New Geographies of Electronic Commerce: An Appraisal of Omni-channel Strategies in the United States Grocery Industry

A Dissertation Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Connor Isaac Reed
May 2024

Acknowledgements

Thank you to my friends and family who have supported me through every triumph and challenge. I would like to acknowledge the industry partnerships that have made this research possible. Thank you to Mr. Michael Simon at *SiteWise Analytics* for equipping me with geospatial software at no cost. Thank you to Lesley Newman *at Synergos Technologies* for providing the demographic data used in this study and thank you to Oszkar Breti at *ChainXY* for providing all of the grocery store location data that formed the foundation of this research. I also wish to express my gratitude to my committee members: Dr. Nicholas Nagle, Dr. Hyun Kim, Dr. Charles Liu, and Dr. William Graves. Their thoughtful feedback and support elevated all of the research documented here and challenged me to grow as a scholar. Finally, thank you to my advisor, Dr. Ronald Kalafsky, for investing so much time and mentorship in me and in this work.

Abstract

This body of doctoral research investigates the unique geography of e-commerce [electronic commerce] offerings among major grocery chains in the contiguous United States. Prior research has made efforts to understand the demographic and spatial dynamics that may influence consumers' decision to shop online or in-store. Far less emphasis, however, has been placed on how firms have made key strategic decisions across space regarding e-commerce. This work leverages spatial econometric modeling and machine learning to understand how the characteristics of trade areas inform grocery chains' decisions about whether to offer pickup, delivery, both, or no offering at a given store location. This research effectively pulls back the curtain on how large grocery chains are designing e-commerce strategies across their networks and carries practical implications for both location planning and competitive defense.

Table of Contents

List of Tables

List of Figures

Chapter One

Introduction

Like the automobile, the shopping mall, and the supercenter before it, electronic commerce (e-commerce) has been, and remains, a seismic disrupter in the retail economy. E-commerce sales are projected to exceed $2 trillion by 2030, accounting for 31% of total retail market share (Walk-Morris, 2022), and 79% percent of consumers shop online at least once per month (Baluch, 2023). The ubiquity of e-commerce has driven structural changes in retail markets. These changes have manifested in a variety of unprecedented trends, including: mass closures of some traditional brick-and-mortar retailers, the reorganization and reuse of mall spaces, the reformulation of firm location strategies across a variety of industries, the advent of new complimentary technologies and tools that rely on or augment the digital marketplace, reductions in search costs and frictions, and changes in the behavior of consumers. This list is far from exhaustive, but provides a sampling of how deeply and systemically e-commerce has impacted *where*, *when*, and *how* we shop.

Implicitly, many of the structural changes that have occurred in the retail economy as a result of e-commerce disruption can often be examined through a spatial lens. These changes have unique geographic footprints. Where are stores closing? Where are e-commerce businesses opening? Where are consumers opting for e-commerce purchases over in-store shopping? How can once-vacant mall spaces be refilled and recycled to increase gravity? There are many locational questions that may be asked about e-commerce, and retail geographers are uniquely equipped to answer these questions. Indeed, there is a rich history of retail geographers addressing spatial patterns and structures tied to previous disruptions in retail such as the automobile (Nelson, 1958; Kane, 1966; Claus et al. 1972), the shopping mall (Scott, 1970; Dawson, 1980; Wrigley, 1988), and the supercenter (Jones & Simmons, 1990; Graff, 1998, 2006, 2007; Birkin et al. 2002). But in recent years, retail geographers have had less of a presence and prevalence in bodies of scholastic research related to e-commerce. With the exception of the retail geography community in the United Kingdom (particularly out of University of Leeds, University of Liverpool, and University of Southampton), the vast majority of spatial research in e-commerce has been taken on by business scholars, economists, and more broadly trained transportation and economic geographers. Surprisingly, the most recent scholastic text on retail geography from Wang and Du (2021) only devotes a small part of a single chapter to any discussion of e-commerce, and does little more than list current trends in digital retailing, without engaging with any contemporary research on geographies of e-commerce. This marks a missed opportunity to produce a timely, comprehensive text in retail geography that makes some effort to recognize the increasing ubiquity and disruption of e-commerce. Previous seminal texts in retail geography have already called for the future development of theories and techniques in the discipline to directly address e-commerce (Currah, 2002; Birkin et al. 2002, 2017; Beckers et al. 2022). It follows that e-commerce must disrupt “business-as-usual” in retail geography, just as it has disrupted “business-as-usual” in the retail economy.

The need for contemporary studies of geographies of e-commerce has, undoubtedly, been made more urgent in the wake of the COVID-19 pandemic. The pandemic had a dramatic impact on the retail economy, as many brick-and-mortar chains saw reduced foot traffic and poor performance as consumers were forced to shop online for goods that may have previously been purchased in-store. Other chains carrying essential goods saw exponential increases in traffic and demand for cleaning supplies, toiletries, and foodstuffs. Over the first year of the pandemic in the United States, from March 2020 through March 2021, e-commerce sales increased by $183 billion (Verdon, 2021). Even now, as the pandemic appears to be in a gradual decline from its heights in previous years, e-commerce sales remain strong, and indicate that the retail economy is more likely to settle into some “new normal” than to return to its previous state before the pandemic. Total e-commerce sales in 2021 were projected to be $147 billion greater than they would have been in a world without COVID-19 (Droesch, 2021). There is an argument to be made that this rate of growth and adoption for e-commerce transactions would have been reached at a much later time, given the past growth that e-commerce has enjoyed for two decades, but the pandemic exponentially expedited this growth. This marks a new opportunity to conduct geographic research to understand the spatial structures and processes underlying this increasing ubiquity of e-commerce and how it affects producers and consumers in a post-pandemic world.

Research Questions

The U.S. grocery industry makes for a particularly compelling environment in which to conduct research on the geographies of e-commerce. This is, in no small part, due to the stratified nature of e-commerce offerings among major grocery chains. Unlike other big box stores and general merchandisers, not every grocery store provides the same e-commerce offerings to its customers. Each grocery chain will make strategic decisions about whether or not a given location in their chain will offer pickup, delivery, both pickup and deliver, or no digital channels at all. It is this essential and complex strategic decision that forms the foundation for this research.

This body of dissertation work is guided by the following research questions:

1. What salient research themes have emerged in geographies of digital consumption? What Demographic characteristics and spatial dynamics are driving consumers to purchase online?
2. Using the U.S. grocery sector as a case study, to what extent do e-commerce location strategies among major grocery chains align (or diverge) from academic research related to geographies of e-commerce consumption? What demographic and geographic forces drive a grocery banner's decision to offer e-commerce at a given store location?
3. If the key drivers of e-commerce offerings at grocery locations can be meaningfully measured and analyzed, can a machine learning model accurately prescribe an appropriate e-commerce strategy based on trade area composition?

These three research questions attempt to set a foundation for scholarship in retail geography focused specifically on e-commerce, and how geographic forces exert influence over the decisions made by retailers in specific markets, decisions that ultimately impact day-to-day consumption in physical and cyber space.

Research Objectives

In the pursuit of answers to the above research questions, the objectives of this dissertation are threefold:

1. Producing a comprehensive and expansive literature review of geographic studies of e-commerce and digital retailing, with particular emphasis on the United States – tracing major trajectories of thought and key contributions from scholars in geography, business, economics, and planning
2. Developing a descriptive and deterministic study of the e-commerce strategies adopted by major grocery chains across the United States using trade area analysis, geostatistics, and spatial econometric modeling
3. Exploring the application of a supervised learning algorithm in prescribing optimal e-commerce offerings for grocery locations based on the composition of trade areas, and the deterministic relationships previously identified in the second research product

Achieving each of these research objectives will begin to address the previous call for increased focus on firm location strategy in the era of e-commerce (Kirby-Hawkins et al., 2019), while also establishing new research directions in retail geography and sounding a call-to-action for increased scholarship on how e-commerce has irrevocably shifted the geographic complexities of consumption. To that end, this dissertation is comprised of three core chapters. The first provides a literature review of previous studies of the demographic characteristics and spatial dynamics that influence consumers' decision to shop online or in-store, while also briefly chronicling the history of spatial studies of e-commerce from the earliest days of the worldwide web. The second chapter quantifies the relationship(s) between various trade area characteristics – including demographics and competition – and the decisions made by major U.S. grocery chains about whether to offer some form of e-commerce at a given location. The third core chapter builds on the previous study by exploring an application of supervised machine learning that can appropriately prescribe the most optimal e-commerce offering for a given trade area based on its composition. This multi-classification random forest model will consider demographics, competition, and geography to gauge whether offering pickup, delivery, both pickup and delivery, or no offering are most appropriate for a given store and its trade area. The model will also produce a hypothetical probability surface that can reflect the probability that a given grocery banner will offer e-commerce in a given market area. A concluding section to this dissertation will provide a brief summation of key findings, implications for the retail industry, and tacks for future scholarship.

Chapter Two

Drivers of E-Commerce Consumption: A Survey of Literature

Introduction

For decades, business scholars and social scientists have been concerned with *where* consumers shop, *how* they shop, and *what* they shop for. Inevitably, the answers to these questions have shifted, and will continue to shift as e-commerce consumption becomes increasingly prevalent in daily life. An additional line of questioning that stems organically from this unprecedented rise in e-commerce concerns *who* is adopting e-commerce as a regular channel of consumption and *why*. The characteristics of people and places have a powerful influence over how goods and services are consumed. Where we shop, how we shop, and what we shop for are all shaped by geographic forces that influence who we are and how we interact with the world around us. E-commerce has irrevocably changed how consume goods and services by introducing new sites of consumption through websites and mobile phones. Consumers are free to order goods from around the world to their doorstep, or to have their weekly groceries delivered within hours from their local grocer. When equipped with a mobile phone, these items can be purchased from almost anywhere, not just the home. This time-space compression has empowered consumers with the ability to often find exactly what they are looking for at the best price possible, and to connect with other consumers who may use the same products or services. E-commerce has also shifted the way goods and services are supplied to the marketplace. Firms must now think actively about alternative delivery methods, including dark stores or localized fulfillment centers. They are also taking steps to more carefully assort offerings that can cater to niche demand that may not be serviced by brick-and-mortar stores in a given market. The consequences of this contestation between e-commerce and brick-and-mortar can already be seen in the retail landscape with the mass closures of retail chains, particularly department stores and general merchandisers. Through all of these effects, certainly, e-commerce has systemically altered the retail economy and the physical retail landscape, introducing and altering existing spatial dynamics of consumption.

Given this inherent spatiality, it is concerning that geographers have not given e-commerce the attention it is due as a meteoric disrupter of the retail economy. This lack of attention is all the more concerning in light of the effects of the COVID-19 pandemic, which dramatically decreased foot traffic to many retail centers from 2020 through 2021 (Koster et al., 2022; Enoch et al., 2022). The pandemic incentivized consumers to pivot to online channels where they may have once shopped at physical stores (Dannenberg et al., 2020; Guthrie et al., 2021; Alhaimer, 2022), and the rate at which consumers opted to shop online was found to be proportional to the incidence of COVID-19 in their respective area (Gao et al., 2020). Interestingly, the impacts of the pandemic on retailing were varied across sectors depending on the good or service. In the early days of the pandemic, consumers began flocking to grocery stores and other food retailers to stock their pantries with a surplus of foodstuffs to weather an uncertain period of fear (Hao et al., 2020; Chenarides et al., 2020). Although, scholars identified some discomfort with entering grocery stores depending on the spread and severity of COVID-19 in a consumer's area (Grashuis et al., 2020). By contrast, goods such as electronics and media saw only incremental lifts in online purchases from the pandemic (Kawasaki et al., 2022). Bryson (2021) describes the heterogeneous effect of COVID-19 on retailing succinctly:

> The impact of the pandemic on the retail and hospitality industries takes two forms. On the one hand, providers of essential services experienced an increase in their cost base given the necessity of introducing COVID-19 prevention routines, but turnover and profitability increased… On the other hand, non-essential retailers, including cafes, restaurants, bars, and hotels had to close for extended periods. (p.203).

Further, the pandemic exacerbated previously existing financial and economic problems that were already plaguing many brick-and-mortar retailers (Nicola et al., 2020; Bryson, 2021). It is because of this that Bryson rightly calls out the challenge in assuming any direct causal relationship between many pandemic effects and retail failures. Many of the forces accelerated by COVID-19 had been building up for decades. For some retailers, cans that had once been kicked down the road were suddenly catapulted into present-day crises. The challenges with attributing causality also extend beyond retailers' balance sheets. Some research from geographers and social science calls out that some of the online shopping trends observed during the pandemic were, surprisingly, driven more by consumer's demographic and psychographic characteristics rather than any concerns about health regulations or the risk of exposure to the coronavirus (Hashem, 2020; Warganegara, 2022; Wieland, 2022). Currently, the global retail economy is still recovering from the aftermath of the pandemic, and it is unclear what retailing ultimately will look like when the dust settles. But, the accelerating effects of COVID-19 on e-commerce consumption and disruption further bolster the need for geographers to shift more of their energy and attention to analyzing the unprecedented changes taking place in the retail landscape.

It is this lack of attention that motivates this paper, which provides a new understanding of the drivers of this rise in e-commerce consumption through a spatial lens. To that end, the paper is divided into two parts. First, it will chronicle the historical foundations of geographical studies of e-commerce, connecting digital retailing to foundational retail location theories and chronicling geographers' views of e-commerce from its earliest days. Second, it will catalog spatial quantitative research concerned with the drivers of e-commerce adoption and consumption in developed retail economies. While such reviews have been created in the past (Cao, 2009), they are becoming outdated, and a more contemporary and comprehensive survey of literature is called for. Attention will be given to two key dimensions that inform consumers' decision to participate in e-retailing: demographics and spatial dynamics. Literature investigating each dimension will be reviewed from the earliest days of e-commerce scholarship in the early 2000s to the present day. These studies span various local, regional, and national contexts including Belgium, China, France, Germany, the Netherlands, Spain, the United Kingdom, and the United States. The most recurrent demographic characteristics of e-commerce consumers are demonstrated to be higher incomes, more education, and younger ages. For spatial dynamics, multiple theoretical foundations will be explored, including the *consumer efficiency theory* and the *innovation-diffusion theory* from Anderson et al. (2003). A surprising theme from this body of literature is the absence of any consensus or consistency in empirically defending either of these theories. The core themes and trajectories of thought outlined in this paper will be used as a launch pad for potential future avenues of scholarship focused on the drivers of e-commerce consumption in developed economies. Finally, new directions for geographic research on e-commerce will be presented, with a call-to-action for retail geographers to re-examine geographies of digital consumption.

History of Spatial Studies of E-commerce

The first researchers to study e-commerce through a spatial lens were not geographers. They were a mix of practitioners working in the field of real estate market research, and scholars of business communication. The period from the late 1990s to the turn of the millennium yielded many publications with largely speculative research content, in which experts tried their best to comment on the Internet revolution, and project how the Internet would change or improve many societal mechanisms. The first collection of these speculative reports with an explicitly geographic focus came from the *International Council of Shopping Centers* [ICSC] *Research Quarterly*. Today, the ICSC remains one of the largest professional organizations in retail real estate, and includes many retail geographers who opted to work in the private sector rather than in academia. In this journal, for example, Dwyer (1996, 1998) provided some of the first geographic and demographic survey research tied to online shopping. Dwyer found that e-shopping is, unsurprisingly, strongly related to a household’s Internet accessibility and income. Li et al. (1999) built on this profile of early Internet shoppers with an additional survey analysis that noted that early adopters of e-shopping tended to be male and well-educated. Dwyer also noted that, among consumers surveyed at this time, there was little appetite for purchasing groceries through online channels. This absence of appetite would later carry serious consequences for the retail economy when many online grocery retailers collapsed during the dot-com bubble (Zook, 2002; Murphy, 2003; Andrews & Currim, 2004). Baker (2000) added that this rejection of web-based grocery shopping stems from three factors: consumers' general desire to be able to interact with their produce to assess quality, the high perishability of many grocery items, and the challenges online grocers of the time faced with pricing schemes and brand recognition. Unlike the online grocery landscape of today, most online grocery companies in the 1990s were not tied to any recognizable grocery banners. They were separate brands that were unfamiliar to consumers, such as WebVan, Streamline, and Homeruns. Rather than stocking pantries, Dwyer found that the majority of online shoppers tend to use the Internet for purchasing tickets and travel arrangements, software and hardware, and general merchandise goods such as books, entertainment, and electronics. This was corroborated by other surveys of e-shopping habits from the time (Nair & Lambert, 2002).

In the above-cited publications, Dwyer also expressed concerns that manufacturers of retail goods could look to the Internet as a means of cutting out traditional retail stores from supply chains entirely. Customers would no longer have to drive to the store to purchase goods. They could order online and the product will be delivered right to their door. Historically, firms with localized markets could enjoy some freedom with pricing due to friction driven by geography and, at times, customer ignorance (Bakos, 1998; Bell, 2014). The Internet removes this friction and commensurate search costs, allowing customers to identify other potential sellers to give business. It seemed to Dwyer that it was possible that this shift in the geography of the marketplace could eventually harm traditional brick-and-mortar retailers, a fear which was eventually realized. This fear was shared by other scholars and industry experts. McMahan (1999) produced one of the most comprehensive pieces of the time offering speculation as to the profound impacts e-commerce could have on the geography of real estate not only in retail, but across other sectors as well. He suggested that:

> Every transaction that occurs on the Internet is a transaction that will not occur in physical space. Over time, this substitution can result in less aggregate demand for physical space, all other things being equal. Even if aggregate demand for a particular property type is not affected, the Internet may require different building configurations and locations, hereby rendering many existing buildings obsolete. (p. 4)

McMahan projected lower overall demand for mall spaces and power centers[1], and major demand shifts in industrial and office spaces to accommodate the rapidly growing ecosystem of e-commerce firms. He also expected little to no impact on localized shopping centers and residential real estate. It is worth noting that, at this time, not all researchers, even those who published with Dwyer through the ICSC, shared Dwyer's concerns about the disruption of physical real estate. Some experts and researchers were confident that physical stores would remain strong and thrive, others believed that some retail formats had no need to explore doing business online to begin with (Sauer & Burton, 1999; Brand, 2000; Futterman, 2000). Still, other scholars and experts kept a more balanced view on how physical retailers should approach the digital world.

[1] Power centers may be defined as: "A center dominated by several large anchors, including discount department stores, off-price stores, warehouse clubs, or stores that offer a great selection in a particular merchandise category at lower prices." (Thrall, 2002, p. 168)

Retailers cannot afford to ignore cyberspace in favor of physical space. Rather, the two spaces must co-exist and complement each other (Jones & Biasiotto, 1999), and for many retailers, this mixing of physical and cyberspaces would eventually take the form of omni-channel retailing[2]. Omni-channel approaches allowed physical retailers to leverage their existing brand online, leverage new distribution and supplier networks, reach new demographics, and drive cross-traffic with consumers purchasing across multiple channels depending on their need, income, and location (Baker, 1999).

While industry experts and a handful of business scholars were publishing survey-based studies and largely speculative outlooks regarding e-commerce and its role in the new economy, Cairncross (1997) published her provocative text, *The Death of Distance.* Today, this text serves as a fascinating time capsule from when it was unclear how the Internet was going to change society, as it was still very much in its infancy. Cairncross comments on the unprecedented ability of e-commerce to destroy borders and barriers in the retail economy. Customers could now access and compare product information and prices on a variety of goods and services without leaving their homes. Retailers could now enjoy lower distribution costs, and a more global reach for marketing and servicing. Cairncross also suggested that e-commerce may spawn entirely new economies that relate to, or directly augment, existing markets. An example she proposed is the rise of vehicle-based jobs, in which the Internet becomes a hub for delivery services, of both people and goods. Today, this projection has been fully realized in the form of rideshare services such as Lyft and Uber, and in the form of food couriers such as DoorDash and

[2] In this form of retailing, consumers may order online and pickup in a store, order online and have their purchased goods delivered to their home, or purchase in-store and have their purchased goods delivered to their home (Wang & Du, 2021).

GrubHub. Cairncross's text would become the foundation (and frequent subject of critique) for many of the first papers published by geographers on e-commerce.

Beginning around 2000, broadly trained economic and urban geographers began to engage seriously with the Internet revolution, and the indefatigable rise of a new cyberspace existing alongside other conceptions of space that geographers had already made great efforts to analyze. Geographers were forced to reckon with an entirely new idea of space that also supported economic and social processes that had previously been relegated to the physical world. Leading this charge was Leinbach and Brunn's *Worlds of E-commerce* (2001), a collection of research from economic geographers in cooperation with economists, examining the emerging geographies of e-shopping and its consequences for the retail economy. Leinbach and Brunn call out speed, accessibility, globalization, and information as the key attributes of e-commerce. Malecki and Gorman (2001) challenged conceptions and assertions from pundits and industry experts that the Internet inherently renders space and geography inconsequential. They achieved this by examining the backbone of the Internet infrastructure of the United States, which has a clear geographic footprint. American cities have a hierarchical organization with respect to how they are connected as nodes to the Internet's physical network infrastructure. They theorized that the relative connectivity of cities to the Internet could become a driver in the location decisions of businesses in the future. This assertion that distance is not "dead" would be echoed by many geographers going forward (Wrigley 2002; Currah 2002; Zook 2002). Dodge (2001) made the first of many efforts to understand the geography of Amazon.com. Though Amazon is recognized as the dominant e-commerce player in the market today, at the time, it was a rising star and innovator. Dodge traced Amazon's rise against a backdrop of neoliberalization, noting Amazon's continued growth into markets beyond the United States, with a retained emphasis on localizing global branches to conform to societal and economic tastes among consumers in foreign markets. Dodge also asked some key questions that will be picked up by geographers in future decades, questions concerning the impact of Amazon on urban spaces, transportation networks, malls, and local shops. At the time, it was not possible to project the nature and magnitude of e-commerce on these geographies which, though ever-changing, were always familiar.

Amid this torrent of theorizations, projections, speculations, exclamations, and surveys that poured out of scholastic communities in urban-economic geography and business research, the smaller sub-discipline of retail geography began producing its own body of research tied to the rise of e-commerce, particularly for the United States and Canada, and the United Kingdom. Hernandez et al. (2001) added context to Dwyer's (1999) projection that the rise of e-commerce could, potentially, endanger the health of physical retail real estate. Hernandez et al. (2000) build a large sample of intercept surveys at three regional malls in Ontario. They found that, generally, mall shoppers are not participating in electronic commerce. For them, the Internet remained only a source for product and business information, not transaction. Shoppers cited concerns over security and privacy, shipping costs and speed, and a desire for a shopping "experience" as the primary deterrents to turning to e-tailing for purchases. This survey, of course, serves only as a snapshot in time. The authors recognized that the nature and extent of multi-channel activity must, inevitably, evolve. "Multi-channel behaviors are heralding a new retail era where shopping center vitality will be determined by operating not only from the right location on the ground but also the right addresses on the Internet."

Birkin et al. (2002) devoted a substantial proportion of their seminal text to parrying Cairncross's *Death of Distance* claims, re-affirming the need for understanding geographies in the face of the imperishable spatiality of e-commerce. First, they argued that even in a future in which e-commerce is fully and holistically intertwined with the retail economy, the vast majority of goods will still be distributed among customers with a defined geography. These geographies may become increasingly complex. Instead of customer locations organized into discrete primary and secondary catchment areas around stores, customer distributions may become stretched over space in patterns that cannot be understood through familiar concepts like trade areas. More sophisticated spatial methodologies and geographic conceptualization will be required to study and understand such changes to retail geographies. Second, they called out that that some products must be grounded in geographical space, such as fuel (including charging stations). Even as last-mile delivery methods continue to innovate and improve, retailers must still seek ways to marry and synergize their digital and physical spaces. Third, they asserted that the increasing complexity of retail distribution networks will implicitly require more sophisticated geospatial analytics to yield optimal business operations. Birkin et al. (2002) also took some time to discuss the geographies of e-commerce adoption in the United Kingdom. Their study aligned with Taylor and Murphy (2004) in showing that adoption is substantially lower in rural areas where physical channels are at a premium.

Birkin et al. concluded their text with a call for retail geographers to pursue theories and methodologies that are better equipped to succeed in the era of e-commerce.

In their text, published alongside Burkin et al. (2002), Wrigley and Lowe (2002) offered some additional perspective on the rise of e-commerce. They draw attention to the importance of logistics and fulfillment in retail supply chains:

> As we have seen… major store-based retailers in many countries have increasingly seen themselves as being as much in the business of distribution and inventory management as retail selling. As such, they potentially have considerable competitive advantage over the purely non-store electronic retailers who, during their rapid emergence in the late 1990s, struggled to achieve profitability, not least because of the cost of fulfillment and the challenges and scale requirements of the buying process and inventory management. And it is these issues which lie at the heart of any assessment of the future of store-based retailing and retail logistics systems described above, during a period in the early twenty-first century when a greater proportion of products can be expected to be purchased 'online'. (p. 94).

