UNIVERSITY OF CALIFORNIA,
IRVINE

Essays on Consumer Demand Modeling

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Economics

by

Colin Reinhardt

Dissertation Committee:
Associate Professor Jiawei Chen, Chair
Associate Professor Ivan Jeliazkov
Professor Emeritus David Brownstone

2023

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to take this opportunity to express my deep gratitude and appreciation to the individuals who have contributed significantly to the successful completion of my dissertation. This is a significant milestone in my academic journey, and I am incredibly grateful to everyone who has helped me along the way.

First and foremost, I would like to thank my advisor, Jiawei Chen, for his unwavering guidance, support, and patience throughout my research. His keen insight and critical feedback have been invaluable in shaping my ideas and helping me to develop a coherent and rigorous dissertation.

I am especially grateful for Jiawei Chen's contribution to Chapter 2 of my dissertation which is derived from a joint paper between myself, Professor Chen, and Saad Andalib Syed Shah, and focuses on the econometric analysis of consumer behavior in the face of localized taxation. His expertise and commitment to our research project were instrumental in the success of this chapter. His comments, advice, and constructive criticism helped me to develop my research ideas and refine my writing, and I am grateful for his dedication to my academic growth.

I would also like to extend my appreciation to Ivan Jeliazkov for his invaluable assistance in creating Chapter 3 of my dissertation. Ivan's expertise in discrete choice modeling was critical to the success of my research. He introduced me to the exciting field of discrete choice modeling and provided me with the tools to analyze consumer behavior. I am grateful for the opportunity to work with him and for his generous support and advice throughout my research.

I would also like to thank David Brownstone for his valuable input and feedback during the early stages of my dissertation. His insightful suggestions helped me to shape my research projects and set me on the right path. I am grateful for his guidance, which has been invaluable in my academic success.

I am also grateful to the faculty and staff in our department, who provided me with a supportive and intellectually stimulating environment. I have had the privilege of working with many brilliant scholars in the department, and their insights and perspectives have broadened my horizons and enriched my research.

Finally, I would like to thank my family and friends for their unwavering support and encouragement throughout my academic journey. They have kept me motivated and inspired, and I am grateful for them in every way.

Overall, completing a dissertation is a team effort, and I am incredibly grateful to have had the opportunity to work with Jiawei Chen, Ivan Jeliazkov, and David Brownstone. Their contributions, support, and guidance have been vital in the completion of this dissertation, and I will always be grateful for their help.

VITA

Colin Reinhardt

EDUCATION

Doctor of Philosophy in Economics **2023**
University of California, Irvine *Irvine, California*

Masters of Arts in Economics **2018**
University of Texas at Austin *Austin, TX*

Bacholers of Arts in Communications **2015**
High Point University *High Point, NC*

TEACHING EXPERIENCE

Teaching Assistant **2018–2023**
University of California, Irvine *Irvine, CA*

FIELDS OF STUDY

Industrial Orginization, Econometrics

ABSTRACT OF THE DISSERTATION

Essays on Consumer Demand Modeling

By

Colin Reinhardt

Doctor of Philosophy in Economics

University of California, Irvine, 2023

Associate Professor Jiawei Chen, Chair

This dissertation comprises three chapters. Chapter 1 is based on my unpublished job market paper "Flavorants and Addiction: An Empirical Analysis of Tobacco Product Bans and Taxation." In this chapter, I investigate the potential impact of a proposed menthol cigarette ban on cigarette consumption and product substitution. The use of menthol cigarettes and other tobacco product flavorants has been a contentious issue, but empirical research on the effectiveness of flavorant bans is limited. To fill this gap, I use aggregate-level retail data and micro-level household data to estimate a random coefficient nested logit demand model that incorporates the effects of addiction and consumer heterogeneity. I pay particular attention to the Black American community and low-income households, who have the highest rates of menthol cigarette consumption. My findings show that in the absence of menthol cigarettes, overall cigarette usage reduces by 13%, and the Black smoking rate falls by 35%. In addition, e-cigarettes and cessation products experience a 4.9% and 1.7% increase in demand, respectively. These results are then contrasted with changes in consumption and consumer surplus stemming from a national cigarette tax and a ban on all flavored nicotine products.

Chapter 2 is based on unpublished work in collaboration with Jiawei Chen and Saad Andalib Syed Shah. In this chapter, we introduce a structural choice model that takes into account households' geographic and product substitution behavior to evaluate the impact of localized

taxation policies. Using detailed retail and household data from Philadelphia's soda tax, we estimate the choice model to examine the relationship between households' demographic characteristics, proximity to the city border, and their tax avoidance behavior - such as switching from taxed to untaxed products or from Philadelphia to non-Philadelphia stores. Our findings show that travel time is a crucial factor for modeling households' heterogeneous responses, with an additional minute of travel time equivalent to adding 47¢ to the product price. By factoring in travel costs and the switch to less preferred products, we find that on average, Philadelphia households experience a loss in consumer surplus more than twice the amount of tax paid, with low-income households bearing the greatest burden.

Chapter 3 examines the importance of accounting for realistic substitution patterns in discrete choice models, particularly in the presence of large alternative sets. While multinomial probit models are the preferred method for modeling correlated unobservables across elements of the choice set, their estimation and identification can be challenging. As a result, researchers often resort to alternative methods or restrictive assumptions that disregard substitution patterns. However, ignoring these patterns can lead to biased estimation results. To address this issue, I propose a structured covariance matrix that models substitution patterns as a function of product similarity, allowing for feasible estimation in the presence of large alternative sets. In addition, I incorporate individual parameter heterogeneity and a two-stage consumer decision process, enabling dynamic and individual-level behavior. To estimate the model, I develop a Bayesian MCMC process that utilizes a Tailored Random Block Metropolis Hastings algorithm. Finally, I conduct a simulation experiment to demonstrate the superior performance of my proposed model compared to alternative estimation methods. My results suggest that restrictive substitution patterns can hinder proper estimation of parameter values, emphasizing the importance of considering realistic substitution patterns in choice models.

Chapter 1

Flavorants and Addiction: An Empirical Analysis of Tobacco Product Bans and Taxation

1.1 Introduction

The leading cause of preventable death in the United States is tobacco usage (CDC Smoking and Tobacco Use, 2020). In particular, cigarette consumption is associated with a variety of aliments including a greater risk of bronchitis, heart disease, and cancer, and contributes to one out of every five deaths. This amounts to over 480,000 preventable deaths each year. Within the past several decades, public policy experts in the US and abroad have relentlessly expanded their efforts to curb tobacco consumption. Minimum age limits, advertising restrictions, and heavy taxation are among the tools employed. However, experts concur that more restrictive policies and regulations are necessary, particularly to advance health equity in the face of unethical marketing practices (FDA, 2021).

This paper is focused on a menthol cigarette ban, recently proposed by the Food and Drug Administration (FDA) (FDA, 2021). In particular, I evaluate the expected impact of the removal of mentholated cigarettes on cigarette, e-cigarette, and cessation products (nicotine patches, gum, and lozenges) using data from 2015 to 2019. For this purpose, I construct a model of consumer demand that takes into account the dynamic effect of nicotine addiction, as well as combining available household- and retail-level data in a way that is internally consistent. I account for unobserved preferences and product substitution through the use of both nesting parameters and random coefficients. Finally, I allow household consumption to differ through observed demographic characteristics.

Differences in demand arising from demographic preference for flavorants in tobacco products is an important issue. Historically, the Black American community has long been the target of marketing and advertising practices promoting the use of menthol cigarettes (Gardiner, 2004). Current national estimates of the Black American smoking rate suggests that menthol purchases make up 74% to 89% of their total cigarette purchases; menthol usage is two to three times that of their non-Black peers (Delnevo et al., 2020). In addition, household income has long been correlated with increased price sensitivity and demand for cigarettes, including regular tobacco and menthol (Evans et al. (1999),Wang et al. (2018)). Calls from lawmakers and laypersons seeking to address health inequalities in disadvantaged communities as the result of marketing practices, encouraged the FDA's recently proposed rule prohibiting the sale of menthol cigarettes.

A major consideration when dealing with banning products for health reasons is the willingness of consumers to substitute to equally harmful products. Many menthol smokers in Ontario, Canada, when presented with a local menthol cigarette ban, chose to switch to regular tobacco cigarettes (Chaiton et al., 2020). Further, pre-ban, a fraction of smokers indicated a willingness to consider electronic smoking devices, which also contain nicotine. E-cigarettes, as they are commonly known, are regarded by as a path to smoking cessation, however may also offer a new path to further nicotine addiction (Kasza et al. (2021), Kasza et al. (2022)). I include both e-cigarettes and traditional cessation products in my model, recognizing e-cigarette's role as a substitute for traditional cigarettes as well as its potential to draw nicotine-quitters from more successful cessation products.

The primary goal of my study is twofold: to identify demographic preferences for product flavorants, and to model realistic substitution patterns in the evaluation of the proposed menthol cigarette ban.

In determining the demographic preferences and product substitution patterns, I construct and estimate a model of consumer demand, for cigarettes, e-cigarettes, and cessation products, in the Random Coefficients Nested Logit (RCNL) framework (Grigolon and Verboven, 2014), with a combination of both retail- and household-level data. The use of the random coefficients allows for a rich set of unobserved heterogeneous and observed demographic preferences, and my nesting structure is particularly suited at measuring the degree of substitution across flavors within product categories ("nests"). Further, I adapt the RCNL structure to account for nicotine addiction's dynamic state dependence (e.g. Caves (2005), Tuchman (2019)). Micro-level household purchase data covers only a small subset of total product purchases, but allows for the accurate identification of addiction, consumer heterogeneity, and flavorant substitution. Aggregate-level retail data lacks information necessary to track household-level purchases, but provides a far less noisy measure of price responsiveness and product market shares, and provides a reliable method to account for endogenous model parameters. I use the availability of household and retail data to my advantage, incorporating them in my modeling procedure in such a way as to be internally consistent, and combine the strength of both datasets.

My estimation follows that described in Grieco et al. (2022), adapted for the RCNL structure with dynamic state dependence. This procedure allows me to recover mean utility and unobserved demand shocks, while accounting for household heterogeneity, addiction, and categorical substitution.[1]

Several key findings result from my parameter estimation. (1) I find that the willingness to switch among product flavorings differs significantly between cigarettes and e-cigarettes, and plays a key role in determining the effectiveness of the various bans considered in my model. Menthol and tobacco cigarettes were found to be closer substitutes for each other when compared to the substitution rate between e-cigarette flavorants. (2) I identify addiction, in the form of dynamic state dependence, to play a significant role in repeated purchasing behavior. (3) Demographic differences strongly determine product preference and consumption behavior. I find Black Americans display greater demand for menthol and flavored products, and low-income households exhibit significantly higher rates of cigarette usage.

Conditioned on the results from my structural estimation, I examine several counterfactual scenarios. (1) My model predicts that weekly cigarette smoking rates would have been 13% lower, on average, during the period from April 2015 to April 2019, with the removal of mentholated cigarettes. Black Americans, in particular, would have experienced a 35% drop in expected weekly smoking rates, during this period, with the removal of menthol cigarettes. (2) In contrast to a menthol cigarette ban, I find a 10.23% cigarette sales tax to be as effective in lowering the average weekly smoking rate, with a smaller reduction in average consumer surplus across all demographic groups—including low-income households, thereby demonstrating greater preference for taxation. In addition, a back-of-the-envelope calculation finds a 10.23% cigarette sales tax would have resulted in an expected weekly tax revenue of $66.1 million, for a total of $1.41 billion over the period from April 2015 to April 2019. (3) Finally, I find that by expanding the flavorant ban to include menthol and flavored e-cigarettes over this same period results in a reduction in weekly cigarette smoking rates similar to the menthol cigarette ban alone, as well as a drop in average weekly e-cigarette usage ranging up to 46% dependent upon supply side assumptions.

[1] Several other works have adapted the same or similar procedures, including Goolsbee and Petrin (2004), Chintagunta and Dubé (2005), Tuchman (2019), and Murry and Zhou (2020).

Interest in flavorant bans has grown alongside the popularity of flavored e-cigarette nicotine products, although to date, research addressing the effects of a ban on menthol and other flavorants remains limited. In regard to a menthol cigarette ban, most empirical research involves either questionnaires of consumer intent (Levy et al. (2021a)) or the study of bans imposed in other countries (Chaiton et al. (2020), East et al. (2022), Fong et al. (2022a)). As such, expectations as to the impact of the proposed menthol ban on US smoking rates relies on projections based upon these works, e.g. Levy et al. (2021b), Fong et al. (2022a), and Issabakhsh et al. (2022). Using Canadian data, Fong et al. (2022a) finds an expected decrease of 7.3% reduction in the number of US smokers. In contrast, both Levy et al. (2021b) and Issabakhsh et al. (2022) rely on the same expert elicitation of consumer intent post ban (Levy et al., 2021a), and these works find an expected reduction in smoking cigarette rates of 15% among all consumers and 35.7% when focusing on the Black American community. I complement these existing works by using both retail-level and household-level data to estimate consumer behavior and preference for flavorants, and by conducting counterfactual analyses based on my structural estimation results.

To the best of my knowledge, Olesiński (2020) is the only structural model examining the impact of a mentholated cigarette ban on consumer demand. However, his results and counterfactual analysis pertain to Polish consumers, and provides an ex ante evaluation of the 2020 European Union menthol ban. I rely on US aggregate- and individual-level data, ranging from 2015 to 2019. Furthermore, my modeling structure differs, in that the inclusion of household-level data allows for a richer set of heterogeneous preferences and I account for addiction in the form of dynamic state dependence.

Current literature of addiction commonly considers two modeling formats: "rational addiction" with forward-looking behavior and myopic models. Myopic models allow for past consumption to affect current consumer behavior, but future consequences of addiction play no role in determining one's actions (Houthakker and Taylor (1970), Mullahy (1985)). Furthermore, under the myopic modeling framework, increases in current and past prices reduce current consumption, while increases in future prices will not affect current consumption (Baltagi and Levin (1986), Jones (1989), Baltagi and Levin (1992)). On the other hand, “rational-addiction” models contend that consumers consider future prices and consequences when making current consumption choices (Becker and Murphy (1988), Gordon and Sun (2015)). Researchers, such as Winston (1980) and Akerlof (1991), have objected to the assumption of perfect foresight present in rational-addiction models. More recently, Hidayat and Thabrany (2011) found rational addiction models inadequate in explaining behavior related to cigarette usage; instead, their findings favor myopic modeling assumptions. In my own work, allowing for forward-looking behavior would inhibit my ability to combine the household- and retail-level data in a way that is internally consistent; therefore, I rely on a myopic framework as detailed in Caves (2005) and Tuchman (2019).

The remainder of this work proceeds as follows. In Section 1.2, I introduce the background information regarding the history of flavored nicotine products, and the reasoning underlying the currently proposed menthol ban. Section 1.3 describes my data sources and provides details on products, households, and markets. Section 1.4 provides supportive evidence, formed from both retail and household data, of preference heterogeneity, product substitution, and addiction. Section 1.5 details the discrete choice model of demand which incorporates addiction as well as both household and retail data. In Section 1.6 I discuss parameter identification and estimation. Model results are presented in Section 1.7. Counterfactual simulations providing changes in product consumption rates under product bans and taxation are provided in Section 1.8. Section 1.9 concludes this work.

1.2 Industry Background

The tobacco industry has long been a creative with product development and marketing, much to the detriment of public health. Industry innovations have included cigarette length and width (with ultra long and ultra slim), filters, low-tar tobacco, and a finer control of nicotine content; The introduction of product flavorants began with the country-wide sale of mentholated tobacco in 1927 and, in 1999, the mass production of flavored (fruity, candy, and mint) cigarettes (Toll and Ling (2005), Mills et al. (2018)). Fueled by the desire for greater market share, industry research conducted by Big Tobacco led to fine-tuned innovations targeting specific consumer groups.[2] Slim cigarettes (in particular "Virginia Slims") are regarded as the first and most successful female-oriented cigarette brand, menthol cigarette print and billboard advertising has been found to primarily target the Black American community, and archived tobacco industry documents detail the development of sweet, fruity, and candy-like flavors to target young smokers (Cumminggs (1999), University of California San Francisco (1999), Toll and Ling (2005), Mills et al. (2018)).

In the past two decades, rising health concerns and increasing negative public opinion towards tobacco products has led to the introduction of tobacco control regulations. In particular, the advent of product bans started with the mass introduction of flavored cigarettes in the early 2000s and the subsequent public outcry. From 1999 to 2006, three flavored products were introduced to the US market by well established tobacco companies, and quickly rose to public prominence—Camel Exotic Blends, Kool's Smooth Fusions, and Salem's Silver label (Lewis and Wackowski, 2006). Decades of research into youth consumption and preference for flavored products by industry powerhouses, such as Philip Morris, R.J. Reynolds, and Brown & Williamson, encouraged this product development. Flavored cigarettes quickly became popular among young smokers, and while overall cigarette sales fell, market shares of flavored products rose—defying the national downward trend (Cumminggs (1999), Lewis

[2] Big Tobacco is a name used to refer to the largest companies in the tobacco industry.

and Wackowski (2006)). However, public concerns over increasing youth tobacco usage pressured congress to action.

The Family Smoking Prevention and Tobacco Control Act, signed into law June 22, 2009, by President Barack Obama, provided the FDA the power to regulate the tobacco industry, and marked the first ban on flavored (fruity, candy, and mint) cigarettes some months later. The Act also prohibited advertising to children and required tobacco companies to obtain FDA approval for new tobacco products.

A mere decade later saw the next proposed flavorant ban, this time in relation to youth e-cigarettes usage. The introduction of more stylish pod system e-cigarettes, innovative social media marketing campaigns, and the promotion of flavored products, particularly to the youth and young adults, contributed to over a 300% increase in e-cigarette unit sales during the period from January 2015 through July 2019 (Nardone et al., 2019).[3] Sales of Juul, the most common pod-based e-cigarette, surged over 600% and contributed much to the overall rise in e-cigarettes during this period, and Juul became the company with the single greatest e-cigarette market share by the end of 2017 (Ali et al., 2020). Juul's small size, sleek USB styled design, variety of flavors, and subtle scent made it particularly appealing to young users (Lee et al. (2020), Vallone et al. (2020)). The term "JUULing" soon became synonymous with the discrete usage of e-cigarettes by teenagers in classrooms, school yards, or restrooms (Ramamurthi et al., 2019).

Concerns over this increased youth e-cigaratte smoking pushed the FDA to act; in January 2020, a ban was placed on the sale of all flavored (fruity, candy, and mint) e-cigarette cartridges. While the ban on flavored e-cigarette cartridges was intended to reduce youth consumption, regulators failed to include disposable style e-cigarettes. And, although beyond the scope of this paper, current research suggest consumers—particularly the young consumers—simply switched to these disposable products (Hickman and Jaspers, 2022).

[3]See Figure 1.2.

In 2022, the FDA proposed a new ban on menthol cigarettes. Similar to the prior two product bans, regulators sought to advance health equity by reducing tobacco-related health disparities and addiction, particularly among disproportionately affected, menthol using, minority communities. Thus, as I shift the focus to mentholated tobacco, I will revisit the theme that flavorants attract specific and potentially vulnerable populations.

1.2.1 Menthol Cigarettes

In 1925, the first menthol cigarette was created by Lloyd "Spud" Hughes who, seeking to alleviate the symptoms of a cold, placed loose tobacco in a tin of medical menthol crystals overnight (Lee and Glantz, 2011). The next day, he found the resulting smoke soothing to his throat; the mentholated cigarette providing a more pleasant, "cooler", experience. Hughes later patented his invention and, after selling the patent to the Axton-Fisher Tobacco Company in 1927, "Spud Menthol Cooled Cigarettes" would remain the sole mentholated nicotine product until the introduction of Kool menthol cigarettes in 1933, by Brown & Williamson.

For the next two decades, Kool became the industry leader in menthol cigarettes; however, during this time, mentholated products represented only 3% of the overall cigarette market (Lee and Glantz, 2011). However, post WWII, Big Tobacco saw new opportunity among the Black American community; a new, wealthier, urban Black community was growing. By the 1960s, advertising of specialized products—shampoo, skin creams, etc.—targeted towards this burgeoning community began in earnest.

Following the years of post-war growth, Black media had reached record-breaking levels. Over 600 radio stations now catered to Black audiences, where less than two decades prior there were only 20, and readership of Ebony magazine, the leader in Black print media, was at an all-time high (Pollay et al., 1992). The surge in print, radio, and television con-

Figure 1.1: Print Advertising of Menthol Cigarettes Targeting the Black Community

(a) Kent Menthol Cigarette Ad (1961) (b) Winston Menthol Cigarette Ad (1970)

sumption among Black audiences was a prime opportunity for the advertising of menthol products by Big Tobacco. Research by Gardiner (2004) found that, by 1962, Ebony magazine contained twice as many menthol advertisements as the similarly popular, among white communities, Life magazine. Despite some initial advertising to white clientele, Black communities soon became the primary focus of mentholated cigarettes; Black American smoking rates of menthol products skyrocketed from 14% in 1968 to 44% by 1975 (Gardiner, 2004).

Today, the impact of race based marketing in the Black community remains clear. Despite a fall in overall smoking rates, Black consumers still display a preference for menthol products at rates 2 to 3 times their non-Black peers (Delnevo et al., 2020). Further, although Black Americans make up approximately 12% of the population, they contribute to about 40% of all menthol related tobacco deaths (CDC Smoking and Tobacco Use, 2020). In acknowledgement of past wrongs, and to reduce further cigarette consumption, the FDA proposed, in April 22, 2022, new product standards to prohibit menthol as a flavorant in cigarettes. To quote acting FDA commissioner Janet Woodcock, M.D., "With these actions, the FDA will help significantly reduce youth initiation, increase the chances of smoking cessation among current smokers, and address health disparities experienced by communities of color, low-income populations, and LGBTQ+ individuals, all of whom are far more likely to use these tobacco products." (FDA, 2021)

1.3 Data

In this section, I provide details pertaining to my retail and household data. In addition, I describe my markets of interest, including demographic information and the formation of retail market shares from available data.

1.3.1 Retail Data

I use the Nielsen retail datasets which cover the period from January 1st, 2015 to July 31st, 2019.[4] Sales information is available for the entirety of 2019, however I do not use the months post July, as some brands began to engage in the voluntary removal of flavored cartridge products in an attempt to appease e-cigarette critics. The data contains store-level information detailing weekly price and quantity sold at the Universal Product Code (UPC) level. Recorded sales include my three primary categories of interest: cigarettes, e-cigarettes, and smoking cessation products (nicotine lozenges, gum, and patches). At the store level, I observe unique location identifiers. I choose to focus on 26,916 stores active every year during the entirety of the period studied.

Products within the e-cigarette category contained a mixture of battery units, starter kits,

[4] All Nielsen material discussed herein was obtained from the Kilts Center for Marketing Data at The University of Chicago Booth School of Business.

refill cartridges, disposable e-cigarettes, and flavored e-juice. In my analysis, I remove from consideration those UPCs pertaining to battery units, starter kits, and flavored e-juice. Battery units and starter kits were removed because these products contain only, or include, the rechargeable smoking device used with refill cartridges. These purchases are generally not repeat, and are significantly more costly. E-juice, on the other hand, contains greater variation in terms of price, inconsistent sizing, and nicotine content. In contrast, cartridge packs and disposable e-cigarettes have standardized quantities, similar prices, and account for 89% of category-level unit sales.

Sold in 3 to 5 cartridge packs, each refill cartridge contains a nicotine content generally equivalent to 1-1.5 cigarette packs, and are priced around \$3 to \$5 a cartridge. I find disposable e-cigarettes are generally sold individually or in packs of 10; each unit contain a nicotine content equivalent to 1-1.5 cigarette packs and are generally priced around \$5 to \$10 per unit. Traditional cigarettes are sold in packs of 20 cigarettes or 10 pack cartons, and prices range from \$3.50 to \$15 a pack, depending upon marketing strategies, and federal, state, and local tax. Finally, smoking cessation products such as nicotine lozenges and gum are sold in sizes ranging from 20 to 100 pieces, with a nicotine content of either 2 mg or 4 mg per piece. I weight the sizes of lozenges and gum to a standardized 4 mg per piece, with 15 pieces costing about \$8.50 and providing the about same nicotine as one cigarette pack. Nicotine patches are most commonly sold in packs of 7 or 14; one patch provides a nicotine content equivalent to 1 pack of cigarettes and costs around \$4. In my subsequent analysis, I consider a pack of cigarettes equivalent to one e-cigarette cartridge, one disposable unit, 15 pieces of 4 mg nicotine gum/lozenges or a single nicotine patch, and I adjust product prices for inflation.[5]

Nielsen's retail datasets also provide information pertaining to product flavor in almost all cases—excepting some e-cigarettes. When product flavor was unavailable, I proceeded

[5] I adjust price to its January 2015 dollar value using the Consumer Price Index for all Urban consumers (CPI-U).

with manual identification. There are 10,344 unique cigarette UPCs (5,667 regular tobacco and 4,677 menthol), 1,630 unique e-cigarette UPCs (668 regular tobacco, 493 menthol, and 469 flavored), and 668 unique smoking cessation product UPCs. Among cigarettes and e-cigarettes, all major brands (overall market share $\geq 1\%$) offer tobacco, menthol, and—in the case of e-cigarettes—flavored product varieties. For the remainder of this work, I aggregate UPCs into products, where each product is a category/flavor combination, and the size of each product is standardized to that equivalent to one pack of cigarettes.

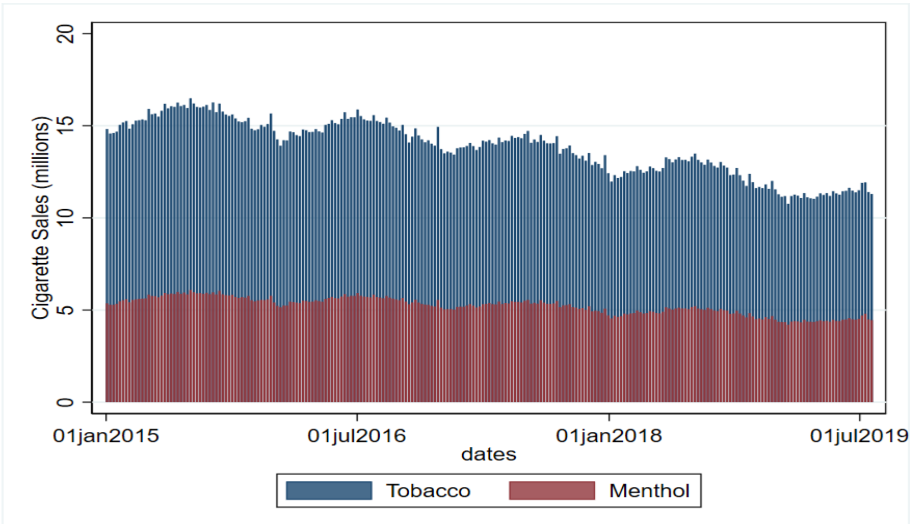
Figure 1.2 plots the trends in cigarette and e-cigarette sales by flavor type from January 2015 through July 2019. Sales from 26,916 stores generate the referenced figure. The plots demonstrate seasonality in cigarette sales, and an overall negative trend. As for E-cigarettes, sales were steadily increasing until January 2018, when a period of rapid growth began, driven primarily by flavored products.

1.3.2 Household Data

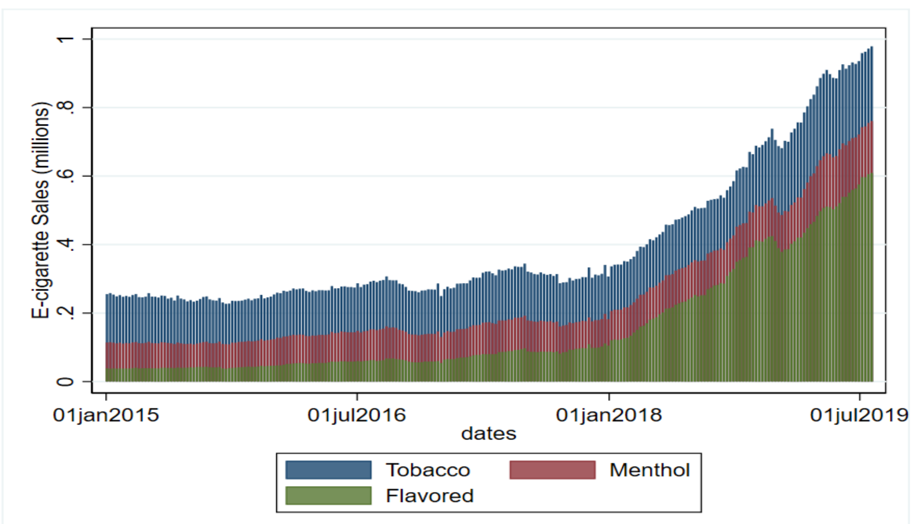
Nielsen provides household purchase data for a sample of US consumers totaling about 50,000 households yearly. Information provided includes cigarette, e-cigarette, and smoking cessation purchases, as well as a household's home county and other demographic data. Pertaining to purchases, I am provided with records that include price, date, quantity, and the unique store identifier where the sale took place, if available.

Between January 2015 and July 2019, I record 17,420 households who engaged in a total of 401,718 purchases of our products of interest. Given the available demographic data, I first generate an indicator for those households recorded as having the racial characteristic "Black (non-Hispanic)"; in my subsequent analysis, this indicator allows me to ascertain the impact of proposed policy changes on the Black American community. I focus on Black as a primary racial characteristic of interest because there exists a well documented difference

Figure 1.2: Weekly Sales Quantities for Cigarettes and E-cigarettes



(a) Weekly Cigarette Sales



(b) Weekly E-cigarette Sales

in preferences among the Black American community, particularly in regard to menthol cigarettes.

Next, I differentiate between low- and high-income households through the use of an indicator variable denoting low income. I define low-income households to be those whose yearly household income falls within 200% of the 2019 federal poverty guideline, which takes into account household size.[6] Table 1.1 reports the joint distribution of households by race and Income.[7] Finally, the average weekly cigarette smoking rate among all households within my panel is 14.7%.

Table 1.1: Household Panel Joint Distribution of Race and Income[a]

	High Income	**Low Income**	**Total**
Black	**6.02%** (6.89%)	**3.97%** (5.66%)	**9.98%** (12.55%)
Non-Black	**54.63%** (63.92%)	**35.39%** (23.54%)	**90.02%** (87.46%)
Total	**60.64%** (70.81%)	**39.36%** (29.20%)	

[a] U.S. household joint distribution included in parentheses for comparison purposes.

1.3.3 Market Formation

I define my markets based upon the Designated Marketing Areas (DMAs) provided by Nielsen. As defined, a DMA consists of a group of counties displaying similar regional characteristics and are considered radio and television markets. Often centered around major metropolitan areas, there exist 210 DMA regions covering the entire continental US, Hawaii,

[6]Nielsen does not provide a continuous measure of household income; I define low-income households to be those falling below the category closest to twice the poverty federal guideline—this difference is never grater than $2,500.

[7]The joint distribution of race and income status for my household data may not match that suggested by the ACS, however by conditioning on these observables the resulting selection bias is removed (see Grieco et al. (2022)).

and parts of Alaska. Defining my markets based upon DMA regions provides several advantages: (1) Nielsen and their provided datasets already contain identifying information as to DMA assignment for both retailers and households. (2) DMAs are generally centered around large urban populations and include surrounding suburban and rural counties—reducing biases that could be present if one only considered, say, major city centers. (3) DMAs form regions of households with similar characteristics and define television/radio markets, therefore demand shocks should be similar across consumers—particularly those stemming from advertising campaigns run at the DMA level.

I begin market formation by first determining total sales and quantity weighted prices at the product/DMA/week level using unique identifiers provided in the store-level data.[8] Next, for population and demographic data, I rely on the 2019 American Community Survey (ACS) 5-year estimates. Note that DMA regions are proprietary to Nielsen; however, from my available retail data, I obtain a list of counties specific to each of the 206 DMAs in which I observe store-level sales. Total household population, defined by racial group, is accessible at the county level provided in the 2019 ACS 5-year estimates. However, to obtain the joint distribution of income status by race, I rely upon the Public Use Microdata Sample from the 2019 ACS 5-year estimates available at the public use microdata area (PUMA) level. I obtain the county-level joint distribution of income status by race as the weighted average of overlapping PUMAs using the PUMA-county crosswalk file from the Missouri Census Data Center.[9] Finally, from the county-level population estimates and joint distribution of income status by race, I obtain county-level population defined by race and income status.

From county specific population distributions defined by race and income, I aggregate to

[8]Similar to Tuchman (2019), my analysis is performed at the week level; I find the average time between purchases, among current smokers, to be less than one week and I do not find significant evidence of stockpiling behavior. For more information, see Appendix A1

[9]I should note that the public use micro sample data could be used to obtain the joint distribution of race and income. However, to avoid introducing greater error via the PUMA to the county conversion, I only calculate the proportion of low-income consumers by race. Data pertaining to the population distribution, in addition to total population, comes from the county-level 2019 ACS 5-year estimates.

the DMA level. A final hurdle arises from determining DMA weekly market shares. My Nielsen retail sample forms a subset of the available stores in each DMA; I do not observe all sales. Therefore, I cannot simply divide observed sales by total population to obtain shares. Instead, I turn to available information pertaining to cigarette smoking rates; countyhealthrankings.org, operated by the University of Wisconsin and Robert Wood Johnson foundation, which provide yearly expected county-level smoking rates for all counties for the years 2016, 2017 and 2018. With this data, I form expected DMA-level smoking rates as the population weighted average of the county-level smoking rates. Then, for each DMA I weight the population such that weekly cigarette market shares best fit DMA expected smoking rates.[10] [11]

My final market sample consists of 100 DMAs with the largest pre-weighted populations, and which displayed strictly positive market shares over all weeks. This provides three major benefits: (1) remaining DMAs form pricing instruments (i.e., Hausman-style instruments as seen in Nevo (2001)), (2) zero market shares complicate estimation, and (3) model runtime is significantly reduced. The markets forming my model provide a mix of all regions—from major urban city centers to rural communities; Finally, 85% of my household sample, 86% of my store sample, and 85% of the US population exist within these 100 DMAs.

1.4 Descriptive Analysis

In this section, I provide supportive evidence for my selection of demographic coefficients through the use of reduced form estimation, figures, and graphs. I also explore the impact of addictive behavior on product selection as supportive evidence for the inclusion of this dynamic element.

[10]The DMA specific population weight applies to all weeks and years; I do not adjust the weight weekly, nor yearly.

[11]My formation of DMA-level weekly product usage rates preclude the existence of illicit sales; I discuss in the impact of this limitation in Appendix A5.

1.4.1 Retail Evidence of Preference Heterogeneity

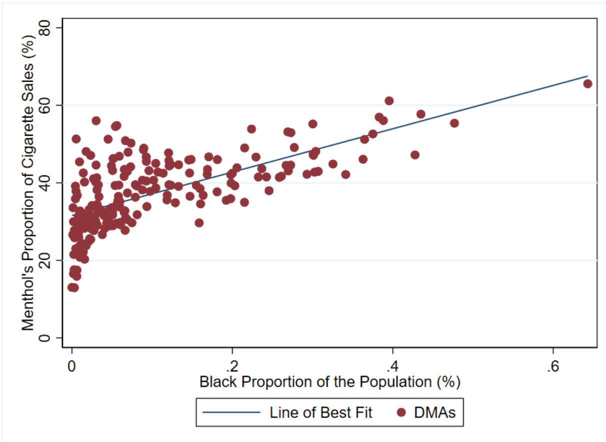
Throughout my analysis, I rely on two primary demographic attributes: income and the prevalence of Black consumers. Prior empirical work provides support for the selection of these demographic variables, especially when considering rates of smoking behavior and the removal of menthol products. I begin by documenting potential systematic differences, or lack thereof, in consumer preferences along these demographic profiles in Figure 1.3.

I show the preference for menthol by considering its proportion of cigarette sales based on each demographic trait—the share of menthol sales to the Black population and the share of menthol sales to low-income households. Demographic preferences for e-cigarettes are similarly considered, however in this case I assign each DMA into quartiles based on the share of each demographic trait. Markets with a greater proportion of Black households have substantially higher sales of menthol cigarettes—not surprising given prior research. However, when considering e-cigarettes, differences in flavorant preference are not apparent between regions of high and low Black populations. In areas with a greater proportion of low-income consumers there is a slight preference for regular tobacco cigarettes, however these regions display a significantly greater demand for regular tobacco e-cigarettes.[12] Finally, I display differences in category preference by observed DMA demographic characteristics in Table 1.4.

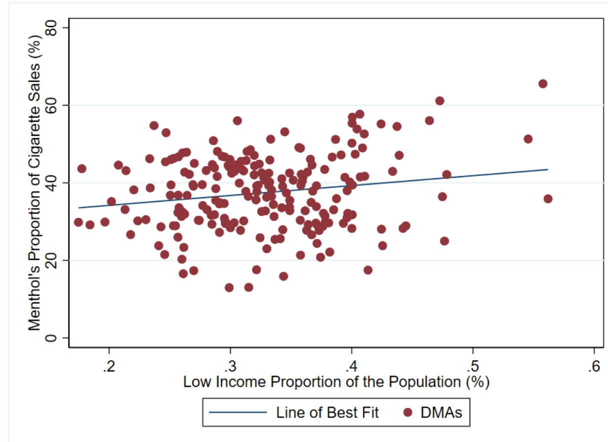
As before, I show differences in preference by assigning markets into quartiles based on each demographic trait. Where the column "High" represents those DMAs falling in the top quartile, and "Low" represents those in the bottom. Markets with a greater share of Black households have substantially higher sales of cigarettes, whereas DMAs with low Black populations display greater preference for cessation products. Areas with a larger proportion of low-income consumers, similar to those with high Black populations, prefer

[12] To avoid confusion, I define the flavor "regular tobacco" to consist of cigarettes and e-cigarettes whose flavor profile is solely tobacco.

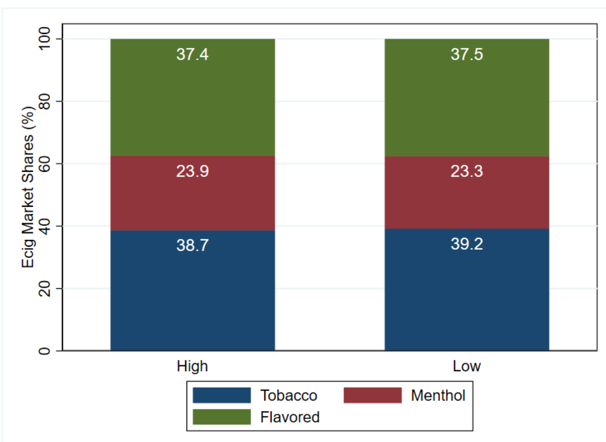
Figure 1.3: Flavorant Choice and DMA Demographics



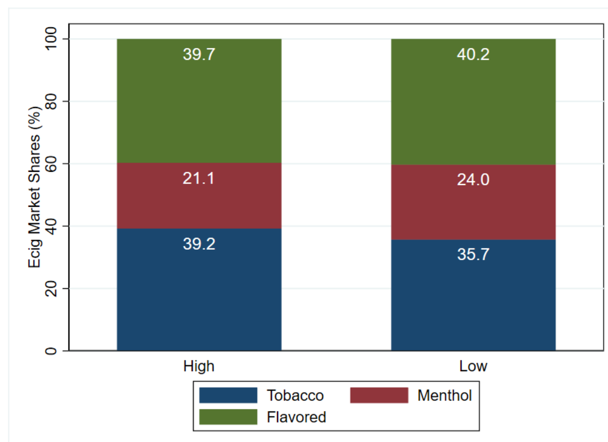
(a) Black and Menthol Cigarette Consumption



(b) Low Income and Menthol Cigarette Consumption



(c) Black and E-cigarette Flavor Choice

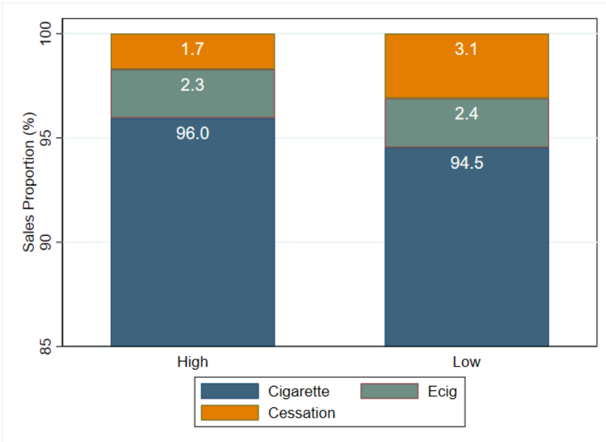


(d) Low Income and E-cigarette Flavor Choice

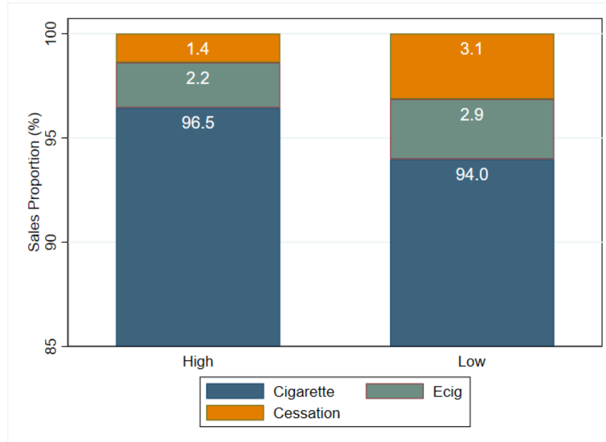
Notes: In the bottom two figures, I compare markets in the top (”High”) and bottom (”Low”) quartile of each demographic trait. For Figure (c) ”High” represents those DMAs with the greatest proportion of Black consumers and for Figure (d) ”High” denotes those DMAs with the highest proportion of low-income households. The top two figures plot each market as a function of demographic attributes and the proportion of menthol sales. These figures are generated from the 206 DMA’s in which I observe store-level sales.

cigarettes. Finally, as a regions’ wealth increases, so does the proportion of sales involving cessation products and e-cigarettes.

