THREE ESSAYS ON LABOR ECONOMICS

By
Md Shahedul Amin

A Dissertation Submitted in Partial Fulfillment of the Requirements
for the Degree of Philosophy in Economics

Middle Tennessee State University

July, 2023

Dissertation Committee:
Charles Baum, Chair
Aaron Gamino
Michael Roach

Dedication

This dissertation is dedicated to the memory of my beloved father, Md Ruhul Amin.

ACKNOWLEDGMENTS

For many detailed comments and conversations, I thank Dr. Charles Baum. I also want to share my gratitude towards Dr. Charles Baum for always providing me a non-judgemental environment and encouraging my growth as an independent researcher by sharing his knowledge and expertise. This dissertation has benefited from very useful conversations at various stages with Dr. Aaron Gamino and Dr. Michael Roach. During the most difficult times, the much needed direction from my dissertation committee made a difference in this current project, and I am indebted to my advisors Dr. Charles Baum, Dr. Aaron Gamino, and Dr. Michael Roach for their invaluable advice, continued guidance and support.

I would like to express my deepest gratitude to my parents, Shahida Begum and the late Md. Ruhul Amin, whose unwavering love, encouragement, and support have been instrumental in my academic pursuits. I am eternally grateful for the values and principles they instilled in me, which have shaped the person and researcher I have become.

Contents

List of Tables

List of Figures

Abstract

My dissertation comprises three chapters. The first chapter examines the impact of Amazon's physical expansion on local business owners, finding no effect on income but an increase in full-time local business owners. The second chapter assesses the effects of e-commerce expansion on the US labor market using instrumental variable techniques to analyze earnings, employment, and establishment numbers. The third chapter studies implementation of Recreational Marijuana Legalization (RML), finding no significant impact on labor outcomes across different skill levels.

The first chapter examines the impacts of the Amazon expansion on US business owners. Amazon's enabling of third-party sellers on its platform gave the company a strong foothold in the e-commerce market, allowing it to expand its physical presence across the United States. This paper examines how the gradual physical expansion of Amazon has affected local independent business owners on the county level outcome. I take natural instrumental variables to approach solving the endogeneity problem, which comes up as a result of taking care of the timing and location issues before setting up the fulfillment center. Using Current Population Survey (CPS) ASEC data from 2000-2021 this paper suggests that closer proximity to a fulfillment center has no predictive power on the income of full-time business owners. On the other hand, this research shows that by bolstering the physical presence across the United States at the county level, amazon motivates previously employed people to be full-time local business owners. Further empirical research using U.S. patent data reveals that the diffusion of e-commerce has no discernible impact on the innovative capacity of either incorporated or non-incorporated firms.

In the second chapter, I empirically test how e-commerce expansion has been impacting US labor market. E-commerce has experienced a remarkable surge in recent years, with top companies dominating the market and generating a significant share of total sales. While e-commerce has undoubtedly brought numerous benefits to consumers and businesses alike, it has also raised concerns about its impact on the labor market. The retail industry represents a significant portion of the total workforce in the US, and the rapid expansion of e-commerce has led to questions about the extent to which it has contributed to local labor markets. Moreover, the adoption of automated fulfillment centers by e-commerce companies such as Amazon has raised concerns about the potential negative spillover effects on other industries. To address the potential endogeneity of Amazon's entry, I employ a rigorous instrumental variable (IV) approach, constructing multiple instruments based on existing literature. In this paper, I aim to address these concerns and explore the impact of physical expansion of e-commerce setups on the local labor force. My analysis focuses on three key outcome variables: earnings, employment, and the number of establishments. By utilizing these IV techniques, I aim to provide

robust estimates of the causal effects of Amazon's presence on the local labor market dynamics.

In the third chapter, my study aims to examine the variation in the implementation of Recreational Marijuana Legalization (RML) across different states, utilizing the CPS Basic Monthly Sample data from 2009 to 2021. Employing a rigorous econometric approach, the Difference-in-Differences (DiD) regression method is employed to estimate the impact of RML on various labor outcomes, accounting for time-varying treatment effects and dynamic effects in states where it has been legalized for an extended period. Prior research provide suggestive and statistically significant evidence that high-skill service workers are less likely to face pre- or post-employment drug screening following the implementation of RML. Additionally, considering the income effect theory and the potential vulnerability of mid-skill workers who lack access to public insurance, this study investigates the impact of RML on their earnings and productivity. By categorizing workers based on their skill levels, the analysis reveals no significant impact on labor outcomes for individuals with high, mid, and low skill levels when employing recently developed DiD estimation methods that account for staggered timing and treatment heterogeneity. This finding starkly contradicts the results obtained through conventional Two-Way Fixed Effects (TWFE) method.