Wrigley and Lowe recognized that the future of retail is uncertain, but argued that the seemingly defensive adoption of the Internet by previously store-based retailers points to an irrevocable shift in the retail economy. Currah (2002) produced research that evaluates the transition of these brick-and-mortar retailers to a multi-channel approach in the Toronto market. In suit with Zook (2001) and Malecki and Gorman (2001), Currah also attempted to uncover spatialities in e-commerce networks in the Toronto metropolitan area. His work joined the chorus of economic geographers who empirically defended clear geographies of e-commerce, and new place-based agents acting to build out and maintain this new facet of the retail economy (Aoyama, 2001; Malecki & Gorman, 2001; Zook, 2000, 2001). Currah also maintained that his work is only a snapshot in time, and recognized that a steady stream of studies from retail and economic geographers will be required to capture all of the spatial processes and evolutions of e-commerce as it continues to mature in the modern economy.

Additional literature on e-commerce from retail geographers at this point in time comes from Wrigley et al. (2002), who critiqued the work of Leinbach and Brunn (2001) and argued, compellingly, that a book whose title suggests it is about e-commerce, devoted much more time and attention to the physical geographies of Internet networks and infrastructure than to any geographies of actual consumption. Geographies of e-commerce consumption is a core research theme that would be picked up by geographers and business scholars as the availability of data and cases became more commonplace in the 2010s. Leinbach and Brunn's text also made no comment on the collapsing of the dot-com bubble. In an effort to course correct their sub-discipline, Wrigley et al. (2002) proposed a new set of research themes to re-orient research efforts, particularly for economic and retail geographers. The first theme, disintermediation and re-intermediation, covered the change in retail supply chains in the wake of e-commerce: the severing of intermediary points of movement or sale and/or the creation of new intermediaries, respectively. The second, the crisis of e-tailing, put the dot-com collapse under the microscope, and sought to understand mistakes made by economists and pundits to evaluate and speculate on the exponential boom of Internet-based companies which grew too big and fell apart in the late 1990s. This theme also evoked questions over the drivers and deterrents of store-based retail firm survival amidst potential cannibalistic or competitive shocks from web-based channels. This work from Wrigley et al. (2002) attempted to refocus the discipline and steer the direction of future research.

It is out of the many disparate foundations laid in this section that various themes of research in the geographies of e-commerce emerged across science disciplines. These broad research themes include, but are not limited to: impacts of e-commerce on retail competition and urban morphology, the development of e-commerce infrastructures and networks, environmental externalities of digital consumption, and *the drivers of e-commerce adoption*. It is this final broad theme that the balance of this paper is concerned with. The subsequent sections examine two of the most essential drivers of e-commerce consumption: demographics and spatial dynamics.

Demographic Forces Driving E-commerce Consumption

This section of the chapter surveys quantitative studies in geography, business, and economics that examine the demographic characteristics that drive consumers to purchase goods and services online. The majority of the studies surveyed echo recurrent trends and patterns in the demographics of e-commerce consumers. The empirics reviewed here defend that e-commerce consumers are, in general, young well-educated consumers with high household incomes. In the face of this overarching trend, there are some interesting deviations as a result of different products, countries, and purchase channels. There is also a diversity in the analytical methods used to quantify each demographic driver, including discrete choice models, structural equation models, segmentation methods, and other tools.

Undoubtedly, regression analysis is the most commonly applied tool to understand demographic drivers of e-commerce consumption. This is likely in no small part due to its ease of interpretation and the potential to derive linear incremental relationships between demographic characteristics and e-commerce activity or propensity to spend. For the majority of these regression-based studies, a discrete choice model is employed in which the dependent variable is some sort of household or individual decision to shop online. This decision is then regressed against various demographic attributes. These studies are catalogued in Table 2.1. The E-commerce Variable column highlights the key explanatory variable(s) tied to e-commerce consumption and/or adoption that were analyzed in the methodology. The Significant Demographic Variables column calls out specific demographic features that were significant in driving e-commerce consumption.

From Table 2.1, it is clear that almost all previous studies at least partially corroborate the narrative that e-commerce consumers are young, well-educated, and affluent. Krizek et al. (2005) apply a logistic regression to household survey data from three American cities. They focus on two dependent variables of interest: a binary variable indicating whether a household purchases online frequently or infrequently, and a binary variable indicating whether the household has ever shopped online. They find that younger households with higher incomes tend to purchase online more frequently. Additionally, education becomes significant in determining whether a household has ever bought online, with more well-educated households opting to use digital channels. Farag et al. (2006a) deploy a logistic regression to survey data from the Netherlands. With a binary dependent variable that indicates whether an individual has ever shopped online, they echo the findings of Krizek et al.: online shoppers are more likely to be younger consumers with higher incomes. Interestingly, they identify a nonlinear relationship

Table 2.1. Regression-based Studies of Demographic Drivers of E-commerce Consumption

Author(s)	Geography	Spatial Dependence	Method	E-commerce Variable	Significant Demographic Variables
Krizek et al. (2005)	United States	-	Logistic Regression	Binary variable denoting whether a household shops online frequently or infrequently	Age Income
Krizek et al. (2005)	United States	-	Logistic Regression	Binary variable denoting whether a household has ever shopped online	Education
Farag et al. (2006a)	Netherlands	-	Logistic Regression	Binary variable indicating whether an individual has ever shopped online	Age Gender (Male) Income Education
Soopramanien & Robertson (2007)	United Kingdom	-	Logistic Regression	Binary variable indicating whether an individual shops online	Age Income
Weltevreden & van Rietenbergen (2007)	Netherlands	-	Multinomial Logit Model	Multinomial variable distinguishing non-eshoppers, online searchers, and e-shoppers	Age Gender (Male)
de Blasio (2008)	Italy	-	Logistic Regression	Probability that a household has purchased goods online	Age Education Income
Ren & Kwan (2009)	United States	-	Logistic Regression	Binary variable indicating whether an individual shops online	Gender (Female) Race
Punj (2011)	United States	-	Logistic Regression	Binary variable indicating whether an individual shops online	Age Education Income
Clarke et al. (2015)	United Kingdom	-	Logistic Regression	Binary variable indicating whether an individual uses the internet often to purchase groceries	Age Gender (Male) Household size Income Vehicle ownership
Van Deursen et al. (2015)	Netherlands	-	Linear Regression	Ordinal measure on a five-point scale ('never' to 'daily') of how often a respondent uses the internet for commercial transactions	Gender (Male) Age Education Income
Van Droogenbroeck & Van Hove (2017)	Belgium	-	Logistic Regression	Binary variable indicating whether a household chooses to adopt a "Collect & Go" program offered by a grocery chain	Age Education Household size Presence of children More than one employed adult
Beckers et al. (2018)	Belgium	-	Logistic Regression	Binary variable indicating whether an individual shops online	Education Gender (Male) Income
Loo & Wang (2018)	China	-	Ordered Logit Regression	Multinomial variable indicating amount of time spent e-shopping daily (0-30 minutes, 30-60 minutes, > 1 hour)	Age Gender (Female) Vehicle ownership

Table 2.1. continued

Author(s)	Geography	Spatial Dependence	Method	E-commerce Variable	Significant Demographic Variables
Zhen et al. (2018)	China	-	Multinomial Probit Regression	Multinomial variable indicating whether a consumer pre-purchases books in-store, pre-purchases online, or shops in-store only	Age Education
Zhen et al. (2018)	China	-	Multinomial Probit Regression	Multinomial variable indicating whether a consumer pre-purchases clothing in-store, pre-purchases online, or shops in-store only	Gender (Female) Income
Etumnu et al. (2019)	United States	-	Logistic Regression	Binary variable indicating whether an individual shops online	Age Education Gender (Male)
Saphores & Xu (2020)	United States	-	Logistic Regression	Binary variable indicating whether an individual shops online	Age Education Gender (Female) Race
de la Llave Montiel et al. (2020)	Spain	Yes	Spatial Durbin Probit Model	Binary variable indicating online retail churn (whether a customer becomes inactive online)	Income
Pernot (2021)	France	-	Logistic Regression	Binary variable indicating whether an individual shops online	Age Education Household size Income
Ganning & Green (2021)	United States	No	Linear Regression	Percent of consumer expenditure via online shopping by county	Age Income Race
Shao et al. (2022)	China	Yes	General Spatial Model (SAC/SARMA)	Online shopping activity as measured by Alibaba's Online Shopping Index	Age Education Income

between age and likelihood of purchasing online. The likelihood of purchasing online increases with age to a point, but begins to decrease after age 33, controlling for other factors. Similarly, in their study of online shopping adoption in the United Kingdom, Soopramanien and Robertson (2007) find that online buyers tend to be younger with higher incomes. Weltevreden and van Rietenbergen (2007) find that younger males are more likely to be e-shoppers in their survey of the Netherlands. However, they stratify two models for car users and "other transport users". Gender is not a significant factor for car users, while age is not a significant factor for users of other modes of transportation. De Blasio's (2008) study of Italian household surveys finds that younger, more educated households with higher income have a higher probability of buying goods online. Punj (2011) uses telephone interviews of American consumers to develop a dependent variable capturing whether or not an interviewed consumer is an online shopper or not. They find that online shoppers tend to have higher income and education levels and tend to be younger. Van Deursen et al. (2015) use linear regression to understand the relationship between demographic characteristics and Dutch survey respondents' responses of how frequently they use online shopping (five-point scale from 'never' to 'daily'). They find that younger, wealthier consumers are more likely to shop online. Their findings from education do deviate from other studies, with less educated consumers responding as more likely to shop frequently online – though it is worth noting that this finding does not account for actual dollars spent online. In a study of Belgium consumers Beckers et al. (2018) use a survey to derive a dependent variable reflecting whether or not a respondent has shopped online in the past year. Again, young, affluent consumers are more likely to shop online. Loo and Wang (2018) model both online shopping and e-working in tandem in Nanjing, China using household survey data. Their ordered logit model demonstrates that younger consumers who do not normally drive a car to shop are more likely to shop online. In their study, the coefficient magnitude of income was relatively small compared to other demographic features. Zhen et al. (2018) also study Chinese consumers. Their trivariate probit model identifies characteristics of shoppers purchasing two different goods: clothing and books. They find that online book shoppers tend to be more educated, with a nonlinear relationship between age and online shopping – similar to Ren & Kwan (2009). Online clothing shoppers were found to have higher incomes. From Etumnu et al. (2019), young, well-educated consumers are also identified as being most likely to shop online. Most recently, Ganning and Green (2021) use a linear regression to regress demographic metrics at the county level on percent of consumer expenditure via online shopping. They find that younger counties with higher incomes are more likely to see greater online spending. Across a swathe of countries, retail sectors, years, and methods – the trend of youth and affluence persists.

This general theme that e-commerce adopters tend to be young, well-educated and affluent holds true even for more inelastic goods such as groceries. Crucially, Clarke et al. (2015) provide one of the first regression-based models for online grocery purchasing, specifically. They use Acxiom's Research Opinion Data to measure e-commerce adoption in the United Kingdom to build a binary dependent variable capturing whether a respondent uses the internet "often" to purchase groceries or not. They identify young, wealthy males in smaller households as being most likely to purchase online often.

Interestingly, they also incorporate a novel data point – number of cars. Even controlling for income, households with more cars are more likely to shop online. In a similar study, Van Droogenbroeck and Van Hove (2017) use a survey of Belgian supermarket shoppers to understand what sociodemographic characteristics impact the adoption of online shopping. In this case, the dependent variable captures the customers' decision to adopt a "Collect & Go" program for the supermarket chain. They identify younger, affluent households with young children and multiple working adults as the most likely to adopt online grocery shopping. Saphores & Xu (2020) use data from the National Household Travel Survey to build a dependent variable indicating whether or not a respondent shopped online for groceries. The design a logit model to demonstrate consumer characteristics that are indicative of E-grocery shopping. They find that online shoppers tend to be younger, well-educated consumers. In a recent survey study out of France, Pernot (2021) uses a dependent variable to indicate whether a respondent purchases goods online and picks them up at a store. They find that e-commerce adopters in France tend to be younger, affluent households with a smaller household size.

A variable for which there appears to be less of an empirical consensus or pattern is gender. Some studies suggest that e-commerce users tend to be male (Farag et al., 2006a; Weltevreden & van Rietenbergen, 2007; van Deuren et al., 2015; Clarke et al., 2015; Beckers et al., 2018; Etumnu et al., 2019). Others yield results that point to females as being more likely to adopt e-commerce than males (Ren & Kwan, 2009; Loo & Wang, 2018; Zhen et al., 2018; Saphores & Xu, 2020). There are any number of factors that could explain this lack of consistency, including geography, nationality, retail sector, or product. Ren & Kwan (2009) produce a particularly interesting study that counters the general theme of youth and affluence discussed previously. They use an Internet diary survey of the Columbus, Ohio metropolitan area to create a binary variable indicating whether a consumer shops online. A logistic regression reveals that women are more likely to purchase online than men. Race also played a role here, as white consumers were more likely to purchase online than black consumers, controlling for other variables. A Poisson regression was also implemented in the study to regress the number of online purchases made within a year on demographic features. Households with children are found to have a higher incidence of online purchases over a year's time, controlling for other factors. Other variables such as income and age were not found to have a strong effect in the study. A few other studies have examined race as a determinant of e-commerce adoption. Saphores and Xu (2020) find that African-American and Asian consumers are more likely to never order goods online compared to white consumers. Conversely, Green and Ganning (2021) find that counties with higher black populations actually see higher online spending, though they acknowledge there may be other regional dynamics driving this counter-intuitive result.

A second theme in these quantitative studies of demographic drivers of e-commerce is spatial dependency. Spatial dependence is a core concept of quantitative analysis in geography which is noticeably absent from the traditional aspatial regression-based studies mentioned previously (Darmofal, 2015; Chi & Zhu, 2020). A relatively smaller and newer group of regression-based studies of e-commerce consumption are beginning to argue the relevance of spatial dependence how consumers shop in the digital world, just as it is relevant in brick-and-mortar retailing (Daunfeldt et al., 2017; Hunneman et al., 2021; Sung, 2022). Shao et al. (2020) use Alibaba's Online Shopping Index to design a regression model to explore the role of accessibility and consumer demographics in the incidence of heightened e-shopping activity. The accessibility piece will be addressed directly in the next section. They find that age, income, and education are all statistically significant demographic factors in driving e-shopping activity. In addition, they defend the necessity of a spatial regression model to capture dependence effects in the dependent variable. The presence of spatial autocorrelation in China's e-commerce landscape has been previously demonstrated (Hai-dong et al., 2017), and must be controlled for in the context of regression. Indeed, spatial dependence violates the autocorrelation assumption, and can even violate other assumptions such as homoscedasticity. These violations may originate from explicit spatial processes or interactions between neighboring units of observation (also called spatial diffusion), or they may be the result of a kind of cluster effect, in which proximal units of observation inherently sharing similar characteristics without interacting directly. In the case of ordinary least squares, failure to account for spatial diffusion effects may yield biased and inconsistent regression parameters. Failure to account for spatial clustering effects can also yield inefficient parameter estimates and, potentially, Type 1 errors (Darmofal, 2015). To control for spatial dependence, they opt for spatial lag model over an aspatial regression. Other regression-based methods that have been applied to e-commerce consumption have incorporated a spatial econometric model specification. In one of the earliest quantitative studies of e-commerce geographies, Jank & Kannan (2005) build a spatial multinomial logit model to measure the impact of geographic factors on web-based book consumption, though their study is not focused on demographic covariates. Wieland (2021a) uses a spatial form of a multinomial discrete choice model in his study of spatial shopping patterns in an omni-channel environment in Germany. A spatial regression model is also used to study online retail churn in Spain by de la Llave Montiel et al. (2020). There remains an outstanding opportunity to produce more studies of e-commerce consumption that recognize and measure potential spatial dependence, revealing new spatial patterns and relationships between demographic features and e-commerce adoption by consumers. Even Ganning and Green (2021) who do not find statistically significant evidence of spatial dependence in their study still recognize the need to test for any presence of spatial autocorrelation in models of retail consumption and consumer behavior.

It is noteworthy that some other methods have been utilized to study demographic features associated with e-commerce adoption, though they are considerably less common. Cao et al. (2013), Driediger and Bhatiasevi (2018), and Hamad & Schmitz (2019) all employ structural equation models to study the impact of demographic and geographic factors on survey respondents' propensity to shop online. Bell and Song (2007) construct a discrete time hazard utility model to measure the impacts of regional, local, and household covariates on trials of Netgrocer's web-based grocery delivery service in the early 2000s. Though this spatio-temporal model is more rigorous and granular than the regression models previously discussed, many of the key results are similar. The utility of net new trials is positively related to demographics that are male, young, wealthy, and white. Hood et al. (2020) conduct a survey of households in Great Britain and use categorical data analysis and inferential statistics to find that grocery e-commerce users tend to be affluent young females, supporting the findings of Saphores and Xu (2020). Wieland (2021b) presents a hurdle model to regress the utility of various shopping options (including online channels) on various demographic and geographic factors. He finds that younger consumers are more likely to shop online, though findings related to gender and employment are inconclusive. It is worth calling out that other empirical studies have not been concerned with identifying the demographic drivers of e-commerce, but rather, segmenting all e-commerce users based on their demographic and behavioral characteristics (Kau et al., 2003; Rohm & Swaminathan, 2004; Harris et al., 2017; Alexiou et al., 2018).

Irrespective of the methodology, this catalog of past empirical studies calls out some key patterns in the demographics of e-commerce users. Though there are a few exceptions, e-commerce adopters are, generally, young, well-educated, and wealthy. It is worth noting that this general consumer profile roughly matches the profiles of early adopters of various technological innovations throughout the 21st century, including online banking (Lee et al., 2003) and ridesharing (Dias et al., 2017). These established trends can form the empirical foundation for future hypotheses in modeling exercises where demographic variables are incorporated in the context of online retailing. Poignantly, these findings remain consistent thematically with some of the earliest studies of e-commerce adopters surveyed in the historical section of this publication (Dwyer, 1996, 1998; Li et al., 1999). Even in the earliest days of e-commerce, affluent consumers were emphasized as the greatest adopters of online shopping. However, it is crucial to note that the demographic composition of e-commerce users must inevitably shift and change as time goes on. Generations with higher adoption rates of e-commerce and other technologies will age up over time. It is feasible to imagine a world in the future in which all age groups use e-commerce equally. Further, demographics are, and will always be, only one piece of the puzzle. Another driving force in e-commerce consumption is spatial dynamics. These dynamics take different forms, and are fueled by two core geographic theories which will be surveyed in-depth in the subsequent section.

Spatial Dynamics Driving E-commerce Consumption

The previous studies in this literature review have focused primarily on historical foundations of e-commerce geographies, or the demographic drivers of e-commerce adoption and consumption in developed retail economies. These studies, however, have not directly addressed the spatial dynamics that play an equally essential role in determining whether a consumer will adopt e-commerce channels in their shopping. This section will survey quantitative studies that have attempted to incorporate these spatial dynamics in their analysis of drivers of e-commerce consumption. This type of review must begin with the two key theories that implicitly or explicitly inform the empirical findings of the many quantitative studies that succeeded them.

Perhaps one of the most substantial works that has explicitly connected e-commerce to geographic theory, Bell's (2014) *Location is (Still) Everything.* Bell catalogs powerful geographic forces in retail resulting from the rise of e-commerce, with some brief case studies and anecdotes to reinforce his theorizations. These spatial forces are: *resistance, adjacency, vicinity, isolation,* and *topography. Resistance* refers to the frictions and barriers encountered by consumers when attempting to search for and purchase goods. E-commerce allows some forms of resistance, particularly those tied to searching across space, to be decreased or eliminated entirely. Consumers are free to hunt for the right good at the right price, even if the good is from a seller who is further away in space. *Adjacency* refers to the foundational concept of spatial dependence (also termed, *homophily* or *Tobler's law of geography*) in which nearby things are related or similar. Consumers who are adjacent to consumers who have already adopted e-commerce are more likely to also adopt e-commerce. *Vicinity* is a force in the digital world that describes the ways in which consumers who share similar beliefs, demographics, and buying behaviors may not always be adjacent in space, but can still interact and influence each other in digital spaces. *Isolation* is not a force necessarily tied to the physical world. Bell describes this force as a kind of "preference isolation" or "preference minority". Online sellers should feel incentivized to cater to more niched goods that cannot be obtained in brick-and-mortar retailers who must always assort to cater to the majority of their market. Offering unique goods that the preference isolated consumer cannot find at their local store marks an opportunity for the e-commerce business. Bell's final force, *topography*, is again abstracted beyond its more traditional understanding via physical geography, to ascribe a digital topography in which customers seamlessly and simultaneously navigate and move through both digital and physical spaces to purchase goods. Consumers check online for reviews before buying products at the store, they test products for features and fit in stores before ordering online. Increasingly, digital and physical marketplaces collide in the customer journey.

Anderson et al. (2003) produced a largely exploratory paper in the early days of online shopping that attempted to project the future consequences of e-commerce through a spatial lens. Many of the projections made by Anderson et al. came to be defining characteristics of e-commerce today, including the re-organization of distribution networks as central hubs with secondary sites for place-specific service, the general decline of some formats of brick-and-mortar retail networks and, most importantly, the idea that location still matters. In addition to these projections, Anderson et al. also introduce us to competing theories of e-retail adoption:

> One can make two plausible hypotheses. First, that urban populations are most likely to adopt e-retail because they are better educated and more likely to use the Internet actively for other purposes, and second, that people in non-metropolitan counties will adopt most rapidly because they have the most to gain from access to the wide variety of goods provided on the Internet. (p. 421).

The first hypothesis has been coined the *innovation-diffusion hypothesis*, the second has been termed the *efficiency hypothesis*. In empirical settings, these two hypotheses seem to be constantly at war with each other, even occurring in tandem on occasion. The overarching theme of research investigating spatial dynamics driving e-commerce adoption and consumption is *volatility* – the absence of any clear consensus on which of hypotheses from Anderson et al. hold true in the modern retail economy.

In similar fashion to demographic variables, spatial variables have been most commonly analyzed via regression and other deterministic models. This review parses these spatial variables into two broad categories: *urbanity* and *accessibility.* Of course, these two concepts are highly collinear in practice, but they do refer to distinct types of variables and/or dimensions incorporated into studies of e-commerce consumption. Urbanity encompasses variables that reflect whether a market or place is urban, suburban, exurban, etc. These include variables such as population density or raw population. Accessibility variables are designed to capture the clustering of amenities, competitors, or other points of interest in a market. These include variables such as retail density, competitive saturation, distance, and travel time. Table 2.2 builds off of Cao's (2009) previous review of spatial attributes of e-commerce consumption, providing an outline of studies of e-commerce that explicitly incorporate variables capturing spatial dynamics. While Table 2.1 catalogs demographic variables used in studies of e-commerce adoption, Table 2.2 provides an inventory of variables capturing spatial dynamics in e-commerce consumption. The E-commerce Variable column marks the key variable(s) of interest in the study tied to e-commerce adoption and/or consumption. The Geographic Variable column captures the specific spatial dynamic captured in a given study's methodology. The Type column classifies this spatial variable under accessibility or urbanity, and the Hypothesis column indicates whether any parts of the study's empirics support the efficiency hypothesis or the innovation diffusion hypothesis. If neither hypothesis is explicitly supported, there is no effect.

