe, Dean Schillinger, Sanjay Basu, Charles E McCulloch, and Kristine A Madsen. 2019. "Sugar-sweetened beverage consumption 3 years after the Berkeley, California, sugar-sweetened beverage tax." *American journal of public health*, 109(4): 637–639.

Leider, JULIEN, ANDREA A Pipito, and LISA M Powell. 2018. "The impact of the Cook County, Illinois, Sweetened Beverage Tax on prices, 2017." *Chicago: Illinois Prevention Research Center, University of Illinois at Chicago*.

Levy, Jordan. 2022. "Philly?s soda tax is under fire again. Who wants to see it end, and who still thinks it?s a win?" https://billypenn.com/2022/02/01/philadelphia–soda–tax–repeal–kenney–oh–rebuild–rhynhart/. (February 1, 2022).

Lewis, Randall A, and David H Reiley. 2014. "Online ads and offline sales: measuring the effect of retail advertising via a controlled experiment on Yahoo!" *Quantitative Marketing and Economics*, 12(3): 235–266.

Lim, Joon Ho, Rishika Rishika, Ramkumar Janakiraman, and PK Kannan. 2020. "Competitive effects of front-of-package nutrition labeling adoption on nutritional quality: evidence from facts up front–style labels." *Journal of Marketing*, 84(6): 3–21.

Lodish, Leonard M, Magid Abraham, Stuart Kalmenson, Jeanne Livelsberger, Beth Lubetkin, Bruce Richardson, and Mary Ellen Stevens. 1995. "How TV advertising works: A meta-analysis of 389 real world split cable TV advertising experiments." *Journal of Marketing Research*, 32(2): 125–139.

Melitz, Marc J, and Sašo Polanec. 2015. "Dynamic Olley-Pakes productivity decomposition with entry and exit." *The Rand journal of economics*, 46(2): 362–375.

Meyers, Linda D, Jennifer Pitzi Hellwig, Jennifer J Otten, et al. 2006. *Dietary reference intakes: the essential guide to nutrient requirements.* National Academies Press.

Miller, Nathan H, and Matthew C Weinberg. 2017. "Understanding the price effects of the MillerCoors joint venture." *Econometrica*, 85(6): 1763–1791.

Moorman, Christine, Rosellina Ferraro, and Joel Huber. 2012. "Unintended nutrition consequences: firm responses to the nutrition labeling and education act." *Marketing Science*, 31(5): 717–737.

Moss, Michael. 2013. *Salt, sugar, fat: How the food giants hooked us.* Random House.

Nelson, Max. 2005. *The barbarian's beverage: a history of beer in ancient Europe.* Routledge.

Nevo, Aviv. 2000. "A practitioner's guide to estimation of random-coefficients logit models of demand." *Journal of economics & management strategy*, 9(4): 513–548.

Nevo, Aviv, and Michael D Whinston. 2010. "Taking the dogma out of econometrics: Structural modeling and credible inference." *Journal of Economic Perspectives*, 24(2): 69–82.

Nickell, Stephen. 1981. "Biases in dynamic models with fixed effects." *Econometrica: Journal of the Econometric Society*, 1417–1426.

Orso, Anna. 2017. "ShopRite admits Philly stores overcharged for the soda tax." https://billypenn.com/2017/01/05/shoprite–admits–philly–stores–overcharged–for–the–soda–tax/. (January 5, 2017).

Polar, Kartika, and Roland Sturm. 2009. "Potential societal savings from reduced sodium consumption in the US adult population." *American journal of health promotion*, 24(1): 49–57.

Powell, Lisa M, and Julien Leider. 2020. "The impact of Seattle?s Sweetened Beverage Tax on beverage prices and volume sold." *Economics & Human Biology*, 37: 100856.

Powell, Lisa M, Julien Leider, and Pierre Thomas Léger. 2020. "The impact of the Cook County, IL, Sweetened Beverage Tax on beverage prices." *Economics & Human Biology*, 37: 100855.