Chapter 1

Amazonomics: Does it Impact Business Owners?

1.1 Introduction

Amazon is becoming more present in people's lives daily by offering the endless promises of the internet and grandiose vision. In 2000, Amazon began allowing third-party sellers access to its platform and eventually named the program Amazon Marketplace. A decade ago, only one-quarter of Amazon's unit sales were generated by third-party merchants; today, however, that number has nearly doubled to nearly half. Bezos (2006) In the period from 1999 to 2018, Amazon grew from having a 10% share of online spending to having a 45% stake, adding to the increase in retail market concentration (Autor et al. (2020)). Over the last two decades, more specifically in the last one, Amazon has rapidly started to build its physical operation across the United States to cover more population in a shorter span of time (Figure 3.1). While doing this the e-commerce giant has met rising opposition from some neighborhoods. Even still, two of the objections that are voiced the most frequently are that Amazon is responsible for putting local businesses at risk and that it eliminates more retail jobs than it generates for a community in the warehouse industry. Amazon now takes a 34 percent cut of the revenue earned by independent sellers through its website by using a variety of fees to pocket the difference. That's up from 30 percent in 2018, and 19 percent in 2014 (Mitchell (2021)). This could be a sign of the upcoming bleak financial climate for independent sellers while the number of locally owned, independently owned , and operated shops in the United States has decreased by 108,000 during the past fifteen years, which is a reduction of 40 percent when assessed relative to population (US Census Bureau)

An independent research institute known as Institute of Local Self-Reliance (ILSR) came to the conclusion, using estimates from industry analysts regarding the margins that Amazon most likely earns on seller advertising and other seller fees, that

Marketplace may have generated operating profits of $24 billion in 2020. This figure is significantly higher than the profit that Amazon reported for Amazon Web Service (AWS), which was only $13.5 billion (Mitchell (2021)). Because of this, Amazon is able to maintain its dominant market position and continue to profit from it. Even though regulators are continuing their careful inspection of the company, there is little evidence to suggest that Amazon's exercise of market power would directly affect independent business owners at this time. Because of this, Amazon can maintain its dominant market position and continue to profit from it. Even though regulators are continuing their careful inspection of the company, there is currently very little evidence to suggest that Amazon's exercise of market power would directly affect the owners of independent businesses.

In this paper, I examine the relationship of how the growing expansion of amazon fulfillment centers on the incomes of different types of business owners during the last two decades. With the entry of Chinese and other foreign sellers into the amazon marketplace in 2015, Amazon's FBA (Fulfillment by Amazon) project has started to be profitable for the first time since its inception. While already in the last fifteen years, the independent business owners across the USA have been showing a declining trend (US Census Bureau), this sudden entry of competitive foreign market sellers may have had some impact over the years on the share of local business owners at the county level– which I would also like to explore in my paper.

In light of this concern, I utilize a database of Amazon fulfillment centers to examine the impact of their presence on local income and employment at the county level. Importantly, I develop an instrumental variable (IV) technique leveraging Amazon's expansion to account for the potential endogeneity of their entry, which can contaminate estimates of the effect on earnings. I believe that my approach significantly enhances the existing body of research on these and related questions, particularly through the implementation of this improved methodology. The most important aspect of this improvement is the implementation of an identification strategy that considers the endogeneity of fulfillment center location and timing of entry, as well as how these factors may be correlated with future changes in earnings or employment. It has been suggested that amazon's strategy was to locate in near the bigger cities from where the tax and labor cost both are higher than average since 2011 (Houde et al. (2017)) , and it is reasonable to expect that it will be situated near the cargo airports since FedEx could achieve the 49% of express delivery (2012) in the domestic USA only by having dominance in air shipping. Furthermore, both FedEx and UPS offer next-day delivery with their air shipping option. This set of information bolsters the concept that if amazon wants to achieve its goal of same-day shipping, setting up its distribution network near cargo airports is the only way to go further.

