Title
E-shopping and Household Travel Before, During, and After the Time of Covid-19

Author
Xu, Lu

Publication Date
2022

UNIVERSITY OF CALIFORNIA,
IRVINE

E-shopping and Household Travel Before, During, and After the Time of Covid-19

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Civil and Environmental Engineering

by

Lu Xu

Dissertation Committee:
Professor Jean-Daniel Saphores, Chair
Professor Wilfred Recker
Professor Michael Hyland
Professor Luyi Gui

2022

DEDICATION

To

my husband, family, and friends

for their endless love, trust, and support.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

Five years ago, I was luckily admitted to the Ph.D. program of Transportation Systems Engineering (TSE) of the Department of Civil and Environmental Engineering at the University of California, Irvine (UCI). It was also the time when I ended my cross-border relationship with my husband, Dr. Jin Yang, who came to pursue his Ph.D. three years earlier than me at Caltech. Along the journey to pursue my Ph.D., I always feel grateful and cherish every moment in this great university, which brought my life with many wonderful people and unforgettable memories.

First, please let me give my deepest gratitude to my great advisor, Prof. Jean-Daniel Saphores. Prof. Saphores is a dedicated, responsible, patient, inspiring, and respectful professor. He is an excellent mentor guiding me not only in research but also in life. I have learned a lot from him in each of our individual and group meetings. I feel a deep sense of pride and luck to work with him in the past five years. Without his continuing guidance, support, help, and encouragement, this research might never have been completed. No matter what I plan to work on in my career, I will benefit tremendously from the study and research experience with him.

I would also like to thank all my committee members, Prof. Wilfred Recker, Prof. Michael Hyland, Prof. Luyi Gui, and Prof. Stephen G. Ritchie for their valuable suggestions and support of my research. The TSE program at the Institute of Transportation Studies (ITS) allows me to learn comprehensive transportation courses, which build a solid foundation for me in the transportation field. I would like to thank Prof. Wilfred Recker, Prof. R. Jayakrishnan, Prof. Michael G. McNally, Prof. Wenlong Jin, Prof. Jean-Daniel Saphores, Prof. Stephen G. Ritchie, and Prof. Michael Hyland for their great lectures, talks, and organized seminars.

I also appreciate the efficient assistance of administrative staff from ITS and the Department of Civil and Environmental Engineering: Ms. Cam Tran, Dr. Zhe Sun, Ms. April Marie Heath, and Ms. Angie Zuniga.

I would like to thank current and previous members of Prof. Saphores's group, Monica Ramirez-lbarra, Bumsub Park, Farzana Khatun, Md Rabiul Islam, Dr. Suman Mitra, Dr. Yunwen Feng, Dr. Ke Wang, and Dr. Sheng-Hsiang Peng for their kind help and support. I would also like to acknowledge Dr. Qinglong Yan, Dr. Zhe Sun, Dr. Karina Hermawan, Dr. Rezwana Rafiq, Dr. Yue Sun, Biling Liu, Dr. Qi-Jian Gan, Dr. Felipe Augusto de Souza, Chenying Qin, Dr. Riju Lavanya, Dr. Jiangbo Yu, Dr. Xiaoxia Shi, Daniel Chuong, Dr. Jun Hyeong Park, Dr. Daisik Nam, and Dr. Roger Lloret-Batlle, for their help and guidance when I started my journey at ITS.

Many thanks to my office-mates: Ximeng Fan, Arash Ghaffar, Deep Shah, Pratiik Malik, Esmaeil (Sina) Dabbagh, Negin Shariat, Guoliang Feng, Marjan Mosslemi, and Khalid Abdulghani for every joyful chat. I would also like to say thank you to my course-mates: Jiaqi Cheng, Steven Ng, Jennifer Kwong, Monica Ramirez-lbarra, Ying Li, Lai Wei, Bai Xue, Eduardo Marino, Kasidit Jirotkun, David Sedath, Kat Lee, Brian Casebolt, Osi, and Rita for all the time studying together. I would also like to say thank you to: Dr. Yiqiao Li, Dr. Yijun Xiang, Dr. Dingtong Yang, Dr. Xuting Wang, Dr. Youngeun Bae, Irene Martinez Josemaria, Sunghi An, Kae Que, Pengyuan Sun, Siwei Hu, Boyuan Jiang, Navjyoth Sarma JS, Koti Reddy Allu, Jerry Tao, Naila Sharmeen, Tanjeeb Ahmed, Dr. De'Von Jennings, Mingqi Yao, etc. I thank all my friends and colleagues, and you are the treasure of my life.

I am lucky to be one of the co-founders of the Civil and Environmental Engineering Graduate Association and a member of the Institute of Transportation Studies Graduate Association. I value all the events we hosted together. I would also like to thank the Beijing

Jiaotong University Alumni Association of Southern California, where I met many excellent alumni. I cherish every event we organized together. In addition, I would like to acknowledge BURN OC, an inspiring running club in Orange County, California, where I trained with these excellent runners and have completed my first half-marathon.

Last but not least, I would also like to thank my beloved families in China especially my parents and brothers for all their support throughout my entire life. Most importantly, I would like to thank my husband Dr. Jin Yang, who is always there for me and supports me no matter what happens. He is my love, friend, and mentor in life. His optimism, talent, and determination to pursue his academic career always inspire me.

Comments:

1) The text of Chapter 2 of this dissertation is a reprint of the material as it appears in Research in Transportation Economics. The co-author, my advisor Prof. Jean-Daniel Saphores, listed in this publication directed and supervised research which forms the basis for the dissertation.

2) The text of Chapter 3 of this dissertation is under review in Transportation. The co-author, my advisor Prof. Jean-Daniel Saphores, listed in this publication directed and supervised research which forms the basis for the dissertation.

CURRICULUM VITA

Lu Xu

EDUCATION

2016-2022	University of California, Irvine (Irvine, CA, USA)
	Ph.D. in Transportation Systems Engineering M.S. in Transportation Systems Engineering
	Beijing Jiaotong University (Beijing, China)
2013-2016	M.S. in Traffic and Transportation Planning and Management
2009-2013	B.S. in Optical Information Science and Technology
	Minor: Traffic and Transportation

FIELD OF STUDY

Travel Behavior Modelling, E-commerce, Urban Logistics, Optimization, Statistical Models, Data Analysis

PUBLICATIONS

[1] **Xu, L.,** & Saphores, J. D. (under review). I E-SHOP, AND THEREFORE I TRAVEL MORE? Evidence from the 2017 U.S. National Household Travel Survey. *Transportation*.

[2] Saphores, J. D., & **Xu, L**. (2021). E-shopping changes and the state of E-grocery shopping in the US-Evidence from national travel and time use surveys. *Research in Transportation Economics*, 87, 100864.

[3] **Xu, L**., Zhang, X.C., Wang, L.L., & Xu, X. (2016). Explore the relationship between customer experience value and high-speed railways' service quality (Chinese version). *Chinese Railways, 03*.

[4] **Xu, L.**, Zhang, X.C., Chen, J.H., & Yu, J. (2014). Transfer Efficiency Evaluation Method of Urban Rail Transit Based on the Generalized Cost Function. *Applied Mechanics and Materials*, 505-506, 542.

CONFERENCES & PRESENTATIONS

[1] **Xu, L**., Saphores, J. D. E-shopping changes and the state of E-grocery shopping in the US-Evidence from national travel and time use surveys. E-shopping workshop among Chinese Scholars, 2021.

[2] **Xu, L**., Saphores, J. D. I E-SHOP, AND THEREFORE I TRAVEL MORE? Evidence from the 2017 U.S. National Household Travel Survey. 100th Transportation Research Board Conference, 2021.

[3] **Xu, L.,** Saphores, J. D. DOES ONLINE SHOPPING REDUCE TRAVEL? Evidence from the 2017 National Household Travel Survey. Best Poster Award, 2019 CHUMmit.

[4] **Xu, L.,** Saphores, J. D. DETERMINANTS OF RESIDENTIAL DELIVERIES FROM INTERNET SHOPPING – An analysis of the 2009 & 2017 National Household Travel Surveys (NHTS). 2019 California Transportation Planning Conference.

SKILLS

Software: TransCAD, ArcGIS, TransModeler, Latex, and Microsoft

Programming: Python, Gurobi, Stata, R, SQL, Octave, and Matlab

LEADERSHIP

2018 Cofounded Civil and Environmental Engineering Graduate Association at UCI

2018-2019 President of Institute of Transportation Studies Graduation Student Association at UCI

2020-2021 Vice president of Beijing Jiaotong University Alumni Association of Southern California

ABSTRACT OF THE DISSERTATION

E-shopping and Household Travel Before, During, and After the Time of Covid-19

by

Lu Xu

Doctor of Philosophy in Civil and Environmental Engineering

University of California, Irvine, 2022

Professor Jean-Daniel Saphores, Chair

During the past two to three decades, and especially during the Covid-19 pandemic, e-shopping has become increasingly popular, changing the way people shop and travel. With increasing concerns about the environmental impacts of transportation, particularly on regional air quality and on emissions of greenhouse gases (GHG), it is important to understand how e-shopping has affected household travel behavior.

In this dissertation, I investigated the influence of e-shopping before, during, and after the pandemic by analyzing data from the 2009 and the 2017 U.S. National Household Travel Surveys (NHTS), from the 2017 American Time Use Survey (ATUS), and from an IPSOS survey of Californians conducted in late May 2021. Understanding changes in shopping is essential to business owners, logistics managers (for adapting supply chains), transportation planners (for mitigating the impacts of warehousing and of additional residential freight deliveries), and policymakers (for helping at-risk and underserved groups).

This dissertation has three parts. In the first part, I estimated zero-inflated negative binomial models to analyze factors that affected residential deliveries before the pandemic based on the 2009 and 2017 NHTS. I found that e-shoppers in the U.S. were more varied in 2017 than in 2009. Households with more females, higher incomes, and more education, received more

deliveries. I also analyzed the 2017 ATUS to explore factors that influence grocery shopping. I found that in-store grocery shoppers were more likely to be female and unemployed but less likely to be younger, to have less than a college education, and to be African American. In contrast, online grocery shoppers were more likely to be female.

In the second part, I studied the impact of e-shopping on household travel using propensity score matching. My analysis of 2017 NHTS data showed that before the pandemic, greater online shopping was associated with more frequent trips and slightly more travel. Furthermore, the extent to which an increase in the number of activities translated into more travel depends on population density, the day of the week, the frequency of online shopping, and the type of activity.

In the third part, I analyzed the impact of the Covid-19 pandemic on grocery shopping frequency in-store, and online with home delivery (e-grocery) or pickup (click-and-pick), to understand how they changed due to the pandemic, and how they may change after, using ordered models and structural equation models. My results showed that Californians kept shopping for groceries in brick-and-mortar stores during the pandemic but less frequently than before. The pandemic accelerated the adoption of e-grocery and click-and-pick with some strong generation effects: younger generations were more likely to experiment with e-grocery and click-and-pick, while older generations relied more on in-store shopping. Education also made a difference, but thankfully race did not impact the use of e-grocery and click-and-pick, and intentions to use e-grocery and click-and-pick (but it did affect in-store grocery shopping before). My results also illustrated the heterogeneity of Hispanics. As expected, tech-savvy households were much more likely to embrace e-grocery and click-and-pick.

Keywords: online shopping, in-store shopping, grocery, travel, Covid-19 pandemic

Chapter 1. Introduction

The popularity of online shopping has been rapidly growing, with e-commerce increasing its share of total retail sales in the U.S. to 10.7% ($576.53 billion) in 2019 compared to only 4.4% ($169.14 billion) in 2010 (United States Census Bureau, 2021). This percentage reached 13.6% ($759.47 billion) in 2020 due to the Covid-19 pandemic. These changes are affecting freight and supply chains management, household travel, land use, and more generally, the environment. Understanding the factors that influence online shopping and how changes in online shopping impact travel is essential to transportation planners and engineers, logistics practitioners, and policymakers. Most recently, the Covid-19 pandemic has accelerated the adoption of online shopping and dramatically impacted travel, leading transportation engineers and planners, and policymakers to inquire about travel and online shopping after the pandemic is over.

In this dissertation, I studied the determinants of online shopping and how it impacted travel before and during the Covid-19 pandemic, and what the post-pandemic situation may look like. I analyzed data from the 2009 and the 2017 National Household Travel Surveys (NHTS), from the 2017 American Time Use Survey (ATUS), and from a survey of Californians conducted in late May 2021 by IPSOS to assess the impact of the pandemic on commuting and food shopping. This dissertation is organized as follows.

In Chapter 2, I explored how deliveries from online shopping changed over time before the pandemic. I estimated zero-inflated negative binomial models using data from the 2009 and 2017 NHTS. In addition, I characterized the profile of grocery shoppers in the U.S. using data from the 2017 ATUS using logit models. A better understanding of the determinants of online shopping alongside e-grocery is beneficial to logistics managers (to better supply warehouses serving residences), to transportation engineers (to maintain residential roads and adequately update design), and to transportation planners (to mitigate the externalities of changing freight flows and accommodate efficient deliveries).

In Chapter 3, I analyzed the impact of online shopping on household travel before the pandemic. This question is particularly relevant for policymakers concerned with urban congestion, air pollution from transportation, and greenhouse gas (GHG) emissions. Unlike other papers in the transportation literature, my unit of analysis is the household since travel and shopping decisions within households are usually interrelated. To tackle the bias from households self-selecting into various levels of e-shopping, I relied on propensity score matching, which enabled me to make a causality link between more online shopping and changes in travel. I classified households who participated in the 2017 NHTS into three groups based on how frequently (per person per month) they shop online: low (up to once), medium (more than once but less than 4 times), and high (over 4 times). I then analyzed the impact of different frequencies of online shopping on different trip purposes either during weekends or during weekdays, for different population densities.

In Chapter 4, I studied the impact of the pandemic on the frequency of grocery shopping in California before and during the pandemic, and what these frequencies could look like after the pandemic. I analyzed data from an online survey of Californians conducted at the end of May 2021 by IPSOS (a top-tier market research firm) using factor analysis, ordered models, and structural equations. My results contribute to the growing interest in grocery shopping during the Covid-19 pandemic since social-distancing restrictions necessitated by the pandemic led to a surge in e-grocery (+63.9% year over year in 2020) (Droesch, 2021). Understanding changes in grocery shopping is important to urban planners (to mitigate increased freight deliveries in residential neighborhoods), logistics managers (to better organize deliveries), and environmental planners concerned about vehicle miles traveled and GHG emissions from transportation.

Finally, in Chapter 5, I summarized the main findings of this dissertation, outlined some limitations, and made some suggestions for future work.

References

Droesch, B. (2021). US Digital Grocery Forecast 2021 - Defining the Key Players and Trends in a Rapidly Evolving Market. Available from: https://www.emarketer.com/content/us-digital-grocery-forecast-2021

United States Census Bureau (2021). Estimated Quarterly U.S. Retail Sales (Adjusted): Total and E-commerce 1999-2021. Available from https://www.census.gov/retail/index.html#ecommerce

Chapter 2. E-shopping Changes and the State of E-grocery Shopping in the U.S. – Evidence from National Travel and Time Use Surveys

2.1 Introduction

By expanding the range of products available to consumers, stimulating competition, and enhancing shopping convenience, e-commerce is changing the way people shop. Its popularity is growing. According to the Pew Research Center (Smith and Anderson, 2016), four out of five Americans have purchased items online at least once (up from 22% in 2000). Globally, e-commerce is taking an increasing share of total retail sales, rising from 7.4% in 2015 to 11.9% in 2018 (eMarketer, n.d.). These changes have widespread implications for freight and supply chains management (Perboli and Rosano, 2019), travel (Calderwood and Freathy, 2014; Suel and Polak, 2018), the environment (Cherrett *et al*., 2017; Dost and Maier, 2018), in-store shopping (Farag *et al*., 2007; Lee *et al*., 2017), and land use planning (Pettersson *et al*., 2018).

The growth of online shopping (e-shopping) is far from homogeneous from a geographic point of view, however. For example, in 2018 online retail sales were approximately 28.6% of total consumer retail sales in China (insideRetail Hong Kong, n.d.), versus less than 10% in the United States (US Census Bureau News, 2019). Moreover, e-shopping in a given sector can differ widely even between countries that are culturally and economically similar. Indeed, despite an average annual growth of 18.7% between 2000 and 2016 (US Census Bureau, 2018a), e-commerce sales of food, beer, and wine[1] in the United States represent currently only 0.35% of

[1] There is no separate category for groceries in the Census data on e-commerce sales. Moreover, data from the Census in Historical Tables 4 and 5 do not distinguish between meals delivered to a home/office and foods bought at a grocery store.

total food and beverage purchases (US Census Bureau, 2018b). By comparison, online sales made up 5.3% of total food retail sales in the UK (Office of National Statistics, 2018), which outlines the need to study e-grocery in the United States, even though this topic has already received much attention elsewhere, especially in Europe.

This chapter has two purposes. The first purpose is to understand changes between 2009 and 2017 – two years selected because of data availability - in residential deliveries from online shopping in the United States. I focus on residential deliveries because national data on grocery deliveries from online purchases in the United States are not publicly available. Understanding residential deliveries from e-shopping is important to logistics managers (so they can supply warehouses serving residences), to transportation engineers (so they can maintain residential roads and adequately update their design), and to transportation planners (so they can mitigate the externalities of changing freight flows and accommodate new delivery options). Better quantifying the traffic and environmental impacts of residential deliveries from online shopping is particularly of concern because soaring uncoordinated deliveries will increase residential traffic congestion, noise, and air pollution, and exacerbate parking shortages in dense urban areas.

The second purpose of this chapter is to characterize U.S. households who are shopping online for groceries, which is salient because of the importance and the challenges of grocery retailing. Yet there is no publicly available national dataset on household deliveries of groceries in the U.S., an indirect way of analyzing e-grocery deliveries is to couple characterizations of e-grocers with an understanding of deliveries from online purchases, as analyzed in the first part of this chapter. Although a number of papers have analyzed online shoppers (e.g., see Brashear *et al*., 2009; Ganesh *et al*., 2010; Crocco *et al*., 2013; Bressolles *et al*., 2014; or Harris *et al*., 2017, and references herein), much of the recent literature has focused on Europe, and there is a dearth of academic research on e-grocery in the United States. Profiles of online shoppers generated by consultants for grocers can also be found online, but they typically rely on univariate analyses and none of these profiles has analyzed datasets representative of the U.S. population.

This study is the first to analyze recent changes in residential deliveries from online shopping in the United States and to examine online grocery (e-grocery) shopping using publicly available survey data.

The point of departure is the latest (2017) National Household Travel Survey (NHTS), which I contrast with its previous edition, the 2009 NHTS. These two national surveys are analyzed because they asked participating households how many deliveries from online shopping they received in the 30 days preceding their assigned survey day. To explain the number of these deliveries and understand how they changed between the 2009 and the 2017 NHTS, I estimate similar zero-inflated mixture models on 2009 (N=134,371)[2] and 2017 (N=123,148) NHTS data, and test differences in model coefficients for these two years.

Since the 2017 NHTS does not include e-grocery shopping data, the 2017 American Time Use Survey (ATUS) dataset is also analyzed to contrast socio-economic characteristics of people who engage in online grocery shopping with conventional grocery shoppers. To do so, I first estimate logit models on a subset (N=2,934) of the 2017 ATUS to consider only households likely to have had access to e-grocery shopping in 2017. Since the number of people who shopped for groceries online in the 2017 ATUS is small, I compare the distributions of selected socio-economic characteristics of e-grocery shoppers with those of conventional grocery shoppers using Kruskal-Wallis tests. Finally, I contrast the distributions of shopping start times

[2] In this paper, (N=number) refers to the size of the sample on which specific models were estimated.

(i.e., when a customer starts browsing to buy groceries) between these two groups, since proponents of e-grocery shopping highlight the convenience of shopping at any time.

This chapter is organized as follows. In the next section, I will introduce a literature review about how e-grocery shopping first emerged, assess what obstacles led to early failures, and explain how they were (at least partly) overcome. Then I briefly discuss characteristics of e-shoppers in the United States, and review some potential impacts of e-shopping and e-grocery. In Section 2.3 and 2.4, I present my data and introduce the modeling approach. In Section 2.5 results are further discussed. In Section 2.6, I summarize all my key findings, discuss potential impacts on local deliveries, outline some limitations of my study, and propose some avenues for future work.

2.2 Background and Literature Review

2.2.1 The first coming of e-grocery in the U.S.

Alternative channels to traditional retail (e.g., mail order) predate by decades the arrival of online shopping but their market share has always been small, especially for groceries (White, 1997). The emergence of the internet was a game changer because it considerably expanded consumer choice, sped up deliveries, and provided new ways to learn about products (Seaman, 1995).

The first big push to develop online grocery shopping took place in the late 1990s when consumers started buying products over the internet. Grocery was an early target because it is the single largest retail sector, most consumers shop for groceries frequently, and many do not particularly enjoy it (Saunders, 2018). Proponents argued that e-grocery would give consumers the freedom to shop at the time and from the place of their choosing, expand consumer choice, and stimulate competition by facilitating price comparisons. Convenience and time saving were seen as major advantages at a time when women's participation in the labor force was rising (Morganosky and Cude, 2000).

E-grocery pioneers were technology companies eager to leverage their knowledge of information technologies to take over what they saw as an underachieving sector (Saunders, 2018). Companies like Webvan (founded in 1996) and HomeGrocer.com (started in 1997) rode the dot.com bubble. They attracted large investments to buy warehouses, delivery vans, and marketing campaigns, and built from scratch purely online businesses. However, when consumer demand failed to meet expectations, investments dried up with the burst of the dot.com bubble and they went bankrupt (Webvan purchased HomeGrocer.com in September 2000, and it filed for bankruptcy in July 2001) (Grunert and Ramus, 2005; Saunders, 2018). Partly as a result of these failures, online grocery shopping has been called "the Bermuda Triangle of e-commerce" - a place where investments vanish without leaving a trace (McDonald *et al*., 2014).

What went wrong? U.S. e-grocery pioneers overlooked several key characteristics of the grocery sector and underestimated the magnitude of the change they wanted to introduce.

First, e-shopping implies that a number of tasks previously undertaken by customers, including picking, packing, and delivering goods, are taken over by the retailer. This adds to the costs of retailers and squeezes their already thin profits.

Second, since groceries are quite diverse and some are perishable, they require more complex logistics (Murphy, 2003). Moreover, delivering to a customer's residence raises new issues. Indeed, if no one is present to receive an order, coming back at a different time is costly. Conversely, if an order is left on a buyer's doorstep, delivered goods could spoil or be stolen. The problem is especially acute for prepared foods, whose temperature, texture, taste, and appearance can quickly change over time.

Third, early e-grocers did not appreciate the difference between buying groceries online and shopping for groceries in a conventional store (Robinson *et al*., 2007). In particular, sensory information (e.g. smell and touch) and interpersonal interactions are lacking online (Hansen, 2005). This is not an issue for search goods (whose characteristics are easily evaluated before purchase), but it matters for experience goods (which can only be evaluated after a purchase) (Nelson, 1970). In brick-and-mortar stores, fresh produce can be touched and smelled so they are search goods, but for online shoppers they become experience goods (Weathers *et al*., 2007).

In spite of its failure, the first wave of e-groceries caught the attention of traditional grocers as online shopping revolutionized shopping in other sectors, such as books or electronics.

2.2.2 The second coming of e-grocery in the U.S.

As the first wave of internet-only grocers was closing, traditional grocers started experimenting with online shopping. Moreover, new start-ups began emerging with innovative solutions to address some of the shortcomings of earlier e-grocery shopping models.

In addition to attractive and easily navigable websites (Freeman and Freeman, 2011), one key to success in e-grocery shopping is low operational costs in order to offer competitive prices and effective delivery services (Kamarainen *et al*., 2001; Anckar *et al*., 2002). This lesson was learned by both traditional grocers and mega retailers such as Walmart or Target, and Amazon, which acquired Whole Foods in 2017 to boost its physical presence.

To keep costs down, some grocers partnered with start-ups that provide a platform to customers who order from the grocers' websites and have their employees pick, pack, and deliver orders in exchange for payments from both grocers and shoppers (they also make money from customer information). The largest of these start-ups is Instacart, created in 2012 (Lien, 2017). By the end of 2018, Instacart had partnerships with over 300 retailers operating over 15,000 grocery stores. A number of other start-ups have been offering similar services, including Deliv, DoorDash, Postmates, or Shipt (acquired by Target at the end of 2017).

To avoid unsecured deliveries, e-grocers have experimented with different alternatives: 1) Click and pick, where consumers order online but pick up at a store or a warehouse; 2) Bring to a local storage area and deliver when customers are home; 3) Allow the delivery person to leave purchases inside a customer's home. This service is offered by smart lock maker August with delivery partner Deliv and several retailers (Macy's, Best Buy, Bloomingdale's, and PetSmart). It is also offered by Amazon for its Prime customers who live in selected cities, and subscribe to Amazon Key. Amazon Key requires buying an Amazon cloud cam and installing a compatible smart lock at home (Wollerton, 2018); and 4) Deliver an order to the customer's car trunk, if he/she is an Amazon Prime subscriber, has an active connected car service plan, drives a GM or a Volvo vehicle from 2015 or newer, and lives in one of 37 U.S. cities (Hawkins, 2018).

Leading e-grocery retailers in 2017 include Wal-Mart, Costco, Sears (which filed for bankruptcy in 2018), Amazon, Kmart, but also Kroger (the largest overall grocer in the US), and Safeway (an Albertsons brand). However, this sector has been changing quickly with Amazon's purchase of Whole Foods and Target's acquisition of Shipt (both in 2017), for example. While in 2017 just over 30% of grocery stores in the U.S. offered home delivery/store pickup of online orders, this percentage had jumped to over 52% by 2019 (Conway, 2020).

2.2.3 E-shoppers and e-groceries

To inform my choices of explanatory variables, I also reviewed papers characterizing people who engage in e-grocery shopping. Early studies reported that online grocery shoppers are typically younger, better educated, and tend to have higher incomes than the general population (Morganosky and Cude, 2000, 2002). They are also more likely to be female because women are typically more involved in grocery shopping (Morganosky and Cude, 2000, 2002). Other studies reported that some seniors and some disabled individuals also shop online for groceries (White, 1997; Anckar *et al*., 2002).

As expected, convenience is a driving force behind e-groceries, but situational factors (such as a recent baby or a deteriorating health) also matter (Hand *et al*., 2009). People comfortable navigating the internet are not necessarily online shoppers, however, and when they shop online, they do not usually discontinue offline shopping (Hand *et al*., 2009). Furthermore, Kang *et al*. (2016) showed that the impact of convenience depends on experience with e-shopping and with the type of product considered. Moreover, although the time requirement to access offline grocery markets has no effect on the adoption of online grocery shopping, it may affect the amount of groceries purchased online.

A number of papers have inquired about the determinants of consumers' channel choice (e.g., see Melis *et al*., 2015; or Wang *et al*., 2015). Melis *et al*. (2015) found that when consumers start buying groceries online, they tend to select the online store from their preferred offline stores; moreover, the offline environment is important when customers are new to online shopping, although it matters increasingly less as they gain more experience with online shopping. The device used for e-grocery shopping also seems to matter: according to Wang *et al*. (2015), m-shopping (i.e., shopping via smartphones or tablets) increases the rate of orders, especially for low-spending customers.

In spite of high growth rates and enthusiasm for online grocery shopping, a recent Gallup survey (Redman, 2018) showed that 84% of U.S. adults have never ordered groceries online, and that 7 out of 10 of those who buy groceries online do so twice a month or less.

2.2.4 Impacts of e-shopping/e-groceries on land use, retailers, and supply chain planning

As e-shopping and e-grocery become more common in the U.S., they may have multiple impacts. Here, the impacts on land use, retailers, and supply chain planning are briefly considered.

As they become increasingly affordable and ubiquitous, information and communication technologies (ICT) are decoupling activities such as work or shopping from specific times and spaces (Kwan, 2007). By decreasing the cost of exchanging information, ICT may weaken agglomeration forces and promote the emergence of decentralized, smaller urban centers (Ioannides *et al*., 2008). However, concrete evidence that internet use (and in particular e-shopping) has impacted urban structure is still lacking (Ioannides *et al*., 2008), possibly because it takes years to substantially change the structure of an urban area, but also because of the complexity of the changes induced by ICT, and more particularly by e-shopping (Nahiduzzaman *et al*., 2019). While a shift in demand from retail space to warehousing or other types of storage space can be expected, the magnitude of that shift is still uncertain. It may be amplified, however, with the widespread adoption of self-driving technologies, which are expected to substantially cut the cost of freight transportation (Wadud, 2017; Andersson and Ivehammar, 2019). This shift may be less important for e-grocery than for e-shopping in general. It will also depend on the dominant form of grocery (e.g., click-and-pick vs. home deliveries) and on the extent to which grocers adopt omni-channel strategies, where the business processes of multiple retail channels are increasingly integrated (Marchet *et al*., 2018).

Although online shopping is often invoked to explain high-profile bankruptcies among U.S. retailers (e.g., Sears, Sport Authority, Payless) and store closures by major retailers such as J.C. Penney and Macy's that took place over the last decade, other factors may have contributed just as much to retail store closures in the U.S., including an excessive number of malls and shifts in consumers' spending habits (Thompson, 2017). Grocery stores have also been affected as they are facing increasing competition from Walmart, Aldi, and Amazon (Meyerson, 2019), but casualties so far have only been small and regional firms that were out of sync with the markets, or could not afford the costly investment to expand online (Meyerson, 2020).

In addition to the obstacles associated with e-shopping in general, such as risks associated with the safety of internet connections and the payment system, or the lack of complete information about online orders (Wat *et al*., 2005), e-grocers need to adopt efficient home delivery solutions that can accommodate the requirements of groceries (Punakivi and Saranen, 2001). This entails ensuring delivery during tight time windows while observing adequate temperature requirements, the ability to promptly respond to demand, and having enough information to avoid failed home deliveries due to a customer's absence (Punakivi and Saranen, 2001), all of which are issues that delivery service companies or the postal service typically do not face. A number of options have been proposed to manage the demand for grocery deliveries based either on time slot allocations or time slots pricing, either in static (i.e., forecast-based) or dynamic (i.e., real-time) settings (Agatz *et al*., 2013; Klein *et al*. (2019).

More generally, brick-and-mortar retailers aiming to be competitive online need to redefine their logistics networks (Wollenburg *et al*., 2018). This involves adapting inventory management, distribution settings (i.e., the number and type of logistics facilities handling online orders), fulfillment strategy, deliveries, and return management policies (Marchet *et al*., 2018).

2.3 Data

As explained in the introduction, this study relied on data from several publicly available datasets. First, to understand how deliveries from online shopping have been changing over the last few years, data from both the 2009 and the 2017 National Household Travel Surveys (NHTS) (FHWA, 2010, 2018) are analyzed. These surveys provide comprehensive national data on households, their members, their vehicles, and daily travel for all purposes and by all modes of transportation. Second, since (to the best of my knowledge) there is no publicly available national dataset on e-grocery in the U.S., data from the 2017 American Time Use Survey (ATUS) is analyzed to characterize U.S. consumers who shop online for groceries. These datasets and variable choices are presented in turn.

2.3.1 Data from the 2009 and 2017 NHTS

Compared to the 2009 NHTS, data for the 2017 NHTS were collected using a new sampling strategy and a new methodology (e.g., data were retrieved using a self-completed web-based survey instead of via an interviewer assisted phone survey) to lower the burden on respondents and improve coverage. In 2009, the sample frame was obtained using random digit dialing of landline phone numbers, but this approach is no longer appropriate because since 2016 over half of American homes have abandoned landlines in favor of cell phones (Blumberg and Luke, 2017). Approximately 45% of households who participated in the 2017 NHTS have no land line, and many include ethnic minorities and younger people (McGuckin and Fucci, 2018).

Overall, the 2017 NHTS collected data from 129,696 households who made 923,572 trips between April 2016 and April 2017. Conversely, the 2009 NHTS collected data from 150,147 households who made 1,167,321 trips between March 2008 and April 2009 (FHWA, 2011).

Both the 2009 and the 2017 NHTS inquired about the number of times each respondent purchased something online and had it delivered in the 30 days preceding their survey day. However, while the 2009 NHTS question specifies deliveries to home, the 2017 NHTS question I analyzed does not. I aggregated individual answers to this question by household to create the dependent variable for my models that explain deliveries from online shopping. Households are the focus here because it is not uncommon for one household member to order goods for other household members, especially if they are children.

Both surveys also asked how frequently respondents use the internet. This information is used to exclude households who stated they never use the internet but receive deliveries from online purchases.

To explain deliveries from online purchases, I relied on a wide range of directly observable household characteristics and on land use variables available in both the 2009 and the 2017 NHTS. First, my models include variables describing household composition (see Table 2.1), and a count of the number of household women over 18, as women are often in charge of shopping. Since the age of household members likely matters, I kept track of the number of children (household members under 18), and of the number of household members 18 or older from each generation, as defined by the Pew Research Center (2018).