Three of the studies catalogued in Table 2.2 provide evidence to support the innovation diffusion hypothesis. Farag et al. (2007) incorporate shops within a ten-minute cycling time as a proxy for retail accessibility, and street address density as a proxy for urbanity to understand frequencies of online purchasing. The results of their structural

Table 2.2. Studies of Spatial Dynamics as Drivers of E-commerce Consumption

Author(s)	Geography	Method(s)	E-commerce Variable(s)	Geographic Variable(s)	Type	Hypothesis
Farag et al. (2005)	Netherlands	Path Analysis	Frequencies of online searching and online buying	Urban versus suburban	Urbanity	No effect
Krizek et al. (2005)	United States	Logistic Regression	Online buying adoption Online buying frequency Online search adoption	Distance to CBD Retail accessibility	Accessibility and Urbanity	No effect
Farag et al. (2006a)	Netherlands	Logistic Regression and Linear Regression	Online buying adoption Online buying frequency	Dummy variables for retail density	Accessibility	Adoption supports Innovation Diffusion Frequency supports Efficiency
Farag et al. (2006b)	United States Netherlands	Logistic Regression	Online buying adoption Online buying frequency	U.S. – shops within walking distance Netherlands – travel time	Accessibility	Dutch adoption supports innovation diffusion
Weltevreden & van Rietbergen (2007)	Netherlands	Multinomial Logit	Online buying adoption	Retail density	Accessibility	No effect
Farag et al. (2007)	Netherlands	Structural Equation Model	Frequency of online searching Frequency of online buying	Shops within 10 minutes by bike Street address density	Accessibility and Urbanity	Innovation Diffusion
de Blasio (2008)	Italy	Logistic Regression	Probability of online purchase	City size	Urbanity	No effect
Ren & Kwan (2009)	United States	Logistic Regression	Online buying adoption	Retail density	Accessibility	Efficiency
Cao et al. (2013)	United States	Structural Equation Model	Online buying frequency	Retail density Population density	Accessibility and Urbanity	Innovation Diffusion in Urban Efficiency in Exurban
Clarke et al. (2015)	United Kingdom	Logistic Regression	Online buying frequency	Population density	Urbanity	Efficiency
Zhen et al. (2018)	China	Multinomial Probit	Online or in-store buying	Travel time to retail Population density	Accessibility and Urbanity	Both
Kirby-Hawkins et al. (2019)	United Kingdom	Quadrant Analysis	Online sales data from retail chain	Population density	Urbanity	Both
Beckers et al. (2018)	Belgium	Logistic Regression	Online buying adoption	Population density	Urbanity	No effect
Hood et al. (2020)	United Kingdom	Descriptive Analysis	Surveyed online grocery shoppers	Population density	Urbanity	Both
Wieland (2021b)	Germany	Hurdle Model	Consumer channel choice	Competitive saturation Travel time City size	Accessibility and Urbanity	Innovation Diffusion
Ganning & Green (2021)	United States	Linear Regression	E-retail spending by county	Retail accessibility Distance to MSA centroid	Accessibility	Efficiency
Shao et al. (2022)	China	General Spatial Model (SAC/SARMA)	Online shopping activity index	Retail density	Accessibility	Innovation Diffusion

equation model suggest that urban consumers are more apt to shop online frequently than rural consumers. Farag et al. (2006a) also reckon with hypotheses of efficiency versus innovation diffusion. Their model of e-shopping includes variables for urbanization and the agglomeration of retail shops. In contrast to Krizek et al., both variables return a strong and significant positive relationship with e-commerce activity, supporting the innovation-diffusion hypothesis. Interestingly, they find that urban consumers are certainly more likely to shop online, however, the majority of online orders still comes from rural consumers. Wieland's (2021) hurdle model includes measures of competitive saturation, travel time, and city size in an analysis of channel choice by German consumers. His results also signal that urban consumers are more likely to select a digital shopping channel. However, Wieland does call out that internet access in Germany is unequally distributed which may be biasing his results. Shao et al. (2022) also find evidence to support the innovation diffusion hypothesis. They incorporate retail density into their spatial regression to capture accessibility in determining online shopping activity. Their results suggest that urban consumers are more likely to consume digitally, in part because smaller cities in China do not have the necessary infrastructure to incentivize e-shopping. Implicitly, all of these studies suggest that rural consumers are have no significant incentive to use e-commerce. Per de Blasio, there is some intuition as to why remote customers may resist adopting e-commerce:

> . . . the propensity to shop on the Internet is a matter not only of cost and convenience but also of culture and infrastructure. First, knowledge about the possibilities of the Internet may be lacking. Insofar as imitation is a crucial factor for online shopping, rural residents might find it more difficult to observe and try e-commerce. Second, there might be inefficiencies in the payment phase or in the parcel delivery service. These obstacles can jeopardize the prospect of e-commerce

> in remote areas as high-quality support services may be more readily available in urban settings. (p. 352).

In opposition to de Blasio's assertion, there are three studies listed above that produce findings that support the efficiency hypothesis. Ren and Kwan (2009) find that shopping accessibility has a negative influence on the adoption of online buying in Columbus, Ohio. Clarke et al. (2015) also yield evidence in favor of the efficiency hypothesis. They find that even in urban areas, online consumers tend to have less accessibility to retail offerings than consumers who shop in-store. Ganning and Green (2021) also confirm the efficiency hypothesis, with the caveat that their conclusions assume no incremental changes to shipping or logistical costs in remote areas. This is, of course, rarely the case, as more remote areas will tend to have higher costs for shipping goods ordered online. They qualify:

> The internet has not erased geography. Geography means low-income, remote consumers still must travel to population centers for necessities, and doing so incentivizes completing all purchases during those trips rather than paying any additional monies for shipping. (p.1267).

As evidenced by Table 2, there is a jarring lack of consistency in how empirical studies have supported Anderson et al.'s (2003) original hypotheses. The conclusions drawn by studies of e-commerce adoption demonstrate that Anderson et al.'s hypotheses are, at the very least, not mutually exclusive, and almost certainly subject to spatial and contextual circumstance. No substantial body of research has definitively favored one hypothesis over the other. Indeed, the majority either support both hypotheses under different contexts, or yield no significant effect to support either hypothesis. Farag et al.

(2006a), Cao et al. (2013), Zhen et al. (2018), Kirby-Hawkins et al. (2019), and Hood et al. (2020) all find evidence that support both hypotheses. Other studies don't find statistically significant evidence of either hypothesis (Farag et al., 2005; Weltevreden & van Rietbergen, 2007; Beckers et al., 2018). Krizek et al. (2005) incorporate variables to test both accessibility and urbanity in determining e-commerce adoption by consumers in the United States. These are proxied through consumers' distance to the nearest central business district and by retail accessibility, respectively. Their results produce no meaningful effect from either spatial dynamic. It may be argued that the relationships between accessibility and urbanity with e-commerce consumption will continue to change in the same way that the demographic composition of e-commerce consumers is also changing. In the future, e-commerce will be ubiquitous, regardless of demography or geography. Wieland (2021b) contends that the innovation diffusion hypothesis and efficiency hypothesis may becoming gradually adjacent and simultaneous.

> Indeed, it is questionable if this spatial difference still holds true when considering 1) the consolidation of e-shopping over more than two decades, and 2) the availability of fast (broadband) internet for the vast majority of households in both rural and urban areas. (p. 348).

In light of the seeming absence of any conclusive empirics regarding spatial dynamics and e-commerce, one would be forgiven for thinking that data related to geographical context (urbanity, density, etc.) is simply irrelevant in geographical analyses of e-commerce. On the contrary, this absence of conclusiveness and empirically defended hypotheses demands the inclusion of spatial dynamics all the more. Future empirical studies of the geographies of e-commerce must continue to situate their stories in geographic contexts – not only to control for spatial dependence in methodological approaches, as was discussed in the demographic section of this review – but to deeper investigate Anderson et al.'s competing (and potentially converging) theories of e-commerce adoption and consumption. This exploration is also crucial for informing strategy and decision making for retailers seeking to implement effective e-commerce strategies across their chains. The existence of these theories in tandem may signal that retailers must adopt different e-commerce strategies for urban and rural markets, but to implement those strategies intelligently, further investigation of the efficiency and innovation diffusion hypotheses is needed.

Conclusion

This literature review has logged a body of empirical literature concerned with methods to determine whether an individual or group of individuals will adopt e-commerce for their shopping needs. The review began with some of the historical foundations of geographical studies of e-commerce, identifying connections across time between modern and aged studies that identify the demographics most likely to use e-commerce. Despite some hyperbole in the early days of e-commerce, the imperishable spatiality of digital retailing has been more than adequately defended by geographers in the 21st century. The second part of this review pivoted to a discussion of two types of data that are shown to be determinant of likelihood to adopt e-commerce: demographics and spatial dynamics. These two sections inventoried studies spanning over two decades, covering multiple countries, and multiple geographic scales – from cities to entire nations. A recurring demographic theme in these quantitative studies is the prominence of wealthy, well-educated individuals as digital consumers. It should be recognized that there remains an opportunity to incorporate spatial dependence into these demographic studies to bolster methodological rigor, and evaluate any evidence of spatial autocorrelation in retail activity that occurs in digital space. The subsequent discussion of spatial dynamics focused on two seemingly contradictory theories laid out by Anderson et al. (2003): the efficiency theory and the innovation diffusion theory. A survey of studies that incorporate variables tied to population density, urbanity, retail agglomeration, and other proxies for place reveal that these theories are not mutually exclusive. There is sufficient evidence that they may coincide, or not occur at all. However, this does not signal that geography is not a key component in determining or predicting e-commerce demand.

This review lays the foundation for novel streams of inquiry related to geographies of e-commerce consumption. One such stream incorporates a temporal dimension into e-commerce consumption research. Panel and time series data can offer a richer, more nuanced view of e-commerce activity across space. This temporal dimension could be incorporated through the analysis of longitudinal data (Driediger, 2019; de la Llave Montiel, 2020), or potentially, a behavioral analysis of how digital and physical consumption decisions are made by consumers - either sequentially or in tandem (Wieland, 2021a; 2021b). A second avenue of future research must be, of course, the COVID-19 pandemic. The pandemic had a dramatic impact on the retail economy, as many brick-and-mortar chains saw reduced foot traffic and poor performance as consumers were forced to shop online for some goods that may have previously been purchased in-store. Some chains that offer essential goods, such as grocers, discounters, and wholesale clubs, carrying saw exponential increases in traffic and demand for cleaning supplies, toiletries, and foodstuffs. Over the first year of the pandemic in the United States, from March 2020 through March 2021, e-commerce sales increased by $183 billion (Verdon, 2021). There is an argument to be made that this rate of growth and adoption for e-commerce transactions would have been reached at a much later time, given the past growth that e-commerce has enjoyed for two decades, but the pandemic exponentially expedited this growth. In this regard, COVID-19 is, in its own right, a driver of e-commerce consumption and competition (Beckers et al., 2021; Pisani, 2021; Wieland, 2022). This marks a new opportunity to conduct geographic research to understand the spatial structures and processes underlying this the rise of e-commerce and how it affects producers and consumers in a post-pandemic world. In retail geography the pandemic remains a new frontier of scholarship, with new research arriving as data connecting the pandemic to retail geographies becomes more available. Further avenues of research could also take a comparative view of the demographics driving a swathe of technological innovations in consumption that have manifested in the 21st century. While some studies point to a general trend of adoption stemming from young, affluent consumers (Lee et al., 2003; Dias et al., 2017; Ali et al., 2020), there remains an opportunity to contrast some of the unique nuances of consumers for these various innovations, particularly digital consumption innovations that may benefit or target less affluent consumers, such as music streaming (Datta et al., 2017). A final line of inquiry related to e-commerce consumption concerns the decision of the firm. Almost all of the literature reviewed here emphasizes the behaviors of consumers: *who* shops *where*, *why* they shop *there*, and *how*. Little emphasis has been placed on how retail firms are making decisions about how to deploy omni-channel offerings across their chains. Even retailers that were once exclusively digital have begun to establish physical store networks (Wang & Du, 2021). Pivoting to the perspective of the firm raises exciting new questions about how they approach e-commerce, and there is a need for researchers to engage more with retail firm data to answer these questions (Kirby-Hawkins et al., 2019). One of the most intriguing questions here may be whether or not retail e-commerce strategies are conforming to key patterns and findings that have been previously established for geographies of e-commerce consumption.

Chapter Three

Identifying the Determinants of E-Commerce Offerings Among Grocery Store Locations in the United States: A Spatial Econometric Analysis

Introduction

In May 2023, the U.S. Census Bureau (2023a) published a report of quarterly e-commerce sales spanning 2014 to 2023. E-commerce sales in the first quarter of the year were estimated to total $272.6 billion, a lift of 3% over the previous year. For the grocery sector, online purchases totaled $86.8 billion in 2022, and e-commerce sales saw year-over-year growth of 26.7% in February 2023 (Verdon, 2023). Research firms project this growth to continue into the future, with online grocery sales expected to maintain a compound annual growth rate of 11.7% over the next five years (Moran, 2023a). Certainly, the growth of e-commerce in the grocery sector can also be largely attributed to the COVID-19 pandemic. Online grocery sales in the United States more than doubled during the pandemic, with average monthly spending increasing by $3.1 billion from pre-pandemic levels (Verdon, 2022). Undoubtedly, e-commerce has become cemented as a new channel for purchasing groceries in the United States, and its prevalence is expected to grow over time. Geographers, economists, and business scholars have begun efforts to better understand this growth, including identifying key patterns and trends in the demography and geography of e-commerce consumption (Ganning & Green, 2021; Saphores & Xu, 2020; Beckers et al., 2018). A majority of these studies involve analyzing how key demographic characteristics or spatial dynamics may influence consumers' decisions to shop online or in stores. Prevailing trends among these studies indicate that consumers who purchase online tend to be younger and more affluent.

Competing theories are attributed to how spatial dynamics may influence online shopping habits. The *efficiency hypothesis* suggests that consumers in non-metropolitan areas will adopt e-commerce more rapidly than consumers in urban areas due to limited accessibility to retail centers in more rural areas (Anderson et al., 2003). The *innovation-diffusion hypothesis* argues that consumers in urban areas will be more likely to embrace e-retailing because they tend to be better educated and are more active Internet users (Anderson et al., 2003). No consensus has been reached empirically to confirm one hypothesis over another, and spatial dynamics remain a key theme of inquiry into geographies of e-commerce across the world.

While scholars have made great efforts to analyze both demographics and spatial dynamics and their relationship with e-commerce adoption across a variety of retail sectors, there remains an outstanding opportunity to pivot focus to the decisions being made by the firms servicing e-commerce consumers (Kirby-Hawkins et al., 2019). To that end, this paper makes an important contribution to existing literature by building off prior studies of consumer choice and shifting focus from the decisions of customers to the e-commerce strategies adopted by grocery chains across their store networks. The findings from this research provide a more nuanced view of the geography of e-commerce in the United States, and yield actionable insights for retail planners and developers who must critically consider the e-commerce strategies of grocery chains to plan shopping center tenancy, food accessibility, local/regional economic development, or competitive defense strategies. The study uses cross-sectional sample of grocery location data to assess the geographic patterns underlying various e-commerce offerings for grocery stores in the contiguous United States. Grocery stores may offer delivery, pickup, both delivery and pickup, or none. Through exploratory spatial data analysis, geostatistics, and spatial econometric modeling, the demographic and geographic patterns underlying these varied e-commerce strategies were then examined against past empirical research focused on consumption patterns to gauge the extent to which grocery chains are building their e-commerce location strategies in correspondence with prevailing demographic trends and spatial dynamics associated with e-commerce adoption and consumption. Crucially, this study will also explores spatial heterogeneity among e-commerce location strategies by running additional analysis and modeling for a regional case study based on delineations set by the U.S. Bureau of Economic Analysis (2023). The regional focus of this case study is the Southeastern United States. Introducing analysis and modeling procedures specific to this region uncovered interesting relationships and patterns that deviate from the broader trends of the national model. Notably, while a model at the national level highlights that trade areas with higher proportions of black population are less likely to have e-commerce offerings among grocery stores, the regional model for the Southeast reveals that trade areas with higher black populations are *more likely* to see grocery stores offering pickup and delivery.

Review of Literature

The prevailing demographic trend of younger more affluent consumers being more likely to purchase through digital channels is established by literature spanning almost two decades. Typically, the demographic characteristics of e-commerce adopters have been analyzed via regression analysis, specifically through discrete choice models.

These studies often regress a consumer's individual decision to shop online or not against their own demographic profile to identify which traits correspond to increased propensity to shop online. These studies are listed in Table 3.1. The E-commerce Variable field identifies the dependent variable in a given study, while the Significant Demographic Variables and Coefficient Sign column denotes which demographic variables were significant, and the sign associated with those variables.

Of the studies catalogued in Table 3.1, seven have focused specifically on e-commerce in the grocery industry. De Blasio (2008) finds that Italian households that are younger and more affluent tend to have a higher propensity to purchase grocery goods online. In the United Kingdom, young, wealthy, smaller households are identified as being most likely to frequently purchase groceries through digital channels (Clarke et al., 2015). This trend persists for consumers of a "Collect & Go" program for a Belgian grocery chain (Van Droogenbroeck & Van hove, 2017). Saphores & Xu (2020) use National Household Travel Survey data to build a logit model that points to younger, well-educated consumers as the most likely respondents to have previously shopped online. Similar survey data for the French grocery market to identifies younger, smaller, affluent households as the most likely consumers to have previously purchased food goods online to pick up at their local store (Pernot, 2021). This body of generally aligned findings inform the first hypothesis of this paper: *grocery locations offering some form of e-commerce to customers are more likely to be located in trade areas with young, affluent consumers with higher levels of education and income, controlling for other factors.*

Table 3.1. Literature Review - Demographics

Author(s)	Grocery Focus	Method	E-commerce Variable	Significant Demographic Variables and Coefficient Sign
Krizek et al. (2005)	No	Logistic Regression	Binary variable denoting whether a household shops online frequently or infrequently	Age [-] Income [+]
Krizek et al. (2005)	No	Logistic Regression	Binary variable denoting whether a household has ever shopped online	Education [+]
Farag et al. (2006a)	No	Logistic Regression	Binary variable indicating whether an individual has ever shopped online	Age [+/-] Gender (Female) [-] Income [+] Education [+]
Soopramanien & Robertson (2007)	No	Logistic Regression	Binary variable indicating whether an individual shops online	Age Income [+]
Weltevreden & van Rietenbergen (2007)	No	Multinomial Logit Model	Multinomial variable distinguishing non-eshoppers, online searchers, and e-shoppers	Age [-] Gender (Male) [+]
de Blasio (2008)	Yes	Logistic Regression	Probability that a household has purchased goods online	Age [-] Education [+] Income [+] Household size [+]
Ren & Kwan (2009)	No	Logistic Regression	Binary variable indicating whether an individual shops online	Gender (Female) [+] Race (White) [+]
Punj (2011)	No	Logistic Regression	Binary variable indicating whether an individual shops online	Age [-] Education [+] Income [+]
Clarke et al. (2015)	Yes	Logistic Regression	Binary variable indicating whether an individual uses the internet often to purchase groceries	Age [-] Gender (Male) [+] Household size [+] Income [+]
Van Deursen et al. (2015)	No	Linear Regression	Ordinal measure on a five-point scale ('never' to 'daily') of how often a respondent uses the internet for commercial transactions	Gender (Male) [+] Age [-] Income [+]
Van Droogenbroeck & Van Hove (2017)	Yes	Logistic Regression	Binary variable indicating whether a household chooses to adopt a "Collect & Go" program offered by a grocery chain	Age [+/-] Education [+] Household size [+]
Beckers et al. (2018)	No	Logistic Regression	Binary variable indicating whether an individual shops online	Education [+] Gender (Male) [+] Income [+]
Loo & Wang (2018)	No	Ordered Logit Regression	Multinomial variable indicating amount of time spent e-shopping daily (0-30 minutes, 30-60 minutes, > 1 hour)	Age [-] Gender (Female) [-]

Table 3.1. continued

Author(s)	Grocery Focus	Method	E-commerce Variable	Significant Demographic Variables and Coefficient Sign
Zhen et al. (2018)	No	Multinomial Probit Regression	Multinomial variable indicating whether a consumer pre-purchases books in-store, pre-purchases online, or shops in-store only	Age [+] Education [+]
Zhen et al. (2018)	No	Multinomial Probit Regression	Multinomial variable indicating whether a consumer pre-purchases clothing in-store, pre-purchases online, or shops in-store only	Gender (Female) [-] Income [+]
Etumnu et al. (2019)	Yes	Logistic Regression	Binary variable indicating whether an individual shops online	Age [-] Education [+] Gender (Male) [+]
Saphores & Xu (2020)	Yes	Logistic Regression	Binary variable indicating whether an individual shops online	Age [-] Education [+] Gender (Female) [+] Race (Black) [-] Income [+]
de la Llave Montiel et al. (2020)	Yes	Spatial Durbin Probit Model	Binary variable indicating online retail churn (whether a customer becomes inactive online)	Income [+]
Pernot (2021)	Yes	Logistic Regression	Binary variable indicating whether an individual shops online	Age [-] Income [+]
Ganning & Green (2021)	No	Linear Regression	Percent of consumer expenditure via online shopping by county	Age [+] Income [+] Race (Black) [+]
Shao et al. (2022)	No	General Spatial Model (SAC/SARMA)	Online shopping activity as measured by Alibaba's Online Shopping Index	Age [+] Education [-] Income [+]

Two demographic characteristics for which there is less empirical consensus regarding their relationship with e-commerce adoption are gender and race. Several studies point to men as the more likely e-commerce users across a variety of product types (Farag et al., 2006a; Weltevreden & van Rietenbergen, 2007; van Deuren et al., 2015; Clarke et al., 2015,; Beckers et al., 2018; Etumnu et al., 2019). Other studies contend that females are more likely than males to adopt e-commerce (Ren & Kwan, 2009; Loo & Wang, 2018; Zhen et al., 2018; Saphores & Xu, 2020). This inconclusiveness may be driven by geography, national culture, retail sector, or product. Ren & Kwan's (2009) study of Internet diary survey for the Columbus, Ohio metropolitan area demonstrates that white customers are more likely to purchase online than Black consumers, controlling for other variables. Similarly, Saphores and Xu (2020) find that African-American consumers are more likely than white consumers to never order goods online, controlling for both education and income. Green and Ganning (2021) find, however, that counties with higher Black populations actually see higher online spending. To apply these previous studies in this paper in order to derive a well-informed hypothesis, literature from retail redlining should also be considered. It is understood that many retail chains in the United States have, historically, avoided or abandoned trade areas with high Black populations in a practice known as retail redlining (Rowlands et al., 2023; Shannon, 2020; Zhang & Ghosh, 2016; Eisenhauer, 2001). The second hypothesis of this study is *grocery locations are less likely to offer e-commerce in trade areas with higher Black population, controlling for education and income.*

A second body of literature anchoring this study is concerned with spatial dependence in e-commerce consumption. Several more recent regression-based studies of e-commerce adoption and consumption have incorporated spatial autocorrelation into their methodologies, and have advocated that accounting and controlling for spatial dependency is just as important in studies of digital retailing as in physical retailing (Daunfeldt et al., 2017; Hunneman et al., 2021; Sung, 2022). Shao et al. (2020), in addition to identifying age, income, and education as significant demographic factors driving e-shopping, also demonstrate the necessity of spatial regression modeling to capture and control for dependence effects in their dependent variable, Alibaba's Online Shopping Index. Hai-dong et al. (2017) also find spatial autocorrelation in China's e-commerce landscape. Failing to control for spatial dependence in regression models of human activity where autocorrelation across space is common, retail being no exception, can result in inefficient models, and violations of regression assumptions, including homoscedasticity (Darmofal, 2015). Wieland (2021a) uses a spatial form of a multinomial discrete choice model in his study of spatial shopping patterns in an omni-channel environment in Germany. A spatial regression model is also used to study online retail churn for a grocery business in Spain by de la Llave Montiel et al. (2020). There remains an outstanding opportunity to produce more studies of e-commerce consumption that control for spatial autocorrelation, yielding more accurate model outputs and findings.

A third body of research that directly informs this study is concerned with the spatial dynamics of e-commerce. Anderson et al. (2003) presents two opposed theories that may explain how e-commerce is adopted across geographic space:

> One can make two plausible hypotheses. First, that urban populations are most likely to adopt e-retail because they are better educated and more likely to use the Internet actively for other purposes, and second, that people in non-metropolitan counties will adopt most rapidly because they have the most to gain from access to the wide variety of goods provided on the Internet. (p. 421).

The first hypothesis has been coined the *innovation-diffusion hypothesis*, the second has been termed the *efficiency hypothesis*. Various empirical studies have come out in favor or against both of these hypotheses. The volatility and inconsistency in empirical studies that have investigated these spatial dynamics suggest further research is required to more fully grasp the nuances behind when each hypothesis may be true or not. These spatial dynamics are typically captured through variables related to geographic accessibility to retail offerings or urban centers, or through measures of urbanity such as population density. Incorporating variables corresponding to one or both of these dimensions of accessibility or urbanity allows empirical studies to test for the efficiency and innovation-diffusion hypotheses. Table 3.2 builds off of a previous literature review from Cao (2009) of the spatial dynamics of e-commerce consumption. The E-commerce Variable field provides the dependent variable or variable of interest in a given study. The Geographic Variables field shows which variables related to accessibility or urbanity are measured against e-commerce consumption in a given study. The Hypothesis column specifies whether any parts of the study's empirics support the efficiency hypothesis or the innovation diffusion hypothesis. If neither hypothesis is supported, there is no effect.