To account for the endogeneity of Amazon's entry, I utilize two sets of instrumental variables: IV A and IV B. IV A represents the interaction of distance from cargo airport with year dummy. The determination of the fulfillment center's location within a county is primarily influenced by the minimum distance from the closest cargo airport, while the temporal aspect of its establishment is typically determined by the utilization of year dummies. IV B corresponds to the distance from the intersection of the interstate highway, as well as the incorporation of year dummies. These instruments are carefully constructed based on the relevant economic literature, taking into account the unique characteristics of Amazon's expansion strategy and the local labor market dynamics. By employing these instruments in my 2SLS models, I can obtain unbiased estimates of the causal effects of Amazon fulfillment centers on earnings, employment, and the number of establishments. By employing a meticulous instrumental variable approach and considering the specific instruments I construct based on existing literature, my study aims to provide a detailed and nuanced understanding of how Amazon's presence affects the local business owners. In addition, there are additional biases that are present, such as the problem that was mentioned earlier, which is that it is impossible to accurately identify the source of any changes that occur at the county level in terms of employment or wages. This is because there are a variety of factors that could be responsible for a change in employment or wage levels within a county that is unrelated to the establishment of an Amazon fulfillment center. The paper makes use of two-way fixed effects regressions, the details of which will be provided later in the paper, in order to take into account these issues.

Although e-commerce can encourage the entry of new firms into a market by facilitating greater concentration, it may also lead to increased dominance among large firms in some market where the industries are characterized by high entry barriers. Using data on U.S. county business patterns from 1994 to 2003, Goldmanis et al. (2010) conclude that as the use of e-commerce spreads among customers, major corporations gain market share at the expense of smaller ones.

Thus, the contribution of this paper to the literature is to find a definitive answer on how a physical expansion of an ever-growing amazon distribution network directly affects independent business owners at the county level. My results indicate that as amazon sets up a fulfillment center in a county it shows no predictive power on the income of non-incorporated business owners. Upon further investigation, my work reveals that in response to the setting up of the amazon fulfillment center the probability of transition to become a local business owner from a full-time employee - is positive and significant. The coefficient of IV which is almost double the coefficient with OLS regression Therefore, it shows the magnitude of omitted variable bias in OLS estimates when I are not dealing with the endogeneity issue of the location of the amazon distribution setup decision.

The most important aspects of endogenous growth theory lies under the assumption of how technology is the exogenous factor that will fuel sustainable economic growth keeping everything else constant. On that note, some academics have made a bold claim, arguing that not only has innovation under-performed during the last few decades, but that it will be extremely challenging, if not impossible, to achieve high levels of economic growth in the future (Cowen (2011); Gordon (2016)) Yet another camp disputes this pessimistic outlook by noting the immense growth potential of entrepreneurs (Guzman and Stern (2020)) and the incredible potential of emerging technologies whose economic influence has yet to be fully realized. As the diffusion of e-commerce is the bedrock of new wave of technology in retail and production sector, how the spread of this technology is contributing to user centered innovation and manufacturer centered innovation is a burning question. In this paper, I try to answer this question by providing empirical results how the diffusion of e-commerce at county level of United States has been affecting innovation level by incorporated and non-incorporated firms.

The remaining parts of the paper are organized as described below. The literature is discussed in the second section. In the third section, I offer some insights into why amazon has been investing in acquiring lands near the cargo airports, as well as how this will impact the foreseeable logistics industry if amazon successfully wants to keep the promise of same-day delivery. Section four outlines my data sources and shows my identification approach. In the fifth and sixth sections, I present not only descriptive statistics but also my most important empirical findings, as well as the outcomes of several different robustness and specification tests. In the final and concluding section, number six, I will go over the consequences of my findings.

1.2 Background of Amazon Fulfillment Center (FC)

Prior to 2005, Amazon only had a few fulfillment centers (FC) under its management, and the company chose its FC locations with the primary goal of maximizing its advantage in terms of sales tax (Houde et al. (2017)). Amazon's priorities have altered as the company has developed over the years. At this point in time, avoiding taxes takes a back seat to ensure timely delivery. As a result, Amazon has been hard at work constructing warehouses in proximity to urban centers in states with significant populations (Houde et al. (2017)).

Amazon spent nearly $50 billion on facilities and acquisitions in the four quarters leading up to April 1, 2021, which is double what Google spent during the same time period (MWVPL). This change is intended to help Amazon compete more effectively against its competitors, such as Walmart and Target, who both use their local stores to facilitate online ordering, followed by either curbside pickup or home delivery of their products. The previous year, Amazon stated that it was working to shorten the delay on Prime shipping, stating that it would bring Prime from a delivery time of two days down to just one day. It is taking measures toward being able to serve Prime members in an even more timely manner by installing fulfillment facilities in local markets.