Table 2.1 Summary statistics for data used to explain deliveries from online shopping

	2017 NHTS (N=123,148)				**2009 NHTS** (N=134,371)			
Variable	Min	Mean	Max	Std. Dev.	Min	Mean	Max	Std. Dev.
Number of deliveries to the household from online shopping in the past month	0	4.709	60	6.700	0	2.105	60	4.194
Household composition								
1 adult without children	0	0.181	1	0.385	0	0.102	1	0.302
2+ adults without children	0	0.214	1	0.410	0	0.213	1	0.410
1 adult with children	0	0.035	1	0.184	0	0.026	1	0.160
1 adult, retired, without children	0	0.140	1	0.348	0	0.130	1	0.336
2+ adults, retired, without children	0	0.242	1	0.428	0	0.276	1	0.447
Count of household members younger than 18	0	0.355	8	0.816	0	0.467	11	0.931
Count of women 18 and over in the household	0	0.948	8	0.511	0	0.956	6	0.476
Age structure (# of household members ≥18)								
Generation Z (born after 1997)	0	0.043	6	0.224				
Generation Y (born 1981 to 1996)	0	0.329	8	0.661	0	0.125	5	0.400
Generation X (born 1965 to 1980)	0	0.392	5	0.682	0	0.324	4	0.643
Baby Boomers (born 1946 to 1964)	0	0.734	7	0.809	0	0.730	4	0.815
Silent generation (born before 1946)	0	0.282	4	0.578	0	0.585	5	0.761
Annual household income								
First quintile	0	0.192	1	0.394	0	0.214	1	0.410
Second quintile	0	0.221	1	0.415	0	0.175	1	0.380
Fourth quintile	0	0.236	1	0.424	0	0.208	1	0.406
Fifth quintile	0	0.172	1	0.378	0	0.200	1	0.400
Count of household members who:								
Work full-time	0	0.756	6	0.797	0	0.736	5	0.784
Work part-time	0	0.208	6	0.460	0	0.222	5	0.471
Highest education achieved in the household								
Less than high school	0	0.020	1	0.138	0	0.049	1	0.216

High school graduate or GED	0	0.133	1	0.340	0	0.210	1	0.408
Bachelor's degree	0	0.263	1	0.440	0	0.237	1	0.425
Graduate or professional degree	0	0.302	1	0.459	0	0.221	1	0.415
Ethnicity of the household head								
Black or African American	0	0.076	1	0.265	0	0.062	1	0.241
Asian	0	0.037	1	0.190	0	0.020	1	0.138
More than one ethnicity	0	0.028	1	0.165	0	0.006	1	0.077
Other (single ethnicity)	0	0.024	1	0.153	0	0.046	1	0.209
Household head is Hispanic/Latino	0	0.069	1	0.253	0	0.067	1	0.250
Count of household members born abroad	0	0.196	9	0.583	0	0.169	7	0.513
Count of household members with a medical condition that makes it difficult to travel outside of home	0	0.191	7	0.448	0	0.218	6	0.463
Household owns dwelling	0	0.757	1	0.429	0	0.873	1	0.333
Number of vehicles per household adult	0	1.138	12	0.630	0	1.111	27	0.571
Land use								
Household resides in lower density census tract	0	0.325	1	0.468	0	0.345	1	0.475
Household resides in higher density census tract	0	0.257	1	0.437	0	0.233	1	0.423
Household MSA has heavy rail	0	0.154	1	0.361	0	0.170	1	0.376

Notes:
1) The dependent variable (deliver) was truncated at 60.
2) MSA stands for Metropolitan Statistical Area.
3) Lower density census tracts have less than 300 people per square mile. Higher density census tracts have over 7,000 people per square mile.

Because income categories differ between the 2009 and the 2017 NHTS, approximate quintiles (20% strata) are created. The number of household adults who work part-time and full-time since working decreases the time available for activities such as shopping is also added. To capture the highest educational achievement in the household, five common categories shown in Table 2.1 are relied on here.

In addition, I included race variables, a Hispanic/Latino indicator, and the number of household members born abroad is counted. The purpose of these variables is to help capture cultural differences or uneven access to online shopping.

People with a medical condition that impairs their mobility would likely benefit from online shopping, so an indicator variable for them is created. Home ownership (a wealth indicator) may also play a role here, and so could the number of vehicles per adult.

Finally, two population density indicator variables (<300 people per square mile and ≥7,000 people per square mile) and an indicator variable for heavy rail are created. The low-density variable could reflect that people in rural areas have fewer shopping options and may therefore benefit more from online shopping. Conversely, the high-density indicator (and the heavy rail variable) could capture impacts from enhanced online shopping and delivery options.

After removing households with missing data and with over 2 deliveries per day because the latter are unusual and influential observations (39 and 117 observations for the 2009 and 2017 NHTS respectively), my final samples have 134,371 households for 2009 and 123,148 for 2017. Summary statistics for both are provided in Table 2.1.

2.3.2 Data from the 2017 ATUS

To get a profile of U.S. consumers who shop for groceries online, data from the 2017 ATUS is analyzed. The main goal of this survey is to understand how noninstitutionalized U.S. residents who are civilians 15 or older allocate their time. ATUS samples individuals who participated in the Current Population Survey. It is conducted annually by the U.S. Census Bureau by phone, using Computer Assisted Interview software.

Although grocery shopping is a household activity, my basic unit of analysis here is the individual because only one person per household participates in ATUS. For each person in my sample, I gathered a broad range of socio-economic variables. Individual characteristics include marital status, gender, generation (based on Pew Research Center definitions), work status, education level, race, Hispanic/Latino status, and presence of a mobility impairment. Household characteristics include the number of children and annual income.

Two dependent variables obtained by combining activity (grocery shopping) and location data are considered. The first dependent variable indicates whether or not a respondent shopped for groceries in a store during their ATUS survey day. The second dependent variable indicates whether or not a respondent shopped for groceries online (i.e., when the respondent shopped for groceries, he/she was neither in a grocery store nor in another store/mall). Out of 10,223 persons in the sample, 1,420 (13.9%) shopped for groceries in a store, but only 59 (0.57%) shopped for groceries online. This low percentage is not surprising since by August 2018, of the 16% of U.S. adults who had ever ordered groceries online, 7 out of 10 did so twice a month or less (Redman, 2018). Interestingly, nobody in the ATUS dataset shopped for groceries both in a store and online on their survey day.

Since e-grocery shopping was not available everywhere in the United States in 2017, the sample was pared down by keeping only respondents who reside in core-based statistical areas[3]

[3] A core-based statistical area (CBSA) is a U.S. geographic area that consists of one or more counties (or equivalents) anchored by an urban center of at least 10,000 people plus adjacent counties that are socioeconomically

where at least one ATUS respondent shopped for groceries online. This reduced my ATUS sample to 2,934 respondents. This sample includes only people who are African American, Asian, or White, and it does not include people with a mobility impairment. As a result, other ethnic variables and the variable indicating the presence of a mobility impairment could not be included in the logit models estimated on that sample.

Summary statistics for two ATUS samples are presented in Table 2.2.

Table 2.2 Summary statistics for ATUS analysis

	N=10,223			N=2,934		
Variable	Min	Mean	Max	Min	Mean	Max
Dependent variables						
Conventional grocery shopping indicator	0	0.148	1	0	0.150	1
E-grocery shopping indicator	0	0.006	1	0	0.017	1
Explanatory variables						
Marital status (baseline: married)						
Never married	0	0.240	1	0	0.266	1
No Spouse	0	0.279	1	0	0.243	1
Gender (1 if female)	0	0.546	1	0	0.545	1
Age (baseline: Baby Boomer)						
Generation Z (born after 1997)	0	0.053	1	0	0.057	1
Generation Y (born 1981 to 1996)	0	0.221	1	0	0.222	1
Generation X (born 1965 to 1980)	0	0.272	1	0	0.298	1
Silent generation (born < 1946)	0	0.132	1	0	0.113	1
Number of household children	0	0.745	11	0	0.763	11
Annual household income (baseline: $60,000 to $99,999)						
Less than $30,000	0	0.268	1	0	0.223	1
$30,000 to $59,000	0	0.265	1	0	0.258	1
Over $100,000	0	0.244	1	0	0.306	1
Work status (baseline: full-time job)						
Part-time job	0	0.125	1	0	0.131	1
Not employed	0	0.393	1	0	0.363	1
Education (baseline: college degree)						
Less than high school	0	0.122	1	0	0.123	1
High school	0	0.235	1	0	0.207	1
Some college / associate degree	0	0.271	1	0	0.241	1

tied to that urban center by commuting. Source: https://catalog.data.gov/dataset/core-based-statistical-areas-national.

Graduate / professional degree	0	0.146	1	0	0.172	1
Ethnicity (baseline: White)						
African American	0	0.147	1	0	0.176	1
Asian	0	0.041	1	0	0.071	1
Other	0	0.021	1	NA	NA	NA
Hispanic/Latino	0	0.155	1	0	0.208	1
Mobility impairment	0	0.037	1	NA	NA	NA

Notes: 1) All variables are binary or count variables so I did not present standard deviations.
2) The smaller sample (N=2,934) was obtained by keeping only ATUS respondents from core-based statistical areas (CBSA) where at least one other ATUS respondent shopped for groceries online. I removed respondents whose ethnicity is "other" (neither African American, Asian, nor White) and who had a "mobility impairment" because none of them shopped for groceries online so the impact of these characteristics on e-grocery shopping could not be estimated.

2.4 Models

In this section, I first describe the mixture models which I relied on to explain changes between 2009 and 2017 in residential deliveries from online shopping in the United States. Then the strategy is explained to characterize Americans who shop at brick-and-mortar grocery stores versus Americans who engage in e-grocery. Combining results from both analyses could allow identifying areas with a high potential for the delivery of groceries from online shopping.

2.4.1 Number of deliveries from online shopping (2017 and 2009 NHTS)

I first explained the number of deliveries from online shopping by household in the last 30 days preceding their assigned survey day, and how it changed between 2009 and 2017. In 2009, 57.8% of households had no deliveries from Internet shopping during the 30 days preceding their survey day; in 2017, this percentage fell to 31.3%. To account for this relatively high percentage of zeros, mixture models (Greene, 2011) are estimated, which assume that my samples are composed of two distinct groups of households. Households in the first group never purchase goods online, and therefore get no deliveries from online shopping. Households in the second group make online purchases from time to time, so the number of their deliveries is modeled via a count model. To model the probability of belonging to either group, a logit model is relied on. The resulting mixture model can be written (Greene, 2011):

$$Pr(N_i = n_i|\boldsymbol{x_i}, \boldsymbol{z_i}) = \begin{cases} \varphi(\gamma'\boldsymbol{z}_i) + \{1-\varphi(\gamma'\boldsymbol{z}_i)\}g(0|\boldsymbol{\beta}'\boldsymbol{x_i}), & \text{if } n_i = 0 \\ \{1-\varphi(\gamma'\boldsymbol{z}_i)\}g(n_i|\boldsymbol{\beta}'\boldsymbol{x_i}), & \text{if } n_i > 0 \end{cases} \quad (1)$$

where:

- N_i designates the random variable that generated n_i, which is the number of packages from online shopping received by household *i* in the 30 days up to their survey day;
- $\varphi(\gamma'\boldsymbol{z}_i)$ is the probability that household *i* belongs to the group that never makes online purchases (the first group);
- $g(n_i|\boldsymbol{\beta}'\boldsymbol{x_i})$ is the probability that household *i* received $n_i \geq 0$ packages from online purchases over the 30 days up to their survey day;
- $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}$ are vectors of unknown coefficients that need to be estimated; and
- $\mathbf{z}_i$ and $\mathbf{x}_i$ are vectors of explanatory variables, respectively for the logit model that identifies households who never buy goods online (the first group) and for the count model.

If modeling counts with a Poisson process, a zero-inflated Poisson (ZIP) mixture model is obtained; using a negative binomial regression model instead gives a zero-inflated negative binomial (ZINB) mixture model (Long, 1997).

To avoid multicollinearity, I checked the variance inflation factors for all my explanatory variables are below 10 (they are).

To assess if the 2017 estimate of the coefficient for an explanatory variable ($\hat{\beta}_{2017}$) differs from its 2009 value ($\hat{\beta}_{2009}$), the test statistic is relied on:

$$Z = \frac{\hat{\beta}_{2017} - \hat{\beta}_{2009}}{\sqrt{SE^2(\hat{\beta}_{2017}) + SE^2(\hat{\beta}_{2009})}}. \tag{2}$$

Under the null hypothesis H_0: $\beta_{2017} = \beta_{2009}$, Z has approximately a standard normal distribution as the difference of two (asymptotically) independent normal random variables, assuming that the 2009 and the 2017 NHTS samples are independent (Greene, 2011).

2.4.2 Characteristics of U.S. e-grocery shoppers (2017 ATUS)

To obtain a baseline profile of Americans who shop at brick-and-mortar grocery stores, a logit model (Greene, 2011) on the full ATUS sample (N=10,223) is first estimated.

Then two more logit models are estimated on the sub-sample (N=2,934) of ATUS respondents who live in core-based statistical areas where at least one person shopped for groceries online. As mentioned above, this sub-sample is constructed because e-grocery shopping was not available everywhere in the United States in 2017. The first logit model estimated on the ATUS sub-sample again characterizes ATUS respondents who shopped at a physical grocery store. It allows me to check that grocery shoppers in this sub-sample do not differ substantially from those in the full ATUS dataset. The second logit model characterizes ATUS respondents who shopped for groceries online.

Since only 59 people shopped for groceries online in the 2017 ATUS, it is difficult to fully capture the determinants of online grocery shopping. Therefore, the distributions of selected socio-economic characteristics of people who shopped for groceries online with those who shopped in stores are also compared using Kruskal-Wallis (KW) tests (Conover, 1999). A KW test assesses whether different samples originate from the same distribution. Finally, the distribution of shopping start times (the self-reported time when ATUS respondents started browsing for groceries online) is analyzed since one argument for e-grocery is the convenience afforded by the ability to shop at any time.

2.4.3 Interpreting results

As discussed above, ZINB (a mixture consisting of a negative binomial regression model with a logit) and logit models are estimated. For a negative binomial regression model, the coefficient of an explanatory variable represents the difference of the logs of expected counts when that explanatory variable is increased by one unit, holding constant all other explanatory variables (Long, 1997). In a logit model, the coefficient of an explanatory variable represents the change in the logit (the log of the probability of 1 divided by the probability of 0) of the probability associated with a unit change in that explanatory variable holding all other predictors constant.

2.5 Results

My results were obtained with Stata 15.1. They are presented in Table 2.3 and Table 2.4, and illustrated from Figure 2.1 to Figure 2.3. In Table 2.3, shaded numbers indicate when the 2017 value of a coefficient differs from its 2009 value. To better link explanations with results presented in Table 2.3 and Table 2.4, estimated coefficients and their statistical significance (see notes below these tables) are occasionally reported.

2.5.1 Number of deliveries from online shopping (2017 and 2009 NHTS)

Before discussing results from count models, it is instructive to contrast monthly household package deliveries from online shopping for 2009 and 2017 (Figure 2.1). Results are weighted to be representative of the U.S. population. As expected, Figure 2.1 shows a sharp reduction in the percentage of households who do not get any deliveries (it drops from 57.8% to 31.3%) together with an increase in the percentage of households who received packages. This increase is especially marked for the 6-10 deliveries category (from 6.6% to 15.6%) and for more than 10 deliveries (from 3.5% to 13.2%).

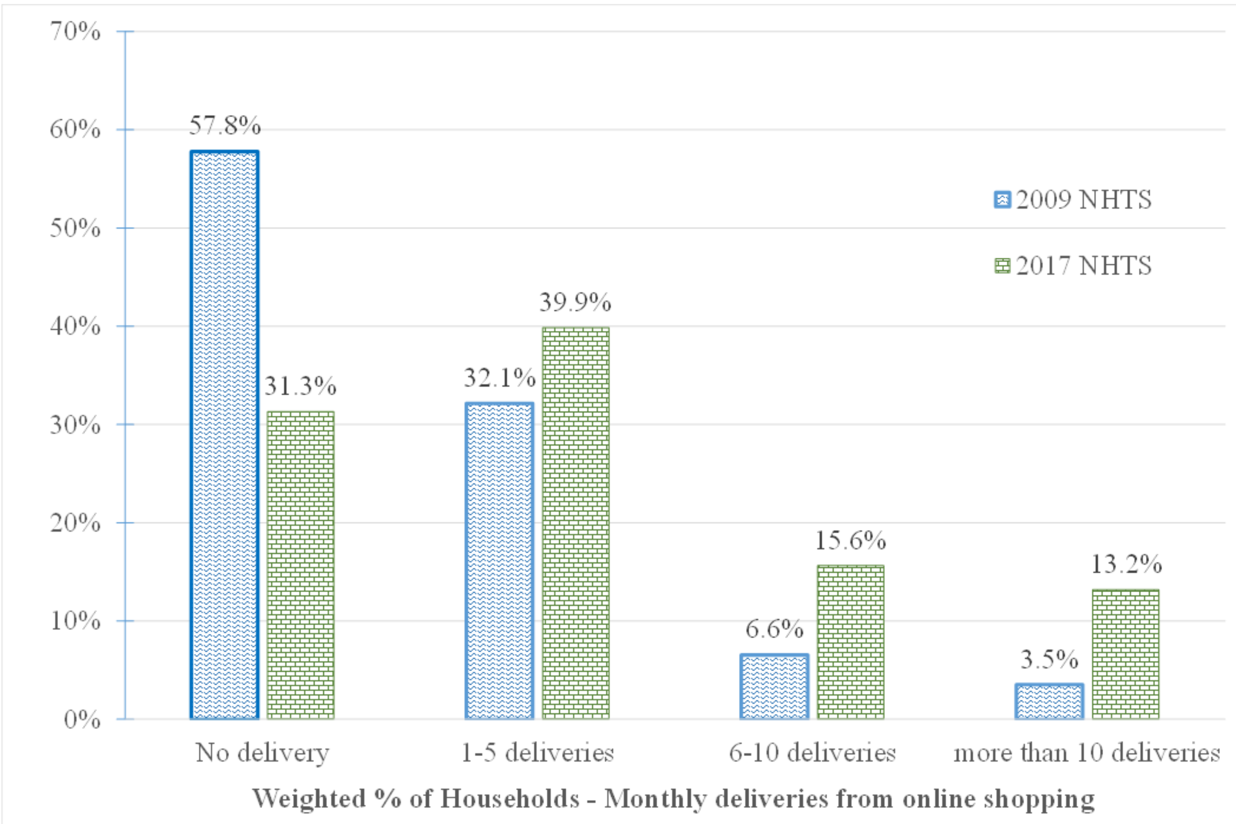


Figure 2.1 Change in monthly household package deliveries from online shopping

Source: 2009 and 2017 National Household Travel Surveys

For robustness, Poisson, negative binomial, ZIP, and ZINB models are estimated for 2009 and 2017 NHTS samples using maximum likelihood. The ZINB models presented in Table 2.3 have the best (lowest) AIC and BIC values.

Table 2.3 Results for Zero-Inflated Negative Binomial models (2009 & 2017 NHTS) explaining deliveries from online shopping

Variable	Count sub-models (ZINB) (number of deliveries from online shopping)		Logit sub-models (households who never buy goods online)	
	2017	2009	2017	2009
Household composition (baseline: 2+ adults, not retired, with children)				
1 adult without children	-0.368***	-0.069	0.304**	0.304***
2+ adults without children	-0.012	0.035	0.157	0.259***
1 adult with children	-0.204***	-0.094*	-0.14	-0.223*
1 adult, retired, without children	-0.465***	-0.084	0.506***	0.677***
2+ adults, retired, without children	-0.006	-0.002	-0.305**	0.041
Count of household members younger than 18	0.039***	0.032***	0.120***	-0.049
Count of women 18 and over in the household	0.051***	-0.005	-0.427***	-0.167***
Age structure (count of household members ≥18)				
Generation Z (born after 1997)	0.154***		-0.185*	
Generation Y (born 1981 to 1996)	0.239***	0.268***	-0.340***	-0.820***
Generation X (born 1965 to 1980)	0.155***	0.338***	-0.006	-0.792***
Baby Boomers (born 1946 to 1964)	0.033**	0.268***	0.371***	-0.544***
Silent generation (born before 1946)	-0.148***	0.133***	1.028***	0.092*
Annual household income (baseline: third quintile)				
First quintile	-0.189***	-0.112***	1.391***	1.083***
Second quintile	-0.124***	-0.091***	0.622***	0.506***
Fourth quintile	0.155***	0.116***	-0.586***	-0.420***
Fifth quintile	0.407***	0.330***	-1.721***	-1.084***
Work Status. Count of household workers who:				
Work full-time	-0.005	-0.049***	-0.211***	-0.147***
Work part-time	0.011	-0.022	-0.361***	-0.275***
Highest education achieved in the household (baseline: some college/associate degree)				
Less than high school	-0.304***	-0.108	1.544***	1.923***
High school graduate or GED	-0.168***	-0.142***	0.773***	0.728***
Bachelor's degree	0.103***	0.114***	-0.490***	-0.471***

Graduate or professional degree	0.172***	0.228***	-0.780***	-0.764***
Ethnicity of the household head (baseline: White)				
Black or African American	-0.336***	-0.208***	0.713***	0.856***
Asian	-0.132***	-0.203***	0.424***	0.251*
More than one ethnicity	0.017	-0.044	0.152	0.270
Other (single ethnicity)	-0.088**	-0.025	0.441***	0.627***
Household head is Hispanic/Latino	-0.136***	-0.108***	0.440***	0.415***
Count of household members born abroad	-0.070***	-0.025*	0.059	0.205***
Count of household members with a medical condition that makes it difficult to travel outside of home	0.109***	0.086***	-0.078**	0.233***
Household owns dwelling	0.000	-0.008	-0.084*	-0.114**
Number of vehicles per household adult	0.048***	0.101***	-0.258***	-0.177***
Land use				
Household resides in lower density census tract	0.031***	-0.002	0.133***	0.145***
Household resides in higher density census tract	0.054***	-0.009	-0.005	-0.081*
Household MSA has heavy rail	0.084***	0.084***	-0.279***	-0.108**
Constant	1.324***	0.418***	-1.689***	0.301**
Ln(α)	-0.217***	0.134***		

Notes:
1) The dependent variable is the number of deliveries to the household from online shopping in the past month. It was truncated at 60.
2) * p<0.05, ** p<0.01, *** p<0.001. To assess statistical significance, robust standard errors are used to mitigate the potential impacts of heteroskedasticity.
3) A shaded cell indicates that the 2017 coefficient value for an explanatory variable differs from its 2009 value (p-value ≤ 0.05).
4) Count sub-models explain the number of deliveries to each household from online shopping. Logit sub-models explain the characteristics of households who never order goods online.
5) Lower density census tracts have less than 300 people per square mile. Higher density census tracts have over 7,000 people per square mile.
6) MSA: Metropolitan Statistical Area.
7) α is the inverse of the scale parameter of the gamma noise variable in the negative binomial count component of a zero-inflated negative binomial model. Stata estimates and reports Ln(α), as well as its statistical significance.
8) The sample size is 123,148 for the 2017 NHTS and 134,371 for the 2009 NHTS.

From Table 2.3, it is found that the impact of household composition on deliveries from internet shopping has been changing over time. Overall, the likelihood to never order online has decreased for most household types, and especially for households with 2 or more retired adults (-0.305**). These households are now less unlikely than baseline households to shop online, possibly because older adults with a spouse are more likely to use modern information technologies (Vroman *et al*., 2015). However, differences in deliveries have sharpened (household composition made little difference in 2009): compared to my baseline households (2+ adults with children), 1-adult households (with and without children, retired or not) received fewer deliveries in 2017.

The impact of the number of children (household members under 18) has also been evolving and it is mixed. In 2017, having more children increased the likelihood of never shopping online, but it also slightly increased the number of deliveries from online shopping.

As reported by Ferrell (2005), having more females in the household matters. In 2009, it decreased the likelihood of never shopping online (-0.167***), without impacting the number of deliveries. In 2017, it both decreased the likelihood of never shopping online (-0.427***), and increased deliveries (0.051***).

The importance of the generational structure of households has also been changing, suggesting a broader adoption of online shopping. In 2009, an increase in the number of household members from younger generations (Gen X and Gen Y, and to a lesser extent Baby Boomers) sharply reduced the likelihood of never shopping online. Furthermore, increasing the number of household adults raised the number of deliveries. By contrast, in 2017, the magnitude of model coefficients for Gen X and Gen Y decreased. As Baby Boomers aged, however, they became more likely to never order goods online (0.371***), although not as much as members of the Silent generation (1.028***). As Parment (2013) explained, unlike Gen Y members, Baby Boomers often prefer to start a purchase with a retailer they trust, before eventually committing to a purchase, either online or in a store.

The impact of household income is monotonic. As their income rises, households are less likely to never order goods online and they tend to get more deliveries (as in Wang and Zhou, 2015). This effect increased across the board between 2009 and 2017. Likewise, as they gain workers, households are less likely to never order goods online, although the impact on deliveries is insignificant in 2017 (and in 2009 for part-time workers).

Similarly, as their level of education increases, households are less likely to never order goods online, and their deliveries increase. Except for households with less than a high school education in 2009, this relationship is monotonic. This result echoes the reported correlation between education and computer proficiency (Burroughs and Sabherwal, 2002).

As reported by Ren and Kwan (2009), race matters. Compared to White households (our baseline), African Americans (0.856*** for 2009) and Asians (0.251* for 2009) are more likely to never order goods online and they tend to receive fewer deliveries. The same holds for Hispanic/Latino households. Conversely, whereas people born abroad were more likely to never order goods online in 2009, it is no longer the case in 2017. Having more foreign-born household members slightly decreases the number of deliveries, however.

People with a medical condition that hinders their mobility appear to take better advantage of the convenience of online shopping. Whereas in 2009 they were more likely to never order goods online (0.233***), this effect disappeared in 2017 and those who shopped online received more deliveries in 2017 (0.109***) than in 2009 (0.086***).

Home ownership decreases the likelihood of never ordering goods online but it does not impact deliveries from online purchases. Likewise, households who have more vehicles per adult are less likely to never order goods online and they tend to receive more packages from online shopping. This result agrees with Zhou and Wang (2014), who reported that online shopping stimulates shopping trips and found that the number of household vehicles is positively correlated with the level of online shopping.

As expected, population density also plays a role here, although its impact is relatively small. In lower density areas (<300 people per square mile), households are more likely to never order goods online, but those who do received slightly more deliveries in 2017 (0.031***). This result illustrates that e-shopping has the potential to increase the range of products available to households currently underserved by brick-and-mortar stores. Conversely, in denser areas (>7,000 people per square mile), households are slightly less likely to never order goods online and they receive more deliveries (in 2017, not 2009). The same holds for households who reside in a Metropolitan Statistical Area (MSA) with heavy rail.

To capture the endogeneity of the land use and vehicle ownership variables, I could have estimated generalized structural equation models (GSEM; Kline, 2010) but ZINB models are adopted instead for two main reasons. First, these endogenous effects are small, as it is verified by estimating ZINB models without land use and vehicle ownership variables. Second, GSEM models with non-continuous endogenous variables (population density is categorical and the presence of heavy rail is binary) are more difficult to interpret.

2.5.2 Characteristics of e-grocery shoppers (2017 ATUS)

Results from my analysis of the characteristics of U.S. grocery shoppers are shown in Table 2.4. In this table, a positive coefficient for a variable indicates that the probability to shop for groceries (or e-groceries for the last column of Table 2.4) increases with that variable.

Table 2.4 Results for logit models characterizing in-store & e-grocery shoppers (2017 ATUS)

Variable	**In-store grocery shopping**		**E-grocery shopping**
	N=10,223	N=2,934	N=2,934
Marital status (baseline: married)			
Never married	0.009	0.017	0.392
No Spouse	0.033	0.065	0.51
Gender (1 if female)	0.390***	0.413***	0.879**
Age (baseline: Baby Boomer)			
Generation Z (born after 1997)	-1.134***	-1.274***	-1.119
Generation Y (born 1981 to 1996)	-0.057	-0.353**	0.783
Generation X (born 1965 to 1980)	0.044	-0.151	0.67
Silent generation (born < 1946)	-0.082	-0.224	0.092
Number of household children	0.01	0.045	0.041
Annual household income (baseline: $60,000 to $99,999)			
Less than $30,000	0.073	0.132	-0.365
$30,000 to $59,000	0.106	0.151	0.075
Over $100,000	0.184**	0.133	-0.225
Work status (baseline: full-time job)			
Part-time job	0.102	0.021	0.493
Not employed	0.303***	0.293**	0.368
Education (baseline: college degree)			
Less than high school	-0.444***	-0.637***	0.295
High school	-0.276***	-0.344**	0.311
Some college / associate degree	-0.177**	-0.315**	-0.212
Graduate / professional degree	0.061	-0.064	-0.266
Ethnicity (baseline: White)			
African American	-0.393***	-0.338**	-0.503
Asian	0.033	0.107	-0.377
Other	0.01	NA	NA
Hispanic/Latino	0.116	0.108	-0.778
Mobility impairment	-1.012***	NA	NA
Constant	-1.971***	-1.819***	-5.187***

Notes: 1) * p<0.10, ** p<0.05, *** p<0.01.
2) For in-store grocery shopping, the dependent variable equals 1 if the ATUS respondent shopped in a brick-and-mortar grocery store on the survey day and zero otherwise. For e-grocery shopping, the dependent variable equals 1 if the ATUS respondent shopped online for groceries on the survey day and zero otherwise.
3) My reduced sample (N=2,934) was obtained by keeping only ATUS respondents from core-based statistical areas (CBSA) where at least one other ATUS respondent shopped for groceries online. I estimated e-grocer characteristics using this sample to avoid the bias that would result from analyzing people who had no access to e-grocery shopping. I removed respondents whose ethnicity is “other” (neither African American, Asian, nor White) and who

had a "mobility impairment" because none of them shopped for groceries online so the impact of these characteristics on e-grocery shopping could not be estimated.

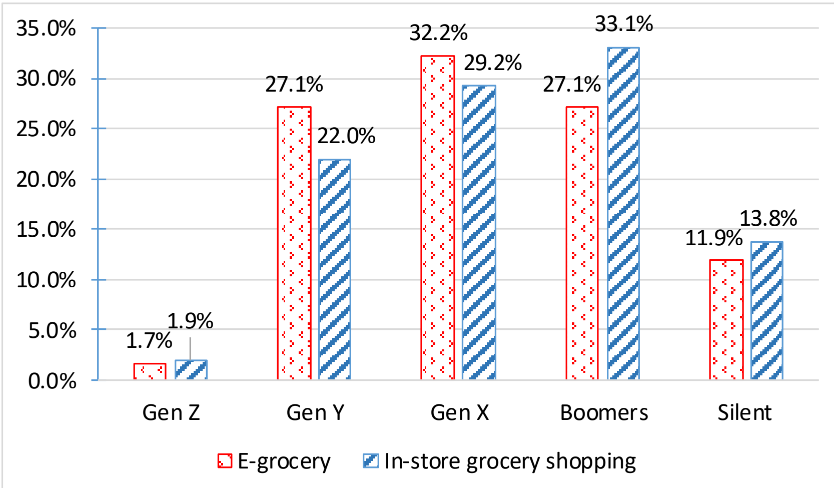
Starting with the logit model that characterizes people who shopped at brick-and-mortar grocery stores on their ASUS survey day (Column 2 in Table 2.4), it is found that women are more likely to shop for groceries than men (0.390***), which is well-known (e.g., see Morganosky and Cude, 2002; or Li *et al.*, 2018). Apart from members of the Z generation who are less likely to shop for groceries, there are no generational effects. Likewise, the number of household children does not matter, and neither does household income, except for members of the highest income group (0.184*). As expected, grocery shoppers are more likely to be "unemployed" (0.303***), likely because they take care of the household while other household adults are at work and homemakers (still usually women) are considered unemployed. Interestingly, people with less education are less likely to shop for groceries. Race matters but only for African Americans, who appear to shop for groceries less frequently than other groups, which agrees with previous findings that African Americans have less access to supermarkets (Morland *et al*., 2002; Beaulac *et al*., 2009). To compensate, poor people in predominantly black neighborhoods or in poor rural areas tend to shop at dollar stores[4] (Whalen, 2018), and rely more on fast food outlets (James *et al*., 2014). Lastly, respondents with a medical condition that impairs their mobility are less likely to shop for groceries (-1.012***).

When focusing on core-based statistical areas where at least one respondent in my sample shopped for groceries online (Column 3 in Table 2.4), two main differences are observed. First, members of the Y generation are less likely to shop for groceries. Second, all income groups become equally likely to shop for groceries.

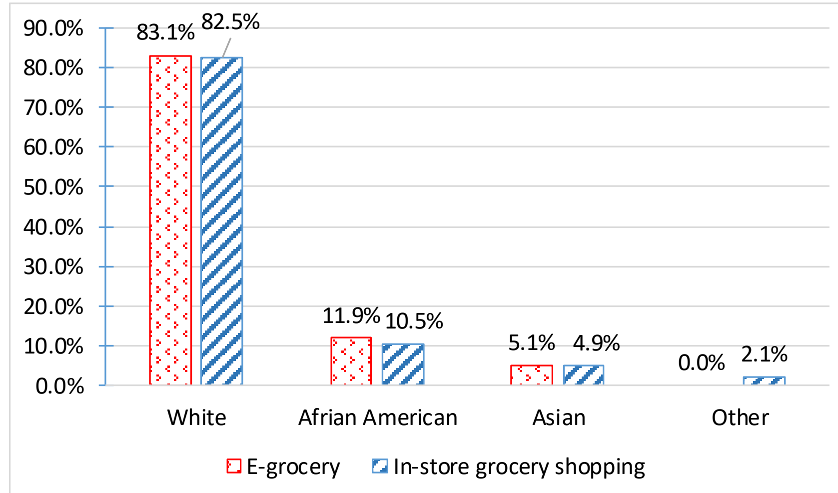
[4] Dollar stores are discount retailers that sell a wide range of products at low prices. Dollar Tree and Dollar General are the largest dollar stores in the United States, with over 14,000 locations each.

By contrast, the only socio-economic characteristic that is statistically significant for people who shopped online for groceries is gender: women are more likely to shop online for groceries (0.879**) than men, and this gender gap appears wider than for in-store shopping. This result echoes the findings of Morganosky and Cude (2002) in their study of 10 U.S. markets based on three datasets collected between 1998 and 2001. However, when interpreting these results, it is important to keep in mind that the underlying dataset (N=2,934) only includes 49 people who shopped for groceries online (lost 10 respondents from my initial sample because their location is unknown).