Table 3.2. Literature Review – Spatial Dynamics

Author(s)	Grocery Focus	Method(s)	E-commerce Variable(s)	Geographic Variable(s)	Hypothesis
Farag et al. (2005)	No	Path Analysis	Frequencies of online searching and online buying	Urban versus suburban	No effect
Krizek et al. (2005)	No	Logistic Regression	Online buying adoption Online buying frequency Online search adoption	Distance to CBD Retail accessibility	No effect
Farag et al. (2006a)	No	Logistic Regression and Linear Regression	Online buying adoption Online buying frequency	Dummy variables for retail density	Adoption supports Innovation Diffusion Frequency supports Efficiency
Farag et al. (2006b)	No	Logistic Regression	Online buying adoption Online buying frequency	U.S. – shops within walking distance Netherlands – travel time	Dutch adoption supports innovation diffusion
Weltevreden & van Rietbergen (2007)	No	Multinomial Logit	Online buying adoption	Retail density	No effect
Farag et al. (2007)	No	Structural Equation Model	Frequency of online searching Frequency of online buying	Shops within 10 minutes by bike Street address density	Innovation Diffusion
de Blasio (2008)	Yes	Logistic Regression	Probability of online purchase	City size	No effect
Ren & Kwan (2009)	No	Logistic Regression	Online buying adoption	Retail density	Efficiency
Cao et al. (2013)	No	Structural Equation Model	Online buying frequency	Retail density Population density	Innovation Diffusion in Urban Efficiency in Exurban
Clarke et al. (2015)	Yes	Logistic Regression	Online buying frequency	Population density	Efficiency
Zhen et al. (2018)	No	Multinomial Probit	Online or in-store buying	Travel time to retail Population density	Both
Kirby-Hawkins et al. (2019)	Yes	Quadrant Analysis	Online sales data from retail chain	Population density	Both
Beckers et al. (2018)	No	Logistic Regression	Online buying adoption	Population density	No effect
Hood et al. (2020)	Yes	Descriptive Analysis	Surveyed online grocery shoppers	Population density	Both
Wieland (2021b)	No	Hurdle Model	Consumer channel choice	Competitive saturation Travel time City size	Innovation Diffusion
Ganning & Green (2021)	No	Linear Regression	E-retail spending by county	Retail accessibility Distance to MSA centroid	Efficiency
Shao et al. (2022)	No	General Spatial Model (SAC/SARMA)	Online shopping activity index	Retail density	Innovation Diffusion

As evidenced by Table 3.2, there is a clear lack of consistency among empirical studies on which hypothesis of spatial dynamics driving e-commerce has greater explanatory validity. These spatial dynamics are examined through a variety of contexts using a variety of spatial variables tied to accessibility and/or urbanity. Several studies offer empirical support for the innovation-diffusion hypothesis. Farag et al. (2007) study shops within a ten-minute cycling time as a measure of accessibility, and street address density as a measure of urbanity in their study of online purchase frequency. Their structural equation model points to urban consumers as more likely to shop online frequently than rural consumers. Farag et al. (2006a) also find that urban consumers are more likely to shop online. A similar study of German consumers identifies urban consumers as primary e-commerce consumers, though the prevalence of urban consumption may be driven in part by unequal distribution of web connectivity in rural versus urban Germany (Wieland, 2021b). The innovation-diffusion hypothesis is also supported by Shao et al. (2022), who incorporate retail density into a spatial regression to capture accessibility in driving e-commerce activity. Their findings assert that urban consumers are more likely to adopt e-commerce, in part because smaller cities in China do not have the necessary infrastructures to support e-retailing. All of these studies would seem to suggest that rurality is not a significant geographic driver of e-commerce use. Some intuition for this pattern is provided by de Blasio (2008):

> . . . the propensity to shop on the Internet is a matter not only of cost and convenience but also of culture and infrastructure. First, knowledge about the possibilities of the Internet may be lacking. Insofar as imitation is a crucial factor for online shopping, rural residents might find it more difficult to observe and try e-commerce. Second, there might be inefficiencies in the payment phase or in the parcel delivery service. These obstacles can jeopardize the prospect of e-commerce

> in remote areas as high-quality support services may be more readily available in urban settings. (p. 352).

In contrast, other studies listed in Table 2 favor the efficiency hypothesis over the innovation-diffusion hypothesis. Ren and Kwan (2009) find that shopping accessibility has a negative effect on rates of online shopping in Columbus, Ohio. In the United Kingdom, online consumers tend to have less accessibility to retail offerings than consumers who shop in-store, even in urban environments (Clarke et al., 2015). The efficiency hypothesis is also supported by Ganning and Green (2021), though they acknowledge a critical assumption that logistical costs do not scale with rurality. This, of course, is rarely the case. More remote areas will tend to have higher shipping costs for online orders. They qualify:

> The internet has not erased geography. Geography means low-income, remote consumers still must travel to population centers for necessities, and doing so incentivizes completing all purchases during those trips rather than paying any additional monies for shipping. (p.1267).

Clearly, there is great variability in how empirical studies have affirmed Anderson et al.'s (2003) competing hypotheses. The conclusions drawn by studies of e-commerce adoption demonstrate that Anderson et al.'s hypotheses are, at the very least, not mutually exclusive, and almost certainly subject to spatial and contextual circumstance. No substantial body of research has definitively favored one hypothesis over the other. Indeed, multiple studies focused on e-commerce in the grocery industry specifically produce findings that would seem to support both hypotheses (Kirby-Hawkins et al., 2019; Hood et al., 2020). While this absence of consensus makes formulating hypotheses challenging, it does punctuate the need for further research. Future empirical studies of the geographies of e-commerce must continue to situate their stories in geographic contexts – not only to control for spatial dependence, as previously discussed – but to deeper investigate Anderson et al.'s competing (and potentially converging) theories of e-commerce adoption and consumption. Because this paper shifts its empirical focus from the decisions of consumers to the decisions of firms, the innovation-diffusion hypothesis is selected over the efficiency hypothesis as the theoretical foundation for two additional hypotheses. This is partially informed by the assumption that many grocery chains may be more willing to invest in e-commerce systems and infrastructure in more urban trade areas where stores may tend to enjoy higher sales volumes. In the first hypothesis, *grocery locations offering some form of e-commerce to customers are more likely to be located in trade areas with higher population densities, controlling for other factors.* In the second hypothesis, *grocery locations offering some form of e-commerce to customers are more likely to be located in trade areas with higher degrees of competitive saturation, controlling for other factors.*

Methodology

To test the four hypotheses provided in the previous section, data from a variety of sources was synthesized and analyzed to uncover geographic and demographic patterns in the e-commerce location strategies of various chains. The following section provides an overview of this data, and outlines methods used to analyze the data. The dataset for this study is a June 2022 cross-sectional sample of 14,278 grocery store locations from a variety of major national and regional grocery chains across the contiguous United States. For each store location, a drive time polygon was constructed to approximate a primary trade area. Demographic and geographic data was then pulled for each drive time polygon to reflect the composition of the trade area associated with a given store. An exploratory spatial data analysis shows e-commerce offerings for specific regions of the United States based on the current sample of stores. The Southeast BEA region was chosen as regional case study based on the diversity of e-commerce offerings within the region and the competitiveness of the grocery market. A Join-Count test (Cliff & Ord, 1981) was also employed to measure spatial dependence in the incidence of e-commerce offerings across the contiguous United States. A spatial probit model was introduced that regresses a grocery chain's decision to offer e-commerce at a specific location or not against demographic and geographic characteristics of the stores respective trade area. Finally, this modeling sequence was re-run for the Southeastern United States to identify unique patterns and nuances in the geographies of e-commerce for the region.

Data Overview

Location data for a sample of 14,278 grocery stores across the contiguous United States is provided by *ChainXY,* a store location data vendor used by many analytics and real estate practices in the retail industry. This cross-sectional sample from June 2022 includes the address for each store, geographic coordinates, chain name, and a field indicating whether the observed store offers grocery pickup, delivery, both, or none. Grocery chains included in the sample include Walmart, Target, Sam's Club, Trader Joe's, Aldi, Ingles Markets, Publix Supermarkets, Albertsons, Jewel-Osco, Giant Food, Safeway, Vons, BJ’s Wholesale Club, Whole Foods Markets, Acme Markets, and Shaw’s. The entire store network within the contiguous United States was included for each chain as of June 2022. Figure 1 shows a map of all store locations included in the sample symbolized by their respective e-commerce offerings.

Additional data for each store was appended through a variety of sources. First, trade areas were drawn for each store location using a drive time engine provided by *SiteWise Analytics*. Because the sample of grocery locations includes a variety of store formats, each with varying degrees of gravity, it was important to stratify the isochrones approximating each store trade area based on store type. Wholesale clubs and supercenters, including Walmart, Sam’s Club, and BJ’s Wholesale Club all received 20-minute drive time trade areas, reflecting consumers’ general willingness to travel further distances to frequent these stores with larger footprints and a variety of offerings in foodstuffs and general merchandise. Supermarkets, including Aldi, Publix, Safeway, etc. were assigned 10-minute drive time trade areas, reflecting their smaller footprints typically designed to cater to localized trade areas or neighborhoods. Specialty grocers such as Whole Foods Markets and Trader Joe’s were assigned 15-minute drive time trade areas, reflecting their slightly lifted gravity from offering a variety of organic, high-quality, locally and ethically sourced products. As an outlier with a similar product assortment to supercenters and wholesale clubs, but with a smaller footprint, Target stores were also assigned 15-minute isochrones. A suite of demographic and geographic data was pulled for each trade area drawn around each store. Demographic data is

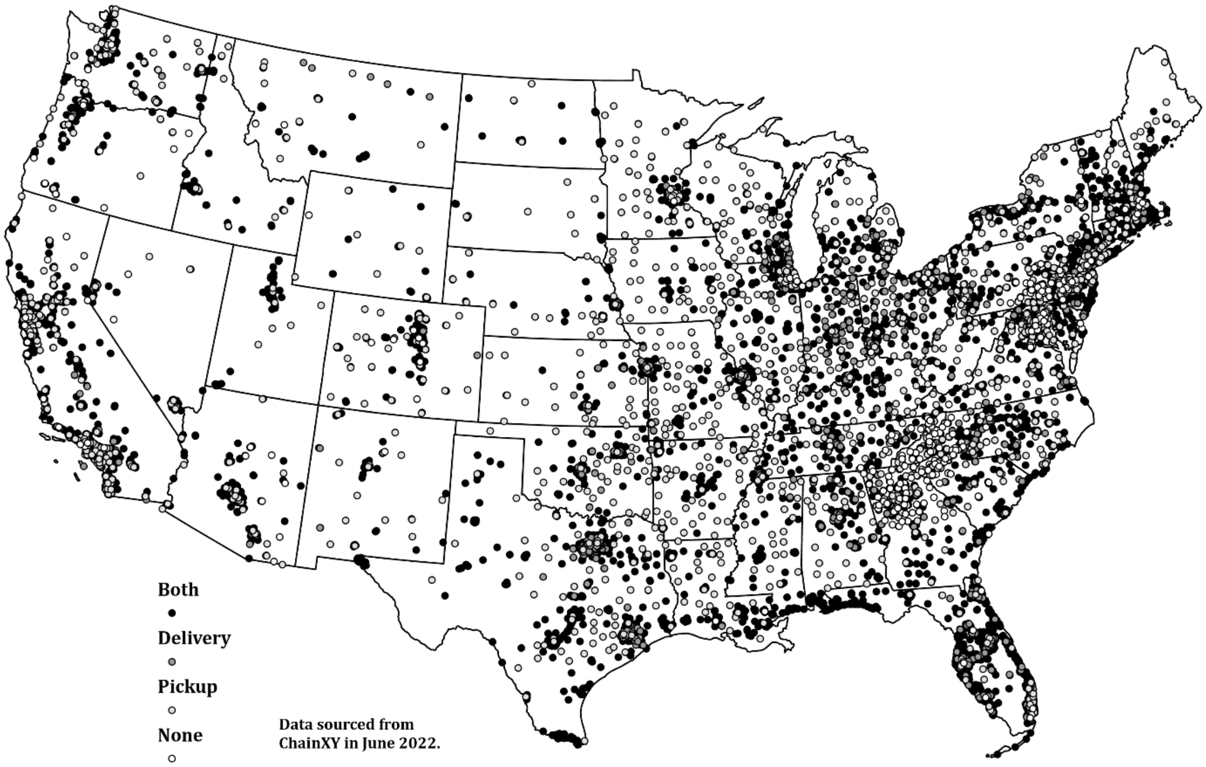

Figure 3.1. Study Area and Store Locations

sourced from *Synergos Technologies*. These trade area characteristics were appended to each store location. Finally, a region was assigned to each store location based on regions delineated by the Bureau of Economic Analysis (BEA, 2023). These regions are provided in Figure 2.

Exploratory Spatial Data Analysis

The exploratory data analysis includes an analysis of e-commerce offerings by BEA region, and a geostatistical analysis of spatial autocorrelation in the incidence of e-commerce grocery offerings across the contiguous United States. A Join-Count test (Cliff & Ord, 1981) was used to evaluate spatial dependence or autocorrelation among grocery locations that do or do not offer e-commerce. The Join-count test is specifically engineered to identify spatial autocorrelation in binary categorical variables. The spatial structure underlying the Join-Count is built on a hybridized spatial weight matrix that uses the *K*-nearest neighbors to each store weighted by their inverse distance squared. The value of *K* (the number of neighboring grocery stores) is determined using a data-driven approach by minimizing the Akaike's Information Criterion diagnostic of the spatial probit model in the following subsection. For this analysis, the categorical field indicating the e-commerce offering for each grocery store location was collapsed into a binary variable. Stores that offer some form of e-commerce (pickup, delivery, both) were assigned a value of 1. Stores that have no e-commerce offering were assigned a 0. This allows for three types of joins to be measured: 1:1, 0:0, and 1:0, where 1:1 and 0:0 denote neighboring grocery stores that have the same e-commerce offering (or lack thereof), while 1:0 denotes neighbors with different offerings. The asymptotically normally

Bureau of Economic Analysis (2023)

BEA Region	States
Far West	California, Nevada, Oregon, Washington
Great Lakes	Illinois, Indiana, Michigan, Ohio, Wisconsin
Mideast	Delaware, Maryland, New Jersey, New York, Pennsylvania
New England	Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island, Vermont
Plains	Iowa, Kansas, Minnesota, Missouri, Nebraska, North Dakota, Oklahoma, South Dakota
Rocky Mountains	Colorado, Idaho, Montana, Utah, Wyoming
Southeast	Alabama, Arkansas, Florida, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, Tennessee, South Carolina, Virginia, West Virginia,
Southwest	Arizona, New Mexico, Oklahoma, Texas

Figure 3.2. BEA Economic Regions

distributed join-count statistics (JC_{11}, JC_{00}, and JC_{10}) capture the number of joins for each type. The Join-Count test has a null hypothesis of no spatial autocorrelation for the evaluated variable. A negative z-value associated with a Join-Count statistic indicates that the number of neighbors of different formats is higher than expected, while a positive z-value indicates that there is a high degree of clustering among stores that do or do not offer e-commerce.

Spatial Econometric Modeling

To examine the e-commerce strategies of grocery locations with respect to demographics and spatial dynamics, a spatial autoregressive probit model with autoregressive disturbances (SARAR) that captures spatial autocorrelation in the incidence of e-commerce offerings across markets was implemented:

$$\boldsymbol{y} = \rho \boldsymbol{W}\boldsymbol{y} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\mu}$$

$$\boldsymbol{\mu} = \lambda \boldsymbol{M}\boldsymbol{\mu} + \boldsymbol{\varepsilon}$$

$$y_i = \begin{cases} 1 \; if \; grocery \; offers \; pickup, delivery, or \; both \\ 0 \; if \; no \; offering \end{cases}$$

where y_i denotes a binary variable indicating whether an observed store location offers e-commerce or not, W and M denote sparse spatial weight matrices, ρ is the spatial lag term, and λ is the spatial error term. X denotes a matrix of explanatory variables for each observation with a commensurate coefficient β for each explanatory variable, and ε denotes the error term. The spatial lag term (Wy) is designed to model the interdependence of store location decisions across space, while the spatial error term ($M\mu$) controls for spatial dependence among model residuals. A positive value for W signals a spread or spillover effect, in which a neighboring store offering e-commerce increases the probability that an observed store will offer e-commerce. A negative value for *W* signals a backwash effect, in which a neighboring store offering e-commerce would decrease the probability that an observed store also offers e-commerce (Chi & Zhu, 2020). A similar logic is applied to *M* which captures the spatial dependence among model errors.

Many estimation techniques have been proposed for spatial probit models (Pinske and Slade, 1998; Klier & McMillen, 2008; McMillen, 1992; LeSage, 2000). In this analysis, the approximate likelihood estimation (ALE) technique from Martinetti and Geniaux (2017) was used. Their R package, *ProbitSpatial*, allows for three different types of spatial probit models to be run: spatial autoregressive probit (SAR), spatial error probit (SEM), and the spatial autoregressive probit model with autoregressive disturbances (SARAR). The SAR and SEM models are effectively subset models of the SARAR, and are implicitly nested within its specification. Under the SAR specification, the spatial error term λ is not statistically significant, signaling no risk of spatial autocorrelation in the model errors. Conversely, under the SEM specification, the spatial lag term ρ is not statistically significant, indicating there is no spatial dependence in the binary dependent variable. The optimal model to fit the data was determined using a data-drive approach by minimizing the AIC after running all three models for varying neighborhood structures from two neighbors up to ten.

Novkaniza et al. (2019) conducted a simulation exercise that demonstrates that ALE yields tighter confidence intervals for parameter estimates, and less biased results than other estimation techniques, including Maximum Likelihood Estimation (McMillen, 1992), Bayesian (LeSage, 2000), and Linearized Generalized Method of Moments (Klier & McMillen, 2008). The coefficients of spatial probit models are challenging to interpret, and were converted to marginal effects that include both direct effects from changes in a store's explanatory variables and indirect effects to capture the spillover of an observed store's explanatory variable impacting its neighbor. Summing direct effects and indirect effects yields a total effect. These direct, indirect and total marginal effects were calculated using the R package, *ProbitSpatial*, and are reported for all spatial probit models.

Dependent variables were comprised of demographic characteristics of consumers within each grocery store's trade area, as well as variables designed to capture spatial dynamics and test the innovation-diffusion hypothesis. These variables are catalogued in Table 3.3. Population density is measured as the population per square mile in a trade area polygon. This variable is designed to capture urbanity, and by extension, test the innovation-diffusion hypothesis. Competitive saturation is also included in the model. This is a commonly used variable in the grocery industry that involves dividing the number or total square footage of competing stores in a trade area by the population as an overall measure of the competitiveness of the market. In this study, competitive saturation is the number of competing grocers per 1,000 population. It should be noted that competing grocers includes all types of grocery locations provided by ChainXY, not just store locations that are included in the analysis sample. The introduction of competition marks an important difference between an analysis of e-commerce geographies from the firm's perspective rather than the consumer's choice. Demographic

Table 3.3. Explanatory Variables

	Variable	Description
Spatial Dynamics	Population Density	Trade area population per square mile
	Competitive Saturation	Number of competing grocery stores within trade area per 1,000 people
Demographics	Median Household Income (Second Quartile)	Binary variable indicating trade area's median household income falls within quartile 2 of sample
	Median Household Income (Third Quartile)	Binary variable indicating trade area's median household income falls within quartile 3 of sample
	Median Household Income (Fourth Quartile)	Binary variable indicating trade area's median household income falls within quartile 4 of sample
	Percent Black	Percent of trade area population that is Black
	Median Age	Median age within trade area
	Average Household Size	Average size of households within trade area
	Percent Bachelor's +	Percent of trade area population with a bachelor's degree or higher

variables for each trade area include median household income, median age, percentage of the population with a bachelor's degree or higher, percent Black, and average household size. Median household income was discretized to a categorical variable to reduce multicollinearity with other variables in the model, including education, race, and household size. Categories for median household income are built based on the quartiles of trade area income values in the sample, with the first quartile being the baseline. Table 3.4 provides summary statistics for each explanatory variable.

Regional Case Study

To investigate key regional differences in geographies of e-commerce location strategies among grocery chains, the spatial probit model outlined in the previous subsection was re-applied to the Southeast (Florida, Georgia, Arkansas, Louisiana, Mississippi, Tennessee, Kentucky, West Virginia, Virginia, Alabama, and the Carolinas). The Southeast region was selected as a case study due to its interesting diversity of e-commerce offerings compared to other regions of the country, and the hyper-competitive grocery market that has developed in the region (Walters, 2018; Bennett, 2019; Llovio, 2021). Indeed, beyond the implications for the industry itself, the degree of competition among retail banners and districts in the Southeast has been established as a recurrent theme of geographic scholarship specific to the region (Buckwalter, 1989; Posey, 1994; Lord, 2000).

The results of modeling this region with a re-optimized spatial weight matrix and probit model specification uncovered patterns in the geography of e-commerce strategies across grocery locations that differ from trends yielded from an analysis at the national

Table 3.4. Summary Statistics for Explanatory Variables

Independent Variable	Minimum	Maximum	Median	Mean	Standard Deviation
Population Density	*9.3*	*55,356.7*	*1,464.4*	*2,418.3*	*3,738.5*
% Population Growth (2020-2025)	*-13.5%*	*48.2%*	*3.8%*	*4.8%*	*6.2%*
% Population Growth (2010-2020)	*-43.6%*	*415.1%*	*7.1%*	*9.2%*	*11.8%*
Median Age	*21.4*	*75.6*	*38.9*	*39.3*	*4.5*
Average Household Size	*1.6*	*4.2*	*2.5*	*2.5*	*0.3*
Unemployment Rate	*0.2*	*24.4*	*3.4*	*3.6*	*1.2*
Median Household Income	*$26,323*	*$178,269*	*$60,176*	*$65,172*	*$20,255*
Average Number of Vehicles	*0.3*	*2.8*	*1.8*	*1.8*	*0.2*
% of Population w/ Bachelor's Degree or Higher	*4.8%*	*87.6%*	*31.4%*	*33.6%*	*13.5%*
% Black Population	*0.0%*	*94.7%*	*7.9%*	*12.6%*	*13.5%*
% Hispanic Population	*0.4%*	*96.9%*	*10.1%*	*15.9%*	*16.0%*
% Male Population	*42.4%*	*66.3%*	*48.9%*	*49.0%*	*1.2%*
Competitive Saturation	*0.00*	*0.89*	*0.13*	*0.14*	*0.06*

level. The results of both models were then benchmarked against prevailing trends in the literature to identify the extent to which existing e-commerce location strategies among American grocery chains correspond to empirically defended characteristics of e-commerce adopters and consumers.

Results

The results of the analysis are provided in the following section. A breakdown of e-commerce offerings by BEA region is provided, indicating which region of the United States may serve as a compelling case study for further investigation. The results of the Join-Count test are also presented for a variety of spatial weight matrix structures. Finally, spatial probit model outputs are presented for the contiguous United States and the Southeast region of the country.

Exploratory Spatial Data Analysis

A regional breakdown of e-commerce offerings among major grocery chains in the contiguous United States is provided in Table 3.5. The Both column provides the percent of stores in the region sample offering both grocery pickup and delivery. The Delivery column shows the percent of stores in the regional sample that offer only delivery. The Pickup column shows the percent of stores in the regional sample that offer only pickup. The None column shows the percent of stores in the region that have no e-commerce offering for customers. A total count of all stores in the regional sample is also provided. For the entire contiguous U.S. sample, 83.8% offer both pickup and deliver, 1.7% offer only delivery, 12.7% offer only pickup, while 1.4% have no e-commerce offering. Each BEA region was found to have its own unique mix of e-commerce

Table 3.5. Grocery E-commerce Offerings by BEA Region

BEA Region	Both	Delivery	Pickup	None	Total Stores
New England	82.1%	4.3%	11.9%	1.7%	582
Mideast	69.4%	8.7%	18.1%	3.9%	1,828
Southeast	80.9%	6.7%	9.1%	3.3%	4,715
Great Lakes	76.6%	13.7%	8.7%	1.0%	1,888
Plains	74.0%	5.3%	20.2%	0.5%	927
Southwest	83.5%	6.2%	9.8%	0.5%	1,569
Rocky Mountain	83.8%	1.7%	12.7%	1.8%	543
Far West	79.5%	4.6%	14.5%	1.4%	2,226
Contiguous U.S.	**78.6%**	**7.1%**	**12.1%**	**2.2%**	**14,278**

offerings among grocery locations. For all regions, the majority of major grocery chain locations offer both pickup and delivery. The Mideast and Southeast regions both show markedly diverse offerings of e-commerce across all stores in the sample, with over 3% offering no e-commerce in both regions. It is this diversity, and the hyper-competitive landscape of grocery stores in the Southeast that drove the selection of this region as case study for this analysis (Walters, 2018; Bennett, 2019; Llovio, 2021).

An additional piece of the exploratory spatial data analysis involved gauging spatial autocorrelation in the incidence of e-commerce offerings (or lack thereof) across the contiguous United States. The categorical variable indicating e-commerce offerings was collapsed from four categories to two. Stores with a value of 1 offer some form of e-commerce to customers in their trade area, while stores assigned a 0 do not have any e-commerce offering. Table 3.6 demonstrates significant spatial autocorrelation in these offerings across the country. Join-count tests were run using a distance-weighted, row standardized weight matrix, testing for a variety of neighborhood structures. The latitude and longitude coordinates of each store location were used to build the matrix. For all evaluated neighborhood structures, the number of observed pairs of grocery stores with e-commerce (JC_{11}) was significantly higher than the expected value ($JC`_{11}$). Additionally, the number of neighboring grocery stores with no e-commerce offerings ($JC_{00}$) was also significantly higher than the expected value ($JC`_{00}$). Finally, the number of neighboring grocery stores with different offerings (JC_{10}) was also significantly higher than expected ($JC`_{10}$). These results held true across all neighborhood structures tested, with only

Table 3.6. Join-Count Statistics

Weight Matrix	Join-Count 1:1 test (E-commerce : E-commerce)		
	JC_{11}	JC'_{11}[a]	*z*-value
$W_{k=2}$	6862.9	6832.4	5.48***
$W_{k=3}$	6864.9	6832.4	6.27***
$W_{k=4}$	6864.5	6832.4	6.39***
$W_{k=5}$	6865.3	6832.4	6.66***
$W_{k=6}$	6866.0	6832.4	6.89***
	Join-Count 0:0 test (No offering : No offering)		
	JC_{00}	JC'_{00}	*z*-value
$W_{k=2}$	12.8	3.35	6.75***
$W_{k=3}$	13.4	3.35	7.58***
$W_{k=4}$	13.4	3.35	7.89***
$W_{k=5}$	13.3	3.35	7.93***
$W_{k=6}$	13.2	3.35	8.00***
	Join-Count 1:0 test (E-commerce : No offering)		
	JC_{10}	JC'_{10}	*z*-value
$W_{k=2}$	263.3	303.3	-6.70***
$W_{k=3}$	260.7	303.3	-7.64***
$W_{k=4}$	261.1	303.3	-7.83***
$W_{k=5}$	260.4	303.3	-8.09***
$W_{k=6}$	259.8	303.3	-8.32***

Note: *** significant at 1%. [a] JC'_{11} is the expected value of JC_{11}; $E[JC_{00}] = JC'_{00}$; $E[JC_{10}] = JC'_{10}$

marginal shifts in the Join-Count statistics as the number of neighbors in the weight matrix increased.