Powell, LISA M, SAMANTHA Marinello, JULIEN Leider, and TATIANA Andreyeva. 2021. "A review and meta-analysis of the impact of local US sugar-sweetened beverage taxes on demand." *UIC Policy, Practice and Prevention Research Center Research Brief*, , (121).

Rehm, Colin D, José L Peñalvo, Ashkan Afshin, and Dariush Mozaffarian. 2016. "Dietary intake among US adults, 1999-2012." *JAMA*, 315(23): 2542–2553.

Rojas, Christian, and Emily Wang. 2017. "Do taxes for soda and sugary drinks work? scanner data evidence from berkeley and washington." *Scanner Data Evidence from Berkeley and Washington (September 23, 2017).*

Rojas, Christian, and Emily Wang. 2021. "Do taxes on soda and sugary drinks work? Scanner data evidence from Berkeley and Washington state." *Economic Inquiry*, 59(1): 95–118.

Rojas, Christian, and Everett B Peterson. 2008. "Demand for differentiated products: Price and advertising evidence from the US beer market." *International Journal of Industrial Organization*, 26(1): 288–307.

Schiff, Mary D, Dara D Mendez, Tiffany L Gary-Webb, J Jeffrey Inman, and Anthony Fabio. 2021. "A decade of drinking: temporal trends in apparent household beer intake and standard drink consumption in the United States." *Substance Use & Misuse*, 1–11.

Schonfeld, and SAI Associates Inc. 2021. "Advertising Sales Ratios." *Schonfeld and Associates, Inc.*, https://saibooks.com/advertising–sales–ratios/. (May 2021).

Seiler, Stephan, Anna Tuchman, and Song Yao. 2021. "The impact of soda taxes: Pass-through, tax avoidance, and nutritional effects." *Journal of Marketing Research*, 58(1): 22–49.

Sethuraman, Raj, Gerard J Tellis, and Richard A Briesch. 2011. "How well does advertising work? Generalizations from meta-analysis of brand advertising elasticities." *Journal of Marketing Research*, 48(3): 457–471.

Shapiro, Bradley T. 2018. "Positive spillovers and free riding in advertising of prescription pharmaceuticals: The case of antidepressants." *Journal of Political Economy*, 126(1): 381–437.

Shapiro, Bradley T, Günter J Hitsch, and Anna E Tuchman. 2021. "TV Advertising Effectiveness and Profitability: Generalizable Results from 288 Brands."

Sharma, Lisa, Stephen Teret, and Kelly Brownell. 2010. "The Food Industry and Self-Regulation: Standards to Promote Success and to Avoid Public Health Failures." *American Journal of Public Health*, 100(2): 240–246.

Silver, Lynn D, Shu Wen Ng, Suzanne Ryan-Ibarra, Lindsey Smith Taillie, Marta Induni, Donna R Miles, Jennifer M Poti, and Barry M Popkin. 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.

Simester, Duncan, Yu Hu, Erik Brynjolfsson, and Eric T Anderson. 2009. "Dynamics of retail advertising: Evidence from a field experiment." *Economic Inquiry*, 47(3): 482–499.

Smith-Spangler, Crystal M, Jessie L Juusola, Eva A Enns, Douglas K Owens, and Alan M Garber. 2010. "Population strategies to decrease sodium intake and the burden of cardiovascular disease: a cost-effectiveness analysis." *Annals of internal medicine*, 152(8): 481–487.

Stack, Martin. 2000. "Local and regional breweries in America's brewing industry, 1865 to 1920." *Business History Review*, 74(3): 435–463.

Stack, Martin. 2003. "A concise history of America?s brewing industry." *Economic History Net Encyclopedia of Economic and Business History, available at¡ http://eh. net/encyclopedia/article/stack. brewing. industry. history. us.*

Statista, Statista. 2021. "Market share of the U.S. alcohol industry by beverage 2000-2020." https://www.statista.com/statistics/233699/market–share–revenue–of–the–us–alcohol–industry–by–beverage/. (April 2021).