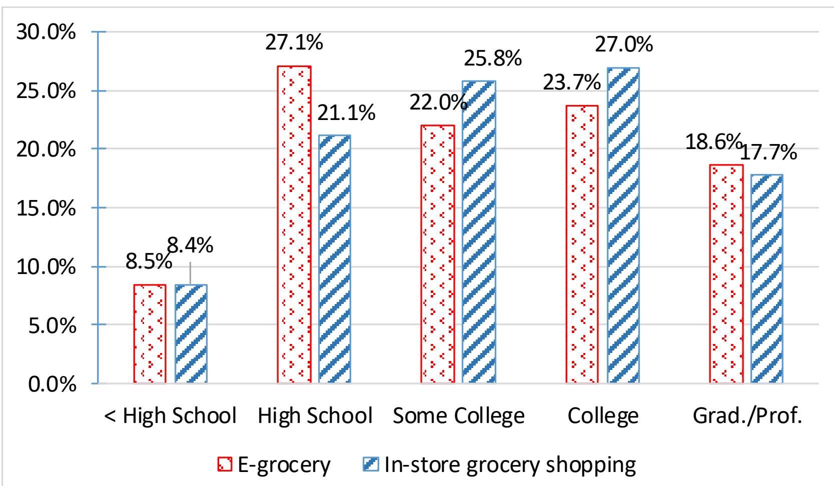
To take full advantage of my relatively small sample of only grocery shoppers, differences in the distribution of selected socio-economic characteristics of people who shop for groceries in stores and online using Kruskal-Wallis (KW) tests are also explored. Figure 2.2 shows the empirical distributions of selected characteristics of the 1,420 people in the 2017 ATUS dataset who shopped for groceries in stores and the 59 who shopped online (so 24 times more people shopped for grocery in stores than online on a given day in 2017). Only the KW test for gender (not shown) was significant and it indicated that the gender difference for grocery shopping online is more marked than for grocery shopping in stores. Crosstabulation analyses using χ^2 tests gave similar results.



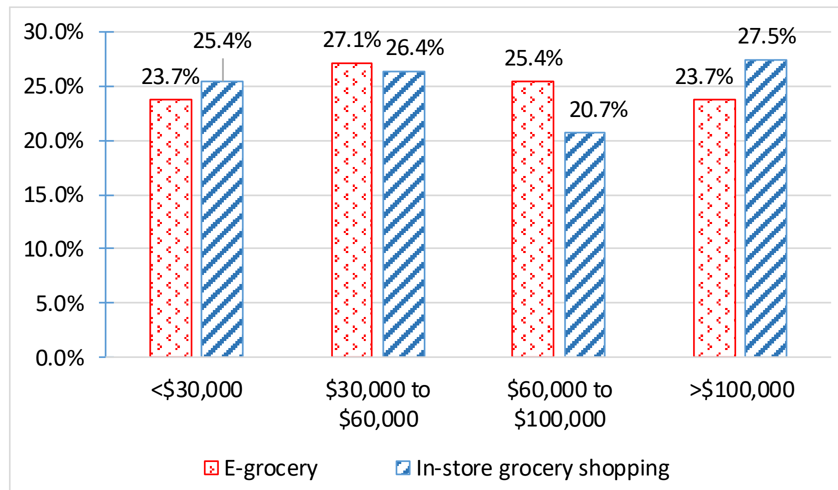
Panel A: E-groceries and age (generations)



Panel B: E-groceries and race



Panel C: E-groceries and education



Panel D: E-groceries and household income

Figure 2.2 Comparison of characteristics between online and conventional grocery shoppers

Source: 2017 American Time Use Survey

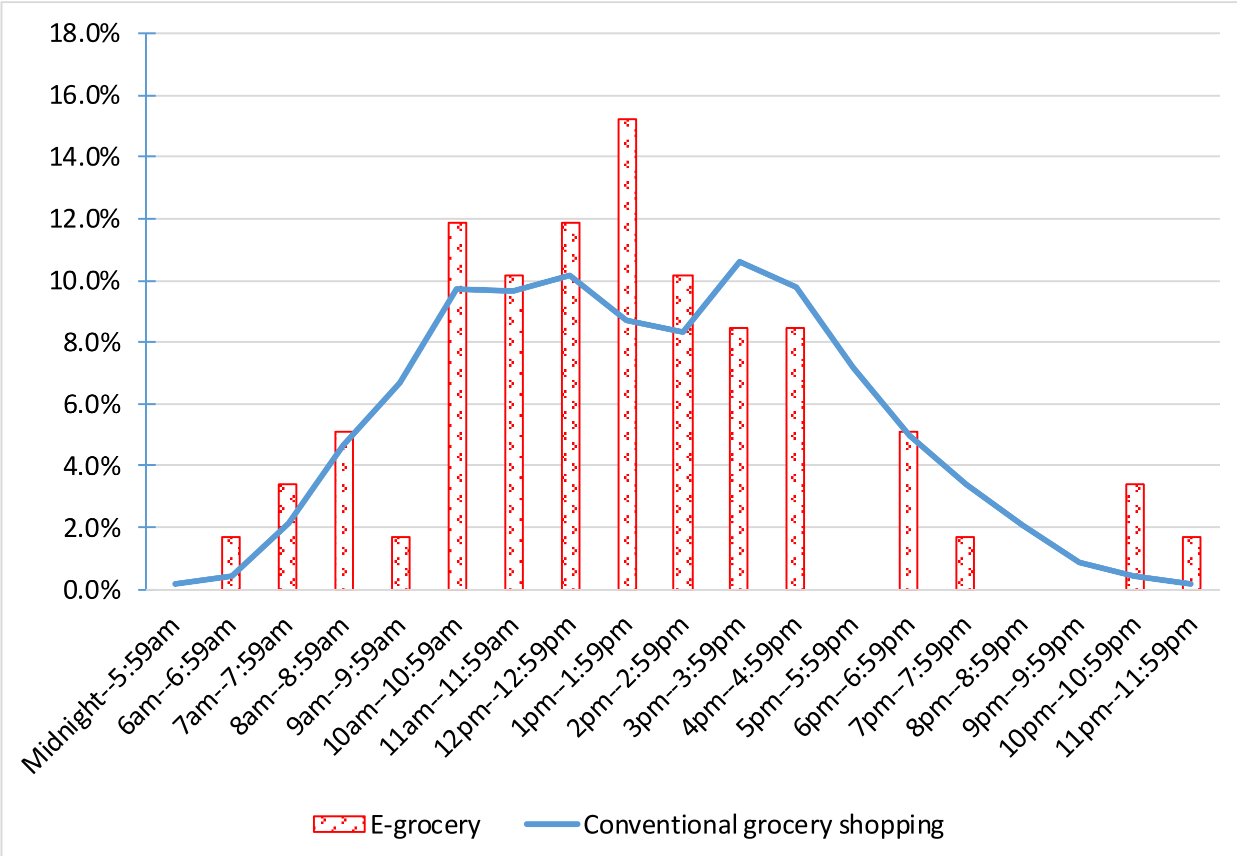


Figure 2.3 Comparison of shopping start times

Notes. 1) Source: 2017 American Time Use Survey. 2) The shopping start time is the time when an ATUS respondent starts looking online for groceries.

Figure 2.3 contrasts the distributions of the time when people start grocery shopping online and in stores. It shows that online shopping activity picks up early afternoon (especially between noon and 2 pm) at a time when conventional grocery shopping tends to subside, and between 10 pm and midnight, when in-store grocery shopping tappers off. These differences highlight the added flexibility and convenience of online grocery shopping.

2.6 Conclusions

In this chapter, I first analyze data from the 2009 and the 2017 National Household Travel Surveys to understand changes in deliveries from online shopping in the US. Results from zero-inflated negative binomial models show that online shopping in the U.S. has been embraced by a much larger percentage of the U.S. population, and that e-shoppers are more varied in 2017 compared to 2009, although differences in the number of deliveries resulting from online shopping are sharper than in 2009. In particular, households with more adult female members receive significantly more deliveries, and so do households with higher incomes and higher educational achievements. Even after controlling for other socio-economic characteristics, I found that minority households are less likely to buy goods online, a disappointing finding that requires further investigations (one possible reason may be more limited access to credit cards). Finally, households with mobility impaired members rely more in 2017 than in 2009 on online shopping to satisfy their needs.

Understanding the determinants of e-shopping is important for supply chain managers so they can adapt the facilities handling online orders, adjust deliveries, and plan product returns. It is also of interest for planners and policymakers concerned with the externalities generated by deliveries of online orders (congestion, traffic accidents, air pollution, and noise) so they can consider appropriate incentives (e.g., subsidies for electric delivery vehicles in denser areas) and regulations (e.g., on the power train technology of delivery vehicles, hours of deliveries, or maximum noise levels).

Since U.S. national household travel surveys do not track what e-shoppers purchase, grocery shopping data from the 2017 ATUS was also analyzed using logit models and non-parametric tests. Consistent with the literature, my results show that in-store grocery shoppers are more likely to be female and unemployed (because homemakers, who are still often female, are considered unemployed), but less likely to belong to younger generations, to have less than a college education, or to be African American, because poor people in predominantly black neighborhoods or in poor rural areas in the U.S. tend to shop at dollar stores and rely more on fast food outlets. By contrast, the only significant socio-economic variable for online grocery shoppers is gender: again, women are more likely to shop for groceries, and the gender gap is larger than for in-store grocery shopping. While this result may be partly due to the small number (only 59) of online grocery shoppers in my sample (N=2,934), the small number of people for groceries online reflects that e-grocery is not common currently in the US: on any given day, people are 24 times more likely to shop for groceries in stores compared to online.

Combining the profile of people who order groceries online with information about households who receive many online orders and data on local stores offering e-grocery could help understand where e-grocery is likely to succeed in the US. Policymakers concerned with access to fresh foods in underserved neighborhoods may then consider subsidizing delivery costs and the creation of grocery packing and delivery jobs. Emergency programs may also be put in place to deliver groceries to groups who cannot go grocery shopping, such as the elderly with impaired mobility or persons at-risk during contagious epidemic diseases (such as COVID-19).

As online shopping becomes even more popular, the need for residential freight deliveries will keep on increasing. However, the magnitude of that increase and its impacts on traffic and the environment depend crucially on how last mile deliveries are organized. If packages are delivered to people’s doorsteps with little coordination, soaring residential freight deliveries will increase congestion, noise, and air pollution, not to mention exacerbate parking shortages in denser urban areas. If, however, last mile deliveries are coordinated, performed as part of existing daily deliveries (e.g., via the U.S. Postal Service), done by bicycle or electric vehicles, and/or go to lockers or neighborhood convenience stores (as is commonly done in Taiwan[5], for example), their external costs could be much reduced (e.g., see Moore, 2019).

The last mile delivery problem is particularly acute for perishable groceries. If local demand is sufficiently high, preferred pricing could foster coordinated local deliveries during specific time windows, which would reduce the need for local freight trips. The widespread adoption of smart home lock systems (see Section 2.2) that allow deliveries to people's fridges when they are not home may also help, although less intrusive solutions such as click-and-pick at grocery stores or deliveries to local convenience stores (where available) may be cheaper and easier to implement.

One limitation of this work is the small number of people who shopped for groceries online in the 2017 ATUS, even though this survey gathered data from a representative sample of over 10,000 Americans (unfortunately only over a single day each), which reflects the current lack of popularity of online grocery shopping in the United States. A second limitation is that my e-grocery dependent variable may have missed some click-and-pick orders.

In future work, it would therefore be of interest to survey U.S. households to gather detailed data about time use and travel, with retrospective questions about online purchases, in order to better capture the links between in-store and online purchases. Until e-grocery becomes more popular in the U.S., researchers could analyze stated preferences of U.S. consumers for e-grocery to understand their preferences for various delivery and price options, and potential obstacles to e-grocery. To better understand the impact of last mile deliveries, it would also be of interest to monitor local freight deliveries in a wide range of neighborhoods. Lastly, I suggest

[5] Personal communication from Professor May Tsai, National Chung Hsing University, May 19, 2019.

analyzing international experiences to learn from creative solutions implemented elsewhere to reduce the external costs of e-grocery (and more generally e-shopping).

2.7 References

Agatz, N., Campbell, A. M., Fleischmann, M., Van Nunen, J., & Savelsbergh, M. (2013). Revenue management opportunities for Internet retailers. *Journal of Revenue and Pricing Management*, 12(2), 128-138.

Andersson, P., & Ivehammar, P. (2019). Benefits and Costs of Autonomous Trucks and Cars. *Journal of Transportation Technologies*, 9(2), 121-145.

Anckar B., Walden P., Jelassi T. (2002). Creating customer value in online grocery shopping. *International Journal of Retail & Distribution Management* 30(4): 211-220.

Beaulac, J., Kristjansson, E., & Cummins, S. (2009). A systematic review of food deserts, 1966-2007. Preventing Chronic Disease, 6(3).

Blumberg, S.J, & Luke, J.V. (2017). Wireless Substitution: Early Release of Estimates from the National Health Interview Survey, July–December 2016. National Center for Health Statistics. Available from https://www.cdc.gov/nchs/data/nhis/earlyrelease/wireless201705.pdf

Brashear, T. G., Kashyap, V., Musante, M. D., & Donthu, N. (2009). A profile of the Internet shopper: Evidence from six countries. *Journal of Marketing Theory and Practice*, 17(3), 267-282.

Bressolles, G., Durrieu, F., & Senecal, S. (2014). A consumer typology based on e-service quality and e-satisfaction. *Journal of Retailing and Consumer Services*, 21(6), 889-896.

Burroughs, R. E., & Sabherwal, R. (2002). Determinants of Retail Electronic Purchasing: A

Multi-Period Investigation1. INFOR: Information Systems and Operational Research, 40(1), 35-56.

Calderwood, E., & Freathy, P. (2014). Consumer mobility in the Scottish isles: The impact of internet adoption upon retail travel patterns. *Transportation Research Part A: Policy and Practice*, 59, 192-203.

Cherrett, T., Dickinson, J., McLeod, F., Sit, J., Bailey, G., & Whittle, G. (2017). Logistics impacts of student online shopping–Evaluating delivery consolidation to halls of residence. *Transportation Research Part C: Emerging Technologies*, 78, 111-128.

Conover, W. J.: Practical Nonparametric Statistics. Wiley, Hoboken, NJ, 3rd edition (1999).

Conway, J. (2020). Share of grocery stores that offered home delivery/store pickup of online orders in the United States from 2017 to 2019. Available from https://www.statista.com/statistics/762746/store-share-online-grocery-delivery-store-pickup-services/#statisticContainer.

Crocco, F., Eboli, L., & Mazzulla, G. (2013). Individual attitudes and shopping mode characteristics affecting the use of e-shopping and related travel. *Transport and Telecommunication*, 14(1), 45-56.

Dost, F., & Maier, E. (2018). E-Commerce Effects on Energy Consumption: A Multi-Year Ecosystem-Level Assessment. *Journal of Industrial Ecology*, 22(4), 799-812.

eMarketer, & Website (retailtechnews.com). (n.d.). E-commerce share of total global retail sales from 2015 to 2021. In Statista - The Statistics Portal. Retrieved May 9, 2019, from https://www.statista.com/statistics/534123/e-commerce-share-of-retail-sales-worldwide/.

Farag, S., Schwanen, T., Dijst, M., & Faber, J. (2007). Shopping online and/or in-store? A structural equation model of the relationships between e-shopping and in-store shopping. *Transportation Research Part A: Policy and Practice*, 41(2), 125-141.

Federal Highway Administration (FHWA) (2011, October). 2009 National Household Travel Survey User's Guide. Version 2. Available from https://nhts.ornl.gov/2009/pub/UsersGuideV2.pdf

Ferrell, C. (2005). Home-based teleshopping and shopping travel: Where do people find the time? *Transportation Research Record: Journal of the Transportation Research Board*, (1926), 212-223.

Freeman M., Freeman A. (2011). Online grocery systems design through task analysis. *Journal of Enterprise Information Management* 24: 440-54.

Ganesh, J., Reynolds, K. E., Luckett, M., & Pomirleanu, N. (2010). Online shopper motivations, and e-store attributes: an examination of online patronage behavior and shopper typologies. *Journal of Retailing*, 86(1), 106-115.

Greene, W.H. (2011). Econometric Analysis (7^{th} edition). Boston, London: Pearson.

Grunert, K.G., Ramus, K. (2005). Consumers' Willingness to Buy Food through the Internet: A Review of the Literature and a Model for Future Research. *British Food Journal*, 107(6): 381-403.

Hand C., Riley F.D., Harris P., Singh J., Rettie R. (2009). Online grocery shopping: The influence of situational factors. *European Journal of Marketing*, 43(9/10): 1205-1219.

Hansen, T. (2005). Consumer Adoption of Online Grocery Buying: A Discriminant Analysis. *International Journal of Retail & Distribution Management*, 33(2): 101-121.

Harris, P., Dall'Olmo Riley, F., Riley, D., & Hand, C. (2017). Online and store patronage: a typology of grocery shoppers. *International Journal of Retail & Distribution Management*, 45(4), 419-445.

Hawkins, A.J. (2018, April 24). Amazon will now deliver packages to the trunk of your car. The Verge. Available from https://www.theverge.com/2018/4/24/17261744/amazon-package-delivery-car-trunk-gm-volvo

InsideRetail Hong Kong. (n.d.). E-commerce share of total retail sales in China from 2014 to 2019. In Statista - The Statistics Portal. Retrieved May 9, 2019, from https://www.statista.com/statistics/379087/e-commerce-share-of-retail-sales-in-china/.

Ioannides, Y.M., Overman, H.G., Rossi-Hansberg, E., and Schmidheiny, K. (2008). The effect of information and communication technologies on urban structure. *Economic Policy*, 23(54): 202–242.

James, P., Arcaya, M. C., Parker, D. M., Tucker-Seeley, R. D., & Subramanian, S. V. (2014). Do minority and poor neighborhoods have higher access to fast-food restaurants in the United States? *Health & Place*, 29, 10-17.

Kamarainen V., Smaros J., Jaakola T., Holmstrom J. (2001). Cost-effectiveness in the e-grocery business. *International Journal of Retail & Distribution Management*, 29: 41-48.

Kang, C., Moon, J., Kim, T., & Choe, Y. (2016, January). Why consumers go to online grocery: Comparing vegetables with grains. In 2016 *49th Hawaii International Conference on System Sciences* (HICSS) (pp. 3604-3613). IEEE.

Klein, R., Neugebauer, M., Ratkovitch, D., & Steinhardt, C. (2019). Differentiated time slot pricing under routing considerations in attended home delivery. *Transportation Science*, 53(1), 236-255.

Kline, R.B. (2010). *Principles and Practice of Structural Equation Modeling*. 3rd edition. The Guilford Press.

Kwan, M. P. (2007). Mobile communications, social networks, and urban travel: Hypertext as a new metaphor for conceptualizing spatial interaction. *The Professional Geographer*, 59(4), 434-446.

Lee, R. J., Sener, I. N., Mokhtarian, P. L., & Handy, S. L. (2017). Relationships between the online and in-store shopping frequency of Davis, California residents. *Transportation Research Part A: Policy and Practice*, 100, 40-52.

Li, Y., Sun, Y., Zhang, Y., & Hu, X. (2018). Online grocery retailing for fresh products with order cancellation and refund options. Journal of Supply Chain and Operations Management, 16(1), 1-16.

Lien, T. (2017). Apoorva Mehta had 20 failed start-ups before Instacart. Los Angeles Times (January 27). Available from https://www.latimes.com/business/technology/la-fi-himi-apoorva-mehta-20170105-story.html

Long, J.S. (1997). Regression Models for Categorical and Limited Dependent Variables. Sage Publications, Inc.

Marchet, G., Melacini, M., Perotti, S., Rasini, M., & Tappia, E. (2018). Business logistics models in omni-channel: a classification framework and empirical analysis. *International Journal of Physical Distribution & Logistics Management*, 48(4): 439-464.

McDonald, R., Christensen, C., Yang, R., Hollingsworth, T. (2014). Amazon Fresh: Rekindling the Online Grocery Market, *Harvard Business School Case*, 615-013.

McGuckin, N., & Fucci, A. (2018). Summary of Travel Trends: 2017 National Household Travel Survey. FHWA Report No. FHWA-PL-18-019. Available from https://nhts.ornl.gov/assets/2017_nhts_summary_travel_trends.pdf

Melis, K., Campo, K., Breugelmans, E., & Lamey, L. (2015). The impact of the multi-channel retail mix on online store choice: does online experience matter? *Journal of Retailing*, 91(2), 272-288.

Meyerson, N. (2019). How Wegmans and H-E-B survived Amazon's onslaught. CNN Business, November 13. Available from https://www.cnn.com/2019/11/13/business/heb-wegmans-publix-hy-vee-regional-grocery/index.html.

Meyerson, N. (2020). This grocery store wanted to compete with Whole Foods. Now it's going bankrupt. CNN Business, January 28. Available from https://www.cnn.com/2020/01/28/business/luckys-market-bankruptcy-aldi-publix-kroger-trnd/index.html.

Moore, A. M. (2019). Innovative scenarios for modeling intra-city freight delivery. *Transportation Research Interdisciplinary Perspectives*, 100024. Available from https://www.sciencedirect.com/science/article/pii/S2590198219300247.

Morganosky, M.A., Cude, B.J. (2000). Consumer response to online grocery shopping. International *Journal of Retail & Distribution Management*, 28(1), 17-26.

Morganosky, M.A., & Cude, B.J. (2002). Consumer demand for online food retailing: is it really a supply side issue? *International Journal of Retail & Distribution Management*, 30(10), 451-458.

Morland, K., Wing, S., Roux, A. D., & Poole, C. (2002). Neighborhood characteristics associated with the location of food stores and food service places. *American Journal of Preventive Medicine*, 22(1), 23-29.

Murphy, A. J. (2003). (Re)solving space and time: fulfilment issues in online grocery retailing. *Environment and Planning A*, 35(7), 1173-1200.

Nahiduzzaman, K. M., Aldosary, A. S., & Mohammed, I. (2019). Framework Analysis of E-Commerce Induced Shift in the Spatial Structure of a City. *Journal of Urban Planning and Development*, 145(3), 04019006.

Nelson, P. (1970). Information and Consumer Behavior. *Journal of Political Economy*, 78(2): 311-329.

Office for National Statistics (2018). Internet Sales as a Percentage of Total Retail Sales. Available from https://www.ons.gov.uk/businessindustryandtrade/retailindustry/timeseries/j4mc/drsi

Parment, A. (2013). Generation Y vs. Baby Boomers: Shopping behavior, buyer involvement and implications for retailing. *Journal of Retailing & Consumer Services*, 20(2), 189-199.

Perboli, G., & Rosano, M. (2019). Parcel delivery in urban areas: opportunities and threats for the mix of traditional and green business models. *Transportation Research Part C: Emerging Technologies*, 99, 19-36.

Pettersson, F., Winslott Hiselius, L., & Koglin, T. (2018). E-commerce and urban planning– comparing knowledge claims in research and planning practice. *Urban, Planning and Transport Research*, 6(1), 1-21.

Pew Research Center (2018). Defining generations: Where Millennials end and post-Millennials begin. Available from http://www.pewresearch.org/fact-tank/2018/03/01/defining-generations-where-millennials-end-and-post-millennials-begin/

Punakivi, M. and Saranen, J. (2001). Identifying the success factors in e-grocery home delivery. International Journal of Retail & Distribution Management, 29(4): 156-163.

Redman, R. (2018, August 13). Gallup: Most Americans just aren't into online grocery. Supermarket News. Available from: https://www.supermarketnews.com/consumer-trends/gallup-most-americans-just-aren-t-online-grocery

Ren, F., & Kwan, M. P. (2009). The impact of geographic context on e-shopping behavior. *Environment and Planning B: Planning and Design*, 36(2), 262-278.

Robinson, H., Dall'Olmo, R., Rettie, R., Rolls-Willson, G. (2007). The Role of Situational Variables in Online Grocery Shopping in the UK, *The Marketing Review* 7(1): 89-106.

Saphores, J. D., & Xu, L. (2021). E-shopping changes and the state of E-grocery shopping in the US-Evidence from national travel and time use surveys. *Research in Transportation Economics*, 87, 100864.

Saunders, N. (2018, October 17). Online Grocery: Lessons from History [Blog post]. Retrieved from https://www.onespace.com/blog/2018/10/online-grocery-lessons-history/

Seaman C.E.A. (1995). Food on the Internet. *Nutrition and Food Science*, 95: 26-9.

Smith, A., & Anderson, M. (2016, December). Online Shopping and E-Commerce. Pew Research Center. Available from http://www.pewinternet.org/2016/12/19/online-shopping-and-e-commerce/

Suel, E., & Polak, J. W. (2017). Development of joint models for channel, store, and travel mode choice: Grocery shopping in London. *Transportation Research Part A: Policy and*

Practice, 99, 147-162.

Suel, E., & Polak, J. W. (2018). Incorporating online shopping into travel demand modelling: challenges, progress, and opportunities. *Transport Reviews*, 38(5), 576-601.

Thompson, D. (2017). What in the World Is Causing the Retail Meltdown of 2017? The Atlantic, April 10. Available from https://www.theatlantic.com/business/archive/2017/04/retail-meltdown-of-2017/522384/.

US Census Bureau News (2019, March 13). Quarterly Retail E-Commerce Sales, 4th Quarter 2018. Available from https://www.census.gov/retail/mrts/www/data/pdf/ec_current.pdf

United States Census Bureau (2018a). Historical Table 5. Estimated Annual Sales for US Electronic Shopping and Mail-Order Houses (NAICS 45411) - Total and E-commerce Sales by Merchandise Line: 1999-2016. Available from https://www.census.gov/data/tables/2016/econ/e-stats/2016-e-stats.html

United States Census Bureau (2018b). Historical Table 4. Estimated Annual US Retail Trade Sales - Total and E-commerce: 1998-2016. Available from https://www.census.gov/data/tables/2016/econ/e-stats/2016-e-stats.html

Valentini, S., Montaguti, E., & Neslin, S.A. (2011). Decision process evolution in customer channel choice. *Journal of Marketing*, 75(6), 72-86.

Vroman, K. G., Arthanat, S., & Lysack, C. (2015). Who over 65 is online? Older adults' dispositions toward information communication technology. *Computers in Human Behavior*, 43, 156-166.

Wadud, Z. (2017). Fully automated vehicles: A cost of ownership analysis to inform early adoption. *Transportation Research Part A: Policy and Practice*, 101, 163-176.

Wang, R. J. H., Malthouse, E. C., & Krishnamurthi, L. (2015). On the go: How mobile shopping affects customer purchase behavior. *Journal of Retailing*, 91(2): 217-234.

Wang, X. C., & Zhou, Y. (2015). Deliveries to residential units: A rising form of freight transportation in the US. *Transportation Research Part C: Emerging Technologies*, 58, 46-55.

Wat, F.K.T., Ngai, E.W.T. and Cheng, T.C.E. (2005). Potential risks to e-commerce development using exploratory factor analysis, International Journal of Services Technology and Management, 6(1): 55-71.

Weathers, D., Sharma, S.L, & Wood, S. (2007). Effects of online communication practices on consumer perceptions of performance uncertainty for search and experience goods. *Journal of Retailing*, 83(4): 393-401.

Whalen, A. (2018). Dollar Stores Planning for Permanent American Underclass, Sell More Groceries Than Whole Foods. Newsweek, 12/07/18. Available from https://www.newsweek.com/dollar-stores-general-tree-america-economy-poor-people-food-poverty-lower-1250186.

White, G.K. (1997). Using the WWW for retail food sales: Discussion and demonstration. *Journal of Food Distribution Research*, 28: 26-29.

Wollenburg, J., Hübner, A., Kuhn, H., & Trautrims, A. (2018). From bricks-and-mortar to bricks-and-clicks. International Journal of Physical Distribution & Logistics Management 48,4(2018),415-438.

Wollerton, M. (2018, May 25). Amazon Key took over my door for 3 months. It wasn't as creepy as I expected. CNET. Available from https://www.cnet.com/news/amazon-key-took-over-my-door-for-3-months-it-wasnt-as-creepy-as-i-expected/

Zhou, Y., & Wang, X. C. (2014). Explore the relationship between online shopping and shopping trips: an analysis with the 2009 NHTS data. *Transportation Research Part A: Policy and Practice*, 70, 1-9.

Chapter 3. I E-shop, and therefore I Travel More? Evidence from the 2017 U.S. National Household Travel Survey

3.1 Introduction

Does more online shopping (e-shopping) increase household travel? This question should be of interest to policymakers concerned with urban congestion in the United States where shopping was responsible for approximately 12.6% of all travel person miles in 2017 (McGuckin and Fucci, 2018). It is especially salient in jurisdictions that have adopted targets to reduce vehicle miles traveled (VMT) in order to reduce their emissions of air pollutants and greenhouse gases (GHG), and to achieve various co-benefits (American Council for an Energy-Efficient Economy, 2020; Litman, 2022). California, for example, is aiming for carbon neutrality in its transportation sector by 2045 (California Environmental Protection Agency, 2020). Despite the adoption in June 2020 of a landmark mandate for fostering the adoption of zero emission vehicles (Shepardson and Groom, 2020), other measures (such as VMT reduction) will likely be necessary if California is to meet its GHG reduction targets. Other states will face similar problems when the United States rejoins the Paris agreement on climate change.

Recent years have seen the rapid expansion of e-commerce around the world. In the United States, estimated annual e-commerce retail trade sales soared from $27.6 billion to $389.1 billion between 2011 and 2016 (United States Census Bureau, 2020). The Covid-19 pandemic further accelerated this trend (World Trade Organization, 2020), and it likely altered the relationship between travel and e-shopping due to travel and in-store shopping restrictions.

While a number of papers have examined whether online shopping decreases household travel and especially VMT (Cao *et al.*, 2012; Jaller and Pahwa, 2020; Lee *et al*., 2017; Shi *et al.*, 2019; Zhou and Wang, 2014), the literature is still inconclusive, partly because of the complexity of this question, and the changing relationship between shopping and travel. Another reason is the wide range of approaches that have been used by researchers to examine the relationship between e-shopping, instore shopping, and travel, which include structural equation modeling (SEM) (Cao *et al*., 2012; Ding and Lu, 2017; Farag *et al*., 2007; Ferrell, 2005; Xi *et al*., 2018; Zhou and Wang, 2014), discrete and continuous regression models (Cao *et al*., 2010a; Farag *et al*., 2003; Ferrell, 2004; Lee *et al*., 2017; Maat and Konings, 2018; Shi *et al*., 2019), and hazard-based duration models (Bhat *et al*., 2003). It is found that most published studies did not account for self-selection, which likely biased their results. Self-selection here means that the characteristics of households who are heavy e-shoppers may differ systematically from those who prefer shopping in brick-and-mortar stores.

In that context, the purpose of this study is to get an understanding of the impact of e-shopping on household travel in the United States by analyzing household-level travel data from the 2017 NHTS (Federal Highway Administration, 2017). To control for self-selection, I relied on propensity score matching (PSM), a well-known method that has so far been under-used in transportation, and controlled for household socio-demographic characteristics, land use around their dwellings, and their vehicle holdings. I also examined separately the relationship between e-shopping and travel for different population density bands, since population density is known to affect both shopping opportunities and the availability of various transportation modes, and therefore the way people travel. Finally, unlike other papers, households are adopted as the unit of analysis because shopping and travel within the household are usually interrelated.

This chapter is organized as follows. In Section 3.2, I review selected studies concerned with the impact of e-shopping on travel, and transportation applications of propensity score matching. Then I introduce the utilized methodology and data in Section 3.3 including all the variables. After that, the model results will be discussed in detail in Section 3.4. Finally, I will summarize my key findings, outline the limitations of this study, and introduce avenues for future work in Section 3.5.

3.2 Background and Literature Review

The impact of online shopping on travel has been receiving increasing attention from researchers and policymakers with the steady rise of e-commerce (United States Census Bureau, 2020) and growing concerns about the impact of transportation on greenhouse gas emissions and air pollution. After reviewing selected papers dealing with the impact of e-shopping on travel to brick-and-mortar stores, I discuss the possible influence of a variety of factors, before summarizing selected applications of PSM to transportation issues.

3.2.1 E-shopping and travel to brick-and-mortar stores

One may expect that customers who are increasingly relying on internet purchases will decrease their patronage of brick-and-mortar stores and as a result reduce their shopping trips. However, the relationship between e-shopping, in-store shopping, and travel is more complex (Dias *et al*., 2020; Lavieri *et al*., 2018; Mokhtarian, 2009, 2004). On one hand, e-shopping could substitute for in-store shopping when travel is costly, difficult, or dangerous (Mokhtarian, 2009). On the other hand, customers may want to physically experience the products they are considering, or travel to stores to pick up online orders and return unwanted items (Cao *et al*., 2012; Ding and

Lu, 2017; Zhai *et al*., 2019). Empirical studies are therefore necessary to assess the net impact of increased e-shopping on travel (Mokhtarian, 2002; Salomon, 1986).

Early studies found that online and store shopping are not perfect substitutes. For example, Visser and Lanzendorf reported that more e-shopping can lead to an overall increase in travel for both individuals and for freight transport (Visser and Lanzendorf, 2003). After analyzing data from the Metropolitan Transportation Commission in California, Ferrell concluded that teleshopping households lead to more shopping trips (Ferrell, 2004), although in a follow-up paper, Ferrell reported that people who substituted home teleshopping for in-store shopping made fewer and shorter shopping trips (Ferrell, 2005). In contrast, Cao found that most of the empirical studies he reviewed show complementarity effects between e-shopping and store shopping frequency (Cao, 2009), a conclusion that was reinforced after the analysis of an online survey of shoppers in the Minneapolis-St. Paul metropolitan area (Cao *et al*., 2010a). Likewise, Zhou and Wang (2014) reported that e-shopping encourages in-store shopping based on their analysis of 2009 NHTS data, a conclusion also supported by Lee *et al.* (2017), but for the former more shopping in brick-and-mortar stores decreases the propensity to shop online (Zhou and Wang, 2014). This intricate complementarity and substitution effects between online and in-store shopping has led some researchers to jointly consider online and in-store shopping to analyze travel behavior and improve travel forecasting (Dias *et al*., 2020; Lavieri *et al*., 2018).

Another motivation for studying the relationship between online and in-store shopping has been to quantify the environmental impacts of more e-shopping, especially as they relate to air pollution and greenhouse gas emissions. For example, Goodchild *et al.* (2018) developed a theoretical model to estimate VMT and CO_2 emissions from household shopping trips and home delivery vehicles. They showed that in-store shopping is better for the environment when household density and the emission ratio (i.e., the relative emissions of a delivery vehicle compared to a personal vehicle) are relatively low, while the reverse holds when they are relatively high (Goodchild *et al*., 2018). Emissions from cooling systems may also make a difference, especially for perishable foods (Heldt *et al*., 2019). Moreover, the choice of delivery vehicles, basket size, the ability to consolidate vehicles, and the time of delivery (e.g., during congested times) also matter (Jaller and Pahwa, 2020).