Not only will Amazon foresee the future as a significant market shareholder of the e-commerce industry but will also foray into its frenetic growth to have a bigger slice of the global trade logistics market (Figure 3.2). According to Dave Clark, CEO of Amazon's worldwide consumer business, on Monday, the company is expected to surpass its longstanding shipping rivals UPS and FedEx by early 2022 and become the largest package delivery service in the United States. Palmer (2022) If they become vertically integrated into this sector, they will seize a significant portion of the market share. To accomplish this, they are incentivizing their sellers to store products in their logistics instead of using FedEx or UPS. The average distance to the consumer fell from 308 miles in 1999 to only 67 miles in 2018. Most of this decline is due to the expansion of the distribution network into the most densely populated states along the coasts in the mid-2010s. (Houde et al. (2017)) This is the goal Amazon has in mind when they are aiming to break the logistics oligopoly the existing companies have. To break the chain and reduce the shipping cost by entering the labor cost expensive areas, they are aiming to take a major share of logistics trade besides being a giant US of e-commerce. The question at hand is how Amazon will adjust a strategy designed for a time of economic expansion to the conditions of the present market. The growth of online sales is decelerating, interest rates are climbing, and some analysts predict that a recession will begin within the next few months. Considering the long-term objectives of enhancing cash flow and expanding market share, the proposed shift towards dominating the logistics industry would confer an unparalleled competitive edge over major retail competitors like Walmart, Target, and Kroger.

Therefore, it is an extremely remote possibility that Amazon selects the locations of its fulfillment centers at random. When it comes to maximizing efficiency, the location of a company's fulfillment centers is one of the most important factors to consider. Since the primary function of fulfillment centers is to make the storage and transportation of goods more streamlined, a company like Amazon, which is known for its lightning-fast customer service, must place particular emphasis on this factor. During the time period under consideration, Houde et al. (2017) estimations suggest that the typical cost of delivering an order placed with Amazon fell from $2 to $0.30. This finding brings light to the potential supply chain efficiencies associated with vertical integration into sortation – signifying the importance of placing fulfillment centers strategically across the United States.

1.3 Literature Review

Innovation is significant component of the role that small businesses play in driving productivity, economic expansion, and a reduction in unemployment (Acs 1999; Robbins et al. (2000); Waite (1973)). Jia (2008) came up with a model in which large discount retailers like Walmart and Kmart make simultaneous moves, and smaller discounters decide whether or not to enter the market. The entry of large chains is contingent upon the local economic conditions, an awareness of the movements of competitors (both large and small), and an incentive to establish outlets in close proximity to markets in order to reduce operating expenses. The entries for smaller stores are based on those for larger chains. Jia (2006) found that areas with a Wal-Mart or a Kmart had two to three fewer establishments catering to residents on a limited budget. Around the middle of the 1980s, the typical county in the United States had fewer than four discounters; therefore, Wal-Mart is primarily responsible for the collapse of these businesses Basker (2007). At the same time, this opens doors of opportunity for others to come in. The "net mortality" or "net entry" of various types of businesses in the same or adjacent locations is therefore an important factor to consider.

Ellickson and Grieco (2013) found that Walmart's impact largely falls on larger businesses, with little on minor players. Walmart's influence is confined, affecting only enterprises within a two-mile radius. The main impact of the decline is concentrated on the intensive margin and on firms that are experiencing a decrease in productivity within the specified radius. New firm entry is unrestricted. Basker (2005) estimated Walmart's sales by utilizing instrumental variables, and their findings included all of the stores. In the five years following Walmart's entry into a county, four smaller competitors go out of business. The impact of Walmart's presence is mitigated by the fact that the typical county contains more than 200 independent businesses.

According to Paruchuri et al. (2009), the entry rate of competing businesses is expected to decrease in close proximity to the Walmart (i.e., within the same zip code), their exit rate is expected to rise, and as a result, their net entry (i.e., entries minus exits) is expected to decrease in comparison to their levels before the Wal-Mart opened. These effects should be strongest initially but then begin to decline after the competitors with the weakest market share exit and cumulative exits begin to attract new competitors who want to enter the market.

