Essays on the Digital Transformation of Retail Grocery Industry

Zhou Yang

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2023

Reading Committee:
Yuya Takahashi, Chair
Dong-Jae Eun
Simha Mummalaneni
Jing Tao

Program Authorized to Offer Degree:
Economics

University of Washington

Abstract

Essays on the Digital Transformation of Retail Grocery Industry

Zhou Yang

Chair of the Supervisory Committee:
Yuya Takahashi
Department of Economics

This dissertation looks at the profound impact of digital transformation in the grocery retail industry. It specifically focuses on the competitive repercussions of Amazon's foray into the grocery market, the effects of its Fulfillment Centers (FCs) on local labor markets, and the integration of image and text information from digital sellers into demand estimation. Spanning three interconnected chapters, the research presents a comprehensive examination of the ongoing shifts revolutionizing the retail landscape.

Chapter 1 probes the entry of Amazon Fresh and its ensuing effects on local grocery stores. An intricate analysis of quarterly grocery scanner data, combined with Amazon Fresh's entry information, reveals intriguing dynamics of retail competition. Interestingly, Amazon Fresh's debut prompts a significant negative volume response and a slight positive price response from incumbent grocery stores, specifically within the juice products category. The competition extends beyond mere volume and price adjustments, revealing that incumbent stores expand their product assortments and enhance service quality to differentiate from digital retailers. This nuanced understanding of the competitive entry effect sheds light on the strategic responses of traditional brick-and-mortar stores to digital competition in a rapidly evolving marketplace.

Chapter 2 broadens the scope to the county level, examining the impacts of Amazon FCs on local labor markets. Employing robust econometric methods, this chapter unveils that

while FCs stimulate higher wages, they also seemingly contribute to a reduction in the overall retail workforce. Additionally, the analysis notes a decrease in juice sales volume and inconsistent price changes across the county. These findings hint at a complex interplay between labor market changes, retail market adjustments, and the ongoing digital transformation.

Chapter 3 offers a comprehensive exploration of the integration of deep learning-derived features into traditional demand estimation methodologies. A comparison of econometric models with machine learning models using juice product data illuminates the efficacy of different techniques in capturing diverse product characteristics. The study further explores advanced methods like double machine learning and convolutional neural networks to address challenges inherent in high-dimensional and sparse data. Notably, we successfully incorporate deep learning into traditional demand estimation, demonstrating a substantial improvement in model performance, particularly in neural networks. This work not only enhances the representation of product features but also bolsters the models' capacity to accurately capture underlying demand dynamics.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to the chair of my supervisory committee, Professor Yuya Takahashi. His steadfast support, insightful guidance, and unwavering patience throughout my Ph.D. journey have not only inspired me but also steered me into the realm of Industrial Organization. His invaluable insights and resourceful suggestions have been instrumental in shaping and refining this dissertation.

My great thanks extend to my supervisory committee members: Professor Dong-Jae Eun, Professor Jing Tao, and Professor Simha Mummalaneni. I am particularly grateful to Professor Eun for his remarkable courses in Industrial Organization, which significantly broadened my research vision and provided me with an indispensable toolkit. His constructive advice helped clarify my research goals from the very beginning. My sincere thanks to Professor Tao for introducing me to the crucial subject of Econometrics and for his consistent kindness and support during my time at the University of Washington. I am indebted to Professor Simha Mummalaneni, my GSR, for his insightful and productive discussions on my research topics. His prompt responses to my research inquiries, as well as his recommendations of pertinent articles and courses, have been invaluable. Notably, it was my research experience with him that fine-tuned my exploration of digital retail markets. Without their concerted efforts and expertise, the completion of this dissertation would not have been feasible.

I am also grateful to numerous professors at the University of Washington who have enriched my academic journey. These include Professors Jacques Lawarrée, Fahad Khalil, Alan Griffith, Xu Tan, Yanqin Fan, Quan Wen, Eric Zivot, Thomas Richardson, Yu-chin Chen, and Stephen Turnovsky, among others. I also wish to acknowledge my peers, namely Lia, Yue, Yuming, Yida, Zihao, Zhe, Sajid, Jackson, Minyan, and Kwong-Yu, for their collaborative spirit and mutual enhancement of our research endeavors.

Lastly, my profound gratitude goes to my parents and grandparents. Their unconditional love has provided me the freedom to chart my own course and become the person I aspire to be. Lastly, but by no means least, my heartfelt appreciation is extended to my fiancée Xiao, who has remained a constant source of love, support, and patience throughout this journey. Her unwavering faith in me, even during the most challenging times, has been a wellspring of strength and motivation. Her companionship and understanding have rendered this journey not only feasible but also profoundly rewarding. I am profoundly grateful for her unwavering support and for sharing every step of this incredible journey with me.

DEDICATION

to people who raise me up,

who get me cry,

who make me strong,

I would not be me without you

WO AI NI MEN.

Chapter 1

THE COMPETITIVE EFFECT FROM DIGITAL TRANSFORMATION IN GROCERY INDUSTRY: THE ENTRY OF AMAZON FRESH

1.1 Introduction

The conventional grocery retailing industry in the United States is significantly influenced by several demographic variables, including population density, lifestyle, local income, and most importantly, local competition. The phenomenon of suburbanization and increased mobility since the 1960s has driven the development of economies of scale. This progress has consequently promoted the dominance of larger stores over smaller ones and initiated the emergence of innovative distribution structures centered around mass retailing(Holmes, 2011). Through their broad scope of operations and the sale of cost-effective store-brand products, mass merchandisers can provide products at lower prices. Walmart's distribution network exemplifies the potential advantages of scaling up operations in this concentrated industry[1].

Over recent decades, the digital transformation has significantly disrupted the U.S. food and grocery industry. The proliferation of the internet, wireless technologies, and mobile devices has spurred the growth of online grocery shopping, offering consumers a convenient alternative to traditional in-store purchases. Technology companies, rather than traditional supermarkets or food manufacturers, were the early adopters of this trend. These companies

[1] As of 2018, Walmart, a leading mass merchandiser in the United States, has achieved a market share of 50 percent or more of grocery sales in 43 metropolitan areas and 160 smaller markets. This dominance can be attributed to the company's strategy of leveraging its massive distribution and supply chain infrastructure to offer competitive pricing, particularly for store-brand products. Walmart has been able to use its economies of scale to drive down costs and improve efficiency, which has enabled it to expand rapidly and gain a significant competitive advantage in many regions of the country.

utilized their expertise in coding and systems to revolutionize the sector. Amazon.com, for instance, launched its Amazon Fresh service to penetrate the online grocery market[2].

The digital transformation of the grocery industry presents numerous benefits. For consumers, it provides a convenient platform to purchase a diverse range of products and services from multiple vendors, often at competitive prices. Search engines and price comparison sites significantly reduce search costs, enabling consumers to easily find and compare various offers for the same product. Electronic markets offer 24/7 shopping from any location, circumventing limitations of physical store hours, distance, and product availability. From a business perspective, digital technologies can enhance product differentiation and mitigate price competition. However, despite the potential advantages, concerns remain about the digital transformation's impacts on employment, the environment, and traditional retailers. For example, critics are concerned that the rise of online grocery shopping might lead to job losses in the traditional retail sector. This paper seeks to understand how such competition influences the market behaviors of traditional grocery stores.

This paper aims to delve into the competitive impacts of digital transformation on incumbent grocery stores, with a particular focus on the revenue implications for juice products. The research offers an all-encompassing understanding of competition, incorporating facets such as sales volume, pricing, assortment optimization, and quality modifications. To quantify the responses of physical stores to the incursion of online grocery platforms, I utilize three distinct specifications: a static two-way fixed effect model, dynamic TWFE, and decomposition methods. These are coupled with recent advancements in econometrics that address potential bias arising from the staggered design of the treatment. In this context, the expansion of Amazon's distribution network serves as a case study for investigating potential bias and heterogeneity in the treatment effects.

[2] Amazon Fresh is an online supermarket that allows consumers to purchase a wide range of grocery and household items and have them delivered directly to their homes. As part of Amazon's broader strategy to expand its retail offerings, Amazon Fresh was launched in 2007 as a way to enter the competitive grocery market. The service has since grown to include a range of grocery items at competitive prices, offering consumers the convenience of online shopping and home delivery.

The findings of the research underscore that digital platforms indeed cannibalize the market share of incumbent grocery stores. However, the reactions of these stores are complex and manifold. A reduced variety of juice products are sold in incumbent grocery stores following the entry of Amazon Fresh into the market, and this impact intensifies over time. Competition also manifests in other dimensions, such as adjustments in pack sizes or product assortments by stores. To account for this, I have expanded the empirical approach to sub-module level data, enabling the examination of price variations for similar products in the same store pre and post-exposure. The stores' response is seen in their differentiation from online services by offering a broader selection of products and items at higher prices (which imply superior quality due to either higher input costs or service markup) in physical stores.

Furthermore, I explore price dispersion of similar products at the onset of competition to determine whether digital transformation equips consumers with enhanced access to price information, subsequently leading to heightened price uniformity, as posited by DellaVigna and Gentzkow, 2019. Employing the procedure utilized by Kaplan et al., 2019, I decompose prices and eliminate trends to attain dispersion devoid of seasonality. The results suggest that digital exposure narrows the price dispersion of incumbent stores, although this needs to be verified at a more granular level[3].

This research contributes to several strands of literature. The study conducted by Basker and Noel, 2009 investigates the competitive entry of Walmart in the context of their physical locations. However, the advent of online competition broadens the playing field as it typically manifests at the zip code or county level. This indicates that the degree of competition is not solely determined by the proximity of stores to distribution centers, thus necessitating the elimination of several confounding factors. Additionally, this research integrates recent advancements in correcting bias from staggered design (Goodman-Bacon, 2021, Callaway and SantAnna, 2021, Sun and Abraham, 2021, Athey and Imbens, 2022, de Chaisemartin and D'Haultfuille, 2020), allowing for heterogeneous treatment effects grouped by entry

[3]The data is aggregated at the sub-module level.

timing. These methodological progressions offer valuable avenues for considering the diverse treatment effects noted at different market strata. Notably, this study also contributes to the expanding literature concerning the behavioral aspects of firm decision-making and their implications for market efficiency, as evidenced by DellaVigna and Gentzkow, 2019 and Hitsch et al., 2019. Both papers highlight that the variation in prices within retail chains across markets is significantly less pronounced than what local market conditions would typically suggest.[4]

The remainder of the paper is structured as follows. Section 2 offers a literature review of recent advances in two way fixed effect models. Section 3 delves into the operational aspects of Amazon Fresh and outlines the data sources harnessed to compile the necessary datasets for estimation. Section 4 introduces the model framework employed in the estimation. Section 5 presents the estimation results from various model specifications and offers explanations. Section 6 addresses the alternative issue of price uniformity resulting from digital competition. Section 7 explores store-level heterogeneity analysis and provides potential explanations. Finally, Section 8 concludes the paper.

1.2 Recent Advances in Methodology

The difference-in-difference (DiD) estimator has been widely utilized in various empirical research fields, including economics, marketing, political science, and public health. The core concept of DiD is to estimate the counterfactual outcomes for treated units by requiring several key assumptions that enable researchers to impute the average untreated outcomes for the treated group. To achieve this, following assumptions must be made.

Firstly, the parallel trend assumption suggests that, in the absence of treatment, the average outcomes for both treated and untreated groups would follow parallel trends. This assumption helps isolate the treatment effect from any pre-existing trends in the data. Sec-

[4] These studies emphasize that, despite differences in local market conditions, price uniformity within retail chains across various markets is more prevalent than expected, indicating firm-level decision-making significantly affects market prices.

ondly, the no anticipation effect assumption ensures that the treatment does not have any causal influence before its actual implementation. This constraint prevents any behavioral changes in the treated population that might arise from the anticipation of treatment, ensuring that the treated group's actions remain consistent with the untreated group until the treatment commences. Lastly, the independent intervention assumption posits that the treatment is unrelated to the outcome in order to guarantee asymptotic consistency. This condition helps to eliminate the possibility of biased estimates arising from a correlation between the treatment and the outcome variable. By satisfying these key assumptions, the difference-in-difference estimator can provide a reliable estimate of the causal effect of an intervention on the treated population in various research contexts.

Numerous recent advancements have emerged in the field of DiD estimators. To better comprehend how the impact of Amazon Fresh's entry can be assessed, I present the staggered design approach. This methodology enables the estimation of heterogeneous treatment effects by modifying the underlying parallel trends assumption to accommodate variations. A growing body of research has concentrated on easing the timing constraint while maintaining the fundamental structure of the stylized model, which includes parallel trends, no anticipation, and independent sampling.

In this context, the staggered design approach offers a valuable perspective for investigating the consequences of Amazon Fresh's entry into the market. By allowing for the estimation of heterogeneous treatment effects, this method acknowledges the potential differences in impact across various market segments or time periods. As researchers continue to refine and adapt the key assumptions in DiD estimators, our understanding of their implications for real-world scenarios, such as the case of Amazon Fresh, will only grow stronger.

Angrist, 2009 asserts that generalized regression DiD model has a good property of "easy to add states or periods to regression setup" and it is easy to add additional covariates. The application on staggered design offers desirable properties over simple 2×2 DiD. Multiple treatment periods introduced in the staggered design can alleviate concerns that the observed treatment effects are caused by some contemporaneous trends instead of parallel trend assumption that simple DiD requires but is hard to prove the existence.

The staggered Difference-in-Difference (DiD) approach, particularly the Two-Way Fixed Effects (TWFE) model, can sometimes yield biased results when applied to settings with more than two time periods and more than two units. Recent literature has demonstrated that the estimated coefficient (β in Equation 1.5) is, in fact, a weighted average of numerous distinct treatment effects. Goodman-Bacon (2021) Goodman-Bacon, 2021 illustrates that in cases where the treatment is binary and the design is staggered, the coefficient can be expressed as:

$$\beta = \sum_{g \neq g', t < t'} v_{g,g',t,t'} \beta^{DID}_{g,g',t,t'} \tag{1.1}$$

where $\beta^{DID}_{g,g',t,t'}$ represents a DiD estimator comparing the evolution of outcomes between two groups, g and g', from a pre-treatment period t to a post-treatment period t'. The non-negative weights $v_{g,g',t,t'}$ sum to one, with $v_{g,g',t,t'} > 0$. Goodman-Bacon Goodman-Bacon, 2021 demonstrates that β can be decomposed as a weighted average of DiDs between cohorts of groups transitioning to treatment (grouped by treatment time) and between periods where their treatment status remains constant. This decomposition can be further expressed in the form of Equation 1.1.

Equation 1.1 reveals that the staggered DiD TWFE approach is essentially a weighted average of all possible two-group/two-period DiD estimators present in the data. The bias emerging from traditional models arises when early treated units are compared to later treated units in the presence of treatment effect heterogeneity.[5]

This understanding of the staggered DiD TWFE approach highlights its limitations and the potential biases that may emerge. By recognizing these issues, researchers can account for these biases in their analyses or consider alternative methods to more accurately estimate treatment effects when heterogeneity is present. This finding highlights the complexities and potential pitfalls associated with the staggered DiD approach, especially when using the

[5] A graphical explanation of the Bacon decomposition theorem is provided in the appendix.

TWFE model. By understanding the nuances of this methodology, researchers can better interpret the results and, if necessary, explore alternative approaches to more accurately estimate treatment effects in their specific context.

Similar to Goodman-Bacon, 2021, Callaway and SantAnna, 2021 defines the ATT for different groups g by treatment timing t (they call this as the "group-time average treatment effect")

$$ATT_g(t) = \mathbb{E}[Y_{i,t}(1) - Y_{i,t}(0)|g] \tag{1.2}$$

$ATT_g(t)$ is the expected difference between the observed outcome variable for treated units at time t and the outcome had the firms not received treatment. This generalized formulation easily allows for heterogeneity in the ATT across groups g or over time t.

In their work, Imai and Kim, 2021 explore scenarios where units receive treatment at different times, without necessarily mandating that treatment be an absorbing state. Their estimators intuitively compare changes in outcomes for units with altered treatment status to those units whose treatment status remains constant over the same periods. This approach yields an interpretable causal effect under generalizations of the parallel trends assumption and an additional "no carryover" assumption. The latter imposes that potential outcomes depend solely on the current treatment status, rather than the entire treatment history.[6]

However, as noted by Bojinov et al., 2021, the no carryover assumption might be restrictive in various settings. For instance, if the treatment involves raising the minimum wage and the outcome is employment, the no carryover assumption requires that employment in period t depends only on whether the minimum wage was raised in period t and not on the history of minimum wage changes. Other studies have examined DiD settings with non-binary treatments. de Chaisemartin and D'Haultfuille, 2020 investigate "fuzzy" DiD settings where all groups receive treatment in both time periods, but the proportion of units exposed

[6]Though there is currently no consensus on this issue, ongoing research in this area aims to either correct the weight issues, as in Goodman-Bacon's approach, or select appropriate control groups, or a combination of the two.

to treatment increases in one group and remains constant in the other.

This chapter seeks to offer a comprehensive understanding of the challenges and potential solutions related to staggered DiD estimators in estimating marketing behavior, particularly within the context of online grocery shopping, by highlighting key contributions in the literature. Recognizing the ongoing debates and proposed alternatives is essential for researchers to select the most suitable method for their specific research questions and settings. This paper primarily concentrates on the traditional TWFE regression and the approaches suggested by Goodman-Bacon, 2021 and Callaway and SantAnna, 2021. While several alternative methods have been proposed to tackle biases associated with the staggered DiD design, they are not examined in this paper. Notable examples include the works of Sun and Abraham, 2021, Athey and Imbens, 2022, Baker et al., 2022, Barrios, 2021, and de Chaisemartin and D'Haultfuille, 2020, which present general overviews of these methods across various settings.

1.3 Industry Background

The online grocery industry has witnessed rapid growth as consumers increasingly value the convenience of home delivery for their groceries. In response, numerous companies now offer delivery services to households in metropolitan areas. One of the early entrants in this market was Amazon Fresh, which started operations in Seattle in 2007 and later expanded to Los Angeles, San Francisco, and New York. Around the same period, FreshDirect also explored expansion opportunities beyond the New York City metropolitan area, with plans to enter additional East Coast markets like Boston, Philadelphia, and Washington D.C. In 2017, Walmart announced its intention to open over 1,000 online grocery pickup locations across the U.S. In this paper, the focus is specifically on the entry decision made by Amazon Fresh, one of the first companies to significantly disrupt the traditional grocery retailing industry with its online grocery delivery service. To eliminate confounding effects from the entries of similar services, the study targets markets where Amazon Fresh was the sole provider of online grocery services. According to a 2020 report by Moodys, the US online food retail business was estimated to be roughly \$20 billion[7].

Amazon Fresh is an online grocery delivery service established in 2007 that has experienced consistent growth since its inception. Unlike Walmart, which benefits from its well-established distribution network, Amazon Fresh utilizes Amazon's existing distribution network to deliver groceries to customers in select metropolitan areas. The markets where AF operates typically exhibit two characteristics: high levels of discretionary income, as observed in the West Coast and New York, and a significant proportion of elderly residents, as seen in Florida. Amazon Fresh relies on a large portion of cold storage facilities, unlike a regular Amazon distribution center, to store grocery items. The last-mile clusters comprise delivery stations, Prime hubs, and Amazon Fresh hubs, which are small in size but high in throughput. These facilities are uniformly distributed to maximize accessibility for last-mile deliveries in medium to high-density consumption markets. With this logistical infrastructure, Amazon Fresh can deliver groceries swiftly and compete with traditional grocery retailers on cost savings. The COVID-19 pandemic led to a surge in demand for online grocery services, as consumers increasingly turned to online shopping for essential items. However, even prior to the pandemic, the digital transformation of the grocery industry was already gaining momentum worldwide.

Amazon's grocery distribution network is a crucial component of their grocery delivery service, Amazon Fresh, which operates both ambient and cold storage centers across the United States. In 2017, Amazon acquired Whole Foods Market, and with it, ownership of their existing retail grocery distribution network, primarily focused on perishable merchandise for retail stores in major markets. This acquisition has intensified local competition not only between online and offline retailers but also between Amazon's own services, including their Prime Now service, which offers similar grocery delivery services. To avoid potential confounding factors introduced by the acquisition of Whole Foods, this study concentrates on the period from 2010 to 2016, during which Amazon Fresh was established through the

[7]https://superfood.digital/online-grocery-store-ecommerce-statistics/

opening of its Fresh distribution centers and before the acquisition of Whole Foods. This enables a clearer analysis of the competitive effects of Amazon Fresh on incumbent grocery stores.

Traditional grocery stores are now competing with online grocery stores by recognizing the value of the connection between online and in-store grocery shopping. They are using this connection to win more sales and compete in todays integrated retail environment[8]. One good example is Walmart. Before 2016, Walmart was already working on online grocery pickup since 2014 that was similar to what Amazon Fresh offered in Seattle. Walmart expanded the service to more than 600 stores and was hoping to quickly expand its automated system if all went well[9].

Utilizing a staggered design in the case of Amazon Fresh is justifiable given its incremental market expansion, which enables researchers to evaluate the causal impact of its market entry on existing grocery stores while considering varying treatment effects. This method adjusts for timing differences, enhances identification by using both cross-sectional and time-series data variations, and provides greater resilience against potential breaches of the parallel trends assumptiona critical aspect when examining complex and evolving markets such as the online grocery sector. Ultimately, the staggered design delivers a more precise and thorough analysis of Amazon Fresh's competitive influence on conventional grocery stores.

1.4 Data

The data for this study are gathered from three main sources: Amazon Fresh Distribution Center data, Nielsen Scanner data, and Consumer Expenditure Surveys. The Amazon Fresh Distribution Center data provides information on the locations and operations of Amazon's grocery distribution centers, including their capacity and proximity to customers. The Nielsen Scanner data covers sales data for grocery products, including brand, price, and quan-

[8] https://www.brickmeetsclick.com/competing-with-online-grocery-retailers–five-ways-grocers-can-win-more-sales

[9] https://www.theverge.com/2017/6/6/15742746/walmart-automated-grocery-pickup-amazonfresh

tity sold of numerous physical grocery stores, across various retail channels. The Consumer Expenditure Surveys offer insights into household spending on groceries, demographics. Collectively, these data sources provide a comprehensive view of the online grocery market in the United States, as well as the competitive landscape and consumer behavior within retailing industry.

Amazon Fresh Distribution Center Data

This study utilizes data from MWPVL[10], which includes information on Amazon's global fulfillment center network, such as the address, size, type, and opening and closing dates of each facility. The research defines the entry of Amazon Fresh into a Designated Market Area (DMA) as the inauguration of the first local Fresh facility within that region. Due to tax policies (Newberry et al., Forthcoming) or serving multiple cities simultaneously[11], some Fresh facilities are located in neighboring states. The study identifies the DMAs of Fresh facilities using the facility code assigned by Amazon. To gauge the market penetration of Amazon Fresh in each DMA, the study calculates the distance from each county's population center[12] to the closest Fresh facility and establishes four distance bands. The Google Maps API is employed to acquire the longitude and latitude of facilities and counties, and a map of Amazon Fresh facilities in 2020 is presented in Figure 1.[13].

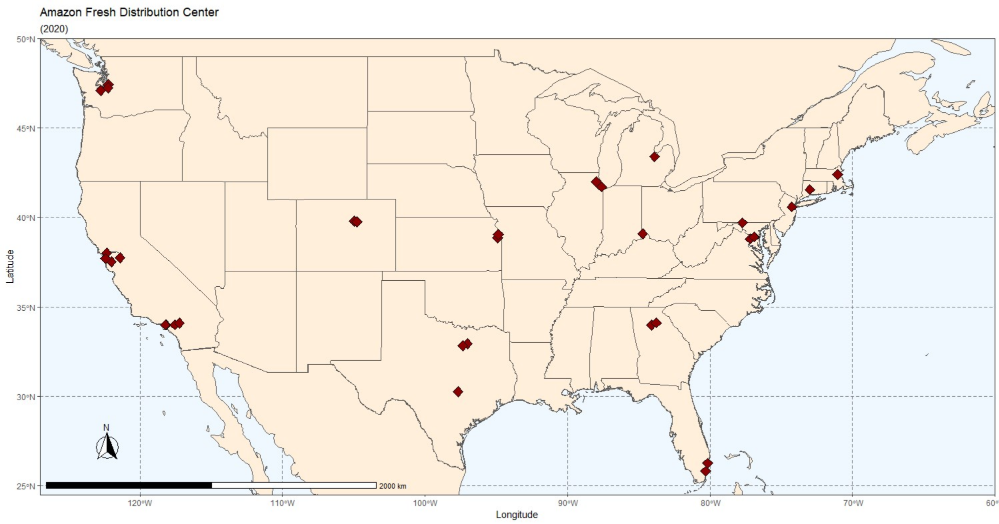
[10]MWPVL is a supply chain consulting company that offers a comprehensive summary of Amazon's Fulfillment center and distribution network. For more information, visit $https://mwpvl.com/html/amazon_com.html$.

[11]Amazon Fresh facilities utilize advanced technology and logistics to optimize the supply chain, minimize delivery times, and enhance the customer experience. They can serve multiple cities at once, particularly on the east coast where metropolitan areas are closely situated.

[12]Population center data is obtained from the U.S. Census: $https://www.census.gov/geographies/reference-files/time-series/geo/centers-population.html$

[13]The complete list of cities and the timing of Amazon Fresh's entry is provided in the Appendix table A.2

Figure 1.1: 2020 Amazon Fresh Distribution Center



Census Data on Consumer Demographics and Population Center

The Consumer Expenditure Surveys (CE) is a comprehensive and nationally representative survey that aims to collect detailed information on household spending, which is then used to develop estimates of the Consumer Price Index. The Consumer Expenditure Surveys (CE) is a data collection effort conducted by the Census Bureau for the Bureau of Labor Statistics (BLS) through two separate surveys: the Interview Survey and the Diary Survey. The Interview Survey collects information on major and recurring expenses, while the Diary Survey focuses on more minor or frequently purchased items. The main purpose of collecting CE data is to revise the relative importance of goods and services in the market basket of the Consumer Price Index, which is a key measure of inflation in the United States. In addition to spending data, the CE survey also collects key demographic and socioeconomic information from the households surveyed. For this study, we used this information to create covariates at the county level.

To calculate the distance between counties and their nearest Fresh distribution center, I utilized population center data collected by the U.S. Census Bureau. The population center data contains geographic information for each county, which enabled us to estimate the distance to the closest Fresh distribution center for each county. The concept of the center of population, as used by the U.S. Census Bureau, refers to a balance point. It represents the point at which an imaginary, weightless, rigid, and flat surface representation of the 50 states[14] and the District of Columbia would balance if weights of identical size were placed on it so that each weight represented the location of one person. This information was crucial for our study, as it allowed us to estimate the distance between counties and Fresh distribution centers, which is an essential variable in our analysis.

Nielsen Scanner Data

The data used in this study is sourced from the Nielsen Retail Scanner data (also known as RMS) provided by the Kilts Center for Marketing at the University of Chicago Booth School of Business. The data set contains more than 2.6 million products identified by 12-digit universal product codes (UPCs). The data is collected from over 40,000 stores across 370 metropolitan statistical areas (MSAs) in the United States, representing approximately 53% of all sales in grocery stores, 55% in drug stores, 32% in mass merchandisers, and 2% in convenience stores. Nielsen categorizes the products into 10 departments, over 100 product groups, and over 1,000 product modules. In this study, we focus on data from the juice product group.