Figure 1.4: Category Choice and DMA Demographics



(a) Black and Category Choice



(b) Low Income and Category Choice

Notes: In the two figures, I compare markets in the top ("High") and bottom ("Low") quartile of each demographic trait. For Figure (a) "High" represents those DMAs with the greatest proportion of Black consumers and for Figure (b) "High" denotes those DMAs with the highest proportion of low-income households. As cigarettes are by far the most popular product, for display purposes both figures start at a y-intercept of 85. These figures are generated from the 206 DMA's in which I observe store-level sales.

1.4.2 Household Evidence of Substitution, Addiction, and Flavorant Heterogeneity

In this subsection, I first present evidence of product substitution through the use of a matrix describing the transitional probability of product purchase. Next, I document consumer addiction through the use of a linear probability model, controlling for time and individual fixed effects. Lastly, I provide figures demonstrating heterogeneous responsiveness in product choice, similar to those shown above. As before, my demographic covariates of interest are Black and low income.

Product Substitution Table 1.2 provides the probability of observing product choice conditional on the last observed inside option purchased. I focus on the last observed inside option purchased, rather than the prior week's purchase, to highlight household product substitution and heterogeneous preference; I discuss weekly continuation of product usage and addiction later in this subsection. The last observed product purchased makes up the first column; each subsequent column provides the conditional probability of transitioning from the last observed purchase to the current product choice, provided a consumer decides to choose an inside option. If a consumer decides not to choose an inside option, then their last observed purchase remains unchanged.

Table 1.2: Product Transition Table

	Current Product Choice					
Last Inside		Cigarette		E-cigarette		
Option Purchased	Cessation	Tobacco	Menthol	Tobacco	Menthol	Flavored
Cessation	**75.48**	15.12	8.36	0.61	0.18	0.24
Cig. Tobacco	0.26	**93.10**	6.03	0.37	0.07	0.16
Cig. Menthol	0.24	10.81	**88.36**	0.10	0.31	0.17
Ecig. Tobacco	0.61	22.12	2.91	**66.78**	1.96	5.61
Ecig. Menthol	0.30	7.82	16.20	3.99	**64.68**	7.01
Ecig. Flavored	0.26	14.62	7.21	8.52	7.84	**61.55**

Notes: In the above table, I present the probability of current product choice ("Current Product Choice") conditioned upon the last observed product chosen ("Last Inside Option Purchased").

If households did not display inherent preferences, then the current purchase probability should be independent of a consumer's last product choice. That is to say, the probability of choice j in the current time period conditional on purchasing choice a in the past should be the same as if they had instead purchased product b. This is obviously not the case observed in Table 1.2. Instead, a consumer's preferred product choice is that of their last observed purchase. This persistence in consumption is strongest among cigarette users, where subsequent purchases almost always consist of the preceding product. The willingness for consumers to switch products within the cigarette category is an important consideration regarding the proposed menthol ban. Individual-level data suggests, when cigarette smokers switch products, it is primarily to an alternative flavor within the same product category—supporting the notion of within-nest substitution among cigarette users.

E-cigarette users also demonstrate persistence in product preference, albeit, not nearly to the degree observed among cigarette smokers. Furthermore, the second most popular present choice of product for past e-cigarette smokers was cigarettes. If products are switched, users of regular tobacco and flavored e-cigarettes preferred to switch to regular tobacco cigarettes instead of products within the e-cigarette category. Smokers of menthol e-cigarettes, when switching products, generally choose menthol cigarettes as their alternative choice—suggesting persistent preference for menthol products. These findings suggest degrees of within-category substitution differ between cigarettes and e-cigarettes.

Unfortunately, for individuals dedicated to smoking cessation, I find that nearly 24% of all purchases of cessation products are eventually followed by a choice of cigarettes. Furthermore, although it is small, there does appear to be a willingness for users of cessation products to switch to e-cigarettes; the probability of choosing e-cigarettes grows in the latter half of the sample, as e-cigarettes rise in popularity, and consumers looking to quit smoking may consider e-cigarettes a viable substitute to other cessation products. Regardless of the methods by which one may attempt to quit, the presence of addiction is clear.

Addiction and Dynamic Dependency Table 1.3 provides an illustration of the dependent nature of nicotine laced products. Testing for the presence of addiction as well as other forms of dynamic state dependency, I analyze the weekly consumption habits of the 17,420 households in my household dataset. Specifically, I consider how past consumption of a nicotine containing product influences future product choice through the use of a linear probability model. To control for individual preferences, time trends, and seasonality, I include household and time fixed effects, and cluster the errors at the household level.

Considering my regression results, I find that consumption in the prior week plays a positive and significant role in determining the probability of purchasing in the current period. This result is unsurprising, as on average 53% of all purchases immediately follow consumption in the prior week. A key result of this regression is supportive evidence that state

Table 1.3: Linear Regression on the Probability of Purchasing

	Coefficient
Purchase in Prior Week	0.104***
	(0.003)
HH FEs	Y
Week FEs	Y
Mean DV	.112
Num HH	17,420
Num Obs	2,622,559

***p<.01, **p<.05, *p<.1

Standard errors clustered at the household level are included in parentheses.

dependence—possibly in the form of addiction—plays a significant role in determining the choice to purchase. However, the impact of prior consumption on the probability of purchasing appears to differ by category.

Table 1.4 presents current categorical choice based upon the prior week's purchase decision. Unlike the transition table presenting product substitution (Table 1.2), Table 1.4 displays current categorical choice as a function of a household's purchase decision during the preceding week, and includes the outside option to highlight how state dependence may differ between categories. I find cessation product purchases are followed by a choice of outside option 78% of the time. Whereas, in the week following a cigarette purchase, households choose the outside option only 47% of the time. Similarly, 50% of all e-cigarette purchases in my household data are followed by a choice of inside option during the next week. These findings, coupled with those displayed in Table 1.3, suggest that dynamic state dependency may differ by category choice in the prior week, affecting both the probability of purchasing an inside option as well as the current choice of product.

Flavorant Preference Finally, in the consideration of within-category choice, I present Figure 1.5 which illuminates a household's flavorant preference dependent on their observed

Table 1.4: Categorical Purchase Probability by Week

Last Week's	**Current Category Choice**			
Category Choice	Outside Op.	Cessation	Cigarettes	E-cigarettes
Outside Op.	**91.47**	0.14	8.20	0.19
Cessation	78.27	**15.88**	5.58	0.26
Cigarettes	46.52	0.08	**53.09**	0.31
E-Cigarettes	49.57	0.16	12.40	**37.86**

Notes: In the above table, I present the probability of current category choice conditioned upon the category choice made during the prior week ("Last Week's Category Choice").

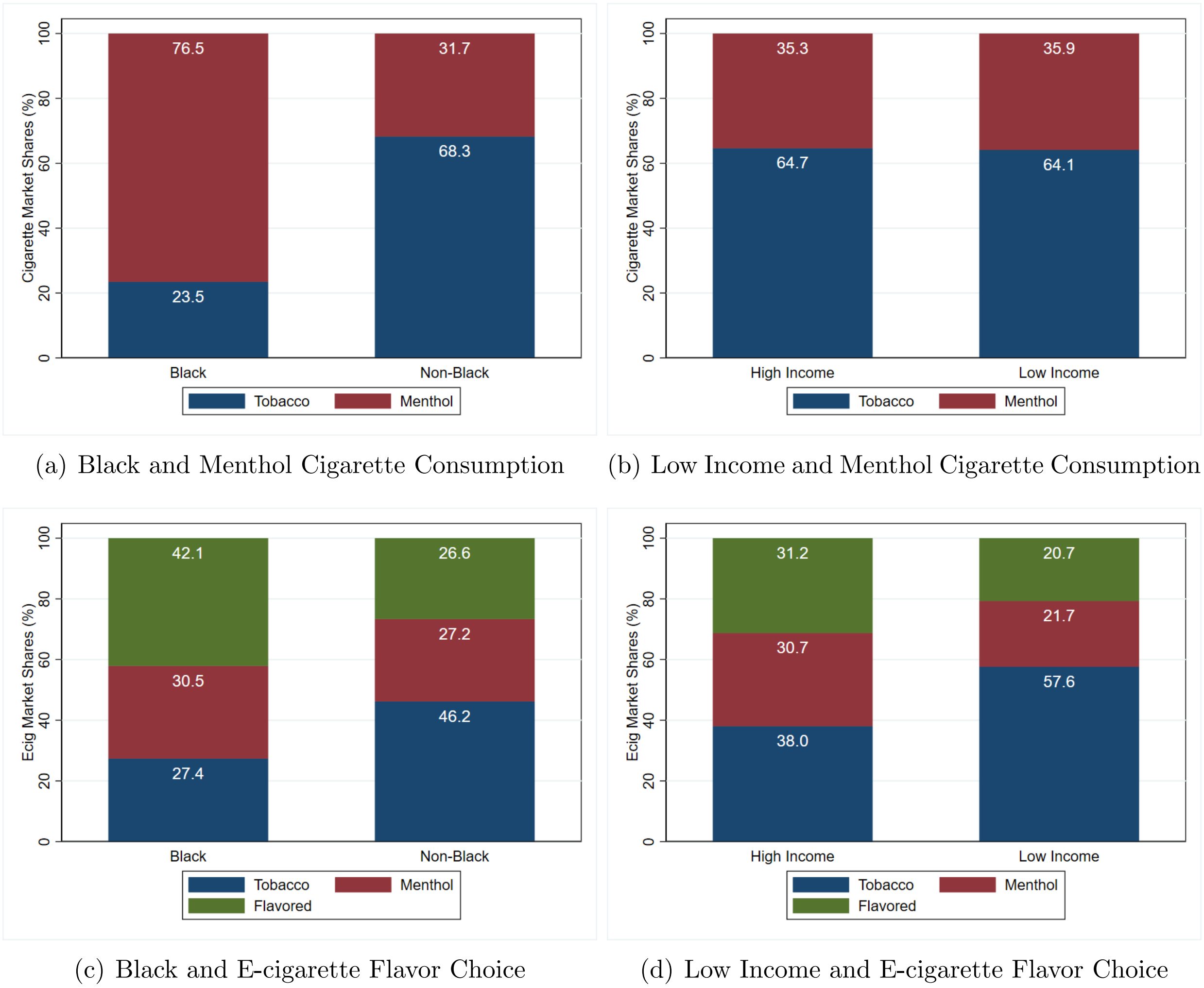
demographic attributes. Similar to the figures in Subsection 1.4.1, I provide bar charts by demographic status providing the sales proportion by flavorant for cigarettes and e-cigarettes.

As observed in the DMA-level data, Black households display a strong preference for menthol cigarettes, with 77% of cigarette purchases by Black consumers consisting of menthol products. Additionally, high and low-income household preference for menthol products appears near identical—similar to the results found in the DMA sales data. In regard to e-cigarettes, both Black and high-income consumers display increased preferences for flavored and menthol products—shunning regular tobacco e-cigarettes. For low-income consumers, this result is similar to that suggested above (Figure 1.3); however, Black households display a clear flavorant preference—for menthol and flavored products—that was not apparent in the retail-level data. This finding stresses the importance of household-level information, and its ability to present a markedly less noisy reference as to demographic product preference.

1.5 Choice Model

I follow the literature detailing demand estimation employing retail-level data (e.g. Berry et al. (1995), Nevo (2000), etc.) in modeling the demand for cigarettes, e-cigarettes, and

Figure 1.5: Product Choice and Household Demographics



smoking cessation products as a function of product characteristics, heterogeneous consumers, demographic information, and addiction. I adjust traditional methods to exploit the availability of household data (Similar to Chintagunta and Dubé (2005),Goolsbee and Petrin (2004), Murry and Zhou (2020), etc.). My work extends the model of addiction, proposed in Tuchman (2019), through the use of a nested framework, inclusion of product flavorants, and modeling of demographic responses. Lastly, my estimation procedure differs in methodology from that performed in Tuchman (2019); rather, I adapt the work of Grieco et al. (2022) in designing my estimation procedure.[13]

The use of retail datasets, coupled with household datasets, allows me to leverage the benefits of both. Specifically, retail data measures demand responsiveness with less noise—particularly for sparsely purchased products. In addition, the retail modeling structure provides a reliable method by which one can account for parameter endogeneity. Household data provides a more accurate estimation of heterogeneity, substitution, and addiction. Therefore, the model I propose utilizes both datasets to their full potential in a way that is internally consistent.

1.5.1 Demand Specification

Let $\mathcal{J}$ represent the set of available products denoted $j = 1, \ldots, J$, where $J = |\mathcal{J}|$, and let $\mathcal{G}$ represent the set of product categories ("nests") denoted $g = 1, \ldots, G$, where $G = |\mathcal{G}|$. Furthermore, consider the outside option to be choice $j = 0$ and a member of group $g = 0$. Then, at the individual level, in week t, a consumer i living in market m obtains the indirect utility from purchasing product $j \in \mathcal{J}$, where product j is a member of group $g \in \mathcal{G}$, given by

$$\begin{aligned} u_{ijmt} =& x_j'\beta_i + \alpha_i p_{jmt} + h_{gmt}'\gamma + \phi \mathbb{I}\big(\sum_{g' \in \mathcal{G}} C_{ig',t-1} > 0\big) + \rho_g C_{ig,t-1} + \xi_{jmt} + \bar{\epsilon}_{ijmt} \\ & \text{where } i = 1, \ldots, H;\ j = 1, \ldots, J;\ t = 1, \ldots, T;\ m = 1, \ldots, M. \end{aligned} \tag{1.1}$$

The $n_1 \times 1$ vector of product characteristics x_j includes elements such as category and flavor constants. Retail price for product j in market m and time t is p_{jmt}. The $n_2 \times 1$ vector h_{jmt} contains market/category and time/category fixed effects. $\mathbb{I}(\cdot)$ is an indicator function,

[13] Tuchman (2019) follows a process described in Chintagunta and Dubé (2005), which involves a four-step estimation procedure; iterating between a maximum likelihood step and the inversion described in Berry et al. (1995). I find in testing that, through the inclusion of numerical gradients, the estimation procedure developed in Grieco et al. (2022) provides a faster and more reliable estimation of the parameters of interest.

and, at the individual level, $C_{ig,t-1}$ signifies the choice of group g, by consumer i, in the prior week.[14] Therefore, ϕ captures the change in demand common across all inside options provided consumption of any nicotine product during the prior week, and ρ_g captures state dependency at the category level. Finally, $\bar{\epsilon}_{ijmt}$ denotes unobserved individual preferences for product j, in market m, at time t, and I allow for common variation in consumer utility through the use of demand shocks (ξ_{jmt}) unobserved by the researcher—but known to the consumer.

I characterize a consumer i living in market m through the use of a $(n + 1) \times d$ matrix of observed demographic attributes, D_i, including race and income. I model unobserved individual preference heterogeneity for product characteristics, v_i, through the use of a multivariate normal distribution; this assumption alleviates the burden imposed by the Independence of Irrelevant Alternatives (IIA) propriety of the logit model. Preferences for product characteristics and prices are as follows:

$$\begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \Pi D_i + \Sigma v_i, \qquad v_i \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{n_1+1}), \tag{1.2}$$

where Π is a $(n_1 + 1) \times d$ matrix that measures the impact of observable demographic attributes on the preference for product characteristics, while Σ captures the covariance of unobserved individual preferences for product characteristics. In practice, I restrict $\Sigma_{jk} = 0$ $\forall j \neq k$, and estimate only the variance of unobserved preference for characteristics.

Furthermore, I follow the work of Grigolon and Verboven (2014) in assuming that unobserved individual preferences for products are correlated across goods of the same category. In my analysis, I observe $G = 3$ product categories: cigarettes, e-cigarettes, and cessation

[14] In essence, it is possible to consider addiction as lasting multiple weeks. However, doing so significantly increases modeling complexity and runtime—particularly in regard to the gradient estimation. In the retail data step, when evaluating the gradient, state dependency requires I model the propagation of changes in demand, over time, due to changes in parameter values—this calculation is computationally intensive. My current procedure strikes a balance between modeling complexity and runtime. I discuss cases of multiple observed product purchases in Subsection 1.5.2

products. Within each category, flavor defines the set of available products. In the case of cigarettes, available flavors are regular tobacco and menthol. E-cigarettes are available in regular tobacco, menthol and flavored products (e.g., fruit, candy, and mint). The choice of cessation, having no within category options, represents a degenerate nest. Finally, my outside option is defined to be group zero. Thus, the unobserved individual preferences, $\bar{\epsilon}_{ijmt}$, for product j, which falls in category g, follows the distributional assumption of a two-level nested logit model, and can be decomposed into

$$\bar{\epsilon}_{ijmt} = \zeta_{igmt} + (1 - \lambda_g)\epsilon_{ijmt} \tag{1.3}$$

where ϵ_{ijt} is iid Type-1 extreme value, the nesting parameter $\lambda_g \in [0, 1]$, and ζ_{igt} has a (unique) distribution such that $\bar{\epsilon}_{ijt}$ is distributed Type-1 extreme value.

The random coefficient nested logit (RCNL) model, described in equations (1.2) and (1.3), can encompass a variety of demand specifications—allowing for correlation in both observed and unobserved preferences. Within nest, perfect substitution is obtained under the case where the category-level nesting parameter equals one. As the category-level nesting parameter tends toward zero, the model reduces to the standard random coefficient specification. Lastly, in modeling different values of λ_g for each category, I allow for products within different nests to display varying degrees of within-group substitution. When accounting for consumer heterogeneity, it is useful to decompose indirect consumer utility into its common, δ_{jmt}, and idiosyncratic, μ_{ijmt}, components (excluding $\bar{\epsilon}_{ijmt}$):

$$\begin{aligned}
&\delta_{jmt} = x_j'\beta + \alpha p_{jmt} + h_{gmt}'\gamma + \xi_{jmt}, \\
&\mu_{ijmt}(\mathbf{C}_{i,t-1}) = \left[x_j', p_{jmt}\right](\Pi D_i + \Sigma v_i) + \phi \mathbb{I}\Big(\sum_{g' \in \mathcal{G}} C_{ig',t-1} > 0\Big) + \rho_g C_{ig,t-1},
\end{aligned} \tag{1.4}$$

where $\mathbf{C}_{i,t-1} = (C_{i0,t-1}, C_{i1,t-1}, \ldots C_{ig,t-1}, \ldots C_{iG,t-1})'$.[15]

[15]Note, $C_{i0,t-1}$ denotes the decision by household i to choose the outside option during the preceeding week.

Thus the probability of a consumer i, living in market m, purchasing product j during time period t is then

$$\pi_{ijmt}(\mathbf{C}_{i,t-1}) = \frac{\exp\Big(\frac{\delta_{jmt}+\mu_{ijmt}(\mathbf{C}_{i,t-1})}{(1-\lambda_g)}\Big)}{\exp\Big(\frac{I_{igmt}(\mathbf{C}_{i,t-1})}{(1-\lambda_g)}\Big)} \times \frac{\exp\big(I_{igmt}(\mathbf{C}_{i,t-1})\big)}{\exp\big(I_{imt}(\mathbf{C}_{i,t-1})\big)}, \tag{1.5}$$

where, after denoting the set of choices available in group g as $\mathcal{J}_g$,

$$I_{igmt}(\mathbf{C}_{i,t-1}) = (1-\lambda_g)\log\sum_{j\in\mathcal{J}_g}\exp\Big(\frac{\delta_{jmt}+\mu_{ijmt}(\mathbf{C}_{i,t-1})}{(1-\lambda_g)}\Big), \tag{1.6}$$

$$I_{imt}(\mathbf{C}_{i,t-1}) = \log\Big(1+\sum_{g\in\mathcal{G}}\exp\Big(I_{igmt}(\mathbf{C}_{i,t-1})\Big)\Big). \tag{1.7}$$

The final equation includes the group composed of the outside option—as the utility from the decision not to purchase is normalized to 0, it is the source of the "1".

1.5.2 Consumer Choice Probabilities

In the household dataset, I consider a consumer i choosing to purchase product j at the weekly level—matching the weekly data available at the retail level. When focusing on household purchases, I do not consider quantity and instead consider purchase incidence—whether at least one unit was consumed. To do otherwise would require strong assumptions to make the model tractable, as retail data does not provide information pertaining to consumer purchase quantities. Furthermore, I derive my retail market shares from observed smoking rates; as such, my model is one of changes in smoking behavior rather than purchase quantities. In the case of multiple distinct products consumed during a single week, I generate duplicate entries for each.[16] Again, to do otherwise is beyond the scope of my model, and this assumption is one innately made by a researcher working solely with retail data (i.e.

[16]Duplicate entries make up less than .02% of weekly observed household-level choices.

Berry et al. (1995), Nevo (2000), etc).

Turning now to the individual choice probabilities, for ease of notation, I consider $\Theta = (\Sigma, \Pi, \phi, \rho_q, \rho_c, \rho_e, \lambda_c, \lambda_e)$. The parameters ρ_q, ρ_c, and ρ_e provide the impact of category-level state dependence for cessation products, cigarettes, and e-cigarettes, respectively. Furthermore, λ_c and λ_e denote the nesting parameters for cigarettes and e-cigarettes, respectively.[17] After integrating out the distribution of unobserved individual attributes, denoted $F_v(v_i)$, the density of a consumer's observed sequence of choices is given by

$$L_i(Y_i|x, p_m, h_m, D_i; \delta, \Theta) = \int \prod_{t=1}^{T_i} \prod_{j=1}^{J} [\pi_{ijmt}(x, p_{mt}, h_{mt}, \delta_{mt}, \mathbf{C}_{i,t-1}, \Theta, D_i, v_i)]^{y_{ijt}} \, dF_v(v_i),$$
$$\text{where } \delta_{mt} = (\delta_{1mt}, \ldots, \delta_{Jmt})', \; x = (x_1, \ldots, x_J)', p_{mt} = (p_{1mt}, \ldots, p_{Jmt})',$$
$$\text{and } h_t = (h_{1mt}, \ldots, h_{Jmt})'. \tag{1.8}$$

I denote Y_i as the observed sequence of a consumer's choices where $y_{ijt} = 1$ if consumer i, living in market m, chooses product j during time period t.

1.5.3 Retail Market Shares

Unlike individual consumer choice probabilities, deriving market shares from aggregate retail sales data introduces a component of difficulty; namely, I do not observe a consumer's prior choice of product. Instead, I am provided with weekly sales data transformed into product-level market shares, which are a function of individual-level smoking behavior. As such, under the assumption of consumer homogeneity for ease of explanation, retail market shares are formed as follows:

$$s_{jmt} = \sum_{g=0}^{G} \pi_{jmt}(C_{g,t-1} = 1) P(C_{g,t-1} = 1), \tag{1.9}$$

[17] As the choice of cessation products is a degenerate nest, it requires no nesting parameter.

where s_{jmt} denotes the market shares of product j in market m and time period t. $P(C_{g,t-1} = 1)$ signifies the probability that group g was chosen in the prior period, and evolves each period according to a simple recursion equation. Under the assumption of homogeneous consumers, the probability of having made a choice contained in group g, this week, is equal to the sum of observed choice shares made within group g, across all possible category decisions from the prior week. Thus, solving for this probability simple;

$$P(C_{g,t} = 1) = \sum_{j \in \mathcal{J}_g} \sum_{g'=0}^{G} \pi_{jmt}(C_{g',t-1} = 1) P(C_{g',t-1} = 1). \tag{1.10}$$

In application, I rely upon the assumption of consumer heterogeneity such that the simulated retail shares now take the form

$$s_{jmt} = \int\limits_{v_m} \int\limits_{D_m} \sum_{g=0}^{G} \pi_{ijmt}(C_{ig,t-1} = 1) P(C_{ig,t-1} = 1) dF_D(D_i) dF_v(v_i). \tag{1.11}$$

I now integrate over the distribution of observable and unobservable consumer attributes denoted $F_D(D_i)$ and $F_v(v_i)$, respectively. In practice, I evaluate the above integrals by Monte Carlo simulation through the use of Halton draws from the empirical distribution of v and D.[18] For each market m, I draw R simulated consumers and evaluate their choices over time such that

$$s_{jmt} = \frac{1}{R} \sum_{R} \sum_{g=0}^{G} \pi_{rjmt}(C_{rg,t-1} = 1) P(C_{rg,t-1} = 1). \tag{1.12}$$

From Eq. (1.10), it follows that for each simulated consumer r the probability of them having

[18] A Halton sequence is a low-discrepancy quasi-random number sequence. See Train (1999).

made a choice in group g, during the current week, is simply

$$P(C_{rg,t}=1)=\sum_{j\in\mathcal{J}_g}\sum_{g'=0}^{G}\pi_{rjmt}(C_{rg',t-1}=1)P(C_{rg',t-1}=1). \tag{1.13}$$

In this context, Eq. (1.13) provides an evolving joint distribution of consumer heterogeneity and consumption status that is easily derived. This recursion equation demonstrates that the consumption behavior of a simulated consumer r relies on each prior time period. Therefore, when performing my demand estimation, I require an initial distribution of consumption status, which I cover in Subsection 1.6.1.

1.6 Identification and Estimation

My objective is to estimate the parameter vectors α, β, γ and Θ corresponding to the mean responses, demographic interactions, unobserved taste heterogeneity, addiction, and nesting parameters. While I am not necessarily interested in the values of δ, they provide the means by which I can recover my mean taste parameters. My estimation proceeds through a two-step process: first, I maximize the individual likelihood function through the use of my household and retail data, and then I perform a two stage least squares (TSLS) regression to estimate my mean utility parameters, α, β and γ.

I rely on a Hausman-style instrument, as used in Nevo (2001), to control for price endogeneity. My identifying assumption is that, by conditioning on market/category and time/category fixed effects, market-specific demand shocks are independent across DMAs. Given this assumption, the average product price across regions not included in my estimation will be independent of my market's demand shocks, but this average will be correlated with my observed prices due to common marginal costs.[19]

[19]In Appendix A3 I compare my model predicted mean utility coefficients with and without the use of my pricing Instrument.

1.6.1 Maximum Likelihood Estimation

Given Eq. (1.8), for any candidate values of δ and Θ the log likelihood of the household data is

$$\mathcal{L}(Y;\delta,\Theta) = \sum_{i=1}^{H} log[L_i(Y_i|x, p_m, h_m, D_i; \delta, \Theta)]. \tag{1.14}$$

In theory, one can estimate δ directly via maximum likelihood, requiring only household data; in practice, this is computationally infeasible.[20] Instead, I rely upon the work of Berry (1994), who shows that for any given value of θ, there exists a unique vector δ such that predicted shares from Eq. (1.12) exactly match those observed in the retail dataset. Thereby, I treat δ as a known function of θ provided retail market shares—as is common practice in discrete choice demand estimation with retail data (Berry et al. (1995), Nevo (2000), Nevo (2001)).

Thus, the log likelihood of the household data, Eq. (1.14), can be rewritten as

$$\mathcal{L}(Y;\delta,\Theta) = \sum_{i=1}^{H} log[L_i(Y_i|x, p_m, h_m, D_i; \delta(\Theta), \Theta)], \tag{1.15}$$

where $\delta(\Theta)$ is provided by the contraction mapping specified in Grigolon and Verboven (2014). When evaluating simulated retail market shares during the contraction mapping step (Eq. (1.12)) I make $R = 200$ Halton draws, per market, from the empirical distribution of v and D. In each time period, the joint distribution of consumer heterogeneity and consumption status, for my simulated consumers, evolves according to the Eq. (1.13).

Consequently, to perform the contraction mapping, I require an initial distribution of consumption status for each simulated consumer. Two techniques govern this decision: (1) I specify the initial distribution as a parameter of interest to be estimated, or (2) I provide

[20] In the household dataset there are many product/time/market combinations lacking observed product purchases, rendering product/time/market-level identification of δ impossible when reliant solely on individual-level data.

an arbitrary initial distribution and forward simulate during a burn-in period (Erdem et al. (2003), Hendel and Nevo (2006b), Tuchman (2019)). I utilize the second procedure, treating the first quarter of 2015 as my burn in period, and provide the initial arbitrary joint distribution as $P(C_{rg1} = 1) = 1/(G + 1)$, $\forall r \in R$. Tests of other arbitrary initial distributions demonstrate convergence to the same steady state well within my allotted burn-in period. Finally, Appendix A2 provides more detail regarding how a unique vector of $\delta(\Theta)$ is derived from my retail data.

After obtaining $\delta(\Theta)$ for a given value of Θ, I evaluate the integral governing the density of a household's observed sequence of choices (Eq. (1.8)) via Monte Carlo simulation. In practice, I utilize 100 Halton draws from the empirical distribution of v.[21] My estimation procedure then searches over the values of Θ that maximize Eq. (1.15).[22] Upon obtaining optimum values, I calculate robust standard errors for $\widehat{\Theta}$ as described in Train (2009), p. 201; sandwiching the covariance of the household-level gradient between the inverted Hessian at the optimum of the likelihood function.

1.6.2 Mean Utility Coefficients

Given $\widehat{\Theta}$ from the maximum likelihood step, the resulting unique vector $\widehat{\delta}$ provides the relationship between a product's mean utility and our covariates of interest—see Eq. (1.4). In my evaluation of this relationship, I proceed with a TSLS regression relying upon the Hausman style instruments discussed above. Standard errors for $(\widehat{\beta}, \widehat{\alpha}, \widehat{\gamma})$ are calculated using a two-step bootstrap procedure where estimation error from the maximum likelihood step is captured by the first stage of the procedure, and the second step accounts for typical sampling er-

[21]Results from Train (1999) show simulation variance with 100 Halton draws to be lower than 1000 random draws in a mixed logit application with a similar number of random coefficients.

[22]My tolerance during the contraction mapping step is set to $1e^{-13}$. For the likelihood maximization algorithm, I set a tolerance of $2e^{-10}$ and provide computed numerical gradients. I consider several randomized starting values when proceeding with the maximization algorithm to rule out local minima. Finally, the RCNL contraction mapping requires a dampening procedure discussed in Grigolon and Verboven (2014).

ror. I begin by taking $B = 1000$ draws from the asymptotic distribution of Θ found in subsection 1.6.1. Next, for each of the thousand draws, Θ_b, I find the corresponding vector, $\delta(\Theta_b)$, and sample with replacement from the set $\{(\delta_{111}(\Theta_b), x_1, p_{111}, h_{111}), \ldots, (\delta_{JMT}(\Theta_b), x_J, p_{JMT}, h_{JMT})\}$ to create a bootstrapped sample of a size equal to the original. Given the bootstrapped sample, I then perform the TSLS regression to estimate $(\beta_b^*, \alpha_b^*, \gamma_b^*)$. Finally, from the distribution of $(\beta_b^*, \alpha_b^*, \gamma_b^*)$, I find the standard errors of my mean utility parameters.

1.7 Estimation Results

Table 1.5 presents the demand estimates of my model's preferred specification using the two-stage process described above. In total, I have 100 markets with 226 time periods each (after removing the burn-in weeks per Subsection 1.6.1) for a total of 135,600 market-level observations.[23] At the individual level, I have 14,712 households (residing in the 100 markets) for a total of 2,100,709 household observations post burn-in. To control for common time and market specific demand shocks, my estimation includes fixed effects at the category/time and category/market level. For presentation purposes, and to avoid perfect collinearity, I exclude the constant for regular tobacco flavor, the final time period, and the last market; these then form my reference categories for my cigarette, e-cigarette, and cessation category level constants.

Dummies representing product flavorant are denoted Menthol and Flavored. Flavored products are only available in the form of disposable or cartridge based e-cigarettes; however, menthol products are available for both e-cigarettes and traditional cigarettes. To account for heterogeneous flavorant preferences across product categories, I include an interaction of menthol and e-cigarettes. On average, consumer valuations of tobacco products exceed that of menthol, but, in terms of e-cigarettes, flavored products are the most preferred on

[23]After burning the first quarter of my sample; the time frame considered under my demand analysis ranges from April 2015 through July 2019.

Table 1.5: RCNL Demand Estimates.[a]

		Means (β)	Std. Dev. (σ)	Demographic Interactions (Π)	
				Low Income	**Black**
Price		-0.759***		-0.017	
		(0.094)		(0.026)	
Cigarette		1.303**	2.036***	0.351**	-0.700***
		(0.606)	(0.028)	(0.164)	(0.090)
E-cigarette		-4.771***	2.281***	0.365*	-1.929***
		(0.352)	(0.075)	(0.220)	(0.329)
Cessation		-1.749**	2.805***		
		(0.889)	(0.086)		
Menthol		-0.718***	1.188***	0.118***	1.055***
		(0.051)	(0.054)	(0.029)	(0.062)
Menthol × Ecig.		-0.348***			
		(0.042)			
Flavored		0.451***		-0.397*	1.040***
		(0.078)		(0.213)	(0.319)
Past Consumption	(ϕ)	0.247***			
		(0.096)			
Cess State Depedence	(ρ_q)	0.958***			
		(0.204)			
Cig State Depedence	(ρ_c)	0.405***			
		(0.099)			
E-cig State Depedence	(ρ_e)	2.672***			
		(0.166)			
Cigarette Nest	(λ_c)	0.768***			
		(0.013)			
E-cigarette Nest	(λ_e)	0.357***			
		(0.086)			
Cat. × Time FEs		Y			
Cat. × Market FEs		Y			
Num HH		14,712			
Num HH Obs		2,100,709			
Num Markets		100			
Num Time Periods		226			
Num Market Level Obs		135,600			

***p<.01, **p<.05, *p<.1

[a] Standard errors are included in parentheses. My estimation includes fixed effects at the category/time and category/market level. For presentation purposes, and to avoid perfect collinearity, I exclude the flavor regular tobacco, the final time period, and the last market—making these my reference levels for the cigarette, e-cigarette, and cessation category dummies. Finally, I explored the inclusion of demographic interactions with cessation as well as a two-layered nesting structure comprising an upper nest (cigarettes, e-cigarettes, cessation) and lower nest choice of flavor, however such changes did not improve model fit.

average. Finally, as expected, average product valuation decreases with price.

Demographic Interactions In addition to average consumer valuation, I allow for a rich set of heterogeneous parameters to account for variations in preferences across demographic groups. The estimates of Π reveals significant variation in demographic valuation. Low-income consumers display greater preference for cigarettes and e-cigarettes and, interestingly, I do not find statistically significant evidence of differences in average price valuation for low-income households. Racial disparities in demand for cigarettes mirror those found in other works (Sakuma et al. (2016), Sakuma et al. (2020)); Black household demand for cigarettes and e-cigarettes is less than that of other consumer types. Preference for flavorants also varies across demographic groups; Black households strongly favor menthol and flavored products. In contrast, while low-income consumers display a slight preference for menthol, other flavored products are less preferred.

Random Coefficients and State Dependence Turning to the estimates of my random coefficients (Σ), all are statistically significant, and account for variation in valuation across households. In addition, past consumption of an inside option plays a positive and significant role in determining consumption status across all product offerings. This result is consistent with the presence of addictive behavior in nicotine containing products. However, dynamic state dependency appears to be primarily focused at the category level, with the values of categorical state dependency (ρ_g) nearly 2 to 10 times larger than the effect of past consumption on the demand for all inside options (ϕ).

Notably, cessation products and e-cigarettes demonstrate the greatest degree of state dependence. For cessation products, I find ρ_q to be twice that of the state dependent parameter for cigarettes, ρ_c, and, in the case of e-cigarettes, ρ_e is roughly 6 times larger than ρ_c. I hypothesize that the differences in state dependency between product categories arises from consumer learning behavior, particularly for goods with small market shares or, in the case of e-cigarettes, products newly introduced.[24] My indicators of prior consumption status may also capture forms of structural state dependence, such as loyalty behavior.[25]

Nesting Parameters I also obtain significant estimates of my nesting parameters for cigarettes and e-cigarettes (λ_c and λ_e), indicating that products of the same category are considered closer substitutes. Interestingly, I find the nesting parameter for cigarettes is greater than twice that of e-cigarettes. This suggests degrees of within-nest substitution differ between product categories. Households consider tobacco and menthol to be close substitutes, whereas e-cigarette flavorants are not held in the same regard. To corroborate this point, I calculate short-run own-price and cross-price elasticities of demand; capturing consumer responsiveness to a one-time price increase during the same week.

Table 1.6: Price Elasticity of Demand.[a]

Average Level		**Own**	**Cross-Elasticity**		
			Same Category	Different Category	All Products
Cigarettes	Tobacco	-4.028	1.682	0.006	0.341
	Menthol	-4.724	2.581	0.006	0.521
	Average	**-4.376**	**2.132**	**0.006**	**0.431**
E-Cigarettes	Tobacco	-4.077	0.854	0.121	0.414
	Menthol	-4.085	0.820	0.178	0.435
	Flavored	-5.153	0.914	0.118	0.436
	Average	**-4.438**	**0.863**	**0.139**	**0.429**
	Cessation	-5.487	-	0.086	0.086

[a] The table above reports own and cross-elasticities at the product and category average level. Cross-elasticities are averaged across products from the same category, different categories, and across all products.

[24] I examined, and did not find, a statistical difference in e-cigarette state dependence pre- and post-2018.

[25] My estimates of ϕ and ρ reflects the net effect of all forms of state dependence.

Price Elasticity Table 1.6 provides the price elasticity of demand. The cross-price elasticity between the focal product and other products is averaged across three groups: the focal product and those that share its same category, the focal product and those in different categories, and the focal product and all other products. Finally, I present own-price and cross-price elasticities of demand at the product and category average level. Consider, the cross-price elasticities of demand averaged across products within the same category compared to the average across products from a different category; tobacco and menthol cigarettes are far more responsive to changes in other product prices when those products exist within the same nest. Similarly, the cross-price elasticity of e-cigarettes is greater when averaged across products within the same nest when compared to the average across products in alternative categories. My cross-elasticity calculations provide supportive evidence of within nest substitution for both cigarettes and e-cigarettes, and suggests sensible substitution patterns across products.

Model estimates imply category average own-price elasticities of demand for cigarettes and e-cigarettes to be -4.376 and -4.438, respectively. In comparison to cessation products, cigarettes and e-cigarettes are generally less elastic. I find that markets with a greater proportion of low-income households have, on average, less elastic demand for cigarettes. This finding is generally in line with literature demonstrating persistence in cigarette consumption among low-income consumers. In terms of product flavorant, I find demand for menthol cigarettes the least elastic in markets with the greatest Black American populations. Interestingly, market-level average own-price elasticity for e-cigarettes—regardless of flavor—does not appear to be significantly correlated with the proportion of low-income households nor Black consumers. Lastly, demand for cessation products is the least elastic in those markets with the greatest percentage of high-income households—suggesting a persistence in preference for cessation products among wealthier consumers. Overall, my calculated own-price elasticities provide sensible variation along consumer demographic distributions and are consistent with consumption differences displayed in Subsection 1.4.1.