Sutton, John. 1991. *Sunk costs and market structure: Price competition, advertising, and the evolution of concentration.* MIT press.

Tellis, Gerard J. 1988. "Advertising exposure, loyalty, and brand purchase: A two-stage model of choice." *Journal of marketing research*, 25(2): 134–144.

Tellis, Gerard J, Rajesh K Chandy, and Pattana Thaivanich. 2000. "Which ad works, when, where, and how often? Modeling the effects of direct television advertising." *Journal of Marketing Research*, 37(1): 32–46.

The Beer Global Market Report, Intrado. 2021. "The Beer Global Market Report." https://www.globenewswire.com/en/news–release/2021/04/01/2203506/0/en/Update–Beer–Global–Market–Report–2021–Including–Market–Size–Share–Growth–Analysis–Global–Beer–Consumption–Statistics.html (April 1, 2021).

Tremblay, Victor J, and Carol Horton Tremblay. 2005. *The US brewing industry: data and economic analysis.* MIT Press.

Trieu, Kathy, Bruce Neal, Corinna Hawkes, Elizabeth Dunford, Norm Campbell, Rodrigo Rodriguez-Fernandez, Branka Legetic, Lindsay McLaren, Amanda Barberio, and Jacqui Webster. 2015. "Salt reduction initiatives around the world–a systematic review of progress towards the global target." *PloS one*, 10(7): e0130247.

Tuchman, Anna E. 2019. "Advertising and demand for addictive goods: The effects of e-cigarette advertising." *Marketing Science*, 38(6): 994–1022.

Unnevehr, Laurian J, and Evelina Jagmanaite. 2008. "Getting rid of trans fats in the US diet: policies, incentives and progress." *Food Policy*, 33(6): 497–503.

USA Today, USAToday. 2020. "Most popular beer brands in America dominated by Anheuser-Busch, Molson Coors." https://www.usatoday.com/story/money/2020/03/18/the–most–popular–beer–brands–in–america/111416118/(December 9, 2015).

Waldman, Evelyn. 1956. "Selection of newspapers for split-run tests." *Journal of Marketing*, 20(3): 271–274.

Wang, Emily, Christian Rojas, and Francesca Colantuoni. 2017. "Heterogeneous Behavior, Obesity, and Storability in the Demand for Soft Drinks." *American Journal of Agricultural Economics*, 99(1): 18–33.

Wang, Emily Yucai. 2015. "The impact of soda taxes on consumer welfare: implications of storability and taste heterogeneity." *The RAND Journal of Economics*, 46(2): 409–441.

West, H. 1939. "Split-run copy testing: a practical, inexpensive technique for judging advertisements-how it works and in what magazines." *Advert. Sell.*, , (32): 26–28.

WHO, World Health Organization, et al. 2013. "Global action plan for the prevention and control of noncommunicable diseases 2013-2020."

World Health Organization, WHO. 2021*a*. "Countries share examples of how tobacco tax policies create win-wins for development, health and revenues." https://www.who.int/news–room/feature–stories/detail/countries–share–examples–of–how–tobacco–tax–policies–create–win–wins–for–development–health–and–revenues/. (April 12, 2021).

World Health Organization, WHO. 2021*b*. "Fact Sheets." https://www.who.int/news–room/fact–sheets.

Xu, Yiqing. 2017. "Generalized synthetic control method: Causal inference with interactive fixed effects models." *Political Analysis*, 25(1): 57–76.

Zhong, Yichen, Amy H Auchincloss, Brian K Lee, and Genevieve P Kanter. 2018. "The short-term impacts of the Philadelphia beverage tax on beverage consumption." *American journal of preventive medicine*, 55(1): 26–34.

Zhu, Chen, Rigoberto A Lopez, and Xiaoou Liu. 2016. "Information cost and consumer choices of healthy foods." *American Journal of Agricultural Economics*.

Zubin, Joseph, and John G Peatman. 1945. "Testing the pulling power of advertisements by the split-run copy method." *Journal of Applied Psychology*, 29(1): 40.