The RMS dataset contains over 100 billion unique observations at the week-store-UPC level. To make the data comparable, I aggregated the weekly data to the quarterly frequency at store sub-module level. sub-modules are defined as subsets of modules with similar package size since they are comparable but not as granular as UPCs. I normalized the prices of product groups to the same units, such as price per ounce for juice. Additionally, we used the size and packaging information provided by Nielsen to calculate the volume of sales for juice products.[15] The store characteristics are obtained from the 2012 to 2016 Nielsen Annual

[14] or 48 conterminous states for calculations made prior to 1960

[15] Since the volume of data is huge, I ignore the upc level difference.

data appendix. These data include corporate parent owner of the store, retailers, store types (Food, Mass Merchant etc.), county code, designated market area. I further calculate chain size of stores and Amazon Fresh entry time.[16]

1.4.1 Store Level scanner data

I begin with 9,056 stores across all markets after excluding those in areas potentially influenced by similar services like Instacart, Peapod, and FreshDirect[17]. To create a balanced panel data, I apply the following criteria for store eligibility: stores must be in operation at both the beginning and end of our study period (2012 Q1 and 2016 Q4) and must have sales records for all 20 quarters during our investigation. Additionally, I exclude stores with insufficient juice sales throughout the study period to avoid issues in calculating prices and indexes necessary for analysis. I also aim to prevent the impact of store exits and entries on our study, as Nielsen's data does not clearly indicate whether missing information is due to no purchases, store closures, or exits from the grocery market. Following these selection criteria, 8,380 stores are included in the study.

Table 1.1 presents the descriptive statistics for various store-sub-module level variables in a quarter. These variables consist of total revenue, total volume, price dispersion, number of visits, number of brands, and number of products. Total revenue refers to the earnings generated by selling juice products, while total volume denotes the quantity of juice products sold in ounces. Price dispersion represents the standard deviation of prices within a store, reflecting the range of product prices offered by the store. The number of visits tallies the customers who purchased juice products, as captured by Nielsen scanner data. The number of brands and products count the unique juice brands and distinct products sold by a store

[16]As for the fresh produce products, there is no consistent unit for items in the fresh produce product group, so I calculate total revenue from scanner data at store level, then aggregate by FIPS code to obtain county level revenue data.

[17]Instacart initially launched its service in San Francisco in 2012. Peapod was founded in Evanston, Illinois in 1989, and later expanded its services to the Chicago metropolitan area. FreshDirect began its operations in New York City in 2002.

Table 1.1: Descriptive Statistics By Package Size

Pkg Size	Variable	Mean	SD	Min	Q1	Median	Q3	Max
Party	Total Volume (kOz)	77.5	234.7	0.2	5.1	17.9	55.9	5284.6
	Price (\$/kOz)	36.2	18.8	12.9	20.6	34.8	44.1	101.5
	Average Volume	906.1	1281.4	101.4	268.8	489.9	957.3	11246.7
	Price Dispersion (\$/kOz)	5.0	5.1	0.0	1.3	3.3	7.3	75.5
	Num Visits	57.8	70.5	2.0	13.0	26.0	75.0	505.0
	Num Brands	2.6	1.9	1.0	1.0	2.0	3.0	13.0
	Num Products	5.7	6.4	1.0	2.0	3.0	7.0	45.0
Large	Total Volume (kOz)	66.5	152.0	0.1	2.2	8.6	52.1	4011.7
	Price (\$/kOz)	53.1	13.6	25.9	42.5	52.6	61.2	100.9
	Average Volume	325.1	329.7	33.6	114.6	204.0	403.3	2386.6
	Price Dispersion (\$/kOz)	12.1	8.8	0.0	6.6	10.6	15.5	135.2
	Num Visits	125.4	188.8	2.0	15.0	41.0	121.0	1844.0
	Num Brands	4.9	5.4	1.0	2.0	3.0	6.0	54.0
	Num Products	12.6	17.4	1.0	2.0	5.0	13.0	152.0
Medium	Total Volume (kOz)	28.5	99.7	0.0	0.5	2.1	11.2	1419.7
	Price (\$/kOz)	105.3	48.0	28.6	64.4	104.0	131.8	254.9
	Average Volume	110.5	109.3	12.5	43.8	72.5	131.0	768.3
	Price Dispersion (\$/kOz)	28.8	27.5	0.0	7.4	22.8	41.7	352.8
	Num Visits	124.9	256.6	2.0	11.0	26.0	76.0	2120.0
	Num Brands	4.8	6.4	1.0	1.0	2.0	5.0	45.0
	Num Products	13.5	26.1	1.0	1.0	3.0	8.0	189.0
Small	Total Volume (kOz)	4.2	11.5	0.0	0.2	0.8	3.3	270.2
	Price (\$/kOz)	320.0	163.4	38.3	209.1	297.6	401.7	1011.8
	Average Volume	38.9	33.6	2.5	17.6	28.6	47.8	258.9
	Price Dispersion (\$/kOz)	66.3	57.8	0.0	8.7	59.9	116.3	778.5
	Num Visits	79.2	149.2	2.0	12.0	25.0	74.0	1348.0
	Num Brands	4.0	5.1	1.0	1.0	2.0	4.0	42.0
	Num Products	8.9	15.6	1.0	1.0	3.0	8.0	122.0

in a quarter, respectively.[18]

The table highlights various disparities among different package sizes. Party-sized products have an average sales volume of 77.5 kOz per store, with a price of 36.2 per kOz. Large-sized products, on average, sell 66.5 kOz with an average price of 53.1 per kOz. Medium-sized products have a mean volume of 28.5 kOz and a mean price of 105.3 per kOz, while small-sized products sell only 4.2 kOz per store at the highest average price of 320 per kOz[19]. The average number of brands and UPCs per store is highest for medium (4.8 brands and 13.5 UPCs) and large packaged products (4.9 brands and 12.6 UPCs), indicating which sizes are most competitive. Price dispersion is highest for small-sized products, as their prices heavily depend on product quality, including factors such as ingredients, packaging, etc. This data provides a comprehensive foundation for understanding the areas where competition may occur.

1.5 Empirical Model

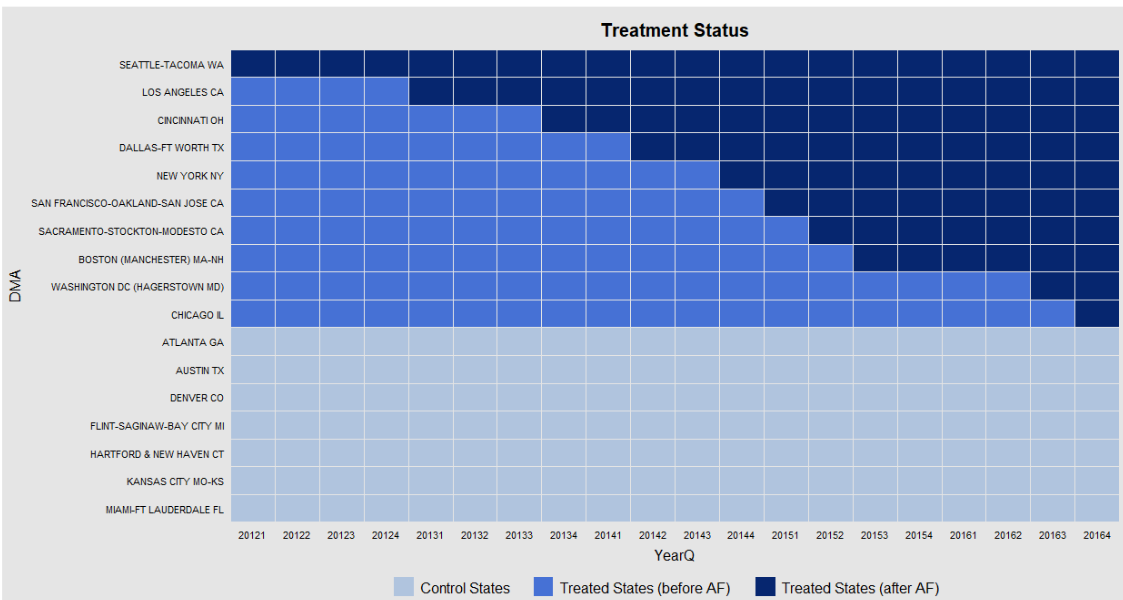
This study investigates the competitive impact of Amazon Fresh's (AF) entry on incumbent grocery stores. Three primary challenges arise in this analysis. The first challenge concerns the limited availability of store-level data, potentially leading to endogeneity due to confounding factors that cannot account for. The second challenge relates to the varying treatment timing across different markets, as AF entered some markets earlier than others, potentially resulting in heterogeneous treatment effects that must be considered. For instance, early-entry markets such as Seattle and Los Angeles may have experienced a smaller impact compared to later-entry markets like Miami, attributable to customers being less acquainted with online grocery shopping and maintaining loyalty to incumbent stores. In

[18]Additionally, I calculate two more variables based on the aforementioned metrics: average volume per visit (total volume divided by the number of visits) and average unit price (total revenue divided by total volume). These variables assess the average consumption and price level of juice products at the store-sub-module level.

[19]Regarding package sizes, small products are under 20Oz per bottle; medium products are above 20 Oz and below 40 Oz; large products are above 40 Oz but less than 96 Oz; party-sized products are over 96 Oz.

contrast, later-entry markets may have faced a more significant impact, as customers were more cognizant of Amazon's reputation and more inclined to embrace online grocery shopping. In addition, the entry of Amazon Fresh could have indirect effects on incumbent stores through increased awareness of online grocery shopping, leading to a change in customer behavior and potentially benefiting other online grocery platforms as well (spillover effects).

Figure 1.2: Amazon Fresh Entry



1.5.1 Assumption

To address the above concerns, I make following assumptions from Goodman-Bacon, 2021 and Callaway and SantAnna, 2021.

Staggered Treatment Adoption Assumption: Staggered treatment adoption implies that once a unit participates in the treatment, they remain treated. This means that once a unit receives the treatment, it stays treated and does not unlearn its treatment experience. For example, this would apply to policies that are implemented in different locations over time. It would also apply to Amazon Freshs market entry. Entering a market usually involves a large cost and no observed exits based on the current data. I focus on this main case because I think it is hard to analyze non-staggered treatment setups within the DiD framework without imposing more restrictions on how the treatment effect varies across time, groups, treatment sequences, etc.

Parallel Trends Assumption based on never-treated units: For all $g = 2, ..., T$, $t = 2, ..., T$ with $t \geq g$

$$E[Y_t(0) - Y_{t-1}(0)|G = g] = E[Y_t(0) - Y_{t-1}(0)|D = 0] \tag{1.3}$$

This is a natural extension of the parallel trends assumption in the two periods and two groups case. This means that, without treatment, the average untreated potential outcomes for the group that received treatment first at time g and for the group that never received treatment would have been parallel in all periods after treatment $t \geq g$.

This parallel trend assumption depends on using the never treated units as a comparison group for all eventually treated units. This requires that (i) a (large enough) never-treated group exists in the data, and (ii) these units are similar enough to the eventually treated units so that they can serve as a valid comparison group. If these conditions are not met, we can use an alternative parallel trend assumption that uses the units that have not yet received treatment as valid comparison groups.

Parallel Trends Assumption based on not-yet treated units: For all $g = 2, ..., T$, $s, t = 2, ..., T$ with $t \geq g$ and $s \geq g$

$$E[Y_t(0) - Y_{t-1}(0)|G = g] = E[Y_t(0) - Y_{t-1}(0)|D_s = 0, G \neq g] \tag{1.4}$$

This assumption implies that we can use the units that have not received treatment by time s ($s \geq t$) as valid comparison groups when estimating the average treatment effect for the group that was treated first at time g. This assumption generally uses more data when constructing comparison groups. However, Callaway and SantAnna, 2021 note that this assumption also imposes some restrictions on some pre-treatment trends across different groups.

No anticipation: This assumption implies that if a unit is not treated in a particular period, their outcome should not be influenced by when they will receive treatment in the future. In other words, units are not expected to behave differently based on their knowledge of their future treatment date before treatment actually begins.

1.5.2 Static Two-Way Fixed Effect (TWFE) Model

To address the concerns outlined in Table 1.2[20], I begin by employing the static staggered design, which has been extensively used in previous literature. This staggered DiD design is a regression-based DiD method commonly applied to situations involving more than two units (treated vs control) and more than two periods (pre-treatment and post-treatment). The regression-based DiD also takes the following two-way fixed form:

$$Y_{it} = \alpha_i + \gamma_t + \beta AF_{it} + \varepsilon_{it} + \theta \tag{1.5}$$

where α_i and γ_t represent unit and time fixed effects. In this application, I estimate the TWFE model using store-level data, with time measured in quarters. The static specification provides a reasonable estimand when there is no heterogeneity in treatment effects across either time or units. β signifies the treatment effect of AF entry on the dependent variable Y_{it}, which is assumed to have a homogeneous treatment effect across different stores and time, irrespective of the duration since the treatment started; θ is the intercept, and ε_{it} is the error term. Each store in the data has a specific date for AF entry, and AF_{gt} is 0 before the entry and 1 afterward. The treatment occurs over time as Amazon opens new Fresh distribution centers, following a staggered design that allows groups to only switch into the treatment group but not out of it. I further modify the TWFE model to include covariates[21].

[20] The table demonstrates that AF enters different markets across the country at various times as part of their strategic expansion. Competitive effects from AF will be influenced not only by market-level characteristics but also by the different timings of entry.

[21] The covariates are at the county level. They include consumer units, vehicles per household, average number of persons in a household, distance bands it is to the closest distribution center, average household annual income, etc.

1.5.3 Dynamic TWFE Model

Apart from the TWFE approach, I also explore the dynamics of the treatment effect using an event study specification, incorporating a series of dummy variables for six months before and after the entry of AF. The following general regression is estimated:

$$Y_{i,t} = \alpha_i + \gamma_t + \beta_k^{-K}\ AF_{i,t}^{<-K} + \sum_{k=-K}^{-2} \beta_k^{lead} AF_{i,t}^{k} + \sum_{k=0}^{L} \beta_k^{lag} AF_{i,t}^{k} + \beta_k^{L+} AF_{i,t}^{>L} + \varepsilon_{i,t} \qquad (1.6)$$

In this equation, $AF_{i,t}^{k}$ represents the event study dummy variable for AF entry, taking a value of 1 if AF's entry is k months away from the initial entry at time t, and 0 otherwise. $AF_{i,t}^{<-K}$ and $AF_{i,t}^{>L}$ assume a value of 1 if the entry is before K or after L months of the initial entry. These represent leads and lags in the TWFE event study. β_k^{lag} can be interpreted as measurements of the average treatment effect for being exposed to treatment for k months, while β_k^{lead} estimates the trends before the treatment occurs. The leads and lags are set to 4 quarters, allowing for an examination of the trends 4 quarters before and after the entry of AF on dependent variables. The event study diagram in a staggered setting for Juice in the data is shown in Figure **??**. This graph indicates a negative trend in volume as AF's entry approaches.

1.5.4 Staggered DiD: Callaway and SantAnna, 2021

In light of the literature review, recent papers have addressed potential issues arising from the assumption of a homogeneous treatment effect in staggered designs. I proceed with another regression following Callaway and SantAnna, 2021, considering observations divided into G groups and T periods, indexed by group segments g and time period t. Let β_{gt} represent the coefficient of treatment AF_{gt}, the treatment in group g at period t. The estimating equation can be expressed as:

$$Y_{igt} = \alpha_g + \gamma_t + \beta_{gt} AF_{gt} + \varepsilon_{gt} \qquad (1.7)$$

Here, α_g denotes the group fixed effect, γ_t represents the time fixed effect, β is the treatment effect of AF entry on dependent variable Y_{igt}, and ϵ_{igt} is the error term. AF_{gt} is defined as in equation 1.5.

Intuitively, under the staggered versions of parallel trends and no anticipation assumptions, we can identify $ATT(g,t)$ by comparing the expected change in outcome for cohort g between periods g - 1 and t to that for a control group not-yet treated at period t. If we average over a set of comparisons G such that $g' > t$ for all $g \in G$, the average treatment effect on the treated would be:

$$ATT(g,t) = \mathbb{E}[Y_{i,t} - Y_{i,g-1}|G_i = g] - \mathbb{E}[Y_{i,t} - Y_{i,g-1}|G_i \in G] \tag{1.8}$$

We can estimate the ATT by replacing expectations with their sample analogs. However, when there are numerous treated periods and/or cohorts, reporting all $A\hat{T}T(g,t)$ might be unwieldy, and each estimate may lack precision. Fortunately, the method described above can be easily extended to estimate any weighted average of the $ATT(g,t)$. For example, we may be interested in an "event-study" parameter that gives the (weighted) average of the treatment effect l periods after adoption across different adoption cohorts, as in Callaway and SantAnna, 2021 and Goodman-Bacon, 2021 in 1.1[22].

1.6 Main Results

In this section, I present the findings from the model setups presented in the previous section, focusing on the competitive effect of Amazon Fresh's entry on various aspects of retail grocery industry. I begin by evaluating the impact of Amazon Fresh's entry on incumbent grocery stores' sales volume, volume-weighted price, product assortment, and price quality at the store level, using the approaches detailed earlier and a log specification. The analysis is centered on juice products, aggregated from Nielsen scanner data.

The regression's dependent variables include the logged average monthly store sales vol-

[22]Details of Bacon decomposition are in Appendix 1.

ume (in ounces), the logged volume-weighted price index (per ounce), the number of brands sold at stores, the number of products (UPCs) sold at stores, and the logged quality index. The findings aim to provide a comprehensive understanding of the competitive effects of Amazon Fresh's entry on incumbent grocery stores by examining various dimensions that characterize store competition, such as sales, prices, assortment, and product or service quality.

1.6.1 Volume Effect

To examine the impact of Amazon Fresh entry on grocery store performance, I first run a regression analysis of the logged quarterly store sales volume on Amazon Fresh entry under various model specifications. At the store level, the influence of Amazon Fresh entry on grocery stores' quarterly volume is negative. Results from the static model indicate that Amazon Fresh entry leads to a -2.1% decrease in the average volume of juice sold in stores. The dynamic specification results suggest a 2.0% decrease in sales volume due to competition.

When applying the method proposed by Callaway and SantAnna, 2021, grouping at entry time to allow for heterogeneous treatment effects, the average treatment effect is -0.8%. By grouping at time since treatment, which allows for heterogeneous treatment effects as time passes, the average treatment effect is -3.1%. This demonstrates that, when accounting for heterogeneous treatment effects, the impact of Amazon Fresh entry is less severe than with traditional estimation methods, suggesting potential bias in the latter. All estimates are significant and provide a good explanation for the observed variability.

Similar to the methods employed by Callaway and SantAnna, 2021 and Goodman-Bacon, 2021, I divide different treatment times into multiple 2×2 DiD designs and implement a series of diagnostic tests to evaluate the robustness of TWFE estimates in a staggered DiD setting[23]. I plot the 2×2 DiD against its weight and calculate the treatment effect for different types of comparisons: treated versus untreated, early treated versus later treated,

[23]The Bacon decomposition is conducted on county-level data due to model complexity

Table 1.2: Volume Effect

Model Spec.	OLS	Static	Static	Dynamic	CS (Time)	CS (Group)
Amazon Fresh	-0.5	-2.1*	-4.3***	-2.0*	-3.1*	-0.8*
SD	(1.3)	(1.0)	(1.0)	(0.8)	(1.3)	(0.8)
95% CI Lower Bound	-3.1	-4.1	-6.4	-0.4	-5.7	-2.4
95% CI Higher Bound	2.2	-0.1	-2.3	-3.6	-0.6	0.8
Unit Fixed Effect	None	Store	Store	Store	Store	Store
Time Fixed Effect	No	Yes	No	Yes	Yes	Yes
adj. R^2	30.2	92.8	92.7	92.8	NA	NA

Note: 1. Standard errors are in parentheses. Coefficients reported in the table are in 10^{-2}. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. 2. All std errors are grouped at county level. 3. All estimates are in percentage. 4. For dynamic TWFE, I only report the estimates that are 1 quarter post treatment. Rest of estimates will be available in the appendix.

and later treated versus early treated. In the baseline Bacon DiD setup, there is also a later versus always treated group.

The results obtained from the Bacon decomposition reveal new insights. They suggest a 4.5% decrease in the volume of juice products, indicating that competition from Amazon Fresh has led to a decline in juice sales at local grocery stores based on the weighted estimates. The volume response remains relatively small for groups treated at different times, such as early treated versus later treated and later treated versus earlier treated, implying that the timing of treatment is not a significant factor.

However, the results also highlight two substantial volume responses: a 5.3% drop in juice sales when comparing stores facing Amazon Fresh competition to those that do not face online competition, and an even larger decline when comparing later treated stores with Seattle (the only always treated market since 2012). As Seattle was the first market for Amazon Fresh, the adoption of online shopping was much lower at the time of entry. When comparing Seattle with other, later treated markets, incumbent grocery stores may face

Table 1.3: Bacon Decompose Weights and Estimates on Volume

Groups	**Weight**	**Estimate**
Earlier vs Later Treated	22.5	-0.8
Later vs Always Treated	6.7	-15.9
Later vs Earlier Treated	9.0	-0.2
Treated vs Untreated	61.8	-5.3

Note: All estimates are in percentage.

more intense competition due to the faster shipping speeds and increased market adoption of online shopping.

1.6.2 Price Effect

Next, I employ the previous specifications, including static, dynamic, and Eq 1.7, to obtain a regression of logged average quarterly store unit price[24] on Amazon Fresh's entry, with store fixed effects and time fixed effects. The results are shown in Table 1.4. For the static setup, the impact on price is not statistically different from 0. However, CS estimators grouped by time since treatment suggest a 3.1% increase in price due to competition, while grouping by entry time results in a competitive effect that only increases the price by 1.1%. The dynamic TWFE also suggests a 0.7% increase in price.

Interestingly, as shown in 1.5, the weighted price response from Bacon decomposition is -1.2%. The results following the Bacon decomposition reveal new insights, suggesting an increase in juice product prices between treated and untreated stores. The price response between these groups demonstrates a 1.7% drop in prices, indicating that competition from Amazon Fresh lowers juice product prices when comparing treated stores with untreated stores. The price response remains relatively small for groups treated at various treatment

[24]In the Nielsen Data, the unit for juice products is ounces. I also deflate the price using the CPI index from FRED. https://fred.stlouisfed.org/series/CPIFABSL

Table 1.4: Price Effect

Model Spec.	OLS	Static	Static	Dynamic	CS (Time)	CS (Group)
Amazon Fresh	0.7***	0.4	8.2***	0.7*	3.1***	1.1***
SD	(0.3)	(0.3)	(0.2)	(0.3)	(0.5)	(0.3)
95% CI Lower Bound	0.2	-0.3	7.8	0.1	2.2	0.6
95% CI Higher Bound	1.2	1.0	8.7	1.3	4.0	1.7
Unit Fixed Effect	None	Store	Store	Store	Store	Store
Time Fixed Effect	No	Yes	No	Yes	Yes	Yes
adj. R^2	37.3	72.3	70.3	73.7	NA	NA

Note: 1. Standard errors are in parentheses. Coefficients reported in the table are in 10^{-2}. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. 2. All std errors are grouped at county level. 3. All estimates are in percentage. 4. For dynamic TWFE, I only report the estimates that are 1 quarter post treatment. Rest of estimates will be available in the appendix.

timings. The inconsistency between different model setups implies that competition may not affect prices as initially anticipated.

Table 1.5: Bacon Decompose Weights and Estimates on Price

Groups	**Weight**	**Estimate**
Earlier vs Later Treated	22.5	0.3
Later vs Always Treated	6.7	-0.4
Later vs Earlier Treated	9.0	-1.4
Treated vs Untreated	61.8	-1.7

Note: All estimates are in percentage.

The observed increase in prices could stem from various factors. Stores may attempt to distinguish themselves by offering premium products, unique items, or specialized services, which could warrant higher prices. They might also incur additional costs due to competition, such as increased marketing expenses, investments in store improvements, or enhancements to their e-commerce platforms. These extra costs might be passed on to consumers in the form of higher prices. Alternatively, when faced with competition, a store could target a customer segment that is less price-sensitive and places greater emphasis on quality, service, or convenience, leading to increased prices for specific products. In the following analysis, I will explore whether stores compete on a more granular level, such as product assortment or quality-based differentiation.

1.6.3 Product Assortment

The marginally positive price effects observed in Table 1.4[25] prompt an investigation into how retailers might compete in other areas. For instance, improvements in service or better product assortment availability might be contributing factors. Consequently, I analyze the distribution of product and brand availability across stores and retail chains. The entry of Amazon Fresh could also influence the product assortment provided by incumbent stores, who may respond by expanding their product selection to differentiate themselves. In such cases, even if prices rise, the change in assortment might offset the price response, potentially masked by the volume-weighted price index.

To explore this hypothesis, I examine whether firms strategically adjust their product assortments in response to competition. Specifically, I focus on the number of unique brands and UPCs sold at stores per quarter as dependent variables, capturing the characteristics of product assortments offered by stores and chains throughout the study period. Using the scanner data, I compute the number of distinct UPCs and brand codes sold at each store in a quarter and employ them separately as measures of product assortment[26].

The findings[27] are displayed in Table 1.6. In the static model, the entry corresponds

---

[25]This table presents the results of the analysis of price effects, indicating a potential increase in service quality or product assortment availability in stores.

[26]These measures help in understanding how retailers might adapt their product offerings in the face of increased competition.

[27]The results for the change in the number of UPCs are presented in Appendix Table A.4.

Table 1.6: Product Assortment Effect: Use Number of Distinct Brands.

Model Spec.	OLS	Static	Static	Dynamic	CS (Time)	CS (Group)
Amazon Fresh	6.6***	-1.3	6.6***	-1.2*	2.4***	-0.2
SD	(0.3)	(0.7)	(0.6)	(0.4)	(0.7)	(0.4)
95% CI Lower Bound	6.1	-2.8	5.4	-2.1	1.0	-1.1
95% CI Higher Bound	7.1	0.2	7.8	-0.4	3.8	0.6
Unit Fixed Effect	None	Store	Store	Store	Store	Store
Time Fixed Effect	No	Yes	No	Yes	Yes	Yes
adj. R^2	81.8	93.2	93.1	93.3	NA	NA

Note: 1. Standard errors are in parentheses. Coefficients reported in the table are in 10^{-2}. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. 2. All std errors are grouped at county level. 3. For dynamic TWFE, I only report the estimates that are 1 quarter post treatment. Rest of estimates will be available in the appendix.

to a 6.6-brand increase. However, when employing CS estimators grouped by time since treatment, the competitive effect results in a 2.4-brand increase, while grouping by entry time reveals a negligible change of -0.2 brands. These outcomes suggest a minor augmentation in product assortment based on brand and product differentiation[28]. Results from both measures (UPC and brand) indicate that the entry of Amazon Fresh prompts incumbent grocery stores to expand their assortments to cater to different consumer preferences and increase the likelihood of retaining customers.

1.6.4 Product Quality

In response to competition from Amazon Fresh, stores may also refine their assortment strategies by differentiating the quality of products they offer. Retailers must consider not only the number of products they display and sell but also other store-level factors, such as

[28] These minor increases could be a result of retailers attempting to differentiate their offerings in response to the competition.

the range of brands and availability of organic items. These elements may cater to varying consumer demands, with stores in lower-income areas potentially offering more affordable options, while those in higher-income areas may stock premium alternatives.

Table 1.7: Product Quality Effect

Model Spec.	OLS	Static	Static	Dynamic	CS (Time)	CS (Group)
Amazon Fresh	7.1***	7.7	7.1*	2.0*	15.6***	8.2***
SD	(1.3)	(3.0)	(3.4)	(1.0)	(2)	(1.3)
95% CI Lower Bound	4.6	1.8	0.5	0	11.7	5.7
95% CI Higher Bound	9.6	13.6	13.7	4	19.6	10.7
Unit Fixed Effect	None	Store	Store	Store	Store	Store
Time Fixed Effect	No	Yes	No	Yes	Yes	Yes
adj. R^2 (%)	8.0	91.6	91.5	91.6	NA	NA

Note: 1. Standard errors are in parentheses. Coefficients reported in the table are in 10^{-2}. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. 2. All std errors are grouped at county level. 3. For dynamic TWFE, I only report the estimates that are 1 quarter post treatment. Rest of estimates will be available in the appendix. 4. All units are in ‰

To examine this aspect, I create a quality measure index based on the average national-level prices of products sold by a store, factoring in package size. The index is higher for stores carrying relatively expensive products but does not depend on the store's own pricing strategy. For product j, I compute a per-unit constant price as the average log price in year y across all stores s carrying the product, divided by the unit size (e.g., 40 oz). This measure is then averaged across the products carried by store s to establish a quality measure index for store s in year y within a sub-module b. Sub-modules are categorized as subsets of modules with similar product sizes to ensure comparability. To generate the final assortment index for store s, I demean the index by sub-module-quarter and subsequently average it over the quarters and sub-modules. The final store-level index is weighted by the revenue of each sub-module. To further refine the index, I apply a min-max normalization. To minimize bias, I include only products that rank in the top 1000 in terms of total revenue during the study period.