1.8 Counterfactual Product Bans and Taxation

I now use my estimates of cigarette, e-cigarette and cessation product demand to measure the effect of the proposed menthol ban—in addition to other counterfactuals. Thus, I can evaluate consumer responsiveness to various product bans, and to provide a taxation rate which results in consumption-level changes equivalent to that resulting from the removal of menthol products. I proceed by first describing my supply-side model, and the assumptions I impose while performing my analysis. Then, provided estimates of counterfactual prices from my supply-side model, I present expected changes in consumption behavior resulting from my varied counterfactual scenarios.

1.8.1 Supply-Side Model

I begin my model of supply-side behavior by considering multi-product firms interested in maximizing their profits. Generating a full supply-side model with true forward-thinking firm behavior would be exceedingly complex given the presence of dynamic state dependence. I simplify by considering firms to be interested in maximizing profits over the finite sum time-periods included in my sample, and I rely upon the fact that changes in consumption behavior resulting from price changes made weeks prior tend towards zero as time progresses. Thus, when considering optimal prices for a given week, I find firms place almost no weight on the resulting changes for profits occurring a quarter or more in the future. As such, in my analysis, only counterfactual prices calculated towards the final weeks of my sample would inherit bias resulting from my specifying a finite time problem (as opposed to considering profit maximization over an infinite number of periods). In practice, I drop the final quarter of my counterfactual analysis, analogous to how I rely upon a burn-in period when forward simulating in my maximum likelihood estimation (see Subsection 1.6.1).[26]

[26] After burning the first quarter and last quarter of my sample, the time frame considered under my counterfactual analysis ranges from April 2015 through April 2019.

Table 1.7: E-Cigarette Brands by Market Share

Brand	**Market Share**	**Owner**
Blu	24.02%	Imperial Brands*
Juul	23.40%	Juul, Altria* (35% Post Dec. 2018)
NJOY	18.22%	NJOY
Logic	11.08%	Logic, JTI* (Post April 2015)
Vuse	7.23%	R. J. Reynolds*
21st Century Smoke	5.46%	21st Century Smoke
FIN	4.23%	FIN
Mark Ten	2.51%	Altria*
Mistic	2.26%	Ballantyne
Other	1.59%	Other

* Tobacco company.

Next, of note, is my decision to consider firms as either operating at the nest level, or to consider a single firm as the producer of both cigarettes and e-cigarettes.[27] If I model firms at the nest level, then I contend that competition exists between products of differing nests, and that manufactures of cigarettes, for instance, do not likewise produce e-cigarettes. Otherwise, I could consider cigarettes and e-cigarettes to be owned and produced by a singular entity interested in maximizing the collective sum of their profits. Consider Table 1.7, which presents brand-level market shares for e-cigarettes sold between January 2015 and July 2019.

I observe that prior to 2019, 55.16% of e-cigarettes sold were by companies not directly owned or operated by Big Tobacco. With the purchase of a 35% stake in Juul by Altria (previously known as Phillip Morris) in late December 2018, the proportion of independent producers fell to 31.76%. The trend towards e-cigarette acquisition by large multinational tobacco manufactures is not surprising. Initially, the industry was composed of small independent companies interested, primarily, in producing products to assist in smoking cessation be-

[27] My choice of modeling cigarettes and e-cigarettes as a composite product inhibits my modeling assumptions. I can either consider producers of cigarettes and e-cigarettes as competitive firms or a single entity.

havior, but Big Tobacco's entry into the market during the early 2010s changed producer incentives, and lead to growing market concentration among the largest players (University of Bath, 2012).

To compensate for both the independence of firms and the growth in market concentration, I consider two versions of my supply-side model. The first defines firms at the nest level (cigarette and e-cigarette producers considered as competitors), and the second models the total acquisition of e-cigarette producers by Big Tobacco, e.g. one firm producing both products. Thus, my findings can be perceived as providing bounds for possible firm responses based upon the proportion of market concentration under traditional producers of tobacco products. Throughout both models of my supply-side analysis, I assume the producers of cessation products are now, and continue to be, independent. Finally, a detailed description of my counterfactual price estimation is provided in Appendix A4.

1.8.2 Counterfactual Simulations

This subsection begins with the proposed menthol cigarette ban; I report expected changes in cigarette and e-cigarette consumption by demographic profile, as well as the average change in cessation product usage upon removal of all non-tobacco cigarettes. Next, I calculate an average national sales tax that results in a similar reduction in smoking rates as those expected under the menthol ban—weighing the pros and cons of bans vs taxation. Lastly, I explore the expansion of the menthol ban to all, non-tobacco, product flavorants—paying particular attention to the expected reduction in e-cigarette usage. All counterfactual scenarios considered in my model rely on supply-side estimates of counterfactual price discussed above, in Subsection 1.8.1.

To obtain average weekly usage rates, I impose my counterfactual scenarios beginning in 2015, and simulate weekly demand over the following four and a half years, for each simulated consumer r. Weighting my counterfactual shares by the market population, and averaging over each week, I determine the weekly average rate of consumption for all products—across all markets. I burn the first and last quarter of my results, and average across all weeks to determine the average change in product usage over the period from April 2015 through April 2019.

Menthol Cigarette Ban Table 1.8 presents the impact of the removal of mentholated cigarettes from a household's choice set. I display smoking rates for cigarettes and e-cigarettes by demographic profile; cessation usage rates are presented as the average across all households. Changes in consumption behavior are displayed under the assumption of both independent and merged (cigarette and e-cigarette) producers. I find that in the absence of menthol cigarettes, weekly cigarette smoking rates reduce, across all households, by 13% (from 15.72 to 13.74 percent) regardless of producer merger status. On average, 67.5% of menthol smokers switched to tobacco cigarettes upon removal of mentholated product offerings; expected consumer surplus, across all households, falls by 15.7 to 15.9% (dependent on merger status) compared to the status quo.

Among Black households, the average reduction in cigarette consumption is far higher; a 35% drop in their average weekly cigarette smoking rate (from 15.41 to 10.00 percent). This result bodes well for the proponents of the proposed menthol ban; it addresses disparities in smoking behavior thought to be influenced by race-based advertising practices. Overall, I find that 52.8% of all Black menthol smokers switched to Tobacco cigarettes when faced with the removal of mentholated products, and expected consumer surplus among Black households falls by 42.7 to 42.9% (dependent on merger status) when compared to the status quo.

Researchers Levy et al. (2021b) evaluated the expected impact of a menthol cigarette ban through the use of a recent expert elicitation on behavioral changes resulting from the removal of mentholated cigarettes. They find an expected decline in cigarette smoking rates of 15%;

Table 1.8: Average Weekly Rate of Product Usage: Menthol Cigarette Ban.[a]

			Independent Producers		Merged Producers	
		Without Ban	With Ban	% Change	With Ban	% Change
Cigarettes	Black	15.41%	10.00%	(-35.12%)	9.99%	(-35.13%)
	Non-Black	15.76%	14.30%	(-9.29%)	14.30%	(-9.31%)
	High Income	14.91%	13.22%	(-11.32%)	13.22%	(-11.33%)
	Low Income	17.75%	15.04%	(-15.24%)	15.04%	(-15.27%)
	Average	**15.72%**	**13.74%**	**(-12.58%)**	**13.74%**	**(-12.59%)**
E-Cigarettes	Black	0.23%	0.25%	(+12.23%)	0.28%	(+22.74%)
	Non-Black	0.48%	0.51%	(+4.38%)	0.53%	(+10.06%)
	High Income	0.43%	0.45%	(+3.75%)	0.47%	(+8.96%)
	Low Income	0.49%	0.53%	(+7.48%)	0.0.57%	(+15.21%)
	Average	**0.45%**	**0.47%**	**(+4.91%)**	**0.50%**	**(+10.90%)**
	Cessation	0.47%	0.48%	(+1.74%)	0.48%	(+1.71%)

[a] The table above reports expected weekly rates of product usage under the assumption of a menthol cigarette ban, averaged across April 2015 through April 2019 and adjusted for market population. I display usage rates for cigarette and e-cigarette by demographic profile; cessation rates are presented as the average across all consumers.

my results suggest a similar—if slightly smaller—reduction. With regard to changes in smoking rates among Black Americans, researchers Issabakhsh et al. (2022) rely upon the same expert elicitation of behavioral changes as in the aforementioned study. Their results suggest an expected 35.7% reduction in the Black smoking rate when compared to the current status quo scenario. Again, my counterfactual study suggests similar changes in cigarette usage among the Black community.

Finally, I find the menthol ban is associated with a rise in the sale of electronic smoking devices, the amount of which differs dependent upon the assumption of independent or merged (cigarette and e-cigarette) producers. Under the assumption of independent producers, I find the menthol ban is associated with at 4.91% rise in the average weekly consumption of e-cigarettes. Unsurprisingly, Black households experience the largest growth in e-cigarette smoking rates—these consumers being most affected by the removal of menthol cigarettes.

Provided the total acquisition of e-cigarette producers by Big Tobacco (one firm producing both products), the rise in average weekly e-cigarette usage more than doubles to 10.90%.

Ultimately, I find the vast majority of smokers who quit cigarettes, provided a menthol ban, do not substitute their consumption to other nicotine products, i.e. e-cigarettes and cessation products. My results mirror those observed in Ontario, Canada, where, despite a fraction of consumers indicating willingness pre-ban (Ontario having banned menthol cigarettes in 2017) to switch to e-cigarettes, research by Chaiton et al. (2020) did not find a significant association between the Ontario's menthol ban and e-cigarette usage. This result bodes well for policymakers concerned with the continuation of addiction through the use of electronic smoking devices post ban.

However, I must note that for much of my sample, the relative share of e-cigarette usage remained quite small; shares post January 2018 seeing a dramatic rise in the proportion of e-cigarettes. As such, the willingness to substitute to e-cigarettes remains very much time-dependent; rising alongside the growth in popularity of electronic smoking products. Nor does my counterfactual model consider that marketing practices by e-cigarette companies, may change in an attempt to draw disfranchised cigarette smokers post-ban.

Cigarette Taxation For decades sin taxes—excise taxes placed on things like tobacco, alcohol and gambling—have been used for health, education, and other public programs; for example, states such as Arizona, New Hampshire, Virginia, and Colorado, use revenue generated from cigarette sales to fund programs from public education to economic revitalization projects. In recent years, tax revenue from tobacco products has fallen with the decline in smoking rates, and the FDA's proposed menthol ban may lead to the steepest reduction yet seen.

As an alternative to the menthol ban, I find that a 10.23% sales tax, imposed in addition to current state and federal-level taxes, leads to a comparable reduction in the average weekly cigarette smoking rate (see Table 1.9). Further, under taxation, the average household faces a reduction in consumer surplus of 13.9% to 14% dependent on merger status; whereas, the proposed menthol ban reduced average surplus by 15.7% to 15.9%. Of greater disparity is the reduction of surplus experienced, on average, by Black households: taxation resulting in an average consumer surplus reduction of 12.9% to 13%, whereas the proposed menthol ban lowers consumers surplus by 42.7% to 42.9%. Black households largely prefer menthol products, and a 10.23% sales tax reduces household consumption far less than the proposed menthol ban among Black consumers, therefore it's only logical that Black Americans would prefer a 10.23% tax to the removal of mentholated cigarettes.

Table 1.9: Average Weekly Rate of Product Usage: Cigarette Tax (10.23%).[a]

			Independent Producers		Merged Producers	
		Without Tax	With Tax	% Change	With Tax	% Change
Cigarettes	Black	15.41%	13.63%	(-11.52%)	13.64%	(-11.50%)
	Non-Black	15.76%	13.76%	(-12.72%)	13.76%	(-12.71%)
	High Income	14.91%	12.98%	(-12.94%)	12.98%	(-12.93%)
	Low Income	17.75%	15.66%	(-11.78%)	15.66%	(-11.77%)
	Average	**15.72%**	**13.74%**	**(-12.57%)**	**13.74%**	**(-12.56%)**
E-Cigarettes	Black	0.23%	0.23%	(+2.38%)	0.24%	(+6.14%)
	Non-Black	0.48%	0.50%	(+2.79%)	0.52%	(+6.40%)
	High Income	0.43%	0.45%	(+2.60%)	0.46%	(+6.15%)
	Low Income	0.49%	0.51%	(+3.15%)	0.53%	(+6.93%)
	Average	**0.45%**	**0.46%**	**(+2.77%)**	**0.48%**	**(+6.39%)**
	Cessation	0.47%	0.48%	(+1.93%)	0.48%	(+1.93%)

[a] The table above reports expected weekly rates of product usage under the assumption of an 10.23% cigarette tax, averaged across April 2015 through April 2019. I display usage rates for cigarette and e-cigarette by demographic profile; cessation rates are presented as the average across all households.

Regardless of demographic group, changes in relative degree of consumer surplus demonstrate clear a preference for taxation rather than an outright product ban. For instance, I find among low-income households — those often most impacted by sales taxation policies —, a 10.23% cigarette sales tax results in a smaller reduction in consumer surplus when compared to the removal of menthol cigarettes. Low-income households face a reduction in consumer surplus ranging from 13.3% to 13.4% under taxation vs a loss of 18.8% to 19% under the menthol cigarette ban. Lastly, under taxation, e-cigarette consumption does not experience the same increase in demand—smokers with a high menthol preference no longer seeking an alternative among e-cigarettes. Again, the assumption of merged producers results in greater e-cigarette usage rates through coordinated pricing strategies among taxed and untaxed products.

As a back-of-the-envelope calculation, I multiply DMA-level weekly smoking rates by market population, weighted by the average number of packs purchased each week among cigarette smokers—provided via the household-level data. I find a 10.23% sales tax generates an average tax revenue of $66.1 million each week, across the 100 DMAs making up my sample, for a total revenue of $1.41 billion over the period from April 2015 through April 2019.[28]

Revenue generated has the potential to replace that lost, at the state and federal level, as a result of reduced smoking rates. However, to paraphrase FDA commissioner Janet Woodcock, the primary objective of the proposed menthol ban is to address health disparities as a result of unscrupulous marketing practices—particularly in communities of color; for this purpose, an outright ban has the greatest effect (FDA, 2021).

Flavorant Ban It is necessary to acknowledge that, pursuant to the successful implementation of the menthol cigarette ban, flavored and menthol e-cigarettes will likely be the FDA's next target. Already, flavored e-cigarettes are only available in disposable form; flavored cartridges were banned in 2020 in an attempt to reduce youth consumption. Further, lawmakers in California, New York, Massachusetts, and New Jersey have passed some form of flavored product restriction, and many other states opting to ban purchasing of flavored products through online marketplaces—avenues of illegal sales to youth and young adults. Therefore, it would be remiss of me to fail to consider the implications of a ban on all—cigarette and e-cigarette—menthol and flavored (fruity, candy, mint) products. Table 1.10

[28]This expected tax revenue should be treated as an upper bound as my model does not consider possible reductions in the number of packs smoked each week; rather, my model is one of smoking incidence. Nor do I address how tax revenue, itself, may be used to fund anti-smoking campaigns and other cessation inducing behavior.

presents my findings.

Table 1.10: Average Weekly Rate of Product Usage: Flavorant Ban.[a]

			Independent Producers		Merged Producers	
		Without Ban	With Ban	% Change	With Ban	% Change
Cigarettes	Black	15.41%	10.00%	(-35.09%)	10.02%	(-34.98%)
	Non-Black	15.76%	14.32%	(-9.18%)	14.34%	(-9.05%)
	High Income	14.91%	13.24%	(-11.21%)	13.26%	(-11.08%)
	Low Income	17.75%	15.06%	(-15.15%)	15.08%	(-15.03%)
	Average	**15.72%**	**13.76%**	**(-12.48%)**	**13.78%**	**(-12.35%)**
E-Cigarettes	Black	0.23%	0.06	(-72.41%)	0.07%	(-71.26%)
	Non-Black	0.48%	0.27%	(-44.65%)	0.28%	(-42.89%)
	High Income	0.43%	0.23%	(-46.81%)	0.24%	(-45.06%)
	Low Income	0.49%	0.27%	(-45.67%)	0.28%	(-43.98%)
	Average	**0.45%**	**0.24%**	**(-46.46%)**	**0.25%**	**(-44.73%)**
	Cessation	0.47%	0.48%	(+1.88%)	0.48%	(+1.86%)

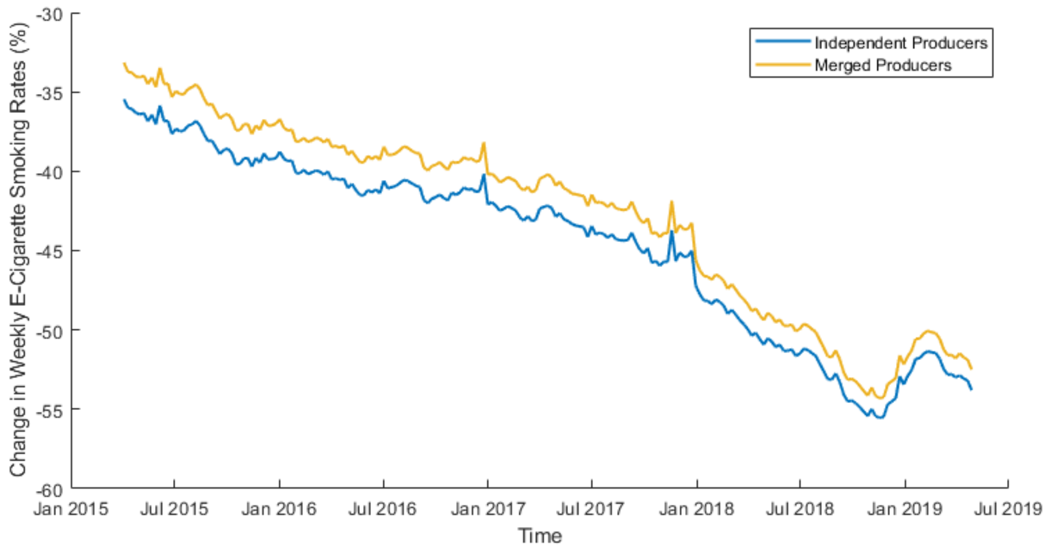
[a] The table above reports expected weekly rates of product usage under the assumption of a flavorant (non-tobacco) ban, averaged across April 2015 through April 2019. I display usage rates for cigarette and e-cigarette by demographic profile; cessation rates are presented as the average across all consumers.

Banning flavorants across all products leads to a similar reduction in average cigarette usage as that seen under the menthol ban. In addition, the fall in Black smoking rates mirror those seen with the earlier menthol ban. Of greater interest is the expected change in weekly e-cigarette usage. On average, a flavorant ban reduces weekly e-cigarette usage by 44.7% to 46.5% (dependent on supply side assumptions). Of course, the average reduction in weekly e-cigarette usage, as a result of a flavorant ban, is very much time dependent.

E-cigarette market shares in the latter half of my sample are dominated by flavored products, whereas, pre-2018, regular tobacco was the primary choice. It then follows, that a flavorant ban's effect on weekly e-cigarette consumption should be considered on a week-by-week basis. Figure 1.6 graphs the weekly expected reduction in e-cigarette sales upon the removal of product flavorants when compared to the status quo scenario.

As the popularity of flavored e-cigarettes grows, so does the impact of a flavorant ban. I find

Figure 1.6: Percent Change in Weekly E-cigarette Consumption Relative to the Status Quo



an average reduction in weekly e-cigarette usage, pre-2018, to be 41.1% assuming independent producers and 39.1% assuming merged producers. Post-2018, the average weekly reduction becomes 51.9% and 50.5% when assuming independent and merged producers, respectively.

1.9 Conclusion

In this paper, I employ a model of consumer demand that incorporates retail- and household-level data, in a way that is internally consistent, to study consumer demand for cigarette and e-cigarette flavorants, and evaluate the impact of the proposed menthol cigarette ban among other counterfactual scenarios.

My work is among the first that analyzes the effect of flavorant bans on demand for cigarettes, e-cigarettes and cessation products, and is the only work that incorporates addiction, categorical substitution, as well as both household- and retail-level data in the study of these effects. I demonstrate that product bans significantly reduce cigarette and e-cigarette consumption, and I find a taxation level which reduces average weekly consumption, among all consumers, by the same rate as the proposed menthol ban. To account for the purchase of e-cigarette companies by cigarette manufactures, I consider my counterfactual results under the assumption of independent and merged producers of cigarettes and e-cigarettes. My results suggest that, across all households, the removal of mentholated cigarettes results in a 13% decrease in the average weekly smoking rate.

Further, by considering a rich set of heterogeneous parameters, I find demographic differences play a key role in responsiveness to product bans; Black households reduce their cigarette consumption by 35% when faced with the removal of menthol cigarettes. In contrast, I find a 10.23% cigarette sales tax as effective, on average, in reducing weekly cigarette smoking rates among all households, and results in a reduction in consumer surplus less than that experienced under the proposed menthol ban (and significantly less when considering Black households).[29] My results suggest, when it comes to e-cigarettes, only a fraction of e-cigarette smokers switch among products. In addition, increases in e-cigarette usage under the proposed menthol cigarette ban are heavily dependent on the assumption of independent or merged (cigarette and e-cigarette) producers; coordination in product pricing playing a key role.

As a final counterfactual scenario, I consider the removal of all menthol and flavored products for both cigarettes and e-cigarettes. I find, on average, the reduction in e-cigarette usage is time dependent, as market shares of flavored e-cigarettes grew rapidly near the end of my sample. As it stands, I find an average reduction in weekly e-cigarette usage, pre-2018, to be 41.1% assuming independent producers and 39.1% assuming merged producers. Post-2018, the average weekly reduction becomes 51.9% and 50.5%, respectively.

Although not considered in this paper, future work has the potential to address youth consumption of product flavorants; my analysis is limited by the unavailability of youth and

[29]The imposition of a 10.23% tax does not cause nearly as great a reduction in cigarette smoking among Black consumers, and therefore may not fulfill the intent of the menthol ban.

young adults in the Nielsen household dataset. Further, I do not address the long term health benefits as the result of the reduction in product usage. Nor do I consider inter brand substitution; rather, my model is one of product usage at the flavor level. Also, beyond the scope of my work is the recent self-regulation by producers designed to avoid government intervention—the effectiveness of which may be a topic of interest. Finally, I form market shares by considering average smoking rates and weekly purchase incidence; I do not consider purchase quantities. Future work has the potential to bridge this gap, forming a model linking both incidence and quantity choice.

Chapter 2

Substitution Patterns and Welfare Implications of Local Taxation: Empirical Analysis of a Soda Tax

2.1 Introduction

Governments of all types levy "sin taxes"—excise taxes imposed on certain goods deemed harmful to society and individuals—with the dual, and oftentimes competing, motives of curbing consumption and raising tax revenue. Examples include taxes on tobacco, alcohol, gambling, drugs[1], junk foods[2], etc. The present study designs a model to evaluate the effects of an increasingly popular category of sin taxes—"soda taxes", which are imposed on sugar-sweetened beverages (SSBs)—by paying particular attention to both cross-border shopping (geographic substitution) and switching to alternative products (product substitution) as forms of tax avoidance.

[1]Such as legal marijuana (Hollenbeck and Uetake, 2021).
[2]See for example Yazzie et al. (2020).

We focus on the SSB tax implemented in the US city of Philadelphia. Philadelphia provides a set of conditions that benefits researchers interested in the effects of SSB taxation. First, Philadelphia is demographically diverse, particularly in terms of income distribution, which allows researchers to better understand the heterogeneous effects of the taxation on the city's rich and poor households. Second, Philadelphia is a large urban center with a substantial set of retail-level and household-level data available. Finally, the city of Philadelphia is both expansive and surrounded by a large suburban population, which provides an ideal setting for studying the effects of geographic and product substitution.

The difference between geographic and product substitution is an important one. For a local government collecting tax revenue, geographic substitution hurts local businesses and lowers tax revenue as consumers take their SSB purchase and with it their grocery shopping to other locations, whereas product substitution leaves consumers' purchases in the same location. For public health agencies, geographic substitution defeats the purpose of the tax as consumers continue to buy unhealthy products and only change where they buy them, whereas product substitution achieves exactly the health objective of the tax by diverting consumption from unhealthy products to healthier ones. A good understanding of the relation between and the magnitudes of geographic and product substitution is then an important prerequisite for sound policymaking, for local governments and public health agencies alike.

Besides SSB taxation, analogous scenarios featuring such tension between geographic and product substitution apply to many policies implemented by states, counties, or cities, including all kinds of sin taxes collected at the local level, other types of local taxes and regulations[3], local subsidies for certain products such as healthy foods and gasoline[4], and

[3]Such as gasoline taxes at the state level and local amusement taxes (Breslow, 2019).

[4]A local subsidy not only induces local consumers to switch from unsubsidized products to subsidized ones, but also incentivizes consumers in other locations to travel to the subsidized location in pursuit of lower prices. For example, when the subsidized gasoline prices in Mexico are noticeably lower than the prices in the US, many US drivers cross the border into Mexico to fill their tanks, leading to a gasoline shortage and temporary suspension of the gasoline subsidy in Mexico's US border region (Garrison and Barrera, 2022).

even restrictions on abortion imposed by various US states[5]. By providing a structural empirical analysis of a local policy that decomposes consumers' heterogeneous substitution responses along the dimensions of geographic and product substitution, this paper offers new insights as well as a useful approach for related policy studies in local taxation, subsidy, and regulation.

The primary goals of this study are then twofold: to estimate consumers' geographic and product substitution as well as welfare changes resulting from an SSB tax, and to provide an empirical framework by which one can evaluate the effects of local taxation or related policies taking into account consumers' multifaceted and heterogeneous substitution patterns.

To quantify the effects of Philadelphia's SSB tax on consumers' product and location choices and their welfare, we construct and estimate a model of consumer demand in the random coefficients nested logit (RCNL) framework (e.g., Grigolon and Verboven (2014), Miller and Weinberg (2017), and Miravete et al. (2018)) using a combination of retail and household data. The random coefficients approach allows rich modeling of heterogeneity in consumer tastes and travel costs, while the nested structure is particularly suited to our analysis of consumers' substitution across beverage categories ("nests"). Aggregate-level retail data lacks the information needed to track individual households' heterogeneous responses to the tax, but measures the aggregate effect of the tax with far less noise and provides a reliable method by which one can account for endogenous variables. Micro-level household data covers only a small subset of all households, but provides an accurate measure of consumer heterogeneity and responsiveness to travel costs. Our empirical approach combines the strengths of the above elements and incorporates the two kinds of data in an internally consistent way.

[5]In states where abortion restrictions are in place, a woman may face a choice among getting an abortion at an out-of-state clinic, switching to an alternative method such as abortion medication by mail, and the "outside option" of using none of the above and giving birth instead. In such cases, the cost associated with traveling out of state to obtain an abortion plays a significant role in determining the woman's ultimate choice.

In estimation, we follow an approach suggested in Grieco et al. (2022) to recover mean utility and unobserved demand shocks while accounting for heterogeneous tastes and cross-border shopping.[6] Our results include estimates of mean responses to SSB taxation and travel time as well as heterogeneous parameters related to preference and substitution. To the best of our knowledge, this paper is the first study that estimates an RCNL model using a combination of aggregate-level and micro-level data.[7]

Several key findings emerge from our analysis. (1) Our demand estimates show that travel time to the alternative region plays a key role in determining households' willingness to cross-border shop, the effectiveness of the taxation, and changes in consumer surplus. On average an extra minute of travel time to reach a store in the alternative region is equivalent to adding 47¢ to the product price. (2) We obtain households' substitution patterns in response to the SSB tax that clearly show both geographic and product substitution are substantial. For each category of Philadelphia SSBs, we quantify to what extent households switch their purchases to a store outside Philadelphia, to an untaxed product in Philadelphia, and to the outside option of no purchase, respectively. (3) We find the SSB tax to be highly regressive. When measured as a percentage of annual income, low-income Philadelphia households on average incur a loss of consumer surplus 4.8 times as large as their high-income counterparts'. (4) Sugar intake from beverages drops significantly for Philadelphia households, by 38% and 35% for high- and low-income households, respectively, attesting to the substantial public health benefit of the tax. (5) Accounting for households' heterogeneous preferences and substitution patterns, we find 3.14¢ per ounce to be the revenue-maximizing tax rate. Compared to this rate, the current tax rate of 1.5¢ per ounce results in 90% of the tax revenue, 72% of the reduction in Philadelphia SSB volume sales, and 68% of the loss in consumer surplus.

[6] Several other papers have used similar methods combining retail and household data, including Goolsbee and Petrin (2004), Chintagunta and Dubé (2005), Tuchman (2019), and Murry and Zhou (2020).

[7] In our estimation process, we found that the inclusion of household data, rather than relying solely on retail data, greatly facilitates the estimation of the RCNL model, particularly the estimation of the nesting parameter (compared to using moment conditions derived from aggregate-level data).

As of July 2022, excluding Cook County in the state of Illinois and the Navajo Nation, all SSB taxes in the US have been implemented at the city level. Given the relatively small area of taxation, these SSB taxation policies are especially vulnerable to tax avoidance behavior in the form of cross-border shopping. Roberto et al. (2019) compare pre- and post-taxation SSB sales in and around Philadelphia, concluding that 24% of the decrease in Philadelphia SSB sales due to the SSB tax is offset by an increase in sales in the surrounding region. Similarly, Seiler et al. (2021) find evidence of cross-border shopping by Philadelphia households to the city's surrounding region, indicating that such behavior offsets 52% of the sales reduction resulting from the city's SSB tax. In the general market for food products, cross-border shopping as a response to sales taxes has been observed in the District of Columbia (Fisher, 1980) and West Virginia (Tosun and Skidmore, 2007), among others.

Literature pertaining to both aggregate-level data (e.g., Thomadsen (2005), Davis (2006), and Houde (2012)) and micro-level data (e.g., McFadden et al. (1977), Capps et al. (2003), Bayer et al. (2007), and Burda et al. (2008)) finds that distance plays an important role in determining product choices. In terms of cross-border shopping, Harding et al. (2012) show that the distance to a lower-tax border affects the pass-through rates of state cigarette taxes, suggesting that consumers engage in cross-state purchasing, which pushes the burden of taxation backwards onto the factors of production. Chandra et al. (2014) find that longer driving distances strongly disincentivize shopping across the US-Canadian border in search of cheaper alternatives. Cross-border shopping as a function of geographic distance has also been identified in Denmark (Bygvrå, 2009) and Norway (Friberg et al., 2018). Our analysis builds upon the idea that distance plays a large role in inhibiting cross-border shopping, and applies it to the policy evaluation of Philadelphia's SSB tax. Our modeling of travel time as a measure of distance within an RCNL model provides a novel approach for incorporating heterogeneous cross-locational substitution patterns into the analysis of consumer choices.

Through the inclusion of geographic and product substitution of beverages in a choice modeling structure, our paper also contributes to the expanding set of SSB taxation literature. Prior works that have considered Philadelphia's SSB tax as well as cross-border shopping, such as Roberto et al. (2019) and Seiler et al. (2021), have used either retail-level or household-level data but not both and have relied on reduced form estimation techniques. We complement those existing works by using both retail-level and household-level data to estimate consumer behavior and aggregate responsiveness to taxation and by conducting counterfactual analyses based on structural estimation results. In the context of structural modeling, Kifer (2015), Wang (2015), Allcott et al. (2019) and Dubois et al. (2020) have used pre-taxation data to predict the effects of hypothetical SSB taxes. We take a different approach by studying the actual implementation of an SSB tax, incorporating both retail-level and household-level data, and accounting for the effects of geographic substitution.

The remainder of this paper proceeds as follows. In Section 2.2, we introduce background information about the Philadelphia SSB tax. We describe our data sources and provide detailed information about the products and market in Section 2.3. Section 2.4 details the discrete choice model of demand that incorporates both the retail and household data. In Section 2.5, we discuss model identification and estimation. Section 2.6 presents the results of our demand estimation. We discuss the effects of the taxation on prices, market shares and consumption in Section 2.7. Changes in consumer surplus and the heterogeneous impact of the taxation by household income level are discussed in Section 2.8. Section 2.9 derives the revenue-maximizing tax rate and explores the effects of alternative taxation schemes. Section 2.10 concludes.

2.2 Philadelphia Soda Tax

On June 16th, 2016, Philadelphia became the second US city to pass an SSB tax, after Berkeley. Initially proposed as a 3¢-per-ounce tax on all sugar-sweetened sodas, the measure garnered widespread support.[8] Supporters of the proposal, such as the American Medical Association, American Heart Association, and other medical groups, argued that such a tax would combat the twin epidemics of obesity and heart disease. Philadelphia ranks as one of the worst cities in the US in terms of type 2 diabetes, heart disease, and obesity. City mayor Jim Kenney predicted the tax would raise $400 million over five years, which would be used to fund universal pre-kindergarten, job creation, and development projects.

Opponents of the proposal claimed that the measure would disproportionately affect the least fortunate. The American Beverage Association, a lobbying group formed of beverage manufacturers and distributors, pushed newspaper, radio and television ads condemning the proposal as regressive—burdening the city's poorest with the largest share of the tax. Interest in the measure was so high that Democratic primary candidates Hilary Clinton and Bernie Sanders weighed in with their opinions for and against the measure, respectively. After months of negotiation, a compromise was reached.

Passing with a city council vote of 13-to-4, the final draft required distributors to pay a 1.5¢-per-ounce tax on all sugar-/artificially sweetened beverages, with the law becoming effective on January 1st, 2017.[9] Thus, the tax applies to not only beverages sweetened with sugar but also diet beverages containing artificial sweeteners. While it may seem surprising to tax artificially sweetened beverages, given that artificial sweeteners have virtually no calories and that diet beverages (beverages with few or no calories) are generally considered healthier alternatives, the city council included diet beverages in the tax to make up for lost revenue as a result of decreasing the tax from the proposed 3¢ per ounce to the actual 1.5¢ per ounce. Most other soda taxes (in Berkeley, CA, Boulder, CO, Seattle, WA, etc.) tax only products with added caloric sweeteners, thus excluding diet beverages. In this paper, we use the term *SSB* to denote a sugar-/artificially sweetened beverage, corresponding to

[8]In the context of beverages, the term *ounce* is a measure of volume and means fluid ounce.

[9]The tax is levied on distributors, and so the price increase observed by consumers is subject to a pass-through rate.

Figure 2.1: Grocery Store Price Tags Indicating Amount of SSB Tax

the coverage of Philadelphia's soda tax. Figure 2.1 illustrates store-level responses to the taxation policy by retailers. The figure shows that the retailers display the amount of SSB tax prominently, contributing to the issue of tax salience, which we discuss in Section 2.6.

2.3 Data

In this section, we describe the data used in our estimation.

2.3.1 Retail Data

Our retail dataset, from Nielsen through the Kilts Center for Marketing at The University of Chicago Booth School of Business, covers the 4-year period from January 1st, 2015 to December 31st, 2018 (Philadelphia's SSB tax took effect at the midpoint of this period on January 1st, 2017). The dataset contains store-level information detailing weekly price and quantity sold at the Universal Product Code (UPC) level. For each store in the dataset, we observe a store identifier, retailer identifier, retailer type as well as the store's ZIP Code prefix (a ZIP Code prefix is the first three digits of a 5-digit ZIP Code). Stores contained within the six ZIP Code prefixes in and around Philadelphia (080, 081, 189, 190, 191, 194) are considered in our analysis. We apply further restrictions by only considering stores that maintained a presence throughout the period of the dataset, whose ZIP Code could be approximated via the household-level data (as detailed later), and whose approximated ZIP Code fell within 8 miles of the nearest ZIP Code in Philadelphia.[10]

Seiler et al. (2021) suggest that cross-border shopping in response to the Philadelphia SSB tax occurs in the region immediately surrounding the city. They find that post SSB taxation, there is a positive, statistically significant increase in SSB sales in stores located 0-6 miles from Philadelphia's border, but not in stores more than 6 miles from the border. Given that the primary purpose of our work is to evaluate the effect of SSB taxation on cross-border shopping and avoidance behavior, we define our market similarly. In practice, we define our market to be the collection of the ZIP Codes in Philadelphia and the surrounding 8-mile band ("city + 8 miles"), where we use the wider 8-mile band to account for the fact that our retail dataset does not provide exact store locations. Appendix B1 shows that sales in stores beyond the 8-mile band surrounding the city do not experience an increase in SSB sales following the implementation of the SSB tax. Our final retail dataset contains 218 stores: 111 stores in Philadelphia and 107 in the surrounding region.[11]

In our retail data, we observe 7,805 UPCs pertaining to eight beverage categories: Carbonated Soft Drinks, Juice, Sports Drinks, Energy Drinks, Coffee, Tea, Flavored Water and Pure Water. All beverage categories, excluding Pure Water, contain both taxed and untaxed products. For each UPC, we have information concerning brand, pack size, container ounces, and flavor (many UPCs relate to variations in pack size and container ounces). We rely on the USDA FoodData Central database along with several food nutrition API services[12] to collect information pertaining to ingredients, sugar content, and caloric value (sugar content

[10]Nielsen data provides ZIP Code information according to the United States Postal Service (USPS) designation. We match these USPS ZIP Codes to their corresponding ZIP Code Tabulation Areas (ZCTAs) as defined in 2016 according to the US Census Bureau. UDSMapper.org, funded by the American Academy of Family Physicians, provides the most up-to-date conversion of USPS ZIP Codes to their corresponding ZCTAs. ZCTA centroids and distances for 2016 are provided by the NBER ZIP Code Distance Database.

[11]In our retail dataset we observe 31 grocery stores, 171 drug stores, and 16 discount stores, which comprise 54%, 30%, and 16% of our observed unit sales, respectively.

[12]world.openfoodfacts.org, chompthis.com, edamam.com, foodrepo.org and nutritionix.com.

and caloric value are reported per a 100ml serving size). Among the UPCs we observe, we remove infrequently purchased items and consider only the 5,259 UPCs whose brand has greater than 0.5% market share in any of the eight beverage categories; such UPCs account for 97.5% of all unit sales.

We then aggregate the UPCs into products, where each product is a brand/SSB status/category/diet status/size combination.[13] SSB status is an indicator denoting the presence of added sugar or artificial sweeteners—these products are subject to the SSB tax if they are sold in Philadelphia. Diet status indicates those products marketed as "diet", "light", "reduced calories", etc. To allow for heterogeneous responsiveness to the tax by product size, we create three size categories in which all products fall: small, for products whose pack size $\times$ container ounces is less than or equal to 20oz; medium, greater than 20oz but less than or equal to 80oz; and large, greater than 80oz. Each size category accounts for roughly a third of all unit sales. In total there are 567 products, of which 377 are SSBs and the other 190 are non-SSBs. Prices are adjusted for inflation.[14]

We use the term *location* to denote Philadelphia or non-Philadelphia (the 8-mile band surrounding Philadelphia). For computational reasons, we aggregate our data from the store-week level to the location-month level; the aggregation over time also helps reduce the potential bias in demand estimation stemming from households' stockpiling behavior (see for example Miller and Weinberg (2017)). In our demand model, to be specified in the next section, we define an alternative in households' monthly choice set to be a product-location combination. Correspondingly, total unit sales, quantity-weighted price, sugar content, and caloric value are considered at the product-location-month level. If every product is available in every location in every month, there would be $567 \times 2 \times 12 \times 4 = 54,432$ observations at the product-location-month level. In reality, not all products are available in both locations

[13]Flavor variations for the same product are aggregated together. Such variations typically have uniform price and similar sugar content and caloric values.

[14]We adjust for inflation by expressing prices as their December 2018 dollar values using the Consumer Price Index for All Urban Consumers (CPI-U).

every month, and as a result our retail dataset has a smaller number of observations, at 41,464.

Table 2.1 provides retail data descriptive statistics, broken down by beverage category and SSB status. We note that we do not account for beverage sales at non-retailer vendors such as restaurants, fast-food outlets, and theaters, as such vendors are not covered in our data.

2.3.2 Household Data

Nielsen provides household purchase data for a sample of US households. Beverage purchases, information pertaining to the number of shopping trips, a household's ZIP Code of residence, and other household demographic data are recorded. The purchase data reports the price paid, number of units purchased, and product UPC. When available, store identifier, retailer identifier, retailer type and store location information are provided. As with the retail data, store location information is provided as a 3-digit ZIP Code prefix.