3.2.2 Factors affecting the relationship between online and in-store shopping

The relationship between online and in-store shopping is affected by multiple factors. For example, online searching is an essential component in the hybrid shopping process, and it was found to have a positive impact on both online and store in-shopping (Cao, 2012). The popularity of e-shopping also depends on convenience, shops accessibility, and cost (Cao, 2009; Farag *et al*., 2006b), and changes in shopping habits may have wide-ranging consequences. For example, same-day deliveries may change shopping in local stores and alter the distribution of commercial land use over time (Xi *et al*., 2020). Therefore, focusing only on in-store versus online shopping on aggregate may mask important differences.

The impact of e-shopping on travel also depends on the type of products sold and on the characteristics of internet shoppers (Hoogendoorn-Lanser *et al*., 2019; Suel *et al*., 2015; Weltevreden, 2007). For example, clothing purchases are more likely to generate store visits than books purchases, simply because many customers would like to try on clothing or shoes they are considering buying, whereas reading reviews of books or electronic devices online may provide sufficient information to prospective buyers (Schmid and Axhausen, 2019; Zhai *et al*., 2017; Zhen *et al*., 2016).

To understand the enduring attraction of in-store shopping, some researchers have argued that it could fulfill some psychological and recreational functions besides obtaining information (Lee *et al*., 2017; Salomon and Koppelman, 1988). Haridasan and Fernando (2018) also pointed out that in-store shoppers tend to value social interactions and personalized attention while online shoppers typically seek product variety, value for money, and the convenience of home deliveries (Haridasan and Fernando, 2018). This strand of the literature argues that understanding shopping behavior requires recognizing that the motivations of shoppers are often complex and may involve a leisure component (Farag *et al*., 2007).

Local preferences also matter. For groceries, for example, Americans were more likely to shop in stores prior to Covid-19 (Saphores and Xu, 2020), while online shopping for groceries (e-grocery) is more popular in many parts of Europe, and notably in London (Suel and Polak, 2017). In countries such as Iran and Netherlands, e-shopping appears to motivate residents to visit stores more frequently, creating a complementary relationship between in-store and e-shopping frequency (Etminani-Ghasrodashti and Hamidi, 2020; Farag *et al*., 2006a). For some cities in China, Shi *et al*. (2019) concluded that in Chengdu e-shopping reduces the frequency of in-store shopping trips, while in Nanjing, Xi *et al*. (2018) reported that e-shopping has a neutral effect on in-store shopping, and that more shopping at brick-and-mortar stores increases the frequency of e-shopping (Shi *et al*., 2019; Xi *et al*., 2018).

The time saved from shopping online may be redirected toward other activities. For example, in Stockholm (Sweden), researchers have found that the time that frequent online shoppers save by shopping online is spent on additional shopping trips but also on trips for other purposes (Hiselius *et al*., 2015).

Finally, it is found that shopping behavior has been changing over time. In an early study, Sim and Koi concluded that e-commerce does not strongly affect traditional retail because Singaporeans prefer in-store shopping (Sim and Koi, 2002). This is no longer the case. According to a 2019 survey, 67% of Singapore shoppers have experience with "*webrooming*" - browsing items online and then purchasing them in-store, and 63% with "*showrooming*" - looking at products in-store and then purchasing them online (Hirschmann, 2019). For some product categories, including for example digital entertainment (e.g., movies, music, and video games), online purchases in Singapore dwarf retail sales (Sam and Sharma, 2015).

3.2.3 Controlling for self-selection

In addition to heterogeneity and temporal change, the diversity of results reported in the literature may also be partly explained by the methods used to analyze the relationship between e-shopping and travel, and particularly the treatment of self-selection.

Since empirical studies of the relationship between e-shopping, in-store shopping, and travel are observational (randomized experiments are impractical in this context), one key challenge when trying to quantify the impact of a "treatment" (here, a high level of e-shopping) on a measure of travel with observational data, is that respondents may self-select. As a result, the characteristics of respondents with the treatment would differ systematically from the characteristics of the respondents without, so a direct measurement of the impact of the treatment would likely be biased. To overcome this problem, researchers have employed a variety of strategies including (but not limited to) statistical control, instrumental variables, Heckman's sample selection method, propensity score matching, joint discrete choice models, structural equations modeling, and longitudinal designs (Cao *et al*., 2009).

Here, I rely on propensity score matching (PSM) (Rosenbaum and Rubin, 1983), because its relative simplicity and its effectiveness have led to its widespread use in a variety of fields to minimize the confounding effects of observed covariates (Austin *et al*., 2018). In transportation, Mishra *et al.* analyzed the effect of carsharing on vehicle holdings and travel behavior, Cao *et al*. investigated the association between the built environment and travel, and Cao and Fan analyzed the impact of population density on travel (Cao *et al*., 2010b; Cao and Fan, 2012; Mishra *et al*., 2015). However, to the best of my knowledge, my study is the first to apply PSM to the analysis of the impact of e-shopping on household travel. My goal here is to quantify shifts in simple measures of travel associated with various degrees of e-shopping while controlling for self-selection using an extensive list of household characteristics, land use, and vehicle holding variables.

3.3 Methodology and Data

3.3.1 Methodology

To implement PSM, I split the sample into two groups of households: a treatment group, where households receive a "high" (defined in "Treatment variables" below) number of deliveries per person per month from e-shopping, and a control group, made up of other households. Then each household in the treatment group is matched with almost identical households in the control group based on socio-economic characteristics, residential land use, and car ownership variables, which are known to impact travel. This procedure is similar in spirit to experiments on pairs of identical subjects, where one randomly receives a treatment and the other a placebo to quantify the impact of the treatment.

Let $y_{1,i}$ denote my dependent variable (a measure of household travel; see below) for household i who is observed in the group with the treatment (here, a high level of e-shopping). Conversely, let $y_{0,j}$ denote the dependent variable for household j observed in the group without the treatment. To assess the impact of a high level of e-shopping on a measure of travel, I would like to calculate $y_{1,i} - y_{0,i}$ for all households in the sample, but this is not possible since only $y_{1,i}$ or $y_{0,i}$ (but not both) is observed. Comparing $y_{1,i}$ and $y_{0,i}$ can therefore be construed as a missing data problem (Rosenbaum and Rubin, 1983), so I treat $y_{k,i}$ ($k \in \{0,1\}$) as an outcome from random variable $Y_{k,i}$.

Let us decompose the expected difference in travel with and without treatment $\Delta\mu$,

$$\Delta\mu = E\left[Y_{1,i} \mid \text{eshop}_i = 1\right] - E\left[Y_{0,i} \mid \text{eshop}_i = 0\right], \tag{1}$$

into a causal effect and a self-selection bias component:

$$\begin{aligned}\Delta\mu = {} & E\left[Y_{1,i} \mid \text{eshop}_i = 1\right] - E\left[Y_{1,i} \mid \text{eshop}_i = 0\right] \\ & + E\left[Y_{1,i} \mid \text{eshop}_i = 0\right] - E\left[Y_{0,i} \mid \text{eshop}_i = 0\right].\end{aligned} \tag{2}$$

In Equations (1) and (2), eshop_i is a binary variable that equals 1 is household i shops online a lot (the treatment here), and 0 otherwise; and the expectation is taken over households indexed by i.

The first difference on the right side of Equation (2) calculates the average causal effect of a high level of e-shopping on the dependent variable for households observed in the treatment group; it is called the average treatment effect on the treated (ATET). The second difference on the right side of Equation (2) is the self-selection bias calculated as the difference between the expected travel of households in the treatment group if they had not received the treatment and the expected travel of households observed in the control group.

If the self-selection bias is zero, $\Delta\mu$ equals the ATET, but for observational data this is unlikely. Indeed, households in the treatment group could have different socio-economic and demographic characteristics compared to households in the control group, land use around their residence may differ, and so could the number of motor vehicles at their disposal. It is also possible that households in the control group do not like to shop online, although much of the travel behavior literature assumes that attitudes and tastes can be explained by socio-economic and demographic characteristics.

To minimize the self-selection bias, the PSM algorithm matches every household in the control group with one or more households in the treatment group based on covariates known to explain travel behavior (socio-economic, demographic, and land use variables as well as vehicle ownership), denoted here by $\mathbf{X}_i$. The goal is to shape the joint distribution of covariates in the control group so that it matches closely their distribution in the treatment group. If the matching is thorough and comprehensive, the outcome variables ($Y_{k,i}$, $k \in \{0,1\}$) become orthogonal to membership in the treatment and control groups conditional on covariates X_i, the bias becomes zero, and $E\left[Y_{k,i} \mid \mathbf{X}_i, \text{eshop}_i = l\right] = E\left[Y_{k,i} \mid \mathbf{X}_i\right]$ (Rosenbaum and Rubin, 1983). As a result, the observed difference $\Delta\mu$ between the treated and control groups in the travel variable considered gives an unbiased estimate of the impact of the treatment (here a high level of e-shopping).

While matching can be done in multiple ways, households are matched based on the propensity score $e_i(\mathbf{X}_i) = \Pr(\text{eshop}_i = 1 \mid \mathbf{X}_i)$, which represents the probability (estimated via a logit model here) that household *i* shops online a lot and receives a high number of deliveries per household member conditional on the control variables $\mathbf{X}_i$. This approach hinges on Rosenbaum and Rubin's (1983) proof that matching based on the probability of treatment conditional on all relevant observed covariates $\mathbf{X}_i$ is sufficient for obtaining an unbiased estimate of a treatment on an outcome variable.

Like Mishra *et al*. (2015), however, I note that PSM can only remove the self-selection bias arising from observed covariates $\mathbf{X}_i$, but not from unobserved variables such as attitudes toward e-shopping, so I cannot exclude the risk of residual omitted variable bias (Mishra *et al*., 2015).

To estimate the potential impacts of e-shopping on various measures of travel using PSM via ATET, I relied on the 'teffects psmatch' command in Stata 15, which implements the Abadie and Imbens (2016) correction of the variance of PSM estimators. This is particularly important for ATET estimates ignoring the estimation error in the propensity score which may result in biased confidence intervals for $\Delta\mu$ (Abadie and Imbens, 2016).

To assess whether propensity score model was correctly specified, several tools are used. First, I compared the means of the matching variables of control and treatments groups using standardized mean differences such as

$$d_x = \frac{\overline{x}_{treatment} - \overline{x}_{control}}{\sqrt{\frac{s^2_{treatment} + s^2_{control}}{2}}}, \tag{3}$$

where the numerator is a difference of sample means and the denominator is the pooled samples standard deviation. Austin (2009) recommended a maximum value of 0.1 for binary variables (Austin, 2009a, 2009b).

Second, to assess if the matching process worked well I calculated variance ratios of the matching variables, in which the ratio of the variances of the propensity score in the two groups should be close to one for a model to be well-specified (Rubin, 2001).

Third, I also relied on the criteria proposed in Rubin (2001) to assess the mean bias and adequacy of matches (good, of concern, or bad). Finally, I compared visually the density plots of each control variable respectively in the treatment and control samples to check to what extent they differ.

3.3.2 Data

To understand the impact of e-shopping on travel, I analyzed data from the 2017 NHTS. The NHTS (known as the National Personal Transportation Survey before 2001) has been conducted at irregular intervals (the last three surveys were in 2001, 2009, and 2017) by the Federal Highway Administration (FHWA) to provide a comprehensive picture of travel trends in the United States and answer travel-related questions of interest to add-on states (i.e., states who helped fund the NHTS in exchange for larger samples and additional questions for their residents).

To improve coverage, potential respondents of the 2017 NHTS were selected using an address-based sampling strategy. Random digit dialing of landline phone numbers, which was used in the 2009 NHTS, is no longer appropriate because since 2016 over half of American homes have abandoned landlines in favor of cell phones (Blumberg and Luke, 2017). To increase the response rate, 2017 NHTS data were retrieved using a self-completed web questionnaire.

Data from the 2017 NHTS are organized in four files (households, persons, trips, and vehicles). These files contain information from 129,696 households corresponding to 264,234 individuals who undertook 923,572 trips in 256,115 vehicles on their assigned survey day.

In this study, households are chosen as units of analysis because household members may purchase items online for each other (e.g., adults will order for their children) (Aguiléra *et al.*, 2012). The NHTS distinguishes between two trip destinations for in-store shopping: buying goods (groceries, clothes, appliances, gas), and buying meals (go out for a meal, snack, carry-out).

Treatment variables

In this study, I focused on responses to the NHTS question “In the past 30 days, how many times did you purchase something online and have it delivered?”. Answers to this question provide the basis for the treatment variable (i.e., if households are very active e-shoppers or not). I divided households into three groups based on their monthly frequency of online purchases per household member: low frequency online shoppers (the baseline group with up to one purchase per month per household member), medium frequency online shoppers (the first treatment group with more than one and up to four online purchase per month per household member), and the high frequency online shoppers (the second treatment group with over four online purchases per month per household member). In each of the PSM models, I contrasted a measure of travel between one of the treatment groups and its control group. Since household travel typically varies substantially between weekdays and weekends (e.g., due to commuting), separately weekdays and weekends travel are analyzed.

Dependent variables

To contrast the travel behavior of households who actively shop online and those who do not, I created and analyzed response variables that reflect the number of activities/trips and the distance traveled on the NHTS survey day for different trip purposes (see Table 3.1).

Table 3.1 Description of response variables and treatment indicators

Variables	Description
Response Variables (household level)	
1) Number of activities for each trip purpose (lower bound)	Count of activities per household where activities with more than one household member count as one
2) Number of activities for each trip purpose (upper bound)	Count of activities summed over all members of each household participating in a trip
3) Trip miles for each trip purpose	Sum over household members of all miles traveled (all modes)
4) Vehicle miles traveled (VMT)	Sum over household members of all vehicle miles traveled
5) Shopping activities % (lower bound)	Percentage of household activities related to shopping where activities with more than one household member count as one
6) Shopping activities % (upper bound)	Percentage of household activities related to shopping where activities are summed over all household members in each trip
7) Shopping mile %	Percentage of total household miles for shopping
8) Shopping VMT %	Percentage of total household vehicle miles for shopping
Trip Purpose	Trip/activity relates to…
Shop	buying groceries, clothes, appliances, and gas
Buy meals	going out for a meal, snack, or carry-out
Exercise	going for a jog/walk, walking the dog, and going to the gym
Health care visit	medical, dental, and therapy visits
Buy services	dry cleaners, banking, car servicing, and pet care
Recreation	visiting parks, going to movie theaters, bars, or museums
Visit friends	visiting friends and relatives
Religious & community visit	religious and other community activities
Regular home activities	home chores and sleep
Work and work-related	working and work meetings
Work from home	working from home with payments
Volunteer activity	volunteer (unpaid) activity
Transport-related	dropping-off/picking up someone and/or changing mode
Attend school/child/adult care	attending school, child care, or adult care
Other errands	running general errands (e.g., post office or library)
Something else	trip/activity not captured by the variables above
Treatment Indicators	
Low frequency online shoppers	≤ 1 online purchase for delivery per household member per month (our control group)
Medium frequency online shoppers	(1, 4] online purchases for delivery per household member per month
High frequency online shoppers	> 4 online purchases for delivery per household member per month

First, I counted the number of trips for different trip purposes, as well as trip start time and end time. Real life complexity and NHTS data limitations led us in some cases to define two variables for a given trip purpose when two or more household members travel together to the same destination for the same purpose. This is the case for example when two household members travel together to the same mall. If they shopped together, that trip should count as one. But if they visited different stores, it could count as two shopping trips. Since this information is unavailable in the NHTS dataset, a lower and an upper bound for the number of trips for different trip purposes are defined.

To analyze the impact of e-shopping on distance traveled, four dependent variables related to household trip miles are generated: “miles traveled”, “VMT”, “shopping trip miles”, and “shopping VMT”. To avoid duplication, I assumed that if two or more household members share the same start and end time with the same trip purpose and the same trip mode, it counts as only one trip for trip distance. “Miles traveled” sums up all trip miles from any mode of transport for a household on their survey day. Likewise, “shopping trip miles” sums trip miles related to shopping using any transportation mode. “VMT” and “shopping VMT” are similar but account only for distance traveled in a motorized vehicle.

Since it could be insightful to look at travel in relative terms, I also examined the impact of e-shopping on the percentage of trips for shopping activities (with lower and upper bounds), on shopping miles percentage, and on shopping VMT percentage. Finally, for completeness, I also tracked miles traveled for other trip purposes.

Control variables

For control variables, I relied on a rich set of household socioeconomic and demographic characteristics that are known to influence travel (Xiao *et al*., 2018). Summary statistics for control variables are presented in Table 3.2 and Table 3.3.

Table 3.2 Summary statistics for binary and categorical control variables

	Number of monthly e-purchases per household member					
	≤ 1		>1 to 4		> 4	
	weekday	weekend	weekday	weekday	weekend	weekday
Number of observations	42,492	10,873	29,131	7,912	14,502	4,120
Household composition						
1 adult without kids	18.55%	17.83%	14.89%	15.03%	22.44%	20.95%
2+ adults without kids	16.86%	16.69%	26.04%	24.96%	32.04%	32.16%
1 adult with kids	4.65%	4.27%	3.35%	2.79%	2.04%	1.82%
2 adults with kids	17.05%	17.37%	25.50%	26.47%	16.19%	17.84%
1 adult, retired without kids	17.34%	18.37%	7.16%	6.56%	7.29%	7.14%
2+ adults, retired without kids	25.56%	25.47%	23.05%	24.19%	20.00%	20.10%
Household head generation						
Z & Y (born after 1981)	12.12%	11.75%	18.59%	18.10%	22.95%	23.86%
X (born 1965 to 1980)	19.57%	19.08%	26.38%	27.57%	25.70%	25.73%
Baby boomers (born 1946 to 1964)	45.02%	44.78%	44.71%	43.67%	43.63%	43.11%
Silent (born before 1946)	23.29%	24.38%	10.32%	10.67%	7.72%	7.31%
Household income quintiles						
1st: <$25,000	27.00%	26.54%	8.99%	8.05%	6.37%	6.50%
2nd: $25,000 to $49,999	27.40%	26.82%	17.81%	17.56%	13.76%	13.47%
3rd: $50,000 to $74,999	18.27%	18.38%	18.92%	18.95%	16.91%	15.87%
4th: $75,000 to $124,999	18.61%	18.82%	30.08%	30.11%	30.53%	30.39%
5th: ≥$125,000	8.71%	9.45%	24.20%	25.34%	32.44%	33.76%
Highest educational attainment in the household						
High school or less	22.47%	21.35%	6.53%	6.18%	4.35%	4.13%
Some college/associate	32.67%	33.26%	24.50%	23.80%	19.87%	19.81%
Bachelor’s degree	23.10%	22.75%	30.57%	30.27%	31.51%	31.87%
Graduate/professional degree	21.76%	22.64%	38.41%	39.75%	44.27%	44.20%
Household ethnicity						
White	79.46%	81.03%	86.47%	86.22%	88.91%	87.84%
Black or African American	11.11%	9.02%	5.05%	4.52%	3.30%	3.06%
Asian	3.66%	3.83%	3.92%	4.35%	3.69%	4.10%
Multiple or other	5.78%	6.13%	4.56%	4.90%	4.10%	5.00%
Household head is Hispanic/Latino	8.41%	8.00%	5.89%	5.64%	4.75%	5.15%
Fewer vehicles than drivers	9.22%	9.46%	7.72%	7.87%	6.65%	8.16%
Population Density in census tract of residence						
<500 people/mi^2	33.61%	32.66%	30.75%	29.90%	28.38%	27.77%
500 to 4,000 people/mi^2	42.14%	40.72%	43.88%	42.16%	41.90%	40.36%
>4,000 people/mi^2	24.25%	26.63%	25.37%	27.93%	29.72%	31.87%

Households are categorized based on the number of monthly e-purchases per household member.

Table 3.3 Summary statistics for count control variables

		Number of monthly e-purchases per household member					
		≤ 1		>1 to 4		> 4	
		weekday	weekend	weekday	weekend	weekday	weekend
Number of observations		42,492	10,873	29,131	7,912	14,502	4,120
Number of adult females in the household	0	16.87%	16.47%	10.84%	9.91%	13.86%	12.99%
	1	75.00%	75.01%	79.40%	80.11%	79.08%	79.49%
	≥2	8.12%	8.52%	9.76%	9.98%	7.06%	7.52%
Number of children in the household	0	80.29%	80.55%	74.24%	73.98%	84.02%	83.11%
	1	8.43%	8.25%	11.53%	11.10%	8.61%	9.30%
	≥2	11.27%	11.20%	14.24%	14.93%	7.38%	7.60%
Number of household members working full-time	0	52.02%	53.73%	33.19%	33.42%	29.93%	29.81%
	1	33.45%	32.00%	40.76%	39.79%	43.10%	41.33%
	≥2	14.54%	14.27%	26.05%	26.79%	26.98%	28.86%
Number of household members working part-time	0	81.98%	81.30%	77.86%	76.97%	81.75%	81.00%
	≥1	18.02%	18.70%	22.14%	23.03%	18.25%	19.00%
Number of household members with a medical condition that impairs their mobility	0	80.99%	81.55%	87.54%	88.09%	89.24%	89.30%
	≥1	19.01%	18.45%	12.46%	11.91%	10.76%	10.70%

This table shows summary statistics for PSM count control variables in the sample. Sample households are categorized based on the numbers of monthly E-purchases per household member.

First, control variables include household structure as households with children will travel differently (i.e., they may travel to school or places for kids to play sports) compared to childless households. I also included the number of female adults since women are more likely to shop (Gould and Golob, 1997), and the number of children in the household since children have different shopping needs.

To capture generational effect, four categorical variables are created based on definitions by the Pew Research Center (2018): Silent Generation (born before 1946), Baby Boomers (born between 1946 and 1964), Gen X (born between 1965 and 1980), Millennials (or Gen Y, born between 1981 and 1996), and Gen Z (born after 1996) (Dimock, 2019).

Household income clearly impacts the ability to shop and travel, so 5 groups are created based on approximate quintiles (20% strata) of annual household income. As education may influence shopping and travel (Cao *et al*., 2012; Zhou and Wang, 2014), four categories are defined to capture the highest educational achievement in the household (see Table 3.2).

Ethnicity may also play a role. I relied here on the ethnicity of the household head and create indictor variables for standard categories, and I created a binary variable to capture self-reported Hispanic status.

In addition, since the residential location may affect driving intensity and frequency (Cao *et al*., 2010b), I included population density in the census tract of household residence with three categories: low (less than 500 people per square mile), medium (500 to 4000 people per square mile), and high (more than 4000 people per square mile). I conjectured that higher population densities are associated with more numerous and more varied shopping opportunities.

The lack of access to motor vehicles could constrain travel (Mitra and Saphores, 2020), so I added a binary indicator for households with fewer drivers than vehicles.

Employment status impacts how much time people have to shop, so count variables are created for the number of household members who work part-time and full-time (Gould and Golob, 1997). Finally, adults with a medical condition that hinders their mobility would likely travel and shop differently, so a count variable is created to capture the number of household members affected.

3.4 Estimation and Results

The models were estimated using Stata 15. Overall results are presented in Tables 3.4 and 3.5, and results stratified by population density are shown from Figure 3.2 to Figure 3.4.

3.4.1 Specification tests

Before running PSM, binary logit models are estimated to confirm that potential control variables could explain the e-shopping categories I created based on the frequency of online purchases. Since travel behavior differs between weekdays and weekends, I estimated separate models for households whose NHTS survey day was on a weekday or on a weekend. I found that most of the control variables are highly significant, which confirmed the necessity to include them in my analysis.

For a PSM model to yield unbiased results, the distribution of each control variable in the treatment and control samples should be statistically equal (the so-called balancing condition). To check that the balancing condition is met for each of the models, I calculated for each control variable and for each model the standardized mean difference between the treatment and the control samples, the mean bias, and the ratio of variances. I relied on the criteria proposed by Rubin (2001) to assess the adequacy of matches ("good", "of concern", or "bad"). For each model, I also compared graphically the density of each control variable with its density in the corresponding control group to check to what extent they may differ.

For weekday models, no match was found to be bad, and only six variables had matches of concern (all of which were for the low population density sample and the group that shops online the most, but deviations were mild). All mean biases were under 2 percent, and the ratios of variances were between 0.8 and 1.25 (typically in [0.9,1.1]), with just six variables with variance ratios slightly over 1.25 or under 0.8.

For weekend models, matches were acceptable but not as good as for weekday models, possibly because sample sizes were smaller. There were no bad matches for any of the variables considered, but the match of a few variables was found of concern as follows (the first digit is the number of variables of concern for the match with medium frequency online shoppers, and the second is the number of variables of concern for the match with high frequency online shoppers): (6, 12) for the whole sample, (3,3) for the low population density sample, and (3,13) for the medium density case. There were no variables of concern for the high population density samples.

Despite these minor matching problems, I am confident that the residual biases in the estimates of the travel difference between treatment and control households are small.

3.4.2 Impact of e-shopping on travel for shopping

PSM results for general household travel characteristics are shown in Table 3.4. For all response variables (Column 1 of Table 3.4), the control group is low frequency online shoppers (at most one monthly online purchase per household member). Medium frequency online shoppers are households with 1 to 4 online purchases per member per month, and high frequency online shoppers are households with more than four monthly e-purchases per member.

In general, positive coefficients in Table 3.4 indicate that households who are more active in online shopping than households in the control group engage in more activities (including shopping), travel more (overall and in vehicles), and have higher percentages of shopping activities and shopping miles. Moreover, these differences are larger for weekdays than for weekends for which most measures of travel are not significant.

Table 3.4 PSM results (average treatment effect on the treated)

	Weekday travel		Weekend travel	
E-shopping frequency	**Medium♠**	**High♥**	**Medium♠**	**High♥**
Number of observations	71,623	56,994	18,785	14,993
Total number of activities (lower bound)	9.00***	7.79***	2.34***	--
Total number of activities (upper bound)	9.70***	7.57***	--	--
Number of shopping activities (lower bound)	2.05***	2.29***	0.70**	--
Number of shopping activities (upper bound)	2.22***	2.42***	--	--
Percentage of shopping activities (lower bound)	0.7%***	0.9%***	--	--
Percentage of shopping activities (upper bound)	0.7%***	0.9%***	--	--
Total miles traveled	57.24***	53.00**	--	--
Total vehicle miles traveled (VMT)	53.28***	--	--	--
Shopping miles	15.64***	20.31**	12.24**	--
Shopping VMT	15.36***	19.71**	13.57***	--
Percentage of shopping miles	0.6%***	1.0%***	1.0%*	--
Percentage of shopping VMT	0.5%**	0.8%***	1.2%**	1.4%*

This table shows the difference in the number of monthly activities/travel households for two treatment groups (medium or high frequency online shoppers) compared to low frequency shopping households (i.e., households with up to one monthly e-purchase per household member) for either weekdays or weekends.

♠ Households who are medium frequency online shoppers make more than one but no more than four monthly e-purchases per household member.

♥ Households who are high frequency online shoppers make over four monthly e-purchases per household member.

"--" indicates statistically non-significant results (p-value>0.10).

*, **, and *** designate statistical significance at the 0.05, 0.01, and 0.001 levels respectively.

More specifically, households in the medium frequency e-shopping group engage in 9 to 10 (9.00***-9.70***) more activities monthly on weekdays compared to households in the control group. Likewise, high frequency e-shopping households engage in up to 8 more activities monthly on weekdays (7.57*** to 7.79***). In other words, households who make more online purchases participate in more activities in their daily life, but only 2 of these activities are related to shopping (2.05*** to 2.22*** for medium frequency online shoppers, and 2.29*** to 2.42*** for high frequency online shoppers). Albeit statistically significant, this represents only a small increase in household shopping activities (0.7%*** and 0.9%*** for medium and high online frequency households respectively). On weekends, however, differences in the number of activities between control and treatment groups are significant only for medium frequency online shoppers, and they are much smaller than on weekdays.

These patterns carry over to miles traveled and VMT, for shopping and for other activities. On weekdays, households in the treatment groups travel more than 50 additional miles monthly (57.24*** and 53.00** for medium and high frequency online households respectively), of which over 90% are in motor vehicles. Interestingly, only a fraction of these miles (one third for medium frequency and two fifths for high frequency online shoppers) are shopping miles, which again corresponds only to a small increase (1 percent or less) in both shopping miles and shopping VMT. On weekends, although medium frequency online shoppers do not travel significantly more overall, they have slightly more shopping miles (12.24** per month), and a slightly higher VMT (13.57*** per month). It is also found that the shopping VMT percentages are marginally higher (1.2% and 1.4% for medium and high frequency online shoppers respectively) for weekends compared to weekdays, which reflects that more shopping often takes place on weekends when people have more time. Other weekend differences are small or insignificant.

Our results also show that more e-shopping does not translate monotonically into more travel and more miles as differences between the high frequency e-shopping group and the control group are not as large as those between the medium frequency e-shopping group and the control group. Differences on weekends between high frequency online shoppers and households in the control group are essentially not significant.

These results are consistent with previous findings that frequent online shoppers tend to shop at a higher rate in-store as well (Lee *et al*., 2017; Zhou and Wang, 2014). They reflect interactions between in-store and e-shopping, and confirm that households who shop online more tend to also have more physical shopping trips, although I do not know if they go to brick-and-mortar stores to examine items, buy, pick up orders, or simply return unwanted items.

3.4.3 Impact of e-shopping on various trip purposes

From the results above, it is found that households who shop more online have a greater number of activities and travel more. To better understand how e-shopping may influence household travel, I contrasted the frequency and mileage of the trip purposes available in the 2017 NHTS for households in two treatment groups and their respective control groups. Results, obtained via PSM, are shown in Table 3.5. This table only shows significant results; non-significant differences between a treatment group and the corresponding control group have been replaced by "- -". To link my discussion more concisely with Table 3.5, pairs of numbers are used to refer to results for medium and high frequency e-shopping households (e.g., (22.7%,29.4%) refers to weekdays' differences in shopping in Table 3.5).

Table 3.5 PSM results for percentages change in travel for different trip purposes

	Number of trips for various activities				Trip miles			
	Weekday travel		***Weekend travel***		***Weekday travel***		***Weekend travel***	
Treatment indicator	(1, 4]♠	> 4♥	(1, 4]♠	> 4♥	(1, 4]♠	> 4♥	(1, 4]♠	> 4♥
Shopping	22.7%	29.4%	30.0%	--	27.3%	38.3%	--	--
Buy meals	14.9%	16.7%	21.2%	--	21.1%	20.3%	--	--
Exercise	10.8%	14.7%	13.7%	--	7.2%	7.7%	--	--
Health care visits	3.4%	6.5%	--	--	8.0%	13.0%	--	--
Buy services	3.9%	6.2%	6.8%	--	--	--	--	--
Recreation	--	--	--	--	--	--	--	--
Visit friends	5.4%	5.1%	--	--	--	--	--	--
Religious & Community	--	-4.0%	-9.9%	--	--	--	--	--
Regular home activities	17.6%	--	28.3%	--	--	--	--	--
Work and work-related	--	--	--	--	--	-30.9%	--	--
Work from home	7.1%	14.7%	--	--	--	--	--	--
Volunteer activities	2.4%	--	--	--	--	--	--	--
Transport related	--	--	--	--	--	--	--	--
Attend school / child / adult care	--	--	--	--	7.3%	4.9%	--	--
Other errands	--	7.9%	--	--	--	--	--	--
Something else	-1.0%	--	--	--	--	--	--	--
Total	87.3%	97.2%	90.1%	--	70.9%	53.4%	--	--

♠ Medium frequency online shoppers: more than one but no more than four monthly e-purchases per household member.
♥ High frequency online shopper: over four monthly e-purchases per household member.
"--" indicates statistically non-significant results (p-value>0.10).
A positive number shows an increase in the number of trips or miles traveled for a group of online shoppers (medium or high frequency online shoppers) compared to low frequency online shoppers; conversely, a negative number shows a decrease.

On weekdays, differences in e-shopping frequency can explain most of the induced trips from e-shopping, with 87.3% (medium frequency online shoppers) and 97.2% (high frequency) for weekdays, and to a lesser extent for miles traveled, with 70.9% and 53.4% for medium and high frequency e-shoppers, respectively. Households who shop online more tend to also shop in stores more (22.7%, 29.4%), and buy more meals (go out for a meal, snack, carry-out) (14.9%, 16.7%), possibly because they are more comfortable with e-shopping and eating outside in general. They also tend to exercise (10.8%, 14.7%), visit friends (5.4%, 5.1%), and buy services (dry cleaners, banking, service a car, pet care) (3.9%, 6.2%) more frequently. In addition, there are slightly more frequent health care visits. One possible explanation for these differences is that households who are very active online buyers have more time flexibility as they work from home more (7.1%, 14.7%).

These weekday differences in the number of activities for online shoppers translate into differences in trip miles, although not across the board for all activity types. Households who shop more online also tend to travel more to shop in brick-and-mortar stores (27.3%, 38.3%) and buy meals (21.1%, 20.3%), but also for health care visits (8.0%, 13.0%), to exercise (7.2%, 7.7%), and to attend school or care for a child or another adult (7.3%, 4.9%). Except for work-related trip miles, which are 30.9% lower for high compared to low frequency online shoppers, the other differences in trip miles are not statistically significant partly because some trip purposes are relatively rare as indicated in Figure 3.1.