As illustrated in Table 1.7, the quality index suggests that stores competing with Amazon enhance their service quality. This index is an all-encompassing measure that captures various factors contributing to a store's overall quality. Retailers may choose to sell organic or high-quality products at elevated prices, or opt to expand their inventory and revamp their store layout to justify increased markups for these enhancements. The models control for differences in location and income through time and store fixed effects. The consideration of this index also relates to the marginally positive price response observed in Table 1.4. In the face of Amazon Fresh's entry, stores optimize their assortment by gravitating toward relatively more expensive products. They aim to provide a diverse range of top-quality products catering to diverse customer preferences and requirements. Furthermore, they maintain a clean, well-organized, and visually appealing store layout, promoting effortless navigation and a pleasant shopping experience.

1.6.5 Heterogeneous Treatment Effects

This section presents the evidence of heterogeneous treatment effects of Amazon Fresh's entry based on model specification described before. Demonstrating heterogeneous treatment effects strengthens the validity and robustness of the study's findings. It shows that the results are consistent across different model specifications, which enhances the credibility of the conclusions drawn from the analysis. Also, accounting for heterogeneous treatment effects allows for a more comprehensive understanding of the digital transformation. It acknowledges that the impact of entry may vary across different groups, contexts, or time periods, and helps identify which factors contribute to these variations.

In Fig 1.3 and Fig 1.4, both trends exhibit similarities and are in line with our estimations. Regarding volume, the pre-entry trend is approximately zero but begins to decline in the post-entry period, with a more pronounced drop evident after 12 quarters following entry. As for price, its pre-entry trend is also around zero, but it starts to rise after 4 quarters in the post-entry period, and the increase becomes more substantial after 9 quarters since entry. The reactions of product assortment and quality align with our estimates as well. The growth in product assortment starts to emerge after about 8 quarters from the entry, while the change in product quality experiences an immediate spike following the entry and resumes its upward trajectory after 10 quarters.

In Fig 1.5 and Fig 1.6, stores confronted with competition display a diverse range of responses. These differences suggest that the treatment effects of entry on various measures vary across markets, emphasizing the importance of accounting for heterogeneity in the study. By considering these variations, we can better understand the nuances in competitive dynamics and more accurately evaluate the impact of market entry on different aspects of the retail industry.

Figure 1.3: Dynamic two way fixed effects results

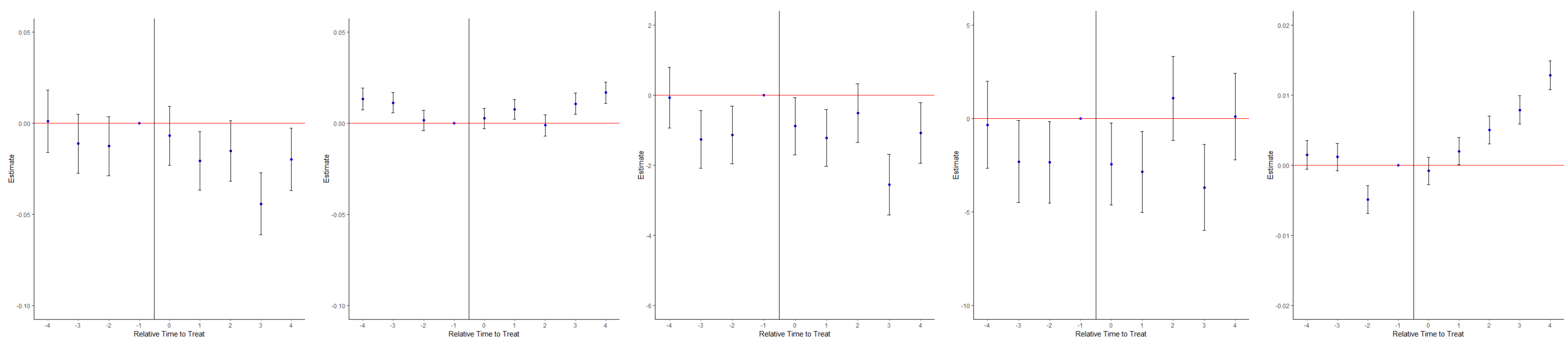
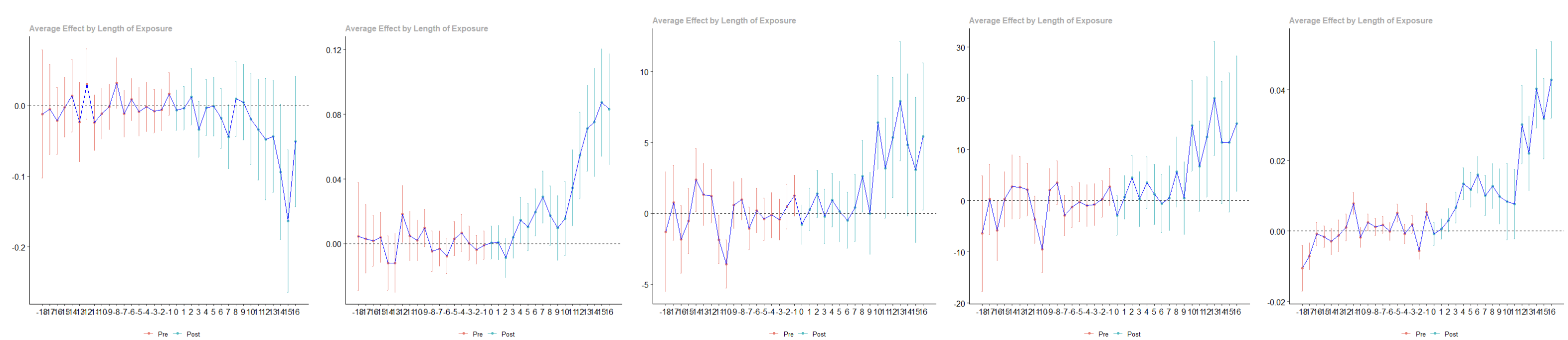


Figure 1.4: Callaway and SantAnna, 2021, Group by Time to Treatment



Note: From left to right, the figures in the table represent the results from measures following: Volume; Price; Brands, showing the number of unique brands available at the store; UPCs, representing the number of distinct Universal Product Codes (UPCs) sold at the store; Product Quality, an index based on the average national-level prices of products sold by the store.

Figure 1.5: Callaway and SantAnna, 2021, Group by Cohorts

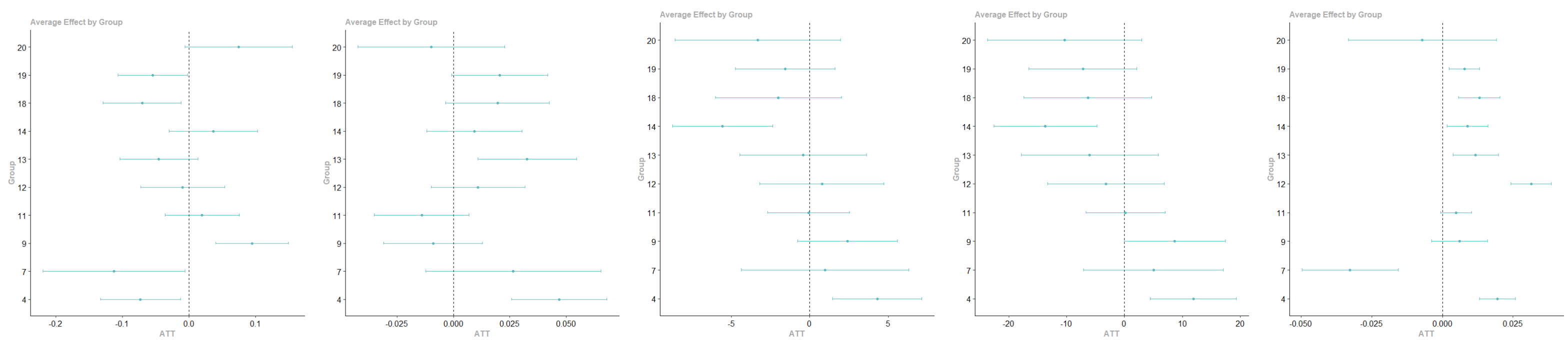
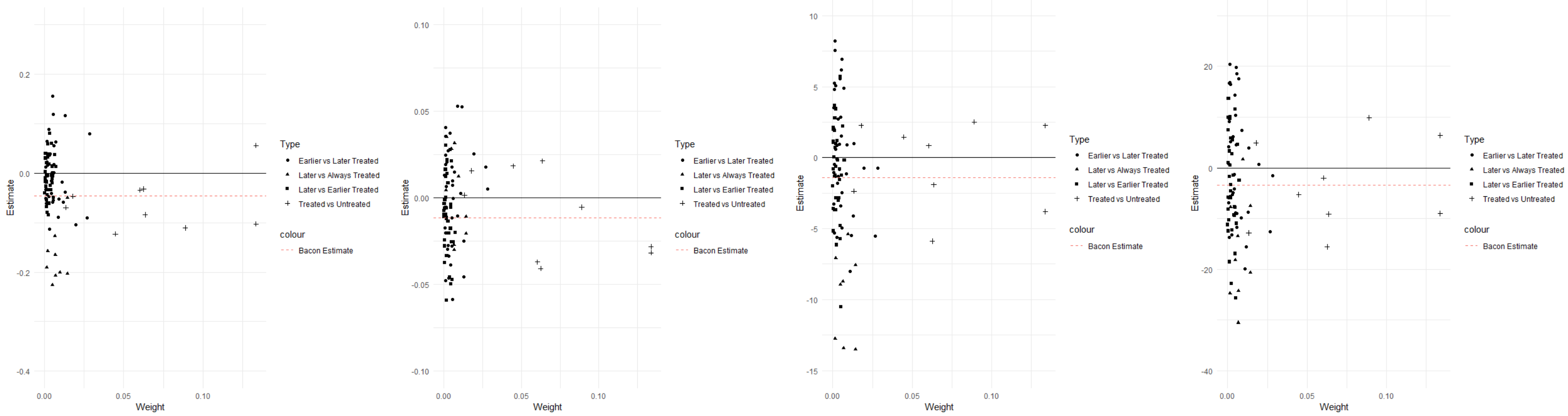


Figure 1.6: Bacon Decompose Distribution



Note: From left to right, the figures in the table represent the results from measures following: Volume; Price; Brands, showing the number of unique brands available at the store; UPCs, representing the number of distinct Universal Product Codes (UPCs) sold at the store; Product Quality, an index based on the average national-level prices of products sold by the store.

1.7 Price Dispersion

In order to investigate the impact of Amazon Fresh on store behavior, I concurrently examine the price dispersion and dynamics of identical goods within the same geographic area and over the same time period. Following Kaplan et al., 2019, I use Nielsen data to construct a dispersion measure, which estimates a comprehensive stochastic process for a store's average price level, as well as the price of a good at a store relative to its average price level. Using the estimated stochastic process, I decompose the variance of the price of the same good within the same time period and geographic area. Subsequently, I employ the models specified in previous sections to explore the effects of online competition, addressing questions examined by numerous scholars, such as DellaVigna and Gentzkow, 2019.

1.7.1 Decomposition Framework

Let p_{jst} denote the quantity-weighted average price of sub-module j at store s in time period t. As defined before, a time period would be a quarter and a good is defined by sub-module. I first decompose the log of each price p_{jst} into three additively separable components: a component that reflects the average price of the good in period t, μ_{jt}; a component that reflects the expensiveness of the store selling the good, y_{st}; and a component that reflects factors that are unique to the combination of store and good, z_{jst}. Formally, we decompose the log of p_{jst} as

$$logp_{jst} = \mu_{jt} + y_{st} + z_{jst} \tag{1.9}$$

I model both the store component of the price, y_{st}, and the store-good component of the price, z_{jst}, as the sum of a fixed effect, a persistent part and a transitory part. This statistical model is motivated by the empirical shape of the auto-correlation functions of y_{st} and z_{jst},

Formally, the statistical model for y_{st} and z_{jst} is given by

$$
\begin{aligned}
y_{st} &= y_s^F + y_{st}^P + y_{st}^T, & z_{jst} &= z_{js}^F + z_{jst}^P + z_{jst}^T,\\
y_{st}^P &= \rho_y y_{s,t-1}^P + \eta_{s,t}^y, & z_{jst}^P &= \rho_z z_{js,t-1}^P + \eta_{jst}^z,\\
y_{st}^T &= \epsilon_{s,t}^y + \sum_{i=1}^{q} \theta_{y,i}\epsilon_{s,t-1}^y, & z_{jst}^T &= \epsilon_{js,t}^z + \sum_{i=1}^{q} \theta_{z,i}\epsilon_{s,t-1}^z,\\
y_s^F &= \alpha_s^y & z_{js}^F &= \alpha_{js}^z
\end{aligned}
$$

where y_s^F and z_{js}^F denote the fixed-effects of the store and of the store-good components, y_{st}^P and z_{jst}^P denote the persistent parts of the store and of the store-good components, and y_{st}^T and z_{jst}^T denote the transitory parts of the store and of the store-good components. The parameters α_s^y and α_{js}^z are random variables with mean zero and variance $\sigma^2_{\alpha_s^y}$ and $\sigma^2_{\alpha_s^z}$. The parameters ρ_y and ρ_z are the auto-regressive parameters of the AR(1) part of the store and store-good components, while $\eta_{s,t}^y$ and $\eta_{s,t}^z$ are the innovations to the AR(1) part and are assumed to be random variables with mean zero and variance $\sigma^2_{\eta_s^y}$ and $\sigma^2_{\eta_s^z}$. Finally, the parameters $\theta_{y,i}$ and $\theta_{z,i}$ are the coefficients of the MA(q) part of the store and store-good components, while $\epsilon_{s,t}^y$ and $\epsilon_{js,t}^z$ are the innovations to the MA(q) part and are assumed to be normal random variables with mean zero and variance $\sigma^2_{\epsilon_s^y}$ and $\sigma^2_{\epsilon_s^z}$. All random variables are independent across goods, stores and times. In our baseline model we set $q = 1$.

1.7.2 Estimation Process

I estimate the parameters of the statistical model presented in 1.7.1 by utilizing data on quantity-weighted average prices, p_{jst}, for a substantial number of goods (j = 1 . . . J) at numerous stores (s = 1 . . . S) within a the same geographic markets at a quarterly frequency (t = 1 . . . T). Due to the vast number of goods, stores, and time periods, as well as the presence of unobserved components in prices, estimating this model via Maximum Likelihood is not feasible. Instead, I employ a multi-stage Generalized Method of Moments approach, which is analogous to techniques commonly used in estimating models of labor earnings dynamics (see Kaplan et al., 2019). I provide their framework here as a reference.

The estimation procedure involves four steps. **Step 1**. Estimate the good-time mean,

μ_{jt}, as the average of the log price, $logp_{jst}$, across all stores s in the market of interest, i.e.

$$\hat{\mu_{jt}} = \frac{1}{S}\sum_{s=1}^{S} logp_{jst} \tag{1.10}$$

Then construct normalized prices as

$$\tilde{p}_{jst} = logp_{jst} - \hat{\mu_{jt}} \tag{1.11}$$

Step 2. Estimate the store component y_{st} by taking sample means of the normalized prices across all goods in store s, i.e.

$$\hat{y}_{st} = \frac{1}{n_{jst}}\sum_{j=1}^{n_{jst}} \tilde{p}_{jst} \tag{1.12}$$

where n_{jst} is the number of goods for which we have data for store s in period t. In some instances $n_{jst} < J$ because not every store-good combination will meet our sample selection requirements in every quarter. I estimate the store-good component z_{jst} as

$$\hat{z}_{jst} = \tilde{p}_{jst} - \hat{y}_{jst} \tag{1.13}$$

The above process leads to a $S \times T$ panel of store components $\hat{y}_{st}$, and a $(J \times S) \times T$ panel of store-good components $\hat{z}_{jst}$ (where there may be missing data for some combinations of (j, s, t)).

Step 3. Construct the auto-covariance matrix of each of these panels up to L lags.

Step 4. Minimize the distance between the theoretical auto-covariance matrices implied by the model and the empirical auto-covariance function from step three. I use a diagonal weighting matrix that weights each moment by $n_{jst}^{0.5}$. However, the main results are not sensitive to using an identity weighting matrix instead.

1.7.3 Variance Decomposition

These findings align with other research studying uniform pricing behavior using Nielsen scanner data. DellaVigna and Gentzkow, 2019 find higher levels of uniform pricing in the data when firms face chain-level decision-making costs. In a comprehensive analysis of the same data, Hitsch et al., 2019 also find that chain factors explain a larger fraction of price dispersion than market factors. They conclude this is partly driven by the fact that, by segmenting the market, chains face relatively more homogeneous demand than the market as a whole. Incumbent grocery stores may not respond to Amazon Fresh's entry in a specific market since most stores in the Nielsen data belong to chains. If pricing decisions are made at the chain level, no price response is expected for such a local market shock unless the stores of a particular chain are densely located in one market. However, this is typically not ideal due to cannibalization concerns. Later, I also examine heterogeneous competitive effects at the store level with stores belonging to chains of different sizes.

Results of price decomposition are shown in figure 1.7.[29] In the next section, I will talk about the characteristics that contribute to store level heterogeniety of competition effects.

1.8 Store Level Heterogeniety

The observed negative causal impact of Amazon Fresh's entry on incumbent grocery stores' sales volume, along with the slightly positive effect on prices, is consistent with the findings of Arcidiacono et al., 2020[30]. Although these results may be somewhat surprising, they concur with studies investigating uniform pricing behavior among retailers using Nielsen data. To verify the robustness of these findings, I construct store-level data incorporating store characteristics to assess response variations at a more granular level. The three store characteristics considered include store type, proximity to the nearest Amazon Fresh distribution center, and chain size[31].

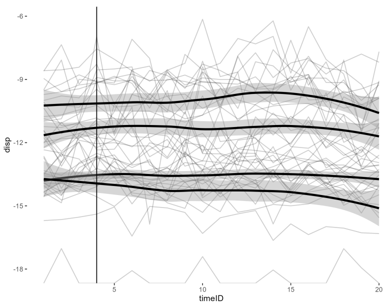
---

[29] I only present the markets that experience the entry of Amazon Fresh during the study period. Controls are not included.

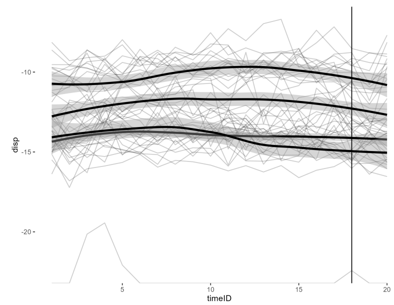
[30] Arcidiacono et al., 2020 explores the competitive consequences of Walmart's entry on nearby grocery stores, utilizing the distance to Walmart as a means to differentiate between treatment and control groups.

[31] Owing to computational constraints related to store-level fixed effects, estimations are conducted on data from 2014 to 2016 for selected markets.

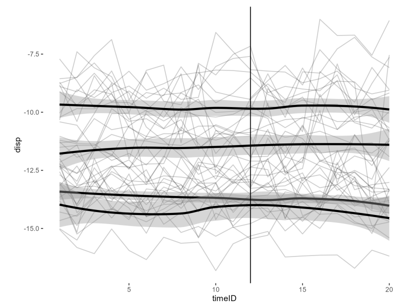
Figure 1.7: Price Dispersion Decomposition



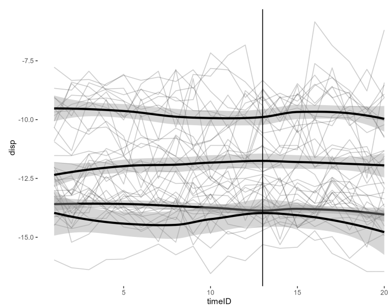
(a) Los Angeles



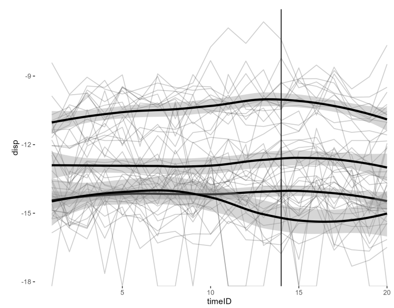
(b) Washington



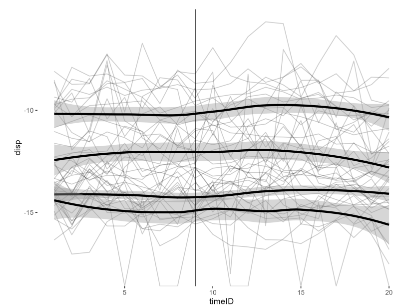
(c) San Francisco



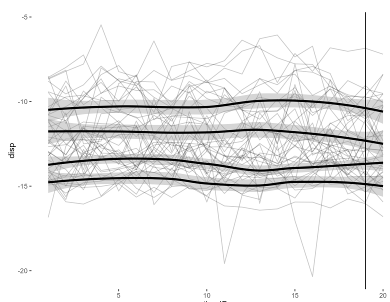
(d) Sacramento



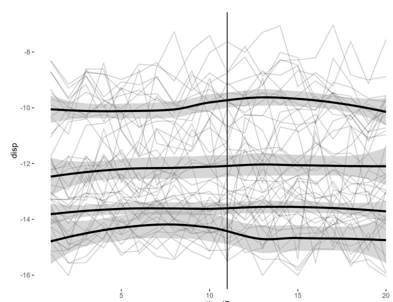
(e) Boston



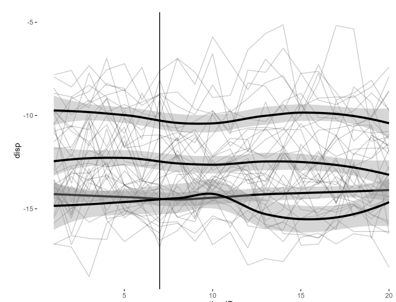
(f) Dallas



(g) Chicago



(h) New York



(i) Cincinnati

1.8.1 Distance to the Closest Amazon Fresh Distribution Center

The penetration level of Amazon Fresh can be affected by a store's proximity to the nearest AF distribution center. Stores situated closer to a distribution center might experience heightened competition, as local consumers can take advantage of swifter delivery and earlier service availability[32]. Such proximity can render Amazon Fresh a more attractive option due to the reduced delivery times.

The disparity in customer awareness could also imply varying impacts. Consumers residing nearer to the AF distribution center might exhibit greater awareness of the service and could be more inclined to embrace online grocery shopping because of its convenience and expedited delivery[33]. A shorter distance between the distribution center and customers enables AF to streamline its delivery routes, curtail transportation expenses, and bolster overall operational efficiency. These benefits empower AF to present competitive prices, thereby further challenging the incumbent stores in the vicinity.

To gauge the influence of Amazon Fresh's entry on the price and volume of juice products in incumbent grocery stores, I utilize a static two-way fixed effects (TWFE) approach. I classify stores into distinct groups based on their distance from the nearest Fresh distribution center and estimate the entry's impact using the equation below:

$$Y_{it} = \alpha_i + \gamma_t + \beta AF_{it} + \varepsilon_{it} \tag{1.14}$$

In this equation, α_i represents the store fixed effect to control for county heterogeneity, while γ_t denotes the month fixed effects to control for temporal shocks or seasonality. Y_{it} refers to the logged store-level sales volume or price per unit, β is the treatment effect of AF's entry on the dependent variable Y_{it}, and ε_{it} is the error term.

The results, as displayed in Table 1.8, reveal that incumbent grocery stores situated within 20 miles of the nearest distribution center experience a 2.1% price increase and a 2.5% sales volume decrease in juice products upon Amazon Fresh's entry. The positive price response due to the competitive effects of entry contradicts my initial expectations, suggest-

[32]Faster delivery times are essential for perishable items such as fresh produce, which plays a significant role in customers' decisions to opt for online grocery shopping.

[33]As a result, incumbent stores in these areas may face stiffer competition as customers gravitate towards AF.

Table 1.8: Model Estimates by Different Distance Bands

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable	Log Price per Unit			Log Total Volume		
Distance (Miles)	< 20	20 to 50	> 50	< 20	20 to 50	> 50
AF	2.112***	0.137	-2.013***	-2.527***	-0.530*	2.499***
	(0.129)	(0.203)	(0.412)	(0.147)	(0.232)	(0.156)
95% CI Lower Bound	1.86	-0.26	-2.82	-2.82	-0.98	1.57
95% CI Higher Bound	2.37	0.53	-1.21	-2.24	-0.08	3.43
Store Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
Month Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
Observations	147744	74424	19932	147744	74424	19932
Number of Stores	4105	2069	554	4105	2069	554

Note: Standard errors are in parentheses. Coefficients reported in the table are in 10^{-2}. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

ing the existence of price-increasing competition[34]. For stores located between 20 and 50 miles from the distribution center, the competitive effects on price and volume from AF are not significantly different from zero, indicating no competitive effects within this distance range. Intriguingly, stores located more than 50 miles away from the Fresh distribution center exhibit significantly positive volume responses and significantly negative price responses. Although these effects are counterintuitive, it is plausible that AF's entry signifies a flourishing grocery market in nearby counties, which in turn positively impacts the revenue of more distant stores.

[34]One possible explanation is that, in price-taking markets, suppliers especially large brands can impose higher prices when consumers have more choices.

1.8.2 Store Type

I utilized the information provided by Nielsen data to focus on three major store types that sell juice products, as these stores compete with Amazon Fresh to varying degrees based on their customer overlap. These stores compete with Amazon Fresh to varying extents, contingent on the degree of customer overlap. Table 1.9 demonstrates that incumbent food grocery stores encounter a 3.0% price increase and a 3.3% volume decrease in juice product sales when faced with Amazon Fresh's entry, signifying direct competition with these stores. Conversely, drug stores exhibit a 0.5% price decline and no significant volume alteration, implying a lack of competitive effects due to their indirect rivalry with Amazon Fresh. Simultaneously, incumbent mass merchandise stores grapple with a moderate level of competition, manifested by a 2.3% price escalation and a 2.7% volume reduction in response to Amazon Fresh's entry.

Table 1.9: Model Estimates by Different Store Types

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable	Log Price per Unit			Log Total Volume		
Channel	Food	Drug	Mass	Food	Drug	Mass
AF	2.988**	-0.491**	2.301***	-3.306***	-0.075	-2.654***
	(0.137)	(0.163)	(0.242)	(0.151)	(0.189)	(0.281)
95% CI Lower Bound	2.72	-0.81	1.82	-3.60	-0.44	-3.21
95% CI Higher Bound	3.26	-0.17	2.78	-3.01	0.29	-2.10
Store Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
Month Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
Observations	81792	112068	48240	81792	112068	48240
Number of Stores	2272	3113	1340	2272	3113	1340

Note: Standard errors are in parentheses. Coefficients reported in the table are in 10^{-2}. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Multiple factors can influence the degrees of competitive effects between Amazon Fresh and various store types. Firstly, the customer base for each store type might differ in aspects such as income levels, shopping inclinations, and priorities[35].

Secondly, each store type generally focuses on a unique set of products or services. Food grocery stores customarily offer a more comprehensive assortment of fresh produce and groceries, while drug stores might emphasize pharmaceuticals, health, and beauty items. Mass merchandise stores provide a diverse range of products, including groceries, clothing, and electronics. Although Amazon Fresh principally competes with food grocery stores in the online grocery domain, its impact on other store types that specialize in distinct product categories may be restricted.

Thirdly, store types can vary concerning their physical locations, operating hours, and accessibility[36]. In conclusion, the competitive interplay between Amazon Fresh and different store types hinges on factors such as customer demographics, product offerings, and convenience. These elements contribute to the observed heterogeneity in competition across disparate store types.

1.8.3 Chain Size

In this analysis, I investigate whether the lack of price response can be attributed to retailer scale. Large chains may be inclined to minimize their reaction to local entry to reduce price competition. Similarly, chain supermarkets could establish relatively uniform retail prices across their stores within a specific market or region, either to save on firm-level adjustment costs or to prevent price comparisons within their chain. This multi-market aspect of price-setting would, in turn, limit the extent to which chains might optimally react to Amazon

[35]For instance, customers who prioritize convenience may gravitate towards Amazon Fresh or drug stores, while those who value an extensive product selection might opt for larger grocery or mass merchandise stores.