Between 2015 and 2018, there were 866 households recorded in the Nielsen data who lived within the 153 ZIP Codes pertaining to our market.[15] Over the course of these 4 years, these households recorded 212,301 purchase opportunities (i.e., store trips) with 68,442 beverage purchases. With the provided household demographic information, we differentiate between low- and high-income households. Specifically, we create an indicator variable for the 365 households whose annual income falls below \$50,000—we provide reasoning for this choice of cutoff in the next subsection. We focus on income as a demographic variable of interest since (1) opponents of the taxation policy argued that low-income individuals would be most negatively affected by the policy, and (2) prior works suggest that low income is correlated with a higher price sensitivity and a greater preference for sugary beverages.

[15] Of these 866 households, 211 were tracked for all 4 years. Nationally, Nielsen records a household attrition rate of about 40% each year. Our estimation uses all the 866 households.

Table 2.1: Retail Data Descriptive Statistics[a]

Category	Number of Products	Market Share in Beverages	Price	Sugar[b] (g/100ml)	Calories (cal/100ml)
Carbonated Soft Drinks	**160**	**36.48%**			
SSB	114	32.6%	$2.20	7.43	27.84
Non-SSB	46	3.88%	$2.13	0.03	0.14
Coffee	**37**	**1.76%**			
SSB	25	1.64%	$2.96	8.13	54.48
Non-SSB	12	0.12%	$3.89	0.10	5.26
Energy Drinks	**40**	**4.63%**			
SSB	38	4.63%	$2.73	6.64	28.03
Non-SSB	2	<0.01%	$1.76	2.93	14.69
Flavored Water	**32**	**2.57%**			
SSB	27	2.5%	$1.52	2.9	10.95
Non-SSB	5	0.07%	$1.37	0.13	0.68
Juice	**145**	**20.25%**			
SSB	87	9.38	$2.16	8.37	35.91
Non-SSB	58	10.87	$3.26	9.69	46.99
Pure Water	**45**	**12.07%**			
SSB	0	0%	–	–	–
Non-SSB	45	12.07%	$2.70	0	0
Sports Drinks	**27**	**8.63%**			
SSB	23	8.57%	$1.81	4.76	18.66
Non-SSB	4	0.06%	$1.10	0	0
Tea	**81**	**13.53%**			
SSB	63	13.01%	$2.04	5.98	25.63
Non-SSB	18	0.52%	$2.24	0.08	0.32

[a]Price, Sugar and Calories are presented as quantity-weighted averages.
[b]Sugar present in non-SSB products is the result of natural processes and is not considered added.

2.3.3 ZIP Codes

We define our market as the 153 ZIP Codes either within Philadelphia or whose centroid is outside Philadelphia but within 8 miles of the nearest Philadelphia ZIP Code centroid;

Table 2.2: Household Distribution by Location and Income Status

Income Status	**Households by Location**		
	All	Philadelphia	Non-Philadelphia
High-Income	659,923	267,727	392,196
Low-Income	535,749	327,112	208,637
Total	**1,195,672**	**594,839**	**600,833**

46 ZIP Codes exist within Philadelphia while the other 107 are in the surrounding 8-mile band. ZIP Code-specific demographic data pertaining to the number of households and the percentage of households whose annual income is below $50,000 is collected from the 2018 5-Year American Community Survey (ACS).[16] We consider those households whose annual income is below $50,000 to be "Low-Income" for the purposes of this study (as also observed in Miravete et al. (2018)). Table 2.2 provides the household distribution by location and income status. The table shows that the two locations have roughly the same number of households, with Philadelphia having more low-income households and non-Philadelphia having more high-income ones.

Rather than using straight-line distance to account for location substitution in our model, we rely on travel time as provided by the Google Maps API service. For each Philadelphia ZIP Code, we find the minimum travel time to drive to a non-Philadelphia ZIP Code, and vice versa.[17] We rely on travel time rather than distance to account for location substitution for two reasons: (1) ZIP Code distances do not account for road and highway placements which can greatly alter consumers' willingness to cross-border shop, and (2) Philadelphia is home to many rivers and bridges which would remain unaccounted for if distance was the metric considered. Furthermore, driving is by far the most popular mode of transportation in and around Philadelphia (see for example Duchneskie (2016)), giving support to calculating

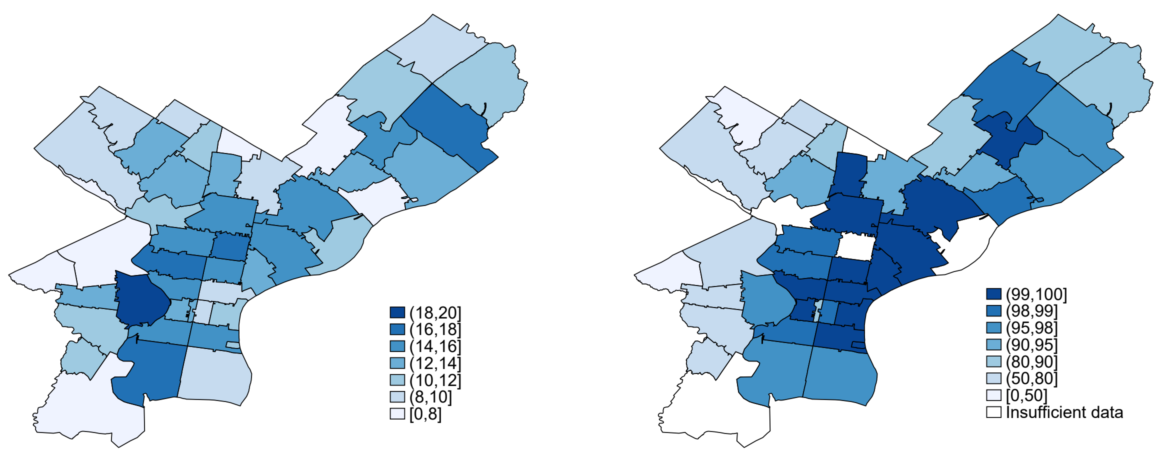
[16] Both the Nielsen data and the ACS report household income in ranges, and the range cutoffs in the two data sources match only at the $50,000 mark.

[17] Travel time between two ZIP Codes is defined as the average time required to drive from one ZIP Code centroid to the other. Using ZIP Code centroids for the calculation is analogous to how ZCTA distances are calculated.

travel time based on driving as an approximation.

Figure 2.2 presents some model-free, suggestive evidence of the importance of travel time. Panel (a) shows, for each Philadelphia ZIP Code, the minimum travel time to a non-Philadelphia ZIP Code. Panel (b) shows, for each Philadelphia ZIP Code, the percentage of beverage purchases made by the ZIP Code's households that are recorded in a store within their home location (Philadelphia). A comparison of the two panels suggests these two variables are positively correlated (a longer travel time to the alternative location is associated with a higher percentage of beverage purchases in the home location), and calculation shows these two variables have a correlation coefficient of 0.53.

Figure 2.2: Travel Time and Beverage Purchases



(a) Travel time (minutes) to alternative location for Phil. ZIP Codes; affected by not only distance but also roads, highways, rivers, bridges, etc.

(b) Percentage of beverage purchases in home location by Phil. households, as observed in the household dataset.

2.3.4 Store Location

As detailed above, the retail dataset does not provide stores' exact locations or full 5-digit ZIP Codes. Instead, we are provided with the stores' 3-digit ZIP Code prefixes (corresponding to the first three digits of the ZIP Codes). There are six ZIP Code prefixes in and around

Philadelphia. Among them, two are entirely within our market: 191 is the ZIP Code prefix for Philadelphia, and 081 corresponds to a region of New Jersey that is entirely within the 8-mile band surrounding Philadelphia. Stores located within the ZIP Code prefixes of 080, 189, 190 and 194 have their locations approximated to determine whether they fall within any of the ZIP Codes pertaining to our market, as follows.

To approximate store locations, we rely on a method similar to that proposed in DellaVigna and Gentzkow (2019) and Goldin et al. (2022). For each store, we observe in the household data the ZIP Codes of residence for the households who make purchasing trips to the store. We then take the store's location to be the average of the centroids of these ZIP Codes, weighted by the total number of trips to the store originating from each of these ZIP Code during the pre-taxation period.[18] In the data, only retailers of the types "Grocery", "Discount Store", and "Drug Store" have unique identifying information that allows for this location approximation. Thus, our final retail and household dataset only considers stores of these types to remain consistent.

2.4 Model

In modeling the demand for beverages as a function of product and household characteristics incorporating consumer heterogeneity and demographic information, we follow the literature on discrete choice demand estimation with retail data (Berry et al. (1995) (BLP), Nevo (2000), etc.), and supplement the traditional method with household data in a process similar to that described in Goolsbee and Petrin (2004), Murry and Zhou (2020), and Grieco et al. (2022).[19] This allows us to leverage the benefits of both datasets: the retail data measures

[18]Centroid locations are given as latitude and longitude. We first convert the centroids to polar coordinates, calculate the weighted average, then convert back to latitude and longitude. There is a slight error introduced, as this conversion assumes a perfectly spherical earth, however given the relative closeness of locations this error is minimal.

[19]Another method is the micro-BLP estimator (Berry et al., 2004). Grieco et al. (2022) suggest that the use of micro-moment conditions, as described in Berry et al. (2004), induces an additional cost in efficiency

responses to the SSB tax with far less noise and allows for a reliable method by which one can account for price endogeneity, while the household data provides a more accurate estimation of heterogeneous parameters, substitution patterns, and responsiveness to travel time. The model we propose utilizes the retail and household data in an internally consistent way.

2.4.1 Demand Specification

Consider household i in month t. The household chooses one of the available beverage options ($j = 1, \ldots, J_t$) or the outside option of no purchase ($j = 0$), where a beverage option is defined as a product-location combination.[20] Household i's indirect utility from choosing beverage option j in month t is given by

$$\begin{aligned} u_{ijt} =& x'_{jt}\beta_i + \alpha_i p_{jt} + h'_{jt}\gamma + \mathbb{1}(A_j \neq A_{z_i})(\phi_i Q_{z_i}) + \xi_{jt} + \bar{\epsilon}_{ijt}, \\ &\text{where } i = 1, \ldots, H_t,\ j = 1, \ldots, J_t,\ t = 1, \ldots, T,\ \text{and } z_i = 1, \ldots, Z. \end{aligned} \tag{2.1}$$

x_{jt} is an $n_1 \times 1$ vector of option j's characteristics in month t, including a constant, Philadelphia dummy variable, category dummy variables, brand dummy variables, sugar content, caloric value, etc. (the full specification is given later in Section 2.6). p_{jt} denotes the retail price for option j in month t. The $n_2 \times 1$ vector h_{jt} contains categorical time trends and month fixed effects. Our month fixed effects are not year-specific; rather, they capture seasonal variation in beverage sales. z_i denotes household i's ZIP Code of residence. A_j and A_{z_i} are indicator variables signifying if option j and ZIP Code z_i are in the Philadelphia location, respectively. Q_{z_i} is the minimum travel time for a household living in ZIP Code z_i to drive to the alternative location (Philadelphia or non-Philadelphia), in which z_i is not located. ξ_{jt} denotes unobserved quality, and $\bar{\epsilon}_{ijt}$ denotes unobserved idiosyncratic preferences. The

relative to a share constrained micro likelihood estimator, the type of estimator applied in this paper.

[20] Product availability varies month to month. Similar to Miravete et al. (2018), if no sales are observed for a beverage option during a specific month, then we assume that option is not present in households' choice set for that month.

indirect utility from choosing the outside option excluding $\bar{\epsilon}_{i0t}$ is normalized to 0.

We characterize household i by a d-vector of demographic attributes D_i, including low-income (below \$50,000) and location (non-Philadelphia). We model unobserved household preference heterogeneity through the use of the multivariate normal distribution. Households' preferences for price, beverage option characteristics, and travel time are as follows:

$$\begin{pmatrix} \alpha_i \\ \beta_i \\ \phi_i \end{pmatrix} = \begin{pmatrix} \alpha \\ \beta \\ \phi \end{pmatrix} + \Pi D_i + \Sigma v_i, \;\; v_i \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_{n_1+2}), \tag{2.2}$$

where Π is an $(n_1 + 2) \times d$ matrix that measures the impact of observable demographic attributes on preferences, while Σ is an $(n_1 + 2) \times (n_1 + 2)$ matrix that captures the covariance of unobserved household preferences. In our study we estimate only the variance of unobserved household preferences, and therefore we restrict $\Sigma_{hk} = 0 \; \forall h \neq k$.

Given the specification in Eq. (2.2), the indirect utility in Eq. (2.1) excluding $\bar{\epsilon}_{ijt}$ can be decomposed into its common and idiosyncratic components, δ_{jt} and μ_{ijt}, respectively, where

$$\begin{aligned} &\delta_{jt} = x'_{jt}\beta + \alpha p_{jt} + h'_{jt}\gamma + \xi_{jt}, \text{ and} \\ &\mu_{ijt} = \left[x'_{jt}, p_{jt}, \mathbb{1}(A_j \neq A_{z_i})Q_{z_i}\right](\Pi D_i + \Sigma v_i) + \mathbb{1}(A_j \neq A_{z_i})(\phi Q_{z_i}). \end{aligned} \tag{2.3}$$

We assume that unobserved idiosyncratic preferences for beverage options, $\bar{\epsilon}_{ijt}$, are correlated within the same beverage category. In our data we observe eight beverage categories (coffee, carbonated soft drinks, energy drinks, flavored water, juice, pure water, sports drinks, and tea), and the outside option of no purchase is defined to be category zero. Thus $\bar{\epsilon}_{ijt}$ follows the distributional assumption of a one-level nested logit model and can be decomposed into

$$\bar{\epsilon}_{ijt} = \zeta_{igt} + (1 - \rho)\epsilon_{ijt}, \tag{2.4}$$

where ϵ_{ijt} is i.i.d. extreme value, $\rho \in [0, 1]$ is the nesting parameter, $g \in \{0, 1, \ldots, 8\}$ is the category that option j belongs to, and ζ_{igt} has a (unique) distribution such that $\bar{\epsilon}_{ijt}$ is distributed extreme value. The nesting parameter ρ measures the correlation in preferences across beverages within the same category. Perfect within-nest substitution is obtained if ρ equals one, while as ρ goes to zero, the model reduces to the standard random coefficients logit specification.

The probability of household i choosing option j belonging to category g in month t is then

$$\pi_{ijt} = \frac{\exp\Big((\delta_{jt} + \mu_{ijt})/(1-\rho)\Big)}{\exp\Big(I_{igt}/(1-\rho)\Big)} \times \frac{\exp\big(I_{igt}\big)}{\exp\big(I_{it}\big)}, \tag{2.5}$$

where the "inclusive values" I_{igt} and I_{it} are given by

$$I_{igt} = (1-\rho)\log \sum_{j\in\mathcal{J}_{gt}} \exp\Big(\frac{\delta_{jt} + \mu_{ijt}}{1-\rho}\Big) \tag{2.6}$$

with $\mathcal{J}_{gt}$ denoting the set of beverage options in category g in month t, and

$$I_{it} = \log\Big(1 + \sum_{g=1}^{8} \exp\Big(I_{igt}\Big)\Big). \tag{2.7}$$

2.4.2 Household Choice Probabilities

In the household dataset, for each household i and each month $t \in \mathcal{T}_i$ during which household i is in the data, we observe the household's O_{it} purchase opportunities (i.e., store trips). During each opportunity, the household chooses one of the available beverage options or the outside option of no purchase.[21] Integrating over the distribution of unobserved household attributes, denoted $F_v(v_i)$, the density of household i's observed sequence of choices is given

[21] We assume that the number of purchase opportunities is independent of observable or unobservable individual characteristics. Such an assumption is necessary for our estimation to be tractable under the BLP framework, and is one innately imposed by researchers working solely with retail data (i.e., Berry et al. (1995), Nevo (2000), etc.).

by

$$L_i(Y_i|x,p,h,Q_{z_i},D_i;\delta,\Theta) = \int \prod_{t\in\mathcal{T}_i}\prod_{o=1}^{O_{it}}\prod_{j=0}^{J_t}[\pi_{ijt}(x_t,p_t,h_t,Q_{z_i},D_i,\delta_t,\Theta,v_i)]^{y_{ijot}}\, dF_v(v_i),$$

where $\delta_t = (\delta_{1t},\dots,\delta_{J_tt})',\ x_t = (x'_{1t},\dots,x'_{J_tt})',\ p_t = (p_{1t},\dots,p_{J_tt})',$ and $h_t = (h'_{1t},\dots,h'_{J_tt})'.$ (2.8)

We summarize the model's heterogeneous taste, travel time, and nesting parameters as $\Theta = (\Pi,\Sigma,\phi,\rho)$, and use Y_i to denote the observed sequence of household i's choices, where $y_{ijot} = 1$ if household i chooses beverage option j during purchase opportunity o in month t.

2.4.3 Retail Market Shares

At the retail level, we use M_t to denote the market size in month t, i.e., the total number of purchase opportunities experienced that month, obtained as the total number of households in the market multiplied by the average number of grocery store trips per household in that month as observed in the household data. We assume a continuum of purchase opportunities of mass M_t, and the household data is assumed to be a finite sample drawn from it.[22]

Consider the set of household-specific characteristics that lead to the purchase of beverage option j in month t, $\{(D_i,z_i,v_i,\bar{\epsilon}_{ijt})|u_{ijt}\geq u_{ikt}\ \forall k=0,1,\dots,J_t\}$. The distribution of $\bar{\epsilon}_{ijt}$ is extreme value as given in Eq. (2.4), which leads to household choice probabilities π_{ijt} given in Eq. (2.5). The distribution of v_i is multivariate normal as given in Eq. (2.2), and the distributions of z_i and $D_i|z_i$ are obtained from the ACS. Integrating over the distributions of v_i, z_i, and $D_i|z_i$, we obtain the predicted market share for beverage option j in month t as

$$s_{jt} = \int_{v_i}\int_{z_i}\int_{D_i} \pi_{ijt}(x_t,p_t,h_t,Q_{z_i},D_i,\delta_t,\Theta,v_i)dF_D(D_i|z_i)dF_z(z_i)dF_v(v_i). \quad (2.9)$$

In assuming a continuum of households, as is routine in the literature, and conditioning on

[22]Appendix B2 provides details about the case of multiple purchases during a single trip.

ξ, through δ, the market share in Eq. (2.9) is deterministic, and the aggregate demand for beverage option j is obtained as $M_t s_{jt}$.

2.5 Identification and Estimation

Our objective is to estimate the parameters α, β, γ, Π, Σ, ϕ, and ρ. While we are not necessarily interested in the value of δ per se, it is required to recover the mean taste parameters α, β, and γ. Thus, our estimation proceeds with two steps. First, we maximize a likelihood function using the retail and household data. This identifies all the parameters except those derived from the mean utility. Next, to estimate α, β, and γ, we use a two-stage least squares (TSLS) regression and instrument p_{jt} with a Hausman style instrument (as seen in Nevo (2001)) to control for correlation with the error term ξ_{jt}.[23]

2.5.1 Maximum Likelihood

In the first stage of our estimation, for any candidate values of Θ and δ, the density of a household's choice history is given by Eq. (2.8), and the corresponding log-likelihood of the household data is

$$\mathcal{L}(Y;\delta,\Theta) = \sum_{i=1}^{H} log[L_i(Y_i|x,p,h,Q_{z_i},D_i;\delta,\Theta)]. \tag{2.10}$$

In theory it is possible to estimate δ directly via maximum likelihood solely with the household-level data; practically, however, this is computationally infeasible considering the large number of beverage options available. Instead, we rely upon the work of Berry (1994) who shows that for any given value of Θ, there exists a unique vector of δ such that the predicted market shares from Eq. (2.9), s_{jt}, exactly match those observed in the retail data,

[23] We calculate the average price of each product across all US stores in the Nielsen data, excluding those in the Philadelphia designated market area (DMA) which contains the market of our demand model, and use this average to instrument the price in our model.

S_{jt}. Consequently, given the retail market shares, we can treat δ as a known function of Θ.[24] Appendix B3 shows in more detail how a unique vector of δ is obtained from our retail data.

Thereby, the log-likelihood of the household-level data shown in Eq. (2.10) can be re-written as

$$\mathcal{L}(Y;\Theta) = \sum_{i=1}^{H} log[L_i(Y_i|x, p, h, Q_{z_i}, D_i, \delta(\Theta); \Theta)], \tag{2.11}$$

where $\delta(\Theta)$ is given by the one-to-one contraction mapping from the retail market share constraint. In performing the contraction mapping, we evaluate the integrals of Eq. (2.9) by Monte Carlo simulation with 4000 Halton draws from the distributions of v, z, and $D|z$ (i.e., 4000 simulated households). Similarly, we use a separate set of 100 Halton draws from the distribution of v when evaluating the integral in Eq. (2.8).[25] Our estimation proceeds by searching for the value of Θ that maximizes Eq. (2.11).[26] Finally, we obtain robust standard errors for Θ by sandwiching the covariance of the household-level gradient between the inverted Hessian at the optimum of the likelihood function.[27]

2.5.2 Mean Utility Coefficients

Given $\hat{\delta}$ resulting from the optimal $\hat{\Theta}$ in the maximum likelihood step, we use the fact that $\delta_{jt} = x'_{jt}\beta + \alpha p_{jt} + h'_{jt}\gamma + \xi_{jt}$ to determine our mean utility parameters. We proceed with

[24] By assuming the aggregate market shares are derived from a continuum of households, the asymptotic variance of the shares is zero. Grieco et al. (2022) shows that this assumption has a cost in terms of both efficiency and inference, unless the household sample size is negligibly small when compared to the size of the market population. This is similar to the efficiency loss of the standard micro-BLP (Berry et al., 2004). In our model $H/N = 0.00072$, where $H = 866$ is the size of the household dataset and $N = 1,195,672$ is the population of households in and around Philadelphia from which those 866 households were drawn; accordingly, the efficiency loss should be minimal. Furthermore, to use a mixed data likelihood estimator as suggested in Grieco et al. (2022) would be too computationally burdensome, as each δ_{jt} must be treated as a parameter of interest in the likelihood estimation.

[25] Results from Train (1999) show simulation variance with 100 Halton draws to be lower than 1000 random draws in a mixed logit application.

[26] Our tolerance for the contraction mapping step is set to $.5e^{-12}$. For the likelihood maximization algorithm, we set a tolerance of $2e^{-10}$ and provide computed numerical gradients. We consider several randomized starting values when proceeding with the maximization algorithm to rule out local minima.

[27] See Train (2009), p. 201.

a TSLS regression relying upon Hausman style instruments, as there is reason to believe p_{jt} may be correlated with the error term ξ_{jt}. Standard errors for $(\hat{\alpha}, \hat{\beta}, \hat{\gamma})$ are calculated using a two-stage bootstrap procedure, where the first stage captures the estimation error from the maximum likelihood step and the second stage captures the typical sampling error. Specifically, we begin by first taking 1000 draws from the asymptotic distribution of Θ. Next, for each draw, Θ_d, we find its corresponding vector $\delta(\Theta_d)$. We then draw with replacement from the sample $\{(\delta_{11}(\Theta_d), x_{11}, p_{11}, h_{11}), \ldots, (\delta_{J_T T}(\Theta_d), x_{J_T T}, p_{J_T T}, h_{J_T T})\}$ to create a bootstrapped dataset (of a size equal to the original sample). Given this bootstrapped sample, we then perform the TSLS regression to estimate $(\alpha_d^*, \beta_d^*, \gamma_d^*)$. From the distribution of $(\alpha_d^*, \beta_d^*, \gamma_d^*)$, we find the standard errors of our mean utility parameters.

2.6 Demand Estimates

Table 2.3 presents the demand estimates of our preferred specification of the RCNL model using the two-step procedure outlined above.[28] To avoid perfect collinearity, we have dropped the category pure water, the brand Aquafina, the size small, and the month of December. On average, consumer valuations for beverages decrease with calories, but increase with sugar content. Excluding juices and flavored water, beverages that contain added sweeteners display a comparative increase in demand. We also observe that consumer valuations decrease with price, conforming to the law of demand.

Tax Salience Considering the price tags displayed in Figure 2.1, where the per-unit price and tax amount are displayed independently and prominently, and the publicity surrounding the SSB tax, we hypothesize that consumer responsiveness to the taxation policy is greater

[28]We considered a three-level nested logit model (Train, 2009) with the choice between beverages and the outside option at the highest level and the choices of beverage category and beverage option at subsequent nodes; however such a model did not improve model fit.

Table 2.3: RCNL Demand Estimates[a]

	Mean	Standard	Demographic Interactions (Π)	
	(α, β, ϕ)	Deviation (Σ)	Low-Income	Non-Phil.
Price	-0.285***	0.010***	-0.006	
	(0.024)	(0.008)	(0.013)	
Calories	-0.011***			
	(0.001)			
Sugar	0.021***	0.051***	0.009	
	(0.004)	(0.004)	(0.006)	
Diet	-0.393***		-0.009	
	(0.039)		(0.042)	
Medium	0.546***		-0.161***	
	(0.054)		(0.062)	
Large	0.964***	0.235***	-0.222***	
	(0.082)	(0.029)	(0.084)	
Tax Amount	-0.157***		0.020	
	(0.029)		(0.055)	
Tax Saving	0.012			
	(0.008)			
SSB × Carb. Soft Drinks	0.224***	0.705***		
	(0.024)	(0.040)		
SSB × Coffee	0.556***			
	(0.048)			
SSB × Energy Drinks	0.580***			
	(0.065)			
SSB × Flavored Water	0.042			
	(0.085)			
SSB × Juice	-0.263***	0.242***		
	(0.024)	(0.030)		
SSB × Sports Drinks	0.798***			
	(0.081)			
SSB × Tea	0.571***			
	(0.048)			
Philadelphia	-0.378***	0.428***		0.754***
	(0.064)	(0.045)		(0.203)
Constant	-4.781***	0.988***	0.877***	
	(0.200)	(0.045)	(0.226)	
Travel Time	-0.127***	0.047***	0.021**	
	(0.014)	(0.007)	(0.010)	
Category Nesting (ρ)	0.685***			
	(0.024)			
Category FEs	Y	N	Y	N
Category Time Trends	Y	N	N	N
Month FEs	Y	N	N	N
Brand FEs	Y	N	N	N

***p<.01, **p<.05, *p<.1

[a]Standard errors are reported in parentheses. Estimates of Category FEs and corresponding Low-Income interactions are provided in Appendix B4.

than that arising solely from a change in price. In fact, Acton et al. (2022) find an increase in perceived costs of SSBs and taxation awareness in countries where a national SSB tax has been implemented.

We therefore include the variable “tax amount”, which provides a quantity-weighted dollar value of the SSB tax to be paid for each Philadelphia beverage option. Similar to Li et al. (2014), who examine gasoline taxes, our negative and significant coefficient for tax amount shows consumer responsiveness to the taxation policy is greater than what the price increase per se suggests. This points to the existence of a tax salience effect, whereby consumers exhibit heightened awareness of and aversion to a highly visible tax, given the extensive media coverage of the tax and retailers' eagerness to inform consumers of the source of such price increases (Figure 2.1). Our estimation results show that on average, the tax salience effect increases consumers' disutility from a price increase due to the tax by $0.157/0.285 = 55\%$.

Likewise, we include the variable “tax saving” which provides, for non-Philadelphia beverage options, their SSB tax amount if sold in Philadelphia. Although only statistically significant at the 85% confidence level, the coefficient for tax saving suggests a small but positive increase in demand for products in the non-Philadelphia location whose counterparts in Philadelphia are subject to the tax. We can interpret this increase in demand as a result of psychological gains from purchasing a product at a lower price than at the alternative location.

Demographic Interactions We also allow for variation in consumer valuations across observed demographic characteristics including income and location, presented in columns 4 and 5 of Table 2.3. The estimation of Π reveals significant differences in consumer valuations for beverage options. For instance, compared to high-income households, low-income households have higher valuations for inside options except for medium- and large-sized tea products (based on the estimates for low-income interactions with constant, sizes, and category fixed effects, the last of which reported in Appendix B4). The disutility from travel time is greater for high-income households, consistent with prior transportation research (e.g., Hymel et al. (2010)) which suggests that high-income households have a higher valuation of their time. Finally, allowing non-Philadelphia households to experience a heterogeneous response to Philadelphia beverage options allows for differing intercepts when considering willingness to cross-border shop.

Random Coefficients and Nesting Parameter We include in our model a rich set of random coefficient parameters (Σ), all of which exhibit statistical significance and sensible results. For instance, the relatively large standard deviation of the random coefficient on sugar content suggests that only 66% of high-income households and 72% of low-income households experience an increase in utility from higher sugar content. The nesting parameter ρ is estimated very precisely, and implies that consumers show a strong correlation across beverages within the same category. To corroborate this point, consider the price elasticity of demand.

Price and Travel Time Elasticities Table 2.4 provides the price elasticity of demand for all households, reporting own- and cross-elasticities averaged at the category-location level, location level, and all beverage options level. Cross-elasticities of demand are reported for beverage options from the same category, same category and same location, same category and different location, and all beverage options. Estimates for the own-elasticity of demand show that households have elastic demand for beverages, with the elasticity ranging from -1.71 to -2.65. Considering the cross-elasticity of demand, we see that it is higher between beverages in the same category, and furthermore it is higher between beverages in the same category and same location when compared to beverages in the same category but different locations. These results delineate a clear order of preference in terms of substitution.

Turning to the travel time elasticity of locational demand, the estimates in Table 2.5 provide

Table 2.4: Price Elasticity of Demand for All Households

Average Level	**Own-Elasticity**	**Cross-Elasticity**			
		Same Category			All Bev. Options
		All Bev. Options	Same Location	Different Location	
Phil. Bev. Options	-2.0910	0.0115	0.0195	0.0035	0.0017
Carbonated Soft Drinks	-1.9999	0.0050	0.0083	0.0017	0.0016
Coffee	-2.6384	0.0446	0.0778	0.0120	0.0020
Energy Drinks	-2.0163	0.0200	0.0348	0.0053	0.0017
Flavored Water	-1.7985	0.0248	0.0411	0.0081	0.0012
Juice	-2.3227	0.0063	0.0106	0.0019	0.0019
Pure Water	-1.9250	0.0201	0.0344	0.0057	0.0019
Sports Drinks	-1.8850	0.0266	0.0452	0.0084	0.0015
Tea	-1.9713	0.0084	0.0141	0.0028	0.0015
Non-Phil. Bev. Options	-2.0225	0.0120	0.0212	0.0028	0.0018
Carbonated Soft Drinks	-1.8685	0.0052	0.0092	0.0011	0.0017
Coffee	-2.6511	0.0482	0.0841	0.0118	0.0022
Energy Drinks	-2.1000	0.0203	0.0342	0.0063	0.0017
Flavored Water	-1.7104	0.0255	0.0461	0.0054	0.0013
Juice	-2.2776	0.0067	0.0120	0.0014	0.0020
Pure Water	-1.9100	0.0214	0.0377	0.0052	0.0020
Sports Drinks	-1.7606	0.0250	0.0442	0.0057	0.0014
Tea	-1.8698	0.0092	0.0165	0.0018	0.0016
All Bev. Options	-2.0564	0.0117	0.0203	0.0031	0.0017

the percentage changes in quantity demanded for beverage options in a household's home and alternative locations, respectively, given a 1% increase in travel time needed to reach the alternative location. These estimates are found by first taking, for each simulated household/beverage option/month combination, 100 draws from the distribution of $\bar{\epsilon}_{ijt}$. Next, the change in choice of beverage location is found by comparing beverage choices given a 1% increase in travel time and holding draws from the distribution of $\bar{\epsilon}_{ijt}$ constant. Finally, the change in locational demand is averaged across simulated households at the location level. These results provide a picture of households who are elastic in travel responsiveness, willing to decrease their propensity to shop in the alternative location when faced with increased travel time.

Table 2.5: Travel Time Elasticity of Locational Demand

	Phil. Households	**Non-Phil. Households**
Phil. Bev. Options	0.14	-1.75
Non-Phil. Bev. Options	-1.36	0.14

Cost of Traveling Lastly, consider the cost of traveling to a grocery store. Taking the average of the ratio of travel time responsiveness to price responsiveness across all simulated households, we find that on average an extra minute of travel time to reach the store is equivalent to adding \$0.47 to the product price.[29] Note that our travel time variable measures the time needed to travel *to* a store in the alternative location, so purchasing at a store 10 minutes away would involve a 20-minute round trip. Also note that the per-minute cost of travel includes not only the cost of time but also the cost of fuel. Other factors such as the depreciation of the car are not significant for the relatively short trips of grocery shopping.

To obtain a back-of-the-envelope figure for the value of time based on the above estimate, consider driving to a store 10 minutes away and coming back. Assuming a speed of 30 miles per hour, a fuel efficiency of 25 miles per gallon, and a gasoline price of \$2.53 per gallon (the average gasoline price in the Philadelphia area in May 2017 (U.S. Bureau of Labor Statistics, 2017)), the fuel cost for the 20-minute round trip is approximately $(30 \times 20/60)/25 \times 2.53 = \$1.01$. Consequently, the value of the 20 minutes spent is approximately $0.47 \times 10 - 1.01 = \$3.69$, implying a value of time equal to $3.69 \times 60/20 = \$11.07$ per hour. This figure falls in the same range as the US Government's practice of valuing people's time between 1/3 and 1/2 of the wage rate based on research in transportation and recreational demand (Goldszmidt et al., 2020). Workers in the Philadelphia area had an average hourly wage of \$26.41 in May 2017 (U.S. Bureau of Labor Statistics, 2018), implying a value of time at $26.41/3 = \$8.80$ per hour using the 1/3 factor and $26.41/2 = \$13.21$ per hour using the

[29] Both price and travel time have random coefficients, and so directly taking the ratio of the mean coefficient for travel time to the mean coefficient for price would give an inaccurate figure, as the ratio of averages differs from the average of ratios.

1/2 factor.[30]

2.7 Pass-Through, Substitution Patterns, and Consumption Changes

In the remainder of this paper, we study the effects of Philadelphia's SSB tax by using our demand estimates to evaluate various counterfactual scenarios. Comparing the outcome under taxation to the counterfactual scenario of no tax, this section examines consumers' substitution patterns and consumption changes brought about by the tax, while the next section analyzes the welfare implications and regressivity of the tax. Then in Section 2.9 we consider the effects of alternative tax rates, alternative tax coverages, and changes in travel time.

As an input for conducting the counterfactual analyses, we first estimate the pass-through rate of the SSB tax.

2.7.1 Pass-Through Rate

Since we do not estimate the supply side of the market, when conducting counterfactual analyses involving changes in the SSB tax, we need to make an assumption about prices under the counterfactual scenarios. To that end, we follow the literature on SSB taxation and estimate a pass-through rate of the tax for constructing counterfactual prices.

Studying Philadelphia's SSB tax, several authors have conducted various analyses on this

[30]Goldszmidt et al. (2020) find a higher value of time at $19 per hour using natural field experiments with the ridesharing company Lyft. One possible reason for the difference between our estimate and that of Goldszmidt et al. (2020) is that Lyft passengers considered in their study may differ from the household sample in our data. Additionally, the value of time may vary across different types of activities.

topic. Cawley et al. (2018) and Roberto et al. (2019) find pass-through rates of 55% and 68%, respectively, while Bleich et al. (2020) and Cawley et al. (2020) find higher pass-through rates of 120% and 105%, respectively. More recently, Seiler et al. (2021) find a pass-through rate of 97%, relying upon their finding that the region more than 6 miles away from Philadelphia does not exhibit an increase in SSB sales in response to Philadelphia's SSB tax. They proceed by treating this region as their control: it is close enough to Philadelphia to experience similar marketing and demand shocks while uninfluenced by cross-border shopping. Similarly, our Appendix B1 demonstrates that the stores present in the Nielsen data that are located in ZIP Codes 8+ miles from Philadelphia do not exhibit a positive SSB sales response to Philadelphia's SSB tax.

We progress with the estimation of the pass-through rate as observed in Seiler et al. (2021), treating the stores 8+ miles from Philadelphia yet still within the surrounding 3-digit ZIP Code prefixes (see Subsection 2.3.1) as the control group for the per-ounce price of SSBs. When performing the pass-through rate analysis, we work with quantity-weighted per-ounce prices at the product-store-week level (similar to Roberto et al. (2019) and Cawley et al. (2020)) rather than aggregating per-ounce prices to the SSB status-store-week level (as in Seiler et al. (2021)), and we obtain category-level estimates of the pass-through rate rather than an average across all categories.

Specifically, for each of the seven categories containing SSBs, we regress price observed at the product-store-week level on the interaction Post-Tax $\times$ Philadelphia as well as store fixed effects and their interactions with the diet, medium, and large dummy variables, week fixed effects, and additional product characteristics including sugar and caloric content. Detailed results are reported in Appendix B5. The interaction Post-Tax $\times$ Philadelphia provides the mean increase in SSB prices in Philadelphia compared to the control group. We find category-level price increases ranging from 1.09¢ per ounce (tea) to 1.59¢ per ounce (energy drinks), corresponding to pass-through rates of 72.7% to 106% of the 1.5¢-per-ounce tax rate. A Wald test rejects ($p < .01$) the null hypothesis of the same pass-through rate across the seven categories containing SSBs. Differences across category-level pass-through rates likely arise from a combination of factors including different price elasticities on the demand side and different levels of competition on the supply side. Furthermore, our category-level estimates fall within the range observed in prior research (e.g., Cawley et al. (2018), Roberto et al. (2019), Bleich et al. (2020), Cawley et al. (2020), and Seiler et al. (2021)).

Our estimation contributes to the expanding literature of tax pass-through and serves as an input for our policy evaluation. For the remainder of this study, we assume category-level pass-through rates equal to those obtained here when performing counterfactual analyses.

2.7.2 Substitution Patterns

As our first counterfactual of interest, we examine how the SSB tax induces categorical and locational substitution. To perform this analysis, similar to how we found travel time elasticity, we simulate 100 draws from the distribution of $\bar{\epsilon}_{ijt}$ for each combination of simulated household i, beverage option j, and post-taxation month $t = 25, \ldots, 48$ (January 2017 to December 2018). We then determine product-level utility with and without the SSB tax holding the $\bar{\epsilon}_{ijt}$ draws constant. That is, product-level utility takes the form

$$u_{ijt}^{\text{with tax}} = \delta_{jt}^{\text{with tax}} + \mu_{ijt}^{\text{with tax}} + \bar{\epsilon}_{ijt}, \text{ and} \tag{2.12}$$

$$u_{ijt}^{\text{without tax}} = \delta_{jt}^{\text{without tax}} + \mu_{ijt}^{\text{without tax}} + \bar{\epsilon}_{ijt}. \tag{2.13}$$

Thus, beverage choice with and without the tax is given by the maximal value of the utilities found in Eqs. (2.12) and (2.13), respectively.

In both equations, the coefficients are the estimated coefficients from our demand estimation, and the household and beverage option characteristics are the observed characteristics.

In Eq. (2.12), the prices are the observed prices, while in Eq. (2.13), for each SSB sold in Philadelphia, the tax amount calculated according the relevant pass-through rate is subtracted from the observed price to obtain the counterfactual price in the no-tax scenario, and the variables "tax amount" and "tax saving" are set to zero for all beverage options.

Holding $\bar{\epsilon}_{ijt}$ to be the same between the two equations when examining beverage choices with and without the tax allows us to isolate the effects of the tax on households' beverage choices. In comparison, tracking how the households in our household dataset actually change their choices from the pre-taxation period to the post-taxation period would not paint an accurate picture of the tax-induced substitution patterns, because households' idiosyncratic preferences $\bar{\epsilon}_{ijt}$, product availability, and demand shocks all have changed between the two periods. Likewise, relying on the retail data would not allow us to track how households switch from one category to another and/or from one location to the other as a result of the tax.

We report our findings in Table 2.6. The first column of the table provides the category market shares for Philadelphia SSBs under the counterfactual scenario of no taxation, averaged across the 24 post-taxation months. The next four columns provide the first, second, third and fourth location $\times$ category $\times$ SSB status choices with taxation, given the household would have chosen the leftmost item of that row without taxation. For example, without taxation, sweetened carbonated soft drinks would have made up 43.3% of the market share for Philadelphia SSBs; with taxation, 66.78% of the households who would have chosen sweetened Philadelphia carbonated soft drinks continue to choose sweetened Philadelphia carbonated soft drinks (no substitution), 13.07% choose sweetened carbonated soft drinks in the non-Philadelphia location (geographic substitution), 12.46% choose the outside option (consumption reduction), and 4.45% choose non-sweetened carbonated soft drinks in Philadelphia (product substitution). Information like this can be particularly useful to policymakers for understanding people's behavior patterns in response to the implementation of a policy.