Spatial Econometric Modeling

The optimal spatial weights matrix with which to build a spatial probit model was determined by running both SAR, SEM, and SARAR specifications of the model for varying neighborhood structures in an effort to minimize the AIC diagnostic. Figure 3.3 shows the results of this exercise. Iteratively calculating AIC for a series of models of different spatial specifications and weight matrices determined that a SARAR model with an adaptive distance matrix of six neighbors per store is the optimal spatial weight matrix for the contiguous United States. This signals that there is evidence of spatial autocorrelation not only in the dependent variables, as confirmed by exploratory data analysis, but also in the model residuals. The coefficients and marginal effects for this model are provided in Table 3.7.

The first two columns provide the model coefficients and p-values for a standard aspatial probit model using maximum likelihood estimation. The third and fourth columns show these same coefficients and p-values for the SARAR probit model estimated using approximate likelihood estimation. Estimates are also provided for the spatial lag and spatial error terms. The final three columns show the direct, indirect, and total marginal effects for each covariate.

Model coefficients are robust across both spatial and aspatial models in terms of both sign and magnitude. Both models show income, education, race, age, and household size as important factors in determining whether an observed grocery store offers e-

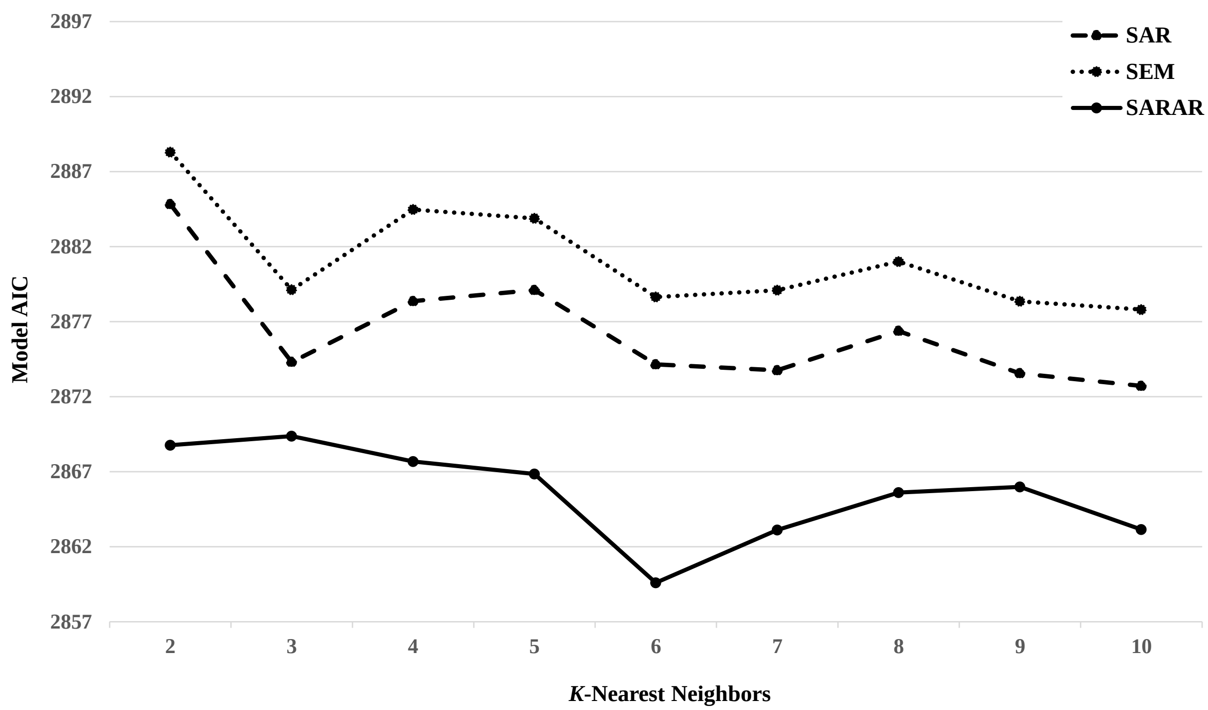

Figure 3.3. Determining Optimal Spatial Weights Matrix

Table 3.7. Spatial Probit Model Results

	Aspatial Probit		SARAR Probit		SARAR Effects		
	Coefficient (β)	*p*-value	Coefficient (β)	*p*-value	Direct	Indirect	Total
Intercept	0.106	0.941	0.136	0.859			
Population Density	0.000	0.766	0.000	0.763	0.0%	0.0%	0.0%
Median Household Income (Quartile 2)	0.312	0.000	0.315	0.000	1.4%	0.2%	1.6%
Median Household Income (Quartile 3)	0.248	0.006	0.249	0.009	1.1%	0.2%	1.3%
Median Household Income (Quartile 4)	-0.092	0.466	-0.709	0.322	-0.3%	0.0%	-0.4%
Percent Bachelor's +	0.808	0.022	0.680	0.034	3.0%	0.5%	3.4%
Percent Male	1.840	0.412	1.726	0.416	7.6%	1.2%	8.7%
Percent Black	-0.469	0.013	-0.441	0.005	-1.9%	-0.3%	-2.2%
Median Age	-0.013	0.046	-0.013	0.027	-0.1%	0.0%	-0.1%
Average Household Size	0.515	0.000	0.449	0.001	2.0%	0.3%	2.3%
Competitive Saturation	-0.287	0.451	-0.268	0.243	-1.2%	-0.2%	-1.4%
Spatial lag (ρ)			0.139	0.000			
Spatial error (λ)			0.111	0.000			
Akaike's Information Criterion	2903.6		2859.6				

commerce or not. Statistical significance is determined by interpreting the reported p-value as the probability that the coefficient β is equal to the estimate given a null hypothesis of $\beta = 0$. Both the coefficient for the spatial lag term and spatial error term are positive and statistically significant, indicating the presence of spatial dependence in both the binary dependent variable and the model errors. The SARAR probit model also shows an overall improved model fit relative to the aspatial probit model.

The marginal effects from the SARAR probit model also provide a more intuitive interpretation of the relationships between each independent variable and the binary dependent variable. The direct effect shows how unit changes in each explanatory variable will change the probability of the dependent variable for each observation. The indirect captures the spillover effect of changes in the explanatory variables of an observed trade area impacting the dependent variable in neighboring trade areas. The total effect is the sum of both direct and indirect effects.

Regional Case Study

The modeling sequence used for a model of e-commerce location strategy among grocers at the national level was re-run for the Southeast region specifically. A new spatial weights matrix specific to the region was optimized by minimizing the AIC of the final model. The optimal model for this region was also a SARAR specification, consistent with the national model. Marginal effects for both the regional and national model are stacked side by side in Table 3.8.

Table 3.8. Marginal Effects of Significant Features – Southeastern BEA Region vs. Contiguous United States

	Southeast (SARAR)			U.S. (SARAR)		
	Direct	Indirect	Total	Direct	Indirect	Total
Population Density	0.0%	0.0%	0.0%	Not Significant		
Median Household Income (Quartile 2)	2.1%	0.5%	2.6%	1.4%	0.2%	1.6%
Median Household Income (Quartile 3)	2.2%	0.5%	2.7%	1.1%	0.2%	1.3%
Percent Bachelor's +	Not Significant			3.0%	0.5%	3.4%
Percent Black	3.6%	0.8%	4.4%	-1.9%	-0.3%	-2.2%
Median Age	Not Significant			-0.1%	0.0%	-0.1%
Average Household Size	Not Significant			2.0%	0.3%	2.3%

Both spatial probit models for the Southeastern region and contiguous United States exhibit key differences not only in which demographic features and spatial variables are significant, but also in the signs of those variables.

Discussion

The following discussion section explores key findings and implications from the above results, and situates those findings against salient themes from existing literature. The discussion is framed geographically, beginning with a discussion of the U.S. grocery landscape overall using results from the Join-Count tests and SARAR probit model. A second subsection zooms in on the Southeastern region of the country specifically, contrasting the drivers of e-commerce locations strategies among grocery chains between the region and the nation.

Geography of E-commerce across the U.S. Grocery Landscape

Across a variety of spatial weight matrix structures, there is evidence of statistically significant clustering among grocery stores that offer e-commerce, and those that do not. For both categories, the Join-Count statistic reflecting the number of like pairs of neighbors exceeds the expected value under a random spatial distribution. Crucially, the Join-Count statistic for differing pairs (ie. a store offering e-commerce neighboring a store with no offering) is also statistically significant across a variety of neighborhood structures, and exceeds the expected value under a random distribution across space. The significance of this final Join-Count statistic signals that there are no substantial geographic divisions between grocery stores offering e-commerce and those that do not. This finding is particularly meaningful in how it related to urbanity and rurality. If there were clear spatial divides between e-commerce offerings and no offerings based on whether a store is located in a more urbanized or rural area, we may expect the Join-Count statistic to be insignificant, particularly as the number of neighbors in the weight matrix decreases. In aggregate, this geostatistical analysis demonstrates that there is significant localized clustering among grocery stores offering e-commerce and stores with no offering. However, there is no evidence that these clusters are disparate across space. This finding is echoed by the SARAR probit model.

The results of the SARAR probit model are salient both in what independent variables are significant and which are not significant. SARAR probit model coefficients and p-values provided in Table 3.7 show that population density is not a statistically significant feature in determining whether an observed grocery store location offers e-commerce or not. This finding does not effectively support either the innovation-diffusion hypothesis or the efficiency hypothesis. This finding rejects the initial hypothesis that grocery chains are more likely to offer e-commerce in urban trade areas over rural trade areas. Based on this model result, grocery chains are apt to offer e-commerce in both urban and rural regions. This may suggest that both the efficiency hypothesis and innovation-diffusion hypothesis are simultaneously true. Although the drive to adopt and the value proposition are different for both hypotheses, the end result is the same. This finding is consistent with other studies in the grocery industry that have seemingly found both hypotheses to be true in tandem (Kirby-Hawkins et al., 2019; Hood et al., 2020). Ultimately, this finding rejects the hypothesis that e-commerce is more likely to be offered in areas with higher population densities.

Additionally, the insignificance of competitive saturation in both the aspatial and SARAR probit models also signals that the overall competitive environment is not a key driver in whether a grocery store location will choose to offer e-commerce. This finding rejects the hypothesis that e-commerce is more likely to be offered in areas with increased competitive saturation. This signals that grocery store locations may not consider implementing e-commerce as a means of differentiation. Rather, the significance of the spatial lag term in the model suggests that grocery locations may be taking cues from their immediate neighbors to determine whether e-commerce should be adopted.

Another variable that was not statistically significant in the model is percent male population. While many previous studies of e-commerce consumption have identified males as a key demographic for adoption (Farag et al., 2006a; Weltevreden & van Rietenbergen, 2007; van Deuren et al., 2015; Clarke et al., 2015; Beckers et al., 2018; Etumnu et al., 2019), other studies have also pointed to females as the key demographic (Ren & Kwan, 2009; Loo & Wang, 2018; Zhen et al., 2018; Saphores & Xu, 2020). The absence of statistical significance in this coefficient would seem to echo the absence of consensus among previous studies, indicating that perhaps gender is not necessarily indicative of a consumers' overall propensity to shop online.

The results of the SARAR probit model indicate that median household income, education, race, and household size are all significant trade area determinants in whether or not a given grocery store location will offer e-commerce. The SARAR probit model returns direct and indirect marginal effects that offer a nuanced and intuitive interpretation of how each of these variables tangibly impacts the probability that an observed store will offer e-commerce based on its trade area composition.

The direct effect for average household size is equal to 2.0%. This indicates that, as the average household size in a given trade area increases by one individual, the probability that the grocery store servicing that trade area offers some form of e-commerce (delivery, pickup, or both) increases by 2.0%. The indirect effect reveals that, as the average household size in a given trade area increases by one individual, the probability that neighboring grocery stores also offer some form of e-commerce increases by 0.3%. The positive effect of household size on consumers' propensity to spend online has been empirically defended previously (de Blasio, 2008; Clarke et al., 2015; van Droogenbroeck & van Hove, 2017). This finding aligns with this prevailing trend in the literature.

Undoubtedly, the most predominant trend in past literature points to young, affluent consumers as the key demographic to purchase through e-commerce channels. Many studies have pointed to age, income, and education as key drivers of e-commerce consumption (Krizek et al., 2005; Farag et al., 2006a; Soopramanien & Robertson, 2007; de Blasio, 2008; Punj, 2011; Clarke et al., 2015; van Droogenbroeck & van Hove, 2017; Zhen et al., 2018; Etumnu et al., 2019; Saphores & Xu, 2020; Pernot, 2021). Indeed, regardless of geography, product, or other contextual factors, this predominant trend would seem to be the most prevalent and powerful force in determining whether a consumer is liable to shop online for goods. The results of this study confirm the hypothesis that grocery store locations in younger, more affluent trade areas are more likely to offer some form of e-commerce channel, aligning with the vast majority of previous studies. The direct effect for the median age variable signals that as the median age of a trade area increases by one year, the probability that the grocery store servicing that trade area offers some form of e-commerce decreases by 0.1%. The indirect effect for this variable is low in magnitude, indicating a negligible spillover effect to neighboring stores. The interpretation for the marginal effects of trade area income are adjusted slightly, as this variable is categorical.

The direct effect for the first categorical median household income variable signals that as the median household income of a trade area increases from the first quartile ($26,323 - $50,686 annually) to the second quartile ($50,686 - $60,176 annually), the probability that the grocery store servicing that trade area offers some form of e-commerce increases by 1.4%. The indirect effect indicates that as the median household income of a trade area increases from the first quartile to the second quartile, the probability that neighboring grocery stores offer some form of e-commerce increases by 0.2%. The direct effect for the second categorical median household income variable signals that as the median household income of a trade area increases from the first quartile to the third quartile ($60,176 - $75,506 annually), the probability that the grocery store servicing that trade area offers some form of e-commerce increases by 1.1%. The indirect effect indicates that as the median household income of a trade area increases from the first quartile to the third quartile, the probability that neighboring grocery stores offer some form of e-commerce increases by 0.2%. Interestingly, the final categorical variable for median household income capturing the fourth quartile of trade area incomes is not significant in the aspatial probit model or the SARAR probit model. It is possible that some trade areas in this upper income quartile ($75,506 - $178,269 annually) are located in more remote areas with older populations that dilute the significance of the variable. It is also possible that these wealthier households have a higher probability of having a stay-at-home spouse, or they tend to consume more prepared food and/or restaurant meals rather than groceries.

The final variable associated with the hypothesis that trade areas with younger, more affluent populations are more likely to have stores that offer e-commerce is education. Education has a strong relative influence over the dependent variable. The direct effect for percent of the trade area population with a bachelor's degree (or higher) is equal to 3.0%. This indicates that, as the percentage of the population in the trade area with a degree increases by 1%, the probability that the grocery store servicing that trade area offers some form of e-commerce increases by 3.0%. The indirect effect reveals that, as this percentage of population with a degree in a given trade area increases by 1%, the probability that neighboring grocery stores also offer some form of e-commerce increases by 0.5%. Though the relative magnitude and significance of age, income, and education vary in the SARAR probit model, the results align with a wealth of previous literature arguing that young, affluent consumers are the key demographic supporting e-commerce consumption across geographies and product lines. Because e-commerce is still a relatively new force in the retail economy and society, an important caveat to this prevailing trend and past empirics is that, inevitably, the demographics of e-commerce consumers will shift and change over time. Younger generations that may currently have higher adoption rates of e-commerce and other technologies will age over time, resulting in a future in which, likely, all age groups and classes may rely on e-commerce equally for their shopping needs.

The final variable considered in the SARAR probit model is race. This study hypothesized that trade areas with higher Black populations are less likely to have access to e-commerce offerings, corresponding to the longstanding pattern of retail chains often overlooking Black populations in planning their store deployments and offerings (Rowlands et al., 2023; Shannon, 2020; Zhang & Ghosh, 2016; Eisenhauer, 2001). The SARAR probit model appears to confirm this hypothesis. The direct effect for the percent Black population variable is -1.9%. This indicates that, controlling for other factors, as the Black population in a trade area increases by 1%, the probability that the grocery store servicing that trade area offers some form of e-commerce decreases by 1.9%. The indirect effect reveals that, as the Black population in a trade area increases by 1%, the probability that neighboring grocery stores also offer some form of e-commerce decreases by 0.3%. These findings suggest that large grocery chains tend to overlook Black populations when selecting which stores in their chain should offer some form of e-commerce. When zooming into a regional analysis of the southeastern United States, the relationship between race and e-commerce offering changes.

The Southeast United States

Running an additional SARAR probit model for the southeastern region of the United States reveals interesting contrasts between how grocery chains are developing e-commerce strategies at a regional level as opposed to the broader national trends. Table 3.8 stacks the statistically significant marginal effects for the sample of grocery locations in the BEA Southeast region against the marginal effects in the previously discussed national model for the contiguous United States. There are many key points of differentiation between the two models. While population is statistically significant in the Southeast SARAR probit model, the magnitude of the direct and indirect effect is negligible. The marginal effects for median household incomes align closely, with income having an even stronger deterministic effect on e-commerce strategies among southeastern grocery stores. Education, age, and household size were all statistically significant and impactful drivers of e-commerce offerings in the national model. However, these independent variables are not significant in the Southeastern region. It is possible that prevailing demographic trends in the south may disrupt or dilute these relationships within the model. For example, Florida has many trade areas with older, retired populations that may consume using e-commerce in urban environments just as much as younger populations, rendering median age an uninformative variable in the model for the southeastern United States. Indeed, Florida has already seen increased use of ride-sharing apps through new adoption programs via public-private partnership (Golant, 2019).

Undoubtedly, the most striking deviation between the national and southeastern regional models is the role of race as a determinant of e-commerce deployment among grocery chains. In the southeastern United States, the percent Black population has a direct effect of 3.6%. This indicates that, controlling for other factors, as the Black population in a trade area increases by 1%, the probability that the grocery store servicing that trade area offers some form of e-commerce increases by 3.6%. The indirect effect reveals that, as the Black population in a trade area increases by 1%, the probability that neighboring grocery stores also offer some form of e-commerce increases by 0.8%. This finding stands in stark contrast to the national model, and rejects the original hypothesis that Black populations were more likely to be excluded or neglected. It is possible the result is driven by the distribution of the Black population. The Black population in the Northern United States is largely concentrated in urban areas, while the Black population in the South is spread across both urban and rural areas (Cox & Tamir, 2022). This finding diverges from Saphores and Xu (2020) and bodies of literature on retail redlining and equity in retail accessibility. It aligns more closely with Ganning and Green (2021) who also identify a positive relationship between percent Black population and online spending across U.S. counties. The positive relationship between race and e-commerce offering in the southeast SARAR probit model certainly indicates there are regional nuances to the impact of race on e-commerce accessibility, and demands further empirical investigation with an emphasis placed on regional heterogeneity.

Conclusion

This study has applied spatial data analysis and spatial econometric modeling to identify the determinants of e-commerce offerings among grocery store locations across the United States. The study offers an important contribution to existing bodies of literature by pivoting away from the conventional mode of study focusing on e-commerce consumption from the consumer’s perspective, opting instead to explore how retail chains are making decisions about where to deploy e-commerce offerings across their network. Emphasis was placed on demographic determinants, and spatial dynamics. A geostatistical analysis revealed significant clustering among stores that do and do not offer e-commerce, signaling the need to incorporate spatial dependence into a regression model. Income, education, age, household size, and race are all relevant forces in determining whether an observed store will offer e-commerce to consumers in its respective trade area. The results of the SARAR probit model for the contiguous United States reveal that grocery stores offering e-commerce tend to be located in trade areas with young, affluent consumers from larger households, with lower Black populations. Interestingly, a regional case study of the southeastern United States reveals very different patterns underlying the geography of e-commerce offerings among grocery stores. In the southeast, age and household size are no longer important deterministic forces. Further, higher Black populations no longer have a negative relationship with the incidence of e-commerce offerings. In point of fact, trade areas with larger Black populations are found to be more likely to offer grocery pickup, delivery, or both.

This crucial finding punctuates the need for additional empirical research into geographies of e-commerce deployment among grocers and other retail chains to build a more nuanced understanding of the forces driving decisions at the firm level. Opportunities remain to focus additional research on more granular spatial heterogeneity. While this study offers a single regional case study, there are still opportunities to investigating how business make crucial decisions about implementing e-commerce in other regions, or in specific markets. For instance, households in Utah are, on average, larger than other states, fueling larger baskets and greater overall demand for groceries per household (World Population Review, 2024). A state-level model could produce a significant effect for household size. Future research is also poised to apply a comparative lens to which products are most likely to be consumed through e-commerce channels, and how these products may change across geographies. The extent to which geographers and other scholars may be able to investigate these pressing questions is, in part, commensurate on the ability of scholars to form important partnerships with practitioners and data providers in the industry to build analyses and studies that go beyond what is possible with publicly available data.

Chapter Four

Determining E-Commerce Strategies Among Grocery Stores in the United States Using a Multi-class Supervised Learning Model

Introduction

By the end of 2023, e-commerce sales in the United States are projected to reach $1.14 trillion, a 10% increase over 2022. While year-over-year growth is expected to settle at 10% or less in the coming years, online sales are projected to maintain double the growth rate of in-store sales (FTI, 2023). The U.S. Census Bureau (2023b) released its quarterly retail e-commerce sales report, showing that e-commerce sales for the second quarter of 2023 were $277.6 billion, increasing 2.1% from the first quarter and increasing 7.5% over the second quarter of 2022. While e-commerce growth in the grocery sector has slowed in recent years, reflecting consumers' continued efforts to rein in spending in response to macro-economic forces, there remain some signs of continued growth over a longer time horizon. Notably, the total number of households in the U.S. that purchased groceries online has increased 5% over last year (Bishop, 2023). Grocery pickup, in particular, saw year-over-year growth of 9.1% in sales (Rajagopal, 2023). Overall, the online grocery market is projected to grow at a compound annual rate of 13.1% between 2022 and 2027, corresponding to an increase of $740.9 billion (PR Newswire, 2023). Leading the charge among U.S. grocery banners to embrace and capitalize on growth in e-commerce is *Walmart*, which accounted for 36% of all U.S. online grocery sales in the second quarter of 2023 (Moran, 2023b). Literature concerned with the demographic and spatial forces driving this unprecedented rise in e-commerce consumption has, historically, placed emphasis on consumer behavior and consumer traits. However, less emphasis has been placed on the decisions of the firm and how demographics and spatial dynamics may inform e-commerce location strategies. This paper makes a key contribution to previous literature by pivoting to the perspective of the firm, investigating the critical decisions made by major U.S. grocery chains regarding whether a given store in their chain should appropriately offer pickup, delivery, both, or no e-commerce offering based on the characteristics of its trade area.

There are a variety of lenses through which this seismic growth in online shopping may be examined. In recent years, geographers, economists, and management scholars have developed a body of literature primarily concerned with patterns underlying the demographics and spatial dynamics of e-commerce consumption, including in the grocery sector (Ganning & Green, 2021; Saphores & Xu, 2020; Beckers et al., 2018). Salient themes from this body of literature highlight young, affluent consumers as the most likely to purchase through online channels. Further, divergent theories of the spatial dynamics of e-commerce consumption have emerged: the innovation-diffusion hypothesis and the efficiency hypothesis. The efficiency hypothesis argues that rural consumers are more likely to adopt e-commerce due to their limited access to retail nodes, while the innovation-diffusion hypothesis contests that urban consumers are more likely to adopt e-commerce because they are, on average, more educated and more active on the Internet (Anderson et al., 2003). Many studies support either one or both of these hypotheses concurrently, suggesting the nuances underlying both theories are far from universal, and may vary depending on regional context, product type, or other consumer-specific traits.

This study leverages a unique dataset from the grocery industry to build upon previous geographic studies of e-commerce consumption. Previous studies have focuses largely on the demographic patterns and spatial dynamics that drive consumers to purchase online. These dynamics encompass geographic variables such as accessibility to retail or population density. This work makes an important contribution to the literature through shifting the focus of analysis to the strategies and decisions made by store banners rather than the behavior of consumers. This shift in focus has been established as an important avenue for future research in geographies of e-commerce (Kirby-Hawkins et al., 2019). The findings here provide a fuller picture of geographies of e-commerce for the U.S. grocery sector by applying the demographic and spatial dynamic elements of prior research to model the e-commerce strategies being adopted by grocery banners for individual store locations. For many national and regional grocery chains in the United States, these decisions are uniquely stratified across grocery store networks, unlike other general merchandise or apparel chains.