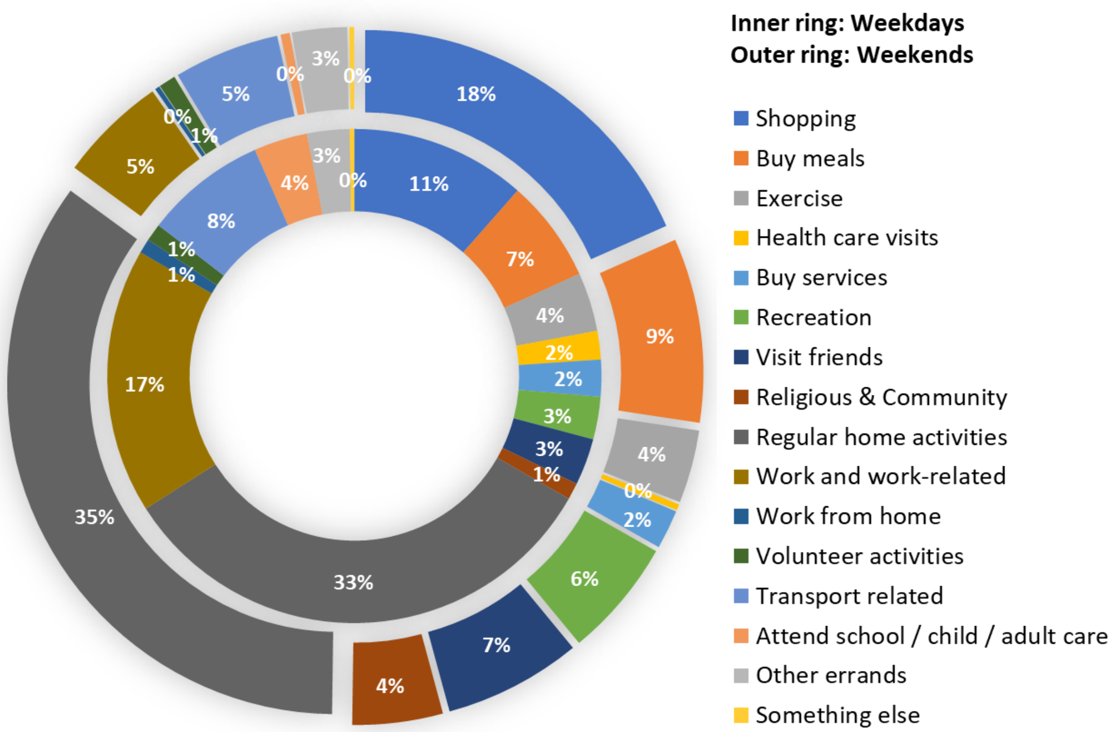


Figure 3.1 Percentages of household trips for various trip purposes
Note: trip purposes are presented clockwise, starting with shopping at 1 o'clock (18% on weekends and 11% on weekdays).

On weekends, I also observe differences in the number of activities, but partly because of sample size, these differences are significant only for medium frequency online shoppers; as reported in Table 3.5, the number of activities and the number of miles traveled does not differ significantly between the high and the low frequency e-shopping households. Compared to low frequency online shoppers, the latter tend to shop more in brick-and-mortar stores (30.0%, - -), buy more meals (21.2%, - -), exercise (13.7%, - -), and buy services (6.8%, - -), all of which increase more than during weekdays. They also have more regular home activities (28.3%, - -), but fewer religious and community activities (-9.9%, - -). However, extra trips do not translate into a significant number of additional miles traveled. One reason may be an increase in trip chaining, which may allow undertaking more activities without increasing travel.

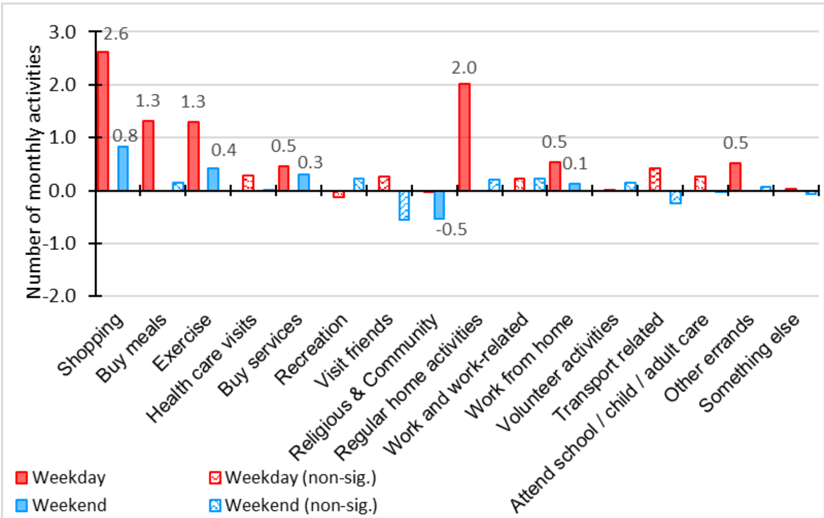
Finally, I note that more e-shopping does not result in more recreation (visiting parks, movies, bars, museums). In addition to recognizing that shopping often includes a recreational component (Farag *et al*., 2007), I cannot discard that some trip purposes have multiple dimensions (e.g., "going for a walk" could be seen as exercise or as recreation).

3.4.4 PSM results by population density bands

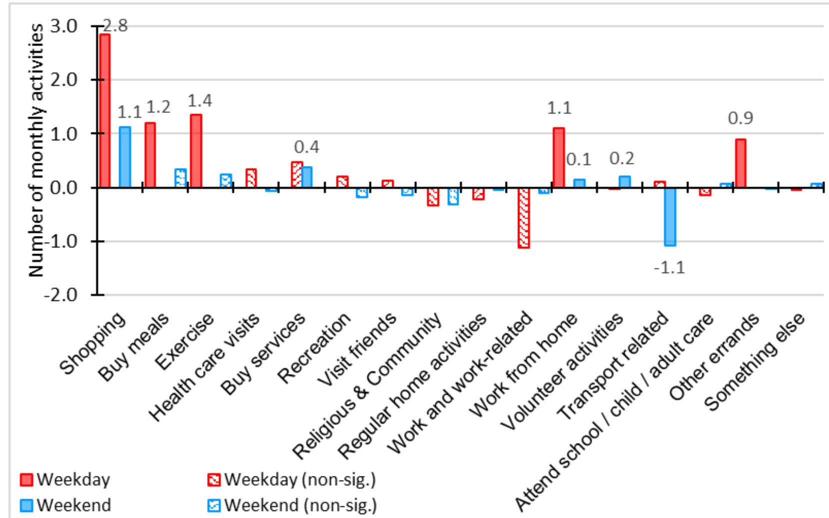
Results presented above control for population density, which is known to impact how people travel (Caliendo and Kopenig, 2005; Heckman *et al*., 1998) by offering additional modes (e.g., transit, micro-mobility options, or shared transportation) and more potential activities. For example, areas with higher population densities often offer a larger number of closer and more diverse shopping opportunities. To further unpack the impact of population density on my results, I repeated analyses after splitting households in the sample into three groups based on the population density in the census tract of their residence: low density for under 500 ppsm (persons per square mile), medium density for between 500 ppsm and 4,000 p/sqm, and high density for over 4,000 ppsm.

PSM results by population density are shown from Figure 3.2 to Figure 3.4. Each figure corresponds to one of three density groups and has 4 panels. Panels A and C respectively present differences in the number of various activities and in the corresponding travel mileage between medium e-shopping frequency households and their (low e-shopping frequency) control group, while Panels B and D present similar information for high e-shopping frequency households versus their control group. As above, weekday and weekend travel are distinguished. For completeness, these figures also display results that are not statistically significant to illustrate travel variability for some trip purposes, many of which are a small percentage of all trip purposes as shown in Figure 3.1.

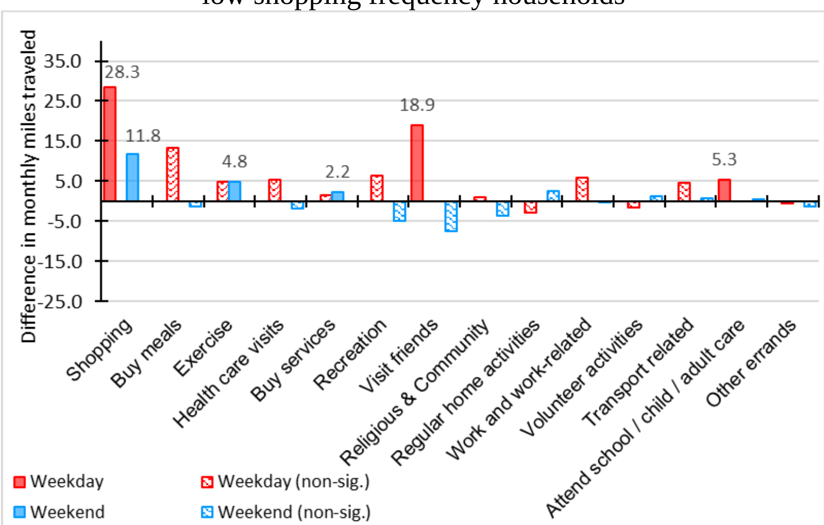
Overall, compared to households in their control groups, active e-shopping households consistently have more activities where they shop, buy meals, exercise, and work from home. Differences in the number of activities are typically larger on weekdays, with a few exceptions (e.g., more regular home activities for medium frequency online shoppers in higher density areas, on Panel A of Figure 3.4). Moreover, the magnitude of these differences depends on population density. For example, the number of extra shopping activities on weekdays for medium frequency e-shopping households is 2.6 in lower density areas, and 1.5 in higher density areas. Other differences in the number of activities appear to be relatively minor. For example, households who e-shop more tend to buy more services, except if they are high frequency e-shoppers who live in high-density areas (Panel B of Figure 3.4).



Panel A Difference in number of monthly activities for medium vs. low shopping frequency households



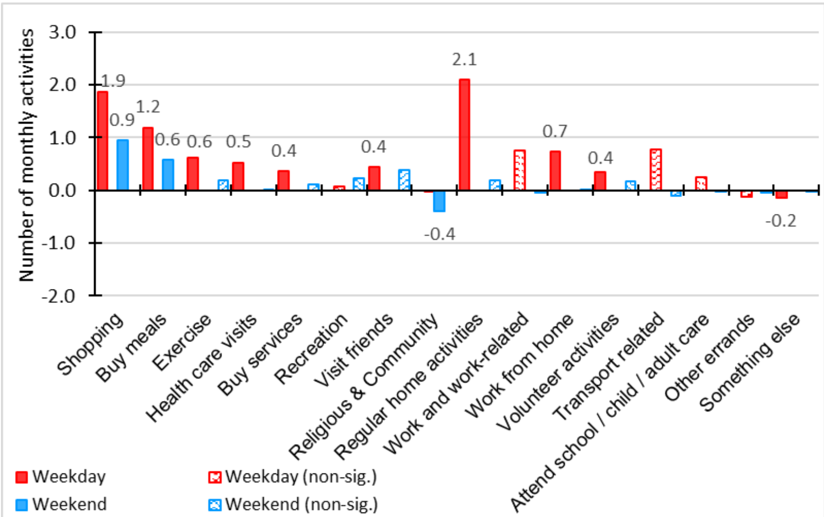
Panel B Difference in number of monthly activities for high vs. low shopping frequency households



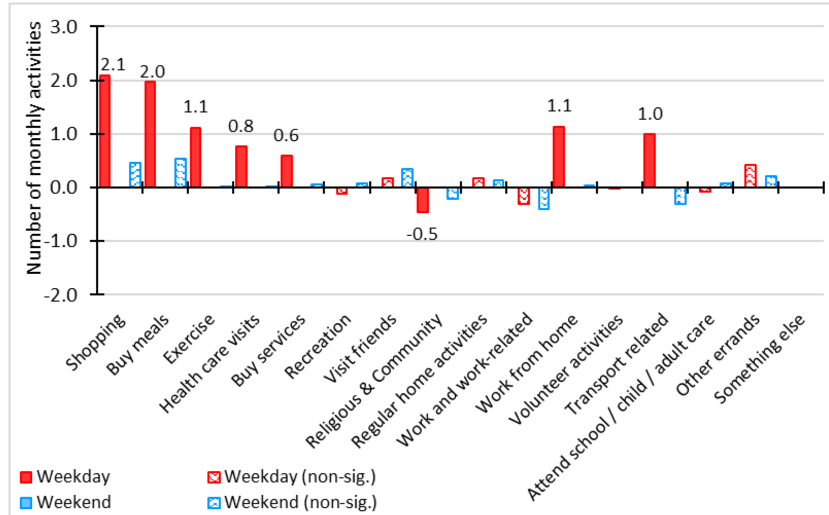
Panel C Difference in monthly miles traveled for medium vs. low shopping frequency households

Panel D Difference in monthly miles traveled for high vs. low shopping frequency households

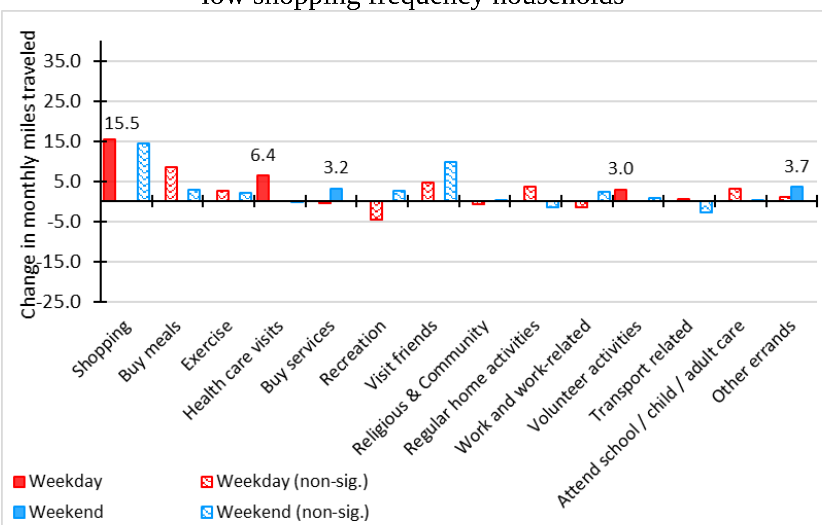
Figure 3.2 Travel differences for low population density areas (under 500 p/sqm)



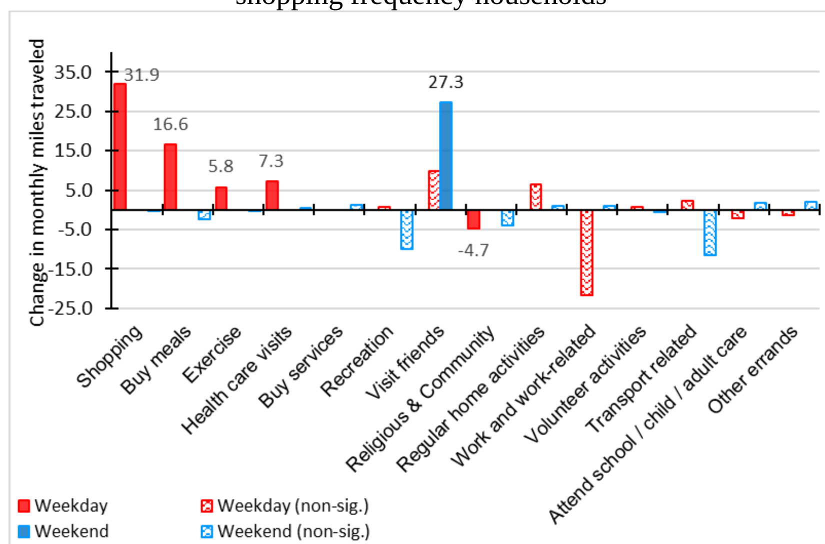
Panel A Difference in number of monthly activities for medium vs. low shopping frequency households



Panel B Difference in number of monthly activities for high vs. low shopping frequency households

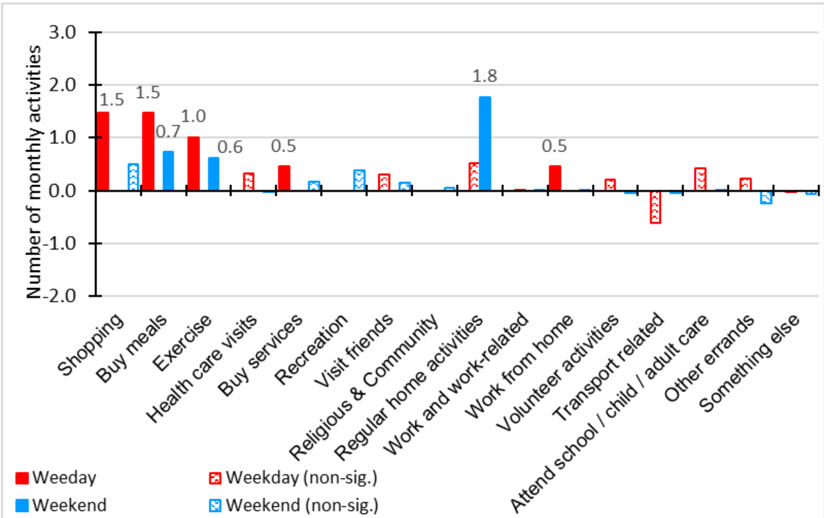


Panel C Difference in monthly miles traveled for medium vs. low shopping frequency households

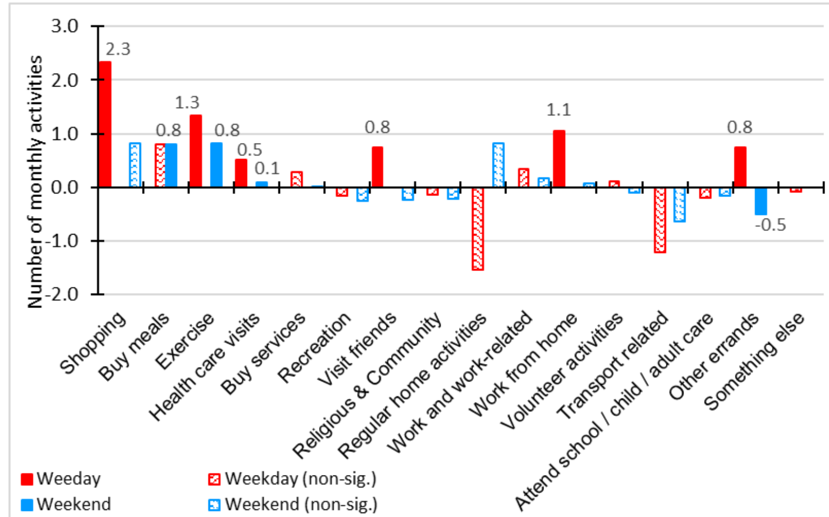


Panel D Difference in monthly miles traveled for high vs. low shopping frequency households

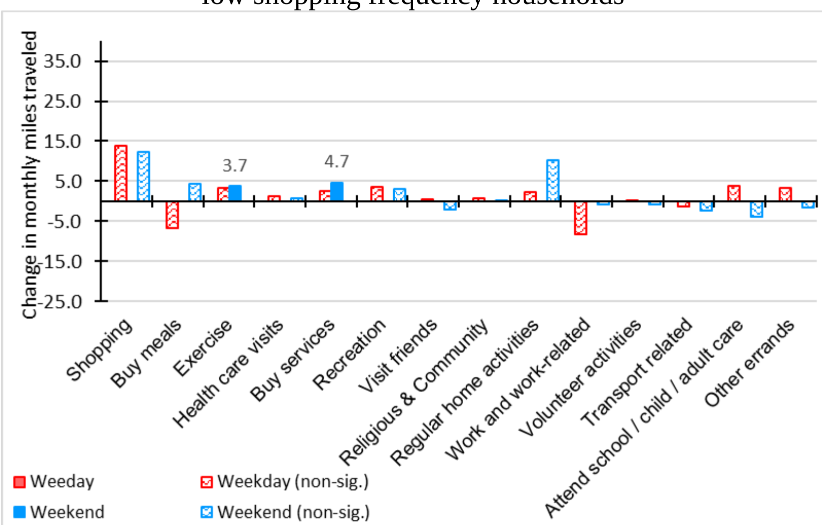
Figure 3.3 Travel differences for medium population density areas (500 p/sqm to 4000 p/sqm)



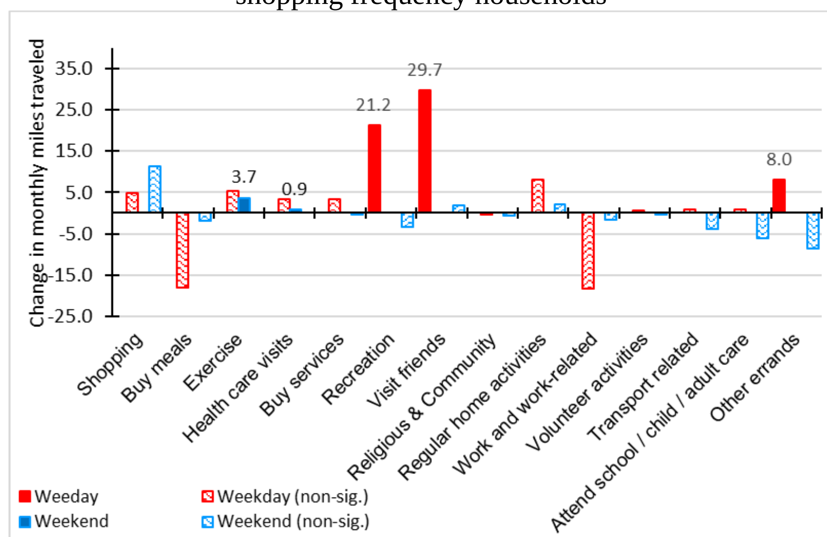
Panel A Difference in number of monthly activities for medium vs. low shopping frequency households



Panel B Difference in number of monthly activities for high vs. low shopping frequency households



Panel C Difference in monthly miles traveled for medium vs. low shopping frequency households



Panel D Difference in monthly miles traveled for high vs. low shopping frequency households

Figure 3.4 Travel differences for high population density areas (over 4000 p/sqm)

However, the extent to which more activities translate into more travel depends on population density, the frequency of e-shopping, and the activities considered. For example, more shopping by medium frequency e-shopping households results in more driving (28.3 more miles per month on weekdays and 11.8 mi on weekends) in lower density areas, which drops to 15.5 more miles per month on weekdays only in the medium density area, and no significant change in driving in higher density areas. Conversely, more shopping by high frequency e-shopping households results in fewer additional miles in lower density areas (+23.6 miles on weekends, no statistically significant difference on weekdays) than in medium density areas (+31.9 miles on weekdays, none on weekends), with no statistically significant differences in travel from shopping in higher density areas (Panel D of Figure 3.4). Finally, I note that there can also be a change in miles traveled for an activity even though the number of monthly occurrences for a household is unchanged: this is the case for recreation for high frequency online shoppers who live in higher density areas (see Panels B and D in Figure 3.4), for example. These results point to the complex relationship between the number and nature of household activities, travel, and time use (which unfortunately is not captured in the NHTS).

3.5 Conclusions

In this chapter, I have analyzed data from the 2017 NHTS to understand the impact of e-shopping on household travel in the United States using PSM to control for self-selection. Although this topic has already received some attention, most previous studies did not account for the impact of self-selection in their analysis, and even when they have, this topic should be revisited because the relationship between e-shopping and travel has been evolving with technological improvements (such as virtual reality tools) and changes in shopping habits.

Results show that more e-shopping is, on average, associated with more travel. Overall, I found that households who e-shop more than once per person per month have on average 2.2 to 2.4 additional monthly shopping trips on weekdays compared to households who e-shop less frequently. Households with avid e-shoppers also travel on average 15.6 to 20.3 more shopping miles per month miles on weekdays (and 12.2 on weekends but only if they are in the medium frequency e-shopping group).

The results further indicate that e-shopping has a deep impact on household travel. By broadening people's access to goods and services and by giving them more time flexibility (baring internet outages, e-shopping is possible 24/7), households who are medium or high frequency online buyers have more activities where they shop, buy meals, exercise, or simply work from home. However, the extent to which an increase in the number of household activities translates into more travel depends on population density, the frequency of e-shopping, and the activities considered. These results point to the complex relationship between the number and the nature of household activities, household travel, and time use, which unfortunately is not recorded in the NHTS.

In addition to the social dimension of shopping (i.e., customers enjoy sharing their shopping experience with others) (Rintamäki *et al*., 2006), I surmise that a couple of factors explain results. First, customers typically like examining experience goods (Nelson, 1970) in person, for example by smelling or assessing the firmness of fresh vegetables (Saphores and Xu, 2020), or by trying on a garment. For the former, customers may buy online generic grocery items (e.g., canned tuna or toilet paper) and shop for produce in a brick-and-mortar store, and for the latter they may buy an item online after experiencing how they look in it in a brick-and-mortar store.

Second, in spite of improvements in website design and the availability of customer reviews (including home-made video-reviews), shoppers are more likely to return items bought online: while 5 to 10 percent of in-store purchases are returned, this jumps to 15 to 40 percent for online purchases, with shoes and clothing at the high end of this range (Reagan, 2019). Not being able to experience a product in-person motivates some shoppers to order multiple sizes and colors of the same clothing item, knowing that they can return unwanted items for free (i.e., with shipping paid by the retailer). Although the value of returns may soon exceed a trillion dollars per year (Reagan, 2019), retailers still hesitate to abandon free return policies because free returns are important to e-shoppers (Charlton, 2020), and the e-shopping environment is very competitive.

While online retailers ponder how to address the increasing cost of returns from e-shopping, policymakers may consider several measures to stem the increase in VMT associated with e-shopping. First, at the local level zoning could be amended in residential areas to allow for the creation of pick-up depots (such as Amazon lockers) coupled with return centers within walking distance of the customers they serve. Second, public funding could be considered to help develop virtual reality tools for shopping. While larger retailers such as Neiman Marcus (its MemoryMirror shows a 360° view of a person wearing virtually various outfits (Steele, 2015)) or Ikea (whose virtual reality showroom enable customers to explore how a piece of furniture would look like in their home (Ikea, 2018)), are already developing some of these tools, they will likely be unavailable to smaller retailers. One could argue that giving smaller retailers access to this technology would help keep retail competitive, as large retailers such as Amazon, Walmart, or Costco continue to increase their market share. Finally, policymakers should strive to make sure that disadvantaged communities have access to fast and reliable internet services so they can benefit from changes in retail (especially communities underserved by supermarkets (Beaulac *et al*., 2009; Morland *et al*., 2002)).

One limitation of this study is that control variables do not include data on personal attitudes towards shopping and traveling, which may impact matching (Rosenbaum and Rubin, 1985), because such data were not available in the 2017 NHTS. A second limitation is that purchases such as groceries, clothing, appliances, and gas were all lumped together in 2017 NHTS questions, which was unfortunate because some of these goods can be purchased online and others not. Third, the 2017 NHTS does not have detailed time use information, which would allow us to better understand the interactions between e-shopping, conventional shopping, and travel.

In addition to understanding the impacts of the Covid-19 pandemic on how people shop online and in brick-and-mortar stores, future research could examine the changing nexus between shopping, online browsing, and traveling by combining time use and travel diary information.

3.6 References

Abadie, A., & Imbens, G. W. (2016). Matching on the Estimated Propensity Score. *Econometrica*, *84*(2), 781–807. https://doi.org/10.3982/ecta11293

Aguiléra, A., Guillot, C., & Rallet, A. (2012). Mobile ICTs and physical mobility: Review and research agenda. *Transportation Research Part A: Policy and Practice*, *46*(4), 664–672. https://doi.org/10.1016/j.tra.2012.01.005

American Council for an Energy-Efficient Economy (2020). *State and Local Policy Database*. https://database.aceee.org/city/sustainable-transportation-planning

Austin, P. C. (2009a). Using the standardized difference to compare the prevalence of a binary variable between two groups in observational research. *Communications in Statistics: Simulation and Computation*, *38*(6), 1228–1234. https://doi.org/10.1080/03610910902859574

Austin, P. C. (2009b). Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. In *Statistics in Medicine* (Vol. 28, Issue 25, pp. 3083–3107). https://doi.org/10.1002/sim.3697

Austin, P. C., Jembere, N., & Chiu, M. (2018). Propensity score matching and complex surveys. *Statistical Methods in Medical Research*, *27*(4), 1240–1257. https://doi.org/10.1177/0962280216658920

Beaulac, J., Kristjansson, E., & Cummins, S. (2009). A Systematic Review of Food Deserts, 1966-2007. *Preventing Chronic Disease*, *6*(3).

Bhat, C. R., Sivakumar, A., & Axhausen, K. W. (2003). An analysis of the impact of information and communication technologies on non-maintenance shopping activities. *Transportation Research Part B: Methodological*, *37*(10), 857–881. https://doi.org/10.1016/S0191-2615(02)00062-0

Blumberg, S. J., & Luke, J. V. (2017). Wireless substitution: Early release of estimates from the National Health Interview Survey, July–December 2016. *National Center for Health Statistics*. https://www.cdc.gov/nchs/data/nhis/earlyrelease/wireless201705.pdf

Caliendo, M., & Kopenig, S. (2005). Some Practical Guidance for the Implementation of Propensity Score Matching. *IZA Discussion Paper*. https://doi.org/10.1111/j.1467-6419.2007.00527.x

California Environmental Protection Agency. (2020). *Climate Action Programs*. https://calepa.ca.gov/climate/

Cao, X. (2009). E-Shopping, Spatial Attributes, and Personal Travel: A Review of Empirical Studies. *Transportation Research Record*, *2135*, 160–169. https://doi.org/10.3141/2135-19

Cao, X. (2012). The relationships between e-shopping and store shopping in the shopping process of search goods. *Transportation Research Part A: Policy and Practice*, *46*(7), 993–1002. https://doi.org/10.1016/j.tra.2012.04.007

Cao, X., Douma, F., & Cleaveland, F. (2010a). Influence of E-Shopping on Shopping Travel: Evidence from Minnesota's Twin Cities. *Transportation Research Record: Journal of the Transportation Research Board*, *2157*(1), 147–154. https://doi.org/10.3141/2157-18

Cao, X., & Fan, Y. (2012). Exploring the influences of density on travel behavior using propensity score matching. *Environment and Planning B: Planning and Design*, *39*(3), 459–470. https://doi.org/10.1068/b36168

Cao, X., Mokhtarian, P. L., & Handy, S. L. (2009). Examining the impacts of residential self-selection on travel behaviour: A focus on empirical findings. *Transport Reviews* (Vol. 29, Issue 3). https://doi.org/10.1080/01441640802539195

Cao, X., Xu, Z., & Douma, F. (2012). The interactions between e-shopping and traditional in-store shopping: An application of structural equations model. *Transportation*, *39*(5), 957–974. https://doi.org/10.1007/s11116-011-9376-3

Cao, X., Xu, Z., & Fan, Y. (2010b). Exploring the connections among residential location, self-selection, and driving: Propensity score matching with multiple treatments. *Transportation*

Research Part A: Policy and Practice, *44*(10), 797–805. https://doi.org/10.1016/j.tra.2010.07.010

Charlton, G. (2020). *Ecommerce Returns: 2020 Stats and Trends*. SaleCycle. https://www.salecycle.com/blog/featured/ecommerce-returns-2018-stats-trends/

Dias, F. F., Lavieri, P. S., Sharda, S., Khoeini, S., Bhat, C. R., Pendyala, R. M., Pinjari, A. R., Ramadurai, G., & Srinivasan, K. K. (2020). A comparison of online and in-person activity engagement: The case of shopping and eating meals. *Transportation Research Part C: Emerging Technologies*, *114*, 643–656. https://doi.org/10.1016/j.trc.2020.02.023

Dimock, M. (2019). *Defining generations: Where Millennials end and Generation Z begins*. Pew Research Center. https://www.pewresearch.org/fact-tank/2019/01/17/where-millennials-end-and-generation-z-begins/

Ding, Y., & Lu, H. (2017). The interactions between online shopping and personal activity travel behavior: an analysis with a GPS-based activity travel diary. *Transportation*, *44*(2), 311–324. https://doi.org/10.1007/s11116-015-9639-5

Etminani-Ghasrodashti, R., & Hamidi, S. (2020). Online shopping as a substitute or complement to in-store shopping trips in Iran? *Cities*, *103*, 102768. https://doi.org/10.1016/j.cities.2020.102768

Farag, S., Dijst, M., & Lanzendorf, M. (2003). Exploring the Use of E-Shopping and Its Impact on Personal Travel Behavior in the Netherlands. *Transportation Research Record: Journal of the Transportation Research Board*, *1858*(1), 47–54. https://doi.org/10.3141/1858-07

Farag, S., Krizek, K. J., & Dijst, M. (2006a). E-shopping and its relationship with in-store shopping: Empirical evidence from the Netherlands and the USA. *Transport Reviews* (Vol. 26, Issue 1, pp. 43–61). https://doi.org/10.1080/01441640500158496

Farag, S., Schwanen, T., Dijst, M., & Faber, J. (2007). Shopping online and/or in-store? A structural equation model of the relationships between e-shopping and in-store shopping. *Transportation Research Part A: Policy and Practice*, *41*(2), 125–141. https://doi.org/10.1016/j.tra.2006.02.003

Farag, S., Weltevreden, J., van Rietbergen, T., Dijst, M., & van Oort, F. (2006b). E-Shopping in the Netherlands: Does Geography Matter? *Environment and Planning B: Planning and Design*, *33*(1), 59–74. https://doi.org/10.1068/b31083

Federal Highway Administration, U. S. D. of T. (2017). *2017 National Household Travel Survey*. https://nhts.ornl.gov

Ferrell, C. E. (2004). Home-Based Teleshoppers and Shopping Travel: Do Teleshoppers Travel Less ? *Transportation Research Record*, *1894*(1), 241–248.

Ferrell, C. E. (2005). Home-Based Teleshopping and Shopping Travel: Where Do People Find the Time ? *Transportation Research Record*, *1926*, 212–223.

Goodchild, A., Wygonik, E., & Mayes, N. (2018). An analytical model for vehicle miles traveled and carbon emissions for goods delivery scenarios. *European Transport Research Review*, *10*(1), 1-10. https://doi.org/10.1007/s12544-017-0280-6

Gould, J., & Golob, T. F. (1997). Shopping without travel or travel without shopping? an investigation of electronic home shopping. *Transport Reviews*, *17*(4), 355–376. https://doi.org/10.1080/01441649708716991

Haridasan, A. C., & Fernando, A. G. (2018). Online or in-store: unravelling consumer's channel choice motives. *Journal of Research in Interactive Marketing*, *12*(2), 215–230. https://doi.org/10.1108/JRIM-07-2017-0060

Heckman, J., Ichimura;, H., Smith;, J., & Todd, P. (1998). Characterizing selection bias using experimental data. *National Bureau of Economic Research*.