Table 2.6: Model Predicted Substitution Patterns

Choice W/o Tax		Top Four Choices With Tax							
Phil. SSBs		**1st Choice**		**2nd Choice**		**3rd Choice**		**4th Choice**	
Carb.	(43.30%)	P Carb. S	(66.78%)	NP Carb. S	(13.07%)	Outside Op.	(12.46%)	P Carb. NS	(4.45%)
Coffee	(2.21%)	P Coffee S	(90.90%)	Outside Op.	(3.90%)	NP Coffee S	(2.91%)	P Coffee NS	(1.67%)
Energy	(6.65%)	P Energy S	(90.58%)	Outside Op.	(4.86%)	NP Energy S	(3.92%)	P Water NS	(0.45%)
Flav.	(2.52%)	P Flav. S	(83.51%)	Outside Op.	(7.17%)	NP Flav. S	(6.71%)	P Flav. NS	(1.26%)
Juice	(18.67%)	P Juice S	(61.38%)	P Juice NS	(18.97%)	Outside Op.	(8.63%)	NP Juice S	(6.12%)
Sports	(7.75%)	P Sports S	(73.27%)	Outside Op.	(12.09%)	NP Sports S	(12.05%)	P Water NS	(1.04%)
Tea	(18.91%)	P Tea S	(69.79%)	Outside Op.	(13.53%)	NP Tea S	(11.75%)	P Tea NS	(1.72%)

P, NP, S, and NS denote Philadelphia, non-Philadelphia, SSB, and non-SSB, respectively.

As expected, for all SSB categories, the primary choice with taxation remains the same as that without. We observe that the categories of Philadelphia SSBs that are the most responsive to taxation are juice, carbonated soft drinks, tea, and sports drinks, as measured by the proportion of households who switch away. Excluding carbonated soft drinks and juice, the primary choice of substitution is the outside option, followed by the same category of SSBs in the alternative location. For Philadelphia SSBs, the proportion of households who transfer their consumption to the same category of SSBs in the alternative location is almost as large as those who switch to the outside option, or, in the case of carbonated soft drinks, larger. This provides clear evidence towards a willingness to cross-border shop in the presence of an SSB tax.

The Philadelphia SSB categories of coffee and energy drinks retain the greatest proportion of original consumers. We hypothesize that this pattern is due to the heterogeneous interaction of taxation policy with product size. Coffee and energy drink products are primarily sold in small, single serving containers with relatively high per-ounce prices; thus, the price increase due to the tax is proportionally smaller than those observed in other categories, where products on average come in larger sizes with lower per-ounce prices.

Supportive evidence is provided in Table 2.7, which displays simulated market shares for SSBs and non-SSBs by size and location with and without taxation. Unlike Table 2.6, for

Table 2.7: Simulated Market Shares by SSB Status, Size, and Location

SSB Status × Size × Bev. Location	Without Tax	With Tax	Difference	% Change
Philadelphia Bev. Options				
Non-SSB × Small	0.63%	0.80%	+0.17	26.97%
Non-SSB × Medium	1.27%	1.59%	+0.32	25.27%
Non-SSB × Large	1.06%	1.20%	+0.13	12.46%
SSB × Small	3.03%	3.48%	+0.45	15.01%
SSB × Medium	4.36%	3.10%	-1.26	-28.88%
SSB × Large	2.89%	0.67%	-2.22	-76.75%
Non-Philadelphia Bev. Options				
Non-SSB × Small	0.63%	0.64%	+0.01	2.33%
Non-SSB × Medium	1.89%	1.94%	+0.05	2.37%
Non-SSB × Large	1.70%	1.71%	+0.01	0.87%
SSB × Small	3.08%	3.26%	+0.18	5.69%
SSB × Medium	5.42%	5.85%	+0.42	7.80%
SSB × Large	3.08%	3.50%	+0.42	13.67%
Outside Option	70.96%	72.26%	+1.30	1.84%

Table 2.7 we do not need to keep track of how each simulated household switches from one choice to another in response to the tax, and so the market shares reported in Table 2.7 are found by averaging the choice probabilities of the original 4000 Halton draws across the 24 post-taxation months without directly simulating product choices. The "without tax" counterfactual is conducted with the effect of taxation removed from the individual-level utility.

From Table 2.7 we observe that the effect of the SSB tax is heterogeneously distributed among differently sized SSBs. The tax increases the market share of small Philadelphia SSBs while decreasing the market shares of medium and large Philadelphia SSBs, with large Philadelphia SSBs seeing the biggest drop. SSBs in the non-Philadelphia location experience an increase in market share regardless of size; so do non-SSBs in Philadelphia. These are intuitive results. Consider Philadelphia SSBs, which are subject to the SSB tax. Compared to small products, large products are typically sold at a "quantity discount" and have a lower per-ounce price. Consequently, the SSB tax—levied at 1.5¢ per ounce—results in proportionally larger price increases for large products, thereby having a more negative impact on large products' market shares. Some of the market share that leaves large Philadelphia SSBs goes to small Philadelphia SSBs due to their proportionally smaller price increases and relatively high substitutability, giving rise to an increase in the market share of small Philadelphia SSBs.

2.7.3 Effects of SSB Tax on Beverage Consumption

We now consider the effects of Philadelphia's SSB tax on households' beverage consumption as well as their cross-border shopping and tax avoidance behavior. For each simulated household in each post-taxation month, we compute the household's expected consumption (in ounces) of Philadelphia SSBs, Philadelphia non-SSBs, non-Philadelphia SSBs, and non-Philadelphia non-SSBs, respectively, based on the model predicted choice probabilities and adjusting the amounts to account for the expected numbers of products and units purchased per trip and the expected number of trips in that month. We then sum over the 24 post-taxation months and compute the average per household over all households, Philadelphia households, and non-Philadelphia households, respectively.[31] We do this twice, without tax and with tax, respectively, and then calculate the differences. The results are reported in Table 2.8.

Turning first to Table 2.8's estimates pertaining to the average across all households in both locations, our counterfactual simulation shows that Philadelphia's SSB tax reduces an average household's purchase of Philadelphia SSBs by 55%. 23% ($= 509/2,219$) of this reduction is offset by an increase in the purchase of non-Philadelphia SSBs, leading to a net reduction equal to 42% of the purchase of Philadelphia SSBs in the no-tax scenario. Of course, considering only the average household does not provide a full picture. Instead, a primary benefit of our structural estimation using a combination of retail and household data is the ability to explore how the taxation policy affects households' behavior conditional on

[31]The same procedure for computing the expected amount for each simulated household and then averaging across simulated households is used in subsequent analyses when we compute the average amount of tax paid, loss in consumer surplus, and sugar and caloric consumption.

Table 2.8: Average Beverage Consumption per Household[a]

SSB Status × Bev. Location	Without Tax	With Tax	Difference	% Change
All Households				
Philadelphia Bev. Options				
Non-SSB	2,158	2,428	+270	12.51%
SSB	4,026	1,807	-2,219	-55.12%
Non-Philadelphia Bev. Options				
Non-SSB	3,332	3,364	+32	0.96%
SSB	4,600	5,109	+509	11.07%
Philadelphia Households				
Philadelphia Bev. Options				
Non-SSB	3,827	4,341	+514	13.45%
SSB	7,097	3,306	-3,791	-53.42%
Non-Philadelphia Bev. Options				
Non-SSB	770	824	+54	7.11%
SSB	1,009	1,483	+474	47.00%
Non-Philadelphia Households				
Philadelphia Bev. Options				
Non-SSB	521	551	+30	5.69%
SSB	1,012	335	-677	-66.85%
Non-Philadelphia Bev. Options				
Non-SSB	5,846	5,856	+10	0.17%
SSB	8,124	8,668	+544	6.69%

[a]In ounces; aggregate amount over the post-taxation period January 2017 to December 2018.

the location of their residence.

As expected, Philadelphia households on the whole favor Philadelphia beverage options. In the case without taxation, 88% of Philadelphia households' SSB purchase is for SSBs sold within the city limits. The implementation of the SSB tax reduces their purchase of Philadelphia SSBs by 53% and increases their purchase of non-Philadelphia SSBs by 47%. Since Philadelphia households' purchase of non-Philadelphia SSBs without taxation is relatively small, a 47% increase in their non-Philadelphia purchase offsets only 13% of the reduction in their Philadelphia purchase. When considering the change in SSB purchase in the two locations combined, Philadelphia households experience an average reduction of 41%.

As observed with Philadelphia households, non-Philadelphia households also prefer beverage options in their home location. In the case without taxation, the purchase of Philadelphia

SSBs accounts for only 11% of non-Philadelphia households' SSB purchase. Furthermore, non-Philadelphia households are more responsive to the SSB tax, reducing their purchase of Philadelphia SSBs by 67% in response to the tax and offsetting 80% of this reduction through an increase in non-Philadelphia SSB purchase. This is an intuitive result, as non-Philadelphia households already live in a region without taxation and travel carries an inherent cost. When considering non-Philadelphia households' SSB purchase in the two locations combined, we find that the tax leads to a drop of only 1.5%.

Finally, from Table 2.2 we know that non-Philadelphia households comprise 50.25% of all households in our market, and from Table 2.8 we find that relative to Philadelphia households, non-Philadelphia households display a greater tendency to transfer their SSB purchase from Philadelphia to the surrounding region in response to the SSB tax. It is then not surprising that a majority (54%) of the increase in the purchase of non-Philadelphia SSBs comes from non-Philadelphia households avoiding the taxed region rather than cross-border shopping by Philadelphia households. Prior studies of SSB taxation typically consider the increase in SSB sales in the surrounding untaxed region to be a result of cross-border shopping by residents of the taxed region. Our results shed light on the multiple sources of such an increase and suggest that SSB taxation may be more effective than previously thought, if we consider that the tax's intended target is those households residing within the city limits.

Two prior papers, Roberto et al. (2019) and Seiler et al. (2021), also use retail scanner data to examine the SSB tax and sales of SSBs in and around Philadelphia. There exist several similarities and dissimilarities between our works. In particular, our counterfactual simulation finds a decrease in volume sales of Philadelphia SSBs greater than that suggested by either prior paper. Roberto et al. (2019) find that volume sales of Philadelphia taxed beverages decline by 51% after the taxation policy and that 24% of this reduction is offset by an increase in volume sales in the surrounding region for a net reduction of 38%, while Seiler et al. (2021) find a decrease of 46% in volume sales of Philadelphia taxed beverages, with 52% of this reduction offset by an increase in volume sales in the surrounding region for a net reduction of 22%. In comparison, we find that the SSB tax results in a 55% reduction in volume sales of Philadelphia taxed beverages, with an increase in volume sales in the surrounding region offsetting 23% of this reduction for a net reduction of 42%.

Differences in the estimated impact of the SSB tax can result from a multitude of factors. Firstly, to the best of our knowledge, our paper is the first to analyze the effects of an SSB tax in a structural context where geographic substitution plays a primary role in determining consumers' choices. The works of Roberto et al. (2019), Cawley et al. (2020), and Seiler et al. (2021), among others, employ reduced form estimations that consider the change from pre-taxation to post-taxation SSB volume sales. Using a structural model, we complement prior works by forming our counterfactual estimation directly on the post-taxation months and incorporating the presence of shocks unrelated to changes in tax policy; thus, we model purchase as it would have been in the post-taxation period barring the presence of taxation. Secondly, both Roberto et al. (2019) and Seiler et al. (2021) use data obtained from IRI whereas our data is provided by Nielsen; differences in the retail stores covered by the different data sources can contribute to differences in the expected outcome. Finally, to more accurately account for households' heterogeneous responsiveness, we rely upon both retail and household data, which is another potential source for differing results between our work and those of others.

2.8 Welfare Implications

In this section, we consider the welfare effects of the SSB tax for consumers shopping at grocery stores, drug stores, and discount stores, including the amount of tax paid by households, the change in their consumer surplus, and the effects on their sugar and caloric consumption. We first present our findings at the location level and later, when focusing on the regressivity

Table 2.9: Average Tax Paid and Loss in Consumer Surplus per Household[a]

	All Households	Phil. Households	Non-Phil. Households
Tax Paid	$27.10	$49.59	$5.04
ΔCS	-$55.83	-$106.32	-$6.27

[a]Aggregate amount over the post-taxation period January 2017 to December 2018.

of the SSB tax, consider them at the income status $\times$ location level.

2.8.1 Welfare Effects by Household Location

We begin by evaluating the average amount of tax paid and loss in consumer surplus per household during the 24 post-taxation months, where the loss in consumer surplus is the difference between the expected utility (the "inclusive value" I_{it} in Eq. (2.7)) without and with taxation, divided by the household's marginal utility of money α_i. Table 2.9 presents our findings averaged across all households, Philadelphia households, and non-Philadelphia households, respectively.

As expected with local taxation, households paying the most taxes are those living within the city limits, with an average Philadelphia household paying over 9 times that of an average non-Philadelphia household. This difference follows from non-Philadelphia households' lower demand for Philadelphia SSBs (as discussed in Subsection 2.7.3) and the fact that they can purchase in the untaxed location without incurring travel costs.

For households in both locations, the expected loss in consumer surplus is noticeably greater than the expected amount of tax paid, but the ratio is heterogeneous across household locations. While an average non-Philadelphia household experiences a loss of consumer surplus equal to 124% of their tax amount, an average Philadelphia household's loss of consumer surplus is 214% of their tax amount. The discrepancy in this ratio between the two locations arises primarily from the difference in how costly geographic substitution is. Switching from

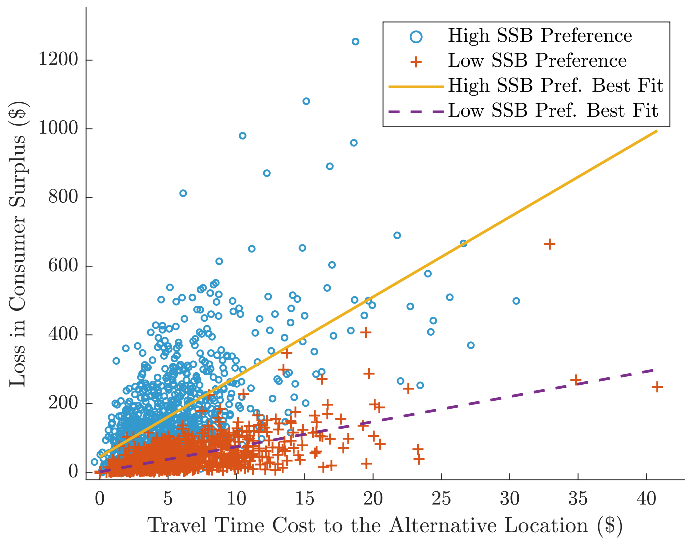
Philadelphia SSBs to non-Philadelphia SSBs in response to the tax necessitates traveling for Philadelphia households but not for non-Philadelphia households, and therefore Philadelphia households incur a larger proportion of their consumer surplus loss in the form of travel costs as opposed to tax paid.

Even among Philadelphia households, there is a large degree of heterogeneity in terms of travel costs, which depend on households' proximity to the city border. In addition, another factor impacting households' welfare changes is their preference for SSBs. Figure 2.3 illustrates how these two factors interact and jointly influence Philadelphia households' loss of consumer surplus resulting from the SSB tax. The figure presents a scatter plot of Philadelphia households' loss in consumer surplus versus their travel time cost to reach the alternative location (equal to a household's travel time to reach the non-Philadelphia location times its marginal disutility of travel time and divided by its marginal utility of money). It shows that across all Philadelphia households, the magnitude of a household's consumer surplus loss increases in the household's travel time cost. Moreover, an increase in travel time cost is particularly detrimental for households with a high preference for SSBs, as they are more "attached" to SSBs and therefore more likely to engage in cross-border shopping to purchase SSBs. In Figure 2.3, among low SSB preference households the line of best fit has a slope of 7.3, whereas among high SSB preference households the slope is much higher at 23.3, implying that a \$1 increase in travel time cost leads to a \$23.3 increase in consumer surplus loss (recall that the consumer surplus loss is the aggregate amount over the 24 post-taxation months).[32] These results thus shed light on the intricate relation between geographic and product substitution as well as the SSB tax's heterogeneous welfare implications for different types of households.

All is not bad for those consumers of SSBs, as we consider how the taxation policy reduces

[32] High SSB preference households are defined as those Philadelphia households whose average utility derived from SSBs, in the simulation without taxation, is greater than the median for Philadelphia households. The rest are low SSB preference households.

Figure 2.3: Loss in Consumer Surplus vs. Travel Time Cost[a]



[a]For simulated Philadelphia households. High SSB preference households are those whose average utility derived from SSBs, in the simulation without taxation, is greater than the median for Philadelphia households. The rest are low SSB preference households.

sugar and caloric consumption—a side benefit of the SSB tax, whose stated primary goal is to generate tax revenue. According to the US Center for Disease Control and Prevention, SSBs are the leading source of added sugars in the American diet, and frequent consumption of sugary drinks is associated with obesity, type 2 diabetes, heart disease, and kidney diseases, among a plethora of other negative health effects.[33] Table 2.10 presents the change in sugar and caloric consumption from beverages during the 24 post-taxation months, averaged across all households, Philadelphia households, and non-Philadelphia households, respectively.

We find that, for an average household living in Philadelphia and the surrounding region, there is an expected reduction in the consumption of sugar by 18%. This effect is strongest for Philadelphia households, who have an average reduction of 36%, whereas non-Philadelphia households—who are not the targeted population of the SSB tax—experience an average reduction of 1.7%. To put this reduction in context, we consider the expected caloric reduction. For Philadelphia households, the implementation of the SSB tax translates to a decrease in caloric intake equal to 27,020 calories—approximately 13.5 days' worth of caloric intake (un-

[33]See https://www.cdc.gov/nutrition/data-statistics/sugar-sweetened-beverages-intake.html.

Table 2.10: Average Sugar and Caloric Consumption from Beverages per Household[a]

	Without Tax	**With Tax**	**Difference**	**% Change**
All Households				
Sugar (g)	20,080	16,483	-3,597	-17.91%
Calories (cal)	81,822	67,774	-14,048	-17.17%
Philadelphia Households				
Sugar (g)	19,148	12,252	-6,896	-36.01%
Calories (cal)	77,212	50,192	-27,020	-34.99%
Non-Philadelphia Households				
Sugar (g)	20,995	20,636	-359	-1.71%
Calories (cal)	86,345	85,030	-1,315	-1.52%

[a]Aggregate amount over the post-taxation period January 2017 to December 2018.

der a 2,000-calories-a-day diet). The sizeable reduction in sugar and caloric consumption among Philadelphia households attests to the substantial public health benefits of the SSB tax. Note that our results only consider the decrease in sugar and caloric consumption from beverages purchased at grocery stores, discount stores, and drug stores; overall reduction will be larger when considering other avenues of purchase. Also note that our analysis does not consider substitution to sugary non-beverage alternatives.

2.8.2 Differences between High- and Low-Income Households

We now consider to what extent households with different income status differ in their amount of tax paid, loss of consumer surplus, and reduction in sugar and caloric consumption. This will in turn inform us about the degree to which the taxation policy exhibits regressive tendencies, which is particularly relevant in this context, as a primary concern for opponents of Philadelphia's SSB tax was its potential impact on the city's poor—households who, as found in past studies, generally display a greater demand for SSBs, the products to be taxed.

From our structural setup, there are several mechanisms by which low-income households may react differently to the implementation of an SSB tax. First, we know from our model

Table 2.11: Price Elasticity of Demand by Income Status

Average Level	**Own-Elasticity**	**Cross-Elasticity**			
		Same Category			All Bev. Options
		All Bev. Options	Same Location	Different Location	
High-Income					
All Bev. Options	-1.8820	0.0101	0.0179	0.0023	0.0015
Low-Income					
All Bev. Options	-2.2696	0.0138	0.0233	0.0042	0.0020

estimates reported in Table 2.3 that low-income households display a greater demand for inside options excluding medium- and large-sized tea products. Second, our results suggest that low-income households incur less disutility in regard to travel, which may result in a greater willingness to cross-border shop. Finally, price sensitivity may differ between those with means and those without.

Price Elasticity of Demand We begin by considering price elasticity of demand by income status. Table 2.11 presents our findings. Unlike in Table 2.4, here we consider own- and cross-elasticities of demand averaged only at the "all beverage options" level to highlight the differences between high- and low-income households in terms of their responsiveness to price increases. As before, cross-elasticities of demand are reported for beverage options from the same category, same category and same location, same category and different location, and all beverage options.

We find that low-income households display a greater price sensitivity than high-income households, with respect to both own-elasticity and cross-elasticities. For example, low-income households' own-elasticity of demand is 21% greater than high-income households'. This finding is intuitive, as one would expect those with less income to display a greater sensitivity to changes in price. However, despite their greater price sensitivity, regardless of location low-income households still display a greater preference for SSBs under taxation, as

Table 2.12: Average SSB Consumption per Household, by Location and Income Status[a]

Income Status × SSB Location	Without Tax	With Tax	Difference	% Change
Philadelphia Households				
High-Income				
Philadelphia SSBs	7,030	3,148	-3,882	-55.22%
Non-Philadelphia SSBs	745	1,192	+447	59.95%
Low-Income				
Philadelphia SSBs	7,152	3,434	-3,718	-51.98%
Non-Philadelphia SSBs	1,223	1,719	+496	40.59%
Non-Philadelphia Households				
High-Income				
Philadelphia SSBs	768	220	-548	-71.41%
Non-Philadelphia SSBs	8,059	8,559	+500	6.20%
Low-Income				
Philadelphia SSBs	1,468	552	-916	-62.41%
Non-Philadelphia SSBs	8,246	8,871	+626	7.59%

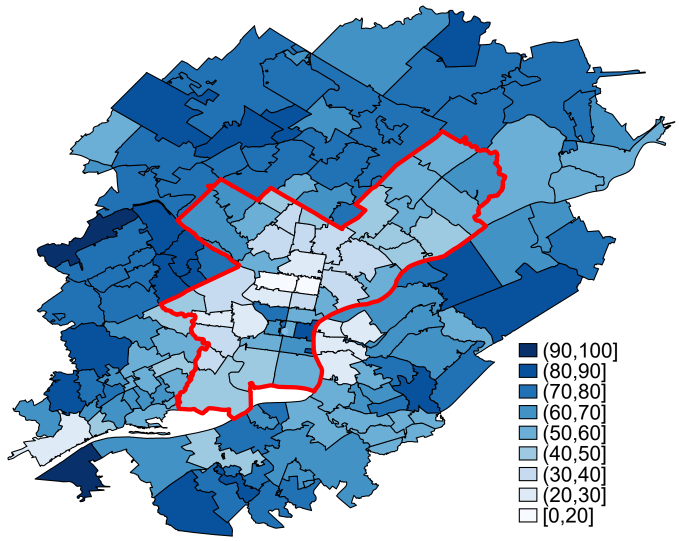
[a]In ounces; aggregate amount over the post-taxation period January 2017 to December 2018.

demonstrated by Table 2.12.

SSB Consumption Table 2.12 shows that low-income households are less responsive to the taxation policy than high-income households. Regardless of household location, low-income households reduce their consumption of Philadelphia SSBs at a lower rate. Among Philadelphia households, low-income households reduce their consumption of Philadelphia SSBs by 52% in response to the tax, 3 percentage points lower than their high-income counterparts. Among non-Philadelphia households, the two types of households exhibit an even greater discrepancy in their responses, with low-income households reducing their consumption of Philadelphia SSBs by 62% in response to the tax, 9 percentage points lower than their high-income counterparts.

High- and low-income households' geographic distribution may go towards explaining the discrepancies between their responses. Figure 2.4 shows the percentage of high-income households for each ZIP Code in and around Philadelphia. From the figure, we observe

Figure 2.4: Percentage of High-Income Households by ZIP Code[a]



[a]The city of Philadelphia is the area outlined in red.

that within Philadelphia, low-income households tend to live near the city center, while outside Philadelphia, low-income households tend to live near the city border. Therefore, when the tax is in effect, among Philadelphia households, low-income households find it more costly to cross-border shop in the non-Philadelphia location since their travel costs would be higher, while among non-Philadelphia households, high-income households find it more beneficial to avoid cross-border shopping in Philadelphia since their saving of travel costs would be higher. Such a pattern therefore offers an explanation for the greater tendency among low-income households in both locations to continue buying Philadelphia SSBs under taxation.

Amount of Tax Paid and Loss in Consumer Surplus Following directly from the differences in purchasing behavior, we consider the differences between high- and low-income households in the amount of tax paid and loss in consumer surplus. Table 2.13 presents these results.

We find that low-income households bear the largest tax burden. Within Philadelphia, low-income households pay 9% more taxes than their high-income counterparts, while out-

Table 2.13: Average Tax Paid and Loss in Consumer Surplus per Household, by Location and Income Status[a]

	All Households	Phil. Households	Non-Phil. Households
High-Income			
Tax Paid	\$21.01	\$47.22	\$3.29
ΔCS	-\$45.29	-\$108.26	-\$2.71
Low-Income			
Tax Paid	\$34.55	\$51.51	\$8.28
ΔCS	-\$68.71	-\$104.75	-\$12.89

[a]Aggregate amount over the post-taxation period January 2017 to December 2018.

side Philadelphia, low-income households pay an astounding 152% more taxes than their high-income counterparts. Among Philadelphia households, the difference in tax paid arises primarily from the pattern that low-income households have a greater preference for SSBs and tend to purchase more SSBs with or without taxation. Among non-Philadelphia households, in addition to the greater preference for SSBs displayed by low-income households, another factor that contributes to the difference in tax paid is a household's home location. As shown in Figure 2.4, outside Philadelphia, low-income households tend to live close to the city border; their proximity to Philadelphia coupled with their lower disutility from travel time (as found in Table 2.3) contributes to their much larger purchase of Philadelphia SSBs, with or without taxation, than their high-income counterparts (as shown in Table 2.12). This, in turn, is the primary driver behind the difference in the amount of tax paid between high- and low-income non-Philadelphia households.

This border proximity also helps explain the significantly larger loss in consumer surplus observed for low-income non-Philadelphia households. For them, compared to their high-income counterparts who tend to live farther away from the city and have a lower preference for SSBs, Philadelphia SSBs are more likely to be the most preferred among all options in their choice set when there is no tax, and therefore the imposition of a tax on Philadelphia SSBs has a more negative impact on their consumer surplus. As supporting evidence, Table 2.12 shows that in response to the tax, low-income non-Philadelphia households on average reduce their Philadelphia SSB purchase by 916 ounces, much higher than the reduction of 548 ounces by their high-income counterparts; in other words, the distortion to the optimal consumption is greater for low-income non-Philadelphia households than for their high-income counterparts.

Interestingly, we find that low-income Philadelphia households experience a smaller loss in consumer surplus compared to their high-income counterparts, although the difference is not large ($104.75 vs. $108.26). Relative to high-income Philadelphia households, low-income Philadelphia households have a greater preference for SSBs, which tends to exacerbate their loss in consumer surplus, but at the same time, as we discuss below, they incur lower travel costs associated with cross-border shopping, which alleviates their loss in consumer surplus. Our simulation shows that despite living closer to the city center, due to their lower disutility from travel time, low-income Philadelphia households' expected cost of travel to the region outside Philadelphia is $5.27, lower than the $6.08 for high-income Philadelphia households, indicating that cross-border shopping is less costly for low-income Philadelphia households. Supporting evidence is provided by Table 2.12, which shows that, with and without taxation, low-income Philadelphia households exhibit a greater tendency to cross-border shop in the non-Philadelphia location, relative to their high-income counterparts.

Regressivity of the SSB Tax Although low-income Philadelphia households on average incur a loss of consumer surplus similar to that for their high-income counterparts, the large income difference between these two groups of households needs to be taken into account when assessing the regressivity of the SSB tax.[34] According to data from the 2018 ACS, the average annual income for low-income Philadelphia households is $22,783, whereas their

[34]Due to the unavailability of joint income and household size data at the ZIP Code level in the ACS, we are not accounting for household size in our model. Research suggests households with less income generally have higher birth rates (Balbo et al., 2013). As such, this would only exacerbate the difference in income per capita between the two types of households, and therefore our estimate of the regressivity of the tax is likely an underestimate.

high-income counterparts have a much higher average of \$112,380.[35] These numbers, together with the loss-in-consumer-surplus numbers reported in Table 2.13, show that when measured as a percentage of annual income, low-income Philadelphia households on average incur a loss of consumer surplus 4.8 times as large as their high-income counterparts', suggesting that the tax is highly regressive. Similarly, among those living outside the city limits, low-income households have an average annual income of \$26,440, whereas high-income households have a much higher average of \$131,974. Therefore, when measured as a percentage of annual income, low-income non-Philadelphia households again incur a much larger loss of consumer surplus than their high-income counterparts. These findings highlight the regressive nature of the SSB tax: those that bear the greatest burden from the tax are those with the least means.

Changes in Sugar and Caloric Consumption Lastly, we consider changes in sugar and caloric consumption from beverages for high- and low-income households by home location. We find that among Philadelphia households, high-income households on average consume less sugar and fewer calories and experience a greater percentage reduction in their consumption, while among non-Philadelphia households, high-income households on average consume less sugar and fewer calories but experience a smaller percentage reduction in their consumption. Detailed results are reported and discussed in Appendix B6.

2.9 Alternative Scenarios

We examine several alternative scenarios to further our understanding of Philadelphia's SSB tax. We first vary the tax rate to identify the one that maximizes the tax revenue. We then

[35] We fit a generalized beta distribution of the second kind (GB2) to the grouped income data from the 2018 ACS (Jorda et al. (2021) show that GB2 is particularly suitable for modeling income distributions). We then compute the average incomes according to the fitted GB2 distribution. We do this for Philadelphia households and non-Philadelphia households, respectively.

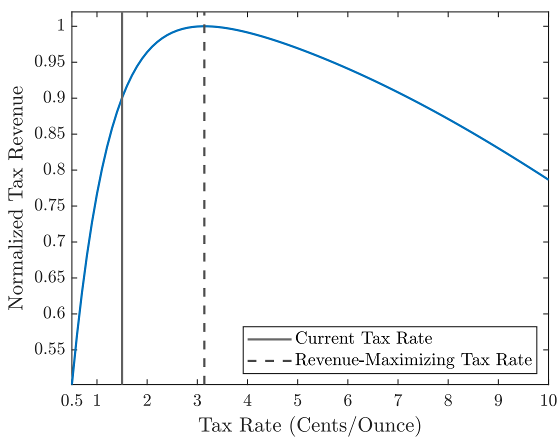
examine the changes in sugar and caloric consumption, consumer surplus, and the revenue-maximizing tax rate if diet products are not subject to taxation (as in the original proposal of the Philadelphia SSB tax). We also explore the impact of taxation on SSB consumption and consumer surplus if not only Philadelphia but also its surrounding region are subject to the same tax (as would be the case if the tax is implemented in a broader region, for instance as a national tax). Lastly, we consider how changes in travel time (resulting from improved roads, for example) would affect SSB consumption and cross-border shopping behavior.

2.9.1 Revenue-Maximizing Tax Rate

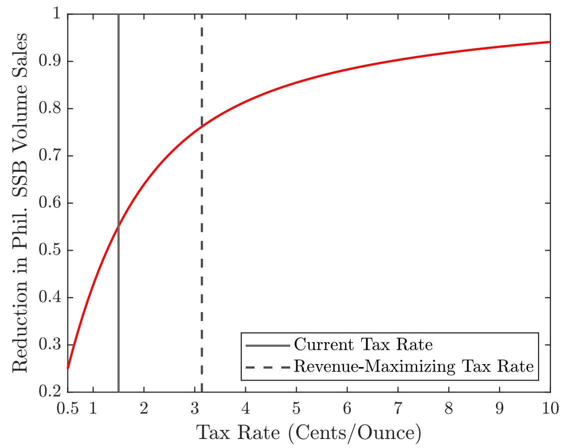
Here we use our estimates of beverage demand and taxation responses to predict outcomes under counterfactual tax rates. To simplify analysis, when computing prices under counterfactual tax rates, we maintain the category-level pass-through rates found earlier. This assumption of a constant pass-through rate is not beyond that observed in prior literature (Allcott et al. (2019) and Seiler et al. (2021)). Unlike prior works, our counterfactual estimates of demand responsiveness to taxation account for consumer heterogeneity in terms of beverage preferences, travel costs, and locational and categorical substitution.

Given any hypothetical tax rate and the corresponding beverage prices computed according to the category-level pass-through rates, we calculate the average amount of SSB tax payment per household during the 24 post-taxation months for each income status/location combination. We use these averages and the demographic distribution of households provided in Table 2.2 to obtain the total tax revenue in each income status/location combination. Summing over the four income status/location combinations provides the total tax revenue for the given tax rate. We also compute the reduction in Philadelphia SSB volume sales and average loss in consumer surplus per household during the 24 post-taxation months in a similar fashion. Figure 2.5 plots those three variables against the tax rate.

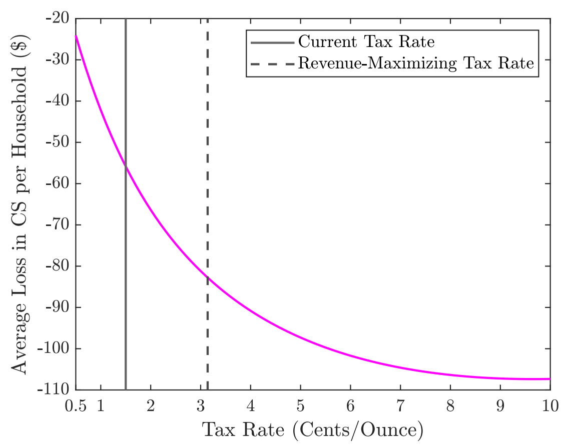
Figure 2.5: Tax Revenue, SSB Volume Sales, and Consumer Surplus: Alternative Tax Rates



(a) Tax Revenue vs. Tax Rate



(b) Reduction in Phil. SSB Volume Sales vs. Tax Rate



(c) Average Loss in CS per Household vs. Tax Rate

Notes: Tax revenue is normalized relative to the maximum, and reduction in Philadelphia SSB volume sales is normalized relative to the volume sales without tax.

We obtain a revenue-maximizing tax rate of 3.14¢ per ounce—much closer to the initially proposed tax rate of 3¢ per ounce than the current tax rate of 1.5¢ per ounce. We find that the current tax rate generates a revenue of $32.5 million during the 24 post-taxation months, which equals 90% of the $36.1 million that would be generated at the revenue-maximizing tax rate, whereas the initially proposed tax rate of 3¢ per ounce would generate a revenue more than 99% of the maximal revenue. We note that these revenue figures account for revenue generated at the stores in our sample, namely grocery stores, discount stores, and drug stores; we do not consider revenue from other sources such as supercenters, gas stations, and dollar stores, nor non-retailer locations such as restaurants, fast-food outlets, and theaters.

With respect to SSB volume sales, under the revenue-maximizing tax rate, there would be a 76% reduction in Philadelphia SSB volume sales, 29% of which would be offset by an increase in the volume sales in the surrounding region, for a net reduction of 54%; under the current tax rate, those figures are noticeably smaller at 55%, 23%, and 42%, respectively. The larger reduction in SSB sales associated with the revenue-maximizing tax rate would be an added benefit for those lawmakers and public health advocates concerned with the consumption of products that contribute to unhealthy lifestyles. However, such a large reduction in SSB sales could be damaging to retailers located in Philadelphia, especially since consumers may take not only their SSB purchase but also their grocery shopping altogether to stores outside Philadelphia.

Furthermore, we find that the revenue-maximizing tax rate demonstrates even bigger regressive tendencies, with 60% of the tax revenue generated coming from low-income households, compared to 57% under the current tax rate. Additionally, the revenue-maximizing tax rate would lead to an average consumer surplus loss of \$82.77, constituting a 48% increase when compared to the current tax rate's average loss of \$55.83.

Differences between the revenue-maximizing tax rate found in our work and those found in other research likely arise from differences in the structure of the demand curve. Of particular note is Seiler et al. (2021), who found a revenue-maximizing SSB tax rate of 1.63¢ per ounce for Philadelphia, when assuming a linear demand curve. Our findings, however, are derived from our demand estimates based on the RCNL modeling structure. We contend that our higher revenue-maximizing tax rate is driven primarily by the persistent consumption of Philadelphia SSBs from a subset of Philadelphia households who lack inexpensive substitutes for SSBs within their home region, many of these households experiencing large travel costs and high SSB preferences. This pattern leads to such households' low price sensitivity with respect to Philadelphia SSBs, which in turn gives rise to a higher figure for the revenue-maximizing tax rate.

In terms of structural modeling, Allcott et al. (2019) consider the optimal SSB tax rate for a government with preferences for wealth redistribution. They determine an optimal tax rate between 1¢ and 2.1¢ per ounce. In contrast to their findings, our analysis is concerned with the revenue-maximizing tax rate rather than the socially optimal tax rate. Furthermore, they focus on a national SSB tax imposed on sugar-sweetened beverages only, whereas our analysis is concerned with the city of Philadelphia and includes diet products containing artificial sweeteners. A more appropriate comparison between our work and Allcott et al. (2019) is found in Appendix B7.1, where we report a revenue-maximizing tax rate of 2.33¢ per ounce under the assumption that diet products are excluded from the tax.

2.9.2 Additional Counterfactuals

Next we consider three additional counterfactual scenarios. Detailed results are reported and discussed in Appendix B7; here we summarize the main findings.

We find that removing diet products from the tax (Appendix B7.1) induces a greater reduction in households' sugar and caloric consumption, reduces households' loss in consumer surplus, and lowers the revenue-maximizing tax rate. The main intuition here is that sugary beverages and their diet counterparts are good substitutes for some households, therefore when diet products are excluded from the tax, these households are able to switch from sugary beverages to their diet counterparts in order to avoid the tax, rather than having to travel for cross-border shopping or switch to less substitutable products.

We also consider a counterfactual in which the tax is levied upon both Philadelphia and its surrounding region (Appendix B7.2), and a counterfactual in which the travel time experienced by all households is varied proportionally from 50% to 200% of the baseline (Appendix B7.3). Results from those counterfactuals indicate that a taxation policy's geographic coverage as well as households' travel costs have significant impact on households' responses and the consumption and welfare outcomes.

Together, our counterfactual analyses show that policymakers need to pay careful attention to the scope of the tax—in terms of product and geographic coverage—as well as households' cross-border shopping behavior when designing taxation policies.

2.10 Conclusion

In this work, we employ a structural modeling framework that combines both retail and household data to study the relationship between local taxation and households' tax avoidance behavior including cross-border shopping and product substitution, focusing on Philadelphia's SSB tax.

We find that travel time to the untaxed region surrounding Philadelphia plays an important role in determining households' substitution patterns. In response to the implementation of an SSB tax, our results quantify households' reduction in the consumption of taxed beverages in Philadelphia and their willingness to seek untaxed products in locales beyond the city border. Accounting for household location, we find that a majority 54% of the rise in SSB sales in the surrounding region is due to an avoidance of Philadelphia SSBs by those residing in the surrounding region, rather than cross-border shopping by Philadelphia households. We also show that price responsiveness alone does not fully account for observed consumer behavior; instead, we provide evidence that tax salience has a noticeable impact on consumer demand, particularly in highly publicized and politicized taxes such as the SSB tax our work studies.

Our model and estimation allow for heterogeneous consumer behavior based on their demographic characteristics and proximity to the city border. Taking into account consumers' heterogeneous responses to the tax, we show that the current tax rate 1.5¢ per ounce is well below the revenue-maximizing tax rate 3.14¢ per ounce. Our results suggest that, without readily available substitutes and facing large travel costs associated with cross-border shopping, a subset of Philadelphia households are persistent in their consumption of Philadelphia SSBs, willing to pay the higher prices resulting from the tax. Their low price sensitivity with respect to Philadelphia SSBs is a main factor behind the high revenue-maximizing tax rate.