[36]Amazon Fresh holds the advantage of online shopping and home delivery, appealing to convenience-seeking customers. In contrast, traditional grocery and mass merchandise stores might supply more extensive in-store services and product variety, while drug stores are often conveniently situated in residential areas or adjacent to pharmacies.

Fresh's entry, which only locally affects one or a few of their stores. Consequently, I examine whether price responses vary depending on whether the incumbent supermarket is part of a national chain, regional chain, local chain, or an independent supermarket.

The results in Table 1.10 indicate that small and medium-sized chains experience more significant competition from Amazon Fresh's entry, while stores belonging to large chains are minimally affected. This finding suggests that Amazon Fresh's entry primarily generates intense local competition, with regional chains being the most impacted. Such competitive effects exhibit positive price responses and negative volume responses.

Table 1.10: Model Estimates by Different Chain Sizes

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent Variable	Log Price per Unit			Log Total Volume		
Chain Size	Small	Medium	Large	Small	Medium	Large
AF	2.030***	5.045***	-0.439**	-2.337***	-5.603***	0.046
	(0.180)	(0.188)	(0.148)	(0.147)	(0.232)	(0.172)
95% CI Lower Bound	1.68	4.67	-0.82	-2.73	-6.02	-0.29
95% CI Higher Bound	2.38	5.41	-0.21	-1.95	-5.19	0.38
Store Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
Month Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
Observations	55368	50328	136404	55368	50328	136404
Number of Stores	1538	1398	3789	1538	1398	3789

Note: Standard errors are in parentheses. Coefficients reported in the table are in 10^{-2}. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Another concern is that large chains might launch their online grocery service in response to Amazon Fresh's entry. These similar online grocery services not only compete with Amazon Fresh but also cannibalize their physical store revenue. Including only a store fixed effect would not be sufficient to account for the effects of stores that initiate online grocery services. Fortunately, according to Cavallo, 2017, online and offline grocery prices do not significantly differ. The study demonstrates that, in large multi-channel retailers, there is minimal difference between online prices collected from a website and offline prices obtained by visiting the physical store. Prices are identical approximately 72% of the time, and while price changes are not synchronized, they exhibit similar frequencies and magnitudes. However, considerable heterogeneity exists across countries, sectors, and retailers.

1.8.4 Future Research

The findings from the previous store-level heterogeneity analysis reveal unexpected competitive effects. Stores anticipating competition from Amazon Fresh's entry exhibit a positive price response and a negative volume response. In addition, results from product assortment and quality index suggest that stores react to the entry by providing more juice and sell higher end products. In this section, I discuss several potential confounds that require future testing and offer possible explanations for the observed store-level estimates.

Firstly, incumbent stores may implement various loyalty programs or customer retention strategies to maintain their customer base in the face of competition from Amazon Fresh. Incumbent stores might engage in more promotional activities, such as special discounts, limited-time offers, or loyalty programs, to retain customers in the face of increased competition from Amazon Fresh. These promotional activities might temporarily affect prices and sales volumes, making it difficult to isolate the direct impact of Amazon Fresh's entry on incumbents.

Another possibility is that customers who continue to shop at these supermarkets may purchase more expensive product bundles, which could obscure any negative price response in our analysis. Even if incumbents do not respond to Amazon Fresh entry by altering their product offerings, the remaining customers might be more likely to buy pricier items or prefer smaller package sizes than those who switch. In such a case, even if prices are reduced, this effect could counterbalance the price response and be masked by our volume-weighted price index.

Another potential confound is that exposure to Amazon Fresh entry might generate a positive causal effect by affecting price dispersion across products. This dual response could result from the digital transformation of grocery shopping, which changes both the distribution of consumers who shop at incumbent stores and the competitive environment. For example, Amazon Fresh's entry may prompt a store to set competitive prices for price-sensitive shoppers and raise prices for brands preferred by quality-sensitive shoppers, leading to an offsetting effect that evens out across products. In line with this, a study of shopper homescan data by Jindal et al., 2018 finds that low-priced value brands perform better at Walmart Supercenters, while premium brands perform better at supermarkets. Additionally, consumers who shop at both types of outlets are more price-sensitive than those who are loyal to a single one. Similarly, a shifting consumer distribution might imply the optimal incumbent response to Amazon Fresh's entry. The distributional change, termed the price sensitivity effect by Chen and Riordan, 2008, suggests that incumbents should raise prices, while the competitive or market share effect implies a price cut. The estimated positive causal effect of entry on prices would thus represent the net of these two effects.

Lastly, the possibility that the positive price response is driven by retailers not having full control over the pricing process is a potential explanation for our findings. Specifically, it is conceivable that incumbent supermarkets are limited in their price-setting ability by contractual (vertical) relationships with manufacturers. The literature does not provide a clear consensus on whether the retailer or the manufacturer has the most power in setting prices. Previous literature has modeled this either as a non-cooperative game, as per (Villas-Boas, 2007), or as a cooperative (Nash) bargaining problem, as per (Draganska et al., 2010). If the manufacturer dominates the relationship, the price response at an individual store may be positive, as the manufacturer seeks higher profits by trading with other platforms.

1.9 Conclusion

This research highlights the importance of employing multiple methods and considering various confounders to thoroughly examine the impact of online grocery services on traditional retail stores. By integrating the traditional DiD design with advanced econometric techniques and accounting for potential confounders, the study provides a more robust and comprehensive understanding of the complex competitive landscape.

In this study, I merge quarterly grocery scanner data with information on the entry of Amazon Fresh to investigate the impact of the digital transformation of online grocery shopping on brick-and-mortar stores. Using the traditional staggered DiD design of TWFE, along with recent advances in econometrics to correct potential bias due to varying treatment timings, I discover that Amazon Fresh's entry leads to a significant negative volume response and, surprisingly, only slightly positive price response for incumbent grocery stores in the juice products category.

Bacon decomposition results further corroborate these findings within treated groups with different treatment timings, while comparisons between treated and untreated groups reveal a positive price response. This suggests a potential price-increasing competition, which starkly contrasts with previous studies that indicate considerable price reductions and welfare gains for consumers through competitive mechanisms.

Incumbent grocery stores not only compete through volume and price but also respond to competition by expanding their product selection and improving the quality of their offerings. The positive price response is partially attributed to increased costs and markup charges due to changes in product assortment and quality. While online grocery services attract customers with convenience, brick-and-mortar stores aim to differentiate themselves by providing superior in-person service. This strategy inevitably raises costs and may force smaller stores that cannot adapt to exit the market.

The positive price response obtained from Bacon decomposition is further scrutinized through store-level heterogeneity analysis. Amazon Fresh's entry appears to have a localized impact, primarily affecting local and regional chains. The estimated effects also suggest that Amazon Fresh mainly competes with stores located within 20 miles of its distribution center. These results warrant further investigation, as other confounding factors such as promotinal activities, manufacturer price-setting behavior, and consumer shopping patterns may change in response to the introduction of online grocery services. The application of recent econometric contributions to correct bias from TWFE estimates offers intriguing implications and has the potential to be further employed in empirical research.

The utilization of multiple methods and the investigation of confounders not only strengthens the credibility of the findings but also contributes to the development of empirical research methodologies. By considering various aspects of the market, researchers and industry practitioners can derive more actionable insights and better understand the repercussions of digital transformation on traditional retail businesses, ultimately guiding future policy-making and business strategy decisions.

Chapter 2

E-COMMERCE EFFECT ON LOCAL MARKETS

2.1 Introduction

The rise of e-commerce giants, such as Amazon, has drastically altered the retail landscape and consumer behavior. This phenomenon, known as the digital transformation of retail, is the focus of this chapter. We aim to explore the intricate impacts of this transformation at the county level, with particular emphasis on the labor market, the retail market, and consumer shopping behavior. These shifts are essential to comprehend as they provide crucial insights into how the retail industry is adjusting to the digital era and how these changes affect communities at a localized level.

An integral part of the e-commerce retailer's strategy is the optimization and expansion of the Fulfillment Center (FC) network. The inauguration of the first FC took place in 1997, and by the close of 2016, the number of FCs had exceeded 90. The establishment of new FCs aims to eliminate long-zone shipping and reduce the order-to-delivery time, which is evident by the majority of FCs being strategically located in densely populated areas such as the east and west coasts. This strategy may draw nearby customers to shop through the e-commerce retailer, favoring the convenience of faster delivery over local brick-and-mortar stores.

We begin our exploration by examining the labor market, focusing on the complex dynamics that emerge when Amazon Fulfillment Centers (FCs) establish their presence in an area. Interestingly, while these centers tend to increase wages, they also appear to decrease the total retail workforce. This dichotomy presents intriguing questions about labor competition and the impact of digital transformation on traditional stores. Next, we scrutinize the retail grocery market at the county level, paying close attention to changes in sales volumes and prices. Incumbent stores are facing the daunting challenge of an increasingly digital marketplace, which seems to be applying pressure on their revenues.

Furthermore, we analyze consumer behavior, noting a shift towards online shopping and a reduction in physical store visits and spending. This evolution offers a clear indication of the ongoing digital transformation and presents questions about the future of traditional shopping habits. Our goal in this chapter is to illuminate these multi-faceted impacts of digital transformation at the county level. By doing so, we seek to provide a comprehensive understanding of these ongoing changes and their implications for various stakeholders in the retail industry.

This research contributes to the literature on the heterogeneous effects of e-commerce on the retail sector, particularly in the grocery retail sector. Earlier studies have reported how e-commerce lowers prices and increases consumer surplus for various products using reduced-form models. This research investigates how e-commerce influences competition, employment, income, entry-exit dynamics, and pricing in the grocery retail sector using a structural model. This sector is particularly relevant for studying e-commerce effects because it accounts for a large share of consumer spending and employment, and e-commerce penetration has been growing rapidly in recent years.

The chapter is organized as follows: Section 2 presents a literature review; Section 3 introduces the data used in the study; Section 4 discusses the impact of Amazon fulfillment centers on local labor markets; Section 5 examines the competitive effects of Amazon Fresh at the county level; Section 6 investigates the changes in consumers' behavior in response to the introduction of Amazon Fresh; and finally, Section 7 concludes the chapter.

2.2 Literature Review

This paper contributes to the growing literature on the competitive effects of e-commerce on traditional retail sectors and expands upon previous research by focusing on the impact of online competition on local labor markets at the county level. Several related studies have laid the groundwork for understanding various aspects of e-commerce and its effects on consumer welfare, pricing, and geography. For instance, Brynjolfsson et al., 2003 demonstrated that online bookstores increased consumer welfare by offering greater product variety, while Dimoka et al., 2012 found that online used-book sales only cannibalized a small portion of new-book purchases.

The paper also connects to the literature that explores the impact of big-box retailers, such as Walmart, on local labor markets at the county level. Basker, 2005 studies the effects of Walmart on retail employment using nationwide data. Basker attempts to account explicitly for endogeneity by instrumenting for the actual number of stores opening in a county in a given year with the planned number. The latter is based on numbers that Walmart assigns to stores when they are planned; according to Basker, these store numbers indicate the order in which the openings were planned to occur. She then combines these numbers with information from Walmart Annual Reports to measure planned and actual openings in each county and year; the data reflect some measurement error in store opening dates. Her results indicate that county-level retail employment grows by about 100 in the year of Walmart entry, but declines to a gain of about 50 jobs in five years as other retail establishments contract or close. In the meantime, possibly because Walmart streamlines its supply chain, wholesale employment declines by 20 jobs in the longer term.

Jia, 2008 tackles the endogeneity of Wal-Mart store locations through a more structural approach. Specifically, she investigates the entry of chain stores such as Wal-Mart and K-Mart into local markets and their effects on the number of small retailers (as opposed to retail employment) by modeling it as a two-stage game. In the first stage, Wal-Mart and K-Mart enter local markets, and in the second stage, small retailers decide whether to enter. Her estimates indicate that during the 1988-1997 sample period, Wal-Mart's expansion accounts for 50% to 75% of the decline in the number of small retailers. Furthermore, she demonstrates that estimating a reduced-form model and neglecting the endogeneity of entry decisions could lead to underestimating the negative effects of Wal-Mart on small retailers by 50-60 percent.

Neumark et al., 2008 examine the impact of Wal-Mart stores on county-level retail employment and earnings, taking into account the endogeneity of location and timing of Wal-

Mart openings, which could potentially bias the evidence against identifying negative effects of Wal-Mart stores. To address the endogeneity issue, they employ a natural instrumental variables approach that stems from the geographic and temporal patterns of Wal-Mart store openings, which gradually expanded from the initial stores in Arkansas. Their findings suggest that the opening of a Wal-Mart store leads to a reduction in county-level retail employment by approximately 150 workers, implying that each Wal-Mart worker replaces around 1.4 retail workers. This equates to a 2.7 percent decrease in average retail employment. Furthermore, the payroll analysis indicates that Wal-Mart store openings result in a decline of roughly \$1.4 million in county-level retail earnings, or a 1.5 percent decrease. It is essential to note that these effects transpired in the context of increasing retail employment, and only imply lower retail employment growth than what would have occurred in the absence of Wal-Mart's influence.

This paper also relates to studies examining the impact of digital platforms, such as Craigslist, on local newspapers' advertising revenue (Seamans and Zhu, 2014). By focusing on county-level changes, this paper contributes to the literature on e-commerce and traditional retail sectors, emphasizing the importance of understanding the impact of online grocery competition on local labor markets. This insight will be crucial in informing policy decisions and preparing the workforce for the ongoing transformation of the retail landscape driven by e-commerce.

Ghose and Yao, 2011 emphasize the importance of price dispersion as an indicator of market efficiency and examine a unique dataset of actual transaction prices from both electronic and offline business-to-business markets. They find a substantially lower price dispersion in electronic markets (0.22%) than previously reported, suggesting that the ”law of one price” can prevail when considering transaction prices instead of posted prices. By developing a theoretical framework that identifies new drivers of price dispersion, such as product cost, order cycle time, own price elasticity, and transaction quantity, they also investigate the moderating role of electronic markets in the relationship between these drivers and price dispersion. The study concludes that electronic markets can enhance consumer surplus by up to \$97.92 million per year, highlighting the efficiency gains resulting from transactions in relatively friction-free markets.

Newberry et al., Forthcoming investigates the distortionary effects of nexus tax laws on Amazon's distribution network investments from 1999 to 2018, emphasizing the roles of network densification and vertical integration into package sortation. While densification reduces shipping costs, it increases facility operating costs in more expensive areas and reduces scale economies in processing shipments. Nexus laws also generate additional sales tax liabilities as the network expands. By combining household spending data across online and offline retailers with detailed information on Amazon's distribution network, the authors quantify these trade-offs using a static demand model and a dynamic investment model. Their findings suggest that Amazon's expansion led to significant shipping cost savings and facilitated the realization of aggregate economies of scale.

In more recent studies, Chava et al., 2022 utilizes an administrative payroll dataset for 2.6 million retail workers to investigate the impact of a major e-commerce firm's staggered rollout of fulfillment centers on traditional retail workers' income in geographically nearby counties. Their findings reveal a 2.4% decrease in income, with hourly workers, particularly part-time hourly workers, experiencing a significant drop in wages due to reduced hours worked. A U-shaped pattern emerges, with both young and old workers facing a sharper decline in wage income. This leads to an increase in credit card delinquency for some workers. By analyzing data for 3.2 million stores, the study finds that sales and employment at proximate stores decrease by 4% and 2.1%, respectively. Additionally, exits, predominantly of young and small stores, increase while entry decreases. In aggregate, the retail sector loses 938 jobs per county per quarter, while the transportation-warehousing sector and food services sector gain 256 and 143 jobs, respectively. These results underscore the impact of creative destruction driven by e-commerce on local labor markets.

Alcedo et al., 2022 investigates e-commerce trends in 47 economies and 26 industries during the COVID-19 pandemic using aggregated and anonymized transaction-level data from Mastercard, which has been scaled to represent total consumer spending. The study finds that the share of online transactions in total consumption increased more in economies with higher pre-pandemic e-commerce shares, thereby widening the digital divide across economies. However, recent data indicates that these spikes in online spending shares are dissipating at the aggregate level, with variation across industries. Specifically, the share of online spending in professional services and recreation has fallen below its pre-pandemic trend, while a more persistent shift to digital channels is observed in retail and restaurants.

Hwang and Park, 2016 analyze the impact of Walmart supercenter conversions on consumer shopping behavior using a difference-in-difference estimator. They find that, after conversion, Walmart experiences a 41% increase in weekly revenue, primarily due to larger expenditures per visit rather than an increase in store visits. On the other hand, competing grocery retailers face a 20% decrease in weekly revenue, mainly attributed to a reduction in store visits. These findings suggest that consumers benefit from reduced shopping costs by making fewer trips overall and increasing their Walmart basket sizes. Additionally, the authors discover that the revenue gains for Walmart from conversion outweigh the small cannibalization loss at existing Walmart supercenters located farther away. Category-level analyses reveal increased spending in pre-existing categories at the converted supercenter, with positive demand externality more pronounced in food categories, primarily due to increased purchase incidence. The implications of these findings are relevant for both academics and retail managers, as they provide valuable insights into consumer behavior changes and the competitive dynamics in the retail industry following the conversion of Walmart supercenters.

2.3 Data

In this chapter, I concentrate on the impact of Amazon Fresh before its merger with Whole Foods in 2017, focusing on the same markets as in Chapter 1. However, the primary difference is that this study investigates county-level competition arising from household shopping behavior and local retail labor market. To examine the influence of Amazon fulfillment centers (FCs) on the local retail industry, the study utilizes county-level data[1]. The study period spans from 2012 to 2016, during which Amazon expanded from 36 to 97 FCs in the United States.

2.3.1 Nielsen Consumer Panel Data

Nielsen acquires consumer panel data through home scanners, providing detailed insights into participating households' shopping behavior from 2012 Q1 to 2016 Q4. This dataset encompasses information on panelist households' local store shopping habits. Although the data for Amazon Fresh and Amazon are not available during the study period, consumer panel data allows us to directly assess the incremental impact of Amazon Fresh on competing retail channels. The study includes 232 counties, and trip-level data is converted into quarterly data. For each household in a county during the study period, quarterly spending, the count of visits, and average per-visit expenditure are computed, offering a comprehensive analysis of consumer shopping patterns.

2.3.2 Amazon Fulfillment Center data

Amazon's distribution network[2] has witnessed remarkable growth in recent years, with the number of fulfillment centers (FCs) increasing rapidly in parallel with the company's rising net revenue. These FCs cater to various purposes and product types, including sortation centers that handle items of all sizes, as well as specialized FCs such as Prime Now Hubs, return processing centers, vendor product management centers, and facilities for frozen merchandise storage.

The size of FCs varies significantly, with large sortation centers often exceeding one

---

[1] Although I take county-level data from the Census, Amazon operates at a more granular zip code level. Nevertheless, the results should remain unbiased as Amazon usually covers most zip codes within a county where it has a presence, and the majority of retail stores are likely located in densely populated areas within each county that correspond to the zip codes served by Amazon.

[2] In addition to the Amazon Fresh distribution center data set, I expand the data to all fulfillment centers (FC) to estimate both the impact of Amazon FCs and Amazon Fresh centers on the local grocery industry.

million square feet and Prime Now Hubs being considerably smaller, some as little as 10,000 square feet. Amazon's FC expansion is evident in its growth from 19 FCs in 2010 to 97 by 2016 and to surpass 100 in 2017. Table 2.1 showcases the cumulative number of Amazon Fresh Fulfillment centers, newly added FCs, the number of current states and counties with FCs, and average FC area added each year.

A key feature of Amazon's FCs is their strategic clustering in densely populated areas, with 89 out of 92 analyzed counties classified as part of metropolitan statistical areas. Many of these counties host multiple FCs, enabling Amazon to efficiently cater to a large customer base and maintain its competitive advantage in the fast-paced e-commerce industry.

Year	Cumulative Counts of FC	Counts of New FC	States with FC	County with FC	Avg area of New FC (k sqft)
2010	19	7	10	379	759
2011	28	9	10	379	841
2012	36	8	12	752	975
2013	44	8	14	1,189	906
2014	55	11	14	1,644	707
2015	67	12	16	1,937	548
2016	97	30	27	1,983	704

Table 2.1: Amazon Fulfillment Center & Amazon Fresh Center

Note: Here I provide the fulfillment center data for both regular Amazon FC and Amazon Fresh distribution center.

2.3.3 County Business Pattern Data

Using the County Business Patterns (CBP) dataset, collected by the US Census Bureau, I can access annual business and worker data across counties and industries, including payments to workers and job numbers. However, the primary limitation of the CBP data in studying Amazon's effects is the inability to compute wages due to the absence of employment breakdown into full-time and part-time workers or skill composition information. As a result, it is not possible to determine whether changes in payrolls reflect changes in pay rates for comparable workers or shifts in skill composition or hours. Nevertheless, I can estimate the effect of Amazon on total retail payrolls.

The study investigates the retail sector as a whole and the general merchandising subsector, which encompasses Amazon and other general department stores. This approach aids in evaluating the results and understanding the potential implications on different retail sectors.

Working with CBP data presents challenges due to federal law prohibiting the publication of data that may disclose an individual employer's operations. As I examine more disaggregated subsets of industries, data is more likely to be suppressed, and the sample becomes smaller. I construct a sample that consistently compares at least some retail industry sectors [3] for the same set of observations, considering all county-year observations with complete employment and payroll data for aggregate retail and the separate retail subsector to which Amazon and Amazon Fresh belong.

To create time-consistent geographical areas, I account for county merges or splits during the sample period. For counties that split, I maintain the original county definition, and for merged counties, I create a single corresponding county throughout. This process results in a file of 156 counties over seven years, to which I merge the CBP and Amazon data. Finally, I assign population data from the US Census Population Estimates to the counties for each year.

[3] I focus on 3-digit NAICS industries that are most likely to compete with the major e-commerce retailer's product catalog. The 3-digit NAICS codes that I classify as retail includes 442 (furniture and home furnishing stores), 443 (electronic and appliance stores), 444 (building material and garden equipment and supplies dealers), 448 (clothing and clothing accessories stores), 451 (sporting goods, hobby, book, and music stores), 452 (general merchandise stores), and 453 (miscellaneous store retailers).

2.4 The Impact of Amazon Fulfillment Center on Local Labor Market

2.4.1 Empirical Model

Influenced by Neumark et al., 2008[4], I estimate models for shifts in retail employment and payrolls. Generally, I gauge the heightened exposure to Amazon Fulfillment centers through a metric of center openings in a county-year cell, represented by the change in center count. I express changes in employment, payrolls, and the number of centers on a per capita basis to mitigate the disproportionate impact of significant employment shifts in exceptionally large counties. By dividing all these changes by the county population, the estimated coefficient for the Amazon Fulfillment center variable still reflects the effect of a center opening on alterations in retail employment or earnings levels.

To account for overall income growth potentially influencing retail demand, I incorporate changes in total payrolls per capita as a control variable. This captures county- and year-specific economic shocks that might coincidentally relate to the distance-time interactions forming the instrumental variable, as these shocks are likely to affect particular regions during specific periods. Furthermore, all models include fixed year effects to accommodate aggregate factors influencing retail employment or earnings shifts that could correlate with Amazon Fulfillment center openings, which tend to increase in frequency later in the sample.

I represent the county-level measures of retail employment, payrolls (per capita), as Y, the number of Amazon FCs (per capita) as AFC, total payrolls per capita as TP, indicator whether they are in the same county as the FC[5], and year fixed effects (in year s) as YR_s. With county j (j = 1,...,J) and year t (t = 1,...,T) indices, the baseline model for the change in the dependent variable for each observation jt is:

$$\Delta Y_{jt} = \alpha + \beta \Delta AFC_{jt} + \gamma \Delta TP_{jt} + \sum_{s=1}^{T} \xi_s YR_s + \epsilon_{jt} \tag{2.1}$$

[4] They examine the impact of increased Walmart exposure on county labor markets.

[5] If FC is in the county or neighbouring county, the new FCs may hire directly those retail workers which are under consideration.

Fixed county differences in the levels of the dependent variables are eliminated in the first-differenced model. However, there may be systematic variation in these first differences across different regions, corresponding to faster or slower growth. To accommodate this in a highly flexible manner, I also estimate the first-difference models, including county fixed effects. This approach allows for a distinct linear trend for each county, transforming the model into

$$\Delta Y_{jt} = \alpha + \beta \Delta AFC_{jt} + \gamma \Delta TP_{jt} + \sum_{s=1}^{T} \xi_s YR_s + \sum_{i=1}^{J} \phi_i CT_i + \epsilon_{jt} \tag{2.2}$$

I define changes in employment, payrolls, and number of stores on a per person basis, to eliminate the undue influence of a small number of large employment changes in extraordinarily large counties. As long as we divide all of these changes by the number of persons in the county, the estimated coefficient on the Amazon Fresh variable still measures the effect of the entry of Amazon Fresh on the change in the level of retail employment or earnings. To control for overall income growth that may affect the level of demand for retail, we include changes in total payrolls per person as a control variable, capturing economic shocks specific to counties and years that could coincidentally be associated with the distance interactions that make up the instrumental variable, since these shocks likely affect specific regions in specific periods. In addition, all models include fixed year effects to account for aggregate influences on changes in retail employment or earnings that might be correlated with Amazon Fresh openings, which occur with greater frequency later in the sample.

Here, fixed effects models can help address endogeneity in panel data analysis, particularly when the source of endogeneity is due to time-invariant unobserved heterogeneity.[6] By utilizing fixed effects, the model accounts for these time-invariant unobserved factors, potentially reducing endogeneity. However, fixed effects models cannot fully address endogeneity arising from other sources, such as time-varying unobserved factors or simultaneous

[6] In other words, fixed effects models control for unobserved factors that are constant over time within an observational unit (e.g., individual, firm, or region) but might be correlated with the independent variables of interest.

causality (reverse causality) between dependent and independent variables.

2.4.2 Endogeneity and Potential Biases

Consistent estimation of Eq 2.1 or Eq 2.2 requires that ϵ_{jt} is uncorrelated with the right-hand-side variables. If the Amazon fulfillment center location decisions are based in part on contemporaneous and future changes in employment or payrolls, then this condition could be violated. This endogeneity is natural, since Amazon would be expected to make location decisions (including the location and timing of store openings) based on current conditions and future prospects to lower their operational cost, which might be related to both employment and payroll. As one example, Amazon may open fulfillment centers where real estate development and zoning have recently become favorable to retail growth and road construction is favorable for transportation.

An instrumental variable is a third variable that exhibits a correlation with the predictor variable but remains uncorrelated with the response variable. This variable assists in estimating the genuine causal effect of the predictor variable on the response variable. In studying Amazon fulfillment center locations, an instrumental variable is required that relates to Amazon's decision to locate its centers while remaining unconnected to other factors that may impact the outcome of interest. One potential instrumental variable could be the proximity to significant transportation hubs and highways, as this factor may influence Amazon's delivery cost and efficiency without directly affecting other variables, such as local employment, wages, or prices.

Another potential instrumental variable could be the number of existing warehouses in a region, as this may indicate Amazon's market share and demand without directly influencing other outcomes such as consumer behavior or satisfaction. These examples of potential instrumental variables must be tested for validity and relevance to your specific research question and data.

2.4.3 Identification

My identification strategy in light of this potential endogeneity is based on the geographic pattern of Amazon fulfillment centers' openings over time. First, Amazon placed FCs in relatively low population states that were close to highly populated areas. For example, Amazon opened two FCs in Nevada, both of which were on the California border close to that states major cities. Second, they also placed FCs in states with relatively high tax rebate and low sales tax. For example, the company opened a FC in New Hampshire, in addition to the one it already operated in Delaware, both of whom are close to major East Coast cities and have zero sales tax.[7]

This pattern of growth generates an exogenous source of variation in the location and timing of Amazon business service in a given market with different level of services that provides natural instrumental variables for online business. I construct an instrument that is similar to Baum-Snow and Kahn, 2000[8] The instrument is composed of cross-sectional variation in the distance between the FCs and the closest Interstate highway entrance and major airports[9] interacting with state-level state subsidy of corporate subsidies[10].