Heldt, B., Matteis, T., von Schmidt, A., & Heinrichs, M. (2019). Cool but dirty food? – Estimating the impact of grocery home delivery on transport and CO2 emissions including cooling. *Research in Transportation Economics*, *January*, 100763. https://doi.org/10.1016/j.retrec.2019.100763

Hirschmann, R. (2019). *Shoppers' behaviors regarding online and offline shopping Singapore 2019*. Statista. https://www.statista.com/statistics/1026257/singapore-shoppers-behaviors-regarding-online-shopping/#:~:text=According to a survey conducted,then purchasing it in-store

Hiselius, L. W., Rosqvist, L. S., & Adell, E. (2015). Travel Behaviour of Online Shoppers in Sweden. *Transport and Telecommunication Journal*, *16*(1), 21–30. https://doi.org/10.1515/ttj-2015-0003

Hoogendoorn-Lanser, S., Olde Kalter, M. J., & Schaap, N. T. W. (2019). Impact of different shopping stages on shopping-related travel behaviour: analyses of the Netherlands Mobility Panel data. *Transportation*, *46*(2), 341–371. https://doi.org/10.1007/s11116-019-09993-7

Ikea. (2018). *IKEA Virtual Interior Designer - A whole new home and retail VR application*. https://present.digital/ikea/

Jaller, M., & Pahwa, A. (2020). Evaluating the environmental impacts of online shopping: A behavioral and transportation approach. *Transportation Research Part D: Transport and Environment*, *80*, 102223. https://doi.org/10.1016/j.trd.2020.102223

Lavieri, P. S., Dai, Q., & Bhat, C. R. (2018). Using virtual accessibility and physical accessibility as joint predictors of activity-travel behavior. *Transportation Research Part A: Policy and Practice*, *118*, 527–544. https://doi.org/10.1016/j.tra.2018.08.042

Lee, R. J., Sener, I. N., Mokhtarian, P. L., & Handy, S. L. (2017). Relationships between the online and in-store shopping frequency of Davis, California residents. *Transportation Research Part A: Policy and Practice*, *100*, 40–52. https://doi.org/10.1016/j.tra.2017.03.001

Litman, T. (2022). Are Vehicle Travel Reduction Targets Justified ? Victoria Transport Policy Institute. chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/viewer.html?pdfurl=https%3A%2F%2Fwww.vtpi.org%2Fvmt_red.pdf&clen=1929668&chunk=true

Maat, K., & Konings, R. (2018). Accessibility or innovation? Store shopping trips versus online shopping. *Transportation Research Record*, *2672*(50), 1–10. https://doi.org/10.1177/0361198118794044

McGuckin, N., & Fucci, A. (2018). *Summary of Travel Trends: 2017 National Household Travel Survey*.

Mishra, G. S., Clewlow, R. R., Mokhtarian, P. L., & Widaman, K. F. (2015). The effect of carsharing on vehicle holdings and travel behavior: A propensity score and causal mediation analysis of the San Francisco Bay Area. *Research in Transportation Economics*, *52*, 46–55. https://doi.org/10.1016/j.retrec.2015.10.010

Mitra, S. K., & Saphores, J. D. M. (2019). How do they get by without cars? An analysis of travel characteristics of carless households in California. *Transportation*. https://doi.org/10.1007/s11116-019-09994-6

Mokhtarian, P. (2009). If telecommunication is such a good substitute for travel, why does congestion continue to get worse? *Transportation Letters*, *1*(1), 1–17. https://doi.org/10.3328/tl.2009.01.01.1-17

Mokhtarian, P. L. (2002). Telecommunications and Travel: The Case for Complementarity. *Journal of Industrial Ecology*, *6*(2), 43–57. https://doi.org/10.1162/108819802763471771

Mokhtarian, P. L. (2004). A conceptual analysis of the transportation impacts of B2C e-commerce. *Transportation, 31*(3), 257–284. https://doi.org/10.1023/B:PORT.0000025428.64128.d3

Morland, K., Wing, S., Roux, A. D., & Poole, C. (2002). Neighborhood Characteristics Associated with the Location of Food Stores and Food Service Places. *American Journal of Preventive Medicine, 22*(1).

Nelson, P. (1970). Information and Consumer Behavior. *Journal of Political Economy*, *78*(2), 311–329.

Reagan, C. (2019). *That sweater you don't like is a trillion-dollar problem for retailers. These companies want to fix it*. CNBC. https://www.cnbc.com/2019/01/10/growing-online-sales-means-more-returns-and-trash-for-landfills.html

Rintamäki, T., Kanto, A., Kuusela, H., & Spence, M. T. (2006). Decomposing the value of department store shopping into utilitarian, hedonic and social dimensions: Evidence from

Finland. *International Journal of Retail & Distribution Management*, *34*(1), 6–24. https://doi.org/DOI 10.1108/09590550610642792

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Matched Sampling for Causal Effects*, *70*(1), 41–55. https://doi.org/10.1017/CBO9780511810725.016

Rubin, D. B. (2001). Using propensity scores to help design observational studies: Application to the tobacco litigation. *Matched Sampling for Causal Effects*, 365–382. https://doi.org/10.1017/CBO9780511810725.030

Salomon, I. (1986). Telecommunications and travel relationships: a review. *Transportation Research Part A: General*, *20*(3), 223–238. https://doi.org/10.1016/0191-2607(86)90096-8

Salomon, I., & Koppelman, F. (1988). A framework for studying teleshopping versus store shopping. *Transportation Research Part A: Policy and Practice*, *4*, 247–255.

Sam, C.-Y., & Sharma, C. (2015). An Exploration into the Factors Driving Consumers in Singapore towards or away from the Adoption of Online Shopping. *Global Business and Management Research: An International Journal*, *7*(1), 60–73.

Saphores, J.-D., & Xu, L. (2021). E-shopping changes and the state of E-grocery shopping in the US - Evidence from national travel and time use surveys. *Research in Transportation Economics*. https://doi.org/10.1016/j.retrec.2020.100864

Schmid, B., & Axhausen, K. W. (2019). In-store or online shopping of search and experience goods: A hybrid choice approach. *Journal of Choice Modelling*, *31*, 156–180. https://doi.org/10.1016/j.jocm.2018.03.001

Shepardson, D., & Groom, N. (2020). *California passes landmark mandate for zero emission trucks - Reuters*. REUTERS. https://www.reuters.com/article/us-california-trucks-electric-idUSKBN23W31N

Shi, K., De Vos, J., Yang, Y., & Witlox, F. (2019). Does e-shopping replace shopping trips? Empirical evidence from Chengdu, China. *Transportation Research Part A: Policy and Practice*, *122*, 21–33. https://doi.org/10.1016/j.tra.2019.01.027

Sim, L. L., & Koi, S. M. (2002). Singapore's internet shoppers and their impact on traditional shopping patterns. *Journal of Retailing and Consumer Services*, *9*(2), 115–124. https://doi.org/10.1016/S0969-6989(01)00029-7

Steele, B. (2015). *Neiman Marcus' digital mirror compares clothes side by side*. https://www.engadget.com/2015-01-13-neiman-marcus-memory-mirror.html

Suel, E., Le Vine, S., & Polak, J. (2015). Empirical application of expenditure diary instrument to quantify relationships between in-store and online grocery shopping: Case study of greater London. *Transportation Research Record*, *2496*, 45–54. https://doi.org/10.3141/2496-06

Suel, E., & Polak, J. W. (2017). Development of joint models for channel, store, and travel mode choice: Grocery shopping in London. *Transportation Research Part A: Policy and Practice*, *99*, 147–162. https://doi.org/10.1016/j.tra.2017.03.009

United States Census Bureau. (2020). *Annual Retail Trade Survey Supplemental E-commerce Tables: 2018*. https://www.census.gov/data/tables/2018/econ/arts/supplemental-ecommerce.html

Visser, E. J., & Lanzendorf, M. (2003). Mobility and accessibility effects of b2c e-commerce: A literature review. *Tijdschrift Voor Economische En Sociale Geografie*, *95*(2), 189–205. https://doi.org/10.1111/j.0040-747X.2004.00300.x

Weltevreden, J. W. J. (2007). Substitution or complementarity? How the Internet changes city centre shopping. *Journal of Retailing and Consumer Services*, *14*(3), 192–207. https://doi.org/10.1016/j.jretconser.2006.09.001

World Trade Organization. (2020). *E-commerce, trade and the Covid-19 pandemic*.

Xi, G., Cao, X., & Zhen, F. (2020). The impacts of same day delivery online shopping on local store shopping in Nanjing , China. *Transportation Research Part A*, *136*, 35–47. https://doi.org/10.1016/j.tra.2020.03.030

Xi, G., Zhen, F., Cao, X., & Xu, F. (2018). The interaction between e-shopping and store shopping: empirical evidence from Nanjing, China. *Transportation Letters,* 1–9. https://doi.org/10.1080/19427867.2018.1546797

Xiao, Z., Wang, J. J., & Liu, Q. (2018). The impacts of final delivery solutions on e-shopping usage behaviour: The case of Shenzhen, China. *International Journal of Retail and Distribution Management*, *46*(1), 2–20. https://doi.org/10.1108/IJRDM-03-2016-0036

Zhai, Q., Cao, X. (Jason), & Zhen, F. (2019). Relationship between Online Shopping and Store Shopping in the Shopping Process: Empirical Study for Search Goods and Experience Goods in Nanjing, China. *Transportation Research Record*, *2673*(11), 38–47. https://doi.org/10.1177/0361198119851751

Zhai, Q., Cao, X., Mokhtarian, P. L., & Zhen, F. (2017). The interactions between e-shopping and store shopping in the shopping process for search goods and experience goods. *Transportation*, *44*(5), 885–904. https://doi.org/10.1007/s11116-016-9683-9

Zhen, F., Cao, X., Mokhtarian, P. L., & Xi, G. (2016). Associations between online purchasing and store purchasing for four types of products in Nanjing, China. *Transportation Research Record*, *2566*(2566), 93–101. https://doi.org/10.3141/2566-10

Zhou, Y., & Wang, X. (Cara). (2014). Explore the relationship between online shopping and shopping trips: An analysis with the 2009 NHTS data. *Transportation Research Part A: Policy and Practice*, *70*, 1–9. https://doi.org/10.1016/j.tra.2014.09.014

Chapter 4. Grocery Shopping in the Age of Covid-19: Results from California

4.1 Introduction

The Covid-19 pandemic has boosted e-commerce retail sales, with increases topping 30% for the last three quarters of 2020 compared to the corresponding quarters in 2019 (U.S. Census Bureau, 2021). Online sales of groceries in the U.S. have increased even faster (+63.9% year over year), although their growth is expected to slow down to 12.3% in 2021, with sales expected to top $244 billion by 2025 (Droesch, 2021) compared to $62.2 billion in 2019. Although online grocery shopping with home delivery (e-grocery) was gaining ground before the pandemic, its explosive growth reflects households' willingness to adopt new shopping options to avoid crowds (Shamshiripour *et al*., 2020; Unnikrishnan and Figliozzi, 2021; Wang *et al*., 2020).

Because fresh groceries are perishable, they require more complex logistics (Figliozzi and Unnikrishnan, 2021; Murphy, 2003), and warehouses/distribution centers near residential areas to provide increasingly fast delivery (Jaller and Pineda, 2017). As e-grocery involves additional labor for picking, packing, and delivering orders, it entails additional costs that need to be kept to a minimum for this option to remain viable in a typically very competitive environment (Figliozzi and Unnikrishnan, 2021; Punel and Stathopoulos, 2017; Shamshiripour *et al*., 2020).

Although grocery shopping fulfills a basic need, few empirical studies published to date have analyzed changes in food purchases resulting from the Covid-19 pandemic and the extent to which these changes may last. They also almost all analyze the first few months of the pandemic. Moreover, most published studies have relied on non-random sampling. This approach is cheaper and often more convenient than conventional surveys, allowing to obtain quick results in a fast-changing environment, but it precludes generalizing results to a target population. This study starts closing some of these gaps.

To the best of my knowledge my study is the first to explore the impacts of the Covid-19 pandemic on grocery shopping in California (changes in channels and frequency of shopping) and how these changes may last after the pandemic is finally over. It covers a longer period of the pandemic (from March 2020 to May 2021) and provides a glimpse at expected post-pandemic grocery shopping at a time (late May 2021) when most hoped that the pandemic would soon be over (which was not to be the case with the emergence of new Covid-19 variants). Most importantly, my results are generalizable to Californians because the data were obtained by surveying California members of KnowledgePanel, the oldest and largest online probability-based panel representative of the U.S. population.

Understanding changes in grocery shopping and particularly residential deliveries from e-grocery is important to logistics managers, to transportation engineers and planners, and to policymakers at a time when California is trying to meet its climate change targets by decreasing energy use in its transportation sector.

This chapter is organized as follows. In the next section, I summarize selected papers about grocery shopping before and during the pandemic. Then I describe how the data were collected, and describe variables and introduce models, before presenting my findings. In the last section, I summarize my contributions, discuss some policy implications, mention some limitations of my work, and propose avenues for future work.

4.2 Background and Literature Review

To inform the modeling choices and contextualize findings, I first review selected papers concerned with trends in grocery shopping before the pandemic and its determinants, before considering selected papers concerned with grocery shopping during the pandemic.

4.2.1 Trends before the pandemic

Before the pandemic, most U.S. households regardless of income did most of their grocery shopping at supermarkets and supercenters (Morrison and Mancino, 2015). Shopping in-person for groceries allows people to experience (touch, smell, and in the case of foods, taste) the products they are considering buying, it exposes them to a broad range of products, and it has a social and recreational dimension. Shoppers are also attracted by the promise of discounts in the form of store coupons, and the social dimension of shopping. However, in-person shopping can be time-consuming, stressful for parents with young children, challenging for people with mobility impairments, and it may result in unhealthy impulse purchases (Devenyns, 2019).

By comparison, online shopping for groceries is much more convenient (it can be conducted anytime and anywhere with an internet connection), and it could help overcome access to nutritious foods experienced by residents of food deserts (provided delivery fees are sufficiently low) (Hand *et al*., 2009; Morganosky and Cude, 2000, 2002). However, online shopping for fresh produce or prepared foods keeps shoppers from experiencing the products they are buying before they receive them, and home deliveries can create time constraints (Saphores and Xu, 2020). As a result, online grocery is sometimes seen as less satisfactory (Morganosky and Cude, 2002).

Before the Covid-19 pandemic, e-grocery was expanding in the U.S., although not as fast as in other parts of the world (Saphores and Xu, 2020): while e-grocery represented almost 4% of grocery sales in the U.S. in 2019 (Begley *et al*., 2020), online food sales in South Korea were 24% of the total food retail market (Korea Agro-Fisheries and Food Trade Corporation, 2020).

Somewhat surprisingly, relatively few academic papers seem to have analyzed the characteristics of grocery shoppers in the U.S. before the pandemic (see Table 4.1). Moreover, the papers I found distinguished between in-store shoppers and online shoppers with home delivery, but few studies analyzed online shopping with store/curbside pick-up (click-and-pick).

Based on the "before Covid-19" papers in Table 4.1, younger and more affluent households were more likely before the pandemic to shop for groceries online and they had more frequent grocery deliveries (Dias *et al*., 2020; Kim and Wang, 2021; Morganosky and Cude, 2000, 2002; Spurlock *et al*., 2020; Suel *et al.*, 2015). Gender, education levels, and the presence of children also played a role (Hand *et al*., 2009; Kim and Wang, 2021; Morganosky and Cude, 2000, 2002; Saphores and Xu, 2020). Other characteristics, such as race, household size, or the number of household workers were significant in some studies. For in-store grocery shopping, women, unemployed status (because homemakers were not considered "employed"), more education, a larger household size, a higher income, and vehicle availability increased the likelihood of in-store grocery shopping, while African Americans and households with more workers were less likely to shop in-store for groceries (Dias *et al*., 2020; Saphores and Xu, 2020).

4.2.2 Grocery shopping during the pandemic

The Covid-19 pandemic and the resulting stay-at-home orders changed grocery shopping habits, priorities, and perceptions (Grashuis *et al.*, 2020; Shamshiripour *et al*., 2020; Wang *et al*., 2020). Health concerns about crowded indoor spaces caused many households to reduce the frequency of their visits to grocery stores and how much time they spent there (Wang *et al*., 2020). Many households began to explore alternatives to in-store grocery shopping and turned online for the first time, especially for essential items (Shamshiripour *et al*., 2020; Wang *et al.*, 2021).

However, restrictions on non-essential travel and a widespread switch to telework caused the percentage of shopping trips to increase as a percentage of all trips (Abdullah *et al*., 2020).

A small but growing literature has explored the adoption of online grocery shopping during the pandemic from different perspectives (Figliozzi and Unnikrishnan, 2021; Grashuis *et al*., 2020; Shamshiripour *et al*., 2020; Wang *et al*., 2021), although only two of these papers (Figliozzi and Unnikrishnan, 2021; Wang *et al*., 2021) analyzed the characteristics of grocery shoppers (bottom half of Table 4.1).

Grashuis *et al*. (2020) conducted an online choice experiment in the U.S. using Amazon's Mechanical Turk to elicit grocery shopping preferences during the pandemic. They explored preferences for shopping alternatives (in-store, online with home delivery, and online with curbside pickup) and their characteristics (time windows, minimum order requirements, and fees) under different scenarios. They reported that preferences for grocery shopping alternatives depend on the evolution of the Covid-19 pandemic, but did not link these preferences with the socio-economic characteristics of participants.

Table 4.1 Summary of selected papers that analyzed factors affecting grocery shopping before and during Covid-19

Reference	Data	Outcome	Significant factors
Before Covid-19			
Kim and Wang (2021)	2018 Mobility Survey (New York City, U.S.)	Frequency of grocery delivery	(+) White household, children, full-time worker; (-) age, car availability.
Dias *et al*. (2020)	2017 Puget Sound Household Travel Survey (greater Seattle region, U.S.)	Frequency of in-store shopping	(+) household size, vehicle availability, building with apartments and condos; (-) income, multiple workers.
		Frequency of grocery delivery	(+) household income, household size, population density, in-person meals; (-) multiple workers, own household tenure
Spurlock *et al*. (2020)	A sample randomly selected addresses in 9 Bay Areas (CA, U.S.)	Frequency of grocery delivery	(+) income
Saphores and Xu (2020)	2017 American Time Use Survey (U.S.)	In store grocery shopping	(+) female, unemployed, education; (-) Gen Z, African American
		Adoption of grocery delivery	(+) female
Suel *et al*. (2015)	Living Costs and Food survey (expenditure diary) (U.K.)	Adoption of grocery delivery	(+) household income; (-) age, household size.
Hand *et al*. (2009)	Qualitative research (90-min focus groups) and large-scale survey (U.K.)	Adoption of grocery delivery	Situational factors: (+) children, health problems, work late, etc.
Morganosky and Cude (2002)♦	Longitudinal data (1998, 1999, 2001 (U.S.))	Adoption of grocery delivery and pick-up	(+) female, education, income; (-) age
Morganosky and Cude (2000)♣	Longitudinal data (1998, 1999 (U.S.))	Primary reason to shop online	(+) income, adults, children; (-) age
		Experience with e-grocery	(+) age; (-) education
During Covid-19			
Wang *et al*. (2021)	Two U.S. surveys on Amazon Mechanical Turk (May and June, 2020)	Adoption of grocery deliveries	(+) male, Hispanic status, home (apartment), number of people over 65 years old, log of population density
		Intention to continue with grocery deliveries	(+) male; (-) age, household income, grocery store density, number of trips on a non-workday
Figliozzi and Unnikrishnan (2021)	Online survey of residents of Portland metro (U.S.) (May and June, 2020)	Change in delivery of groceries	(+) subscription, disability; (-) age

♦ indicates that significant factors are based on an analysis of percentages; ♣ indicates that significant factors are based on a chi-square analysis.

Shamshiripour *et al*. (2020) conducted an online survey of preferences for grocery shopping in the Chicago metropolitan area between 04/25/20 and 06/02/20. The percentage of their respondents who tried online grocery shopping jumped from 20% to 33% because of Covid-19, and 13% relied on e-shopping as their primary way of getting groceries during the pandemic versus only 2% before. Moreover, 59% of their respondents plan on relying at least partly on e-grocery after the pandemic is over. However, they did not report how the socio-economic characteristics of their respondents influenced their results.

After conducting a survey in the greater Portland metropolitan area in May and June 2020, Figliozzi and Unnikrishnan (2021) analyzed the impact of socio-demographic characteristics and attitudes toward e-commerce on home deliveries and expenditures for 7 product categories including groceries during the Covid-19 lockdown. They found that the availability of a subscription for deliveries, age, and disability are the most important factors that affect the likelihood of home delivery for food (grocery and prepared meals).

Wang *et al*. (2021) focused on how long the increase in demand for online shopping would last. After conducting a survey of U.S. residents using Amazon's Mechanical Turk, they reported that male new adopters are more likely to continue grocery shopping after the pandemic. However, older households, people with a higher income, and residents of areas with more grocery stores are less likely to continue the delivery of their groceries after the pandemic is over. Overall, half of the new adopters of grocery deliveries are likely to discontinue this service.

One limitation of the above studies is that survey respondents were not randomly selected, which invites selection bias and makes it risky to generalize findings to a target population. I also note that these surveys were completed during the early stage of the pandemic (typically between May and June 2020), and as of the end of 2021, it is unfortunately still ongoing. Since shopping behavior may be changing along the way, data from more recent periods should be collected to compare grocery shopping trends. Moreover, to the best of my knowledge there is no published study yet that focuses on California, in spite of its importance of this state in the U.S.

4.3 Data

4.3.1 Survey data

Data for this Chapter were collected in late May 2021 via a random survey conducted by IPSOS (the world's third-largest market research company) of California members of KnowledgePanel, the oldest and largest probability-based U.S. panel (IPSOS, 2021). Note that KP members are invited (they cannot just volunteer) using an address-based mail sampling frame based on the Delivery Sequence File of the U.S. Postal Service. Special care is taken to recruit harder-to-reach groups, such as African Americans, Latinos, Veterans, Americans with disabilities, LGBTQ and non-binary people, rural residents, and non-Internet households. A tablet with a mobile data plan is provided to non-Internet households so they can take online surveys.

Surveying with KnowledgePanel offers (KP) several advantages (IPSOS, 2021). First, it helps to address non-response bias because survey cooperation rates (the % of invited panel members who take a survey) typically exceed 70%. Second, it overcomes the self-selection bias inherent in online surveys because respondents are chosen based on their characteristics, which are recorded when they enroll and updated annually. Third, participant fatigue is minimized by ensuring that panelists take no more than two to three KP surveys per month on average.

The questionnaire had two parts. Part I inquired about commuting, telework, and travel before, during, and potentially after the pandemic. In Part II, I explored how Californians shopped for groceries and prepared meals before and during the pandemic, and how that may change after the pandemic is over.

The questionnaire was first written in English and pre-tested on graduate students. A pilot study of 25 California members of KP was conducted by IPSOS in May 2021 to test the survey instrument, which was then modified to incorporate the feedback received. To include Californians who rely primarily on Spanish (according to the U.S. Census Bureau, two-thirds of the ~44% of Californians who speak a language other than English use Spanish (Yarbrough, 2015)), the survey was translated in Spanish and pre-tested it with native speakers. Both versions of the survey were administered in late May 2021. Data collection was stopped at the end of May 2021 after receiving answers from 1,026 respondents.

In Part II of the survey, I asked the respondents how often their households used the following grocery shopping options before and after the March 2020 stay-at-home Executive Order from California's governor: i) In-person grocery shopping in a brick-and-mortar store or a farmers market; ii) Online purchase of groceries with home delivery (e.g., via Amazon Fresh, Instacart, or Costco grocery); iii) Online order of groceries with store pick-up (via drive-thru, in-store pickup, or curbside pickup); and iv) Other (with a request to briefly describe). I also asked them how often their household is likely to use these alternatives after the Covid-19 pandemic is over. In this chapter, I analyzed answers to these questions.

To better capture how grocery shopping evolved as a result of the pandemic, I estimated two sets of models: 1) models explaining the frequency of grocery shopping for different options before, during, and after the pandemic (based on expectations); and 2) models explaining changes in the frequency of different grocery shopping options during versus before, and after versus before the pandemic. For all of these models, I considered options i, ii, and iii above.

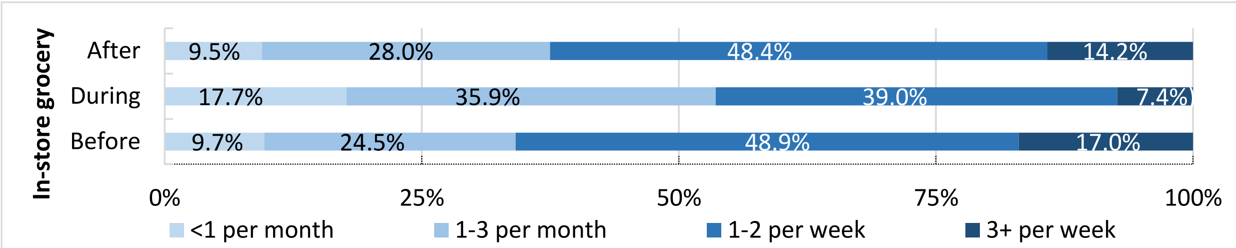
4.3.2 Dependent variables

4.3.2.1 Explaining the frequency of grocery shopping

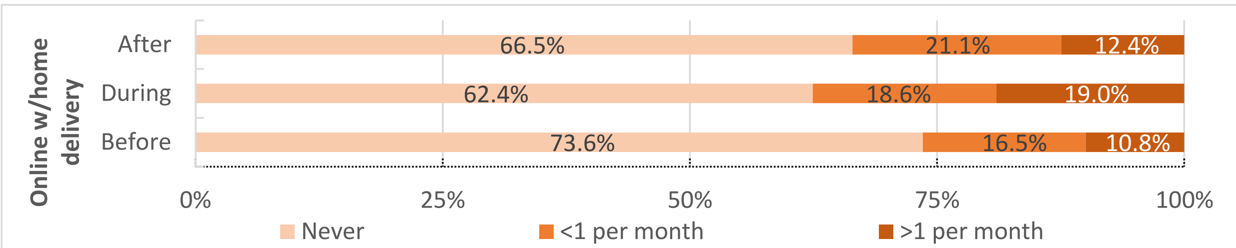
Our survey asked “how often did your household use the following grocery shopping options?” before the March 2020 stay-at-home Executive Order from Governor Newsom, and since the March 2020 stay-at-home Executive Order from Governor Newsom. It also asked “After the Covid-19 pandemic is over, how often do you think your household will use the following grocery shopping options?”. For each of these three questions, six answers were possible: 1. “Never”; 2. “Occasionally but less than once a month”; 3. “1-3 times a month”; 4. “1-2 times a week”; 5. “3 or more times a week”; and 6. “I do not know”. Only a small percentage (up to 4.8% depending on the question) of respondents did not know or refused to answer these questions, so for simplicity, I dropped these observations from models.

For in-store shopping, relatively few (61, or < 6.0%) respondents answered “Never”, which was too small to estimate multivariate models so I merged that category with “Occasionally but less than once a month,” leading to four frequencies: “less than once a month”, “1 to 3 times a month”, “1 to 2 times per week”, and “3 or more times a week”.

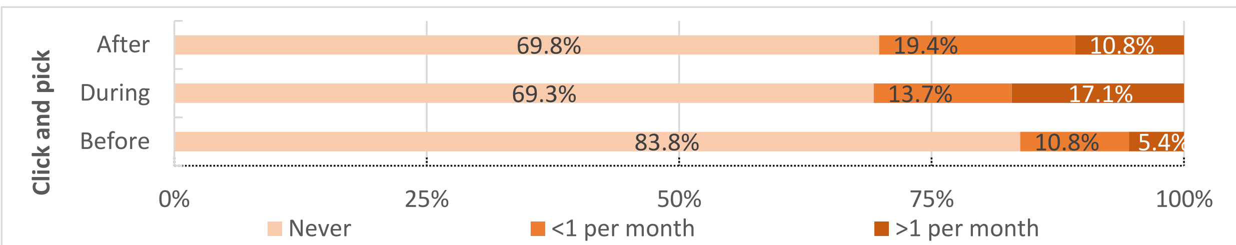
For e-grocery with home delivery and click-and-pick, only small numbers of respondents selected the top two frequencies (“1 to 2 times a week: and “3 or more times a week”) so I merge them with “1-3 times a month” leading to three frequencies: “never”, “less than once a month”, and “more than once per month”. Sankey diagrams showing detailed responses to frequency questions are presented at the beginning of the results section. Summary statistics for dependent variables are shown in Figure 4.1.



Panel A. In-store grocery shopping



Panel B. Online shopping with home delivery (e-grocery)



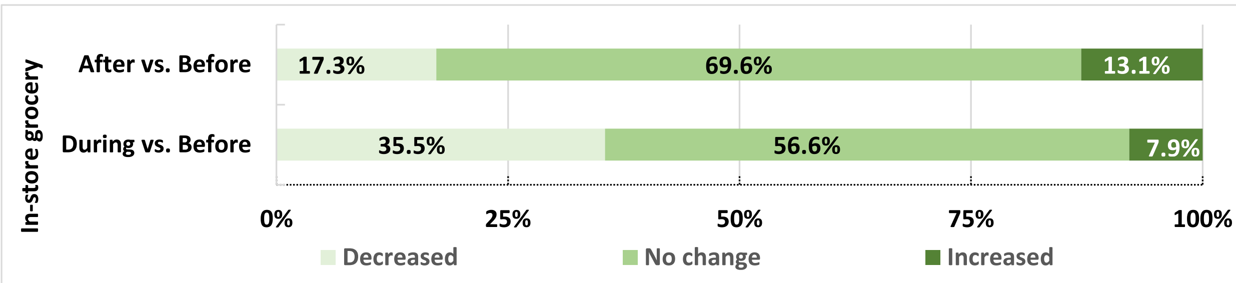
Panel C. Online shopping with store pick-up (click-and-pick)

Figure 4.1 Dependent variables of frequency models

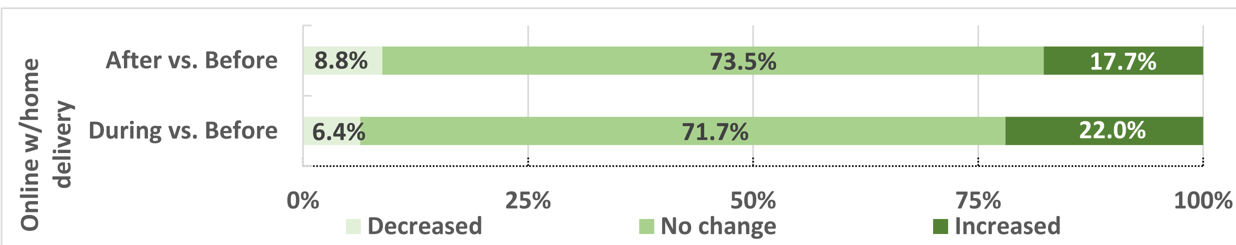
4.3.2.2 Explaining changes in the frequency of grocery shopping

Since the dependent variables for the models that explain the frequency of grocery shopping for different alternatives (in-store, e-grocery, and click-and-pick) have different structures, it is not straightforward to characterize respondents who increased or decreased their frequency of grocery shopping for each alternative. For simplicity, therefore I estimated for each grocery shopping alternative two logit models where the dependent variable equals 1 if the shopping frequency increased compared with before the pandemic, and 0 otherwise: one for during the pandemic, and another for after the pandemic based on what respondents are anticipating. Figure 4.2 shows increases and decreases in shopping frequency for each alternative.

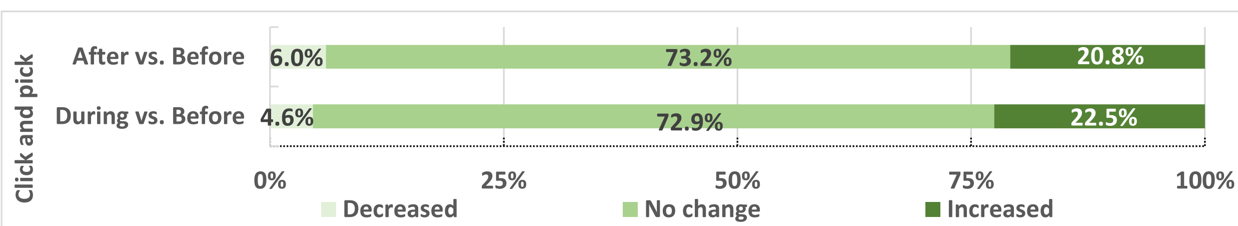
For in-store grocery shopping, I also estimated logit models to explain decreases in frequency, but this was not possible for e-grocery and click-and-pick because very few Californians decreased their grocery shopping frequency for these two alternatives, as shown on Panels B and C of Figure 4.2.



Panel A. In-store grocery shopping



Panel B. Online shopping with home delivery



Panel C. Online shopping with store pick-up (click-and-pick)

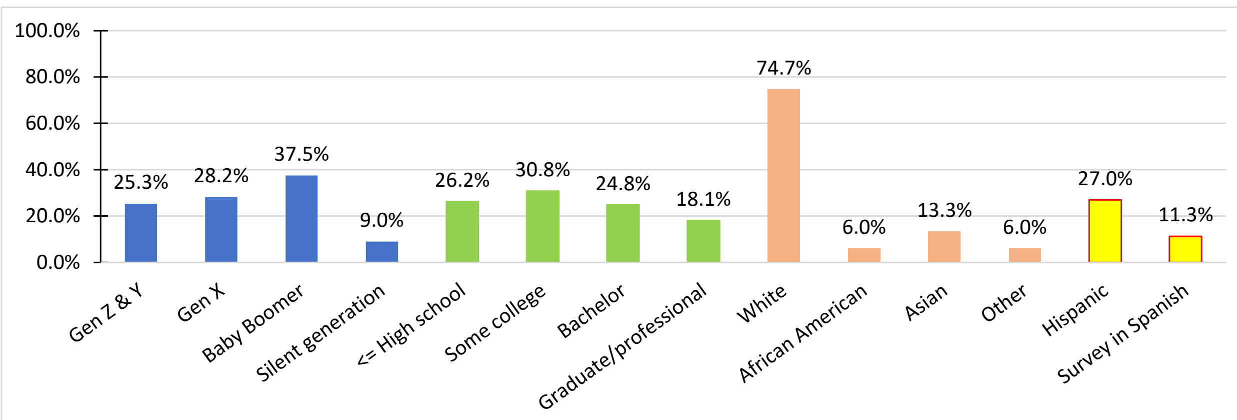
Figure 4.2 Summary statistics for change models dependent variables

4.3.3 Explanatory variables

Explanatory variables for frequency models include socioeconomic characteristics, attitudes toward communication technologies, the number of grocery stores nearby, and for models characterizing grocery shopping frequencies during and after the pandemic, the cumulative number of Covid-19 cases in the county of residence divided by the county population. Out of 1,026 respondents, 44 were lost to missing information, which left with 983 respondents. A few more observations were lost for each model because some respondents did not know or did not answer the question characterizing the dependent variable.