Despite all of this, many local officials are still interested in providing financial support for Amazon's growth because they believe it will bring jobs and revenue to cities and towns that are struggling, particularly during a pandemic that has devastated local businesses and government budgets. But according to the Economic Policy Institute, new warehouses don't bring wider employment growth to a local economy. They either inhibit other forms of economic activity or have an effect that is insignificant enough that it cannot even be measured (Jones and Zipperer (2018)). The growth of Amazon's seller fees has outpaced that of any other major revenue stream, including Amazon Web Services (AWS) (Mitchell (2021)). Among other things, it allows Amazon to exert control over pricing far beyond its own platform, which suggests that its market power in e-commerce is even greater than its roughly 50 percent market share would suggest.[1] The impact that Amazon has

1. The House Judiciary Committee's digital markets investigation concluded, "Although Amazon is frequently described as controlling about 40% of U.S. online retail sales, this market share is likely understated, and estimates of about 50% or higher are more credible." See: "Investigation of Competition in Digital Markets," U.S.House of Representatives, subcommittee on antitrust,

on conventional retail is another way that it ravages local economies. According to research conducted in 2015, retail outlets that had their sales displaced by Amazon sales ended up owing an additional $528 million in property taxes. And this is before I even begin to discuss the particular tax breaks that have been provided to Amazon, which typically take the form of reductions in the company's property taxes (Civic Economics (2015)). New research finds that Amazon has negotiated lucrative public subsidies for more than half of the 77 fulfillment centers and other large warehouses it built between 2005 and 2014. After Amazon began collecting sales tax in 2016, Baugh et al. (2018) tracked the spending of 275,000 households and found that customers reduced their spending on the website by 9.4 percent overall and by 29.1 percent on items priced over $250 after the company announced it would begin collecting sales tax.

Using the first date of planning for a Walmart discount store as a proxy for the entry of Walmart into counties in the United States has allowed researchers to circumvent the issue of temporal endogeneity (Basker (2005), Basker (2007)). It has been argued that the opening date of a store can be manipulated to coincide with favorable conditions, whereas the date on which the store initially plans to open cannot be manipulated in this way. This is because the opening date of a store can be manipulated to coincide with favorable conditions. Using a measure that acts as a proxy for the date that a store's initial planning began, she argues that while the precise timing of a store's opening can be manipulated to coincide with favorable conditions, planning is done far enough in advance that it is not likely to be endogenous to a growth spurt — or sudden decline — exactly coinciding with Walmart's entry. This is because planning is done in sufficient advance. One of the most significant limitations of this IV strategy is the fact that the "instrument" (store planning date) is only defined for locations in which a Walmart store was eventually opened. This is one of the most significant limitations. Because of this, it is challenging to arrive at any definitive conclusions regarding the effect that the presence of a Walmart store has on communities that do not have one. Both Dube et al. (2007) and Neumark et al. (2008) exploit the fact that counties enter Walmart's sphere at different times, depending on their distance from Benton County. These researchers did their research independently and published their findings in 2005.

Because it captures an "intent to treat" and therefore, in principle, allows estimating what Angrist and Imbens (1995) call a "local average treatment effect," the distance instrument is intuitively appealing because it is clearly exogenous to Walmart's entry. Additionally, it has the potential to be much more powerful than the planning date that was used by Basker (2005), Basker (2007).

I am putting together an interaction variable that will tell me about the minimum distance from each county to the nearest cargo airport with year dummy which will tell me how with time Amazon kept changing it's fulfillment center location strategy.

commercial and administrative law of the committee on the judiciary, 2020, at 297.

The minimum distance from the nearest cargo airport will determine the location, and year dummy, in most cases, will determine when the fulfillment center will enter the county, just as Neumark et al. (2008) and Dube et al. (2007) did for Walmart's entry. Moreover, i am using another set of instrumental variable which captures the minimum distance from each county to the nearest intersection of interstate highway. Previous later suggests that placing fulfillment center near interstate highway is a very common scenario amongst the supply chain related companies (Dablanc and Rakotonarivo (2010); Kang (2017)). Therefore, I am interacting the distance from intersection of interstate highway with year dummy as my second set of IV.

To this day, there are an increasing number of academics, policymakers, and advocates for public interest who are arguing for the restoration of the more comprehensive set of concerns regarding amazon. The success of big companies often displaces weaker rivals in a given geographic area. Although Amazon benefits consumers in many ways, the impact of this retailer on local businesses has been mixed. On the one hand, some small businesses have been helped by Amazon's lower prices and improved inventory control. On the other hand, the number of jobs in the local retail industry has declined and the number of independent retail stores has closed in many communities. This overall effect is partially determined by the proximity of small local businesses to the Amazon fulfillment center as well as the overlap that exists between the two.