The choice of FC locations by the major e-commerce retailer is not random. To optimize fulfillment time, the major e-commerce retailer chooses locations that serve large populations and minimize long-zone shipping within the distribution network. This, in turn, utilizes

[7]Newberry et al., Forthcoming argue that such choices of states are also advantageous since sales tax rates are positively correlated with population (across states, correlation of between 0.35 and 0.4 across 2006 to 2013), so that entry into a small state near a large state has limited tax implications for only a small population, while allowing the firm to serve both states populations more efficiently.

[8]Baum-Snow and Kahn, 2000 investigates the impact of large public transit projects in the US aimed at reducing private vehicle dependency and reversing the decline in public transit use. The study utilizes a unique panel dataset for five major cities that upgraded their rail transit systems in the 1980s, using distance as a proxy for transit access. The results indicate that new rail transit has a modest effect on usage and housing values, providing tangible benefits to nearby residents. However, these benefits are not evenly distributed. The authors examine which demographic groups are over represented in transit growth areas and analyze changes in transit usage among different demographic groups.

[9]The distance is calculated as the travel distance to the closest Interstate highway entrance and airports using Google map.

[10]Chava et al., 2022 constructed a similar IV using distance to closest USPS interacting with the state level generosity of Amazon.

existing logistics and shipping infrastructure. I use the distance of each FCs to it's closest Interstate highway and major airports to approximate shipping infrastructure. However, this network of shipping infrastructure does not vary greatly over time. Corporate subsidies offered by local and state governments may play an important role in determining the location of FCs. Using the Good Jobs FirstSubsidy Tracker Database, I find that between 2010 and 2016, the major e-commerce retailer received $129 million in tax rebates, tax credits, and property tax abatement from various state and local governments[11]. I calculate State Generosity as the logged value of cumulative corporate subsidies offered by the state from 1990 to 2016 (the end of our sample period) scaled by cumulative state revenue.

In particular, the instrument that is composed of cross-sectional variation in the distance between the FCs and the closest Interstate highway entrance and major airports interacting with state-level state subsidy of corporate subsidies follows the Eq 2.3

$$\Delta AFC_{jt} = a + b\Delta TP_{jt} + \sum_{i=1}^{J}\sum_{s=1}^{T} \lambda_i DIST_i \times \mu_s Subsidy_s + \epsilon_{jt} \tag{2.3}$$

Given this specification, the endogenous effect of Amazon FCs openings is identified in Eq 2.2 by using the instruments for exposure to Amazon FCs. Throughout, I report standard errors that cluster on state and year, which are robust to heteroskedasticity of the error across state-year cells, and to spatial autocorrelations across counties within states.

2.4.4 Descriptive Statistics

The CBP data covers all counties in the US. Based on my previous analysis, the data is built only on counties belonging to a major metropolitan area defined by Nielsen. I further filter out counties with populations below 10,000 and 300 annual employees. The descriptive table is shown in Tab 2.2. I further filter out counties that the has less than 10,000 in population[12]

[11]https://www.washingtonpost.com/us-policy/2019/02/16/amazon-paid-no-federal-taxes-billion-profits-last-year/

[12]Minimum population is greater than 9,999.

and 300 in annual employees[13]. The descriptive table is shown in Tab 2.2.

Table 2.2: Descriptive Statistics

Variable	**Mean (SD)**	**Q1**	**Median**	**Q3**
Total Employees	21,774 (49,096)	1,805	5,220	19,795
Total Annual Payroll ($1 million)	508.6 (1,211)	44.7	120.1	429.9
Total Number of Establishments	1,728 (3,477)	312	640	1,624
Number of Establishments (Employee size 1-9)	1200 (2458)	244	468	1,105
Number of Establishments (Employee size 10-99)	475.6 (947.6)	60	156	464
Number of Establishments (Employee size 100-499)	51.7 (101.7)	8	16	52

N = 936

Note: 1. All data are gathered by mid March that year.

The table presents summary statistics for CBP data on establishments and their employment sizes. The variables include Total Employees, Total Annual Payroll, Total Number of Establishments, and the Number of Establishments categorized by employee size (1-9, 10-99, and 100-499). The table displays the Mean and Standard Deviation (SD) values, as well as the first quartile (Q1), median, and third quartile (Q3) for each variable.

On average, there are 21,774 total employees in the observed establishments, with a standard deviation of 49,096. The Q1 value is 1,805, the median is 5,220, and the Q3 value is 19,795. The mean total annual payroll is $508.6 million, with a standard deviation of $1,211 million. The Q1, median, and Q3 values for total annual payroll are $44.7 million, $120.1 million, and $429.9 million, respectively.

The mean total number of establishments is 1,728, with a standard deviation of 3,477. The Q1, median, and Q3 values are 312, 640, and 1,624, respectively. For establishments with employee sizes 1-9, the mean number of establishments is 1,200, with a standard deviation of 2,458, and Q1, median, and Q3 values are 244, 468, and 1,105. Establishments with

[13] Minimum number of annual employee is greater than 299.

employee sizes 10-99 have a mean number of 475.6, a standard deviation of 947.6, and Q1, median, and Q3 values of 60, 156, and 464. Finally, for establishments with employee sizes 100-499, the mean number is 51.7, the standard deviation is 101.7, and the Q1, median, and Q3 values are 8, 16, and 52.[14]

2.4.5 OLS and IV Estimates

We now delve into the findings derived from the Ordinary Least Squares (OLS) and Instrumental Variables (IV) estimates. The OLS estimates illuminate the impact of Amazon Fulfillment Centers (FCs) by comparing the outcomes in counties and years where FCs were established versus those where they were not. On the other hand, the IV estimates elucidate the effect by contrasting outcomes in counties and years with a high versus a low predicted probability of FC openings. Therefore, for instance, a decline in retail employment subsequent to Amazon FC openings will be inferred if retail employment decreased or grew at a slower pace in county-year pairs situated within geographic regions (defined by proximity to the nearest regional airport) and years marked by a higher likelihood of FC openings[15].

The dependent variable in this analysis is the per capita change in retail employment at the county level, while the measure for Amazon FCs is the per capita change in the number of stores. We present estimates excluding and including county-specific time trends, with the latter being the preferred specification. The estimates for retail employment and earnings appear to be largely unaffected by the inclusion of county fixed effects. This observation strengthens the validity of the assumption that the instrumental variables are uncorrelated with the error terms[16].

According to the OLS estimates, an Amazon FC opening correlates with a reduction of

[14]On top of it, the data also shows that there are 149 counties have establishment with more than 500 employees in the facility.

[15]This strategy relies on the assumption that geographic regions near regional airports and years with higher predicted probability of FC openings are more likely to see Amazon FCs. These areas and years are compared to those with less likelihood of FC openings.

[16]This means that the changes in retail employment and earnings observed are not due to unmeasured county-level factors that might also affect the opening of Amazon FCs.

31.3 workers in county employment as per the County Business Patterns (CBP) data. The estimated coefficient for total payrolls per person is positive, implying a 2.3% average payroll increase in the county due to enhanced competition.

Contrastingly, the IV estimates—which, given the validity of the identification strategy, can be interpreted as causal effects of FC openings on retail employment—suggest employment declines in the aggregate retail sector as workers transition to different sectors. In the absence of county-specific trends, the estimates indicate that an Amazon FC opening reduces county-level employment by approximately 175 workers. Incorporating county-specific trends reduces this estimate to 152. On a county basis, this translates to a 2.9% reduction in retail employment attributable to an Amazon FC opening, potentially reflecting a mix of other retail establishment closures and employment reductions at these establishments. It's crucial to understand that these estimates do not signal absolute declines in retail employment but rather indicate that retail employment was lower than it would have been in the absence of FC openings[17].

Regarding worker payroll, the IV estimates consistently indicate a 1.8% increase in average payroll (occasionally significant), which, although smaller, qualitatively aligns with the OLS estimates. It's plausible to expect payroll increases as Amazon's entry empowers workers to bargain and negotiate more effectively. However, the evidence suggesting that the estimated payroll increase is significantly less than the average size of an FC implies that Amazon's presence curtails employment at other retailers in the remaining retail sector. The observation that the OLS estimates of employment effects for the aggregate retail sector are generally negative, and the IV estimates negative, is consistent with the idea that Amazon strategically locates FCs in regions where retail growth is on an upward trajectory.

[17] In other words, these estimates measure the difference between the observed level of retail employment and the level that would have been expected in the absence of Amazon FC openings.

2.5 How Does Amazon Fresh Change the Local Competition?

Chapter 1 of this dissertation centers around assessing the multifaceted impact of Amazon Fresh at the store level. This analysis encompasses an extensive range of competitive parameters including, but not limited to, price, sales volume, product assortment, and product quality. The chapter initiates its discourse by evaluating the competitive landscape at the county level.

The first part of this chapter implements a two-way fixed effects model to study the effect of Amazon Fresh's entry on price and sales volume at county level. This comprehensive analysis enables a broad understanding of how Amazon Fresh's entry into the marketplace affects the dynamics of competition within the county. Through this lens, the introduction of Amazon Fresh appears to heighten competition substantially at the county level.

More specifically, the arrival of Amazon Fresh influences several crucial performance indicators within the retail grocery sector. These include the sales volume, price per ounce, and total revenue of incumbent grocery stores. The effects are multidimensional, and understanding the interaction between these variables provides a more comprehensive view of the competitive impact Amazon Fresh has on the retail grocery landscape.

The goal of this section is to unpack these interactions and provide detailed insights into how Amazon Fresh's entry reshapes the competitive dynamics at the county level, setting the groundwork for a more profound examination of store-level impacts in the subsequent sections of the dissertation. I run regressions of the equations as follows:

$$Log(Sales_{it}) = \alpha_i + \gamma_t + \beta AF_{it} + \theta\varepsilon_{it} \tag{2.4}$$

$$Log(Price_{it}) = \alpha_i + \gamma_t + \beta AF_{it} + \theta\varepsilon_{it} \tag{2.5}$$

$$Log(Revenue_{it}) = \alpha_i + \gamma_t + \beta AF_{it} + \theta\varepsilon_{it} \tag{2.6}$$

In the above equations, my aim is to scrutinize the effects of competition at the county level within the grocery retail sector, using the entry of Amazon Fresh as a case study. These regression structures remain consistent with the methodologies delineated in Chapter 1. For a more comprehensive understanding of these methods, I would recommend referring back to the dedicated section in Chapter 1.

The fundamental difference in this context, however, lies in the application of these regressions. Specifically, I employ these regressions on county-level data, focusing on juice and fresh produce categories. The purpose of this approach is to evaluate the impact of the first Amazon Fresh distribution center on the dynamics of the local grocery retail business. This subtle shift in focus allows us to explore the implications of digital transformation at a broader geographical level, providing a wider perspective on the shifts occurring in the industry.

The results outlined in Table 2.3 clearly illustrate the competitive dynamics introduced by a new fulfillment center (FC) in a county. On average, each existing store within a county that encounters the first-ever Amazon FCwhether in their own county or in the adjacent onesexperiences a decrease in juice sales volume by approximately 4.3% to 10% [18]. These consistent estimates signal the tangible business diversion away from local stores towards the e-commerce behemoth. This shift has significant economic implications, particularly for the stability of local establishments and the state of the local labor market within the affected industry.

Upon applying Equation 2.6, a similar pattern emerges in the context of fresh produce revenue. This metric also experiences a decrease, by approximately 3%. In broader terms, these results suggest that the establishment of an Amazon FC in the vicinity induces a behavioral shift in consumers[19]. It appears that customers, enticed by the convenience and potential cost savings offered by a major e-commerce retailer, begin to migrate some of their

[18] This is a significant decrease, suggesting that the entry of Amazon FC has a considerable impact on the sales of local stores.

[19] This behavioral shift refers to changes in shopping habits, such as opting to shop online rather than in-store.

Table 2.3: County Level Competition Effects

Models	(1)	(2)	(3)	(4)	(5)	(6)
AmazonFresh Entry	-0.1004	-0.0727	-0.0689	-0.0727	-0.0694	-0.0428
Volume(Juice)	(0.0193)	(0.0177)	(0.0170)			(0.0199)
AmazonFresh Entry	-0.0183	0.0112	0.0114	0.0111	0.0113	-0.0073
Price(Juice)	(0.0066)	(0.0079)	(0.0077)			(0.0055)
AmazonFresh Entry	0.0615	-0.0293	-0.0301	-0.0293	-0.0296	-0.0591
Revenue(FP)	(0.0186)	(0.0272)	(0.0272)			(0.0158)
Time FE	No	Yes	Yes	Yes	Yes	Yes
Unit FE	Yes	Yes	Yes	Yes	Yes	Yes
Control	No	No	Yes	No	Yes	No

Note: Model (1) reports estimates with only county fixed effect as in equation 1.5; model (2) reports estimates from TWFE model without covariates; model (3) reports estimates from TWFE with covariates; model (4) reports estimates following Goodman-Bacon, 2021; model (5) reports estimates following Goodman-Bacon, 2021; model (6) reports estimates following Callaway and SantAnna, 2021. SE are in the parenthesis.

purchases away from traditional, local brick-and-mortar stores. This trend underscores the profound and transformative impact of digital commerce on traditional retail landscapes, a phenomenon that warrants deeper investigation, especially given its potential to reshape local economies and communities[20].

2.6 How Does Amazon Fresh Change Consumer Shopping Behavior?

From the perspective of consumers, the introduction of Amazon Fresh will increase consumer welfare as it will bring more options to local shoppers and potentially make the market more

[20]The digital transformation of retail has far-reaching effects beyond just consumer behavior, including impacts on employment, real estate, and local economic health.

competitive. But the process may look completely different from the competition with local grocery stores. As discussed by Huang et al., 2012, the market expansion of Walmart has increased consumer surplus as Walmart takes advantage of the economies of scale to squeeze profits which brings down prices leading to higher surplus. As you can imagine, the new service of Amazon Fresh would also make the local grocery market more competitive thus increase the consumer surplus. As found in chapter 1, the entry of Amazon Fresh leads to local incumbents serving better quality of products with a larger assortment availability, but I have not looked at how consumers' shopping behavior change after the introduction of such service.

In this section, I use Nielsen consumer panel data to answer the question of how Amazon Fresh changes the shopping behavior. I estimate similar equation as in Eq 2.4 by replacing the dependent variable with average consumer consumer spending per visit and average number of visits during a quarter from a quarterly frequency data. I also apply a

$$Y_{i,t} = \alpha_i + \gamma_t + \beta_k^{-K}\ AF_{i,t}^{<-K} + \sum_{k=-K}^{-2} \beta_k^{lead} AF_{i,t}^{k} + \sum_{k=0}^{L} \beta_k^{lag} AF_{i,t}^{k} + \beta_k^{L+} AF_{i,t}^{>L} + \varepsilon_{i,t} \tag{2.7}$$

With K = 4 to capture the year before and after the entry of Amazon Fresh on local shoppers' behavior. The two way fixed effects model[21] indicates the average spending of consumers drop by 8.4 per visit. The results in Table 2.4

In the context of the event study, the findings indicate that Amazon Fresh's entry into local markets significantly influences consumers' average spending per household on juice per quarter. The coefficients, especially the indicators for the pre and post periods, suggest a downward trend in average spending both before and after Amazon Fresh's debut. The pre-period coefficient of -12.7 implies that there was a decrease of $12.7 in the average household spending on juice in the quarters leading up to Amazon Fresh's entry, relative to the quarter of entry. Furthermore, the -14.5 post-period coefficient signifies that this trend endures

[21]Standard errors are following Stata type and clustered at designated market level defined by Nielsen.

Table 2.4: County Level Competitive Effects

	Estimate	Std. Error	t-value	p-value
Pre	-12.678	17.755	-0.714	0.475
‘rel_quarter_-4‘	-17.419	21.986	-0.792	0.428
‘rel_quarter_-3‘	-8.942	21.323	-0.419	0.675
‘rel_quarter_-2‘	-23.576	21.467	-1.098	0.272
rel_quarter_0	-8.274	21.467	-0.385	0.700
rel_quarter_1	-21.069	21.555	-0.977	0.328
rel_quarter_2	-19.047	22.564	-0.844	0.399
rel_quarter_3	-29.507	23.542	-1.253	0.210
rel_quarter_4	-15.009	23.654	-0.635	0.526
Post	-14.477	19.013	-0.761	0.446
Multiple R-squared(full model): 0.6058, Adjusted R-squared: 0.5824				
F-statistic(full model): 25.89 on 260 and 4379 DF, p-value: 2.2e-16				

even after Amazon Fresh has established its presence, underlining a lasting impact on the performance of local stores.

The coefficients linked with the ”rel_quater” variables offer a more detailed perspective on the impact on average spending at specific periods around Amazon Fresh’s entry. For instance, the negative coefficient for in two quarters before the entry suggests a $23.6 decrease in average spending two months prior to Amazon Fresh’s entry. However, the relatively high p-values linked with these coefficients imply that these specific monthly effects are not statistically significant.[22]

Nevertheless, the F-statistic and the associated p-value for the projected model suggest that the event-specific (quarterly) effects are not statistically significant. This implies that while the overall trend of reduced average spending associated with Amazon Fresh’s entry

[22]The R-squared value for the full model indicates that the model, which takes into account Amazon Fresh’s market entry among other factors, can explain approximately 60.6% of the variation in average spending on juice. This significant proportion signals a good fit of the model to the data.

is evident, the specific timing effects around the event are less certain. In general, these results suggest that Amazon Fresh's entry has a substantial negative impact on the average household spending on juice at local stores. However, subsequent research could delve deeper into the temporal effects and other potential confounding factors to provide a more nuanced understanding of this impact.

2.7 Conclusion

This chapter broadens the exploration of digital transformation at the county level, focusing on three critical areas: the labor market, the retail market, and consumer shopping behavior. Each of these areas has experienced distinct yet interconnected shifts due to the digitalization brought about by entities such as Amazon.

In the labor market, I find that the advent of Amazon Fulfillment Centers (FCs) corresponds with an increase in workers' salaries. However, this benefit is tempered by a reduction in the workforce within the retail industry. This trend could be attributed to Amazon's competitive pull, attracting workers away from traditional retail. Alternatively, it could be a consequence of incumbent stores closing as they struggle to compete in the increasingly digital marketplace.

Turning to the county-level retail grocery market, I observe a decrease in juice volume, indicating a potential shift in consumer purchasing patterns. Price changes in this market present an inconsistent picture, suggesting complex underlying dynamics. Generally, the revenue of incumbent stores appears to be on a downward trajectory, which could be a contributing factor to the observed labor market changes.

Lastly, when examining consumer behavior, a clear trend towards digital transformation emerges. Consumers are gradually transitioning towards online shopping, reducing both the frequency of their physical store visits and their in-store spending. This shift signifies an adaptation to the digital market, with consumers taking advantage of the convenience and efficiency offered by online platforms like Amazon.

In conclusion, the digital transformation catalyzed by Amazon and similar entities has multifaceted and profound impacts on county-level markets. While it brings some advantages, such as higher salaries for workers, it also presents challenges, including a reduction in the retail workforce and lower revenues for traditional stores. As consumers continue to adapt to this new digital era, it will be crucial to understand these changes further and develop strategies to navigate this evolving landscape successfully.

Chapter 3

DEMAND ESTIMATION USING DEEP FEATURES

3.1 Introduction

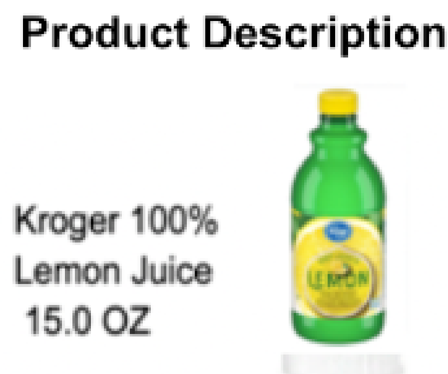
In recent years, the increasing availability of data has opened new avenues for enhancing existing econometric models. With the proliferation of internet access, consumer behavior has shifted towards browsing and comparing products online before making purchases, whether in-store or online[1]. This trend has made it imperative for economists to integrate data from both online and offline sources into their analyses. Consider the modern shopping experience: it is highly likely that you'll encounter attractive product images and detailed descriptions that influence your purchasing decisions. This chapter aims to leverage these everyday experiences by incorporating image and text information into demand estimation models.

Traditionally, demand estimation models have primarily relied on economist-derived characteristics such as price, flavor, and brand. These discrete choice models attempt to predict consumer preferences but often fall short of capturing the rich and diverse array of characteristics influencing consumer choices in today's digital shopping landscape. By integrating image and text data, researchers can create more comprehensive and accurate representations of product characteristics, leading to improved demand estimations. Images and text descriptions are a valuable resource for uncovering the nuances of product aesthetics, branding, and other qualitative factors influencing consumer decisions. For example, images offer a wealth of visual information that can highlight key design elements and visual cues resonating with consumers. On the other hand, text descriptions provide insights into market

[1]Cavallo, 2017 finds that online prices are identical to their offline counterparts about 70% of the time (on average across countries), thereby justifying the treatment of both on-line and off-line retail sector as the single sector. This indicates online retails do have a significant impact on off-line retail.

positioning and unique selling points that distinguish products from their competitors.

Figure 3.1: An example of information consumers see from juice products online



Moreover, incorporating image and text descriptions into demand estimation models enables researchers to explore complex relationships and interactions between product characteristics. This inclusion can lead to the discovery of previously overlooked patterns and trends, ultimately improving the accuracy and predictive power of demand estimations. By leveraging advancements in machine learning and deep learning techniques, researchers can efficiently process and analyze large amounts of image and text data, further enhancing the potential of these models.

In this study, we use sales data from local grocery stores to construct a demand model for juice products. This model will then be used to estimate product elasticities. We generate product attributes from text and images using state-of-the-art deep learning models and combine these attributes with other characteristics to build the demand model.

Specifically, we utilize a pre-trained ResNet50[2] model to generate embeddings for juice

[2]ResNet50 is one of the most popular neural networks to extract image features. We also experiment the performance from various image extraction techniques. However, as the depth of CNNs increases, they face the problem of vanishing gradients, which can lead to degraded performance during training. To address this challenge, He proposed the Residual Network (ResNet) architecture in 2015, which introduced a novel skip-connection mechanism that enables the network to learn residual functions instead of the underlying

product images and a sentence transformer (Reimers & Gurevych, 2019) model to convert word descriptions. In the second phase, we construct the demand function using features generated from these embeddings. We apply various methods including linear regression, logit, random forest, xgboost, double machine learning, and neural networks to estimate the quantity of juice sold.

Our research reveals a significant enhancement in model performance upon the integration of image and text features into demand estimation models. This integration leads to a substantial reduction in training loss, indicating a more accurate and reliable model. It implies that the model is better equipped to learn from the data and make accurate predictions, a benefit directly attributable to the incorporation of these additional features.

Furthermore, the integration of image and text features results in the estimation of elasticities that are more aligned with those found in existing literature. This finding underscores the value of using image and text data in improving the validity of our model. The closer alignment with established findings lends further credibility to our approach, suggesting that the inclusion of these additional features allows for a more comprehensive and nuanced understanding of the factors that influence demand.

By incorporating image and text features into demand estimation models, we are effectively bridging the gap between traditional econometrics and the vast, untapped potential of unstructured data. The contribution of this paper is three-fold. Firstly, it pioneers the integration of image and text data into traditional demand estimation models, thus offering a more holistic approach to predicting consumer preferences. Secondly, it utilizes a variety of machine learning and deep learning techniques to handle the high dimensionality and complexity of image and text data. Finally, it provides a comprehensive comparison of various models' performance in demand estimation, contributing to the ongoing discussion on the intersection of machine learning and econometrics in demand estimation.

The remainder of the paper is structured as follows. Section 2 discusses the relevant

---

mapping directly.

literature on demand estimation, double machine learning, and deep learning, forming the foundation for our methodology. Section 3 introduces the dataset and outlines our feature engineering process. In Section 4, we detail our models, while Section 5 delves into the empirical performance of various configurations. Finally, Section 6 summarizes our findings and proposes directions for future research.

3.2 Literature Review

3.2.1 Demand Estimation

Demand estimation serves as a crucial tool in contemporary Empirical Industrial Organization (EIO) literature for the analysis of various markets with limited data. Berry and Haile, 2021a highlights several advantages of demand and supply estimation. Firstly, it enables economists to deduce demand elasticities[3], which prove valuable for assessing market power, competition, and welfare implications of different policies or interventions. Secondly, it permits the calculation of market markups for firms, reflecting the extent to which firms charge above their marginal costs and unveiling sources of market power, such as product differentiation, entry barriers, or antitrust issues. Thirdly, it facilitates counterfactual analysis by simulating market responses to various scenarios like mergers, regulations, or innovations.

Moreover, Aguirregabiria, 2018 asserts that demand estimation helps overcome data limitations frequently encountered in IO research. For instance, while we may lack direct information about firms' costs or production functions, we can estimate them indirectly using first-order conditions connecting marginal costs and marginal revenues. Likewise, in the absence of detailed data on consumer preferences or characteristics, as in Gandhi and Nevo, 2021, we can estimate them indirectly employing discrete choice models that associate consumers' choices with observable product attributes.

Demand estimation is applicable to a broad array of industries and markets involving differentiated products, such as automobiles, airlines, telecommunications, healthcare, re-

[3]Elasticities measure consumer sensitivity to changes in prices or other determinants of their choices.

tailing, media, and many more. Estimating demand and supply models for these industries offers insights into consumer valuation of different product features, competition among firms based on prices and qualities, and the impact of market structure on efficiency and innovation outcomes.

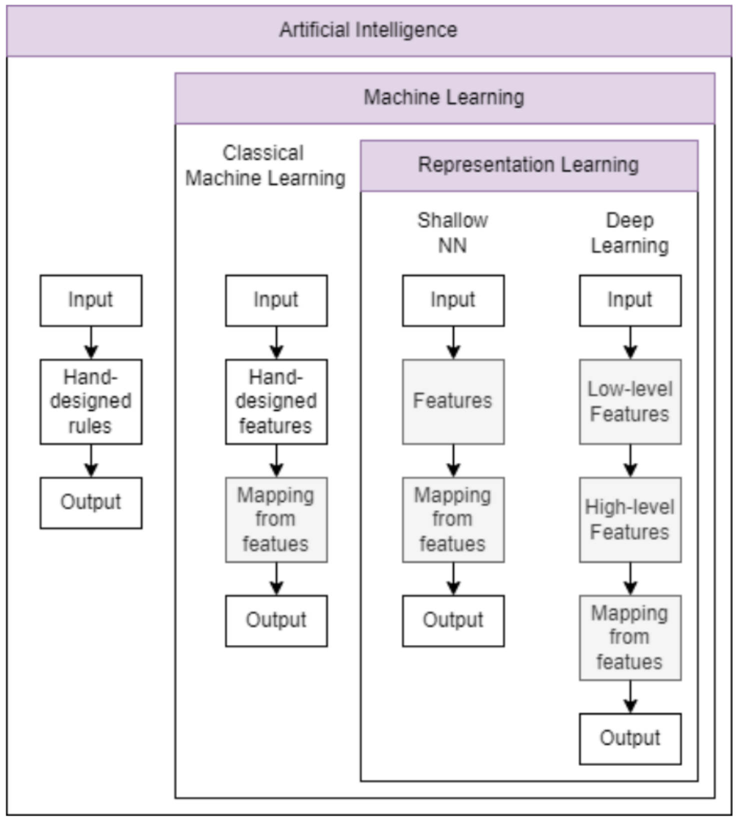
In this study, we aim to compare traditional demand models with machine learning and deep learning techniques to investigate their potential in contemporary empirical IO research.

3.2.2 Deep Learning

Speaking of which, deep learning, which has gained significant attention in recent years due to its ability to analyze large and complex data sets nowadays, is used in many aspects to extract information from images and text descriptions that consumers observe everyday. One of the key features of deep learning is its use of deep convolutional nets, which have brought about breakthroughs in processing images, video, speech and audio, whereas recurrent nets have shone light on sequential data such as text and speechLeCun et al., 2015.