This study employs a 2022 cross-sectional sample of grocery location data in combination with a Random Forest model to prescribe an appropriate e-commerce strategy for a given market based on trade area characteristics and spatial context. Grocery stores may offer delivery, pickup, both delivery and pickup, or no e-commerce at all. The results of the Random Forest model will yield class probabilities that will determine the most appropriate e-commerce offering in a given trade area. From the model results, a discussion will be framed using two scenarios. First, an analysis of model error scenarios will reveal cases of potential missed opportunities and mis-investment among grocery banners whose current e-commerce strategy for select stores are not optimized for the surrounding market. Second, the model will be applied to sample hexagonal grid to derive a surface layer of e-commerce probabilities across the southeastern region of the United States. This surface can serve as a tool for designing competitive defense strategies.

Review of Literature

This study is informed by literature in geography, economics, and management scholarship. Certainly, the largest of these bodies encompasses the aforementioned analyses of demographic and spatial forces that drive consumer decisions to shop online or in-store. The vast majority of these studies use some sort of regression analysis to capture the deterministic effects of various demographic and geographic characteristics on a consumer's decision to shop online. These regression analyses are catalogued in Table 4.1 for demographic features. The *E-commerce Variable* field shows the dependent variable for each study, while the Significant Demographic Variables and Coefficient Sign column highlights significant variables, and the sign associated with those variables.

Seven of the studies inventoried in Table 4.1 focus specifically on the grocery sector. In De Blasio's (2008) study of household consumption in Italy, younger consumers with more education and income are found to be more likely to purchase groceries online. Similar studies produced the same universal finding in the United

Table 4.1. Literature Review - Demographics

Author(s)	Grocery Focus	Method	E-commerce Variable	Significant Demographic Variables and Coefficient Sign
Krizek et al. (2005)	No	Logistic Regression	Binary variable denoting whether a household shops online frequently or infrequently	Age [-] Income [+]
Krizek et al. (2005)	No	Logistic Regression	Binary variable denoting whether a household has ever shopped online	Education [+]
Farag et al. (2006a)	No	Logistic Regression	Binary variable indicating whether an individual has ever shopped online	Age [+/-] Gender (Female) [-] Income [+] Education [+]
Soopramanien & Robertson (2007)	No	Logistic Regression	Binary variable indicating whether an individual shops online	Age Income [+]
Weltevreden & van Rietenbergen (2007)	No	Multinomial Logit Model	Multinomial variable distinguishing non-eshoppers, online searchers, and e-shoppers	Age [-] Gender (Male) [+]
de Blasio (2008)	Yes	Logistic Regression	Probability that a household has purchased goods online	Age [-] Education [+] Income [+] Household size [+]
Ren & Kwan (2009)	No	Logistic Regression	Binary variable indicating whether an individual shops online	Gender (Female) [+] Race (White) [+]
Punj (2011)	No	Logistic Regression	Binary variable indicating whether an individual shops online	Age [-] Education [+] Income [+]
Clarke et al. (2015)	Yes	Logistic Regression	Binary variable indicating whether an individual uses the internet often to purchase groceries	Age [-] Gender (Male) [+] Household size [+] Income [+]
Van Deursen et al. (2015)	No	Linear Regression	Ordinal measure on a five-point scale ('never' to 'daily') of how often a respondent uses the internet for commercial transactions	Gender (Male) [+] Age [-] Income [+]
Van Droogenbroeck & Van Hove (2017)	Yes	Logistic Regression	Binary variable indicating whether a household chooses to adopt a "Collect & Go" program offered by a grocery chain	Age [+/-] Education [+] Household size [+]
Beckers et al. (2018)	No	Logistic Regression	Binary variable indicating whether an individual shops online	Education [+] Gender (Male) [+] Income [+]
Loo & Wang (2018)	No	Ordered Logit Regression	Multinomial variable indicating amount of time spent e-shopping daily (0-30 minutes, 30-60 minutes, > 1 hour)	Age [-] Gender (Female) [-]

Table 4.1. continued

Author(s)	Grocery Focus	Method	E-commerce Variable	Significant Demographic Variables and Coefficient Sign
Zhen et al. (2018)	No	Multinomial Probit Regression	Multinomial variable indicating whether a consumer pre-purchases books in-store, pre-purchases online, or shops in-store only	Age [+] Education [+]
Zhen et al. (2018)	No	Multinomial Probit Regression	Multinomial variable indicating whether a consumer pre-purchases clothing in-store, pre-purchases online, or shops in-store only	Gender (Female) [-] Income [+]
Etumnu et al. (2019)	Yes	Logistic Regression	Binary variable indicating whether an individual shops online	Age [-] Education [+] Gender (Male) [+]
Saphores & Xu (2020)	Yes	Logistic Regression	Binary variable indicating whether an individual shops online	Age [-] Education [+] Gender (Female) [+] Race (Black) [-] Income [+]
de la Llave Montiel et al. (2020)	Yes	Spatial Durbin Probit Model	Binary variable indicating online retail churn (whether a customer becomes inactive online)	Income [+]
Pernot (2021)	Yes	Logistic Regression	Binary variable indicating whether an individual shops online	Age [-] Income [+]
Ganning & Green (2021)	No	Linear Regression	Percent of consumer expenditure via online shopping by county	Age [+] Income [+] Race (Black) [+]
Shao et al. (2022)	No	General Spatial Model (SAC/SARMA)	Online shopping activity as measured by Alibaba's Online Shopping Index	Age [+] Education [-] Income [+]

Kingdom (Clarke et al., 2015), Belgium (Van Droogenbroeck & Van hove, 2017), and France (Pernot, 2021).

These regression analyses have also focused on spatial dynamics, attempting to identify whether Anderson et al.'s (2003) efficiency or innovation-diffusion hypotheses are applicable to how consumers purchase online across space. The results of these explorations have been far less aligned than investigations of demographic patterns, with some scholars asserting the efficiency hypothesis, others confirming the innovation-diffusion hypothesis, and some finding both true in tandem. This signals that more research is needed to understand the patterns underlying these dynamics, and a more robust understanding of the circumstances under which each hypothesis may be true depending on market, region, product, consumer behavior, etc. Table 4.2 shows the results of regression analyses that have examined spatial dynamics in relation to e-commerce consumption. The *E-commerce Variable* field shows the dependent variable. The *Geographic Variables* field reflects which variables were designed to measure each of Anderson et al.'s (2003) hypotheses. The *Hypothesis* column specifies whether any parts of the study's empirics support the efficiency hypothesis or the innovation diffusion hypothesis. If neither hypothesis is supported, there is no effect. Table 4.2 demonstrates the absence of consensus among scholars regarding the validity of each spatial dynamic as a driving force of online consumption.

This study will build off of this foundational literature in two key ways. Firstly, it will pivot focus to the decisions made by grocery banners about which e-commerce strategy should be adopted at a given store location, answering a previous call for more

Table 4.2. Literature Review – Spatial Dynamics

Author(s)	Grocery Focus	Method(s)	E-commerce Variable(s)	Geographic Variable(s)	Hypothesis
Krizek et al. (2005)	No	Logistic Regression	Online buying adoption Online buying frequency Online search adoption	Distance to CBD Retail accessibility	No effect
Farag et al. (2006a)	No	Logistic Regression and Linear Regression	Online buying adoption Online buying frequency	Dummy variables for retail density	Adoption supports Innovation Diffusion Frequency supports Efficiency
Farag et al. (2006b)	No	Logistic Regression	Online buying adoption Online buying frequency	U.S. – shops within walking distance Netherlands – travel time	Dutch adoption supports innovation diffusion
Weltevreden & van Rietbergen (2007)	No	Multinomial Logit	Online buying adoption	Retail density	No effect
de Blasio (2008)	Yes	Logistic Regression	Probability of online purchase	City size	No effect
Ren & Kwan (2009)	No	Logistic Regression	Online buying adoption	Retail density	Efficiency
Clarke et al. (2015)	Yes	Logistic Regression	Online buying frequency	Population density	Efficiency
Zhen et al. (2018)	No	Multinomial Probit	Online or in-store buying	Travel time to retail Population density	Both
Beckers et al. (2018)	No	Logistic Regression	Online buying adoption	Population density	No effect
Ganning & Green (2021)	No	Linear Regression	E-retail spending by county	Retail accessibility Distance to MSA centroid	Efficiency
Shao et al. (2022)	No	General Spatial Model (SAC/SARMA)	Online shopping activity index	Retail density	Innovation Diffusion

research with a focus on firm decisions and strategy (Kirby-Hawkins et al., 2019). Secondly, while regression models are ideal for intuitive readings of deterministic relationships, using a Random Forest can provide more predictive (or prescriptive) power for determining optimal e-commerce offering of a given grocery store trade area. The gains of machine learning algorithms in predictive power over traditional econometric models have been well-documented by geographers and regional scientists (Goulard et al., 2017; Kopczewska, 2021). The deployment of a Random Forest model in this study follows a recurrent trend in retail geography of applying machine learning methods to solve complex problems and analyze retail geographies in new ways (Comber et al., 2019; Aversa et al., 2020; Ballantyne et al., 2021; Rose & Dolega, 2022). The results of this modeling exercise will not only pull back the curtain on how national and regional grocery chains make critical decisions about e-commerce offerings across their store networks, they will also provide a new tool to allow retail geographers in the industry to pre-emptively understand the e-commerce strategies of competitors across space to inform competitive defense strategies.

Data and Methodology

In this study, a Random Forest model is used to classify a given grocery store location's optimal e-commerce offering: delivery, pickup, both, or none, based on the characteristics of its trade area. The analysis follows a sequence of identifying prescriptive features to include in the model based on extant literature, tuning hyperparameters using a training data set, and cross validating through the use of a test data set. After the modeling exercise, some misclassified stores are examined at the market level to understand why they were misclassified, yielding insights about how specific grocery banners may be building e-commerce strategies in certain markets or regions, resulting in missed opportunities or mis-investment. Finally, the model is applied to a sample hexagonal grid over the southeast United States to derive a surface area of optimal e-commerce offerings based on the demographic and competitive characteristics of each grid cell.

Data Overview

For this study *ChainXY*, a location data vendor for the retail industry, provided a sample of 14,278 grocery stores across the contiguous United. The data is a cross-sectional sample from June 2022 that contains the address for each store, geographic coordinates, chain name, and an e-commerce field denoting whether a grocery store offers grocery pickup, delivery, both, or none. Grocery chains in the sample include Walmart, Target, Sam's Club, Trader Joe's, ALDI, Ingles Markets, Publix Supermarkets, Albertsons, Jewel-Osco, Giant Food, Safeway, Vons, BJ's Wholesale Club, Whole Foods Markets, Acme Markets, and Shaw's. Figure 4.1 shows a map of all store locations included in the sample symbolized by their respective e-commerce offerings.

Additional data for each store was appended through a variety of sources. Store trade areas were built for each grocery store using a drive time engine from *SiteWise Analytics*. Because the sample of grocery locations includes a variety of store formats, each with varying degrees of gravity, each trade area is built using a different drive time polygon based on store type. These drive times are informed by industry conventions and the author's experience working in real estate and location research in the U.S. grocery

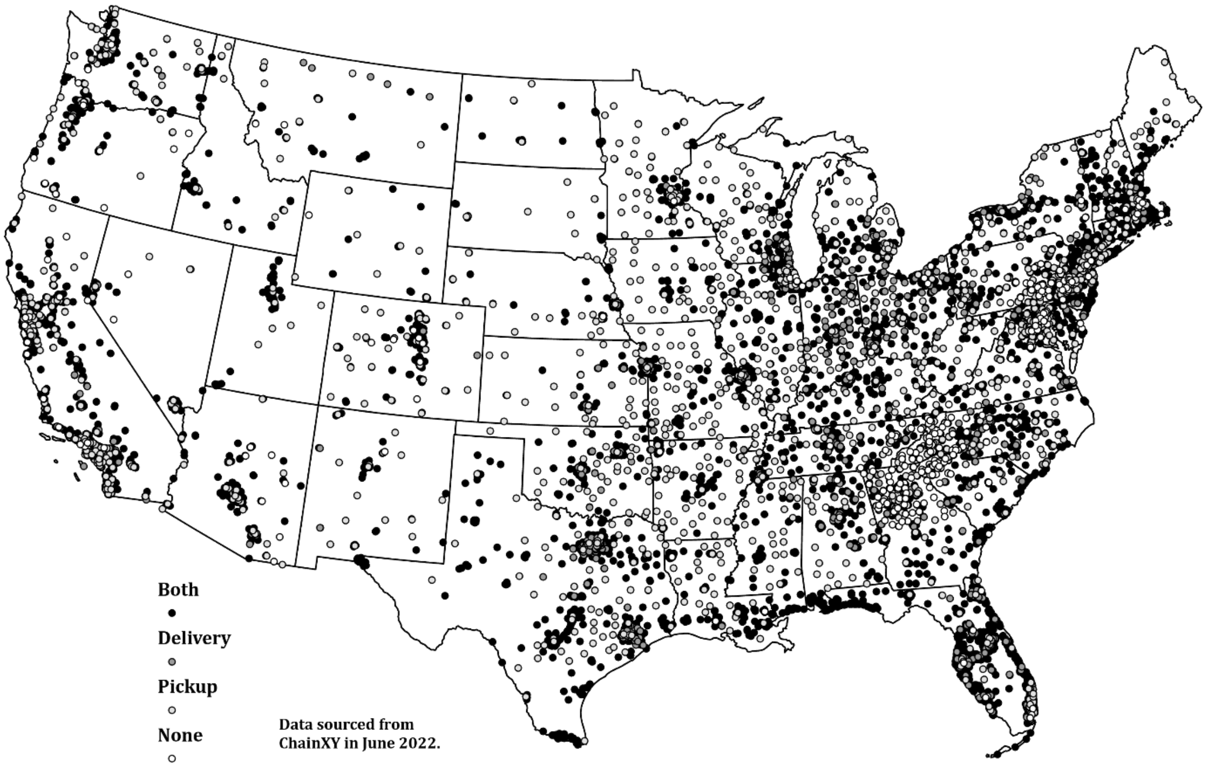

Figure 4.1. Study Area and Store Locations

industry. Walmart, Sam's Club, and BJ's Wholesale Club all received 20-minute drive time trade areas, reflecting consumers' general willingness to travel further distances for their lower prices, extensive selection, and buy-in-bulk offerings. Aldi, Publix, Safeway, etc. were assigned 10-minute drive time trade areas, as they tend to have smaller footprints designed to serve a more localized neighborhood trade area. Specialty grocers such as Whole Foods Markets and Trader Joe's were assigned 15-minute drive time trade areas. This slight lift in the drive time corresponds to their elevated offering of organic, high-quality, locally and ethically sourced products. As an outlier with a similar product assortment to supercenters and wholesale clubs, but with a smaller footprint, Target stores were also assigned a 15-minute drive time. Demographic data sourced from *Synergos Technologies* was pulled for each store trade area, and additional variables were calculated such as population density and competitive saturation.

Sampling and Feature Selection

An essential preliminary step in preparing a classification model is correcting for class imbalances in the training data. Table 4.3 provides a breakdown of how many store locations within the sample have adopted each e-commerce offering for their respective market.

To correct for this imbalance, Chawla et al.'s (2002) SMOTE algorithm was applied to the training set to balance the classes, allowing the Random Forest model to learn from more observations for each class to mine deterministic relationships between the target variable and each feature. The SMOTE algorithm combines both upsampling and downsampling to correct for class imablances, while also being augmented to handle

Table 4.3. Breakdown of E-commerce Offerings

Pickup	**Delivery**	**Both**	**None**	**Total Sample**
1,724	1,016	11,228	310	14,278

categorical data. Note that the algorithm is not applied to the testing data to preserve the original imbalance of classes. A suite of 19 features were selected for the model based on previous studies of consumer choice related to online shopping, including variables to capture the most salient trend in the literature: young affluent consumers being most likely to adopt e-commerce channels. Each feature is inventoried in Table 4.4.

Population density captures the population per square mile within each trade area. Two population growth features reflect the percent growth in population for the previous 10 years, and the projected growth from 2020 to 2025. The competitive saturation variable calculates the number of competing grocery stores per 100,000 population. Note that this statistic includes other grocery banners beyond those within the sample. The e-commerce demand variable is calculated using Synergos Technologies's *Market Outlook* data, and reflects the proportion of overall retail demand attributable to online shopping. The region variable is appended to each store based on economic regions defined by the Bureau of Economic Analysis (2023).

Random Forest Model

Breiman's (2001) Random Forest algorithm leverages multiple decision trees for classification or prediction. Bootstrap aggregation is used to build decision tree models from multiple samples. The algorithm then combines the resultant predictions and yields a final classification or prediction through voting. This method is uniquely equipped to handle nonlinear, multimodal, and categorical data. It is also designed to handle spatial autocorrelation by introducing random variation (Redo et al., 2012).

Table 4.4. Random Forest Model - Selected Features – Summary Statistics for Full Sample

Feature	Minimum	Maximum	Median	Mean	Standard Deviation
Population Density	*9.3*	*55,356.7*	*1,464.4*	*2,418.3*	*3,738.5*
% Population Growth (2020-2025)	*-13.5%*	*48.2%*	*3.8%*	*4.8%*	*6.2%*
% Population Growth (2010-2020)	*-43.6%*	*415.1%*	*7.1%*	*9.2%*	*11.8%*
Median Age	*21.4*	*75.6*	*38.9*	*39.3*	*4.5*
Average Household Size	*1.6*	*4.2*	*2.5*	*2.5*	*0.3*
Unemployment Rate	*0.2*	*24.4*	*3.4*	*3.6*	*1.2*
Median Household Income	*$26,323*	*$178,269*	*$60,176*	*$65,172*	*$20,255*
Average Number of Vehicles	*0.3*	*2.8*	*1.8*	*1.8*	*0.2*
% of Population w/ Bachelor's Degree or Higher	*4.8%*	*87.6%*	*31.4%*	*33.6%*	*13.5%*
% Black Population	*0.0%*	*94.7%*	*7.9%*	*12.6%*	*13.5%*
% Hispanic Population	*0.4%*	*96.9%*	*10.1%*	*15.9%*	*16.0%*
% Male Population	*42.4%*	*66.3%*	*48.9%*	*49.0%*	*1.2%*
Competitive Saturation	*0.00*	*0.89*	*0.13*	*0.14*	*0.06*
E-commerce % of Consumer Demand	*1.2%*	*28.0%*	*9.5%*	*10.5%*	*6.4%*
Longitude	*Numeric variable capturing store longitude coordinate*				
Latitude	*Numeric variable capturing store latitude coordinate*				
Region	*Categorical variable: BEA Economic Region*				
Type	*Categorical variable: Type of store – wholesale club/supercenter, specialty grocer, supermarket, or other*				
Banner	*Categorical variable: Name of chain*				

After applying stratified sampling and using the SMOTE algorithm to adjust for class imbalances, a Random Forest model was repeatedly cross-validated using 3 folds and 20 unique hyperparameter combinations. This *k*-folds cross-validation randomly subsets the training dataset into *k* groups with a similar number of observations in each group. The Random Forest is then fitted on *k* - 1 folds and benchmarked against the withheld sample to gauge performance. This process is repeated *k* times. The hyperparameters for the Random Forest that were tuned using cross-validation included the total number of tree models, the number of predictors randomly sampled at each split for each new tree model, and the minimum number of observations in a given node that are required for the node to continue splitting into net new tree models.

The Random Forest in this study aims to maximize *area under the curve* (AUC) as an overall metric of model performance. Unlike accuracy, this score is not concerned with the total number of correctly classified observations – in this case, stores. Rather, AUC is a metric derived from a receiver operating characteristic (ROC) curve. This curve plots false positive classifications along an x-axis with a true positive rate on a y-axis. In this plane, a perfectly diagonal line extending from the origin reflects a random guess at classifying each observation. The higher the ROC curve is in the upper left quadrant of the plane, the more predictive and powerful the classification model. AUC captures the area under this curve, and is thus, a more appropriate performance metric here than conventional classification accuracy. In this multi-classification model, the Hand-Till estimator is used to derive a generalized AUC score (Hand & Till, 2001). This method, ostensibly, iterates through AUC calculations for each class as though it is a binary classification. The method then accounts for pairwise comparisons between classes to generate a single metric that generalizes the conventional two-class AUC for a multi-class problem. The results of the modelling and cross-validation are provided in the following section. Emphasis is placed on key performance metrics, model performance across observations, and variable importance.

Results

The Random Forest model deployed in this study produced an optimal classification scheme on testing data with the following parameters: 997 trees, 15 sampled predictors, and 38 minimum observations required for additional splitting at each node. ROC curves derived for each class are provided in Figure 4.2. These curves show strong model performance on the test data, with a final model AUC score of 0.95. The confusion matrix provided in Figure 4.3 shows the predicted e-commerce offering of each store in the test data set against the actual offering. The top right cell shows the number of stores with no e-commerce offerings that the Random Forest model correctly classified as having no offering. The other cells in the first column show those stores that were mis-classified. Of the total number of stores in the test data set, 89.9% were correctly classified. It is possible to break down accuracy based on certain categorical variables within the test data. Table 4.5 and Table 4.6 display model accuracies by grocery banners and BEA economic regions, respectively. Table 4.5 shows high model accuracy for the majority of grocery banners included in the store sample. The model had weaker accuracy for select smaller regional supermarkets, including Ingles Markets and

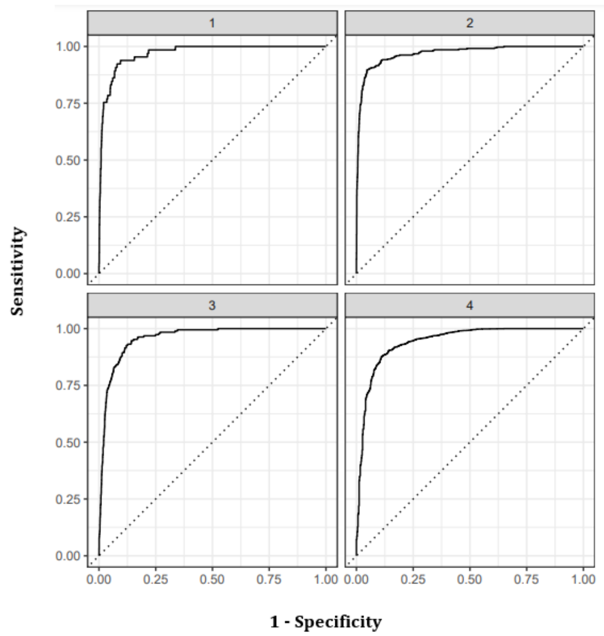

Figure 4.2. Random Forest Model – Test Data Class ROC Curves

Actual Offering

Predicted Offering	*No Offering*	*Pickup*	*Delivery*	*Both*
No Offering	36	15	0	18
Pickup	19	267	0	40
Delivery	2	0	137	102
Both	8	61	52	2099

Figure 4.3. Random Forest Model – Test Data Confusion Matrix

Table 4.5. Random Forest Model – Test Accuracy by Grocery Banner

Grocery Banner	Model Accuracy
Acme Markets	100%
Albertsons	95%
ALDI	68%
BJ's Wholesale Club	90%
Giant Food	76%
Ingles Markets	76%
Jewel-Osco	94%
Publix	97%
Safeway	86%
Sam's Club	100%
Shaw's	100%
Target	92%
Trader Joe's	96%
Vons	95%
Walmart	94%
Whole Foods Market	87%

Giant Food. The model was least accurate in determining e-commerce offerings at Aldi locations. In Table 4.6, model accuracy is displayed by regions within the contiguous United States. Model accuracy is strong across all regions, with no extreme outliers or evidence of spatial heterogeneity in performance. Notably, the variation in accuracy is minimal.

The final exhibit in this results section displays variable importance among the demographic and competitive feature fed into the model. Figure 4.4 shows variable importance factors for all trade area features. The variable importance factors on the x-axis of Figure 4.4 capture the increase in the model's prediction errors after permuting the feature. In this case, a feature is considered important if randomizing its values across observations increases the overall model error. This permutation approach was developed by Breiman (2001) for Random Forest models, and was augmented to a model-agnostic metric by Fisher et al. (2018). Population density is shown to be the most important demographic feature in determining the optimal e-commerce offering for a given trade area- followed by income, education, and overall e-commerce demand. Features capturing competitive saturation in trade areas, population growth, and vehicle access are demonstrably less important in the model. Figure 4.4 would seem to support Anderson et al.'s (2003) innovation-diffusion hypothesis implicitly. The importance of population density suggests that grocery chains are implementing increased e-commerce offerings in their more urban trade areas. However, we cannot know if this decision is fueled by data-driven research or external logistical/economic considerations.