Based on our demand estimates, we calculate the average amount of tax paid and the average loss in consumer surplus for households at different locations and different income levels. Taking into account travel costs and the switch to less preferred products, Philadelphia households on average incur a loss in consumer surplus more than twice the amount of tax paid, with low-income households bearing the largest burden. When measured as a percentage of annual income, low-income Philadelphia households on average incur a loss of consumer surplus 4.8 times as large as their high-income counterparts', suggesting that the tax is highly regressive.

These findings are especially relevant for governmental entities weighing the benefits of a revenue-generating, healthy-habit-inducing tax against the drawbacks of a strongly regressive taxation policy. Additionally, through counterfactual analyses in which we vary the geographic coverage of the tax as well as travel times to the alternative region, we provide supportive evidence for the notion that policymakers must carefully consider geographic coverage and geographic substitution when assessing the effects of local policies.

Lastly, our model's applicability extends beyond the context studied in this work. Any local tax or subsidy susceptible to cross-border shopping offers an opportunity for study under our framework, which facilitates rich modeling and sensible estimation of individuals' heterogeneous travel costs and substitution patterns, as well as the policy's potentially vastly different welfare implications for different individuals.

Chapter 3

Two-Stage Structured Probit Demand Estimation for Application to Large Choice Sets

3.1 Introduction

Three issues are increasingly in the forefront of the estimation of individual level demand models: first, the presence of dynamic consumer behavior, second, random coefficients, and third, realistic substitution patterns. Ignoring these issues can lead to biased and inconsistent estimates of the causal relationship between observed covariates and consumer choice. Many models of consumer choice employ overly restrictive assumptions designed to simplify estimation that may otherwise be infeasible... for instance, logit specifications or restrictive substitution patterns in the case of probit models. Given that the foundation of firm behavior relies on the demand functions of various marketplace agents, it is of the utmost importance that derived consumer demand be as accurate as possible for economic and marketing analysis. The purpose of this article is to develop a structured two-stage multinomial random coefficient probit model that allows for the estimation of large alternative sets without overbearing substitution pattern restrictions.

In this paper, the proposed model allows a household's purchase decision and product choice to be separately estimated with common unobservables that affect both outcome decisions. The main methodological contribution of this paper is to structure the covariance between product level utilities in terms of observed characteristics to maintain the feasible estimation of substitution patterns between product choices in the presence of large alternative sets. Finally, household level heterogeneity is introduced through the use of random coefficients, which allow for differing responses to covariates common to both stages of the modeling process. This portion of the model is standard, however it should be noted that this paper builds upon the work of Chib et al. (2009) and Fong et al. (2014) by estimating the individual level parameters of both stages simultaneously through a Gibbs sampling process. These household level parameters in both the purchase incidence and product choice stage are allowed to be correlated through the distribution of individual level heterogeneity.

Beginning with the consumer's problem, it is assumed an individual follows two decisions steps when making a purchase. First, the individual must decide to purchase a product in a specific choice set. Second, conditioned on the decision to purchase, the individual chooses a product within the set of considered alternatives. The benefits to this two-stage process is clear. Models that focus only on the product choice stage of estimation, without accounting for the decision to purchase, are justifiable only if unobservables that drive the purchase incidence are uncorrelated with product choice. If the unobservables are correlated, then the single stage estimation can lead to biased estimates. **?** best describes this as sample selection bias; where the observed product choice is no longer a randomly selected sample. While correcting for this bias can be accomplished by allowing the no purchase decision to make up the outside option of a consumer's choice set, there are several advantages to this paper's multi-stage estimation.

First, the two-stage process allows for the introduction of dynamic consumer behavior, which would otherwise not be considered if only the static one-stage model was estimated. For example, Hendel and Nevo (2006a) maintain that a traditional static demand model is likely to mis-measure long-run own-price and cross-price elasticities when applied to storable goods due to intertemporal substitution. It is entirely conceivable that a price decrease may result in a large demand increase due to forward purchasing behavior. If this is the case, then the true long-run response to a permanent price decrease may be less than that suggested by a static demand model. Hendel and Nevo hold that this mis-specification results from a static model's lack of relevant history, such as consumer inventories and past pricing.

Furthermore, segmenting the purchasing process allows for differing responses in both the purchase incidence and product choice equation; this is especially important in the presence of heterogeneous parameters. Bucklin and Gupta (1992) consider the behavior of two different consumers. One consumer tends to buy in the category of interest at regular intervals, but switches among products when in the presence of promotional activity (this consumer could be considered as insensitive at purchase incidence, and sensitive at the product level). A second consumer consistently purchases the same product irregularly, but more so in the presence of promotional activity (this individual is considered as sensitive at the purchase incidence level and insensitive at the product level). By conceptualizing the purchasing process through these two-stages, researchers may better understand how promotional activity (price-cuts, advertising, displays, etc.), in conjunction with demographic characteristics, determines a consumer's willingness to switch between products and/or increases their propensity to purchase in the category of interest.

For example, Buklin and Gupta (1992) posit that if 90% of consumers are insensitive at the product choice level, and instead are sensitive at the purchase incidence level, then promotional activity merely subsidizes a product with their own loyal consumers who are forward buying. In contrast, if 90% of the consumers are sensitive at the product level, and insensitive at the decision to purchase, then promotional activity encourages switching behavior. In support of the hypothesis that consumers have differing sensitivities to promotional activity during different parts of the decision process; Buklin and Gupta find evidence that individuals with high sensitivities to promotional activity at the purchase incidence are not always sensitive at the brand level.Finally, a plethora of additional research (Bucklin and Lattin (1991), Chintagunta (1993), Arora et al. (1998), Bucklin et al. (1998), Andrews and Currim (2009), etc.) demonstrates the importance of such multistage consumer decision models.

However, most studies of multistage consumer models, and in particular those that estimate the purchase incidence and product choice, generally rely on the Extreme Value type 1 distribution. This approach is first seen as a nested logit model in Bucklin and Lattin (1991) and Bucklin and Gupta (1992), where an individual's purchasing behavior is modeled as the product of two probabilities; the probability of purchasing and the conditional choice of alternative. The error term on the purchase probability and choice of alternative are then assumed to be distributed Extreme Value Type 1. Other models, such as those in Chintagunta (1993) and Arora et al. (1998), assume a translog bivariate utility function between the purchase incidence and alternative choice equations with an Extreme Value Type 1 error. This results in a model almost identical to that of the aforementioned nested logit specification. While these models are advantageous in terms of computational burden; Chib et al. (2004) discusses the myriad of problems that arise from the logit specification.

First, in a nested logit framework, the link between the purchase incidence and brand choice is achieved through the *inclusive value* - the expected utility for product choice alternatives. Acting as a relative measure of product attractiveness, the inclusive value restricts the household level responsiveness to observed product-level covariates during the purchase incidence stage. For instance, during the product choice stage price may be more important than observed advertising, however during the purchase incidence stage the opposite may be true. Since the product level covariates enter the purchase incidence stage through the inclusive value, such behavior cannot be observed. This restriction can impede efforts to discover how consumers may employ differing responses to economic and marketing variables during the steps of the decision process. In contrast, under the probit framework, observables such as price and promotional activity can make up both the covariates in the purchase incidence and product choice stage of estimation. Furthermore, the inclusive value provides the most influence in the purchase incidence level to the product that provides the highest utility in the product choice stage of the model; thus further restricting a consumer's responsiveness to observables.

Finally, the structure of a nested logit model ignores the presence of unobserved correlations between the purchase incidence and product choice utilities. For example, a weekend dinner party can incentivize a consumer to purchase cheaper store brand products; resulting in an unobserved correlation between the purchase incidence and product choice. This sample selection effect can bias a researcher's findings unless captured by a flexible correlation structure between the purchase incidence and product choice utilities. While the translog utility function can allow for correlation structures that arise from unobservables common to the purchase incidence and brand choice utilities, its structure results from the first order conditions of product choice and purchase incidence decisions, and is restrictive in nature.

In lieu of the nested logit and translog utility, Chib et al. (2004) introduced a two-stage random coefficient probit model that demonstrated a better in-sample fit compared to the nested logit alternative. The authors maintain that the probit error assumption allowed for a flexible correlation structure between the purchase incidence and product choice utilities, removing bias that would otherwise be present. However, unlike the model proposed in this paper, Chib et al. (2004) avoid introducing substitution patterns between the product level utilities. They argue that such correlations are difficult to identify when the number of alternatives is large. However, ignoring substitution patterns between product level utilities is unrealistic and an overly restrictive assumption.

The challenge of this paper was therefore to incorporate substitution patterns into the covariance matrix, while maintaining the feasibility of estimation offered by more restrictive probit or logit models. This is accomplished by structuring the covariance between product level utilities based on the importance of perceived distances between choice alternatives. This is similar to the strategy presented in Dotson et al. (2018) and Cohen (2010), where the authors structure the probit covariance as a correlation matrix defined in terms of different measures of product similarity. They find that modeling substitution patterns in this way provides a better in-sample fit when compared to the logit specification. Finally, in the model proposed in this paper, heterogeneity is introduced through the use of random coefficients estimated via a Gibbs sampler as shown in Fong et al. (2014), and modified as described in Chib et al. (2009) to fit the two-stage estimation process. Modeling household heterogeneity in this way allows researchers to identify how demographic characteristics may influence consumer behavior.

This paper's model is estimated using Bayesian Markov Chain Monte Carlo (MCMC) techniques similar to those seen in Fong et al. (2014), Chan and Jeliazkov (2009), Chib et al. (2009), and Chib and Ramamurthy (2010). The remainder of this paper is organized as follows; in section 2, the model is described in detail, as well as common identification restrictions. In section 3 the algorithm behind the MCMC estimation process is detailed. In section 4, a comparison between the proposed model and unstructured probit specification is then provided. Section 5 concludes with some directions for future research.

3.2 Model

Assume that there exist J alternatives and N individuals in a panel data set containing purchase history with T_i time periods for $i = 1, \ldots, N$. This paper posits that decision to purchase among the bundle of alternatives is itself a choice of interest. Each time period, it is assumed an individual i faces two potential decisions; first, whether to purchase a product in a specific set of alternatives. Second, conditioned upon the decision to purchase, individual i must choose one alternative among the $j = 1, \ldots, J$ alternatives. Let $y_{\mathbf{1}it}$

be the observed decision to purchase and $\boldsymbol{y}_{\mathbf{2}it} = [y^1_{\mathbf{2}it}, \ldots, y^j_{\mathbf{2}it}, \ldots, y^J_{\mathbf{2}it}]'$ denote the $J \times 1$ vector corresponding to the observed choice of alternative for subject i and $t = 1, \ldots, T_i$. Allowing $z_{\mathbf{1}it}$ to denote the latent utility accompanying the decision to purchase, it follows that $y_{\mathbf{1}it} = I(z_{\mathbf{1}it} > 0)$ where $I(\cdot)$ is the indicator function. Conditioned on $y_{\mathbf{1}it} = 1$, and with $\boldsymbol{z}_{\mathbf{2}it} = [z^1_{\mathbf{2}it}, \ldots, z^j_{\mathbf{2}it}, \ldots, z^J_{\mathbf{2}it}]'$ denoting the latent utilities of choice alternatives, it must be that $y^j_{\mathbf{2}it} = I(z^j_{\mathbf{2}it} > \max\{z^{-j}_{\mathbf{2}it}\})$. Else, if $y_{\mathbf{1}it} = 0$ then $\boldsymbol{y}_{\mathbf{2}it} = \mathbf{0}$; the decision to purchase determines if an individuals choice of alternative is observed.

The latent utilities $z_{\mathbf{1}it}$ and $\boldsymbol{z}_{\mathbf{2}it}$ are assumed to follow the multivariate regression detailed below:

$$\begin{aligned} z_{\mathbf{1}it} &= \mathbf{x}'_{\mathbf{1}it}\boldsymbol{\beta}_{\mathbf{1}i} + \varepsilon_{\mathbf{1}it} \\ \boldsymbol{z}_{\mathbf{2}it} &= \mathbf{X}_{\mathbf{2}it}\boldsymbol{\beta}_{\mathbf{2}i} + \boldsymbol{\varepsilon}_{\mathbf{2}it} \end{aligned} \tag{3.1}$$

where $\mathbf{x}_{\mathbf{1}it}$ and $\mathbf{X}_{\mathbf{2}it} = [\mathbf{x}^{1'}_{\mathbf{2}it}, \ldots, \mathbf{x}^{j'}_{\mathbf{2}it}, \ldots \mathbf{x}^{J'}_{\mathbf{2}it}]'$ are a $k_1 \times 1$ and $J \times k_2$ matrix of covariates and product characteristics, respectively. The parameters $\boldsymbol{\beta}_{\mathbf{1}i}$ and $\boldsymbol{\beta}_{\mathbf{2}i}$ represent a $k_1 \times 1$ and $k_2 \times 1$ vector of individual i's coefficients corresponding to the preceding matrices for a total of $k_1 + k_2 = k$ individual regression coefficients. Finally, $\varepsilon_{\mathbf{1}it}$ represents the error term for the purchase incidence equation, while $\boldsymbol{\varepsilon}_{\mathbf{2}it}$ is the $1 \times J$ vector representing the error for the product level latent utility. It is assumed that $[\varepsilon_{\mathbf{1}it}, \boldsymbol{\varepsilon}_{\mathbf{2}it}]' \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega})$, where $\varepsilon_{\mathbf{1}it}$ and $\boldsymbol{\varepsilon}_{\mathbf{2}it}$ are jointly distributed with mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{\Omega}$. This joint distribution between the decision to purchase and the alternative of choice is an important feature of the model, reducing selection bias from potential sources such as unobserved advertising, sales, changing consumer tastes and preferences, etc.

In general, the parameters $\boldsymbol{\beta}_{\mathbf{1}i}$, $\boldsymbol{\beta}_{\mathbf{2}i}$ and $\boldsymbol{\Omega}$ are non-identified without any additional assumptions about the level and scale of the latent utilities (Train (2009)). It should noted, however, that the level of $z_{\mathbf{1}it}$ is already normalized as the outside option is defined to be the decision not to purchase corresponding to a utility of 0. Only the level of $\boldsymbol{z}_{\mathbf{2}it}$, the latent utility corresponding to choice of alternative, needs to be normalized.

To set the level of $\boldsymbol{z}_{\mathbf{2}it}$ one of the J alternatives is taken to be the outside option with a utility 0. Let the 1st alternative be considered the outside option, then define $\widetilde{\boldsymbol{z}}_{\mathbf{2}it}$ to be the $J-1$ vector of relative utilities where $\widetilde{\boldsymbol{z}}^j_{\mathbf{2}it} = \boldsymbol{z}^j_{\mathbf{2}it} - \boldsymbol{z}^1_{\mathbf{2}it}$ for $j = 2, \ldots, J$. It then follows that $y^1_{\mathbf{2}it} = I(\max\{\widetilde{\boldsymbol{z}}_{\mathbf{2}it}\} < 0)$ and $y^j_{\mathbf{2}it} = I(\widetilde{\boldsymbol{z}}^j_{\mathbf{2}it} > \max\{0, \widetilde{\boldsymbol{z}}^{-j}_{\mathbf{2}it}\})$ for $j = 2, \ldots, J$. The final model analyzed is:

$$\begin{aligned} z_{\mathbf{1}it} &= \mathbf{x}'_{\mathbf{1}it}\boldsymbol{\beta}_{\mathbf{1}i} + \varepsilon_{\mathbf{1}it} \\ \widetilde{\boldsymbol{z}}_{\mathbf{2}it} &= \widetilde{\mathbf{X}}_{\mathbf{2}it}\boldsymbol{\beta}_{\mathbf{2}i} + \widetilde{\boldsymbol{\varepsilon}}_{\mathbf{2}it} \end{aligned} \tag{3.2}$$

where $\widetilde{\mathbf{X}}_{\mathbf{2}it} = [\widetilde{\mathbf{x}}^{1'}_{\mathbf{2}it}, \ldots, \widetilde{\mathbf{x}}^{j'}_{\mathbf{2}it}, \ldots \widetilde{\mathbf{x}}^{J'}_{\mathbf{2}it}]'$, with each element $\widetilde{\mathbf{x}}^j_{\mathbf{2}it} = \mathbf{x}^j_{\mathbf{2}it} - \mathbf{x}^1_{\mathbf{2}it}$, and $\widetilde{\varepsilon}^j_{\mathbf{2}it} = \varepsilon^j_{\mathbf{2}it} - \varepsilon^1_{\mathbf{2}it}$ for $j = 2, \ldots, J$. It is assumed $[\varepsilon_{\mathbf{1}it}\ \widetilde{\boldsymbol{\varepsilon}}_{\mathbf{2}it}]' \sim \mathcal{N}(\mathbf{0}, \widetilde{\boldsymbol{\Omega}})$; $\varepsilon_{\mathbf{1}it}$ and $\widetilde{\boldsymbol{\varepsilon}}_{\mathbf{2}it}$ are jointly distributed with the mean vector $\mathbf{0}$ and covariance matrix $\widetilde{\boldsymbol{\Omega}}$. Tthe scale of the latent utility is set by normalizing the first two diagonal entries of $\widetilde{\boldsymbol{\Omega}}$ to one. Under these assumptions $\widetilde{\boldsymbol{\Omega}}$ can be broken down into the following components:

$$\widetilde{\boldsymbol{\Omega}} = \begin{bmatrix} 1 & \widetilde{\boldsymbol{\Omega}}'_{21} \\ \widetilde{\boldsymbol{\Omega}}_{21} & \widetilde{\boldsymbol{\Omega}}_{22} \end{bmatrix}, \text{ where } \widetilde{\boldsymbol{\Omega}}_{21} = \begin{bmatrix} \sigma_{21} \\ \vdots \\ \sigma_{J1} \end{bmatrix} \text{ and } \widetilde{\boldsymbol{\Omega}}_{22} = \begin{bmatrix} 1 & \sigma_{32} & \cdots & \sigma_{J2} \\ \sigma_{32} & \sigma_{33} & \cdots & \sigma_{J3} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{J2} & \sigma_{J3} & \cdots & \sigma_{JJ} \end{bmatrix}. \tag{3.3}$$

Each element of $\widetilde{\boldsymbol{\Omega}}_{21}$ is the normalized covariance between the decision to purchase and the corresponding choice of alternative; that is, σ_{j1} represents $\mathrm{cov}(\varepsilon_{\mathbf{1}it}, \widetilde{\varepsilon}^j_{\mathbf{2}it})$ = $\mathrm{cov}(\varepsilon_{\mathbf{1}it}, \varepsilon^j_{\mathbf{2}it})$ - $\mathrm{cov}(\varepsilon_{\mathbf{1}it}, \varepsilon^1_{\mathbf{2}it})$ for alternatives $j = 2, \ldots, J$. While $\widetilde{\boldsymbol{\Omega}}_{22}$ is the $(J-1) \times (J-1)$ covariance matrix, associated with the product level latent utilities of the $J-1$ alternatives, that helps define the substitution patterns between product choice.

To simplify future notation, consider this paper's model written in vector form:

$$\underset{J\times 1}{\mathbf{Z}_{it}} = \underset{J\times k}{\mathbf{X}_{it}}\ \underset{k\times 1}{\boldsymbol{\beta}_i} + \underset{J\times 1}{\boldsymbol{\varepsilon}_{it}}, \qquad \boldsymbol{\varepsilon}_{it} \sim \mathcal{N}(\mathbf{0}, \widetilde{\boldsymbol{\Omega}}) \tag{3.4}$$

with $\mathbf{Z}_{it} = \begin{bmatrix} z_{\mathbf{1}it} \\ \widetilde{\boldsymbol{z}}_{\mathbf{2}it} \end{bmatrix}$, $\mathbf{X}_{it} = \begin{bmatrix} \mathbf{x}'_{\mathbf{1}it} & \mathbf{0} \\ \mathbf{0} & \widetilde{\mathbf{X}}_{\mathbf{2}it} \end{bmatrix}$, $\boldsymbol{\beta}_i = \begin{bmatrix} \boldsymbol{\beta}_{\mathbf{1}i} \\ \boldsymbol{\beta}_{\mathbf{2}i} \end{bmatrix}$, $\boldsymbol{\varepsilon}_{it} = \begin{bmatrix} \varepsilon_{\mathbf{1}it} \\ \widetilde{\boldsymbol{\varepsilon}}_{\mathbf{2}it} \end{bmatrix}$, and $\boldsymbol{y}_{it} = \begin{bmatrix} y_{\mathbf{1}it} \\ \boldsymbol{y}_{\mathbf{2}it} \end{bmatrix}$

3.2.1 Random Coefficients

To account for heterogeneity across consumers, a random coefficient framework is adopted, with variations assumed to be driven by both an observed and unobserved individual component. This allows $\boldsymbol{\beta}_i$ to be modeled with the following multivariate regression framework:

$$\boldsymbol{\beta}_i = \boldsymbol{\Lambda}\boldsymbol{H}_i + \boldsymbol{\eta}_i, \quad \boldsymbol{\eta}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{V}_\beta), \quad i = 1, \ldots, N \tag{3.5}$$

where $\boldsymbol{H}_i$ is a $h \times 1$ vector including an intercept term and individual i's specific demographic characteristics. $\boldsymbol{\Lambda}$ is a $k \times h$ matrix of coefficients corresponding to the aforementioned vector $\boldsymbol{H}_i$. The $J \times 1$ error term $\boldsymbol{\eta}_i$ accounts for the unobserved individual component, and is assumed to be distributed with mean vector $\mathbf{0}$ and covariance matrix $\mathbf{V}_\beta$. This specification allows the parameter values to vary based on observed demographic variables, and whose dispersion is determined by the variance of the unobserved component. As in Fong et al. (2014), the following prior distributions for $\boldsymbol{\Lambda}$ and $\mathbf{V}_\beta$ are assumed:

$$\boldsymbol{\Lambda}|\mathbf{V}_\beta \sim \mathcal{MN}(\boldsymbol{\Lambda}_0, \mathbf{V}_\beta, A_d^{-1}) \tag{3.6}$$

where $\mathcal{MN}(\boldsymbol{\Lambda}_0, \mathbf{V}_\beta, A_d^{-1})$ is a matrix normal distribution with mean $\boldsymbol{\Lambda}_0$, row covariance $\mathbf{V}_\beta$ and column covariance A_d^{-1}. Finally,

$$\mathbf{V}_\beta^{-1} \sim \mathcal{W}(v, V) \tag{3.7}$$

where $\mathcal{W}(v, V)$ is a wishart distribution with the mean vV and v degrees of freedom.

3.2.2 Covariance Matrix

As discussed above, this paper assumes a structured covariance matrix to allow for the feasible estimation of substitution patterns based on product similarity in the face of large alternative sets. The following structural assumptions reduces the dimensionality of $\widetilde{\mathbf{\Omega}}$ from $\frac{J(J+1)-4}{2}$ to $2J + C - 3$ unknowns.

To structure $\widetilde{\mathbf{\Omega}}$, a measure of product similarity is required. Let $\mathbf{X}^* = (\mathbf{x}_1^*, \ldots, \mathbf{x}_c^*, \ldots, \mathbf{x}_C^*)$ be a $(J-1) \times C$ matrix of time and individual invariant product characteristics. To maintain identification of the structured covariance, it must be that $C \leq \frac{J(J-3)+2}{2}$, generally not an issue for large alternative sets. From $\mathbf{X}^*$ the following distance measure is formed:

$$d_c^{jh} = \frac{\sqrt{(\widetilde{x}_c^{*,j} - x_c^{*,h})^2}}{\mathrm{sd}(\mathbf{x}_c)} \tag{3.8}$$

between products h and j for characteristic c. Using this distance measure, $\widetilde{\mathbf{\Omega}}_{22}$is structured as follows:

$$\sigma_{jh} = \begin{cases} \sum_{c=1}^{C} exp\{-d_c^{jh}\}\gamma_c & \text{if} \quad h \neq j \\ \sigma_{jj} & \text{if} \quad h = j \end{cases} \tag{3.9}$$

where $2 \leq h, j \leq J$ and γ_c is the coefficient on the distance measure for characteristic c. The $C \times 1$ vector of coefficients, $\boldsymbol{\gamma}$, is assumed to have the prior distribution:

$$\boldsymbol{\gamma} \sim \mathcal{N}(\boldsymbol{\gamma}_0, \mathbf{V}_\gamma). \tag{3.10}$$

The variance of $\widetilde{\mathbf{\Omega}}_{22}$ is left to be freely determined, noting that $\sigma_{22} = 1$ for identification purposes. It is assumed the variance of $\widetilde{\mathbf{\Omega}}_{22}$ has the prior distribution:

$$(\sigma_{33}, \ldots, \sigma_{JJ})' \sim \mathcal{N}(\boldsymbol{\sigma}_0^2, \mathbf{V}_\sigma). \tag{3.11}$$

This structured format gives a straightforward interpretation of the coefficient, γ_c. First, note that as products are more similar, i.e. $d_c^{jh} \to 0$, then $e^{-d_c^{jh}} \to 1$. As products are more dissimilar, $d_c^{jh} \to \infty$, then $e^{-d_c^{jh}} \to 0$. Thus, γ_c represents the role that similarity plays in determining the covariance between products in $\widetilde{\mathbf{\Omega}}_{22}$. For instance, if $\gamma_c >> 0$, one can infer that similarity in this characteristic implies greater substitution patterns between products. If $\gamma_c = 0$, then similarity in this characteristic plays no role in determining the substitution, and if $\gamma_c << 0$ then products who share this characteristic have more negative substitution patterns than those that are more dissimilar.

Finally, following the advice of Chib et al. (2004), it is assumed that covariance between the decision to purchase, and product level latent utility is left to be freely determined. From (3.3) it is assumed that $\widetilde{\mathbf{\Omega}}_{21}$ is freely determined with the assumed prior distribution:

$$\widetilde{\mathbf{\Omega}}_{21} \sim \mathcal{N}(\widetilde{\mathbf{\Omega}}_{21}^{0}, \mathbf{V}_{\mathbf{\Omega}_{21}}) \tag{3.12}$$

3.3 Estimation

For notational convenience, let $\boldsymbol{\theta} \equiv (\boldsymbol{\beta}_i, \mathbf{\Lambda}, \mathbf{V}_\beta, \widetilde{\mathbf{\Omega}})$, and define $\boldsymbol{\theta} \backslash \boldsymbol{\theta}_k$ to mean the elements of $\boldsymbol{\theta}$ excluding $\boldsymbol{\theta}_k$. Thus, the proposed MCMC estimation is as follows:

Algorithm

step 1) Initialize $\mathbf{Z}_{it}, \boldsymbol{\beta}_i, \mathbf{\Lambda}, \mathbf{V}_\beta,$ and $\widetilde{\mathbf{\Omega}}$.

step 2) Draw $\mathbf{Z}_{it} \,|\, \boldsymbol{\theta}$ from a truncated normal distribution.

step 3) Draw $\boldsymbol{\beta}_i \,|\, \mathbf{Z}_{it}, \boldsymbol{\theta} \backslash \boldsymbol{\beta}_i$ from a multivariate normal distribution.

step 4) Draw $\mathbf{\Lambda} \,|\, \mathbf{Z}_{it}, \boldsymbol{\theta} \backslash \mathbf{\Lambda}$ from a matrix normal distribution.

step 5) Draw $\mathbf{V}_\beta \,|\, \mathbf{Z}_{it}, \boldsymbol{\theta} \backslash \mathbf{V}_\beta$ from a inverse wishart distribution.

step 6) Draw $\widetilde{\mathbf{\Omega}} \,|\, \mathbf{Z}_{it}, \boldsymbol{\theta} \backslash \widetilde{\mathbf{\Omega}}$ using a Tailored Metropolis Hastings approach.

As in Chib et al. (2009) the unobserved outcomes are not involved in the algorithm; that is, the missing $\widetilde{\boldsymbol{z}}_{\boldsymbol{2}it}$ are not involved in the estimation if the individual decided not to purchase. Alternatively, if an individual chose not to purchase then the unobserved outcomes, $\widetilde{\boldsymbol{z}}_{\boldsymbol{2}it}$, could be simulated through a series of conditional draws. However, as Chib et al. (2009) demonstrates, simulating the missing $\widetilde{\boldsymbol{z}}_{\boldsymbol{2}it}$ is computationally burdensome and worsens the mixing of the Markov chain. Finally, if some covariates are missing when the corresponding observations are not observed, then there follows no clear way to sample the missing $\widetilde{\boldsymbol{z}}_{\boldsymbol{2}it}$. With this in mind, the joint posterior distribution of the structured model is detailed below:

$$\begin{aligned}
&\boldsymbol{\pi}(\boldsymbol{\theta}, \mathbf{Z}_{it} \,|\, \boldsymbol{y}_{it}) \\
&\qquad \propto \Big(\prod_{i} \big[\prod_{t \in P_i} f(\mathbf{Z}_{it} \,|\, \theta) I(z_{\mathbf{1}it} > 0) \big] \big[\prod_{t \in P_i^{\mathbf{c}}} f(z_{\mathbf{1}it} \,|\, \theta) I(z_{\mathbf{1}it} < 0) \big] \Big) \mathcal{N}(\boldsymbol{\beta}_i \,|\, \boldsymbol{\Lambda} \boldsymbol{H}_n, \mathbf{V}_\beta) \\
&\qquad \times \mathcal{MN}(\boldsymbol{\Lambda} \,|\, \boldsymbol{\Lambda}_0, \mathbf{V}_\beta, A_d^{-1}) \, \mathcal{W}(\mathbf{V}_\beta^{-1} \,|\, v, V) \, \mathcal{N}(\widetilde{\boldsymbol{\Omega}}_{21}^{0}, \mathbf{V}_{\boldsymbol{\Omega}_{21}}) \, \mathcal{N}(\boldsymbol{\gamma}_0, \mathbf{V}_\gamma) \, \mathcal{N}(\boldsymbol{\sigma}_0^2, \mathbf{V}_\sigma)
\end{aligned} \tag{3.13}$$

where P_i denotes the set of t_i individual specific time periods in which the subject i chose to purchase. Accordingly, $P_i^{\mathbf{c}}$ denotes the set of t_i^c individual specific time periods in which the subject i chose not to purchase. An in depth examination of the sampling process is now shown below.

3.3.1 Sampling $\mathbf{Z}_{it}$

Given $\boldsymbol{\theta}$, $\mathbf{Z}_{it}$ is sampled as follows:

If $y_{\mathbf{1}it} = 1$ then:

$$
\begin{aligned}
&z_{\mathbf{1}it} \,|\, \widetilde{\boldsymbol{z}}_{\mathbf{2}it}, \boldsymbol{\theta} \sim \mathcal{TN}_{(0,\infty)}\Big(\mathbf{E}(z_{\mathbf{1}it} \,|\, \widetilde{\boldsymbol{z}}_{\mathbf{2}it}), \mathrm{var}(\varepsilon_{\mathbf{1}it} \,|\, \widetilde{\boldsymbol{z}}_{\mathbf{2}it})\Big) \\
&\widetilde{z}^{j}_{\mathbf{2}it} \,|\, z_{\mathbf{1}it}, \widetilde{\boldsymbol{z}}^{-j}_{\mathbf{2}it}, \boldsymbol{\theta} \sim \begin{cases} \mathcal{TN}_{(\max\{0, z^{-j}_{\mathbf{2}it}\}, \infty)}\Big(\mathbf{E}(\widetilde{z}^{j}_{\mathbf{2}it} \,|\, z_{\mathbf{1}it}, \widetilde{\boldsymbol{z}}^{-j}_{\mathbf{2}it}), \mathrm{var}(\varepsilon^{j}_{it} \,|\, z_{\mathbf{1}it}, \widetilde{\boldsymbol{z}}^{-j}_{\mathbf{2}it})\Big), & \text{if} \quad y^{j}_{\mathbf{2}it} = 1 \\ \mathcal{TN}_{(-\infty,\ \max\{0, z^{-j}_{\mathbf{2}it}\})}\Big(\mathbf{E}(\widetilde{z}^{j}_{\mathbf{2}it} \,|\, z_{\mathbf{1}it}, \widetilde{\boldsymbol{z}}^{-j}_{\mathbf{2}it}), \mathrm{var}(\varepsilon^{j}_{it} \,|\, z_{\mathbf{1}it}, \widetilde{\boldsymbol{z}}^{-j}_{\mathbf{2}it})\Big), & \text{otherwise.} \end{cases}
\end{aligned}
$$

If $y_{\mathbf{1}it} = 0$ then:

$$
z_{\mathbf{1}it} \,|\, \boldsymbol{\theta} \sim \mathcal{TN}_{(-\infty,0)}\Big(\mathbf{E}(z_{\mathbf{1}it}), \mathrm{var}(\varepsilon_{\mathbf{1}it})\Big).
$$

As stated above, this estimation strategy relies only upon observed outcomes; when $y_{\mathbf{1}it} = 0$, the missing $\widetilde{\boldsymbol{z}}_{\mathbf{2}it}$ are not simulated.

3.3.2 Sampling $\boldsymbol{\beta}$

The joint posterior distribution (3.13) implies that $\boldsymbol{\beta}_i \,|\, \mathbf{Z}_{it}, \theta \backslash \boldsymbol{\beta} \sim \mathcal{N}(b_n, B_n)$, such that

$$
\begin{aligned}
b_i =& B_i \Big[\sum_{t \in P_i} \mathbf{X}'_{it} \widetilde{\mathbf{\Omega}}^{-1} \mathbf{Z}_{it} + \sum_{t \in P^{\mathbf{c}}_i} \mathbf{R} \mathbf{x}_{\mathbf{1}it} z_{\mathbf{1}it} + \mathbf{V}^{-1}_{\beta} \mathbf{\Lambda} \boldsymbol{H}_i \Big], \\
B_i =& \Big[\sum_{t \in P_i} \mathbf{X}'_{it} \widetilde{\mathbf{\Omega}}^{-1} \mathbf{X}_{it} + \sum_{t \in P^{\mathbf{c}}_i} \mathbf{R} \mathbf{x}_{\mathbf{1}it} \mathbf{x}'_{\mathbf{1}it} \mathbf{R}' + \mathbf{V}^{-1}_{\beta} \Big]^{-1}, \\
\mathbf{R} =& \begin{bmatrix} \mathbf{I} \\ \mathbf{0} \end{bmatrix}.
\end{aligned}
$$

Where $\mathbf{R}$ is a $J \times k_1$ matrix and $\mathbf{I}$ is an identity matrix of dimension k_1. The vector $\mathbf{R}$ selects observations from which consumer purchase incidence and product choice decisions are known. In this way, the estimate of $\boldsymbol{\beta}_n$ relies only upon observed outcomes.

3.3.3 Sampling $\boldsymbol{\Lambda}$

From the joint distribution (3.13), it can be shown that $\boldsymbol{\Lambda} \,|\, \mathbf{Z}_{it}, \theta \backslash \boldsymbol{\Lambda} \sim \mathcal{MN}(d, \mathbf{V}_{\boldsymbol{\beta}}, D)$ where

$$\begin{aligned} d =& D\Big[\mathbf{B}\mathbf{H}' + \boldsymbol{\Lambda}_0 A_d\Big], \\ D =& \Big[\mathbf{H}\mathbf{H}' + A_d\Big]^{-1}, \end{aligned}$$

and $\mathbf{B} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_N)$ and $\mathbf{H} = (\boldsymbol{H}_1, \ldots, \boldsymbol{H}_N)$.

3.3.4 Sampling $\mathbf{V}_{\boldsymbol{\beta}}$

From the joint distributions (3.13), it can be found that $\mathbf{V}_{\boldsymbol{\beta}}^{-1} \,|\, \mathbf{Z}_{it}, \theta \backslash \mathbf{V}_{\boldsymbol{\beta}} \sim \mathcal{W}(q, Q)$ where

$$\begin{aligned} q =& v + N + h, \\ Q =& \Big[\big(\mathbf{B} - \boldsymbol{\Lambda}\mathbf{H}\big)\big(\mathbf{B} - \boldsymbol{\Lambda}\mathbf{H}\big)' + \big(\boldsymbol{\Lambda} - \boldsymbol{\Lambda_0}\big) A_d \big(\boldsymbol{\Lambda} - \boldsymbol{\Lambda_0}\big)' + V^{-1}\Big]^{-1}, \end{aligned}$$

and $\mathbf{B}$, $\mathbf{H}$ and h are as defined before.

3.3.5 Sampling $\widetilde{\boldsymbol{\Omega}}$

Consider $\Phi \equiv (\widetilde{\boldsymbol{\Omega}}_{21}, \boldsymbol{\gamma}, \sigma_{33}, \ldots, \sigma_{JJ})$ where $\widetilde{\boldsymbol{\Omega}}_{21}$, $\boldsymbol{\gamma}$, and $(\sigma_{33}, \ldots, \sigma_{JJ})$ are as defined in section 3.2.2. The conditional densities of Φ have no known family of distribution. Hence, Φ is estimated with a tailored M-H step as first introduced in Chib (1995) and Chib and Jeliazkov (2001), and later expanded upon in Chib and Ramamurthy (2010) as a Tailored Randomized Block Metropolis-Hastings algorithm (TaRB M-H). The TaRB M-H process employed helps reduce the degree of autocorrelation between iterations of the sampling process, and efficiently explores the posterior distribution. As laid out in Chib and Ramamurthy (2010), the algorithm is as follows: at each iteration, h, of the sampling process Φ is broken down into a random number of randomly ordered blocks, B_h, and each block is then estimated using a tailored Metropolis Hastings step described below.

Let $\Phi_{h,b}$ be the b^{th} block of $\Phi = (\Phi_{h,1}, \ldots, \Phi_{h,b}, \ldots, \Phi_{h,B_h})$ for iteration h. Next, to form the tailored proposal density for $\Phi_{h,b}$ first one must find $\hat{\Phi}_{h,b}$, where

$$\hat{\Phi}_{h,b} = \arg\max_{\Phi_{h,b}} \{\pi(\mathbf{Z}_{it} \,|\, \theta \backslash \widetilde{\mathbf{\Omega}}, \Phi_{h,b}, \Phi_{h,-b})\, p(\Phi)\} \tag{3.14}$$

where $\pi(\mathbf{Z}_{it} \,|\, \theta \backslash \widetilde{\mathbf{\Omega}}, \Phi_{h,b}, \Phi_{h,-b})$ is as defined above, and $p(\Phi)$ are the assumed prior distributions on the elements of Φ. Once $\hat{\Phi}_{h,b}$ has been found, let $\mathbf{V}_{h,b}$ be the negative inverse hessian of $\Phi_{h,b}$ evaluated at $\hat{\Phi}_{h,b}$. Thus the tailored proposal distribution is as follows

$$f(\Phi_{h,b} \,|\, \boldsymbol{\theta} \backslash \widetilde{\mathbf{\Omega}}, \Phi_{h,-b}, \mathbf{Z}_{it}) = f_T(\hat{\Phi}_{h,b}, \mathbf{V}_{h,b}, \kappa) \tag{3.15}$$

where $f_T(\cdot)$ is a multivariate t-density with $\kappa > 2$ degrees of freedom.. Given a candidate draw, $\Phi^c_{h,b}$, from this tailored proposal distribution, the draw is accepted with the probability

$$\min\left\{1, \frac{\pi(\mathbf{Z}_{it} \,|\, \theta^c)\, p(\Phi^c_{h,b})\, f(\Phi_{h,b} \,|\, \boldsymbol{\theta} \backslash \widetilde{\mathbf{\Omega}}, \Phi_{h,-b}, \mathbf{Z}_{it})}{\pi(\mathbf{Z}_{it} \,|\, \theta)\, p(\Phi_{h,b})\, f(\Phi^c_{h,b} \,|\, \boldsymbol{\theta} \backslash \widetilde{\mathbf{\Omega}}, \Phi_{h,-b}, \mathbf{Z}_{it})}\right\}. \tag{3.16}$$

Finally, the algorithm is completed by repeating this process for each block in every iteration h.

3.4 Simulation Study

In this section, the behavior of the proposed model and the MCMC sampler is demonstrated through a simulated market with $J = 7$ alternatives. The data is generated with $N = 200$ and 800; with a purchase incidence rate of 52%. Contrasting the proposed model is a two-stage probit, whose covariance matrix is estimated through the techniques described in Chan and Jeliazkov (2009), and a model with a structured covariance matrix similar to Chib et al. (2004), where the covariance between product level utilities are normalized to zero.