One of the most powerful applications of DL is in the area of feature extraction. This involves the automatic identification of patterns and features in data, which can then be used to train models for a wide range of tasks. In the case of natural language processing (NLP), DL models can be trained to automatically extract features from text data, such as sentiment, topic, and syntax. This has led to significant advances in areas such as text classification, sentiment analysis, and machine translation. In the case of image processing, DL has revolutionized the field of computer vision. DL models can be trained to automatically extract features from images, such as edges, textures, and shapes. This has led to significant improvements in tasks such as image recognition, object detection, and facial recognition. Overall, DL has the potential to revolutionize a wide range of fields by automating the process of feature extraction and allowing models to learn and adapt to complex patterns in data. As such, it is a rapidly growing field with significant potential for future developments and applications.

Figure 3.2: Relationships Between AI, ML and DL. (Liu, 2023)



3.2.3 Double Machine Learning

Double machine learning (DML) is a novel method for estimating causal effects of treatments or policies when there are many potential confounders that may affect both the treatment assignment and the outcome. DML combines the advantages of machine learning (ML) techniques, such as flexibility and scalability, with the rigor of causal inference methods, such as orthogonality and consistency. DML can handle high-dimensional and non-parametric settings where traditional methods may fail or be impractical.

The main idea of DML is to use ML methods to estimate nuisance parameters, such as the conditional expectation of the outcome given the confounders and the conditional probability of treatment given the confounders, and then use orthogonalization techniques to obtain unbiased and efficient estimates of the causal parameters. DML can also estimate heterogeneous treatment effects by allowing for interactions between the treatment and covariates. DML has been developed in a series of papers by Chernozhukov et alChernozhukov et al., 2016 and has been applied to various domains such as economics, health care, and social sciences.

Orthogonalization techniques are methods for removing the influence of confounders from the estimation of causal effects. They can be seen as a generalization of regression adjustment, where instead of directly regressing the outcome on the treatment and confounders, we first orthogonalize both the outcome and the treatment with respect to the confounders. This means that we remove any linear or nonlinear dependence of the outcome and treatment on the confounders by subtracting their conditional expectations given by ML models. Then we regress the orthogonalized outcome on the orthogonalized treatment to obtain an unbiased estimate of the causal effect.

Orthogonalization techniques have several advantages over standard regression adjustment. They can handle high-dimensional and non-parametric settings where regression adjustment may suffer from overfitting or misspecification. They can also estimate heterogeneous treatment effects by allowing for interactions between the orthogonalized treatment and covariates. Moreover, they can be easily implemented using any ML method for estimating nuisance parameters, such as random forests, neural networks, or lasso.

3.3 Data

In this study, we focus on juice products as a case example to apply our methods, utilizing a comprehensive dataset sourced from Nielsen's consumer panel database.[4] Our analysis targets sales data for 717 juice products with images and text description available. The Nielsen dataset provides a wealth of information, including a large percentage of price data

[4]The dataset encompasses weekly scanning data collected from over 100,000 households across 208 markets in 23 cities throughout the United States. These households are equipped with handheld scanners by AC Nielsen to record barcodes of every product they purchase. The data primarily covers purchases from food stores, drug stores, mass-merchandise stores, and convenience stores.

embedded within the scanner data and a smaller portion in the Consumer Panel Data. To obtain a more complete view of price information, along with quantity and market characteristics for the juice products, we combine both Panel and Scanner data sources. This integration enables us to access information on the month, price, volume, product features, and market conditions for each juice product.

Variable	**Mean**	**Min**	**1st Q**	**Median**	**2nd Q**	**Max**
Volume (Oz)	195.3	2.7	50.7	104.8	222.6	20289
Price ($ per Oz)	0.148	0.015	0.043	0.086	0.200	1.776

Table 3.1: Summary Statistics of Juice Scanner Data

The word description data for the juice products are also sourced from Nielsen, available as part of the additional product information provided within the consumer panel database. These textual descriptions offer insights into the marketing strategies, unique selling points, and various attributes of the products, such as flavor profiles, nutritional content, and packaging details.

By incorporating this textual data into our analysis[5], we can extract meaningful features through deep learning techniques, such as Sentence Transformers, which allow us to create high-dimensional word vectors representing the semantic content of the descriptions. These word vectors can then be combined with the image features and other numerical data to create a richer and more comprehensive set of inputs for our demand estimation models.

In addition to the data provided by Nielsen, we also collect product images for the juice products by web scraping from major retailers. This is done by matching the product barcodes (UPC) in the Nielsen dataset with the corresponding product listings on these retail websites. This image data enriches our dataset with visual information, which can then be used to extract image features through deep learning techniques, such as ResNet (He et al.,

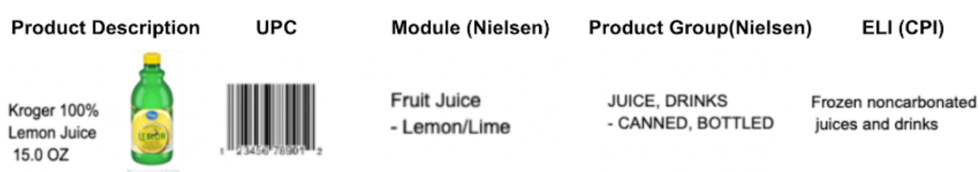
[5]Top descriptions are shown in 3.2.

Variables	Top 3 Frequent Description
Brand	GATORADE
	V8
	WELCH's
Bottle Size (OZ)	20
	32
	64
Product Module	FRUIT DRINKS-OTHER CONTAINER
	FRUIT JUICE-REMAINING
	VEGETABLE JUICE AND DRINK REMAINING

Table 3.2: Summary of Category Variables

2015). [6]

Figure 3.3: Grabing Data Online



Our dataset includes price, volume, market characteristics, and visual features for the

[6]As we proceed with our data analysis, it will be necessary to keep track of three levels of aggregation. It is easiest to understand these levels of aggregation by means of an example drawn from our data. At the lowest level we have a product which we identify using the UPC. For example, a bottle of "Kroger 100% Lemon Juice 15.0 oz" (this product description and the photo are scraped from an online website) has a UPC of 1111070205. Each UPC in turn belongs to a module in the Nielsen dataset "Fruit Juice - Lemon/Lime." At the highest level, we have a "product group," which in this case would be "Juice, Drinks - Canned, bottled," which contains not just lemon juice, but also other modules like "Soft Drinks. At the same time, we use the product group description and product module description to match with the CPI entry level item (ELI).

juice market and also integrates word description data. The combination of these diverse data sources enables us to capture a more holistic understanding of the factors influencing consumer preferences in the juice market, ultimately enhancing the accuracy and predictive power of our demand estimation models.

3.4 Feature Extraction with Deep Learning

Deep learning, with its capacity to undertake intricate tasks with minimal human supervision, has emerged as a transformative technology. This development has stimulated an increasing interest in incorporating deep learning techniques into traditional econometric models, such as those employed for demand estimation. In this study, we aim to leverage these techniques to extract detailed and meaningful features from online product descriptions and images, using these attributes to estimate demand elasticities for a variety of products.

Online product information presents a current and diverse dataset that can enhance the precision and predictive capability of demand estimation models. By integrating these supplemental data sources, our objective is to formulate a more comprehensive depiction of product characteristics influencing consumer decisions. Specifically, we utilize deep learning algorithms to process textual product descriptions and analyze images. This approach allows us to capture nuanced aspects of branding, design, and other qualitative factors often neglected in conventional numerical models.

In pursuit of this objective, we employ pre-trained deep learning models, including Residual Networks (ResNet) for image feature extraction and Sentence Transformers for textual feature extraction. These models are recognized for their efficacy in extracting high-level semantic information from input data. By amalgamating these extracted features with traditional numerical data, we generate a richer set of inputs for our demand models. This approach enables us to investigate complex relationships and interactions between product characteristics and consumer preferences, ultimately offering a more robust and nuanced analysis of market demand.

3.4.1 Word Features Extraction

Word2vec is a technique employed for generating vector representations of words that capture their semantic and syntactic properties. It utilizes a two-layer neural network architecture, which takes a substantial volume of text as input and outputs a set of fixed-length vectors for each word. This results in a generic numeric vector for web descriptions gathered for each product. To augment the quality of vectors, facilitating the categorization and classification of various product types, we employ sentence2vec (Pagliardini et al., 2018). This method generates vector representations of sentences based on word2vec. The advantage of sentence2vec lies in its ability to understand semantic similarities between sentences by integrating word vectors in diverse ways. This results in more accurate and meaningful sentence vectors that can be used to summarize information, thus serving our objective of demand estimation.

In this study, we generate our text features following the sentence2vec process. We employ sent2vec[7], a Python library that provides a plethora of options for word and sentence embeddings. We select Word2Vec for our word embedding method and averaging for our sentence embedding method. Our chosen pre-trained model is $sent2vec_wiki_bigrams$[8], sourced from the package. We apply the sentence embedding model to each product's web description, obtaining a 788-dimensional vector representation for each sentence. These vectors subsequently serve as text characteristics for demand estimation using various methods.

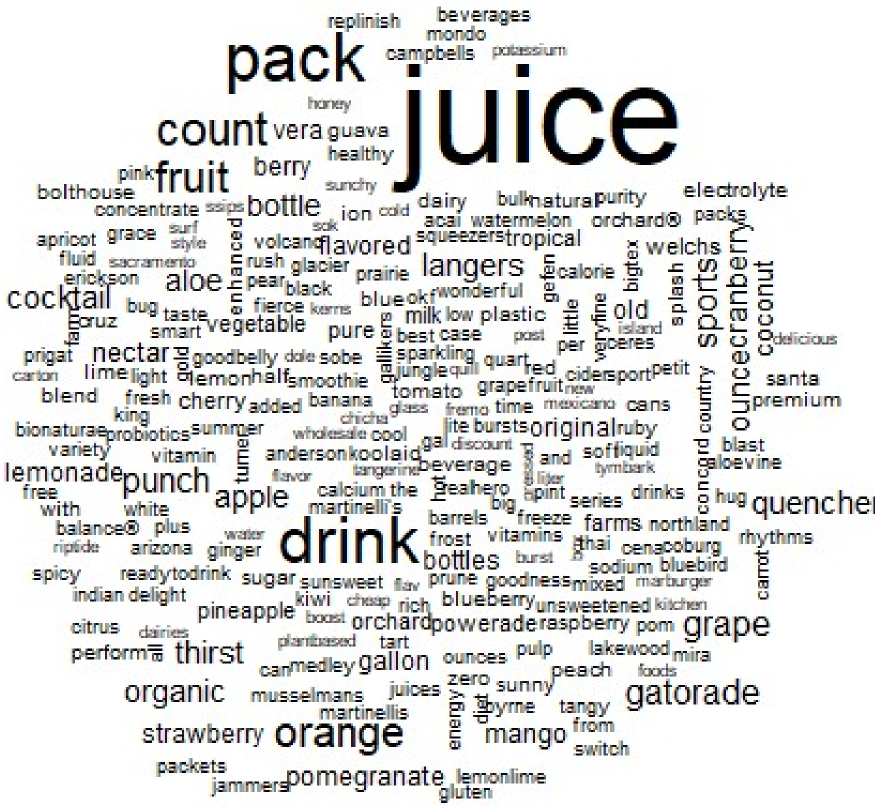
The sentence2vec technique operates through a flow of processes. Starting with a large corpus of text, it first creates a vocabulary of words, assigning a unique vector to each word. It then scans the text using a sliding window, predicting each word based on its context (surrounding words) in the window. Through this process, the vectors are adjusted such that semantically similar words end up with similar vectors. This word embedding model is then used to generate sentence embeddings. A sentence embedding is typically created

---

[7]If you are interested, visit https://pypi.org/project/sent2vec/ for more details.

[8]A bigram feature constitutes a pair of words that occur together, such as I ate, ate banana, or banana bread. These are trained on a large corpus of text spanning various domains

Figure 3.4: The Frequency of Words in the Data



by combining the word description of products. This could be done through averaging, summation, or other complex methods like considering the order of words. These sentence embeddings are then used as input features for our demand models.[9]

3.4.2 Image Features Extraction

ResNet introduced a breakthrough concept that residual blocks can be used to improve the performance of very deep neural networks. These residual blocks allow the network to learn

[9]The flowchart of sentence2vec is presented in the appendix **??**.

a residual mapping, which is then added to the original input to produce the output of the block. This mechanism enables the network to pass the gradient directly from the output to the input, alleviating the vanishing gradient problem and enabling the network to effectively train much deeper architectures.

ResNet can be employed to create image feature vectors that capture high-level semantic information about the input images. By leveraging the pretrained ResNet models, we can extract feature vectors from one of the intermediate or final layers of the network. These feature vectors can then be used as input for various machine learning tasks, such as demand estimation.

We want to estimate the demand for different products based on their characteristics in images. For example, we want to know how many customers would buy a certain juice based on how it looks. To do this, we need to convert the images into numerical features that can be used in our estimation model and the method is introduced in Im2VecReddy et al., 2021.

It is a method that can convert raster images into vector graphics that are scalable, editable, and compact. It uses a neural network that consists of an encoder-decoder architecture with attention and a differentiable renderer. The encoder-decoder learns to predict a sequence of vector commands that can reconstruct the input image. The differentiable renderer converts the vector commands into a raster image that can be compared with the input image using a loss function. Im2Vec does not require any vector supervision, meaning it does not need any ground-truth vector graphics for training. It can learn from any raster dataset. Im2Vec can be applied to various tasks such as vectorization, editing, interpolation, and generation of fonts, logos, emojis, icons, and other graphic designs. It can produce high-quality vector graphics that preserve the details and topology of the input images. It can also handle complex shapes and curves that are challenging for existing methods.

We follow the below process to create vectors of image features:

- Obtain a pretrained ResNet, Im2Vec model (e.g., ResNet-50) from a deep learning library, such as TensorFlow or PyTorch.

- Prepare the input images for feature extraction by resizing and normalizing them according to the requirements of the pretrained ResNet model.

- Pass the input images through the pretrained ResNet model up to a chosen intermediate or final layer (e.g., the last average pooling layer or a fully connected layer) to obtain the feature vectors.

- Optionally, apply dimensionality reduction techniques (e.g., PCA) to reduce the size of the feature vectors while retaining most of the relevant information.

- Use the extracted image feature vectors as input for the demand estimation model, either alone or in combination with other features (e.g., word descriptions, numerical features).

The intial dimensions of vectors from ResNet and Im2Vec are 2048 for each product image.[10] These vector embeddings capture the visual characteristics of the products that influence customer preferences. By using these embeddings as inputs in our estimation model, we can better predict the demand for each product.

3.5 Model

Our objective is to discern the most effective methodologies for incorporating information, derived from deep learning, into traditional demand estimation procedures. Guided by Bajari et al., 2015, we initiate our analysis by juxtaposing econometric models with machine learning models in a standard demand estimation setup, using juice product data for illustrative purposes. This juxtaposition permits us to evaluate the aptitude and limitations of each approach in capturing the impact of various product characteristics on consumer preferences.

[10]These embeddings are further compressed into 128, 256, 512, 1024 due to test the performance on various dimensions.

As we broaden our analytical framework, we delve into advanced techniques such as double machine learning (Chernozhukov et al., 2016). This methodology is specifically formulated to tackle challenges associated with the increased dimensionality and sparsity of data[11]. Additionally, we contemplate the application of other cutting-edge techniques, such as convolutional neural networks and deep instrumental variables (Deep IV) (Hartford et al., 2017), to further probe the potential of demand estimation models.

Let's now establish the fundamental structure of characteristic demand estimation models. Suppose there exist J competing products in the markets, each possessing observable characteristics X_j. Let the demand for product j in market m at time t be represented by a general log form of demand function:

$$\log Q_{jhmt} = f(p_{mt}, X_{mt}, D_{mt}, \epsilon_{jmt}) \tag{3.1}$$

Here, X is the matrix of observed characteristics, p is the vector of logged prices, D represents a vector of demographics for each market at time t, and ϵ is an idiosyncratic shock. This specification is very general, allowing for nesting via the stratification of the error term. Assuming there are H nests of products; for instance, in the juice example, two nests could be organic and regular juices. By grouping alternatives into nests and accounting for the correlation within these nests, we allow substitution patterns to vary in a reduced-form way across different classes of products.

One significant hurdle with this approach is that the characteristics space is typically manually crafted, which might not comprehensively represent the product features. With the assistance of deep learning, we aim to integrate this method into traditional demand estimation. We accommodate various setups to Eq3.1 to investigate the model performance with the inclusion of additional features.

[11]The regression may encompass thousands of variables on one side.

3.5.1 Linear Regression

We start our analysis with a linear approach, where we approximate Equation 1 using a flexible linear form as follows:

$$\log Q_{jmt} = \alpha p_{mt} + \beta X_{mt} + \gamma D_{mt} + \epsilon_{jmt} + \eta_{hm} + \xi_{mt} \tag{3.2}$$

In this equation, the matrix X represents the high-dimensional and potentially sparse data that encompasses information from dummy variables categorizing product attributes such as category and flavor. Furthermore, the dimensions of the final embeddings derived from image and word data can vary significantly, ranging from 128 to as high as 2048[12].

In our linear regression model, we also incorporate nested structures through the term η_{hm}, which captures the influence of unobserved product characteristics shared by products within the same nest. Additionally, we account for seasonality effects by including the term ξ_{mt}, which captures variations in demand patterns across different time periods.

By employing this linear approach, we aim to assess the effectiveness of a relatively simple model in capturing the complex relationships between product characteristics and consumer preferences. This baseline analysis will serve as a foundation for our subsequent exploration of more advanced modeling techniques, allowing us to evaluate their added value in the context of demand estimation.

3.5.2 Logit

As discussed in the literature review, there is a large strand of papers focus on logit models. Starting with traditional BLP style (Berry et al., 1995) to Conlon and Mortimer, 2013's work to allow incomplete product availability. They are widely used to model discrete choice situation with limited data observed by econometricians. The error term is assumed to follow a Type I Extreme Value distribution.

[12]This upper bound could be even larger depending on the structure of the deep learning neural network.

The BLP model is a widely used discrete choice model in the industrial organization literature for estimating demand in differentiated product markets. The model extends the multinomial logit framework by incorporating random coefficients, which capture consumer heterogeneity and allow for flexible substitution patterns. The utility of individual i for product j in market t is given by:

$$u_{ijt} = \delta_{jt} + \mu_{ijt} + \epsilon_{ijt}, \tag{3.3}$$

The market share s_{jt} of product j in market t is derived by integrating the individual-specific probabilities of choosing product j over the distribution of consumer heterogeneity:

$$s_{jt} = \int \frac{\exp(u_{ijt})}{1 + \sum_{k=1}^{J} \exp(u_{ikt})} dF(\nu_i, \xi_i), \tag{3.4}$$

where u_{ijt} is the utility of individual i for product j in market t, J is the number of products, and $F(\nu_i, \xi_i)$ is the joint distribution of individual-specific random coefficients (ν_i and ξ_i).

To calculate the quantity demanded for each product in a demand estimation model, you first need to estimate the market shares s_{jt} of each product using the market share equation. Once you have the market shares, you can multiply them by the total market size M_t to obtain the quantity demanded for each product:

$$q_{jt} = s_{jt} \cdot M_t. \tag{3.5}$$

The total market size M_t can be measured by the total number of consumers, total sales, or total expenditure in market t.[13]

3.5.3 Feature Selection under regularization

When the dimension of right hand side is high. In machine learning they usually add feature selection technique to assign weights on different features. Lasso and Ridge are regularization

[13] The whole process is built upon Conlon's pyBLP package.

techniques used in linear regression to prevent overfitting and improve model generalization. They add a penalty term to the linear regression objective function. Here are the objective functions for Lasso and Ridge regression[14]. The objective function for Lasso regression is given by:

$$\min_{\beta_0,\beta} \sum_{i=1}^{n}(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} |\beta_j|, \tag{3.6}$$

where y_i is the observed response for the i-th observation, β_0 is the intercept, x_{ij} is the value of the j-th predictor for the i-th observation, β_j is the coefficient of the j-th predictor, n is the number of observations, p is the number of predictors, and λ is the regularization parameter that controls the amount of shrinkage applied to the coefficients.

3.5.4 Ensemble Methods

Random Forest, Boosting, and Bagging are popular ensemble learning methods used in various prediction and classification tasks, including demand estimation. These methods combine the outputs of multiple base models to improve overall model performance, reduce overfitting, and increase generalization capabilities.

Random Forest is an ensemble method that constructs a multitude of decision trees and combines their predictions through a majority voting mechanism for classification tasks or by averaging for regression tasks. In demand estimation, Random Forest can be used to model non-linear relationships between the demand and various explanatory variables, capturing complex interactions and providing robust predictions. The method's ability to handle a large number of variables with different scales, missing data, and outliers makes it particularly suitable for practical demand estimation problems.

Given the input features p_{mt} (price), X_{mt} (product characteristics matrix), and D_{mt} (image and word embeddings), as well as the nested structure η_{hm} and seasonality ξ_{mt}, we

[14]The main difference between Lasso and Ridge regression lies in the penalty term: Lasso uses an L_1 penalty (absolute values of coefficients), while Ridge uses an L_2 penalty (squared values of coefficients).

can train a Random Forest model to predict the log quantity $\log Q_{jmt}$.

For each decision tree in the Random Forest ensemble, we would:

- Select a random bootstrap sample from the training data.
- Grow the decision tree by recursively splitting the nodes based on the input features, while randomly selecting a subset of features at each split.
- Continue the process until the tree reaches a predefined depth or a minimum number of samples per leaf node.

Boosting is an iterative ensemble method that combines multiple weak learners (usually decision trees) to create a strong learner. The main idea behind boosting is to sequentially train base models by assigning higher weights to instances that were misclassified or had larger errors in previous iterations. In the context of demand estimation, boosting algorithms like AdaBoost or Gradient Boosting can be applied to model complex, non-linear relationships and interactions between variables, improving prediction accuracy and generalization performance compared to single models.

Bagging (Bootstrap Aggregating) is an ensemble method that trains multiple base models (commonly decision trees) on different subsets of the data, obtained by bootstrapping (random sampling with replacement). The final prediction is obtained by averaging the predictions of all base models for regression tasks or by majority voting for classification tasks. In demand estimation, Bagging can be used to reduce the variance and overfitting of single decision trees or other base models, resulting in more accurate and stable demand predictions.

Overall, these ensemble methods offer valuable tools for demand estimation, as they can handle complex and high-dimensional data, capture non-linear relationships, and provide more robust and accurate predictions than single models. Their versatility and robustness make them particularly suitable for real-world demand estimation problems with diverse data sources, noise, and uncertainty.

3.5.5 Double Machine Learning

Double machine learning (DML) can be employed to estimate causal effects in the presence of many potential confounders. In the context of this study, we can formulate a partially linear model using the DML framework. The advantage of using double machine learning is that we can utilize all different regressors to estimate the model. Let P_{mt} denote the treatment variable prices that would want to estimate elasticity, and W_{jmt} denote the confounders, which could include product characteristics, image and word embeddings, nested structure, and seasonality. The partially linear model can be written as:

$$\log Q_{jmt} = \alpha P_{jmt} + g(W_{jmt}) + \epsilon_{jmt}, \tag{3.7}$$

$$P_{jmt} = m(W_{jmt}) + \xi_{jmt} \tag{3.8}$$

where α represents the causal effect of the treatment variable P_{mt} on the log quantity $\log Q_{jmt}$, $g(W_{jmt})$ is an unknown function of the confounders, and ϵ_{jmt} and ξ_{jmt} are error terms. The goal is to estimate the causal parameter α.

To implement the DML method, we split the data into K cross-validation folds and proceed with the following steps for each fold.

- Train ML models to estimate the nuisance parameters: $E[\log Q_{jmt}|W_{jmt}]$ and $E[P_{jmt}|W_{jmt}]$. Use the remaining $K-1$ folds as the training set and the current fold as the validation set.

- Orthogonalize the log quantity and price with respect to the confounders by subtracting their estimated conditional expectations, i.e., compute $\tilde{Q}jmt = \log Qjmt - \hat{E}[\log Q_{jmt}|W_{jmt}]$ and $\tilde{P}jmt = Pjmt - \hat{E}[P_{jmt}|W_{jmt}]$.

- Regress the orthogonalized log quantity $\tilde{Q}jmt$ on the orthogonalized treatment variable $\tilde{P}jmt$ to obtain an unbiased estimate of the causal effect $\hat{\alpha}_k$.

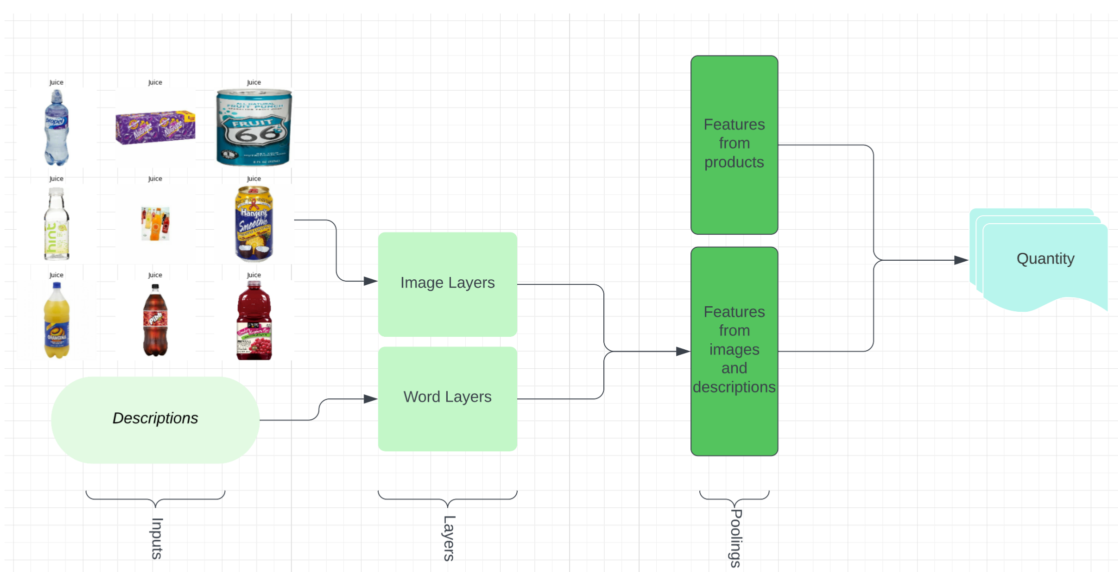
- Average the estimates $\hat{\alpha}k$ across all K folds to obtain the final DML estimate of the causal effect: $\hat{\alpha}DML = \frac{1}{K}\sum_{k=1}^{K}\hat{\alpha}_k$.

By following this procedure, we can efficiently estimate the causal effect of the treatment variable on the log quantity, while accounting for the influence of the confounders and handling high-dimensional and non-parametric settings.

3.5.6 Neural Network

In order to capture the potentially complex and non-linear relationships within our dataset, we opted to utilize a Neural Network (NN) for our analysis. The choice of NN is motivated by its capability to approximate any function, given a sufficient amount of data and a suitable architecture. The flexibility of NNs enables us to uncover underlying structures that traditional linear models might overlook. However, we acknowledge that this flexibility comes with a risk of overfitting, which we have attempted to mitigate through various strategies described later in this section.

Figure 3.5: Neural Network Structure



Our dataset comprises of N observations, each with D features. These features include numerical data such as ..., categorical data like ..., and high-dimensional inputs derived from images and text descriptions of the juice products. Prior to inputting the data into our model, we performed several preprocessing steps. The numerical data were standardized, and the categorical data were one-hot encoded. Text descriptions were converted to word vectors using the Word2Vec model, while images were processed through a pre-trained Convolutional Neural Network to extract feature vectors.

Our NN comprises of multiple layers designed to handle different types of input data. The first part of the network is made up of dense layers that process numerical and categorical data. The second part consists of two branches, each made up of multiple layers designed to process image and text vectors. The image and text branches incorporate dropout layers to prevent overfitting. The outputs of all branches are then concatenated and passed through additional dense layers, culminating in a single output neuron.