Table 4.6. Random Forest Model – Test Accuracy by Economic Region

BEA Economic Region	Model Accuracy
Far West	91%
Great Lakes	86%
Mideast	86%
New England	89%
Plains	91%
Rocky Mountain	89%
Southeast	89%
Southwest	92%

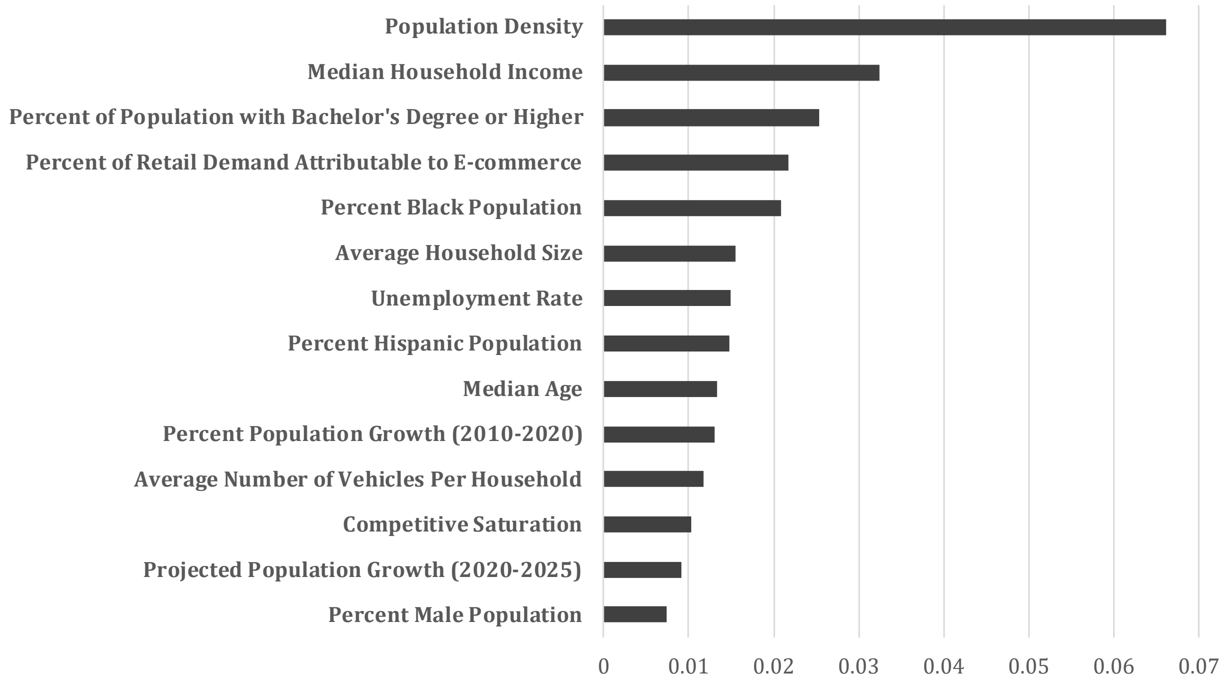

Figure 4.4. Random Forest Model – Variable Importance

The following section will offer further discussion of these model results, with additional context added through a discussion of specific types of model errors, followed by an additional scenario of applying the model in competitive real estate strategy to support location analysis and competitive defense for a hypothetical grocery chain.

Discussion

A grocery banner's decision regarding whether to offer pickup, delivery, both, or neither to customers in the communities it serves is a critical one. The Random Forest model developed for this study effectively pulls back the curtain on how various national and regional grocery chains are reckoning with this decision in markets across the United States. The model serves not only to explore and describe how these decisions are made, but it can also generate new insights that can inform praxis in location strategy and real estate market research. To this end, one of the most noteworthy results of the Random Forest model is not only which e-commerce offerings were correctly classified at a given grocery store, but also which were misclassified, and what those misclassifications represent in practice.

Case Study 1: Missed Opportunities and Mis-investments

There are two types of misclassifications that are of particular interest in this case study. The first is a case in which the Random Forest model classified the e-commerce offering for a given store as both pickup and delivery, a full complement of e-commerce options for consumers, but the true classification of the store in the sample show the store has no e-commerce offering at all. This misclassification may be considered a "missed opportunity". These are stores that, based on the composition of their trade areas, could and should be offering grocery pickup and delivery their customers. However, at the time this sample was taken, were not doing so. Across both training and test sets, 22 observations were mis-classified in this way. The top five are shown below in Table 4.7 based on the class probability that the stores should be offering both pickup and delivery. The discrepancy between class probabilities is clear here, with all stores having a relatively small probability of not offering e-commerce channels to customers based on the characteristics of their trade areas. The application of this model shows these chains markets where they are missing an opportunity to provide e-commerce to their customers. However, it should be noted that a grocery banner's motivations for not offering e-commerce can extend beyond considerations of trade area composition or demographics. It is possible e-commerce is not offered at these locations for various other reasons, including but not limited to, corporate politics, unsustainable economics, carbon footprint, logistical challenges, etc.

The second type of misclassification of interest here is the opposite of the first case. These are errors in which a store is classified as having no offering based on trade area composition, but the store currently offers a full complement of e-commerce offerings available to its customers. Across both training and test sets, this type of misclassification occurs for 32 stores. The top five are shown in Table 4.8 based on the class probability that the stores should be offering no e-commerce to their customers. This classification error marks an instance of a store offering a full complement of e-commerce offerings in a trade area that may not require such a robust offering. Consider the first three stores in Table 4.8. These are all stores in peripheral small- to mid-sized

Table 4.7. Stores in Sample Offering No E-commerce That Are Classified as Offering Pickup & Delivery

Grocery Banner	Location	Probability of No Offering	Probability of Pickup & Delivery
BJ's Wholesale Club	Lady Lake, Florida	1.0%	96.6%
BJ's Wholesale Club	Johnston, Rhode Island	21.7%	78.0%
Sam's Club	Tampa, Florida	19.3%	77.2%
Sam's Club	Linden, New Jersey	15.8%	74.3%
Safeway	Chehalis, Washington	9.7%	68.5%

Table 4.8. Stores in Sample Offering Pickup & Delivery That Are Classified as No Offering

Grocery Banner	**Location**	**Probability of No Offering**	**Probability of Pickup & Delivery**
Whole Foods Market	Ft. Collins, Colorado	79.1%	20.6%
BJ's Wholesale Club	Hyannis, Massachusetts	78.7%	20.9%
Safeway	Grants Pass, Oregon	73.3%	22.8%
Whole Foods Market	New York, New York	72.9%	26.3%
Whole Foods Market	New York, New York	70.3%	28.8%

city markets across the country that tend toward lower population densities. For these types of markets, investing in systems and infrastructure to provide customers with both grocery pickup and delivery options may be a mis-investment of capital. Whole Foods Stores in New York City being classified as having no e-commerce offering is, indeed, counter-intuitive and inconsistent with the overall feature importance returned by the model results. Whole Foods Markets is one of the banners for which the model performs with the lowest accuracy, which might explain why such dramatic misclassifications are common for Whole Foods stores. This reflects that the Random Forest model has the most difficulty in understanding the e-commerce strategy of Whole Foods across space.

Both of these misclassification errors support a compelling case study in which insights from the Random Forest model may be applied to inform e-commerce strategy for grocery chains. In the most extreme cases, grocery banners must consider carefully whether there offering in a given market is appropriate given the characteristics of the trade area. The following case study provides an additional application of the model to support pre-emptive real estate strategy planning and competitive defense.

Case Study 2: E-commerce Probability Surface

In this second case study, the Random Forest model is applied to net new trade areas to derive a spatial surface that captures the probability that a given store will offer a full complement of e-commerce offerings in a given trade area. It is assumed here that *Safeway*, a supermarket chain with a store network spanning the western half of the United States, is hypothetically planning a new deployment of stores in the southeast region. Existing grocers in the southeast may be thinking critically about how to defend against this new competitive threat in their markets, including how best to optimize their e-commerce offerings across the chain to pre-emptively respond to *Safeway's* strategy. The Random Forest model is well-equipped to support this strategic challenge.

In this exercise, the Random Forest model is applied to a contiguous hexagonal grid covering the southeast region of the United States. Hexagons are built using an area of 43.7 square miles, the median area of all supermarket trade areas contained in the training and test data sets. The same suite of demographic and competitive features is pulled for each hexagonal area using *SiteWise Analytics*. The Random Forest model is then fed this hexagonal grid as a substitute for the test data, yielding class probabilities for each hexagon within the grid. The resultant grid is a probability surface across the southeastern United States, with the probabilities that a *Safeway* supermarket in each trade area would offer pickup, delivery, both, or neither. The class probabilities for offering both pickup and delivery are mapped in Figure 4.5.

Intuitively, the probability surface for offering both pickup and delivery shows higher probabilities around metropolitan markets, reflecting *Safeway's* e-commerce location strategy. This type of probability surface is a valuable tool for southeastern grocery chains in preparing competitive defense strategies to meet the (hypothetical) network expansion of *Safeway* into the Southeast region.

While this case study is hypothetical, it demonstrates a valuable application of the Random Forest multi-classification model as a tool for competitive defense and pre-emptively anticipating the e-commerce strategies that may be adopted by competitors in a net new market. For the retail geographer in practice, this model is designed to determine

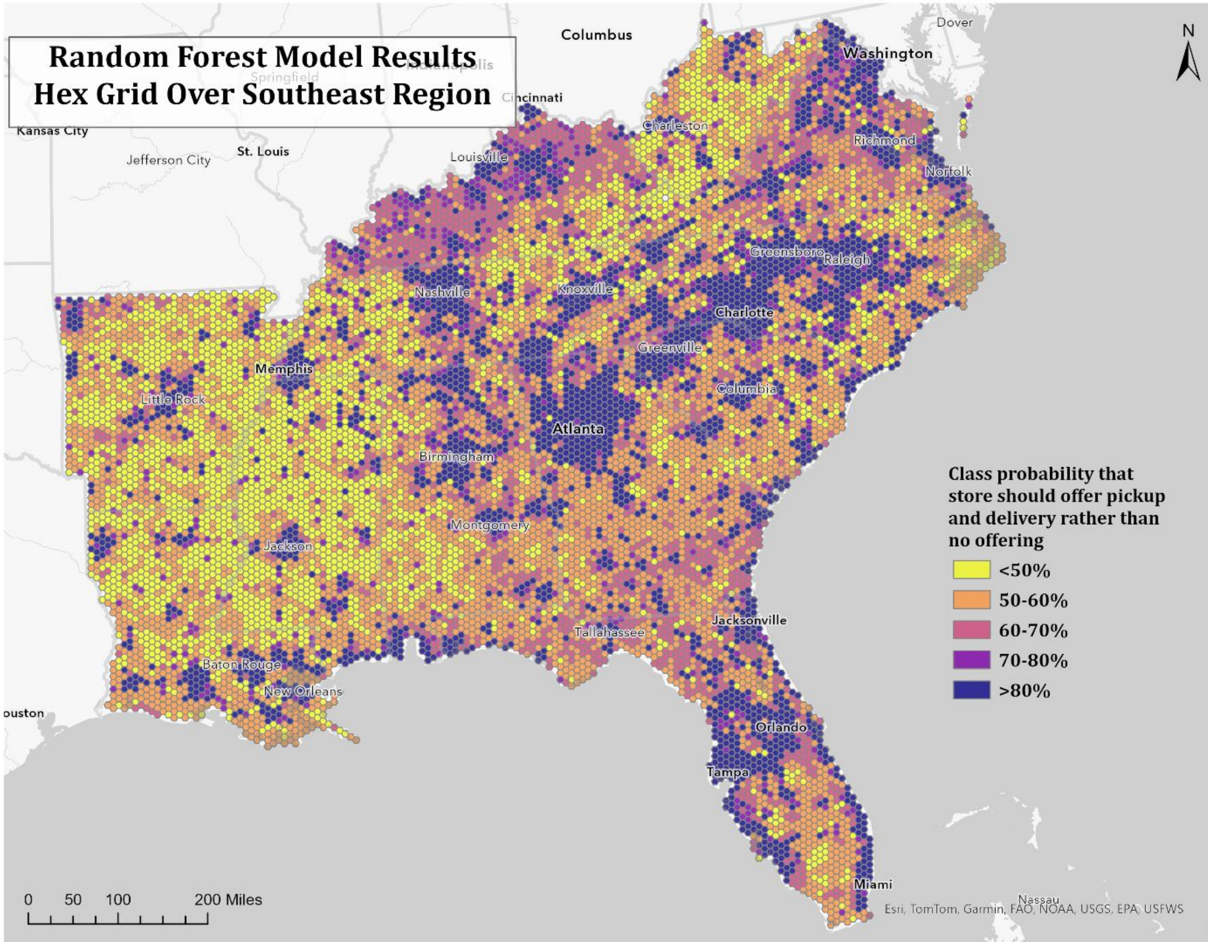

Figure 4.5. ***Safeway*** **E-commerce Probability Surface – Southeast United States**

how likely it is that a competitor opening in a given market will offer e-commerce to consumers.

Conclusion

This analysis has built upon previous literature from geography, economics, and business scholarship concerned with the drivers of e-commerce consumption by making a crucial pivot from the lens of consumer behavior to the decisions of the firm (Kirby-Hawkins et al., 2019). A Random Forest model was designed to model the decisions made by U.S. grocery chains at the local level about whether a given store in their network should offer pickup, delivery, both, or no e-commerce options to its customers. The model can accurately prescribe an optimal e-commerce offering for a given store in a given trade area based on the demographic and competitive composition of the trade area. Variable importance factors reveal that grocery chains in the United States are considering income levels, urbanity, and education when making decisions about whether to offer e-commerce channels for shoppers in a given trade area. While the focus of this paper is on firm location decisions rather than consumer behavior, these trade area characteristics build off of previous literature in two ways. First, the prevailing narrative that more affluent consumers are more likely to adopt e-commerce (De Blasio, 2008; Clarke et al., 2015; Van Droogenbroeck & Van hove, 2017; Pernot, 2021) is corroborated by the variable inflation factors of the Random Forest model. That these findings align signals that many major grocery chains in the United States are, to an extent, aware of this relationship and are designing their e-commerce strategies accordingly. Further, the importance of population density as a feature in the model supports Anderson et al.'s

(2003) innovation-diffusion hypothesis that urbanite consumers are more likely to adopt e-commerce than rural consumers, in line with empirical studies that have previously examined the relationship between population density or urbanity and a consumer's propensity to shop online (Farag et al., 2006a; Farag et al., 2006b, Zhen et al., 2018; Shao et al., 2022). This finding is not only supported by the model's variable importance factors. It is also evident in the probability surface derived in the second applied scenario, which emphasizes major metropolitan areas across the Southeast as being particularly suitable for a full complement of e-commerce offerings in the hypothetical case of a *Safeway* expansion.

The misclassification errors of the model can tell interesting stories about cases where grocery banners may have opted for a sub-optimal e-commerce strategy in a given trade area based on its characteristics. Stores with no e-commerce offering that are classified to have a full complement of both pickup and delivery in the model may mark "missed opportunities" to introduce digital offerings in a trade area. Conversely, stores that offer a full complement of pickup and delivery based on a chain's regional e-commerce strategy that are classified to have no offering may mark cases of "mis-investment" in which a banner could enjoy some cost savings by removing e-commerce from their store.

Finally, this analysis provided a case study of applying the Random Forest model to a hex grid surface to derive the probability that a given store will offer e-commerce is hexagonal areas across a region. This probability surface layer is a powerful tool for retail geographers and strategists who wish to pre-emptively respond to the e-commerce strategies of competitors before they even break ground on a new location.

This paper paves avenues for future research that continues to emphasize analysis of firm decisions regarding e-commerce strategies. Crucially, this study identifies that the e-commerce strategies of major U.S. grocery chains do generally align with the prevailing demographic characteristics that drive consumers to purchase groceries online. Grocery chains are aware that these consumers tend to be more affluent and have planned their strategies accordingly. However, further investigation is needed into the relationship between grocery e-commerce strategies and Anderson et al.'s (2003) theories of e-commerce adoption. Variable importance factors in this study suggest that firms are following the innovation-diffusion hypothesis. However, there may exist opportunities to cater e-commerce offerings to more rural locations in store networks to offer increased convenience and accessibility to customers who live far from urban centers and retail customers. An analysis focuses more narrowly on these rural opportunities may reveal new strategies that may be appropriate for grocery chains wishing to implement e-commerce across more of their networks. Future research may also investigate regional dynamics. While the overall accuracy of the Random Forest model in this study was fairly stable across regions, there remains an outstanding opportunity to test models at smaller geographic scales to uncover new deterministic relationships and dynamics for chains.

Another avenue of future research should focus on the geographies of third-party vendor service areas. These vendors like Instacart and *DoorDash* have deployed their own service areas across urban markets that are subject to not only consumer demand, but also labor supply. Their business models allow consumers to purchase groceries (and other restaurant and general merchandise items) through an app. An independent courier is then dispatched to deliver the item. The ways in which these third-party vendors interplay with grocery stores offering their own e-commerce offerings remains unexplored in academic research, and outstanding questions still remain about how third party delivery services alter or expand geographies of e-commerce in the grocery industry.

A final compelling research direction would be to conduct more rigorous spatial analysis on the e-commerce probability surfaces that can be derived from this type of modelling exercise. Building a hex grid across the entire country could allow for additional predictive modeling and spatial analysis to understand prevailing geographic patterns in the probabilities that stores should offer e-commerce or not. This type of surface analysis could also be further operationalized into dashboard environments in which chains could test different scenarios for different competitors with different trade area sizes. This type of dynamic research product would be valuable for many grocery chains who are continuing to critically consider optimal strategies in e-commerce as online shopping continues its ascent as one of the primary ways that consumers buy and shop for goods and services.

Chapter Five
Conclusion

This body of work found in this dissertation endeavors to contribute to a growing body of geographic scholarship concerned with the spatial forces driving consumers to shop online, and the tandem decisions made by firms to cater to and spur this demand. In comparison to other seismic disruptors in the retail environment such as the mall and the supercenter, geographers have had markedly less to say regarding e-commerce. Only in recent years has a stronger push been made to analyze the new spatial dynamics underlying consumption in physical and digital worlds, driven in no small part by the COVID-19 pandemic. The pandemic's acceleration of macroeconomic (and for that matter, microeconomic) forces and alternations to consumers' day-to-day shopping habits infused the retail geographers with a renewed energy to investigate the many geographic complexities underlying online shopping. Many such studies have, historically, focused specifically on consumer behavior, remarking on how demographic characteristics and spatial dynamics may influence an individual's decision to consume online or in-store. In alignment with Kirby-Hawkins et al. (2019), the body of research composed here places much-needed emphasis on firm-level decisions rather than the behaviors of the consumer – pulling back the curtain on the location strategies that ultimately drive where and how we shop.

The objectives of this dissertation were to: 1) produce a timely comprehensive literature review of literature review of geographic thought focused on digital consumption, 2) a spatial econometric of e-commerce consumption that shifts perspective to the decisions of the firm, and how those decisions may be supported by the demographic characteristics, competitive landscape, and spatial dynamics of grocer trade areas, and 3) the development of a machine learning algorithm that uses a prescriptive approach to optimize grocery location strategy. Achieving these objectives has generated new knowledge in retail geography regarding the spatial dynamics underscoring e-commerce strategies in the U.S. grocery industry. Further, understanding these dynamics informs the construction of a more complex supervised learning algorithm designed to appropriately assign an optimal e-commerce strategy (delivery, pickup, both, or none) to a given store location based on its trade area composition.

Industry Takeaways and Future Research

The findings of this dissertation establish exiting new opportunities for practitioners and analysts in the retail and marketplaces industry to apply analytical methods commonly used in sales forecasting and location planning to strategic problems in e-commerce. These methods can be used not only to derive an optimized location strategy for the business, but also to investigate what factors may drive those same decisions for their competitors. The research in this dissertation identifies household incomes, education, ethnicity, age, and household size to all be significant factors in determining whether or not a given grocery store should offer some form of e-commerce offering or not. It also found, however, that these broader trade area trends exhibit regional heterogeneity – establishing an avenue for future research that focuses on smaller geographies and more nuanced variations in determinants among regions or specific market areas. Further, practitioners in the industry must be aware that, as populations and demographics continue to shift over time, these deterministic relationships between demographics and spatial dynamics and e-commerce strategy will inevitably shift in turn.

Random forest modeling is also found to be an excellent tool for prescribing an optimal e-commerce offering based on these trade area characteristics – able to classify stores more granularly by determining if pickup, delivery, both, or neither should be offered for a given market. There is an outstanding opportunity for the grocery industry to leverage this type of modeling in their own planning practices to achieve cost savings, improve competitive defense, and optimized e-commerce deployment over large store networks.

In addition to these industry implications, this combined body of research also proposes new trajectories of inquiry for scholars in retail geography, economics and business. There are, as a matter of course, intangible business-specific factors that may drive e-commerce location strategies to the same extent, or even more, than the composition of trade areas or the potential for successful implementation. These factors can be political, environmental, financial, or even cultural. An intuitive sequel to this body of work should apply qualitative research methods to connect with relevant stakeholders in the U.S. grocery industry to develop a clearer picture of how e-commerce strategies across chains are being implemented. For example, perhaps pickup or delivery is not offered at a given store to cut costs, to reduce carbon footprint, or to weather labor shortages. There are any number of reasons beyond what is informed by the trade area that may be found by a more high-touch qualitative study of grocery chains. This type of study would align well with the history of qualitative research in retail geography concerned with how retail chains invest in new technologies and methodologies to optimize their real estate portfolios (e.g. Aversa et al., 2018; Wood & Reynolds, 2011).

Another compelling direction for new research would leverage game theory to focus more explicitly on the competitive dynamics that drive retail chains to implement certain e-commerce strategies. There is a game-like structure that may capture how retail chains observe the actions of their competitors and attempt to either match or surpass their competitors' strategies in terms of cost or convenience to the consumer. Grocery stores deciding to offer pickup or delivery based on the decisions of their competition makes fertile ground for microeconomic research. While this type of dependence effect is captured implicitly in the spatial econometric model used in the second chapter, it is not investigated or analyzed in full.

A final avenue of research concerns the unique geographies of third-party delivery service areas and their influence on whether stores in a market may offer delivery or pickup to consumers. This study may include stores and restaurants beyond the grocery sector. Third-party delivery services like Postmates, UberEats, DoorDash, GrubHub, and Instacart have become prevalent in almost all U.S. cities, allowing retailers to outsource last-mile delivery to consumers. Both national and local retailers must make important decisions about whether to partner with a third-party courier as part of their convenience offering to consumers, and if so, which courier. Sitting at the intersection of retail and transportation geography, such a study would deliver compelling findings regarding the spatial dynamics underscoring the relationship between retailers and these couriers. How are courier service areas derived? Are these areas optimized for every market? Can a methodology appropriately match a store with a third-party courier based on overlap in service area and primary trade area?

In summation, this body of dissertation work has taken important steps to introduce novel geographic scholarship concerned with the unique spatial complexities of shopping online in the 21st century. The strategic decisions made by some of the largest grocery chains in the United States have been analyzed – gauging deterministic relationships between trade area composition and e-commerce strategy for a given store location. This work has been augmented with a more complex Random Forest model capable or adeptly prescribing an optimal e-commerce strategy for a given store and its trade area. Optimistically, this work will spur the development of new studies of the geographies of e-commerce, and set a fresh foundation for new analysis of the spatial dynamics of e-commerce by both scholars and practitioners in the industry.

Bibliography

Alhaimer, R. (2022). Fluctuating attitudes and behaviors of customers toward online shopping in times of emergency: The case of Kuwait during the COVID-19 pandemic. *Journal of Internet Commerce*, *21*(1), 26-50.

Ali, S., Khalid, N., Javed, H., & Islam, D. (2020). Consumer adoption of online food delivery ordering (OFDO) services in Pakistan: The impact of the COVID-19 pandemic situation. *Journal of Open Innovation, 7*(10).

Anderson, W. P., Chatterjee, L., & Lakshmanan, T. R. (2003). E-commerce, transportation, and economic geography. *Growth and Change, 34*(4), 415-432.

Andrews, R. L. & Currim, I. S. (2004). Behavioral differences between consumers attracted to shopping online versus traditional supermarkets: Implications for enterprise design and marketing strategy. *International Journal of Internet Marketing and Advertising*, *1*(1), 38-61.

Aoyama, Y. (2001). Structural foundations for e-commerce adoption: A comparative organization of retail trade between Japan and the United States. *Urban Geography, 22*(2), 130-153.

Aversa, J., Hernandez, T., & Doherty, S. (2020). Spatial big data and business location decision-making: Opportunities and challenges. In E. Vaz (Ed.), *Regional intelligence: Spatial analysis and anthropogenic regional challenges in the digital age* (205-224). Springer.

Baker, M. (1999). Multi-channel retailing. *International Council of Shopping Centers Research Quarterly*, *6*(3), 13-17.

Baker, M. (2000). Digital distribution. *International Council of Shopping Centers Research Quarterly, 7*(3), 21-27.

Baker, M. (2000). Online grocery shopping: Time for a stock-taking. *International Council of Shopping Centers Research Quarterly, 7*(1), 16-19.

Bakos, Y. (1998). The emerging role of electronic market places on the Internet. *Communications of the Association for Computing Machinery, 41*(8), 35-42.

Ballantyne, P., Singleton, A., Dolega, L., & Credit, K. (2021). A framework for delineating the scale, extent and characteristics of American retail centre agglomerations. *Environment and Planning B*, *49*(3), 1112-1128.

Beckers, J., Cárdenas, I, & Verhetsel, A. (2018). Identifying the geography of online shopping adoption in Belgium. *Journal of Retailing & Consumer Services*, *45*, 33-41.