The purchase incidence equation is taken to be a function of two covariates: an intercept and a random normal. Letting the first alternative be the outside option in the product choice equation; a model with 6 product-specific constants, one $\mathcal{U}(\{0,1\})$ to simulate promotional activity, and a constant value plus a random uniform to represent the price is considered. $T = 50$ time periods are simulated for each consumer, where in every time period all consumers faces the same set of covariates. An example of a choice outcome for a single consumer appears as:

$$\mathbf{Y} = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & -.78 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & .47 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1.52 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & .2 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1.1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1.39 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 2.9 \end{bmatrix}$$

where the first entry of $\mathbf{Y}$ corresponds to the purchase incidence, and the following entries correspond to the choice of alternative excluding the outside option.

To simulate heterogeneity, $\boldsymbol{\beta}_i$ is distributed with the mean vector $\overline{\boldsymbol{\beta}}$ and covariance matrix $\mathbf{V}_{\boldsymbol{\beta}}$, where $\mathbf{V}_{\boldsymbol{\beta}} = \mathbf{I}_{10} * .4$. Substitution patterns between choice alternatives are formed as described in section 3.2.2 with a set of three characteristics.

3.4.1 Estimation Results

The posterior means for the both the structured and unstructured covariance sampler are found from an MCMC run of length 60,000 after a burn in period of 20,000. The computational burden of the structured sampler, which utilizes the TaRB M-H process, varies by the number of individuals and time-period length. For example, with $N = 200$ and $T = 50$ the computational cost is approximately 400 seconds per 1000 iterations of the sampler. The unstructured two-stage probit's computational burden is lesser, taking about 200 seconds per 1000 iterations given the same N and T as before.

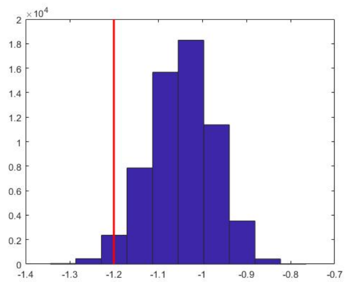
During the estimation process, a set of diffuse priors (similar to those used in Rossi et al. (1996) and Fong et al. (2014)) are considered for the model. Let Λ_0, γ_0, $\widetilde{\mathbf{\Omega}}_{21}^{0}$ be zero vectors, σ_0 be a vector of ones, $\mathbf{V}_{\mathbf{\Omega}_{21}} = \mathbf{I}_6$, $\mathbf{V}_\sigma = \mathbf{I}_5$, $A_d^{-1} = .001 * \mathbf{I}_{10}$, $v = 12$, and $V = \mathbf{I}_{10}$.

3.1 records the posterior mean and standard deviation of $\overline{\boldsymbol{\beta}}$ for both the structured and unstructured two-stage probit estimation. 3.2 gives the posterior mean and standard deviation of γ and the elements of $\widetilde{\mathbf{\Omega}}_{21}$. Finally, 3.3 records the elements of $\widetilde{\mathbf{\Omega}}_{22}$, allowing for a comparison of the estimated substitution patterns between the unstructured and structured model. The models presented, from left to right, are the proposed model, the unstructured probit, and a structured probit whose covariance matrix has had product level substitution patterns normalized to zero.

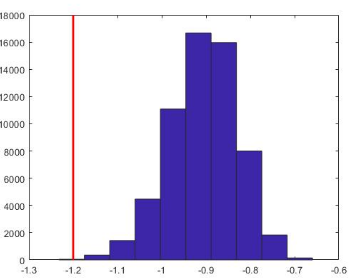
Since point estimates can be less informative about model parameters, figures 3.1-3.3 give the histograms of several posterior densities across models. In 3.1, the top three histograms gives the posterior density for $\overline{\boldsymbol{\beta}}_{10}$ when $N = 200$ and the bottom three for $N = 800$; from left to right is displayed the proposed model, the unstructured probit, and the structured probit. 3.2 gives the histograms for $\sigma_{5,5}$ for $N = 200$ and 800; whereas 3.3 displays the histograms for γ_1. Finally, to illustrate that model convergence is achieved with the TaRB M-H algorithm applied to the proposed covariance matrix, the trace plots for γ_1 are detailed in 3.4.

Figure 3.1: Posterior Density for $\overline{\boldsymbol{\beta}}_{10}$

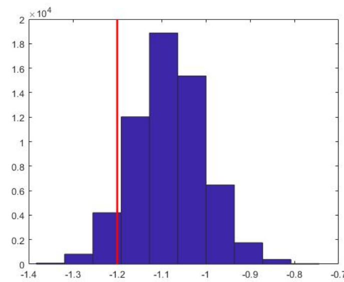
$N = 200$



Proposed

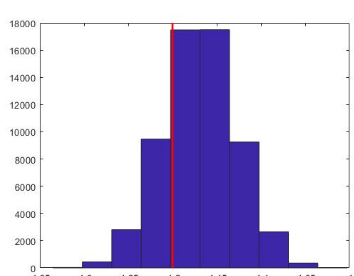


Unstructured

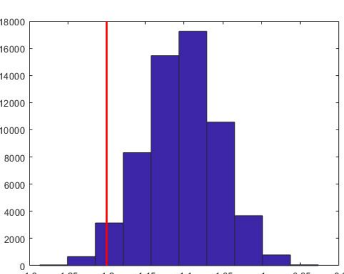


Structured

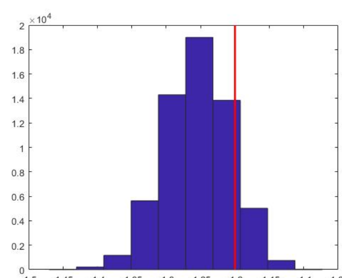
$N = 800$



Proposed



Unstructured



Structured

Table 3.1: Coefficient Estimates

Parameter	True Value	N = 200 Proposed Model	N = 200 Unstruct. Probit	N = 200 Struct. Probit	N = 800 Proposed Model	N = 800 Unstruct. Probit	N = 800 Struct. Probit
$\overline{\boldsymbol{\beta}}_1$ (s.d.)	0.700	0.674 (0.050)	0.676 (0.050)	0.674 (0.050)	0.689 (0.027)	0.690 (0.023)	0.688 (0.027)
$\overline{\boldsymbol{\beta}}_2$ (s.d.)	0.600	0.475 (0.047)	0.476 (0.047)	0.476 (0.047)	0.577 (0.024)	0.577 (0.021)	0.575 (0.024)
$\overline{\boldsymbol{\beta}}_3$ (s.d.)	0.900	0.749 (0.124)	0.668 (0.122)	0.492 (0.118)	0.919 (0.071)	0.897 (0.053)	0.739 (0.071)
$\overline{\boldsymbol{\beta}}_4$ (s.d.)	1.300	1.181 (0.124)	1.083 (0.115)	1.019 (0.120)	1.252 (0.071)	1.204 (0.053)	1.064 (0.072)
$\overline{\boldsymbol{\beta}}_5$ (s.d.)	0.400	0.301 (0.125)	0.305 (0.110)	0.062 (0.113)	0.440 (0.067)	0.460 (0.058)	0.191 (0.065)
$\overline{\boldsymbol{\beta}}_6$ (s.d.)	0.800	0.802 (0.104)	0.706 (0.102)	0.608 (0.094)	0.820 (0.064)	0.775 (0.049)	0.600 (0.062)
$\overline{\boldsymbol{\beta}}_7$ (s.d.)	1.500	1.447 (0.132)	1.253 (0.121)	1.159 (0.134)	1.450 (0.078)	1.363 (0.065)	1.210 (0.086)
$\overline{\boldsymbol{\beta}}_8$ (s.d.)	2.000	1.891 (0.141)	1.638 (0.142)	1.705 (0.147)	2.00 (0.081)	1.900 (0.073)	1.854 (0.086)
$\overline{\boldsymbol{\beta}}_9$ (s.d.)	0.550	0.490 (0.050)	0.424 (0.048)	0.514 (0.054)	0.513 (0.029)	0.482 (0.028)	0.566 (0.033)
$\overline{\boldsymbol{\beta}}_{10}$ (s.d.)	-1.200	−1.046 (0.073)	−0.906 (0.077)	−1.083 (0.080)	−1.170 (0.041)	−1.103 (0.036)	−1.254 (0.048)

Table 3.2: Covariance Parameters

Parameter	True Value	N = 200 Proposed Model	N = 200 Unstruct. Probit	N = 200 Struct. Probit	N = 800 Proposed Model	N = 800 Unstruct. Probit	N = 800 Struct. Probit
γ_1 (s.d.)	0.500	0.551 (0.090)	-	-	0.422 (0.067)	-	-
γ_2 (s.d.)	0.300	0.251 (0.115)	-	-	0.353 (0.089)	-	-
γ_3 (s.d.)	-0.250	−0.339 (0.151)	-	-	−0.191 (0.091)	-	-
σ_{21} (s.d.)	-0.200	−0.121 (0.106)	−0.066 (0.114)	−0.251 (0.107)	−0.211 (0.066)	−0.196 (0.055)	−0.334 (0.064)
σ_{31} (s.d.)	0.200	0.321 (0.104)	0.221 (0.109)	0.264 (0.124)	0.280 (0.069)	0.242 (0.063)	0.276 (0.072)
σ_{41} (s.d.)	-0.100	0.072 (0.109)	0.043 (0.137)	0.017 (0.119)	−0.068 (0.067)	−0.073 (0.063)	−0.106 (0.074)
σ_{51} (s.d.)	0.050	0.123 (0.098)	0.084 (0.114)	0.021 (0.112)	0.068 (0.106)	−0.013 (0.053)	−0.050 (0.067)
σ_{61} (s.d.)	0.200	0.120 (0.101)	0.055 (0.118)	0.122 (0.109)	0.256 (0.077)	0.241 (0.066)	0.315 (0.080)
σ_{71} (s.d.)	0.150	0.201 (0.100)	0.148 (0.124)	0.124 (0.096)	0.070 (0.071)	0.081 (0.064)	0.049 (0.077)

The findings presented in tables 3.1-3.3 suggest that, under the assumption that similarity can determine substitution patterns, the model proposed in this paper offers an alternative to more restrictive covariance assumptions. The proposed probit model estimates are closer

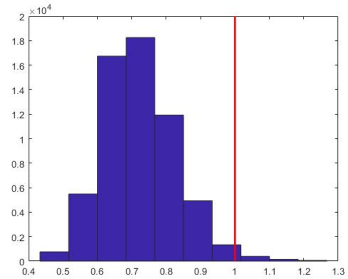
Table 3.3: Covariance Parameters, Cont'd

Parameter	True Value	$N = 200$			$N = 800$		
		Proposed Model	Unstruct. Probit	Struct. Probit	Proposed Model	Unstruct. Probit	Struct. Probit
σ_{33} (s.d.)	1.250	1.039 (0.131)	0.696 (0.159)	0.755 (0.092)	1.242 (0.085)	1.125 (0.132)	1.189 (0.111)
σ_{44} (s.d.)	1.300	1.061 (0.136)	0.851 (0.194)	0.982 (0.092)	1.209 (0.093)	1.191 (0.123)	1.282 (0.112)
σ_{55} (s.d.)	1.000	0.723 (0.105)	0.484 (0.097)	0.580 (0.081)	0.923 (0.074)	0.766 (0.096)	0.930 (0.096)
σ_{66} (s.d.)	1.200	0.953 (0.149)	0.822 (0.172)	0.745 (0.084)	1.111 (0.102)	1.028 (0.118)	0.990 (0.103)
σ_{77} (s.d.)	1.000	0.657 (0.113)	0.459 (0.116)	0.488 (0.070)	0.968 (0.082)	0.838 (0.112)	0.899 (0.105)
σ_{32} (s.d.)	0.146	0.129 (0.038)	0.329 (0.118)	-	0.164 (0.028)	0.287 (0.080)	-
σ_{42} (s.d.)	0.277	0.230 (0.135)	0.393 (0.108)	-	0.263 (0.069)	0.386 (0.067)	-
σ_{52} (s.d.)	0.007	-0.022 (0.055)	0.083 (0.120)	-	0.029 (0.030)	0.012 (0.075)	-
σ_{62} (s.d.)	0.616	0.630 (0.093)	0.682 (0.094)	-	0.557 (0.069)	0.577 (0.065)	-
σ_{72} (s.d.)	0.257	0.223 (0.084)	0.199 (0.110)	-	0.305 (0.065)	0.295 (0.072)	-
σ_{43} (s.d.)	0.109	0.095 (0.028)	0.061 (0.133)	-	0.121 (0.019)	0.126 (0.103)	-
σ_{53} (s.d.)	0.484	0.475 (0.079)	0.233 (0.091)	-	0.442 (0.049)	0.358 (0.086)	-
σ_{63} (s.d.)	0.220	0.175 (0.087)	0.078 (0.140)	-	0.279 (0.065)	0.118 (0.108)	-
σ_{73} (s.d.)	0.472	0.443 (0.091)	0.262 (0.105)	-	0.453 (0.050)	0.438 (0.093)	-
σ_{54} (s.d.)	0.220	0.175 (0.087)	0.154 (0.098)	-	0.279 (0.065)	0.340 (0.084)	-
σ_{64} (s.d.)	0.541	0.561 (0.085)	0.487 (0.139)	-	0.470 (0.063)	0.500 (0.091)	-
σ_{74} (s.d.)	0.044	0.026 (0.034)	0.024 (0.113)	-	0.055 (0.018)	0.119 (0.098)	-
σ_{65} (s.d.)	0.109	0.095 (0.028)	0.017 (0.111)	-	0.121 (0.019)	0.040 (0.102)	-
σ_{75} (s.d.)	0.378	0.355 (0.096)	0.228 (0.081)	-	0.342 (0.051)	0.140 (0.078)	-
σ_{76} (s.d.)	0.176	0.148 (0.057)	0.087 (0.095)	-	0.210 (0.043)	0.256 (0.086)	-

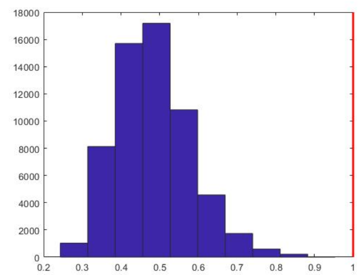
to the true mean values, whereas those of the unstructured probit exhibit a greater degree of bias, likely resulting from the infeasible estimation of a full probit covariance matrix. As N increases to 800, we observe the parameters of the unstructured and proposed model values converging towards the true values; this is unsurprising, as one would expect increased observations to overcome the issues associated with estimating an unstructured probit model. In contrast, increased observations do not provide the same boost to parameter estimation

Figure 3.2: Posterior Density for $\sigma_{5,5}$

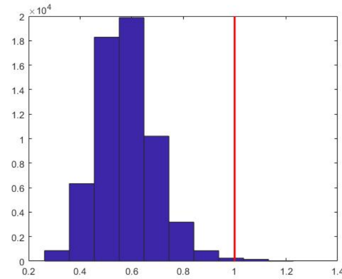
$N = 200$



Proposed

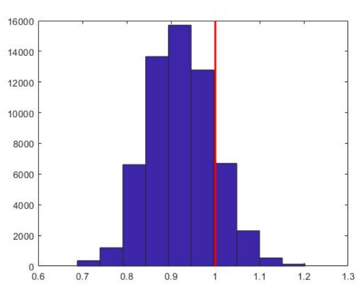


Unstructured

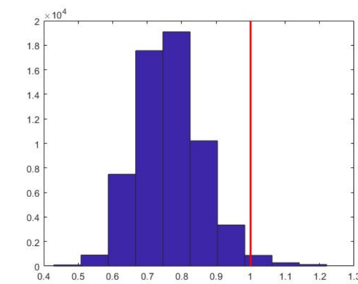


Structured

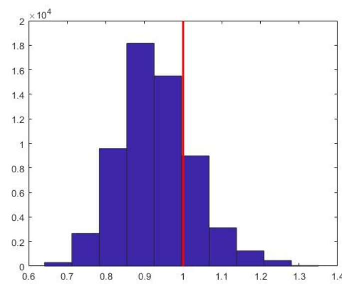
$N = 800$



Proposed



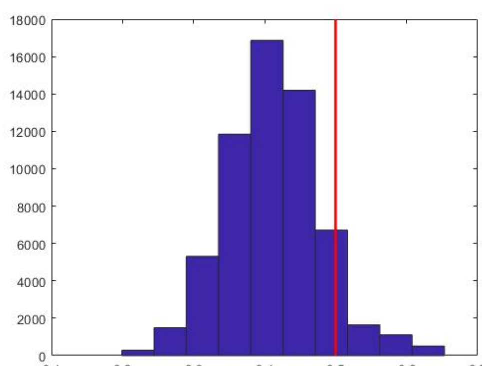
Unstructured



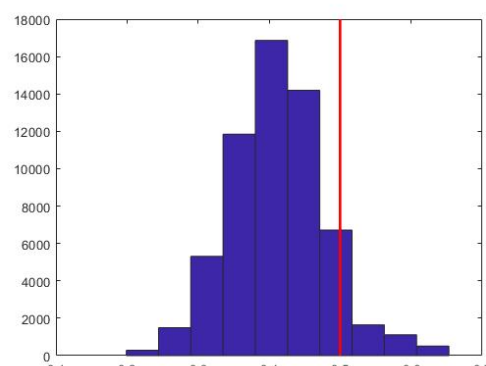
Structured

Figure 3.3: Posterior Density for γ_1

$N = 200$



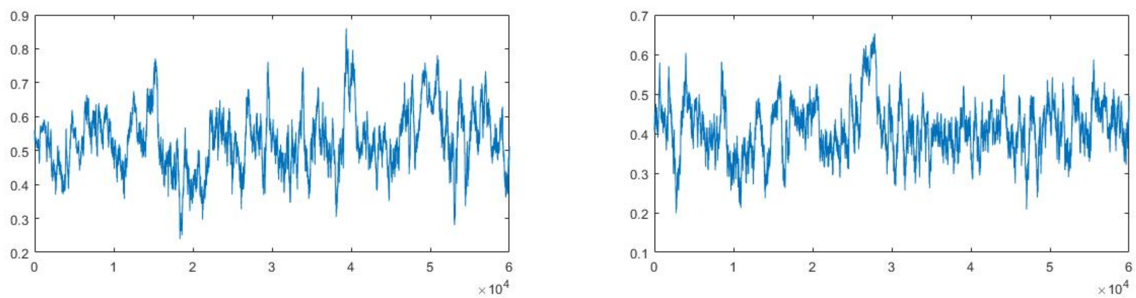
$N = 800$



in the most restrictive fully structured model. This is likely due to the overly restrictive substitution pattern assumptions, and suggests that such assumptions (as seen in Chib et al.

Figure 3.4: Trace Plots for γ_1

$N = 200$ $\qquad$ $N = 800$



(2004)) may inhibit proper estimation of parameter values.

3.5 Conclusion

In this paper, a method of estimating a two-stage multinomial probit model of demand with substitution patterns and individual parameter heterogeneity is proposed. In contrast to other more restrictive modeling assumptions, this paper serves as a demonstration as to how product similarity can be used to structure a probit covariance matrix. In addition, described in this work is a method by which random coefficients can be introduced into the two-stage estimation process.

In comparison to the existing literature, the model proposed in this paper displays three key attributes: first, as in Chib et al. (2004) the probit structure allows for correlation between the purchase incidence and product choice equations. Second, building upon the work of Chib et al. (2009) and Fong et al. (2014) this paper demonstrates how individual level coefficients in a multistage model can be estimated for a probit distribution through a Gibbs sampling process. Finally, unlike prior works, the proposed model allows for the feasible estimation of substitution patterns between product level utilities through a measure of product similarity as defined above. An MCMC process is used to determine the posterior distribution of parameter values.

The model proposed in this paper was then compared to both an unstructured probit model estimated through the techniques detailed in Chan and Jeliazkov (2009) and a structured probit whose restrictions were similar to those seen in Chib et al. (2004). Simulating markets of 200 and 800 individuals, the proposed model outperformed both the unstructured and restrictive probit designation. Surprisingly, as the number of individuals grew, the expected parameter values found from the posterior distribution of the model similar to that in Chib et al. (2004) demonstrated a larger degree of bias; suggesting assumptions made to maintain the feasible estimation of a probit covariance can be overly restrictive.

Finally, a multitude of methods for defining product similarity exist. For instance, the work of Cohen (2010) and Dotson et al. (2018) demonstrate how similarity can be used to define a probit correlation matrix; as in this paper, these methods of defining product similarity can be extended to estimate full probit covariance matrices. Given the both the flexibility of Bayesian probit modeling and the benefits of the TaRB MH algorithm, it stands to reason that the feasible estimation of high-dimensional multinomial probit models and their covariance matrices are achievable through such strategies. The primary intention of this paper was then to offer a guide for such an estimation, and to demonstrate one method by which a researcher could employ similarity to feasible estimate substitution patterns in lieu of restrictive covariance assumptions.

Bibliography

Acton, Rachel B, Lana Vanderlee, Jean Adams, Sharon I Kirkpatrick, Lilia S Pedraza, Gary Sacks, Christine M White, Martin White, and David Hammond. 2022. "Tax awareness and perceived cost of sugar-sweetened beverages in four countries between 2017 and 2019: findings from the international food policy study." *International Journal of Behavioral Nutrition and Physical Activity* 19 (1): 1–18.

Akerlof, George A. 1991. "Procrastination and Obedience." *The American Economic Review* 81 (2): 1–19.

Ali, Fatma Romeh M, Megan C Diaz, Donna Vallone, Michael A Tynan, Jamie Cordova, Elizabeth L Seaman, Katrina F Trivers, Barbara A Schillo, Brandon Talley, and Brian A King. 2020. "E-cigarette unit sales, by product and flavor type—United States, 2014–2020." *Morbidity and Mortality Weekly Report* 69 (37): 1313.

Ali, Fatma Romeh M., Brian A. King, Elizabeth L. Seaman, Donna Vallone, and Barbara Schillo. 2022. "Impact of Massachusetts law prohibiting flavored tobacco products sales on cross-border cigarette sales." *PLOS ONE* 17 (9): 1–9.

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

Andrews, Rick, and Imran Currim. 2009. "Multi-stage purchase decision models: Accommodating response heterogeneity, common demand shocks, and endogeneity using disaggregate data." *International Journal of Research in Marketing - INT J RES MARK* 26.

Arora, Neeraj, Greg M. Allenby, and James L. Ginter. 1998. "A Hierarchical Bayes Model of Primary and Secondary Demand." *Marketing Science* 17 (1): 29–44.

Balbo, Nicoletta, Francesco C Billari, and Melinda Mills. 2013. "Fertility in advanced societies: A review of research." *European Journal of Population/Revue europeenne de demographie* 29 (1): 1–38.

Baltagi, Badi H., and Dan Levin. 1986. "Estimating Dynamic Demand for Cigarettes Using Panel Data: The Effects of Bootlegging, Taxation and Advertising Reconsidered." *The Review of Economics and Statistics* 68 (1): 148–155.

Baltagi, Badi H., and Dan Levin. 1992. "Cigarette taxation: Raising revenues and reducing consumption." *Structural Change and Economic Dynamics* 3 (2): 321–335.

Bayer, Patrick, Fernando Ferreira, and Robert McMillan. 2007. "A Unified Framework for Measuring Preferences for Schools and Neighborhoods." *Journal of Political Economy* 115 (4): 588–638.

Becker, Gary S., and Kevin M. Murphy. 1988. "A Theory of Rational Addiction." *Journal of Political Economy* 96 (4): 675–700.

Berry, Steven, James Levinsohn, and Ariel Pakes. 1995. "Automobile Prices in Market Equilibrium." *Econometrica* 63 (4): 841–890.

Berry, Steven, James Levinsohn, and Ariel Pakes. 2004. "Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market." *Journal of Political Economy* 112 (1): 68–105.

Berry, Steven T. 1994. "Estimating Discrete-Choice Models of Product Differentiation." *The RAND Journal of Economics* 25 (2): 242–262.

Bleich, Sara N, Hannah G Lawman, Michael T LeVasseur, Jiali Yan, Nandita Mitra, Caitlin M Lowery, Ana Peterhans, Sophia Hua, Laura A Gibson, and Christina A Roberto. 2020. "The Association Of A Sweetened Beverage Tax With Changes In Beverage Prices And Purchases At Independent Stores." *Health Affairs* 39 (7): 1130–1139.

Breslow, Samantha K. 2019. "Chicago and Cook County Amusement Tax." *The National Law Review* IX (114). https://www.natlawreview.com/article/chicago-and-cook-county-amusement-tax.

Bucklin, Randolph E., and Sunil Gupta. 1992. "Brand Choice, Purchase Incidence, and Segmentation: An Integrated Modeling Approach." *Journal of Marketing Research* 29 (2): 201–215.

Bucklin, Randolph E., Sunil Gupta, and S. Siddarth. 1998. "Determining Segmentation in Sales Response across Consumer Purchase Behaviors." *Journal of Marketing Research* 35 (2): 189–197.

Bucklin, Randolph E., and James M. Lattin. 1991. "A Two-State Model of Purchase Incidence and Brand Choice." *Marketing Science* 10 (1): 24–39.

Burda, Martin, Matthew Harding, and Jerry Hausman. 2008. "A Bayesian mixed logit–probit model for multinomial choice." *Journal of Econometrics* 147 (2): 232–246, Estimating demand systems and measuring consumer preferences.

Bygvrå, Susanne. 2009. "Distance and cross-border shopping for alcohol – evidence from Danes' cross-border shopping 1986–2003." *Nordic Studies on Alcohol and Drugs* 26 (2): 141–163.

Capps, Cory, David Dranove, and Mark Satterthwaite. 2003. "Competition and Market Power in Option Demand Markets." *The RAND Journal of Economics* 34 (4): 737–763.

Caves, Kevin W. 2005. *Empirical studies of supply and demand dynamics*. Ph.D. dissertation, `https://www.proquest.com/dissertations-theses/empirical-studies-supply-demand-dynamics/docview/305033880/se-2`, Copyright - Database copyright ProQuest LLC; ProQuest does not claim copyright in the individual underlying works; Last updated - 2021-10-14.

Cawley, John, and David Frisvold. 2015. "The Incidence of Taxes on Sugar-Sweetened Beverages: The Case of Berkeley, California." Working Paper 21465, National Bureau of Economic Research.

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

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.

CDC Smoking and Tobacco Use. 2020. "Tobacco-Related Mortality." October, `https://www.cdc.gov/tobacco/data_statistics/fact_sheets/health_effects/tobacco_related_mortality/index.htm`.

Chaiton, Michael, Ismina Papadhima, Robert Schwartz, Joanna E Cohen, Eric K Soule, Bo Zhang, and Thomas Eissenberg. 2020. "Product substitution after a real-world menthol ban: a cohort study." *Tobacco regulatory science* 6 (3): 205.

Chan, Joshua Chi-Chun, and Ivan Jeliazkov. 2009. "MCMC Estimation of Restricted Covariance Matrices." *Journal of Computational and Graphical Statistics* 18 (2): 457–480.

Chandra, Ambarish, Keith Head, and Mariano Tappata. 2014. "The Economics of Cross-Border Travel." *The Review of Economics and Statistics* 96 (4): 648–661.

Chib, Siddhartha. 1995. "Marginal Likelihood from the Gibbs Output." *Journal of the American Statistical Association* 90 (432): 1313–1321.

Chib, Siddhartha, Edward Greenberg, and Ivan Jeliazkov. 2009. "Estimation of Semiparametric Models in the Presence of Endogeneity and Sample Selection." *Journal of Computational and Graphical Statistics* 18 (2): 321–348.

Chib, Siddhartha, and Ivan Jeliazkov. 2001. "Marginal Likelihood From the Metropolis–Hastings Output." *Journal of the American Statistical Association* 96 (453): 270–281.

Chib, Siddhartha, and Srikanth Ramamurthy. 2010. "Tailored randomized block MCMC methods with application to DSGE models." *Journal of Econometrics* 155 (1): 19 – 38.

Chib, Siddhartha, P.B. Seetharaman, and Andrei Strijnev. 2004. "Model of Brand Choice with a No-Purchase Option Calibrated to Scanner-Panel Data." *Journal of Marketing Research* 41 (2): 184–196.

Chintagunta, Pradeep K. 1993. "Investigating Purchase Incidence, Brand Choice and Purchase Quantity Decisions of Households." *Marketing Science* 12 (2): 184–208.

Chintagunta, Pradeep K., and Jean-Pierre Dubé. 2005. "Estimating a Stockkeeping-Unit-Level Brand Choice Model That Combines Household Panel Data and Store Data." *Journal of Marketing Research* 42 (3): 368–379.

Cohen, Michael. 2010. "A Structured Covariance Probit Demand Model." Research Reports 149970, University of Connecticut, Food Marketing Policy Center. 10.22004/ag.econ.149970.

Cumminggs, Kenneth Michael. 1999. "Expert Witness Report In Refrence to: Falise, ET AL. V. The American Tobacco Company, ET AL." `https://www.industrydocuments.ucsf.edu/docs/qqyc0178`, Philip Morris Records; Master Settlement Agreement.

Davis, Peter. 2006. "Spatial Competition in Retail Markets: Movie Theaters." *The RAND Journal of Economics* 37 (4): 964–982.

DellaVigna, Stefano, and Matthew Gentzkow. 2019. "Uniform Pricing in U.S. Retail Chains." *The Quarterly Journal of Economics* 134 (4): 2011–2084.

Delnevo, Cristine D, Ollie Ganz, and Renee D Goodwin. 2020. "Banning Menthol Cigarettes: A Social Justice Issue Long Overdue." *Nicotine & Tobacco Research* 22 (10): 1673–1675.

Dotson, Jeffrey P., John R. Howell, Jeff D. Brazell, Thomas Otter, Peter J. Lenk, Steve MacEachern, and Greg M. Allenby. 2018. "A Probit Model with Structured Covariance for Similarity Effects and Source of Volume Calculations." *Journal of Marketing Research* 55 (1): 35–47.

Dubois, Pierre, Rachel Griffith, and Martin O'Connell. 2020. "How Well Targeted Are Soda Taxes?" *American Economic Review* 110 (11): 3661–3704.

Duchneskie, John. 2016. "Philly's daily commute: How do your habits compare?." *The Philadelphia Inquirer*. https://www.inquirer.com/philly/business/transportation/405506645.html.

East, Katherine A., Jessica L. Reid, Robin Burkhalter, Loren Kock, Andrew Hyland, Geoffrey T. Fong, and David Hammond. 2022. "Evaluating the Outcomes of the Menthol Cigarette Ban in England by Comparing Menthol Cigarette Smoking Among Youth in England, Canada, and the US, 2018-2020." *JAMA Network Open* 5 (5): e2210029–e2210029.

Erdem, Tülin, Susumu Imai, and Michael Keane. 2003. "Brand and Quantity Choice Dynamics Under Price Uncertainty." *Quantitative Marketing and Economics (QME)* 1 (1): 5–64.

Evans, William N, Jeanne S Ringel, and Diana Stech. 1999. "Tobacco taxes and public policy to discourage smoking." *Tax policy and the economy* 13 1–55.

FDA. 2021. "FDA Commits to Evidence-Based Actions Aimed at Saving Lives and Preventing Future Generations of Smokers." `https://www.fda.gov/news-events/press-announcements/fda-commits-evidence-based-actions-aimed-saving\-lives-and-preventing-future-generations-smokers`.

Fisher, Ronald C. 1980. "Local Sales Taxes: Tax Rate Differentials, Sales Loss, and Revenue Estimation." *Public Finance Quarterly* 8 (2): 171–188.

Fong, Duncan, Sunghoon Kim, Zhe Chen, and Wayne Desarbo. 2014. "A Bayesian multinomial probit model for the analysis of panel choice data." *Psychometrika* 81.

Fong, Geoffrey T et al. 2022a. "Impact of Canada's menthol cigarette ban on quitting among menthol smokers: pooled analysis of pre–post evaluation from the ITC Project and the Ontario Menthol Ban Study and projections of impact in the USA." *Tobacco Control.*

Fong, Geoffrey T., Gang Meng, Lorraine V. Craig, and Janet Chung-Hall. 2022b. "Illicit cigarette purchasing after implementation of a menthol ban in Canada: Findings from the 2016-2018 ITC Four Country Smoking and Vaping Surveys." *Tobacco Prevention & Cessation* 8 (Supplement): .

Friberg, Richard, Frode Steen, and Simen A. Ulsaker. 2018. "Hump-shaped cross-price effects and the extensive margin in cross-border shopping." Discussion Paper Series in Economics 20/2019, Norwegian School of Economics.

Gardiner, Phillip. 2004. "The African Americanization of menthol cigarettes in the United States." *Nicotine & tobacco research : official journal of the Society for Research on Nicotine and Tobacco* 6 Suppl 1 S55–65.

Garrison, Cassandra, and Adriana Barrera. 2022. "Mexico suspends gasoline subsidy as U.S. drivers cross border for deals." *Reuters.* https://www.reuters.com/world/americas/mexico-suspends-subsidy-gasoline-border-region-with-us-2022-04-02/.

Goldin, Jacob, Tatiana Homonoff, and Katherine Meckel. 2022. "Issuance and Incidence: SNAP Benefit Cycles and Grocery Prices." *American Economic Journal: Economic Policy* 14 (1): 152–78.

Goldszmidt, Ariel, John A. List, Robert D. Metcalfe, Ian Muir, V. Kerry Smith, and Jenny Wang. 2020. "The Value of Time in the United States: Estimates from Nationwide Natural Field Experiments." Working Paper 28208, National Bureau of Economic Research.

Goolsbee, Austan, and Amil Petrin. 2004. "The Consumer Gains from Direct Broadcast Satellites and the Competition with Cable TV." *Econometrica* 72 (2): 351–381.

Gordon, Brett R., and Baohong Sun. 2015. "A Dynamic Model of Rational Addiction: Evaluating Cigarette Taxes." *Marketing Science* 34 (3): 452–470.

Grieco, Paul LE, Charles Murry, Joris Pinkse, and Stephan Sagl. 2022. "Conformant and Efficient Estimation of Discrete Choice Demand Models." working paper, The Pennsylvania State University.

Grigolon, Laura, and Frank Verboven. 2014. "Nested Logit or Random Coefficients Logit? A Comparison of Alternative Discrete Choice Models of Product Differentiation." *The Review of Economics and Statistics* 96 (5): 916–935.

Harding, Matthew, Ephraim Leibtag, and Michael F. Lovenheim. 2012. "The Heterogeneous Geographic and Socioeconomic Incidence of Cigarette Taxes: Evidence from Nielsen Homescan Data." *American Economic Journal: Economic Policy* 4 (4): 169–98.

Hendel, Igal, and Aviv Nevo. 2006a. "Measuring the Implications of Sales and Consumer Inventory Behavior." *Econometrica* 74 (6): 1637–1673.

Hendel, Igal, and Aviv Nevo. 2006b. "Sales and consumer inventory." *RAND Journal of Economics* 37 (3): 543–561.

Hickman, Elise, and Ilona Jaspers. 2022. "Evolving chemical landscape of e-cigarettes, 2021." *Tobacco Control* 31 (e1): e1–e2.

Hidayat, Budi, and Hasbullah Thabrany. 2011. "Are smokers rational addicts? Empirical evidence from the Indonesian Family Life Survey." *Harm reduction journal* 8 6.

Hollenbeck, Brett, and Kosuke Uetake. 2021. "Taxation and Market Power in the Legal Marijuana Industry." *The RAND Journal of Economics* 52 (3): 559–595.

Houde, Jean-François. 2012. "Spatial Differentiation and Vertical Mergers in Retail Markets for Gasoline." *American Economic Review* 102 (5): 2147–82.

Houthakker, Hendrik S, and Lester D Taylor. 1970. "Consumer demand in the United States." *Mass., Harvard University Press.*

Hymel, Kent, Kenneth Small, and Kurt Van Dender. 2010. "Induced demand and rebound effects in road transport." *Transportation Research Part B: Methodological* 44 1220–1241.

Issabakhsh, Mona, Rafael Meza, Yameng Li, Zhe Yuan, Luz Maria Sanchez-Romero, and David T Levy. 2022. "Public health impact of a US menthol cigarette ban on the non-Hispanic black population: a simulation study." *Tobacco Control.*

Jones, Andrew M. 1989. "A SYSTEMS APPROACH TO THE DEMAND FOR ALCOHOL AND TOBACCO." *Bulletin of Economic Research* 41 (2): 85–106.

Jorda, Vanesa, José María Sarabia, and Markus Jäntti. 2021. "Inequality measurement with grouped data: Parametric and non-parametric methods." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 184 (3): 964–984.

Kasza, Karin A, Kathryn C Edwards, Andrew Anesetti-Rothermel, MeLisa R Creamer, K Michael Cummings, Raymond S Niaura, Akshika Sharma, Stephanie R Pitts, Sara K Head, Colm D Everard et al. 2022. "E-cigarette use and change in plans to quit cigarette smoking among adult smokers in the United States: Longitudinal findings from the PATH Study 2014–2019." *Addictive Behaviors* 124 107124.

Kasza, Karin A, Kathryn C Edwards, Heather L Kimmel, Andrew Anesetti-Rothermel, K Michael Cummings, Raymond S Niaura, Akshika Sharma, Erin M Ellis, Rebecca Jackson, Carlos Blanco et al. 2021. "Association of e-cigarette use with discontinuation of cigarette smoking among adult smokers who were initially never planning to quit." *JAMA network open* 4 (12): e2140880–e2140880.

Kifer, Avigail. 2015. "The Incidence of a Soda Tax, in Pennies and Pounds." Working Paper 1-017, Kilts Center for Marketing at Chicago Booth – Nielsen Dataset Paper Series.

Lee, Stella Juhyun, Vaughan W. Rees, Noam Yossefy, Karen M. Emmons, and Andy S. L. Tan. 2020. "Youth and Young Adult Use of Pod-Based Electronic Cigarettes From 2015 to 2019: A Systematic Review." *JAMA Pediatrics* 174 (7): 714–720.

Lee, Youn Ok, and Stanton A Glantz. 2011. "Menthol: putting the pieces together." *Tobacco Control* 20 (Suppl 2): ii1–ii7.

Levy, David T et al. 2021a. "An Expert Elicitation on the Effects of a Ban on Menthol Cigarettes and Cigars in the United States." *Nicotine & Tobacco Research* 23 (11): 1911–1920.

Levy, David T et al. 2021b. "Public health impact of a US ban on menthol in cigarettes and cigars: a simulation study." *Tobacco Control*.

Lewis, M. Jane, and Olivia Wackowski. 2006. "Dealing with an innovative industry: A look at flavored cigarettes promoted by mainstream brands." *American Journal of Public Health* 96 (2): 244–251.

Li, Shanjun, Joshua Linn, and Erich Muehlegger. 2014. "Gasoline Taxes and Consumer Behavior." *American Economic Journal: Economic Policy* 6 (4): 302–342.

McFadden, Daniel, Antti Talvitie, and Associates. 1977. *Demand Model Estimation and Validation*. Institute of Transportation Studies, Berkeley CA.

Miller, Nathan H., and Matthew C. Weinberg. 2017. "Understanding the Price Effects of the MillerCoors Joint Venture." *Econometrica* 85 (6): 1763–1791.

Mills, Sarah D., Lisa Henriksen, Shelley D. Golden, Rachel Kurtzman, Amanda Y. Kong, Tara L. Queen, and Kurt M. Ribisl. 2018. "Disparities in retail marketing for menthol cigarettes in the United States, 2015." *Health & Place* 53 62–70.

Miravete, Eugenio J., Katja Seim, and Jeff Thurk. 2018. "MARKET POWER AND THE LAFFER CURVE." *Econometrica* 86 (5): 1651–1687.

Mullahy, John. 1985. *Cigarette Smoking: Habits, Health Concerns, and Heterogenous Unobservables in a Microeconomic Analysis of Consumer Demand.* Ph.D. dissertation, University of Virginia, `https://www.proquest.com/dissertations-theses/cigarette-smoking-habits-health-concerns/docview/303428328/se-2`.

Murry, Charles, and Yiyi Zhou. 2020. "Consumer Search and Automobile Dealer Colocation." *Management Science* 66 (5): 1909–1934.

Nardone, Natalie, Gideon St Helen, Newton Addo, Sandra Meighan, and Neal L Benowitz. 2019. "JUUL electronic cigarettes: nicotine exposure and the user experience." *Drug and alcohol dependence* 203 83–87.