We trained the NN using the mean squared error (MSE) loss function, which is suitable for our regression task. The Adam optimizer was used for its efficiency and low memory requirements. We used a portion of the data as a validation set during training to monitor the model's performance and prevent overfitting. The model's architecture and hyperparameters were selected based on their performance on the validation set.

Model performance was evaluated using both the Mean Squared Error (MSE) and Mean Absolute Error (MAE) on a separate test set. These metrics provide a comprehensive view of the model's predictive accuracy, with MSE emphasizing larger errors and MAE providing a more robust measure against outliers. The R2 score, indicating the proportion of variance in the dependent variable that is predictable from the independent variables, was also calculated.

3.6 Results

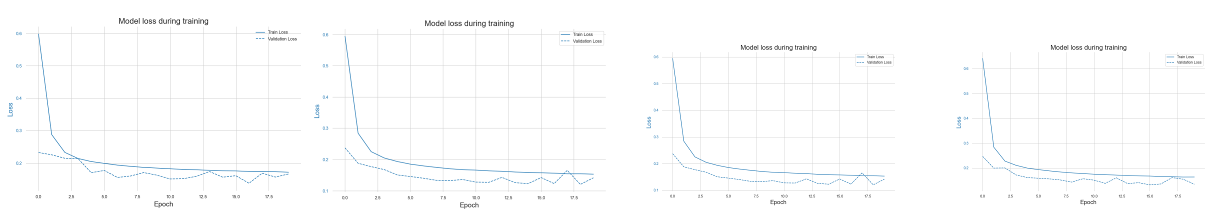
The loss function L used for training our neural network model is the Mean Squared Error (MSE) which is defined as follows:

$$L(y, \hat{y}) = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \tag{3.9}$$

where y_i is the actual value of the target variable for the i^{th} observation, $\hat{y}_i$ is the predicted value of the target variable for the i^{th} observation, N is the total number of observations in the dataset.

In Table 3.3[15], the estimated price elasticities of demand for juice are observed to span a range from 0.8 to 3.0, contingent on the specifics of the model configuration. The integration of supplementary text and image characteristics notably augments the performance of the models, as manifested by the diminished mean squared error (MSE).

Furthermore, the elasticity estimates extracted from these augmented models gravitate towards the reference elasticity of 1.37 for soft drinks, as documented by Guerrero-López et al., 2017. This indicates that our models are not only more statistically adept, but also exhibit a greater alignment with preceding empirical observations, thereby strengthening their external validity.



(a) (b) (c) (d)

Figure 3.6: Word 256, Image 256

We also discern a distinct pattern of escalating Adjusted R^2 values as additional features are incorporated into the models. This infers that these supplementary features inject significant explanatory power, thereby further enhancing the models' capability to accurately

[15]The results presented correspond to the best performing setup after experimenting with a variety of parameter configurations for each model.

Table 3.3: Comparison of Results from Multiple Model Setups

(1)	OLS	Lasso	Random Forest	XGBoost	Linear DML(Lasso)	LDML (RF)	LDML (XGB)	Sparse LDML(Lasso)	SLDML (RF)	SLDML (XGB)
MSE	0.364	0.408	0.102	0.262	0.407	0.118	0.348	0.407	0.165	0.339
Price	-1.44	-1.71			-1.36	-0.68	-1.34	-1.36	-1.032	-1.41
R^2	67.3	63.4	90.9	76.5						
(2)	OLS	Lasso	Random Forest	XGBoost	Linear DML(Lasso)	LDML (RF)	LDML (XGB)	Sparse LDML(Lasso)	SLDML (RF)	SLDML (XGB)
MSE	0.250	0.341	0.073	0.203	0.341	0.094	0.220	0.341	0.077	0.224
Price	-2.82	-1.53			-1.12	-1.52	-1.94	-1.12	-1.82	-1.14
R^2	77.6	69.4	93.5	81.8						
(3)	OLS	Lasso	Random Forest	XGBoost	Linear DML(Lasso)	LDML (RF)	LDML (XGB)	Sparse LDML(Lasso)	SLDML (RF)	SLDML (XGB)
MSE	0.240	0.305	0.073	0.203	0.306	0.092	0.218	0.306	0.077	0.221
Price	-2.96	-1.83			-1.24	-1.40	-1.48	-1.24	-0.81	-1.66
R^2	78.5	72.7	93.4	81.8						
(4)	OLS	Lasso	Random Forest	XGBoost	Linear DML(Lasso)	LDML (RF)	LDML (XGB)	Sparse LDML(Lasso)	SLDML (RF)	SLDML (XGB)
MSE	0.204	0.289	0.073	0.199	0.291	0.093	0.215	0.291	0.077	0.218
Price	-8.79	-1.74			-1.27	-1.36	-1.71	-1.27	-0.80	-1.85
R^2	81.6	72.7	93.5	82.1						

* Note: 1. Results are all significant at 5% level. 2. R^2 are reporint in %. 3. The embeddings of image and word are 256. 4. (1) Only numerical and categorical features; (2) features in (1) with text features; (3) features in (1) with image features; (4) features in (1) with both text and image features.

Table 3.4: Comparison of Results from Multiple Model Setups

(1)	OLS	Lasso	Random Forest	XGBoost	Linear DML(Lasso)	LDML (RF)	LDML (XGB)	Sparse LDML(Lasso)	SLDML (RF)	SLDML (XGB)
MSE	0.33	0.37	0.09	0.24	0.37	0.11	0.31	0.37	0.15	0.31
Price	-1.30	-1.54			-1.22	-0.61	-1.21	-1.22	-0.93	-1.27
R^2	70.6	67.1	91.8	78.9						
(2)	OLS	Lasso	Random Forest	XGBoost	Linear DML(Lasso)	LDML (RF)	LDML (XGB)	Sparse LDML(Lasso)	SLDML (RF)	SLDML (XGB)
MSE	0.23	0.31	0.07	0.18	0.31	0.08	0.20	0.31	0.07	0.20
Price	-2.54	-1.38			-1.01	-1.37	-1.75	-1.01	-1.64	-1.03
R^2	79.8	72.5	94.2	83.6						
(3)	OLS	Lasso	Random Forest	XGBoost	Linear DML(Lasso)	LDML (RF)	LDML (XGB)	Sparse LDML(Lasso)	SLDML (RF)	SLDML (XGB)
MSE	0.22	0.27	0.07	0.18	0.28	0.08	0.20	0.28	0.07	0.20
Price	-2.67	-1.65			-1.12	-1.26	-1.34	-1.12	-0.73	-1.50
R^2	80.7	75.4	94.1	83.6						
(4)	OLS	Lasso	Random Forest	XGBoost	Linear DML(Lasso)	LDML (RF)	LDML (XGB)	Sparse LDML(Lasso)	SLDML (RF)	SLDML (XGB)
MSE	0.204	0.289	0.073	0.199	0.291	0.093	0.215	0.291	0.077	0.218
Price	-8.79	-1.74			-1.27	-1.36	-1.71	-1.27	-0.80	-1.85
R^2	81.6	78.7	94.5	86.1						

* Note: 1. Results are all significant at 5% level. 2. R^2 are reporint in %. 3. The embeddings of image and word are 512. 4. (1) Only numerical and categorical features; (2) features in (1) with text features; (3) features in (1) with image features; (4) features in (1) with both text and image features.

decipher the intricate dynamics of juice demand. The magnitude of this enhancement is of considerable importance as it signifies that the features derived via deep learning effectively embody complex patterns and interrelationships that traditional variables fail to capture.[16]

The addition of deep learning features also had a particularly notable impact on the performance of the neural network models (Fig 3.6). As indicated in Table 3.3, the neural network models, including Random Forest and XGBoost, demonstrated considerable improvements in predictive accuracy following the incorporation of these features. This is a testament to the ability of these models to effectively leverage the high-dimensional, non-linear information contained within the deep features. The neural networks' capacity for modeling complex interactions and capturing intricate patterns in the data appears to synergize well with the rich, nuanced information encapsulated by the deep features. This combined capacity leads to a more accurate representation of product demand, highlighting the potential benefits of integrating deep learning techniques into traditional econometric analysis.

3.7 Conclusion

In this study, we embarked on an exploratory journey to uncover the most effective methods of leveraging information from deep learning in traditional demand estimation. By comparing econometric models with machine learning algorithms, we focused on a typical demand estimation scenario using a rich dataset of juice products. Our research journey led us to expand our methods to encompass techniques like double machine learning, which effectively dealt with the rise in data dimensionality and sparsity.

We then successfully incorporated the high-dimensional and sparse matrices from various data types, including dummy variables categorizing product categories and flavors, image and word embeddings, and more into our models. This process was facilitated by a blend of machine learning and econometric techniques, striking a balance between traditional statistical

[16] Additional results are in the appendix.

methods and modern computational power.

As part of our methodology, we employed a variety of machine learning techniques including Random Forests, Lasso regression, and XGBoost, as well as a custom-designed regression neural network. All these models were tailored and fine-tuned to effectively accommodate the unique structure of our dataset.

Our results highlighted the importance and potential of incorporating deep learning features into traditional econometric models. We found that the inclusion of image and text features not only improved the accuracy of our models but also provided more robust elasticity estimates. This, in turn, reinforced our belief in the potential of a synergistic approach, combining traditional econometrics with innovative machine learning techniques to handle complex, high-dimensional data.

This paper provides a roadmap for integrating machine learning techniques with econometric models, a promising direction for future research. We hope that our findings and methodologies can serve as a catalyst for further innovation and exploration in the field of demand estimation.

REFERENCES

Aguirregabiria, V. (2018). *Empirical industrial organization: Models, methods and applications*. Cambridge University Press.

Alcedo, J., Cavallo, A., Dwyer, B., Mishra, P., & Spilimbergo, A. (2022). E-commerce during covid: Stylized facts from 47 economies.

Angrist, J. D. (2009). *Mostly harmless econometrics : An empiricist's companion*. Princeton University Press.

Arcidiacono, P., Ellickson, P. B., Mela, C. F., & Singleton, J. D. (2020). The competitive effects of entry: Evidence from supercenter expansion. *American Economic Journal: Applied Economics*, *12*(3), 175–206. https://doi.org/10.1257/app.20180047

Athey, S., & Imbens, G. W. (2007). Discrete choice models with multiple unobserved choice characteristics. *International Economic Review*, *48*(4), 1159–1192. Retrieved February 15, 2023, from http://www.jstor.org/stable/4542008

Athey, S., & Imbens, G. W. (2022). Design-based analysis in difference-in-differences settings with staggered adoption [Annals Issue in Honor of Gary Chamberlain]. *Journal of Econometrics*, *226*(1), 62–79. https://doi.org/https://doi.org/10.1016/j.jeconom.2020.10.012

Bajari, P., Nekipelov, D., Ryan, S. P., & Yang, M. (2015). *Demand estimation with machine learning and model combination* (Working Paper No. 20955). National Bureau of Economic Research. https://doi.org/10.3386/w20955

Baker, A. C., Larcker, D. F., & Wang, C. C. (2022). How much should we trust staggered difference-in-differences estimates? *Journal of Financial Economics*, *144*(2), 370–395. https://doi.org/https://doi.org/10.1016/j.jfineco.2022.01.004

Barrios, J. M. (2021). Staggeringly problematic: A primer on staggered did for accounting researchers. *SSRN Electronic Journal*.

Basker, E. (2005). Job Creation or Destruction? Labor Market Effects of Wal-Mart Expansion. *The Review of Economics and Statistics*, *87*(1), 174–183. https://ideas.repec.org/a/tpr/restat/v87y2005i1p174-183.html

Basker, E., & Noel, M. (2009). The evolving food chain: Competitive effects of wal-mart's entry into the supermarket industry. *Journal of Economics & Management Strategy*, *18*(4), 977–1009. https://doi.org/https://doi.org/10.1111/j.1530-9134.2009.00235.x

Baum-Snow, N., & Kahn, M. E. (2000). The effects of new public projects to expand urban rail transit. *Journal of Public Economics*, *77*(2), 241–263. https://doi.org/https://doi.org/10.1016/S0047-2727(99)00085-7

Berry, S., Levinsohn, J., & Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica*, *63*(4), 841–890. Retrieved April 19, 2023, from http://www.jstor.org/stable/2171802

Berry, S. T., & Haile, P. A. (2021a). Chapter 1 - foundations of demand estimationfor the handbook of industrial organization. we benefited from detailed comments from chris conlon, jean-pierre dubé, aviv nevo, and frank verboven. jaewon lee provided capable research assistance. In K. Ho, A. Hortaçsu, & A. Lizzeri (Eds.), *Handbook of industrial organization, volume 4* (pp. 1–62). Elsevier. https://doi.org/https://doi.org/10.1016/bs.hesind.2021.11.001

Berry, S. T., & Haile, P. A. (2021b). *Foundations of Demand Estimation* (Cowles Foundation Discussion Papers No. 2301). Cowles Foundation for Research in Economics, Yale University. https://ideas.repec.org/p/cwl/cwldpp/2301.html

Bojinov, I., Rambachan, A., & Shephard, N. (2021). Panel experiments and dynamic causal effects: A finite population perspective. *Quantitative Economics*, *12*(4), 1171–1196. https://doi.org/https://doi.org/10.3982/QE1744

Brynjolfsson, E., Hu, Y. (, & Smith, M. D. (2003). Consumer surplus in the digital economy: Estimating the value of increased product variety at online booksellers. *Management science*, *49*(11), 1580–1596.

Callaway, B., & SantAnna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of econometrics*, *225*(2), 200–230.

Cavallo, A. (2017). Are online and offline prices similar? evidence from large multi-channel retailers. *American Economic Review*, *107*(1), 283–303. https://doi.org/10.1257/aer.20160542

Chava, S., Oettl, A., Singh, M., & Zeng, L. (2022). Creative destruction? impact of e-commerce on the retail sector. (30077). https://doi.org/10.3386/w30077

Chen, Y., & Riordan, M. H. (2008). Price-increasing competition. *The Rand journal of economics*, *39*(4), 1042–1058.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., & Robins, J. (2016). Double/debiased machine learning for treatment and causal parameters. https://doi.org/10.48550/ARXIV.1608.00060

Conlon, C. T., & Mortimer, J. H. (2013). Demand estimation under incomplete product availability. *American Economic Journal: Microeconomics*, *5*(4), 1–30. https://doi.org/10.1257/mic.5.4.1

de Chaisemartin, C., & D'Haultfuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, *110*(9), 2964–96. https://doi.org/10.1257/aer.20181169

DellaVigna, S., & Gentzkow, M. (2019). Uniform Pricing in U.S. Retail Chains. *The Quarterly Journal of Economics*, *134*(4), 2011–2084. https://ideas.repec.org/a/oup/qjecon/v134y2019i4p2011-2084..html

Dimoka, A., Hong, Y., & Pavlou, P. A. (2012). On product uncertainty in online markets: Theory and evidence. *MIS quarterly*, *36*(2), 395–426.

Draganska, M., Villas-Boas, S., & Klapper, D. (2010). A larger slice or a larger pie? an empirical investigation of bargaining power in the distribution channel. *Marketing science (Providence, R.I.)*, *29*(1), 57–74.

Dynan, K. E. (2000). Habit formation in consumer preferences: Evidence from panel data. *American Economic Review*, *90*(3), 391–406. https://doi.org/10.1257/aer.90.3.391

Gandhi, A., & Nevo, A. (2021). Chapter 2 - empirical models of demand and supply in differentiated products industrieswe thank pierre dubois, phil haile, ali hortaçsu, felipe kup barbieri de matos, jing tao and three referees for helpful comments and insightful suggestions. In K. Ho, A. Hortaçsu, & A. Lizzeri (Eds.), *Handbook of industrial organization, volume 4* (pp. 63–139). Elsevier. https://doi.org/https://doi.org/10.1016/bs.hesind.2021.11.002

Ghose, A., & Yao, Y. (2011). Using transaction prices to re-examine price dispersion in electronic markets. *Information Systems Research*, *22*(2), 269–288. https://doi.org/10.1287/isre.1090.0252

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of econometrics*, *225*(2), 254–277.

Guerrero-López, C. M., Unar-Munguía, M., & Colchero, M. A. (2017). Price elasticity of the demand for soft drinks, other sugar-sweetened beverages and energy dense food in chile. *BMC public health*, *17*(1), 180–180.

Hartford, J., Lewis, G., Leyton-Brown, K., & Taddy, M. (2017). Deep iv: A flexible approach for counterfactual prediction (Proceedings of Machine Learning Research). *Proceedings of Machine Learning Research*, *70*. https://www.microsoft.com/en-us/research/publication/deep-iv-a-flexible-approach-for-counterfactual-prediction/

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep residual learning for image recognition. *CoRR*, *abs/1512.03385*. http://arxiv.org/abs/1512.03385

Hitsch, G. J., Hortaçsu, A., & Lin, X. (2019). *Prices and promotions in u.s. retail markets: Evidence from big data* (Working Paper No. 26306). National Bureau of Economic Research. https://doi.org/10.3386/w26306

Holmes, T. J. (2011). The diffusion of wal-mart and economies of density (Manuscript received January, 2008; final revision received March, 2010.). *Econometrica*, *79*(1), 253–302.

Huang, Q., Nijs, V., Hansen, K., & Anderson, E. (2012). Wal-mart's impact on supplier profits. *Journal of Marketing Research*, *49*(2), 131–143. https://doi.org/10.1509/jmr.10.0256

Hwang, M., & Park, S. (2016). The impact of walmart supercenter conversion on consumer shopping behavior. *Management Science*, *62*(3), 817–828. https://doi.org/10.1287/mnsc.2014.2143

Imai, K., & Kim, I. S. (2021). On the use of two-way fixed effects regression models for causal inference with panel data. *Political Analysis*, *29*(3), 405–415. https://doi.org/10.1017/pan.2020.33

Jia, P. (2008). What happens when wal-mart comes to town: An empirical analysis of the discount retailing industry (Manuscript receivedăAugust, 2006; final revision receivedăJanuary, 2008.). *Econometrica*, *76*(6), 1263–1316.

Jindal, P., Zhu, T., Chintagunta, P. K., & Dhar, S. K. (2018). Point-of-sale marketing mix and brand performance the moderating role of retail format and brand type. *SSRN Electronic Journal*.

Kaplan, G., Menzio, G., Rudanko, L., & Trachter, N. (2019). Relative price dispersion: Evidence and theory. *American economic journal. Microeconomics*, *11*(3), 68–124.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature (London)*, *521*(7553), 436–444.

Liu, X. (2023). Deep learning in marketing: A review and research agenda (K. Sudhir & O. Toubia, Eds.). *20*, 239–271. https://doi.org/10.1108/S1548-643520230000020014

Neumark, D., Zhang, J., & Ciccarella, S. (2008). The effects of wal-mart on local labor markets. *Journal of urban economics*, *63*(2), 405–430.

Newberry, P. W., Seim, K., & Houde, J.-F. (Forthcoming). Nexus tax laws and economies of density in e-commerce: A study of amazons fulfillment center network. *Econometrica*.

Pagliardini, M., Gupta, P., & Jaggi, M. (2018). Unsupervised learning of sentence embeddings using compositional n-gram features. *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*. https://doi.org/10.18653/v1/n18-1049

Reddy, P., Gharbi, M., Lukac, M., & Mitra, N. J. (2021). Im2vec: Synthesizing vector graphics without vector supervision. https://doi.org/10.48550/ARXIV.2102.02798

Reimers, N., & Gurevych, I. (2019). Sentence-bert: Sentence embeddings using siamese bert-networks. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. https://arxiv.org/abs/1908.10084

Seamans, R., & Zhu, F. (2014). Responses to Entry in Multi-Sided Markets: The Impact of Craigslist on Local Newspapers. *Management Science*, *60*(2), 476–493. https://doi.org/10.1287/mnsc.2013.1785

Sfeir, G., Rodrigues, F., & Abou-Zeid, M. (2021). Gaussian process latent class choice models. https://doi.org/10.48550/ARXIV.2101.12252

Sun, L., & Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, *225*(2), 175–199. https://doi.org/https://doi.org/10.1016/j.jeconom.2020.09.006

Villas-Boas, S. B. (2007). Vertical relationships between manufacturers and retailers: Inference with limited data. *The Review of economic studies*, *74*(2), 625–652.

Appendix A

APPENDIX

A.1 Appendix 1

A.1.1 Summary Statistics of Scanner Data

Table A.1: Summary Statistics of Scanner Data

	Juice		FreshProduce	
	Total	**Range**	**Total**	**Range**
Number of Stores	8128		2657	
Number of Years	7	2010-2016	7	2010-2016
Number of County	128		245	
Number of Markets	17		17	
Store type	**Total**	**Percent**	**Total**	**Percent**
Food	2584	31.8%	225862	93.4%
Drug	3711	45.7%	3094	1.3%
Mass Merchandise	1554	19.1%	12831	5.3%
Others	279	3.4%		
Distance to Closest center	**Total**	**Percent**	**Total**	**Percent**
Less than 20 Miles	379506	60.2%	144745	59.9%
Between 20 and 50 Miles	195856	31.1%	68983	28.5%
Between 50 and 100 Miles	45531	7.2%	22235	9.2%
Over 100 Miles	9828	1.6%	5824	2.4%
Chain Size	**Total**	**Percent**	**Total**	**Percent**
Over 2000 Stores	333171	52.8%	3022	1.2%
Between 500 and 2000 Stores	151568	24%	91526	37.9%
Less than 500 Stores	145982	23.1%	147239	60.9%

A.1.2 List of Amazon Fresh Distribution Center

Distribution Center		
DMA Name	Entry Date	Number of center (by 2020)
Seattle	Aug 2007	3
Los Angeles	Jun 2013	4
Cincinnati	Sep 2013	1
New York	Nov 2014	1
San Francisco	Jan 2015	3
Sacramento	Mar 2015	1
Boston	Jun 2015	1
Washington center	Jun 2016	4
Dallas	Jun 2016	2
Chicago	Jul 2016	2
Kansas City	Oct 2016	2
Atlanta	Nov 2016	3
Miami	Jan 2017	2
Denver	Aug 2017	2
Hartford-New Haven	Aug 2017	1
San Antonio	Aug 2017	1
Detroit	Apr 2018	1
Austin	Feb 2020	1

Table A.2: Fresh Distribution Center

Note: The treatment group are the markets that have Amazon Fresh entry before 2017. Miami, Denver, Hartford, San Antonio, Detroit have AF entered after 2017 and they are in the control group.

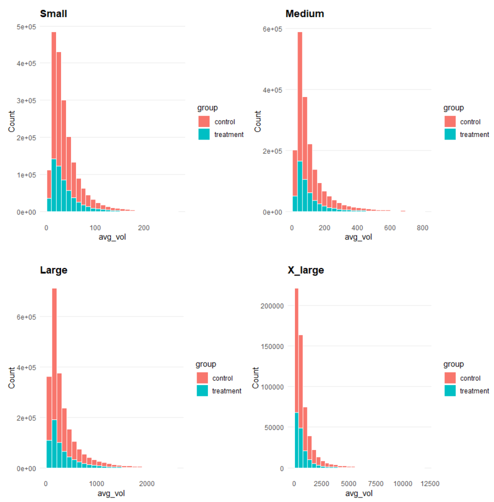
A.1.3 TWFE DID Estimates

Models	(1)	(2)	(3)	(4)	(5)	(6)
Amazon Fresh Entry	-0.1004	-0.0727	-0.0689	-0.0727	-0.0694	-0.0428
Volume(Juice)	(0.0193)	(0.0177)	(0.0170)			(0.0199)
Amazon Fresh Entry	-0.0183	0.0112	0.0114	0.0111	0.0113	-0.0073
Price(Juice)	(0.0066)	(0.0079)	(0.0077)			(0.0055)
Amazon Fresh Entry	0.0615	-0.0293	-0.0301	-0.0293	-0.0296	-0.0591
Revenue(FP)	(0.0186)	(0.0272)	(0.0272)			(0.0158)
Time FE	No	Yes	Yes	Yes	Yes	Yes
Unit FE	Yes	Yes	Yes	Yes	Yes	Yes
Control	No	No	Yes	No	Yes	No

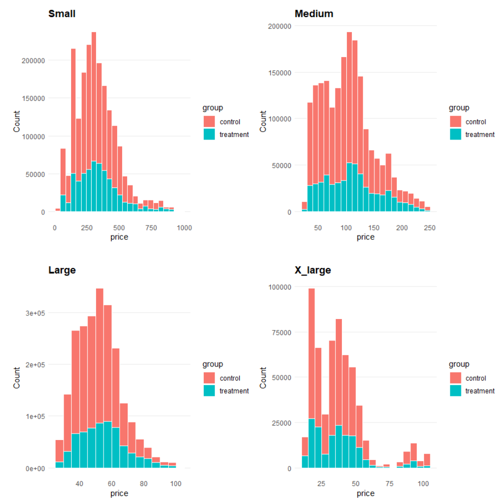
Table A.3: TWFE DID Estimates

Note: Model (1) reports estimates with only county fixed effect as in equation 1.5; model (2) reports estimates from TWFE model without covariates; model (3) reports estimates from TWFE with covariates; model (4) reports estimates following Goodman-Bacon, 2021; model (5) reports estimates following Goodman-Bacon, 2021; model (6) reports estimates following Callaway and SantAnna, 2021. SE are in the parenthesis.

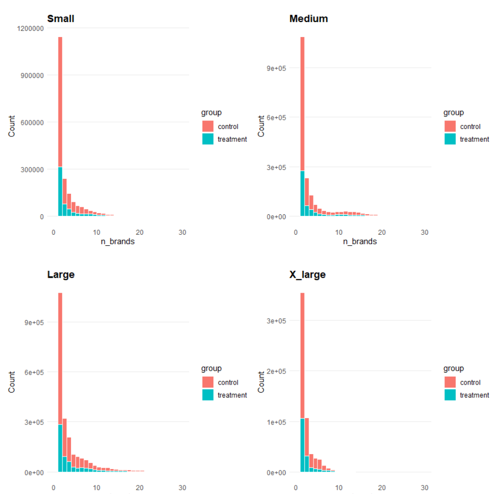
A.1.4 Summmary Stats



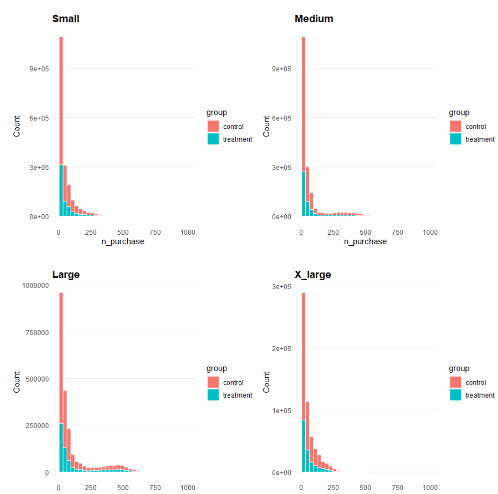
(a) Average Volume



(b) Price



(c) Total Brands



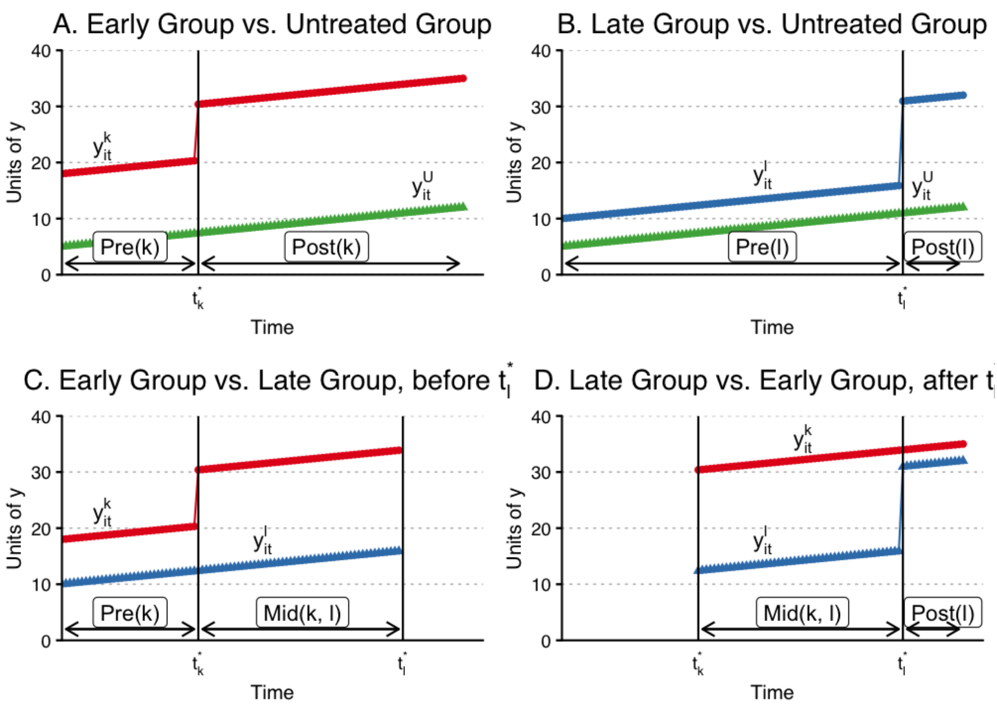
(d) Total Visits

Figure A.1: Summmary Stats

A.1.5 Goodman-Bacon Decompose

Bacon decomposition[1] is a procedure used to decompose a variable into fixed, persistent, and transitory components. It's a popular technique in econometrics for dealing with panel data, as it allows for the analysis of dynamic, non-stationary processes.