Beckers, J., Weekx, S., Beutels, P., & Verhetsel, A. (2021). COVID-19 and retail: The catalyst for e-commerce in Belgium? *Journal of Retailing and Consumer Services*, *62*, 102645.

Beckers, J., Birkin, M., Clarke, G., Hood, N., Newing, A., & Urquhart, R. (2022). Incorporating e-commerce into retail location models. *Geographical Analysis*, *54*, 274-293.

Bell, D. R. (2014). *Location is (still) everything*. Houghton Mifflin Harcourt.

Bennett, T. (2019, January 15). Georgia's grocery market is 'extremely competitive'. The Shelby Report. https://www.theshelbyreport.com/2019/01/03/georgia-market-extremely-competitive/

Birkin, M., Clarke, G., & Clarke, M. (2002). *Retail geography & intelligent network planning.* John Wiley & Sons, Ltd.

Birkin, M., Clarke, G., & Clarke, M. (2017). *Retail location planning in an era of multi-channel growth*. Routledge.

Bishop, D. (2023, October 11). *Tracking online grocery's growth*. Brick Meets Click. https://www.brickmeetsclick.com/tracking-online-grocery-s-growth-

Brand, D. (2000). Digital distribution. *International Council of Shopping Centers Research Quarterly*, *7*(3), 21-27.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*, 5-32.

Bryson, J. R. (2021). COVID-19 and the immediate and longer-term impacts on the retail and hospitality industries: dark stores and turnover-based models. In J. R. Bryson,

L. Andres, A. Ersoy & L. Reardon (Eds.), *Living with pandemics: Places, people and policy* (pp. 202-217). Edward Elgar Publishing, Ltd.

Buckwalter, D. W. (1989). Effects of competition on the pattern of retail districts in the

Bureau of Economic Analysis. (2023). *Statistical Areas*. https://apps.bea.gov/regional/docs/msalist.cfm

Cairncross, F. (1997). *The death of distance*. Orion Publishing Group, Ltd.

Cao, X. (2009). E-shopping, spatial attributes, and personal travel: A review of empirical studies. *Transportation Research Record*, *2135*, 160-169.

Cao, X., Chen, Q., & Choo, S. (2013). Geographic distribution of e-shopping: Application of structural equation models in the Twin Cities of Minnesota. *Transportation Research Record*, *2383*, 18-26.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, *16*, 321-357.

Chenarides, L., Grebitus, C., Lusk, J. L., & Printezis, I. (2020). Food consumption behavior during the COVID-19 pandemic. *Agribusiness*, *37*, 44-81.

Chi, G. & Zhu, J. (2020). *Spatial regression models for the social sciences*. SAGE Publications.

Chocarro, R., Cortiñas, M., & Villanueva, M. (2013). Situational variables in online versus offline channel choice. *Electronic Commerce Research and Applications*, *12*, 347-361.

Clarke, G., Thompson, C., & Birkin, M. (2015). The emerging geography of e-commerce in British retailing. *Regional Studies, Regional Science*, *2*(1), 371-391.

Cliff, A. D. & Ord, J. K. (1981). *Spatial processes: Models and applications*. Pion.

Cox, K. & Tamir, C. (2022, April 14). *Place and community*. Pew Research Center.

https://www.pewresearch.org/race-ethnicity/2022/04/14/black-americans-place-and-community/

Comber, S., Arribas-Bel, D., Singleton, A., Dong, G., & Dolega, L. (2019). Building hierarchies of retail centers using Bayesian multilevel models. *Annals of the American Association of Geographers*, *110*, 1150-1173.

Currah, A. (2002). Behind the web store: The organizational and spatial evolution of multichannel retailing in Toronto. *Environment and Planning A, 34*(8), 1411-1441.

Dannenberg, P., fuchs, M., Riedler, T., & Wiedemann, C. (2020). Digital trasition by COVID-19 pandemic? The German food online retail. *Tijdschrift voor Economische en Sociale Geografie*, *111*(3), 543-560.

Datta, H., Knox, G., & Bronnenberg, B. J. (2017). Changing their tune: How consumers' adoption of online streaming affects music consumption and discovery. *Marketing Science*, *27*(1), 5-21.

Darmofal, D. (2015). *Spatial analysis for the social sciences*. Cambridge University Press.

Daunfeldt, S., Mihaescu, O., Nilsson, H., & Rudholm, N. (2017). What happens when IKEA comes to town? *Regional Studies*, *51*(2), 313-323.

Davies, A., Dolega, L., & Arribas-Bel, D. (2019). Buy online collect in-store: Exploring grocery click & collect using a national case study. *International Journal of Retail & Distribution Management*, *47*(3), 278-291.

Davies, R. L. (1976). *Marketing geography*. Routledge.

de Blasio, G. (2008). Urban-rural differences in Internet usage, e-commerce, and e-banking: Evidence from Italy. *Growth and Change*, *39*(2), 341-367.

de la Llave Montiel, M. A. & López, F. (2020). Spatial models for online retail churn: Evidence from an online grocery delivery service in Madrid. *Papers in Regional Science*, *99*, 1643-1665.

Darmofal, D. (2015). *Spatial analysis for the social sciences*. Cambridge University

Press.

Daunfeldt, S., Mihaescu, O., Nilsson, H., & Rudholm, N. (2017). What happens when IKEA comes to town? *Regional Studies*, *51*(2), 313-323.

Dennis. S. (2020). *Remarkable retail*. LifeTree Media, Ltd.

Dias, F. F., Lavieri, P. S., Garikapati, V. M., Astroza, S., Pendyala, R. M., & Bhat, C. R. (2017). A behavioral choice model of the use of car-sharing and ride-sourcing services. *Transportation*, *44,* 1307-1323.

Dodge, M. (2001). Finding the source of Amazon.com: Examining the store with the 'Earth's biggest selection'. In T. R. Leinbach & S. D. Brunn (Eds.), *Worlds of E-commerce: Economic, geographical and social dimensions* (pp. 167-180). John Wiley & Sons, Ltd.

Driediger, F. & Bhatiasevi, V. (2019). Online grocery shopping in Thailand: Consumer acceptance and usage behavior. *Journal of Retailing and Consumer Services*, *48*, 224-237.

Dwyer, T. (1996). Non-store retailing: Implications of the Internet. *International Council of Shopping Centers Research Quarterly, 3*(3), 7-13.

Dwyer, T. (1998). Retailing on the World Wide Web. *International Council of Shopping Centers Research Quarterly, 5*(1), 10-12.

Eisenhauer, E. (2001). In poor health: Supermarket redlining and urban nutrition. *GeoJournal*, *53*, 125-133.

Enoch, M., Monsuur, F., Palaiologou, G., Quddus, M. A., Ellis-Chadwick, F., & Morton, C. (2022). When COVID-19 came to town: Measuring the impact of the coronavirus pandemic on footfall on six high streets in England. *Environment and Planning B, 49*(3), 1091-1111.

Etumnu, C., Widmar, N. O., Foster, K., & Ortega, D. (2019). Proceedings from the *2019 Agricultural & Applied Economics Association annual meeting.* Atlanta, GA.

Farag, S., Schwanen, T., & Dijst, M. (2005). Empirical investigation of online searching and buying and their relationship to shopping trips. *Transportation Research Record*, *1926*, 242-251.

Farag, S., Weltevreden, J., van Rietbergen, T., Dijst, M., & van Oort, F. (2006a). E-shopping in the Netherlands: Does geography matter? *Environment & Planning B: Planning and Design*, *33*, 59-74.

Farag, S., Krizek, K., & Dijst, M. (2006b). E-shopping and its relationship with in-store shopping: empirical evidence from the Netherlands and the USA. *Transport Reviews*, *26*, 43-61.

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 A*, *41*, 125-141.

Fischer, A., Rudin, C., & Dominici, F. (2018). Allmodels are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously. *Journal of Machine Learning Research*, *20*(177), 1-81.

FTI. (2023, July 10). *U.S. online retail sales to reach $1.14 trillion in 2023*. FTI Consulting, Inc. https://www.fticonsulting.com/about/newsroom/press-releases/fti-consulting-us-online-retail-sales-to-reach-114-trillion-in-2023

Futterman, R. K. (2000, May). E-commerce and its impact on the retail real estate market. *Real Estate Weekly*.

Ganning, J. P. & Green, T. (2021). The online marketplace: Zero-order city or new source of social inequality? *Growth and Change*, *52*, 1251-1270.

Gao, X., Shi, X., Guo, H. & Liu, Y. (2020). To buy or not buy food online: The impact of the COVID-19 epidemic on the adoption of e-commerce in China. *PLoS ONE*, *15*(8).

Golant, S. M. (2019). Stop bashing the suburbs: Mobility limitations of older residents are less relevant as connectivity options expand. *Journal of Aging Studies*, *50*, 100793.

Goulard, M., Laurent, T., & Thomas-Agnan, C. (2017). About predictions in spatial autoregressive models: Optimal and almost optimal strategies. *Spatial Economic Analysis*, *23*(2-3), 304-325.

Grashuis, J., Skevas, T. & Segovia, M. S. (2020). Grocery shopping preferences during the COVID-19 pandemic. *Sustainability*, *12*, 5369.

Guthrie, C., Fosso-Wamba, S., & Arnaud, J. B. (2021). Online consumer resilience during a pandemic: An exploratory study of e-commerce behavior before, during and after a COVID-19 lockdown. *Journal of Retailing and Consumer Services*, *61*, 102570.

Hai-dong, Z., Xin-wu, Y., Hongbo, L., & Ping, L. (2017). Study of spatial pattern of county-level e-commerce and its influential factors in the Zhejiang Province, China. *International Journal of u- and e-Service, Science and Technology*, *10*(4), 1-14.

Hand, D. J. & Till, R. J. (2001). A simple generalization of the Area Under the ROC Curve for multiple class classification problems. *Machine Learning*, *45*, 171-186.

Hao, N., Wang, H. H., & Zhou, Q. (2020). The impact of online grocery shopping on stockpile behavior in COVID-19. *China Agricultural Economic Review*, *12*(3), 459-470.

Hashem, T. N. (2020). Examining the influence of COVID-19 pandemic in changing customers' orientation towards e-shopping. *Modern Applied Science*, *14*(8), 59-76.

Hernandez, T., Gomez-Insausti, R., & Biasiotto, M. (2001). Non-store retailing and shopping center vitality. *Journal of Shopping Centre Research*, 57-81.

Hood, N., Urquhart, R., Newing, A., & Heppenstall, A. (2020). Sociodemographic and spatial disaggregation of e-commerce channel use in the grocery market in Great Britain. *Journal of Retailing and Consumer Services*, *55*, 102076.

Hunneman, A., Elhorst, J. P., & Bijmolt, T. (2021). Store sales evaluation and prediction using spatial panel data models of sales components. *Spatial Economic Analysis*, *17*(1), 127-150.

Jank, W. & Kannan, P. K. (2005). Understanding geographical markets of online firms using spatial models of customer choice. *Marketing Science*, *24*(4), 623-634.

Jones, K. & Biasiotto, M. (1999). Internet retailing: Current hype or future reality? *International Review of Retail, Distribution and Consumer Research*, *9*(1), 69-75.

Kawasaki, T., Wakashima, H., & Shibasaki, R. (2022). The use of e-commerce and the COVID-19 outbreak: A panel data analysis in Japan. *Transport Policy*, *115*, 88-100.

Kirby-Hawkins, E., Birkin, M., & Clarke, G. (2019). An investigation into the geography of corporate e-commerce sales in the UK grocery market. *Environment and Planning B*, *46*(6), 1148-1164.

Klier, T. & McMillen, D. P. (2008). Clustering of auto supplier plants in the United States: Generalized Method of Moments spatial logit for large samples. *Journal of Business & Economic Statistics*, *26*, 460-471.

Krizek, K. J., Li, Y., & Handy, S. L. (2005). Spatial attributes and patterns of use in household-related information and communications technology activity. *Transportation Research Record*, *1926*, 252-259.

Kopczewska, K. (2022). Spatial machine learning: New opportunities for regional science.

Koster, H. R. A., van Ommeren, J., Tang, C. K., & Bras, N. COVID-19 restriction policies and shopping streets. *PLoS ONE*, *17*(7).

Lee, E., Lee, J., & Eastwood, D. (2003). A two-step estimation of consumer adoption of technology-based service innovations. *Journal of Consumer Affairs*, *37*(2), 256-282.

Leinbach, T. R. & Brunn, S. D. (Eds.). (2001). *Worlds of E-commerce: Economic, geographical and social dimensions*. John Wiley & Sons, Ltd.

Leinbach, T. R. & Brunn, S. D. (2001). E-commerce: Definitions, dimensions, and constraints. In T. R. Leinbach & S. D. Brunn (Eds.), *Worlds of E-commerce: Economic, geographical and social dimensions* (pp. xi-xviii). John Wiley & Sons, Ltd.

LeSage, J. P. (2000). Bayesian estimation of limited dependent variable spatial autoregressive model. *Geographical Analysis*, *32*, 19-35.

Llovio, L. (2021, July 8). Grocers look for an edge as competition gets even tighter. Business Observer. https://www.businessobserverfl.com/news/2021/jul/08/grocers-look-for-an-edge-as-competition-gets-even-tighter/

Li, H., Kuo, C., & Rusell, M. G. (1999). The impact of perceived channel utilities, shopping orientations, and demographics on the consumer's online buying behavior. *Journal of Computer-Mediated Communication*, *5*(2).

Loo, B. & Wang, B. (2018). Factors associated with home-based e-working and e-shopping in Nanjing, China. *Transportation*, *45*, 365-384.

Lord, J. D. (2000). Retail saturation: What the experts say. *Southeastern Geographer, 40*(1), 106-115.

Malecki, E. J. & Gorman, S. P. (2001). E-commerce: Definitions, dimensions, and constraints. In T. R. Leinbach & S. D. Brunn (Eds.), *Worlds of E-commerce: Economic, geographical and social dimensions* (pp. 87-105). John Wiley & Sons, Ltd.

Martinetti, D., & Geniaux, G. (2017). Approximate likelihood estimation of spatial probit models. *Regional Science & Urban Economics*, *64*, 30-45.

McMahan, J. (1999). The impact of E-commerce on real estate. *Real Estate Issues, Winter*.

McMillen, D. P. (1992). Probit with spatial autocorrelation. *Journal of Regional Science, 32*(3), 335-348.

Moran, C. D. (2023a, January 31). Online grocery sales will increase at 12% annual rate over 5 years, report says. Grocery Dive.

https://www.grocerydive.com/news/online-grocery-sales-will-increase-at-12-annual-rate-over-5-years-report/641578/#:~:text=Over%20the%20next%20five%20years,Brick%20Meets%20Click%20and%20Mercatus

Moran, C. D. (2023b, August 30). *Walmart's e-grocery sales share hits record high.* Grocery Dive. https://www.forbes.com/sites/joanverdon/2023/03/22/grocery-home-furnishings-continue-to-drive-ecommerce-growth/?sh=614143c26ce7

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

Nair, B. & Lambert, J. (2002). Internet retail: Did anything happen while no one was looking? *International Council of Shopping Centers Research Quarterly, 5*(4), 8-11.

Newing, A., Hood, N., & Sterland, I. (2020). Planning support systems for retail location planning. In J. Stillwell & S. Geertman (Eds.), *Handbook on planning support science* (pp. 459-470). Edward Elgar.

Nicola, M., Alsafi, Z., Sohrabi, C., Kerwan, A., Al-Jabir, A., Iosifidis, C., Agha, M., & Agha, R. (2020). The socio-economic implications of the coronavirus pandemic (COVID-19): A review. *International Journal of Surgery*, *78*, 185-193.

Novkaniza, F., Djuraidah, A., Fitrianto, A., & Sumertajaya, I. M. (2019). Simulation study for comparison of spatial autoregressive probit estimation methods. *IOP Conference Series: Earth and Environmental Science*, *299.*

Pernot, D. (2021). Internet shopping for everyday consumer goods: An examination of the purchasing and travel practices of click and pickup outlet customers. *Research in Transportation Economics*, *87*, 100817.

Pinske, J. & Slade, M. E. (1998). Contracting in space: An application of spatial statistics to discrete-choice models. *Journal of Econometrics*, *85*(1), 125-154.

Posey, A. S. (1994). Grocery retail location and function in Columbia, South Carolina.

Southeastern Geographer, *32*(2), 92-107.

PR Newswire (2023, June 23). *The online grocery market to grow by USD 740.88 billion from 2022 to 2027; the rapid growth in e-commerce to boost market growth – Technavio*. Yahoo! Finance. https://finance.yahoo.com/news/online-grocery-market-grow-usd-051500679.html?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAAF0q78DHUichFWRZjOGTYy6kpouyQchHQocPR-1hwdUkxPTkfJpNFUkX9QAH1MTO0EoHaqO_oiabuFSb6hd9kLZVWZMNsw-qWzHo9W9AjQBkgvZCCLQ1ZCnTzHShZM2bkY-5dB-9aDu5OidSk-kQy7uKHaDLk6OjeAj3QyW4GFAJ

Prisani, N. (2021, May 13). *How COVID-19 will change the geography of competition*. MIT Sloan Management Review. https://sloanreview.mit.edu/article/how-covid-19-will-change-the-geography-of-competition/

Punj, G. (2011). Effect of consumer beliefs on online purchase behavior: The influence of demographic characteristics and consumption values. *Journal of Interactive Marketing*, *25*, 134-144.

Rajagopal, A. (2023, June 13). *May online grocery sales dip overall, but pickup sees 9.1% growth*. Supermarket News. https://www.supermarketnews.com/online-retail/may-online-grocery-sales-dip-overall-pickup-sees-91-growth

Redo, D. J., Aide, T., & Clark, M. L. (2012). The relative importance of socioeconomic and environmental variables in explaining land change in Bolivia, 2001-2010. *Annals of the American Association of Geographers*, *102*(4), 778-807.

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

Rose, N. & Dolega, L. (2022). It’s the weather: Quantifying the impact of weather on retail sales. *Applied Spatial Analysis and Policy*, *15*, 189-214.

Rowlands, D. W., Donoghoe, M., & Perry, A. M. (2023, April). *What the lack of premium grocery stores says about disinvestment in Black neighborhoods.* Brookings Institute. https://www.brookings.edu/research/what-the-lack-of-premium-grocery-stores-says-about-disinvestment-in-black-neighborhoods/

Sauer, C. & Burton, S. (1999). Is there a place for department stores on the Internet? Lessons from an abandoned pilot. *Journal of Information Technology, 14,* 387-398

Saphores, J. & 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*, *87*, 100864.

Shannon, J. (2020). Dollar stores, retailer redlining, and the metropolitan geographies of precarious consumption. *Annals of the American Association of Geographers, 2020*, 1-19.

Shao, R., Derudder, B., & Witlox, F. (2022). The geography of e-shopping in China: On the role of physical and virtual accessibility. *Journal of Retailing and Consumer Services, 64,* 102753.

Soopramanien, D. & Robertson, A. (2007). Adoption and usage of online shopping: An empirical analysis of the characteristics of "buyers" "browsers" and "non-internet shoppers". *Journal of Retailing and Consumer Services*, *14*, 73-82.

Sung, H. (2022). Estimating the spatial impact of neighboring physical environments on retail sales. *Cities*, *123*, 103579.

Thrall, G. I. (2002). *Business geography and new real estate market analysis*. Oxford University Press.

United States Census Bureau. (2023a). Quarterly retail e-commerce sales: 1st quarter 2023. https://www.census.gov/retail/mrts/www/data/pdf/ec_current.pdf

United States Census Bureau. (2023b). Quarterly retail e-commerce sales: 2nd quarter

2023.
https://www.census.gov/retail/mrts/www/data/pdf/ec_current.pdf

van Deursen, A., van Dijk, J., & ten Klooster, P. M. (2015). Increasing inequalities in what we do online: A longitudinal cross section analysis of Internet activities among the Dutch population (2010 to 2013) over gender, age, education, and income. *Telematics and Informatics*, *32*, 259-272.

van Droogenbroeck, E. & van Hove, L. (2017). Adoption of online grocery shopping: Personal or household characteristics? *Journal of Internet Commerce*, *16*(3), 255-286.

van Meeteren, M. & Poorthuis, A. (2018). Christaller and "big data": Recalibarting central place theory via the geoweb. *Urban Geography*, *39*(1), 122-148.

Verdon, J. (2021, May 15). *Adobe: Pandemic shift is permanent, e-commerce to hit $1 Trillion in 2022*. Forbes.
https://www.forbes.com/sites/joanverdon/2021/03/15/adobe-pandemic-shift-is-permanent-e-commerce-to-hit-1-trillion-in-2022/?sh=683deafd44ce

Verdon, J. (2022, March 15). The pandemic changed how we shop for groceries, Adobe report shows. Forbes.
https://www.forbes.com/sites/joanverdon/2022/03/15/the-pandemic-changed-how-we-shop-for-groceries-adobe-report-shows/?sh=601f22fc4bb1

Verdon, J. (2023, March 22). Grocery, home furnishing continue to drive ecommerce *growth*. Forbes.
https://www.forbes.com/sites/joanverdon/2023/03/22/grocery-home-furnishings-continue-to-drive-ecommerce-growth/?sh=614143c26ce7

Walters, H. (2018, March 16). BI-LO closings come amid fierce grocery competition for *shoppers*. Greenville News.
https://www.greenvilleonline.com/story/money/business/2018/03/16/highly-competitive-grocery-market-factor-bi-lo-closings/431713002/

Wang, S. & Du, P. (2021). *Retail geography*. Routledge.

Warganegara, D. L. & Hendijani, R. B. (2022). Factors that drive actual purchasing of groceries through e-commerce platforms during COVID-19 in Indonesia. *Sustainability*, 24, 3235.

Weltevreden, J. & van Rietbergen, T. (2007). E-shopping versus city centre shopping: the role of perceived city centre attractiveness. *Tijdschrift voor Economische en Sociale Geografie*, *98*(1), 68-85.

Weltevreden, J., Atzema, O., Frenken, K., de Kruif, K., & van Oort, F. (2008). The geography of Internet adoption by independent retailers in the Netherlands. *Environment and Planning B, 35*, 443-460.

Wieland, T. (2021a). Spatial shopping behavior in a multi-channel environment: A discrete choice model approach. *Region*, *8*(2), 1-27.

Wieland, T. (2021b). Identifying the determinants of store choice in a multi-channel environment: A hurdle model approach. *Papers in Applied Geography*, *7*(4), 343-371.

Wieland, T. (2022). Pandemic shopping behavior: did voluntary behavioral changes during the COVID-19 pandemic increase the competition between online retailers and physical retail locations? *Papers in Applied Geography,* DOI:10.1080/23754931.2022.2106581.

World Population Review. (2024). *Average household size by state 2024*. World Population Review. https://worldpopulationreview.com/state-rankings/average-household-size-by-state

Wrigley, N. & Currah, A. (2006). Globalizing retail and the 'new e-conomy': The organizational challenge of e-commerce for the retail TNCs. *Geoforum, 37*(3), 340-351.

Wrigley, N. & Lowe, M. (2002). *Reading retail: A geographical perspective on retailing and consumption spaces.* Oxford University Press, Inc.

Wrigley, N., Lowe, M., & Currah, A. (2002). Retailing and E-tailing. *Urban Geography,* *23*(2), 180-197.

Zhang, M. & Ghosh, D. (2016). Spatial supermarket redlining and neighborhood vulnerability: A case study of Hartford, Connecticut. *Transactions in GIS*, *20*(1), 79-100.

Zhang, Y., Chen, J., & Zhang, S. (2017). Analysis of spatial pattern and influencing factors of e-commerce. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XLII-2/W7*, 1461-1467.

Zhen, F., Du, X., Cao, J., & Mokhtarian, P. L. (2018). The association between spatial attributes and e-shopping in the shopping process for search goods and experience goods: Evidence from Nanjing. *Journal of Transport Geography*, *66*, 291-299.

Zheng, Z., Morimoto, T., & Murayama, Y. (2020). Optimal location analysis of delivery parcel-pickup points using AHP and network Huff model: A case study of Shiweitang sub-district in Guangzhou City, China. *International Journal of Geo-Information*, *9*, 193.

Zook, M. A. (2000). The web of production: The economic geography of commercial Internet content production in the United States. *Environment and Planning A, 32*(3), 411-426.

Zook, M. A. (2001). Old hierarchies or new networks of centrality? The global geography of the Internet content market. *American Behavioral Scientist, 44*(10), 1679-1696.

Zook, M. A. (2002). Hubs, nodes, and bypassed places: A typology of E-commerce regions in the United States. *Journal of Economic and Human Geography*, *93*(5), 509-521.

Vita

Connor Reed was born in 1994 and grew up in Muncie, Indiana. He received his PhD in Geography from the University of Tennessee in 2024. He holds a Bachelor of Science (2017) in economic consulting from the Kelley School of Business at Indiana University and a Master of Science (2019) in applied geography from the University of North Texas. Reed has previously worked in location planning and site selection for *BJ's Wholesale Clubs*, *H-E-B* grocers, *Brixmor Property Group*, and other real estate research and consulting firms.