National Research Council. 2015. *Understanding the U.S. Illicit Tobacco Market: Characteristics, Policy Context, and Lessons from International Experiences.* The National Academies Press.

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. 2001. "Measuring Market Power in the Ready-to-Eat Cereal Industry." *Econometrica* 69 (2): 307–342.

Olesiński, Bartosz. 2020. "The Analysis of the Tobacco Product Bans Using a Random Coefficients Logit Model." *Central European Journal of Economic Modelling and Econometrics* 12 (2): 113–144.

Paraje, Guillermo, Michal Stoklosa, and Evan Blecher. 2022. "Illicit trade in tobacco products: recent trends and coming challenges." *Tobacco Control* 31 (2): 257–262.

Pollay, Richard W., Jung S. Lee, and David Carter-Whitney. 1992. "Separate, but Not Equal: Racial Segmentation in Cigarette Advertising." *Journal of Advertising* 21 (1): 45–57.

Ramamurthi, Divya, Cindy Chau, and Robert K Jackler. 2019. "JUUL and other stealth vaporisers: hiding the habit from parents and teachers." *Tobacco control* 28 (6): 610–616.

Roberto, Christina A., Hannah G. Lawman, Michael T. LeVasseur, Nandita Mitra, Ana Peterhans, Bradley Herring, and Sara N. Bleich. 2019. "Association of a Beverage Tax on Sugar-Sweetened and Artificially Sweetened Beverages With Changes in Beverage Prices and Sales at Chain Retailers in a Large Urban Setting." *JAMA* 321 (18): 1799–1810.

Rossi, Peter, Robert McCulloch, and Greg Allenby. 1996. "The Value of Purchase History Data in Target Marketing." *Marketing Science* 15 321–340.

Sakuma, Kari-Lyn K., Jamie Quibol Felicitas-Perkins, Lyzette Blanco, Pebbles Fagan, Eliseo J. Pérez-Stable, Kim Pulvers, Devan Romero, and Dennis R. Trinidad. 2016. "Tobacco use disparities by racial/ethnic groups: California compared to the United States." *Preventive Medicine* 91 224–232.

Sakuma, Kari-Lyn K, John P Pierce, Pebbles Fagan, France T Nguyen-Grozavu, Eric C Leas, Karen Messer, Martha M White, Amanda S Tieu, and Dennis R Trinidad. 2020. "Racial/Ethnic Disparities Across Indicators of Cigarette Smoking in the Era of Increased Tobacco Control, 1992–2019." *Nicotine & Tobacco Research* 23 (6): 909–919.

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.

Stoklosa, Michal. 2019. "No surge in illicit cigarettes after implementation of menthol ban in Nova Scotia." *Tobacco Control* 28 (6): 702–704.

Thomadsen, Raphael. 2005. "The Effect of Ownership Structure on Price in Geographically Differentiated Industries." *RAND Journal of Economics* 36 908–929.

Toll, B A, and P M Ling. 2005. "The Virginia Slims identity crisis: an inside look at tobacco industry marketing to women." *Tobacco Control* 14 (3): 172–180.

Tosun, Mehmet, and Mark Skidmore. 2007. "Cross-Border Shopping and the Sales Tax: An Examination of Food Purchases in West Virginia." *Topics in Economic Analysis & Policy* 7 1877–1877.

Train, Kenneth. 1999. "Halton sequences for mixed logit." working paper, Department of Economics, UC Berkeley.

Train, Kenneth E. 2009. *Discrete choice methods with simulation*. Cambridge university press.

Tuchman, Anna E. 2019. "Advertising and Demand for Addictive Goods: The Effects of E-Cigarette Advertising." *Marketing Science* 38 (6): 994–1022.

University of Bath. 2012. "Tobacco Tatics." `https://tobaccotactics.org`.

University of California San Francisco. 1999. "Truth Tobacco Industry Documents." `https://www.industrydocuments.ucsf.edu/tobacco/`.

U.S. Bureau of Labor Statistics. 2017. "Average Energy Prices, Philadelphia-Wilmington-Atlantic City – May 2017." News release.

U.S. Bureau of Labor Statistics. 2018. "Occupational Employment and Wages in Philadelphia-Camden-Wilmington – May 2017." News release.

Vallone, Donna M., Alison F. Cuccia, Jodie Briggs, Haijun Xiao, Barbara A. Schillo, and Elizabeth C. Hair. 2020. "Electronic Cigarette and JUUL Use Among Adolescents and Young Adults." *JAMA Pediatrics* 174 (3): 277–286.

Wang, Ao. 2021. "A BLP Demand Model of Product-Level Market Shares with Complementarity." The Warwick Economics Research Paper Series (TWERPS) 1351, University of Warwick.

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.

Wang, Teresa W, Kat Asman, Andrea S Gentzke, Karen A Cullen, Enver Holder-Hayes, Carolyn Reyes-Guzman, Ahmed Jamal, Linda Neff, and Brian A King. 2018. "Tobacco product use among adults—United States, 2017." *Morbidity and Mortality Weekly Report* 67 (44): 1225.

Winston, Gordon. 1980. "Addiction and backsliding: A theory of compulsive consumption." *Journal of Economic Behavior & Organization* 1 (4): 295–324.

Yazzie, Del, Kristen Tallis, Caleigh Curley, Priscilla R. Sanderson, Regina Eddie, Timothy K. Behrens et al. 2020. "The Navajo Nation Healthy Diné Nation Act: A Two Percent Tax on Foods of Minimal-to-No Nutritious Value, 2015–2019." *Preventing Chronic Disease* 17:200038.

Appendix A

Chapter 1

A1 Purchase Frequency and Stockpiling

In analyzing the frequency of cigarette purchases and potential stockpiling behavior, I calculate both the number of days between cigarette store trips and the occurrence of short-lived price reductions, "sales".[A1] As suggested in Hendel and Nevo (2006b), if significant storage behavior is observed, cigarette sale occasions should be positively correlated with the number of days between store trips (as households increase their stock of stored products when prices are reduced). Controlling for outliers in my sample—particularly on-again, off-again smokers—I subset my sample to those store trips where the difference between the current and next purchase date is less than or equal to 4 weeks. I find the average number of days between each trip to be 6.77, and 68% of all cigarette store trips fall within 7 days of a prior purchase.

To address cigarette storability, I consider a regression of the number of days until the next

[A1] I define cigarette sale occasions similar to how they are defined in Hendel and Nevo (2006b)—any time in which weekly cigarette price falls at least 5 percent below the modal price in each DMA. Weekly cigarette DMA-level price is taken to be the quantity weighted average price of all observed sales at the DMA/week level.

store trips on cigarette sales occasions. To control for individual preferences, time trends, and seasonality, I include household and week fixed effects, and cluster the errors at the household level. Table A1 presents my results. I find the regression coefficient for sale occasions to be negative and statistically insignificant—suggesting temporary price reductions are uncorrelated with cigarette purchase frequency. Therefore, I conclude storability does not appear to play a significant role in determining time between cigarette purchase occasions.

Table A1: Days Until Next Store Trip Regressed on Cigarette Sales occasions

	Coefficient
Sale Occasion	-.093
	(0.083)
Week FEs	Y
HH FEs	Y
Mean DV	3.994
Num HH	10,344
Num Obs	487,307

***p<.01, **p<.05, *p<.1

Standard errors clustered at the household level are included in parentheses.

A2 Retail Data Step Estimation Procedure

Provided a candidate draw of Θ, for each market m and week t, I need to solve for $\delta_{mt} = (\delta_{1mt}, \ldots, \delta_{Jmt})'$ such that

$$\begin{aligned} &s_{jmt}(\delta_{mt}; \Theta) = S_{jmt}, \\ &\text{for } j = 1, \ldots, J \text{ and } m = 1, \ldots, M, \end{aligned} \tag{A1}$$

where $s_{jmt}(\cdot)$ are the predicted retail market shares from Eq. (1.11) and S_{jmt} are the observed retail market shares. In solving this system of equations, I require two steps to be performed iteratively each period, starting from $t = 1$, as state dependency causes the current period purchase probabilities to rely on prior consumption status.

Thus, for a given period, I start by calculating the left-hand side of (A1). In practice, I rely upon Monte Carlo integration where Eq. (1.11) is approximated by

$$s_{jmt}(\delta_{mt};\Theta) = \frac{1}{R}\sum_{R}\sum_{g=0}^{G}\pi_{rjmt}(C_{rg,t-1}=1)P(C_{rg,t-1}=1). \tag{A2}$$

Each simulated household $r = 1, \ldots, R$ is represented by Halton draw from the empirical distributions of v and D, respectively. I draw $R = 200$ simulated households per market to compute Eq. (A2). Finally, $\pi_{rjmt}(\cdot)$ denotes the individual-level purchase probability conditioned upon prior consumption status $C_{rg,t-1}$ as well as x, p_{mt}, h_{mt}, δ_{mt}, Θ, D_r, and v_r.[A2]

Next, I invert the system of equations (A1) to obtain δ_{mt}. This system of equations is non-linear, and I solve it numerically. Grigolon and Verboven (2014) provides the contraction mapping algorithm, based on that described in Berry et al. (1995), for the random coefficients logit model with the inclusion of nesting parameters. In the case of a two-level nested model, the algorithm iteratively solves

$$\begin{aligned} &\delta_{mt}^{k+1} \equiv \delta_{mt}^{k} + (1-\boldsymbol{\lambda})[\ln(S_{mt}) - \ln(s_{mt}(\delta_{mt}^{k};\Theta))],\ k = 1, 2, \ldots, \\ &\text{where } S_{mt} = (S_{1mt}, \ldots, S_{Jmt})' \text{ and } s_{mt} = (s_{1mt}, \ldots, s_{Jmt})', \end{aligned} \tag{A3}$$

until the relative difference between δ_{mt}^{k+1} and δ_{mt}^{k} is less than my tolerance of $1\mathrm{e}^{-13}$. Note, $\boldsymbol{\lambda}$ represents a $1 \times J$ vector of nesting parameters where each element, $j = 1, \ldots, J$, is given by λ_g such that $j \in \mathcal{J}_g$.

After obtaining a unique δ_{mt}, in market m for a given period t, the evolving joint distribution of consumer heterogeneity and consumption status for the period $t+1$ is defined by Eq. (1.13). Once the inversion has been completed iteratively for each $t = 1, \ldots, T$, across all

[A2] In $t = 1$ prior consumption status is assumed to be $P(C_{rg1} = 1) = 1/(G+1)\ \forall r \in R$, and for subsequent weeks evolves according to Eq. (1.13). I treat the first quarter of my sample as a burn-in period, and derive my results only from data resulting from post burn.

markets, a unique $\delta(\Theta)$ has been obtained, and I proceed to the evaluation of our household-level log-likelihood.

A3 Comparison of Results With and Without Pricing Instrument

Table A2: Mean Utility Estimates With and Without Pricing Instrument.[a]

	Mean Utility	
	Price IV	**OLS**
Price	-0.759***	-0.321***
	(0.094)	(0.028)
Cigarette	1.303**	-1.511***
	(0.606)	(0.188)
E-cigarette	-4.771***	-6.701***
	(0.352)	(0.159)
Cessation	-1.749**	-5.687***
	(0.889)	(0.329)
Menthol	-0.718***	-0.789***
	(0.051)	(0.053)
Menthol × Ecig.	-0.348***	-0.272***
	(0.042)	(0.033)
Flavored	0.451***	0.098
	(0.078)	(0.064)
Category × Time FEs	Y	Y
Category × Market FEs	Y	Y
Num HH	15,223	15,223
Num HH Obs	2,317,585	2,317,585
Num Markets	100	100
Num Time Periods	226	226
Num Market Level Obs	135,600	135,600

***p<.01, **p<.05, *p<.1

[a] Standard errors are included in parentheses. My estimation includes fixed effects at the category/time and category/market level; for presentation purposes, and to avoid perfect collinearity, I exclude the flavor tobacco, the final time period, and the last market.

Table A2 presents a comparison of my results with, and without, my pricing instrument. As discussed in Section 1.6, to account for the possible correlation between the price variable and unobserved demand shocks, I use an instrumental variable technique. Specifically, I take the average product price over all DMAs not included in my estimation to be my pricing instrument.

The use of this instrument generates substantial changes in my estimation. Category dummies for cigarettes and cessation products now enter utility positively, and the parameter value for e-cigarettes rises by a sizable amount. Moreover, the mean price response, in terms of absolutes, increases significantly (more than doubles). These differences are those I would expect if (1) there exists simultaneity between price and demand, and (2) my instrument successfully corrects for this existence. Finally, with the inclusion of my instrument, all parameters remain statistically significant at the 95% level or greater.

A4 Supply-Side Model

In this appendix, I detail how I calculate counterfactual prices provided in my demand estimates found in Section 1.7. To begin, under the assumption that prices are set optimally, marginal cost is inferred from observed prices, market shares, and expected price sensitivity. Specifically, I assume that prices are set at the firm level, where each firm sets their product prices to maximize the total profits over the weeks in my finite sample. In this case, the FOCs are given by the vector $\frac{\partial \pi^f}{\partial p_{ft}}$ with the element corresponding to product j in the set F_j of products sold by firm f in time t (I drop the m subscript, assuming prices are set at the market level) being

$$0 = \frac{\partial \pi^f}{\partial p_{jt}} = \frac{\partial}{\partial p_{jt}} \sum_{k=1}^{T} \sum_{n \in F_j} S_{nk}(p_{nk} - mc_{nk}) = S_{jt} + \sum_{k=1}^{T} \sum_{n \in F_j} \frac{\partial S_{nk}}{\partial p_{jt}}(p_{nk} - mc_{nk})$$

which can be rewritten in vector form as

$$0 = S + \Delta(p - mc), \tag{A4}$$

for $S = (S_{11}, \dots, S_{J1}, \dots, S_{JT})'$, $p = (p_{11}, \dots, p_{J1}, \dots, p_{JT})'$, and $mc = (mc_{11}, \dots, mc_{J1}, \dots, mc_{JT})'$. Finally, Δ is a $(J \times T) \times (J \times T)$ matrix made up of $J \times J$ blocks, $\Delta_{k,t}$ for $k, t = 1, \dots, T$, such that

$$\Delta = \begin{bmatrix} \Delta_{1,1} & 0 & 0 & 0 & 0 \\ \vdots & \ddots & 0 & 0 & 0 \\ \Delta_{k,1} & \ddots & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \Delta_{T,1} & \dots & \Delta_{T,t} & \dots & \Delta_{T,T} \end{bmatrix}, \tag{A5}$$

with the (n, j) elements of $\Delta_{k,t}$ equal to $\frac{\partial S_{nk}}{\partial p_{jt}}$ if both n and j are owned by the same firm, and zero otherwise. Thus, the vector of marginal costs for all products, across all weeks, is

$$mc = \Delta^{-1} S + p. \tag{A6}$$

Once the vector of marginal costs has been obtained, I can predict the impact of changes such as the removal of flavorants or the impact of cigarette taxes. I assume that these changes do not impact my demand parameters or marginal costs. Thus, provided a gradient vector comprising the first order conditions of my firm's profit maximization equation, I find the vector of firm prices such that $\hat{p}_f$ maximizes firm prices. In application, I iterate between the firms, maximizing each firm's profits with respect to the other firm's choice of prices. I continue iterating until $\hat{p}_f$ converges for each firm.[A3]

A5 Illicit Cigarette Sales

A possible source of bias in my weighting procedure, when forming DMA-level weekly product usage rates, is the presence of illicit cigarette sales. Research by the Committee on

[A3]My tolerance for convergence is set to 1e-7.

the Illicit Tobacco Market, appointed by the National Research Council and tasked by the FDA, suggests that the sale of illegal cigarettes makes up 8.5% of the total cigarette market (National Research Council, 2015).[A4] At the DMA-level, if the sale of illegal cigarettes remains a constant proportion of total cigarette sales over the course of my sample period, then the population weight will account for the sale of illicit products when forming my market/time-level product usage rates. In this case, my observed retail sales can act as a proxy for illicit consumption. Supporting this notion, Paraje et al. (2022) suggests that the world-wide market for illicit cigarettes, as a percentage of total consumption, has largely stabilized over the past decade; with the consumption of illicit products trending similarly to that of legal sales. However, research by the National Research Council (2015) found the total proportion of illegal cigarette sales rose slightly over the latter half of their sample period—from 7.1 percent in 2003 to 8.5% by 2011.

Further, of greatest concern to the formation of my market shares is the impact of DMA-level price on the market for illicit cigarettes, as rising product price is considered a primary motivation for the trade in illegal cigarettes (National Research Council, 2015). In this case, legal and illegal sales may no longer trend similarly, and my observed sales can no longer serve as a proxy for illicit consumption.

To this regard, I find that brand specific pricing strategies remain largely consistent across all markets. Therefore, general increases in price may not encourage substitution to the illicit cigarette market, as the presence of profit maximizing smugglers implies that illicit cigarette prices increase alongside that of their legal counterpart. However, localized price changes (predominantly in the form of taxation) have a possibility of encouraging cross-border shopping and smuggling operations. If localized taxation increases the proportion of illicit cigarette sales in a market, then my market shares formation procedure may under-

[A4]Estimates of the size of the illicit cigarette market ranges from 8.5 to 21 percent. The low end, 8.5 percent, is the committee's own estimate and is found by comparing total tax paid sales with self reported consumption.

weight responsiveness to changes in price—stressing the importance of accounting for price endogenity.

Further, the sale of illicit products may also bias my counterfactual results—bans and taxation considered are common motives for illicit trade. However, to date, empirical research has not found an increase in illegal sales after the implementation of a menthol ban. In consideration of Massachusetts' 2020 menthol ban, Ali et al. (2022) found no significant impact on cross-border sales of neighboring states, where menthol products remain accessible to consumers and smugglers interested in menthol cigarettes. Similarly, an analysis of the 2015 Nova Scotia menthol ban found no significant increase in the seizure of illicit cigarettes pre- and post-ban; suggesting that the sales of illegal cigarettes is unlikely to be increasing in response to the removal of mentholated products (Stoklosa, 2019). Finally, Fong et al. (2022b) compared the purchases of Canadian smokers pre- and post-ban, in their respective provinces, and found no increase in the reported purchasing of illicit products.

Although sales of illicit products may not respond significantly to the removal of mentholated tobacco, what remains less clear is consumer responsiveness to my counterfactual taxation scheme. The National Research Council (2015) suggests much of the growth in the illicit tobacco market is a function of taxation—smugglers purchasing products in low tax states/territories and selling in high tax locations. However, my counterfactual taxation scheme is proposed at the federal level, subjecting all markets to an increase in price, and Paraje et al. (2022) hypothesizes that, on a global scale, common reductions in cigarette affordability have largely stymied growth of illicit trade and led to similar reductions both legal and illegal sales. Overall, due to the nature of illegal sales, the degree to which changes in observed cigarette sales can act as a proxy for illicit transactions remains largely unknown, and my results reflect an expectation formed by the assumption that my counterfactual scenarios do not significantly change illicit consumption behavior.

Appendix B

Chapter 2

B1 SSB Sales in Stores 8+ Miles from Philadelphia

To assess the exclusion of stores beyond the 8-mile band surrounding Philadelphia, we examine SSB sales in stores within the 3-digit ZIP Code prefixes (080, 181, 189, 190, 191, 192 and 194) pertaining to Philadelphia and its surrounding region. Specifically, we estimate the impact of the SSB tax on SSB sales within Philadelphia, 0-8 miles from Philadelphia, and 8-10 miles from Philadelphia, respectively. In addition to the 218 stores in the "city + 8 miles" region referenced in Section 2.3.1, we also observe the sales of 25 stores 8-10 miles from the city, and 93 stores 10+ miles from the city which we use as the control group for this analysis. SSB sales are aggregated at the store-week level for estimation purposes.

Results in Table B1 provide evidence that stores 8-10 miles from the city border are not affected by cross-border shopping, as their SSB sales do not demonstrate a positive response to the Philadelphia SSB tax. The variable "Post-Tax × (8-10 miles from Philadelphia)" has a negative coefficient, inconsistent with what would be expected if the taxation induced cross-border shopping in this region. Seiler et al. (2021, p. 35) report a similar finding.

Table B1: Regression of SSB Volume Sales[a]

Dependent Variable: SSB Weekly Volume Sales (in Ounces)	
Post-Tax × Philadelphia	-51626.97***
	(1629.77)
Post-Tax × (0-8 miles from Philadelphia)	10256.10***
	(1643.79)
Post-Tax × (8-10 miles from Philadelphia)	-15138.82***
	(2612.22)
Store FEs	Y
Week FEs	Y
Observations	103,503
Weeks	209

***p<.01, **p<.05, *p<.1
[a]Robust standard errors are reported in parentheses. SSB sales are aggregated at the store-week level.

B2 Multiple Purchases During a Single Trip

During some observed purchase opportunities, households buy multiple units of the same product or choose to purchase multiple different products. However, in our retail data, information pertaining to individual-level purchase variety and amounts is unavailable—we only observe aggregate store sales. To make our model tractable under a discrete choice framework, and to reconcile with the retail data, a couple of assumptions are required. In the case where we observe multiple distinct product purchases during a single trip, we treat them as arising from multiple purchase opportunities. Furthermore, when focusing on household purchases, we follow the example of Tuchman (2019) and consider purchase incidence—whether at least one unit was purchased—instead of purchase quantity.

Current literature involving the purchase of multiple units or multiple products under the BLP framework considers bundling units of the same, or different, goods together as a sort of composite product (e.g., Wang (2021)). This approach would be computationally infeasible in our case given the large number of beverage products. As such, our rationalizations described above (1) simplify our estimation, (2) make the model tractable under the BLP framework, and (3) are simply following those innately made by researchers working solely with retail data (i.e., Berry et al. (1995), Nevo (2000), etc.).

B3 Estimation Procedure During the Retail Data Step

Provided a candidate draw of Θ, for each month $t = 1, \ldots, T$ we need to solve for $\delta_t = (\delta_{1t}, \ldots, \delta_{J_t t})'$ such that

$$s_{jt}(\delta_t; \Theta) = S_{jt}, \text{ for } j = 1, \ldots, J_t, \tag{B1}$$

where $s_{jt}(\cdot)$ are the predicted retail market shares from Eq. (2.9) and S_{jt} are the observed retail market shares. In solving this system of equations, we require two steps.

We start by calculating the left-hand side of (B1). In practice, we rely upon Monte Carlo integration where Equation (2.9) is approximated by

$$s_{jt}(\delta_t; \Theta) = \frac{1}{R} \sum_{r=1}^{R} \pi_{rjt}(x_t, p_t, h_t, Q_{z_r}, D_r, \delta_t, \Theta, v_r). \tag{B2}$$

Each simulated household $r = 1, \ldots, R$ is represented by Halton draws of v_r, z_r, and D_r from the distributions of v, z, and $D|z$, respectively. We draw $R = 4000$ simulated households to compute Eq. (B2).

Next, to obtain δ_t, we must invert our system of equations (B1). For the RCNL model, this system of equations is non-linear and is solved numerically. Grigolon and Verboven (2014) provides the contraction mapping algorithm for the random coefficients logit model with nesting parameters. In the case of a one-level nested model, the algorithm iteratively solves

$$\begin{aligned} &\delta_t^{k+1} \equiv \delta_t^k + (1-\rho)[\ln(S_t) - \ln(s_t(\delta_t^k; \Theta))], \; k = 1, 2, \ldots, \\ &\text{where } S_t = (S_{1t}, \ldots, S_{J_t t})' \text{ and } s_t = (s_{1t}, \ldots, s_{J_t t})', \end{aligned} \tag{B3}$$

until the relative difference between δ_t^{k+1} and δ_t^k is less than our tolerance of .5e^{-12}. Once the inversion has been completed for each $t = 1, \ldots, T$, a unique $\delta(\Theta)$ has been obtained, and we proceed to the evaluation of our household-level log-likelihood.

B4 Category Fixed Effects

Table B2 provides estimates for the category fixed effects not reported in the RCNL Demand Estimates table found in Section 2.6. The first column of Table B2 provides the variable of interest, followed by the mean utility and low-income interaction, respectively. The category pure water was dropped to avoid perfect collinearity.

Table B2: RCNL Demand Estimates - Category Fixed Effects[a]

	Mean Utility	**Low-Income Interaction**
Carb. Soft Dr.	0.42**	-0.18
	(0.11)	(0.18)
Coffee	-1.82***	-0.40
	(0.18)	(0.31)
Energy Dr.	-2.45***	1.26***
	(0.29)	(0.35)
Flav. Water	-1.45***	-0.25
	(0.17)	(0.27)
Juice	0.85***	-0.58***
	(0.11)	(0.20)
Sports Dr.	-1.73***	-0.34
	(0.23)	(0.22)
Tea	-0.22*	-0.72***
	(0.12)	(0.20)

***p<.01, **p<.05, *p<.1

[a]Robust standard errors are reported in parentheses.

Table B3: Pass-Through Rate of SSB Tax, by Category

	Dependent Variable: SSB Per-Ounce Price (in Cents)						
	Carb. Soft Dr.	Coffee	Energy Dr.	Flav. Water	Juice	Sports Dr.	Tea
Post-Tax × Philadelphia	1.221***	1.201***	1.590***	1.239***	1.173***	1.250***	1.091***
	(0.004)	(0.048)	(0.029)	(0.056)	(0.012)	(0.011)	(0.011)
Product Characteristics	Y	Y	Y	Y	Y	Y	Y
Store FEs	Y	Y	Y	Y	Y	Y	Y
Store FEs × Diet/Med./Large	Y	Y	Y	Y	Y	Y	Y
Week FEs	Y	Y	Y	Y	Y	Y	Y
Observations	2,150,691	142,320	550,246	162,800	844,697	268,785	777,840
Stores	229	229	229	229	229	229	229
Weeks	209	209	209	209	209	209	209
Products	114	25	38	27	87	23	63

***p<.01, **p<.05, *p<.1

There are 111 stores in Philadelphia and 118 in the region more than 8 miles from Philadelphia. The 209 weeks span the 4-year period from 2015 to 2018. In total, 377 products across seven categories are subject to the SSB tax if sold in Philadelphia. Prices are aggregated to the product-store-week level for estimation purposes. For the regression in each category, we also include store fixed effects and their interactions with the diet, medium, and large dummy variables, week fixed effects, and additional product characteristics including sugar and caloric content. Standard errors are reported in parentheses and clustered at the product-store-week level.

B5 Category-Level Pass-Through Rates

To estimate category-level pass-through rates, for each of the seven categories containing SSBs, we regress price observed at the product-store-week level on the interaction Post-Tax × Philadelphia as well as store fixed effects and their interactions with the diet, medium, and large dummy variables, week fixed effects, and additional product characteristics including sugar and caloric content. Table B3 presents our results.

Table B4: Average Sugar and Caloric Consumption from Beverages per Household, by Location and Income Status[a]

	Without Tax	With Tax	Difference	% Change
Philadelphia Households				
High-Income				
Sugar (g)	17,953	11,144	-6,809	-37.93%
Calories (cal)	73,062	46,096	-26,966	-36.91%
Low-Income				
Sugar (g)	20,118	13,152	-6,966	-34.63%
Calories (cal)	80,579	53,515	-27,064	-33.59%
Non-Philadelphia Households				
High-Income				
Sugar (g)	19,826	19,655	-171	-0.86%
Calories (cal)	82,488	81,908	-580	-0.70%
Low-Income				
Sugar (g)	23,167	22,458	-709	-3.06%
Calories (cal)	93,513	90,832	-2,681	-2.87%

[a]Aggregate amount over the post-taxation period January 2017 to December 2018.

B6 Changes in Sugar and Caloric Consumption for High- and Low-Income Households

Table B4 reports changes in sugar and caloric consumption from beverages for high- and low-income households by home location. We find that among Philadelphia households, high-income households on average consume less sugar and fewer calories and experience a greater percentage reduction in their consumption. The pattern is different for non-Philadelphia households, where high-income households on average consume less sugar and fewer calories but experience a smaller percentage reduction in their consumption. Differences in the outcomes in response to the taxation are best explained by Table 2.12, where we observe that, in terms of the total volume of Philadelphia and non-Philadelphia SSBs consumed, low-income Philadelphia households are less responsive to the tax compared to their high-income counterparts, but the opposite is true for low-income non-Philadelphia households, who experience a larger volume reduction in SSB consumption—and therefore a larger reduction in sugar and caloric consumption—compared to their high-income counterparts.

B7 Additional Counterfactual Analyses

Here we report the results from three additional counterfactual analyses.

B7.1 No Taxation on Diet Products

We now analyze the counterfactual policy in which diet products are not subject to the SSB tax—as was originally proposed. The Philadelphia City Council has specified that their SSB taxation policy is first and foremost a revenue-generating scheme; generally, however, taxes imposed on sweetened beverages are designed to reduce consumption, as in the case of Berkeley, CA, Boulder, CO, and Seattle, WA, among others. Thus, except for Philadelphia, diet products are normally excluded from SSB taxation, being regarded as healthier alternatives to sugary products. We are therefore interested in how a policy under which diet products remain untaxed in Philadelphia would change households' consumption and welfare as well as the revenue-maximizing tax rate.

As before, we consider average sugar and caloric consumption from beverages per household during the 24 post-taxation months. We find that under the alternative policy, Philadelphia households would on average reduce their sugar intake by 40%, greater than the 36% reduction under the current policy. They would experience an average reduction of 29,358 calories—approximately 14.7 days' worth of caloric intake, a 9% increase compared to the 13.5 days under the current policy. These results show that from a public health perspective, leaving diet products untaxed is more beneficial by inducing a greater reduction in households' sugar and caloric consumption.

With respect to changes in consumer surplus, the alternative policy would leave households better off when compared to the current policy. Among Philadelphia households, we find an average consumer surplus loss of \$80.95 and \$80.09 for low- and high-income households, respectively, noticeably smaller than the \$104.75 and \$108.26 under the current policy (Table 2.13). Sugary beverages and their diet counterparts are good substitutes for some households, therefore when diet products are excluded from the tax, these households are able to switch from sugary beverages to their diet counterparts in order to avoid the tax, rather than having to travel for cross-border shopping or switch to less substitutable products, and thus households' average loss in consumer surplus is reduced.

In addition to lessening the loss in consumer surplus, the alternative policy also has an impact on the revenue-maximizing tax rate. The greater availability of untaxed substitutes in the form of diet products leads to households' higher price sensitivity with respect to sugary beverages, and we find that the revenue-maximizing tax rate falls from 3.14¢ per ounce under the current policy to 2.33¢ per ounce under the alternative policy. The tax revenue generated under the respective revenue-maximizing tax rate falls from \$32.5 million to \$24.6 million.

We note that our revenue-maximizing tax rate of 2.33¢ per ounce under the alternative policy is similar to the 2¢-per-ounce SSB tax rate in Boulder, CO, where the SSB tax includes only those products with added sugar, thus excluding diet products. Our revenue-maximizing tax rate falls slightly above the range of optimal tax rates found by Allcott et al. (2019), who study a national tax imposed on sugar-sweetened beverages only and determine an optimal tax rate between 1¢ and 2.1¢ per ounce. Different from our setting, their optimal tax rate is derived from a model interested in government redistribution of wealth.

B7.2 Both Locations Taxed

Next, we turn to our counterfactual analysis regarding the changes in SSB consumption and consumer surplus if the tax is levied upon both Philadelphia and its surrounding region. This counterfactual scenario can be interpreted as a national or multi-state SSB taxation policy (the region surrounding Philadelphia includes elements of both the state of Pennsylvania and the state of New Jersey), which removes Philadelphia households' ability to avoid taxation by cross-border shopping in the surrounding region. To create our counterfactual, for each beverage option sold in the non-Philadelphia location, we calculate the tax amount based on the tax rate, adjust the price accordingly based on the relevant pass-through rate, and set the variable "tax saving" to zero.

Results pertaining to the effects of this alternative tax coverage on households' beverage consumption are presented in Table B5. We observe that, given the widened tax coverage, non-Philadelphia households now experience a reduction in SSB consumption similar to those living in Philadelphia. Interestingly yet intuitively, Philadelphia households' consumption of Philadelphia SSBs becomes less responsive to the levying of an SSB tax: they reduce their consumption of Philadelphia SSBs by 3,553 ounces when both locations are taxed, compared to 3,791 ounces (Table 2.8) when the tax covers Philadelphia only. This result is primarily driven by Philadelphia households who have strong preference for SSBs: when the widened tax coverage removes their ability to avoid taxation through cross-border shopping, they instead continue to purchase SSBs in their home location, willing to pay the higher prices. Additionally, Philadelphia households' purchase of non-Philadelphia SSBs decreases by 496 ounces, compared to an increase of 474 ounces when the tax covers Philadelphia only.

The removal of households' ability to exploit the geographic nature of local taxation policies has a direct impact on households' loss of consumer surplus. Under the widened tax coverage, the loss in consumer surplus for low- and high-income Philadelphia households is \$128.35 and \$123.93, respectively, representing an increase of 23% and 14% when compared to the current tax coverage (Table 2.13). With cross-border shopping no longer a viable strategy for tax avoidance, a lower disutility from travel time no longer benefits low-income Philadelphia households as much, and their loss of consumer surplus is now greater than their high-income counterparts'. Finally, among non-Philadelphia households, we find a loss

Table B5: Average Beverage Consumption per Household: Both Locations Taxed[a]

SSB Status × Bev. Location	Without Tax	With Tax	Difference	% Change
All Households				
Philadelphia Bev. Options				
Non-SSB	2,158	2,486	+328	15.17%
SSB	4,027	2,012	-2,015	-50.04%
Non-Philadelphia Bev. Options				
Non-SSB	3,332	3,881	+549	16.49%
SSB	4,600	2,286	-2,314	-50.30%
Philadelphia Households				
Philadelphia Bev. Options				
Non-SSB	3,827	4,404	+577	15.09%
SSB	7,097	3,544	-3,553	-50.06%
Non-Philadelphia Bev. Options				
Non-SSB	770	887	+117	15.29%
SSB	1,009	513	-496	-49.20%
Non-Philadelphia Households				
Philadelphia Bev. Options				
Non-SSB	521	603	+82	15.70%
SSB	1,012	507	-505	-49.88%
Non-Philadelphia Bev. Options				
Non-SSB	5,846	6,819	+973	16.64%
SSB	8,124	4,027	-4,097	-50.44%

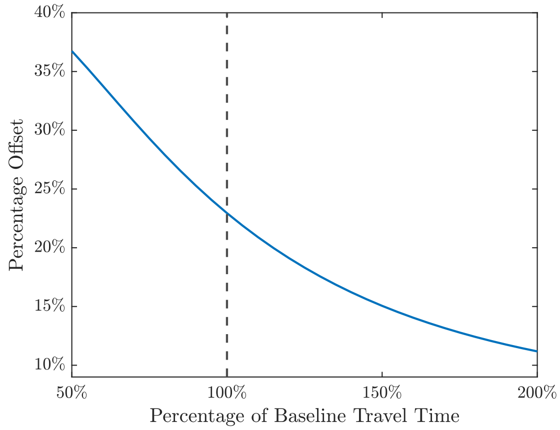
[a]In ounces; aggregate amount over the post-taxation period January 2017 to December 2018.

of consumer surplus in the amount of \$149.20 for low-income households and \$141.38 for high-income households when both locations are taxed.

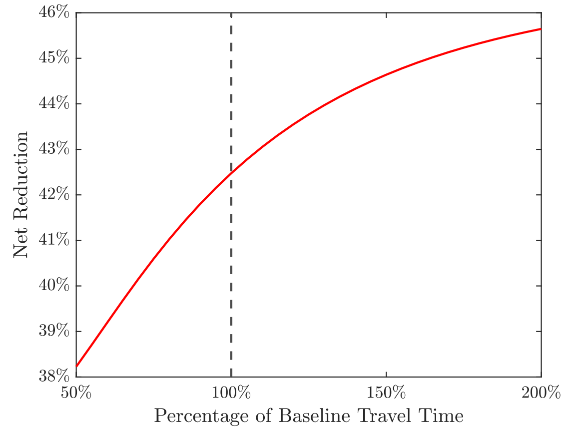
B7.3 Travel Time Changes

We now consider how changes in travel time affect both the willingness to cross-border shop and expected consumer surplus under the current taxation policy. This counterfactual provides an analysis of the effects of increased ease of transportation, for example due to improved roads, reduced traffic, better traffic control, etc. We calculate households' consumption of beverages while proportionally varying the travel time experienced by all households. Our findings are shown in Figure B1, with travel time being varied from 50% to 200% of the baseline. The figure presents the percentage of the decrease in Philadelphia SSB consumption that is offset by an increase in non-Philadelphia SSB consumption, the net

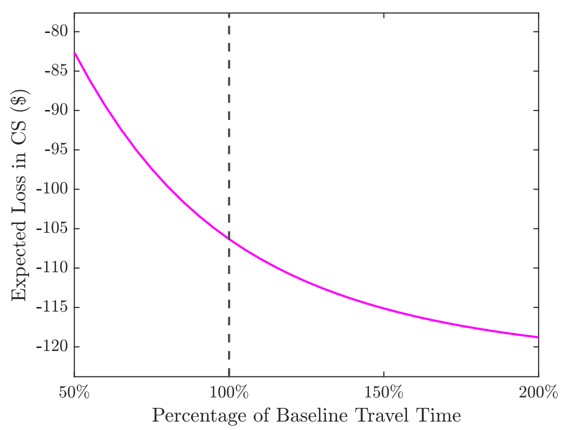
Figure B1: SSB Consumption and Consumer Surplus: Changes in Travel Time



(a) Percentage of Phil. SSB Reduction Offset



(b) Net Reduction in SSB Consumption



(c) Phil. Household Expected Loss in CS

Notes: The percentage of Philadelphia SSB reduction offset (Panel a) measures the percentage of the decrease in Philadelphia SSB consumption that is offset by an increase in non-Philadelphia SSB consumption. The net reduction in SSB consumption (Panel b) measures the net reduction—after accounting for the offset—as a percentage of Philadelphia SSB consumption in the no-tax scenario.

reduction in SSB consumption—after accounting for the offset—as a percentage of Philadelphia SSB consumption in the no-tax scenario, and the expected loss of consumer surplus for Philadelphia households.

Travel time plays a large role in determining the degree to which households cross-border shop. When travel time is halved, we find that 37% of the reduction in Philadelphia SSB consumption due to taxation is offset by an increase in non-Philadelphia SSB consumption. However, when travel time is doubled, only 11% of the reduction is offset. As such, travel time ties directly into the net effect of SSB taxation on the consumption of taxed products. When travel time is halved, the net reduction in SSB consumption equals 38% of Philadelphia SSB consumption in the no-tax scenario. In comparison, when travel time is doubled, we find a net reduction of 46%. At this point, few Philadelphia households engage in cross-border shopping; instead, much of the rise in non-Philadelphia SSB consumption is driven by non-Philadelphia households, for whom purchasing in the non-Philadelphia location does not involve travel costs.

Our findings provide supporting evidence towards the effectiveness, or lack thereof, of SSB taxation policies in regions of differing sizes. For instance, Cawley and Frisvold (2015) suggest that one possible reason the SSB tax pass-through rate found in Berkeley is so low compared to other localities is consumers' ability to evade city-level taxes through cross-border shopping. Berkeley's land area is only 10.4 square miles (compared to Philadelphia's 134 square miles), and the authors note that the average US consumer travels 5.2 miles when shopping for groceries. As such, we would expect Berkeley residents to act similarly to Philadelphia households residing minutes from the city border. Comparatively, residents of large cities may experience longer travel time when seeking to cross-border shop and therefore, as our findings suggest, their rate of tax avoidance by cross-border shopping may be significantly smaller.

Finally, travel time and the ease of cross-border shopping have a direct impact on the loss in consumer surplus resulting from SSB taxation. We focus on Philadelphia households' expected change in consumer surplus, as they reside in the taxed region and experience the greatest change in utility resulting from a change in the ease of travel. As expected, a lower travel time directly implies a smaller loss of consumer surplus associated with SSB taxation, as increased ease of travel allows for greater tax avoidance behavior. When travel time is increased from 50% to 200% of the baseline, an average Philadelphia household's expected loss in consumer surplus increases by 43.8% from \$82.6 to \$118.8, compared to \$106.3 at the baseline. While an increase in the ease of travel is beneficial for consumers, from the perspective of the Philadelphia government, providing for methods by which households can more easily access the untaxed region is contrary to the stated revenue-maximizing intentions of its SSB taxation.