For models explaining changes in the frequency of grocery shopping, I also added variables characterizing changes in the number of hours worked at home, in household income, and in the number of household vehicles.

Figure 4.3 and Table 4.2 show summary statistics for explanatory variables.



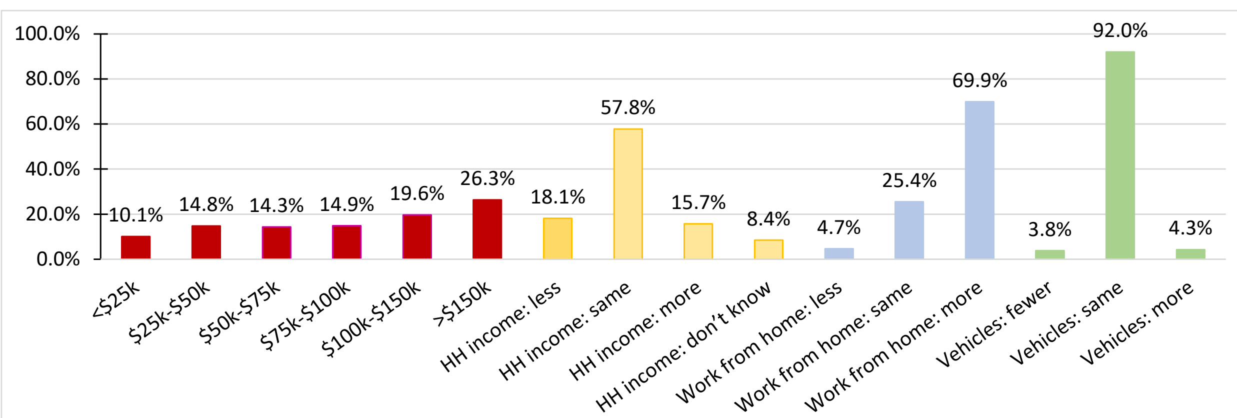


Figure 4.3 Summary statistics for binary explanatory variables (N=983)

Table 4.2 Summary statistics for non-binary explanatory variables (N=983)

Variables	Min	Mean	Max	Std. dev.
Household size	1	2.78	10	1.49
Technological-savviness factor	-1.28	0	4.37	0.87
Log(number of grocery stores within 5 mi of home zip code centroid)	0.00	3.04	7.02	2.25
(Cumulative number of Covid-19 cases in county of residence divided by county population)*100	1.77%	8.54%	17.31%	3.00%

4.3.3.1 Socio-economic characteristics

My literature review showed that age impacts how people shop for grocery (Figliozzi and Unnikrishnan, 2021; Kim and Wang, 2021; Morganosky and Cude, 2002; Saphores and Xu, 2020; Suel *et al*., 2015; Wang *et al*., 2021). To understand generational differences, I created four binary variables based on definitions from the Pew Research Center (Dimock, 2019): Silent Generation (born before 1946), Baby Boomers (born between 1946 and 1964), Gen X (born between 1965 and 1980), and Millennials (or Gen Y, born between 1981 and 1996) & Gen Z (born after 1996).

Education and race also shape grocery shopping behavior (Morganosky and Cude, 2000, 2002; Saphores and Xu, 2020). To capture educational attainment, I relied on four common categories: less than high school, some college or associate degree, Bachelor’s degree, and Master’s degree or higher. Race categories include White, African American, Asian, and other. I also added a binary variable for Hispanics (based on a variable collected by IPSOS), and another one to track respondents who took the Spanish version of the survey.

In addition, explanatory variables include household size as it impacts how much food needs to be purchased (Dias *et al*., 2020; Kim and Wang, 2021; Suel *et al*., 2015).

Household income determines the purchasing power of households and it is widely used in the literature (Dias *et al*., 2020; Kim and Wang, 2021; Morganosky and Cude, 2000, 2002;

Spurlock *et al*., 2020; Suel *et al*., 2015; Wang *et al*., 2021). As shown in Figure 4.3, I organized household income into 6 categories.

To investigate changes in grocery shopping during Covid-19, I also created binary variables that reflect perceived changes in household income ("decreased", "unchanged", "increased", and "don't know"), whether a respondent worked more at home (which would give more time to cook and lead to more grocery purchases), and changes in the number of household vehicles, which may restrict travel to a grocery store (Dias *et al*., 2020).

4.3.3.2 Familiarity with information and communication technologies

With the emergence of online shopping, attitudes toward technology and technological savviness are increasingly important to understand how people shop (Unnikrishnan and Figliozzi, 2021). To capture familiarity with communication technologies, I conducted a factor analysis of 12 questions that IPSOS routinely asks KN members. Details are provided below.

4.3.3.3 Grocery stores within 5 miles

Grocery store availability may affect shopping frequency (Wang *et al*., 2021). To create a measure of availability, I relied on data from the Supplemental Nutrition Assistance Program (Food and Nutrition Service USDA, 2021), which records the address of groceries stores in California. Using GIS, I counted the number of grocery stores within 5 miles of the centroid of the residential ZIP code of each respondent. I selected 5 miles because a USDA study (Ver Ploeg *et al*., 2015) reported that Americans travel on average 3.9 mi to reach their preferred grocery store, even though the nearest one is usually within 2 mi. The extra mile reflects the uncertainty about the exact home location of the respondents (I know only their ZIP code).

4.3.3.4 Covid-19 severity in the county of residence

Where and how frequently Californians shopped for groceries may have been impacted by the severity of the Covid-19 epidemic, in addition to public health restrictions. For each respondent, I therefore included in models the ratio of the cumulative count of Covid-19 cases (until mid-May 2021) in their county of residence divided by that county's population. The data come from California's Covid-19 portal (https://covid19.ca.gov/data-and-tools/#Databases). I also considered a variable based on the cumulative number of deaths from Covid-19, but it was highly correlated with the variable based on the number of cases.

4.4 Models

4.4.1 Explaining the frequency of grocery shopping

Since the models explaining the frequency of grocery shopping in California include an endogenous variable that reflects how savvy respondents are with communication technologies, it would seem natural to rely on generalized structural equations modeling (GSEM). However, the GSEM software I am familiar with cannot handle ordered dependent variables. For simplicity, I therefore estimated the "technological savviness" factor using exploratory factor analysis and included the predicted value of that factor in ordered models.

4.4.1.1 Factor Analysis

To condense answers to the twelve questions (see Table 4.3) about attitudes towards communication technology that IPSOS asks KN members (this information is updated annually), I relied on exploratory factor analysis. Responses to these questions were on a four-point Likert scale: 1 "Do not agree", 2 "Somewhat agree", 3 "Agree", and 4 "Strongly agree".

To select the number of factors, I adopted the Kaiser criterion (I retained only factors corresponding to eigenvalues above 1 (Fabrigar and Wegener, 2012)), which yielded one factor.

To assess its validity, I applied common statistical tools: the Bartlett test for sphericity, the Kaiser-Meyer Olkin (KMO) statistic, and Cronbach's alpha (α). Bartlett's test of sphericity checks whether the correlation matrix of the variables differs significantly from the identity matrix; if not (if it fails to reject the null hypothesis of the Bartlett test), the factor is inappropriate. The KMO statistic (Kaiser and Rice, 1974) measures the proportion of the variance that may be common to the variables considered for factorization; a lower proportion is better and leads to a higher KMO value. An adequate KMO (KMO ranges between 0 and 1) value should exceed 0.8, although values as low as 0.6 are acceptable. Finally, Cronbach's alpha (which also ranges between 0 and 1) indicates how well a set of variables measures a single underlying construct; α values below 0.5 are usually deemed unacceptable, with desired values above 0.65. Finally, factor loading estimates the strength and direction of the influence of the common factors on the measured variable (Fabrigar and Wegener, 2012). I kept variables with a factor loading greater than 0.3, which led to exclude Question 10 ("I like to buy electronics or technology from a physical retail store").

Table 4.3 Questions about comfort with technology

Question	Factor loading
1. Others rely on me for advice about technology	0.5879
2. I often buy a new technology or, device, as soon as it goes on sale	0.5978
3. I like surfing the internet for fun	0.4903
4. I tend to watch less TV on a traditional television because I watch video online	0.4571
5. I like to post online video content that I create (such as on YouTube)	0.5262
6. I use social networking to communicate with others more than email and instant messenger	0.5120
7. I am fine with advertising on mobile phones	0.4154
8. I would pay to watch a TV show or movie to avoid commercials	0.3817
9. I have had to delay some technology purchases because I didn't have the money	0.4140
11. I like to buy technology brands that are environmentally friendly	0.3865
12. I always buy the lowest priced electronics or technology	0.4133

Table 4.3 shows the component of technology savviness with the corresponding factor loadings. I rejected the null hypothesis for Bartlett's test for sphericity (p-value<0.001), obtained a KMO value of 0.838 (adequate), and calculated a Cronbach's alpha value of 0.75. These results jointly validate the "technological savviness" factor.

4.4.1.2 Ordered Models

To explain ordered limited dependent variables, a simple starting point is an ordered logit model (Long, 1997). Assuming there are M possible choices, the probability that respondent *i* choses an alternative higher than $m \in \{1, \ldots, M\text{-}1\}$ is given by

$$\Pr(Y_i > m) = \frac{\exp(\boldsymbol{X}_i\boldsymbol{\beta}' - \tau_m)}{1 + \exp(\boldsymbol{X}_i\boldsymbol{\beta}' - \tau_m)} \tag{1}$$

where $\boldsymbol{X}_i$ is a vector of observed explanatory variables, $\boldsymbol{\beta}$ is a vector of unknown parameters, and the τ_ms (m=1, …, M-1) are unknown thresholds to estimate jointly with $\boldsymbol{\beta}$.

For respondent *i* and alternative $m \in \{1, \ldots, \text{M-1}\}$, the odds Ω_{im}, is defined by

$$\Omega_{im} = \frac{\Pr(Y_i > m)}{\Pr(Y_i \leq m)} = \exp(\boldsymbol{X}_i\boldsymbol{\beta}' - \tau_m), \tag{2}$$

so the ratios of the odds with another respondent, denoted by *j* is

$$\frac{\Omega_{im}}{\Omega_{jm}} = \frac{\exp(\boldsymbol{X}_i\boldsymbol{\beta}' - \tau_m)}{\exp\left(\boldsymbol{X}_j\boldsymbol{\beta}' - \tau_m\right)} = \exp(\boldsymbol{X}_i\boldsymbol{\beta}' - \boldsymbol{X}_j\boldsymbol{\beta}'), \tag{3}$$

which does not depend on alternative *m*. This property is called the "proportional odds" (or "parallel line") assumption, and it is an implication of the ordered logit model. However, Brant tests (Brant, 1990) showed that the proportional odds assumption does not hold here, so I relied instead on generalized ordered logit models (GOL) (Peterson and Harrell, 1990), where the βs can depend on the alternative considered, so that

$$\Pr(Y_i > m) = \frac{\exp(\boldsymbol{X}_i\boldsymbol{\beta}'_m - \tau_m)}{1 + \exp(\boldsymbol{X}_i\boldsymbol{\beta}'_m - \tau_m)}, \tag{4}$$

In the GOL model, the explanatory variables for which the proportional odds assumption holds have the same βs for all alternatives, and different βs otherwise, yielding a more concise model.

With *M* alternatives, a GOL model requires estimating *M*-1 equations., which can be seen as a series of cumulative logistic regressions (Williams, 2016), so results can be interpreted using odds ratios. Indeed, for respondent *i* and alternative $m \in \{1, \ldots, M\text{-}1\}$, the odds Ω_{im} for a GOL model is given by

$$\Omega_{im}(\boldsymbol{X}_i) = \frac{\Pr(Y_i > m)}{\Pr(Y_i \leq m)} = \exp(\boldsymbol{X}_i\beta'_m - \tau_m), \tag{5}$$

so $\Omega_{im}(\boldsymbol{X}_i, x_l + 1)$ denotes the odds obtained by increasing variable *l* by one unit, I have:

$$\Omega_{im}(\boldsymbol{X}_i, x_l + 1) = \frac{\Pr(Y_i > m)}{\Pr(Y_i \leq m)} = \exp(\boldsymbol{X}_i\beta'_m + \beta_{lm} - \tau_m). \tag{6}$$

Combining Equations (5) and (6) gives the expression of the odds ratio for variable *l*:

$$OR_{lm} = \frac{\Omega_{im}(\boldsymbol{X}_i, x_l + 1)}{\Omega_{im}(\boldsymbol{X}_i)} = \exp(\beta_{lm}). \tag{7}$$

Since higher values of the dependent variable correspond to higher shopping frequencies for each alternative considered, I see from Equation (7) that a higher value of β_{lm} increases the odds that a respondent will shop for groceries more frequently with that alternative. I therefore reported results using odds ratios after estimating unknown parameters ($\beta_m s$ and τ_ms) via maximum likelihood.

4.4.2 Explaining changes in the frequency of grocery shopping

To explain changes in grocery shopping frequency during versus before and after versus before the pandemic, I adopted generalized structural equation modeling (GSEM) (Kline, 2015). The structure of the conceptual generalized structural equation model is shown in Figure 4.4. An arrow represents a causal relationship between exogenous (independent) variables and an endogenous (response) variable. I assume that an increase (decrease) in grocery shopping frequency for one of the three alternatives considered (in-store, online shopping with home delivery, or click-and-pick) can be explained by socio-economic characteristics of the respondent, changes in the time worked at home, changes in household income and in the number of household vehicles, the cumulative (from the start and up until the survey date) number of Covid-19 cases in the county of residence divided by the county population, the number of grocery stores within five miles of the centroid of a respondent's ZIP code, and a latent technological savviness factor generated from the 11 questions on technology attitudes retained from exploratory factor analysis.

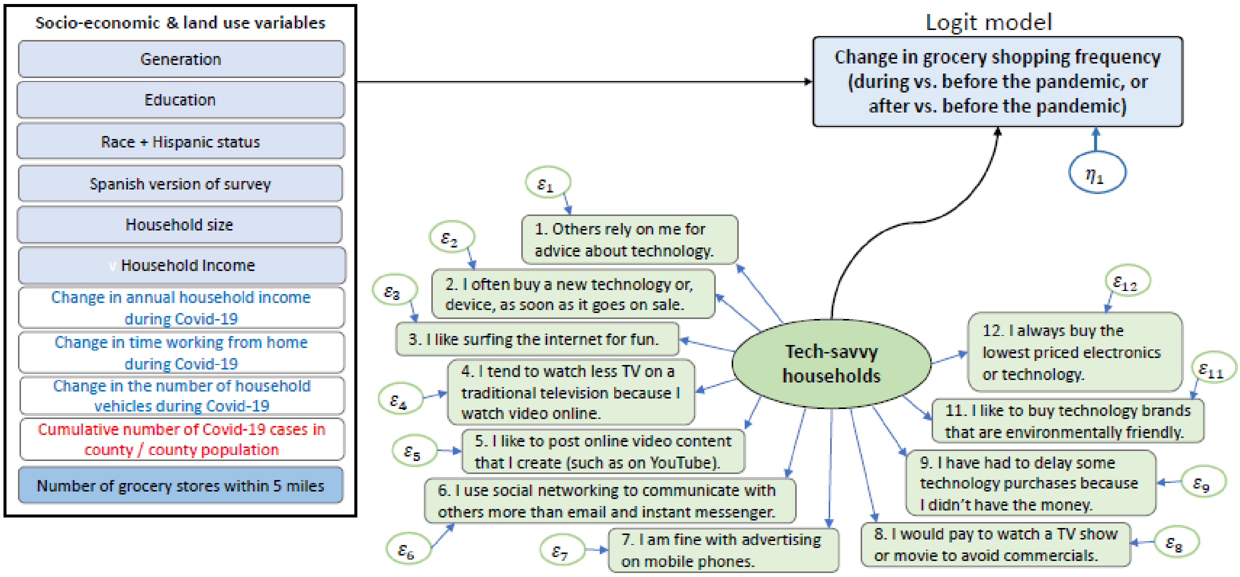


Figure 4.4 Conceptual model for grocery shopping frequency change

This model can be written:

Regression equations for latent technology savviness factor ($j \in \{1, \ldots, 12\} \setminus \{10\}$):

$$\boldsymbol{Y_j} = \boldsymbol{T\Lambda_j} + \boldsymbol{\varepsilon_j}, \tag{8}$$

Logit model for change in grocery shopping frequency:

$$\boldsymbol{S_i} = \begin{cases} \mathbf{1} \text{ if } \boldsymbol{S_i^*} > \mathbf{0}, \\ \mathbf{0} \text{ if } \boldsymbol{S_i^*} \leq \mathbf{0}, \end{cases} \quad \boldsymbol{S^*} = \boldsymbol{X\Gamma_1} + \boldsymbol{T\Gamma_2} + \boldsymbol{\eta_1}, \tag{9}$$

In the above:

- $\boldsymbol{Y_j}$ is an n × 1 vector of the 11 variables used to estimate the latent technological savviness variable $\boldsymbol{T}$;
- $\boldsymbol{S}$ is an n × 1 vector of 0s and 1s; $\boldsymbol{S}_i$=1 if the frequency of grocery shopping increased (decreased) during the pandemic and 0 otherwise;
- $\boldsymbol{X}$ is an n × p_k matrix of personal, household, and land use characteristics and Covid-19 severity around places of residence;

- $\Lambda_1, \Lambda_2 \ldots\ldots\ldots \Lambda_{12}$, $\boldsymbol{\Gamma}_1$, and $\boldsymbol{\Gamma}_2$ are unknown model parameters to estimate; and
- $\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_{12}$, and $\boldsymbol{\eta}_1$ are n × 1 error vectors.

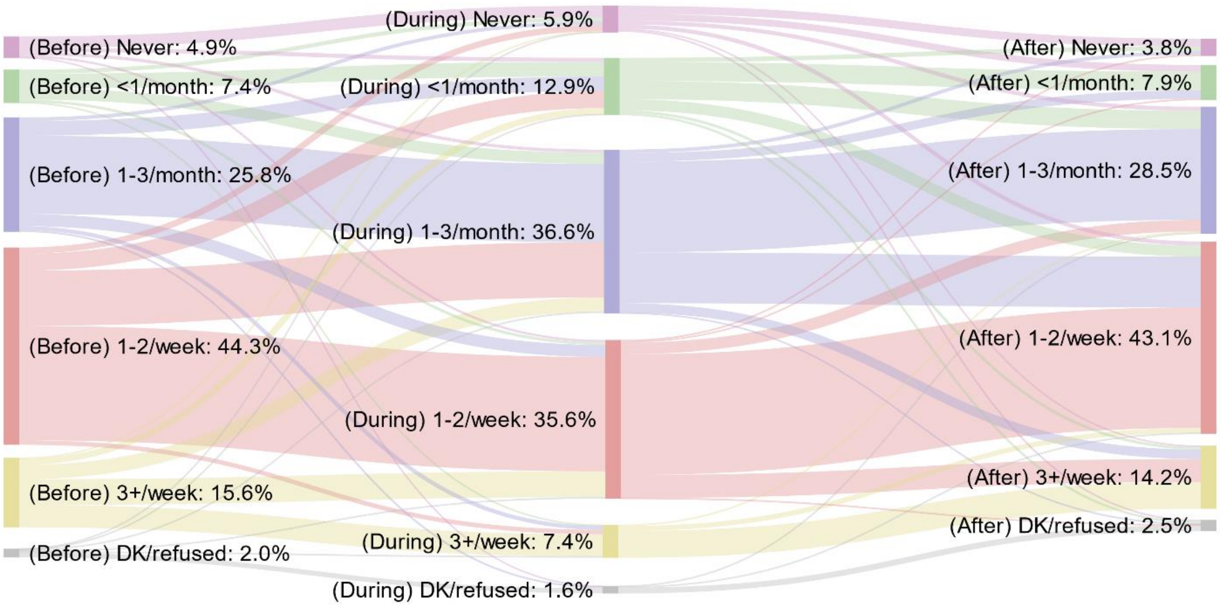
Since the model is recursive, it is identified (Kline, 2015). Unknown model parameters were estimated by minimizing the difference between the sample covariance and the covariance predicted by the model (Bollen, 1989).

4.5 Results

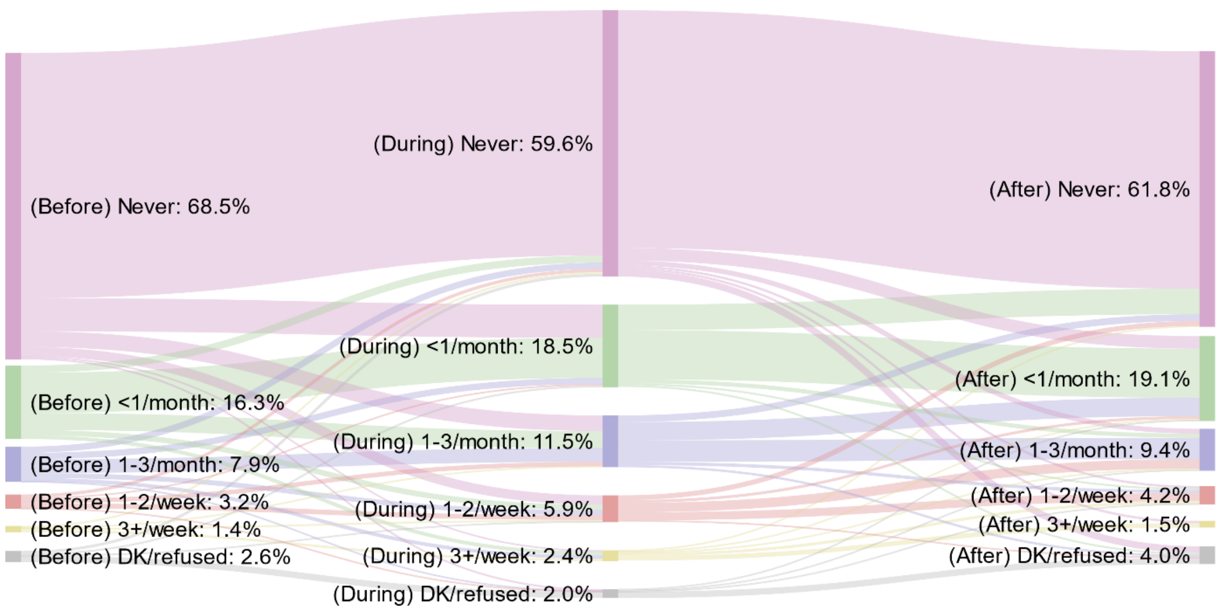
Results were obtained with Stata 17.0. Before estimating models, I checked for multicollinearity using variance inflation factors (VIF). Maximum VIF values were all below 3, so multicollinearity is not an issue here.

4.5.1 A graphical overview

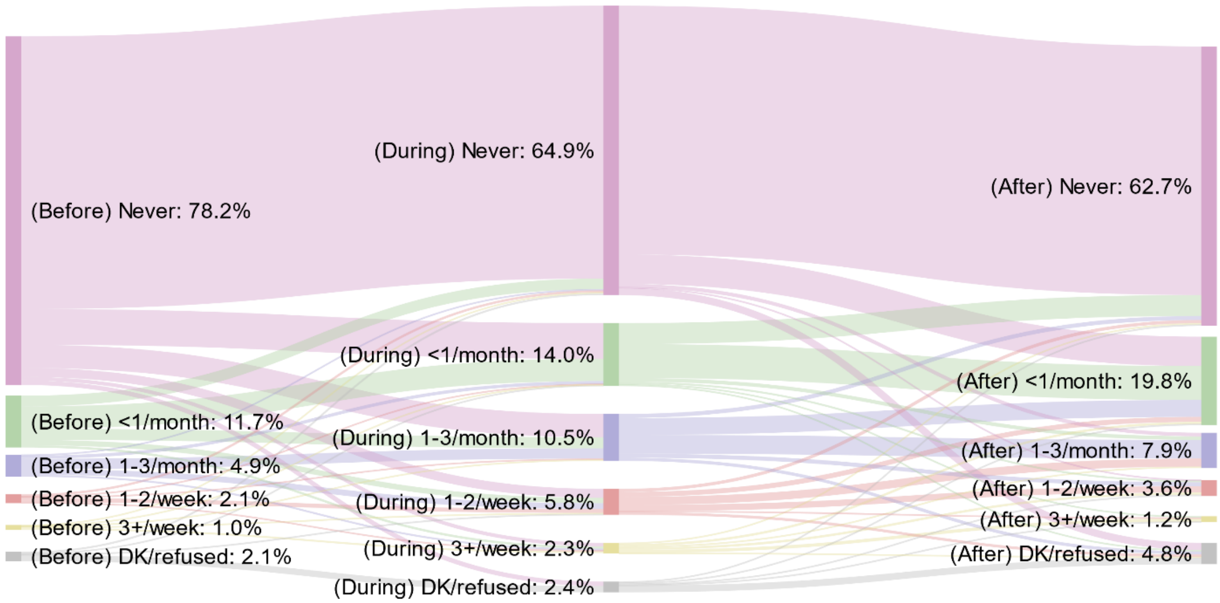
To match the sample to the California population, IPSOS calculated sample weights obtained by raking the following distributions of Californians aged 18 and over from the 2019 American Community Survey: gender by age, race and Hispanic status, education, household income, and language proficiency (for English and Spanish). I used these weights to calculate what percentage of Californians engaged in different forms of grocery shopping (in-store, e-grocery, and click-and-pick) and with what frequency before, during, and after the pandemic, and displayed the information on Sankey diagrams (see Figure 4.5).



Panel A. In-store shopping frequency changes



Panel B. E-grocery frequency changes



Panel C. Click-and-pick frequency changes

Figure 4.5 Grocery shopping changes in California before, during, and after the pandemic

Notes. "DK" stands for "I don't know". The thickness of a flow is proportional to the percentage of people who moved from one frequency to another between two consecutive periods.

Starting with in-store shopping (Panel A), I see that only a small percentage of Californians stopped grocery shopping in-person during the pandemic (the "never" category increased from 4.9% before to 5.9% during). Instead, Californians simply went to grocery stores less frequently, as shown by the change in flows from the two largest frequencies (weekly grocery shopping). Indeed, during the pandemic, the percentage for "1-2/week" dropped from 44.3% to 35.6%, and from 15.6% to 7.4% for "3+/week", while the frequency for "<1/month" rose from 7.4% to 12.9%, and the frequency for "1-3/month" jumped from 25.8% to 36.6%. After the pandemic, many of the respondents expect to shop for groceries as frequently as before, with only small decreases compared to before for the highest frequencies (14.2%-15.6%=-1.4% for "3+/week", and 43.1%-44.3%=-1.2% for "1-2/week"), and small increases for the lower ones

(28.5%-25.8%=+2.7% for “1-3/month” and 7.9%-7.4%=+0.5% for “<1/month”). I even note that the percentage of Californians who never shop for groceries in a store is expected to drop from 4.9% to 3.8%, although there is a slight uptick in non-respondents.

This does not mean, however, that the pandemic will not have lasting changes on grocery shopping in California (Panels B and C). The most significant pandemic change is that the pandemic motivated many households to try new ways to shop for groceries during the pandemic: the gain was a solid 8.9% for e-grocery (a 68.5%-59.6% drop in those who never tried it), and even higher (+13.3%) for click-and-pick (78.2%-64.9%). Likewise, households already familiar with these alternatives used them both more often across the board, with the largest gains for monthly shopping (e.g., +3.6% for “1-3/month” for e-grocery, and +5.6% for “1-3/month” for click-and-pick). Most importantly, most of the gains for e-grocery and especially click-and-pick appear here to stay, with all expected after-pandemic frequencies higher than pre-pandemic frequencies, except for “never”, which rebounds slightly for e-grocery (61.8% after versus 59.6% during, but much better than 68.5% before) but keeps on improving for click-and-pick (62.7% after versus 64.9% during and 78.2% before). This is good news for many traditional grocers who invested in their websites for e-shopping and in vehicles and warehouses for home deliveries (Redman, 2021).

These shifts did not uniformly affect all Californians, however, as revealed by my multivariate models.

4.5.2 Results for shopping frequency models

Table 4.4 presents frequency model (generalized ordered logit) results, obtained with robust standard errors, for in-store grocery shopping, e-grocery, and click-and-pick for before, during, and after the pandemic. As mentioned above, for the in-store grocery shopping, I was able to consider four shopping frequencies (“<1/month” for less than once a month, “1-3/month” for one to three times a month, “1-2/week” for once to twice times a week, and “>2/week” for more than twice a week) (M=4) but I was able to analyze only three frequencies for e-grocery and click-and-pick (“Never”, “<1/month”, and “>1/month”) (M=3) because weekly frequencies were much less common. A “•” indicates that a coefficient was not statistically significant, and numbers correspond to odds ratios as described by Equation (7).

Counting “•” as a (non-significant) value, a cell in Columns I-III contains either one or three values. If there are three values, they correspond respectively to the odds ratios for 1) More than “<1/month” (which is “≥1 month”) versus “<1/month”; 2) More than “1-3/month” (which is “>3/month”) versus up to 3 times a month; and 3) More than “1-2/week” (which is “>2/week”) versus up to twice a week. If a cell in Columns I-III contains a single value, then all three of these odds ratios are the same. Likewise, Columns IV-IX contain either one or two values. If there are two values, they correspond respectively to the odds ratios: 1) more than “never” versus “never”; and 2) more than “<1/month” (which is “≥1/month) versus “<1/month”.

Since Table 4.4 shows odds ratios, I should recall that a value around one indicates that the corresponding variable has little practical impact on shopping frequencies, even if it is statistically significant. Conversely, a value over one indicates the shopping frequency considered will tend to increase with an increase of the corresponding explanatory variable, while a value below one means that the shopping frequency considered will tend to decrease with an increase of the corresponding explanatory variable.

Table 4.4 Frequency models (generalized ordered logit models) results

	In-store (N=961)			E-grocery (N=942)			Click-and-pick (N=937)		
Column #	I	II	III	IV	V	VI	VII	VIII	IX
Variable	**Before**	**During**	**After**	**Before**	**During**	**After**	**Before**	**During**	**After**
Generation (baseline=Baby boomers)									
Gen Z & Y	0.457‡	•	0.578‡	1.450*	•	•	1.663*	1.982‡	2.318‡
Gen X	0.642‡	•	0.667†	1.412*	•	•	1.571*	1.814‡	2.152‡
Silent gen	0.624†	•	•	•	0.555*	•	•	0.544*/•	•
Education (baseline=BA / BS)									
≤high school	0.514†/0.608†/•	•	0.278‡/0.536‡/•	•	•	•	•	•	•
Some college	•	•	0.447†/0.717*/•	•	•	•	•	•	•
Grad./prof.	•	0.642†	•	•	1.415*	•	•	•/1.777†	•
Race (Baseline=White)									
Black	0.373‡/0.467‡/•	•	0.511†	•	•	1.648*	•/2.806†	•	•
Asian	0.385‡/0.652†/•	•	0.692*	•	•	•	•	•	•
Other	•	•	•	•	•	•	•	•	•
Hispanic	0.684*	•	•	•	•	•	1.712†	•/1.646*	•
Spanish survey	1.634*	1.606*	•	•	•	•	•	•	•
HH size	•	•	•	1.093*	•	•	1.186‡	1.102*	1.090*
Annual household income (baseline=$50k-$75k)									
<$25	•	•/•/3.007‡	•/•/1.851*	•	•	•	•	•	•
$25k-$50k	•	•	•	•	•	•	•	0.600*	•
$75k-$100k	1.513*	•	1.607†	•/0.357†	•	•	•	•	•
$100k-$150k	•	1.690†	1.604†	•	•	•	•	•	•
>$150k	1.635†	•/1.887‡/•	•/•/2.219‡	•	•	•	•	•/0.557*	•
Tech-savvy	•	•	•	1.337‡	1.456‡	1.571‡	1.805‡	1.737‡/1.362‡	1.626‡
Log(# grocery stores in 5 mi)	•	•	•	•	•	•/1.130*	•	•	•
Covid severity	NA	•	•	NA	•	•	NA	•	•

1. ‡, †, and * indicate statistical significance at the 0.01, 0.05, and 0.1 levels respectively.
2. "NA" indicates that a variable was not included in the corresponding model.
3. Cells that show only one coefficient have the same coefficients across different frequencies.
4. For in-store shopping, I modeled four shopping frequencies ("<1 a month", "1-3 times a month", "1-2 times a week", and "≥3 times a week:) vs. only three for e-grocery and click-and-pick ("Never", "<1 a month", ">1 a month").

Starting with in-store shopping, I see that compared to households with Baby boomers, households with respondents who belong to Gen Z & Y (0.457‡), Gen X (0.642‡), and the Silent generation (0.624†) shopped in grocery stores less frequently before the pandemic. Although generational differences seem to vanish during the pandemic, they reappear after for Gen Z & Y (0.578‡) and Gen X (0.667†), but not for households with older members. Looking at e-grocery (Columns IV-VI) and click-and-pick (Columns VII-IX), I see that younger generations embraced more online shopping, while households with older members just appear to shop less frequently during the pandemic and to behave like Baby boomers before and after the pandemic.

Indeed, before the pandemic, households with Gen Z & Y or Gen X members were more likely to engage in e-grocery (1.450* and 1.412* respectively) and click-and-pick (1.663* and 1.571* respectively). While these generational differences disappeared during the pandemic for e-grocery, they continued for click-and-pick (1.982‡ for Gen Z & Y and 1.814‡ for Gen X), while households with older members used both e-grocery (0.555*) and click-and-pick (0.544* but only for lower frequencies) less than other generations. After the pandemic, generational differences appear to vanish for e-grocery but not for click-and-pick as households with Gen Z & Y (2.318‡) and Gen X (2.152‡) members stated their intention to rely on click-and-pick much more than older households. These results are not surprising because members of younger generations are more used to relying on technology in their everyday life (Figliozzi and Unnikrishnan, 2021; Wang *et al*., 2021). Moreover, click-and-pick does not require a delivery fee, which is one of the major concerns about the home delivery of groceries (Figliozzi and Unnikrishnan, 2021; Punel and Stathopoulos, 2017; Shamshiripour *et al*., 2020), and it does not impose to be present at home during a delivery window.