Figure A.2: GoodMan-Bacon DiD in Different Groups



In the context of a panel data setting, consider y_{it} as the observed variable of interest for unit i at time t, we can express this as:

$$y_{it} = \alpha_i + \rho y_{i,t-1} + \eta_{it} + \theta \eta_{i,t-1} + \epsilon_{it} \tag{A.1}$$

where α_i is the fixed effect, $\rho y_{i,t-1}$ represents the persistent component, η_{it} and $\theta \eta_{i,t-1}$

[1]Named after its creator, economist Francis Bacon, it is also known as the Bacon-Huber-Wooldridge (BHW) decomposition.

capture the transitory component, and ϵ_{it} is the idiosyncratic error term[2].

The fixed effect (α_i) represents the time-invariant characteristics of each unit. The persistent component ($\rho y_{i,t-1}$) represents a process that is highly autocorrelated and changes slowly over time. The transitory component (η_{it} and $\theta\eta_{i,t-1}$) represents changes that are not permanent and revert to the mean quickly.

Bacon decomposition is a useful tool for understanding the dynamics of a variable over time and can provide valuable insights into the nature of the data being studied. It allows for the identification of the parts of the variation that are intrinsic to the units (fixed), that are due to persistent shifts (persistent), or that are temporary (transitory), thus offering a comprehensive picture of the variable's behavior over time.

[2]It is assumed that the idiosyncratic error is not serially correlated and has zero mean.

Figure A.3: Bacon Decomposition for Juice on Volume

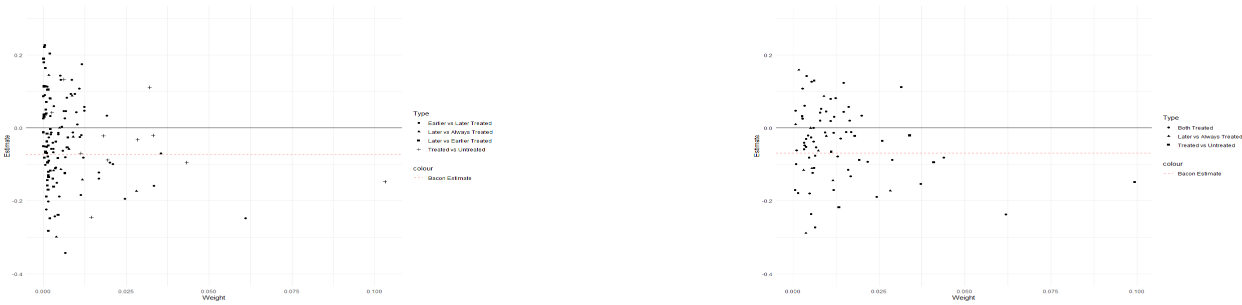
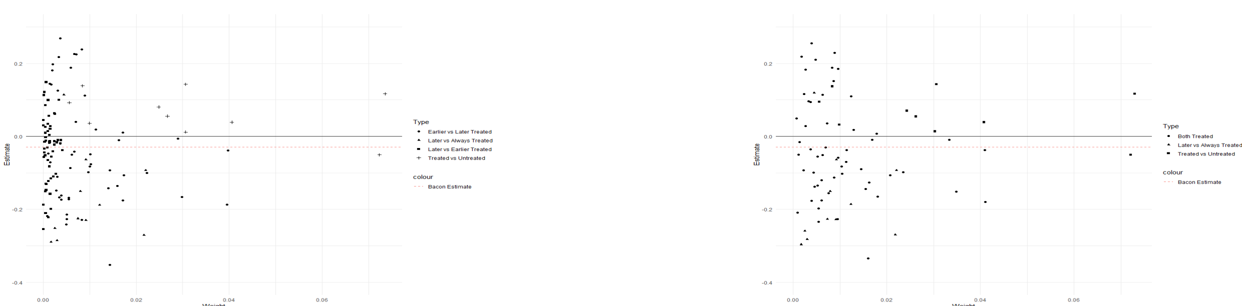


Figure A.4: Bacon Decomposition for Juice on Price

Figure A.5: Bacon Decomposition for Fresh Produce on Revenue



A.1.6 Event Study

Figure A.6: Event Study

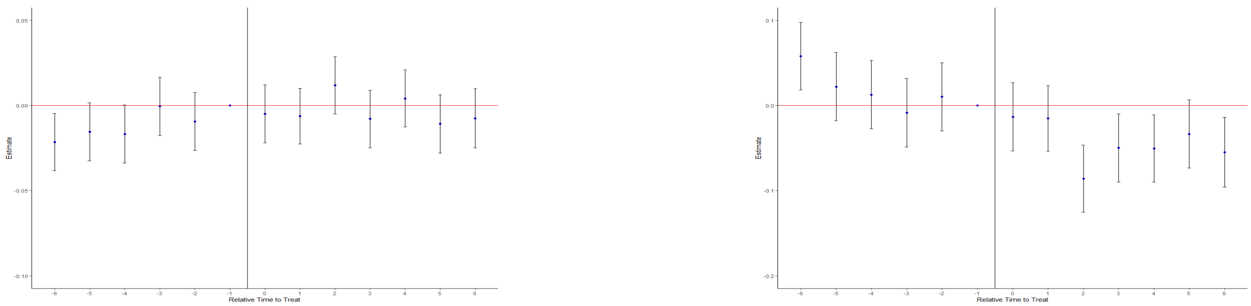
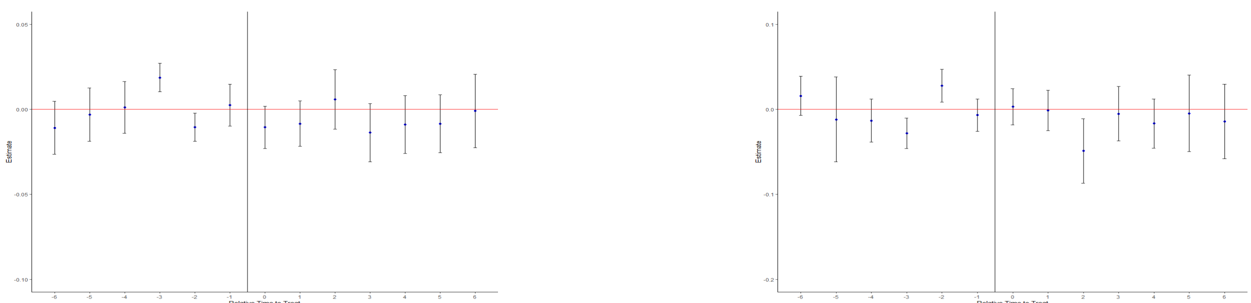


Figure A.7: Event Study: Callaway and SantAnna, 2021



A.1.7 Product Assortment

Table A.4: Product Assortment Effect: Use number of distinct UPCs

Model Spec.	OLS	Static	Static	Dynamic	CS (Time)	CS (Group)
Amazon Fresh	2.6***	-1.1	15.2***	-2.9*	6.2***	-1.3
SD	(0.6)	(1.7)	(1.7)	(1.1)	(1.7)	(1.2)
95% CI Lower Bound	1.6	-4.4	11.9	-5.1	2.9	-3.6
95% CI Higher Bound	3.8	2.2	18.5	-0.7	9.5	1.0
Unit Fixed Effect	None	Store	Store	Store	Store	Store
Time Fixed Effect	No	Yes	No	Yes	Yes	Yes
adj. R^2	83.5	94.4	94.3	94.4	NA	NA

Note: 1. Standard errors are in parentheses. Coefficients reported in the table are in 10^{-2}. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. 2. All std errors are grouped at county level. 3. For dynamic TWFE, I only report the estimates that are 1 quarter post treatment. Rest of estimates will be available in the appendix.

A.2 Appendix 2

IV

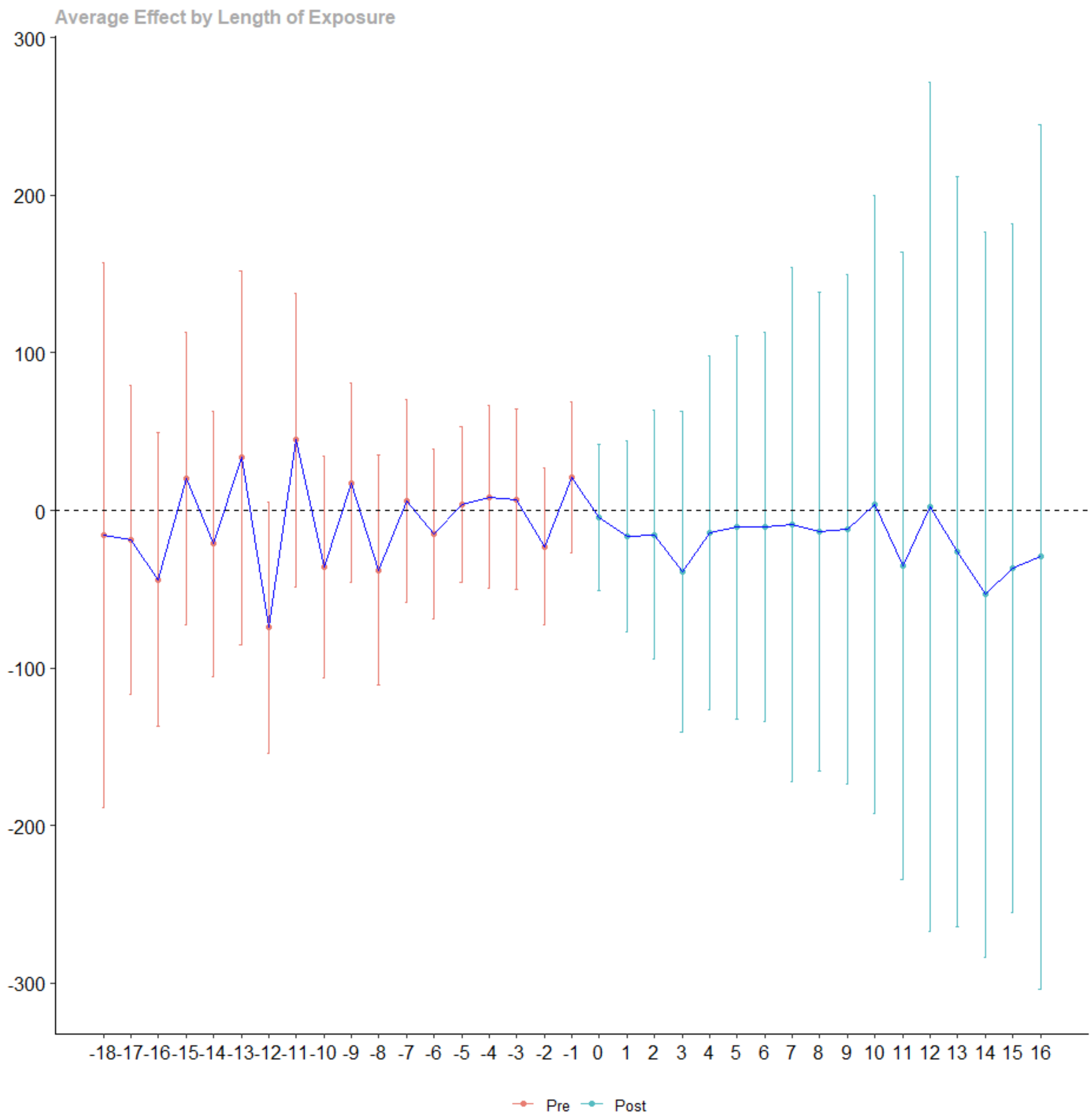
A potential alternative instrumental variable might be the quantity of preexisting warehouses in a given area, as this could serve as an indicator of Amazon's market share and consumer demand. However, this variable may not necessarily influence other outcomes, such as customer behavior or satisfaction. These are merely a few examples of potential instrumental variables, and their applicability and validity may need to be tested depending on your specific research question and data.

The influence of Amazon's distribution network and the progression of freight digitalization is the central theme of this paper. It delves into how Amazon's distribution network impacts freight flows and urban logistics, utilizing the proximity to major transport hubs and highways as an instrumental variable to gauge Amazon's presence's effect on truck traffic.

The paper titled "Machine Learning Instrument Variables for Causal Inference" puts forward a unique algorithm for constructing instrumental variables from exogenous data candidates using machine learning. This methodology is utilized to estimate the causal influence of the number of existing warehouses in a region on consumer demand for online retail.

The paper "Understanding Instrumental Variables in Models with Essential Heterogeneity" scrutinizes the attributes and limitations of instrumental variables in models where responses to interventions are heterogenous, and agents engage in treatments with partial knowledge of their unique responses. It explores how to interpret and test instrumental variables in such contexts.

A.2.1 How Does Amazon Fresh Change the Local Competition?


Average Effect by Length of Exposure
300
200
100
0
-100
-200
-300
-18 -17 -16 -15 -14 -13 -12 -11 -10 -9 -8 -7 -6 -5 -4 -3 -2 -1 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
Pre
Post

A.3 Appendix 3

A.3.1 Data Aggregation

As we conduct our data analysis, it is crucial to keep track of three different levels of aggregation. This can be best understood with the help of an example drawn from our dataset. At the lowest level, we have individual products identified by their unique Universal Product Code (UPC). For instance, we may have a bottle of "Kroger 100% Lemon Juice 15.0 oz" with a UPC of 1111070205. Each UPC belongs to a module within the AC Nielsen dataset, such as "Fruit Juice - Lemon/Lime."

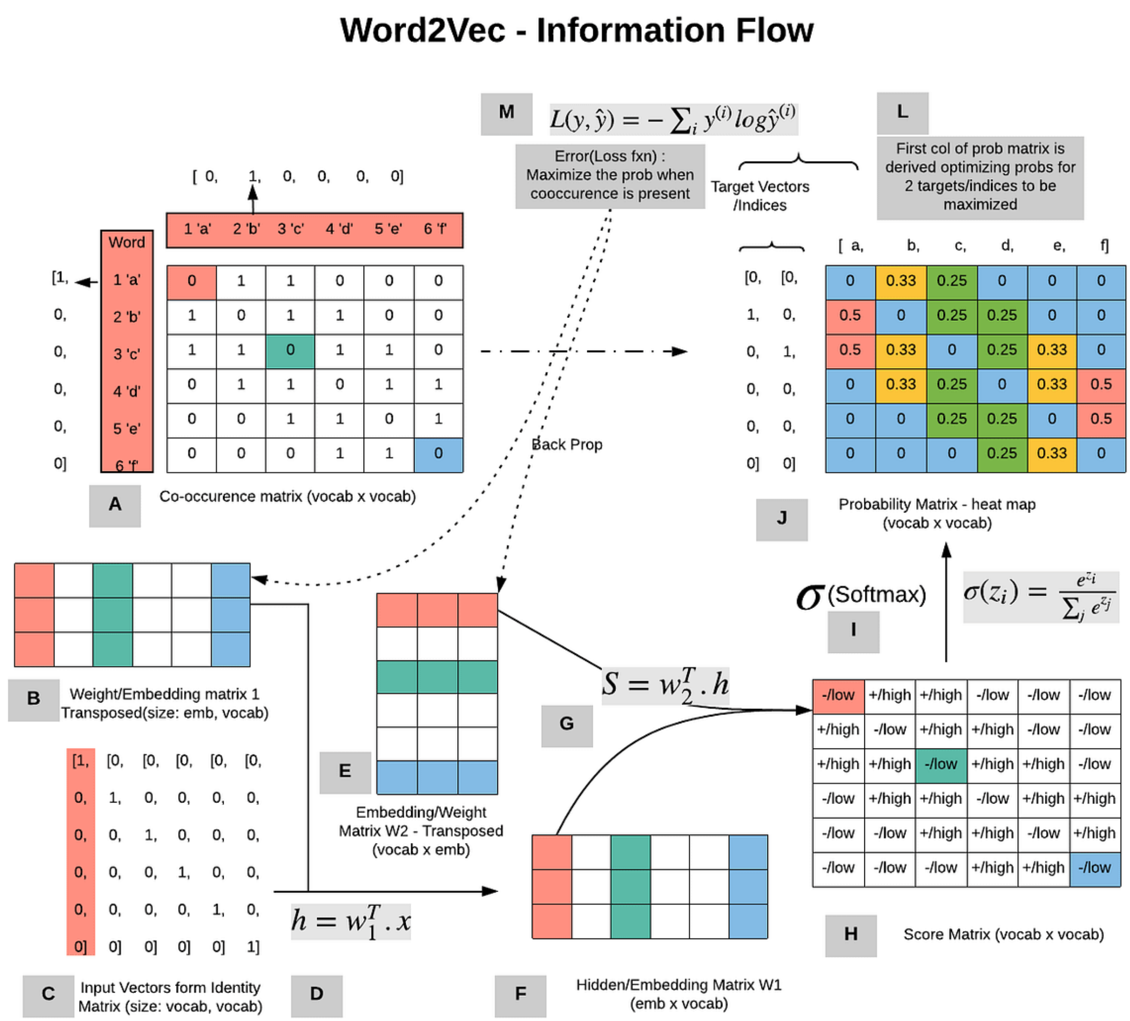
Moving up the hierarchy, the next level of aggregation is the "product module" which groups together similar products with the same UPC codes. For example, "Fruit Juice - Lemon/Lime" is a product module that includes all lemon and lime juice products. The module provides a way to aggregate data for all the products within it.

The highest level of aggregation is the "product group," which encompasses a set of related product modules. For example, "Juice, Drinks - Canned, bottled" is a product group that contains multiple modules, including "Fruit Juice - Lemon/Lime" and "Soft Drinks."

To help match with the CPI entry level items (ELIs), we use the product group description and product module description. The CPI defines ELIs as the goods and services that underlie the index and are used to calculate inflation. Our sample covers approximately 104 of the 305 ELIs, and by matching the product descriptions to the ELIs, we can gain insights into how changes in the prices of different products impact the CPI.

By analyzing the data at these different levels of aggregation, we can gain a deeper understanding of how consumers' purchasing behaviors and preferences influence the prices of specific products, product modules, and product groups, and how these price changes, in turn, impact the overall inflation rate as reflected in the CPI.

A.3.2 Word2vec Flowchart



A.3.3 Sentence2vec Details

In this section, we describe the process of using sentence2vec to create word vectors, which can be employed to capture semantic relationships between words and sentences. The flow of the process can be outlined as follows:

- Data preparation: Gather a large corpus of text from various sources, such as articles, books, and web pages. Preprocess the text by tokenizing it into sentences and words, removing stop words, and performing stemming or lemmatization as necessary.

- Training a word2vec model: Train a word2vec model using the preprocessed text corpus. Word2vec is a widely-used neural network-based technique for creating word embeddings, which are vector representations of words capturing their semantic and syntactic properties. The model learns to map words to high-dimensional vectors based on their co-occurrence patterns in the text.

- Sentence embedding using sentence2vec: Utilize sentence2vec, a method that extends word2vec to generate sentence embeddings. Sentence2vec creates vector representations of sentences by combining the word vectors of the words in a sentence. The sentence embeddings can capture the semantic similarity between sentences, as they account for the relationships between the words in a sentence.

- Choosing an aggregation method: Select an appropriate aggregation method to combine word vectors into sentence vectors. Common aggregation methods include averaging, summation, or more complex techniques such as weighted averaging based on word importance or frequency.

- Applying the sentence2vec model: Apply the sentence2vec model to your text data, transforming each sentence into a high-dimensional vector representation. These sentence embeddings can be used for various natural language processing tasks, such as text classification, clustering, or sentiment analysis.

- Evaluation and fine-tuning: Evaluate the performance of the sentence2vec model on relevant tasks or benchmarks. Fine-tune the model's hyperparameters, such as the size of the word and sentence vectors, the learning rate, and the context window size, to improve the quality of the embeddings.

By following this process, we can leverage sentence2vec to create word and sentence vectors that capture the semantic relationships between words and sentences to create the word descriptions vecotrs used as inputs for future Neural Networks.

A.3.4 Discrete Choice Models

Discrete choice models (DCMs) are widely employed in modern Industrial Organization for analyzing and modeling consumer behavior. Their applications span demand estimation and prediction, product recommendation, and market segmentation. DCMs rest on the assumption that consumers choose among a set of products based on their preferences for product attributes, such as price, quality, and brand.

One of the main strengths of DCMs is their ability to capture the heterogeneity in consumer preferences and estimate the relative importance of different product attributes. DCMs can also predict the market share of different products under different scenarios, such as changes in product attributes, prices, and promotions. DCMs are based on a well-established theoretical framework and offer interpretability of the model estimates, which is particularly important for policy and decision-making purposes.

Another strength of DCMs is their ability to handle complex and dynamic choice situations, such as multiple and nested choices, product substitutions, and attribute non-attendance. DCMs can incorporate different types of preference heterogeneity, such as taste heterogeneity, scale heterogeneity, and error heterogeneity. DCMs can also account for temporal dynamics, such as habit formation and learning Dynan, 2000.

As expected, nothing is always perfect. DCMs have several limitations that need to be taken into account before its application. As stated in Berry and Haile, 2021b, a primary challenge in demand estimation is the endogeneity of prices. There's a risk of mis-specification and bias, particularly in the presence of endogeneity and omitted variable bias. DCMs assume that consumers are rational and make choices based on their preferences, but this assumption may not always hold in practice. DCMs may also suffer from measurement error and response bias, particularly in self-reported surveys. Another limitation is that Demand for a given good, of course, typically depends on the prices and characteristics of all related goods. It requires DCMs rely on strong assumptions about the distribution of unobserved factors and the linearity of the utility function[3]. DCMs are also sensitive to the choice set and the level of detail in the product attributes.

DCMs are a powerful tool for demand estimation that offer interpretability, flexibility, and the ability to capture heterogeneity in consumer preferences. However, DCMs also have several limitations, particularly in the presence of endogeneity and bias. The integration of DCMs with other methods, such as machine learning, offers a promising direction for future research in demand estimation.

Recent research has explored the integration of DCMs with other methods, such as machine learning, to overcome some of these limitations and leverage their strengths. Hybrid models that combine DCMs with machine learning techniques, such as latent class analysis Sfeir et al., 2021 and gradient boosting, have been proposed to improve the accuracy and interpretability of demand models. Another approach is to use machine learning to preprocess the data and extract relevant features for DCMs, such as using deep learning to extract latent representations of images or text. This article is trying to explore the use of deep learning in demand estimation to see if they can outperform traditional DCMs.

A.3.5 Additional Results

---

[3] In Athey and Imbens, 2007, more unobserved characteristics may require strict functional form of the utility.

Table A.5: Comparison of Results from Multiple Model Setups

(1)	OLS	Lasso	Random Forest	XGBoost	Linear DML(Lasso)	LDML (RF)	LDML (XGB)	Sparse LDML(Lasso)	SLDML (RF)	SLDML (XGB)
MSE	0.3465	0.3525	0.0945	0.252	0.3885	0.1045	0.3255	0.3615	0.1575	0.296
Price	-1.365	-1.617			-1.281	-0.6405	-1.171	-1.281	-0.9765	-1.1935
R^2	69.9	66.2	92.1	79.7						
(2)	OLS	Lasso	Random Forest	XGBoost	Linear DML(Lasso)	LDML (RF)	LDML (XGB)	Sparse LDML(Lasso)	SLDML (RF)	SLDML (XGB)
MSE	0.2415	0.3255	0.0735	0.189	0.3255	0.084	0.21	0.3255	0.0735	0.21
Price	-2.667	-1.449			-1.0605	-1.4385	-1.8375	-1.0605	-1.722	-1.0815
R^2	78.9	71.5	93.8	82.5						
(3)	OLS	Lasso	Random Forest	XGBoost	Linear DML(Lasso)	LDML (RF)	LDML (XGB)	Sparse LDML(Lasso)	SLDML (RF)	SLDML (XGB)
MSE	0.231	0.2835	0.0665	0.189	0.294	0.084	0.21	0.294	0.0665	0.21
Price	-2.8035	-1.7325			-1.176	-1.323	-1.407	-1.176	-0.7645	-1.575
R^2	79.7	74.9	93.4	82.5						
(4)	OLS	Lasso	Random Forest	XGBoost	Linear DML(Lasso)	LDML (RF)	LDML (XGB)	Sparse LDML(Lasso)	SLDML (RF)	SLDML (XGB)
MSE	0.2142	0.30345	0.06885	0.20905	0.30555	0.09765	0.22575	0.30555	0.07245	0.2289
Price	-9.2295	-1.827			-1.3335	-1.428	-1.7955	-1.3335	-0.840	-1.9425
R^2	80.58	77.76	94.725	85.695						

* Note: 1. Results are all significant at 5% level. 2. R^2 are reporint in %. 3. The embeddings of image is 256 and word is 512. 4. (1) Only numerical and categorical features; (2) features in (1) with text features; (3) features in (1) with image features; (4) features in (1) with both text and image features.

Table A.6: Comparison of Results from Multiple Model Setups

(1)	OLS	Lasso	Random Forest	XGBoost	Linear DML(Lasso)	LDML (RF)	LDML (XGB)	Sparse LDML(Lasso)	SLDML (RF)	SLDML (XGB)
MSE	0.3465	0.3885	0.0945	0.252	0.3885	0.1155	0.3255	0.3885	0.1575	0.3255
Price	-1.365	-1.617			-1.281	-0.6405	-1.271	-1.281	-0.9765	-1.3335
R^2	74.13	70.455	96.39	82.845						
(2)	OLS	Lasso	Random Forest	XGBoost	Linear DML(Lasso)	LDML (RF)	LDML (XGB)	Sparse LDML(Lasso)	SLDML (RF)	SLDML (XGB)
MSE	0.2415	0.3255	0.0735	0.189	0.3255	0.084	0.21	0.3255	0.0735	0.21
Price	-2.667	-1.449			-1.0605	-1.4385	-1.8375	-1.0605	-1.722	-1.0815
R^2	83.79	76.125	98.91	87.78						
(3)	OLS	Lasso	Random Forest	XGBoost	Linear DML(Lasso)	LDML (RF)	LDML (XGB)	Sparse LDML(Lasso)	SLDML (RF)	SLDML (XGB)
MSE	0.231	0.2835	0.0735	0.189	0.294	0.084	0.21	0.294	0.0735	0.21
Price	-2.8035	-1.7325			-1.176	-1.323	-1.407	-1.176	-0.7665	-1.575
R^2	84.735	79.17	98.685	87.78						
(4)	OLS	Lasso	Random Forest	XGBoost	Linear DML(Lasso)	LDML (RF)	LDML (XGB)	Sparse LDML(Lasso)	SLDML (RF)	SLDML (XGB)
MSE	0.204	0.289	0.073	0.199	0.291	0.093	0.215	0.291	0.077	0.218
Price	-8.79	-1.74			-1.27	-1.36	-1.71	-1.27	-0.80	-1.85
R^2	81.6	78.7	94.5	86.1						

* Note: 1. Results are all significant at 5% level. 2. R^2 are reporint in %. 3. The embeddings of image and word are 1024. 4. (1) Only numerical and categorical features; (2) features in (1) with text features; (3) features in (1) with image features; (4) features in (1) with both text and image features.