Education also matters, but mostly for in-store grocery shopping. Before the pandemic, households with respondents with a high-school education or less were less likely to shop in grocery stores (0.514†/0.608†/•, so not for the highest frequency) than more educated households. Although this difference vanished during the pandemic, respondents expect it to come back after (0.278‡/0.536‡/•), and to also affect households with some college education (0.447†/0.717*/•, again not for the highest frequency). Conversely, households with respondents with a graduate or a professional degree shopped less in stores for grocery (0.642†) during the pandemic. Instead, they resorted more to e-grocery (1.415*) and click-and-pick (•/1.777†), possibly to avoid crowds, which is a top concern during the pandemic (Shamshiripour *et al.*, 2020; Wang *et al.*, 2020).

Interestingly, results suggest that race is not influencing grocery shopping during the pandemic, although it did before and to a lesser extent it may after. Before the pandemic, African Americans (0.373‡/0.467‡/•) and Asians (0.385‡/0.652†/•) seemed less involved in lower frequencies of in-store shopping (confirming results in Saphores and Xu, 2020), with the former possibly compensating with more click-and-pick (•/2.806†). After the pandemic, differences in in-store grocery shopping seem poised to reappear for both (0.511† for African Americans, 0.692* for Asians), although the former may compensate with more e-grocery (1.648*). Although these variables are broadly used in social sciences, I readily acknowledge that they mask substantial heterogeneities.

Results for the Hispanic variable and for the Spanish survey indicator suggest that Hispanics are far from homogeneous. Although Hispanics on average frequented brick-and-mortar grocery stores less (0.684*) than non-Hispanics before the pandemic, the reverse was true for respondents who took the Spanish version of the survey (1.634*), including during the pandemic (1.606*). Hispanics also relied more than non-Hispanics on click-and-pick before (1.712†) and during the pandemic (•/1.646*), although this difference vanishes after the pandemic. I noted no such differences for respondents who took the survey in Spanish.

An increase in household size does not seem to impact the frequency of in-store grocery shopping, although it increased e-grocery frequency before the pandemic (1.093*) and positively impacts click-and-pick frequencies before (1.186‡), during (1.102*), and after the pandemic (1.090*).

Household income influences shopping potential, but its impact on grocery shopping is non-linear. Before the pandemic, compared to those with an annual income of $50k-$75k, households with annual incomes of $75k-$100k or over $150k were likely to shop more frequently in stores for groceries (1.513* and 1.635† respectively). During the pandemic, this shifted to the top two income brackets, but also to the bottom one (•/•/3.007‡), possibly as these customers were trying to benefit from specials or store coupons. And after the pandemic, most income groups expect to go back to stores more than the baseline group (see Column III in Table 4.4). Conversely, for both e-grocery and click-and-pick, income differences just result in a handful of adjustments, which suggest that income is not the main driver behind the adoption of e-grocery and click-and-pick in California.

Instead, as expected (also see Unnikrishnan and Figliozzi, 2021), a higher level of familiarity with information and communication technologies results in a greater adoption of both e-grocery (1.337‡ before, 1.456‡ during, and 1.571‡ after the pandemic) and click-and-pick (1.805‡ before, 1.737‡/1.362‡ during, and 1.626‡ after Covid-19).

Unlike in Grashuis *et al*. (2020), the indicator of the severity of the pandemic does not seem to have an impact on the frequency of and the channel for grocery shopping in California, possibly because it was overshadowed by restrictions adopted by the state authorities at various stages of the pandemic. Likewise, the number of grocery stores within 5 miles of the centroid of the residential ZIP code of respondents is not statistically significant (with one small exception).

4.5.3 Results for frequency change models

Results for grocery shopping frequency change models are presented in Table 4.5. They were obtained using quasi-maximum likelihood with the Huber-White sandwich estimator to relax the assumption that errors are identically and normally distributed (Rabe-Hesketh *et al*., 2004), as many of explanatory variables are binary. Since I relied on a logistic model to explain an increase (or a decrease) in the frequency of grocery shopping, Table 4.5 shows odds ratios.

First, I observe some generational effects. Indeed, households with respondents from Gen Z & Y (1.897*) and from Gen X (2.032†) increased more the frequency of their in-store grocery shopping during the pandemic compared to Baby Boomers (our baseline), and they were much more likely to decrease it less (0.623† and 0.669†, respectively). However, these differences vanish after the pandemic. Likewise, compared to Baby boomers, households from Gen Z & Y respondents increased more the frequency of their use of click-and-pick during the pandemic versus before (1.482*), and they are planning to do the same once the pandemic is over (2.080‡), along with Gen X households (1.774†). Concurrently, compared to Baby boomers, households with members from the Silent generation decreased less the frequency of their grocery shopping trips during the pandemic versus before (0.596*), a trend that will likely continue after (0.285†). They were also less likely to increase more their use of click-and-pick during the pandemic (0.474*). Interestingly, there are no significant generational difference differences for e-grocery.

Overall, these results confirm the increasing interest for click-and-pick, especially by younger generations, and the digital divide.

Second, education also impacted changes in grocery shopping. Households with graduate/professional adults decreased more their frequency of in-store shopping during the pandemic (1.555†), and increased e-grocery (1.697†) more than other households; they are also likely to decrease its frequency less after the pandemic (0.408†). Conversely, households from adults with only some college education are likely to increase their e-grocery frequency less (0.631*) after the pandemic. While click-and-pick was widely adopted during the pandemic, households with less education (≤high school), were less likely to use it more (0.590*).

Third, I find no difference in grocery shopping frequency change for Black households compared to Whites. However, households with Asian members increased their in-store shopping frequency more during the pandemic (2.008*) and will likely do the same after (1.694*). There are no race differences in Table 4.5 for e-grocery and click-and-pick. Just as in Table 4.5, results for Hispanics and for the survey in Spanish indicator reflect the diversity of Hispanics in California: while Hispanics are more likely to decrease less their frequency of in-store shopping (0.563*) and decrease more their use of click-and-pick (1.971*) after the pandemic, Californians who are primarily Spanish speakers are planning on doing the opposite (2.423† and 0.348* respectively).

Table 4.5 shows no impact of household size on grocery shopping frequency changes, and no clear trends linked to annual household income.

Table 4.5 Results of models of change (generalized structural equation models)

	In-store (N=961)				E-grocery (N=942)			Click-and-pick (N=937)		
	During-before (D-B)		After-before		D-B	After-before		D-B	After-before	
Column #	I	II	III	IV	V	VI	VII	VIII	IX	X
Variable	**Increase**	**Decrease**	**Increase**	**Decrease**	**Increase**	**Increase**	**Decrease**	**Increase**	**Increase**	**Decrease**
Generation (baseline=Baby boomers)										
Gen Z & Y	1.897*	0.623†	•	•	•	•	•	1.482*	2.080‡	•
Gen X	2.032†	0.669†	•	•	•	•	•	•	1.774†	•
Silent gen	•	0.596*	•	0.285†	•	•	•	0.474*	•	•
Education (baseline=BA / BS)										
≤high school	•	•	•	•	•	•	•	0.590*	•	•
Some college	•	•	•	•	•	0.631*	•	•	•	•
Grad./prof.	0.343†	1.555†	•	•	1.697†	•	0.408†	•	•	•
Race (Baseline=White)										
Black	•	•	•	•	•	•	•	•	•	•
Asian	2.008*	•	1.694*	•	•	•	•	•	•	•
Other	•	0.559*	0.222†	•	•	•	•	•	•	•
Hispanic	•	•	•	0.563*	•	•	•	•	•	1.971*
Spanish survey	•	•	•	2.423†	•	•	•	•	•	0.348*
HH size	•	•	•	•	•	•	•	•	•	•
Annual household income (baseline=$50k-$75k)										
<$25	•	0.437‡	•	•	•	•	•	•	0.440†	•
$25k-$50k	•	•	•	•	•	•	•	0.474†	•	•
$75k-$100k	0.280‡	•	•	•	•	•	•	•	•	•
$100k-$150k	•	•	•	•	•	•	•	•	•	•
>$150k	•	•	0.401†	•	•	•	•	•	•	•
Change in household income (baseline=no change)										
Decrease	•	1.542†	•	1.510*	1.576†	1.706†	•	1.640†	1.644†	•
Increase	•	•	•	0.595*	•	•	•	•	•	•
Does not know	•	•	•	•	•	•	•	•	•	•
Change in work from home (baseline=no change)										
Decrease	•	•	•	•	•	2.066*	•	•	•	•
Increase	•	1.460†	•	•	1.615†	1.575†	•	•	1.398*	•

Change in the # of household vehicles (baseline=no change)										
Decrease	2.195*	•	•	•	•	•	2.287*	•	•	•
Increase	•	•	•	•	•	•	•	•	•	0.203*
Tech-savvy	•	1.422†	•	1.451†	1.681‡	1.698‡	•	1.777‡	1.771‡	1.956†
Log(# grocery stores in 5 mi)	•	1.078*	•	•	•	•	•	•	•	•
# Covid cases	•	•	•	•	•	•	•	•	•	•

1. This table shows odds-ratios.
2. ‡, †, and * indicate statistical significance at the 0.01, 0.05, and 0.1 levels respectively.
3. D-B stands for "During – before".
4. There were too few instances of grocery shopping frequency decrease between the pandemic period and before for e-grocery and click-and-pick to estimate a logit model with explanatory variables.

However, households who saw their income drop during the pandemic decreased more their frequency of in-store shopping (1.542†), and they increased more than households with no income change their frequency of e-grocery (1.576†) and click-and pick (1.640†). This trend appears likely to continue after the pandemic (1.510*, 1.706†, and 1.644† respectively). Conversely, households who saw an increase in income are likely to decrease less in-store shopping after the pandemic (0.595*). Likewise, compared to households whose time working at home did not change, households who spent more time working from home decreased more their frequency of in-store shopping (1.460†) while increasing more their frequency of e-grocery (1.615†). These households are planning to increase more their frequency of e-grocery (1.575†) and click-and-pick (1.398*) after the pandemic.

One unexpected result came from the variable that captures a decrease in household vehicles: it shows a larger (compared to the no-change baseline) increase in the frequency of in-store shopping during the pandemic (2.195*) and a larger decrease in e-grocery (2.287*) after the pandemic. I surmise that the decrease in household vehicles did not limit the mobility of most of these households, but unfortunately the survey data does not allow it to be sure.

Finally, reinforcing the findings from Table 4.4, I see that more tech-savvy households decreased more their in-store grocery shopping frequency during the pandemic (1.422†) and increased more their frequency of both e-grocery (1.681‡) and click-and pick (1.777‡). These trends are likely to continue (1.451†, 1.698‡, and 1.771‡), although with a correction for click-and-pick (1.956†). The number of grocery stores within 5 miles and the severity of the pandemic played no role in changes of grocery shopping frequency.

4.6 Conclusions

In this chapter, I analyzed the impacts in California of the Covid-19 pandemic on grocery shopping frequency and its changes based on a random survey of 1,026 Californians conducted in late May 2021, using both ordered and generalized structural equation models. Compared to papers published to date, my study covers a much longer period of the pandemic (March 2020 to late May 2021) and respondents are representative of the Californian population, which enables me to generalize results to the whole state. To the best of my knowledge, the work is the first to analyze the impact of the pandemic on grocery shopping in California. Since shopping for groceries is essential for meeting daily needs, it is important to understand the short- and long-term impacts of the Covid-19 pandemic. The results have implications for grocers, logistics managers, land-use planners, and transportation planners so they can better plan their business, adapt supply chains, consider modifying land use regulations, and adapt parking and curbside management practices.

My results show that during the pandemic California households still relied heavily on in-store grocery shopping although they decreased the frequency of their grocery shopping trips. Industry news confirms that traditional grocers did well during the pandemic. For example, Kroger, which is the largest supermarket chain by revenue in the U.S., increased its total sales by 14.2% during 2020, excluding asset sales (Redman, 2021). However, the pandemic accelerated experimentation with e-grocery (+8.9% of new households), and especially click-and-pick (+13.3% of new households). After the pandemic is (finally) over, the respondents expect to shop slightly less frequently in brick-and-mortar stores compared to before, and to increase their frequency of use of e-grocery, and especially click-and-pick, with increases for both in post-pandemic compared to pre-pandemic frequencies.

The rich set of explanatory variables in the models (socio-economic characteristics, attitudes about information and communication technology, a measure of grocery stores proximity, and a variable that reflects Covid-19 severity) enabled me to better understand the heterogeneity of these changes. While Baby boomers and Silent generation households are more attached to in-store grocery shopping, Gen Z, Y, and X households have been more willing to experiment with and adopt e-grocery, and especially click-and-pick. While households with a graduate or professional education grocery shopped less in-store during the pandemic in favor of e-grocery (possibly to avoid riskier, crowded spaces), they appear more likely to decrease their reliance on the latter after the pandemic. Moreover, while African American and Asian households were less likely than Whites to grocery shop in-stores during the pandemic and after, there was no race difference in the online shopping options (e-grocery and click-and-pick). One potential barrier, however, is tech-savviness (but also access to a reliable internet connection, an aspect I did not explore here), which is an important factor in the adoption and use of both e-grocery and click-and-pick.

The first implication of my findings is that online grocery shopping appears to promote equity in access to fresh food. Indeed, the changes in grocery shopping induced by Covid-19 is potentially better access to fresh grocery for African Americans, and more generally poorer households, who have been underserved by supermarkets (Beaulac *et al*., 2009; Morland *et al*., 2002). As explained by Meyersohn (2020), supermarkets have been courting suburban customers, who are predominantly White, and ignoring inner-city residents, who tend to belong to racial minorities (such as Black Americans), making it more expensive for them to access fresh and healthy food. This potential benefit for online shopping is limited by age, however, as older households relied less on e-grocery and click-and-pick during the pandemic, as already noted in the literature (Figliozzi and Unnikrishnan, 2021; Kim and Wang, 2021; Morganosky and Cude, 2000, 2002; Suel *et al*., 2015; Wang *et al*., 2021). Impaired mobility and less tech-savviness are the main obstacles for older people to adopt e-grocery (not to mention delivery fees for people on fixed, limited income). Since click-and-pick requires travel and may expose them to danger during a pandemic, policymakers may consider assistance with online shopping and grocery delivery programs for the elderly, not only during, but also after the pandemic.

Second, I found that click-and-pick is likely to be gaining in popularity after the pandemic is over, especially among younger generations, more so than online shopping with home delivery (e-grocery in this chapter). One reason is that click-and-pick is typically fee-free (Figliozzi and Unnikrishnan, 2021; Pourrahmani and Jaller, 2021; Wang and Zhou, 2015). Moreover, it does not require to be at home during a time window like e-grocery, and it is easy to include in a trip chain as part of a commuting trip, for example. The increasing popularity of click-and-pick suggests that online shopping for groceries will not decrease household travel. Fostering e-grocery, which has the potential to decrease travel would entail removing some of the obstacles currently hampering its more widespread adoption, i.e., cost, excessive delivery window, and concerns about the consistency of the quality of fresh goods (produce, meat, and fish). For grocers and delivery companies to provide more accurate and narrower delivery windows would require neighborhood warehouses from which orders could be quickly delivered, possibly on electric cargo bikes. In many areas in California, this may require changing the zoning law to allow small warehouses closer to or even within residential areas. Giving customers confidence in the consistency of the quality of fresh goods would help overcome the lack of sensory information inherent in online shopping, and require supply chain investments. Decreasing the cost of e-grocery would entail more automation for collecting and packing orders, using robots for example, and increasing market share to serve as many customers as possible per mile traveled.

Finally, the growing popularity of online shopping increases the importance of managing curbside space more efficiently in urban areas to reduce congestion and the emission of air pollutants and greenhouse gases, in retail areas with click-and-pick and in residential areas with e-grocery. Curbside parking provides temporary spaces for freight loading and unloading and pick-up and drop-off for shoppers. If they were properly designed, dedicated areas for freight loading and unloading in commercial areas could help reduce emissions of air pollutants by delivery vehicles (Jaller, 2021). Many retailers (such as Walmart, Target, and Kroger) have provided reserved parking spots for customer pick-up (Kavilanz, 2020). My results show a likely increase in the popularity of both e-grocery and click-and-pick after the pandemic, for the latter especially among younger generations. Therefore, providing sufficient curbside space for pick-up or deliveries, with strict enforcement of time limits could help avoid congestion, excessive queueing, while reducing air pollution and illegal parking. Changes in shopping (not just for grocery) invite revisiting planning requirements for parking, and particularly for the ratio of longer-term parking to pick-up and delivery parking spaces. In residential areas, noise and air pollutant emissions should probably receive increasing attention with the rise of home deliveries, at least until the widespread electrification of delivery vehicles or their replacement by cargo bikes or even robots.

There are several limitations of my study. First, to keep the survey short, I did not ask for detailed information about the socio-economic characteristics (e.g., education, race, or Hispanic status) of the other members of the respondents' households. Second, to get a more complete picture of the impact of the pandemic on grocery shopping, it would have been of interest to know how much respondents typically spent on groceries before and during the pandemic for the different channels I considered (in-store, e-grocery, and click-an-pick), although I suspect that many of respondents would not know this information without additional analysis. Third, it would have been useful to know if changes in the number of household vehicles led some of the respondents to lose some mobility.

One possibility for future work would be to analyze vehicle miles traveled from trips related to in-store grocery shopping, e-grocery, and click-and-pick, both for households and for grocers/delivery firms. In particular, it would be of interest to quantify the environmental impacts of these grocery shopping channels on transportation and freight systems, from the point of view of energy use, air pollution, and greenhouse gas emissions. Another possibility would be to contrast urban, suburban, and rural areas to explore the impacts of the Covid-19 pandemic on grocery shopping.

4.7 References

Abdullah, M., Dias, C., Muley, D., & Shahin, M. (2020). Exploring the impacts of COVID-19 on travel behavior and mode preferences. *Transportation Research Interdisciplinary Perspectives*, *8*, 100255. https://doi.org/10.1016/j.trip.2020.100255

Beaulac, J., Kristjansson, E., & Cummins, S. (2009). Peer Reviewed: A Systematic Review of Food Deserts, 1966-2007. *Preventing Chronic Disease*, *6*(3). Retrieved from /pmc/articles/PMC2722409/

Begley, S., Marohn, E., Mikha, S., & Rettaliata, A. (2020). Digital disruption at the grocery store. Retrieved from McKinsey Company website: https://www.mckinsey.com/industries/retail/our-insights/digital-disruption-at-the-grocery-store

Bollen, K. A. (1989). *Structural Equations with Latent Variables.* John Wiley & Sons.: John Wiley & Sons.

Brant, R. (1990). Assessing Proportionality in the Proportional Odds Model for Ordinal Logistic Regression. *Biometrics*, *46*(4), 1171–1178. https://doi.org/10.2307/2532457

Devenyns, J. (2019). Impulse snack purchases decline as shopping goes online. Retrieved from FOODDIVE website: https://www.fooddive.com/news/impulse-snack-purchases-decline-as-shopping-goes-online/550395/

Dias, F. F., Lavieri, P. S., Sharda, S., Khoeini, S., Bhat, C. R., Pendyala, R. M., … Srinivasan, K. K. (2020). A comparison of online and in-person activity engagement: The case of shopping and eating meals. *Transportation Research Part C: Emerging Technologies*, *114*, 643–656. https://doi.org/10.1016/j.trc.2020.02.023

Dimock, M. (2019). Defining generations: Where Millennials end and Generation Z begins. Retrieved from Pew Research Center website: https://www.pewresearch.org/fact-tank/2019/01/17/where-millennials-end-and-generation-z-begins/

Droesch, B. (2021). US Digital Grocery Forecast 2021 - Defining the Key Players and Trends in a Rapidly Evolving Market. Retrieved from https://www.emarketer.com/content/us-digital-grocery-forecast-2021

Fabrigar, L. R., & Wegener, D. T. (2012). *Exploratory factor analysis*. Oxford University Press.

Figliozzi, M., & Unnikrishnan, A. (2021). Exploring the impact of socio-demographic characteristics, health concerns, and product type on home delivery rates and expenditures during a strict COVID-19 lockdown period: A case study from Portland, OR. *Transportation Research Part A: Policy and Practice*, *153*, 1–19.

https://doi.org/10.1016/j.tra.2021.08.012

Food and Nutrition Service USDA. (2021). Where Can I Use SNAP EBT? Retrieved from https://www.fns.usda.gov/snap/retailer-locator

Grashuis, J., Skevas, T., & Segovia, M. S. (2020). Grocery Shopping Preferences during the COVID-19 Pandemic. *Sustainability*, *12*(13), 5369. https://doi.org/10.3390/SU12135369

Hand, C., Riley, F. D. O., Harris, P., Singh, J., & Rettie, R. (2009). Online grocery shopping: The influence of situational factors. *European Journal of Marketing*, *43*(9), 1205–1219. https://doi.org/10.1108/03090560910976447

IPSOS. (2021). KnowledgePanel. Retrieved from https://www.ipsos.com/en-us/solutions/public-affairs/knowledgepanel

Jaller, M. (2021). *Curbside Management Is Critical for Minimizing Emissions and Congestion*. https://doi.org/10.7922/G27W69H5

Jaller, M., & Pineda, L. (2017). *Warehousing and Distribution Center Facilities in Southern California : The Use of the Commodity Flow Survey Data to Identify Logistics Sprawl and Freight Generation Patterns A Research Report from the National Center*.

Kaiser, H. F., & Rice, J. (1974). Little Jiffy, Mark IV. *Educational and Psychological Measurement*, *34*, 111–117.

Kavilanz, P. (2020). Curbside pickup is great for everyone, but it's a headache for malls. Retrieved from CNN Business website: https://www.cnn.com/2020/06/29/business/curbside-pickup-malls-pandemic/index.html

Kim, W., & Wang, X. (Cara). (2021). To be online or in-store: Analysis of retail, grocery, and food shopping in New York city. *Transportation Research Part C: Emerging Technologies*, *126*, 103052. https://doi.org/10.1016/j.trc.2021.103052

Kline, R. B. (2015). *Principles and practice of structural equation modeling*. Guilford publications.

Korea Agro-Fisheries & Food Trade Corporation. (2020). 2020 food industry key statistics. Retrieved from https://www.atfis.or.kr/article/M001040000/view.do?articleId=3545

Long, J. S. (1997). *Regression Models for Categorical and Limited Dependent Variables. Advanced Quantitative Techniques in the Social Sciences, Volume 7*. Sage.

Meyersohn, N. (2020). How the rise of supermarkets left out black America. Retrieved from CNN Business website: https://www.cnn.com/2020/06/16/business/grocery-stores-access-race-inequality/index.html

Morganosky, M. A., & Cude, B. J. (2000). Consumer response to online grocery shopping. *International Journal of Retail & Distribution Management*, *28*(1), 17–26.

Morganosky, M. A., & Cude, B. J. (2002). Consumer demand for online food retailing: Is it really a supply side issue? *International Journal of Retail & Distribution Management*, *30*(10), 451–458. https://doi.org/10.1108/09590550210445326

Morland, K., Wing, S., Roux, A. D., & Poole, C. (2002). Neighborhood characteristics associated with the location of food stores and food service places. *American Journal of Preventive Medicine*, *22*(1), 23–29. https://doi.org/10.1016/S0749-3797(01)00403-2

Morrison, R. M., & Mancino, L. (2015). Most U.S. Households Do Their Main Grocery Shopping at Supermarkets and Supercenters Regardless of Income. Retrieved from Food Choices & Health website: https://www.ers.usda.gov/amber-waves/2015/august/most-us-households-do-their-main-grocery-shopping-at-supermarkets-and-supercenters-regardless-of-income/

Murphy, A. J. (2003). (Re)Solving Space and Time: Fulfilment Issues in Online Grocery

Retailing: *Environment and Planning A*, *35*(7), 1173–1200. https://doi.org/10.1068/A35102

Peterson, B., & Harrell, F. E. (1990). Partial Proportional Odds Models for Ordinal Response Variables. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *39*(2), 205–217. https://doi.org/10.2307/2347760

Pourrahmani, E., & Jaller, M. (2021). Crowdshipping in last mile deliveries: Operational challenges and research opportunities. *Socio-Economic Planning Sciences*, (February), 101063. https://doi.org/10.1016/j.seps.2021.101063

Punel, A., & Stathopoulos, A. (2017). Modeling the acceptability of crowdsourced goods deliveries: Role of context and experience effects. *Transportation Research Part E: Logistics and Transportation Review*, *105*, 18–38. https://doi.org/10.1016/j.tre.2017.06.007

Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2004). Generalized multilevel structural equation modeling. *Psychometrika 2004 69:2*, *69*(2), 167–190. https://doi.org/10.1007/BF02295939

Redman, R. (2021). Kroger "generated strong momentum" in 2020 fiscal year, CEO Rodney McMullen says. Retrieved March 4, 2021, from Supermarket News website: https://www.supermarketnews.com/retail-financial/kroger-generated-strong-momentum-2020-fiscal-year-ceo-rodney-mcmullen-says

Saphores, J.-D., & Xu, L. (2020). E-shopping changes and the state of E-grocery shopping in the US - Evidence from national travel and time use surveys. *Research in Transportation Economics*, (March). https://doi.org/10.1016/j.retrec.2020.100864

Shamshiripour, A., Rahimi, E., Shabanpour, R., & Mohammadian, A. (Kouros). (2020). How is COVID-19 reshaping activity-travel behavior? Evidence from a comprehensive survey in Chicago. *Transportation Research Interdisciplinary Perspectives*, *7*, 100216. https://doi.org/10.1016/j.trip.2020.100216

Spurlock, C. A., Todd-Blick, A., Wong-Parodi, G., & Walker, V. (2020). Children, Income, and the Impact of Home Delivery on Household Shopping Trips. *Transportation Research Record*, *2674*(10), 335–350. https://doi.org/10.1177/0361198120935113

Suel, E., Le Vine, S., & Polak, J. (2015). Empirical Application of Expenditure Diary Instrument to Quantify Relationships between In-Store and Online Grocery Shopping: Case Study of Greater London. *Transportation Research Record*, *2496*, 45–54. https://doi.org/10.3141/2496-06

U.S. Census Bureau. (2021). *Quarterly e-commerce retail sales 2nd quarter 2021*. Retrieved from http://www2.census.gov/retail/releases/historical/ecomm/07q4.pdf

Unnikrishnan, A., & Figliozzi, M. (2021). Exploratory analysis of factors affecting levels of home deliveries before, during, and post- COVID-19. *Transportation Research Interdisciplinary Perspectives*, *10*, 100402. https://doi.org/10.1016/j.trip.2021.100402

Ver Ploeg, M., Mancino, L., Todd, J. E., Clay, D. M., & Scharadin, B. (2015). *Where Do Americans Usually Shop for Food and How Do They Travel To Get There? Initial Findings from the National Household Food Acquisition and Purchase Survey*. Retrieved from https://www.ers.usda.gov/publications/pub-details/?pubid=79791

Wang, X. C., & Zhou, Y. (2015). Deliveries to residential units: A rising form of freight transportation in the U.S. *Transportation Research Part C: Emerging Technologies*, *58*, 46–55. https://doi.org/10.1016/j.trc.2015.07.004

Wang, X., Kim, W., Holguín-Veras, J., & Schmid, J. (2021). Adoption of delivery services in light of the COVID pandemic: Who and how long? *Transportation Research Part A: Policy and Practice*, *154*, 270–286. https://doi.org/10.1016/j.tra.2021.10.012

Wang, Y., Xu, R., Schwartz, M., Ghosh, D., & Chen, X. (2020). COVID-19 and Retail Grocery

Management: Insights from a Broad-Based Consumer Survey. *IEEE Engineering Management Review*, *48*(3), 202–211. https://doi.org/10.1109/EMR.2020.3011054

Williams, R. (2016). Understanding and interpreting generalized ordered logit models. *Journal of Mathematical Sociology*, *40*(1), 7–20. https://doi.org/10.1080/0022250X.2015.1112384

Yarbrough, B. (2015). Language besides English spoken by 44 percent in California, more than half in Santa Clara County. Retrieved from The Mercury News website: https://www.mercurynews.com/2015/11/04/language-besides-english-spoken-by-44-percent-in-california-more-than-half-in-santa-clara-county/

Chapter 5. Summary and Conclusions

In this dissertation, I have presented three essays that analyzed online shopping and household travel before, during, and after the pandemic using a variety of statistical tools applied to data from the 2009 and 2017 NHTS, the 2017 ATUS, and a May 2021 survey of Californian households.

In Chapter 2, I explored changes in residential deliveries in the U.S. between 2009 and 2017. I found that e-shopping was embraced by a larger percentage of the U.S. population in 2017 and that e-shoppers were more varied in 2017 than in 2009. Households with higher incomes and higher educational achievements were more willing to receive more deliveries in both 2009 and 2017. Households with more female members were less likely to never shop online in both years. In addition, households with mobility-impaired members relied more on e-shopping in 2017 compared to 2009. Although these households were more likely to never shop online in 2009, this was no longer the case in 2017. Understanding the characteristics of people who shop online should help supply chain managers better handle online orders, organize deliveries, and plan package returns. It should also help transportation planners and engineers better address traffic congestion, air pollution, and noise, especially in residential areas. It may also inform the regulations of delivery vehicles, and labor laws for warehouse workers and delivery drivers.

In Chapter 2, I also characterized the profile of U.S. grocery shoppers. I found that grocery shoppers before the pandemic were more likely to be females and unemployed (recall that homemakers are not officially employed), but less likely to be from younger generations.

They were also likely to have less than a college education, or to be African American. The only variable that affected e-grocery shoppers was gender: females were more likely to use e-grocery. Better understanding e-grocery is important because it could help provide food in neighborhoods underserved by supermarkets and to households with mobility impairments. Policymakers may consider subsidizing e-grocery costs, which would help create jobs in grocery packing and delivery in these areas. Second, the increasing popularity of online shopping has led to increases in deliveries to residential areas, which exacerbated parking, air pollution, and noise in denser urban areas.

In Chapter 3, I investigated how e-shopping was impacting household travel using data from the 2017 NTHS. My results showed that more e-shopping was, on average, associated with more travel. Medium (high) frequency e-shopping households had on average 10 (7) more activities per month on weekdays than low-frequency e-shopping households, which translated into ~57 (53) additional miles. Of these additional activities, 2 to 3 were for shopping, which corresponded to 15 to 20 additional monthly miles. In addition, e-shopping has a deep impact on household travel. Households with more e-shopping had more activities, including buying meals, purchasing services, exercising, and working from home. The extent to which an increase in the number of household activities resulted in more travel depends on population density, the frequency of online shopping, the day of the week, and the type of activity. These results unpack the complex relationship between shopping activities and travel. It was more of a complementary relationship between online shopping and store shopping.

Understanding the impact of e-shopping on travel is particularly relevant for policymakers concerned with urban congestion, air pollution from transportation, and greenhouse gas (GHG) emissions. To help reduce travel induced by e-shopping, policymakers should encourage the creation of neighborhood depots where households can easily pick up and return unwanted orders. Additionally, it would be useful to foster the development of virtual reality tools to shop from home and reduce the probability of unwanted returns.

One limitation of my work in Chapter 3 is that 2017 NHTS data did not include personal attitudes towards shopping and traveling. Moreover, online purchases of groceries, clothing, appliances, and gas were all lumped together, which is unfortunate because some of these goods can be purchased online and others not (e.g., gasoline). In addition, the data did not include detailed time use information, which would have allowed a better understanding of the interactions between e-shopping, in-store shopping, and travel.

In Chapter 4, I investigated the impact of the Covid-19 pandemic on grocery shopping frequencies in Californians during and after the pandemic. After weighting survey answers to map respondents to the California population, I found that most Californians continued to shop in-person for groceries (fewer than 6% never went during the pandemic), although not as frequently as before. Online ordering of groceries with home delivery (e-grocery) saw a boost in popularity (+8.9% of new users), but not as much as click-and-pick for groceries (+13.3% of new users). Most of the gains from online grocery shopping will likely remain after the pandemic, as many traditional grocers made e-grocery an integral part of their business.

Results from my models showed that Baby boomers shopped for groceries in stores more frequently than other generations before the pandemic, while younger generations (Gen Z & Y and Gen X) used e-grocery and click-and-pick more often. Except for older households (members of the Silent generation), these differences mostly disappeared during the pandemic for in-store and e-grocery, but not for click-and-pick. For the latter, younger generations were more likely to rely on click-and-pick, a difference likely to continue after the pandemic is over.

Education also played a role, with less educated households shopping less in person before the pandemic (and likely after), while highly educated households shopped less in grocery stores during the pandemic in favor of more e-grocery. Although there were no race differences in frequency or grocery shopping channels during the pandemic, my results showed that African Americans and Asians were shopping less frequently in grocery stores before the pandemic (lower frequencies) than Whites, a difference likely to reappear after the pandemic. In addition, tech-savvy households, households with reduced income during the pandemic, and households with more time working from home increased their shopping frequency in terms of e-grocery and click-and-pick during the pandemic. They may continue doing so afterward.

My results should draw attention to equity issues for underserved groups and elderly people, who were underserved before the pandemic and were most at risk during the pandemic. Moreover, the increasing demand for fast and on-time deliveries of grocery highlights the need to upgrade grocery delivery channels in the United States, which are less developed than in countries such as China or South Korea. The rise in delivery trips in residential areas is likely to increase congestion, illegal parking, noise, and GHG emissions (especially in dense urban areas), so it is urgent to review and upgrade curbside and parking management policies.

One limitation of my work in Chapter 4 is that the IPSOS survey did not fully characterize household adults apart from the respondent (i.e., their age, race, Hispanic status, and education are not available). It would also have been fruitful to have information about attitudes toward grocery shopping, travel, and the pandemic.

Much remains to do to better understand the changing nexus among in-store shopping, online shopping, and travel. Future work could collect and analyze detailed information about in-person and online shopping. It would also be very valuable to incorporate data about shoppers' attitudes into models explaining the selection of different shopping channels for different product categories. A second interesting line of inquiry would be to quantify the energy, air pollution, and greenhouse gas impacts of various forms of shopping (in-person, online with home delivery, or online with curbside or store pick-up) in different environments characterized by different land use metrics. Finally, more work is needed to understand potential obstacles to the generalization of online shopping, particularly for grocery in areas traditionally underserved by supermarkets because of its importance to health.