1.4 Data and Descriptive Statistics

I collect the data for this study in several steps. I estimate the effects that Amazon fulfillment centers have on business owners using data from the 2000–2021 of the march ASEC sample of Current. My outcomes are counts of the number of people who are business owners, along with income variables such as total wage income in logarithmic form by taking care of it with cpi99 variable. Besides, total wage income I also try to see if due to having an efficient distribution network of the e-commerce platform, there is any downward pressure on the working hours of existing and new business owners. I also convert the hours into logarithmic form. I estimate instrumental variables (IV) models that leverage the predictable geographic expansion patterns of Amazon fulfillment geographically situated near an opportunity zone where a cargo airport is available. Specifically, I instrument Amazon fulfillment with the interaction of the distance ring from the nearest cargo airport, with year dummies for setting up an amazon distribution network from the local/state government. The average wage is defined as the annual earnings divided by weeks worked and usual hours worked per week. The summary statistics of the treated counties and control counties are discussed in detail in Table 3.1. Demographic information includes the proportion of females, average age, the proportion of non-white, and proportion with a high school education or less—for each of the three groups of workers. This information is then merged with the county-level database of Amazon fulfillment center openings.

1.4.1 Definition of Business Owner in the Current Population Survey (CPS)

Fairlie and Fossen (2020) established a measure of business activity by using the data collected by the CPS throughout time. This measure accounts for all new business owners, including those who own incorporated or unincorporated enterprises, as well as those who are employers or do not employ anyone. Following their method, I use the information on people's primary jobs, which I define as the jobs in which they put in the most hours of work, to determine whether individuals are business owners throughout each month. Therefore, people who start their own businesses on the side will not be counted if they also work more hours at a job that pays them a wage and salary. In order to ensure a reasonable work commitment to the new company, the criterion of working 15 or more hours per week (which is equivalent to roughly two or more days per week) has been selected.

1.4.2 Cargo airports

The FAA annually asks airports to report their annual cargo landings. Please note that all-cargo reporting is limited to aircraft operations dedicated to the exclusive transportation of cargo. Here, to track the air cargo data I use one annual database sourced from the Federal Aviation Administration (Federal Aviation Administration) Since years 2000-2021 I have collected the data for all the cargo airports that carry the highest tonnage of cargo goods throughout the whole time. This eventually summed up to a total of 100 cargo airports situated in different counties across the united states.

1.4.3 Distance between counties

The Haversine formula is used to determine the great-circle distance between any two places inside a given county. These distances are referred to as county distances. This dataset of counties is uploaded on NBER public website which also provides the distance between counties which I eventually used to find the minimum distance from counties with cargo airports to other counties. (NBER). If a county has at least one cargo airport in it, then the distance variable for that county is zero for my estimation.

1.4.4 Distance from intersection of interstate

This study aims to measure the distance from each county in the USA to the nearest intersection of interstate highways using geospatial analysis techniques. To achieve this goal, two essential shapefiles are required: a shapefile of USA counties (Census Bureau (2023)) and a shapefile of interstate highways (Federal Highway Administration (2023)) that intersect or pass through the counties of interest. The data was obtained from the US Census Bureau. The shapefiles should be loaded into a Python geospatial library, such as GeoPandas, to facilitate the analysis. To ensure the accuracy of the analysis, it is important to ensure that both shapefiles are in the same coordinate reference system (CRS). If not, one of the shapefiles must be reprojected using the to_crs method in GeoPandas to reproject the shapefile to a different CRS. This step is crucial to guarantee that the measurements between the two shapefiles are consistent.The sjoin_nearest method in GeoPandas was used to join the county and interstate shapefiles based on the nearest intersection. This method will create a new GeoDataFrame that includes the county name and the distance to the nearest interstate intersection for each county.

1.4.5 Fulfillment centers

Fulfillment centers came from MWPVL International, a company that specializes in the supply chain, logistics, and distribution. (MWVPL). This unique source provides information on all of Amazon's fulfillment centers, including their locations, opening years, types, and sizes in square feet. In addition, the data that was collected by MWPVL International identifies the type of fulfillment centers that are used by Amazon. For example, the data shows whether the facilities are normal large sortation centers or Prime Now Hubs. With the use of these statistics on the various types of fulfillment centers, it is possible to determine whether there is a major difference between the various types of fulfillment centers in terms of the impact they have on the employment and wage situation in the surrounding area. The data collection also includes information regarding the amount of smaller fulfillment facilities that have either been shut down entirely or merged with other, larger centers.

1.4.6 US Patent Data

The United States Patent and Trademark Office is at the forefront of technological advancement and achievement in the United States. The ongoing need for patents and trademarks demonstrates the resourcefulness of the people who create new businesses and products in the United States. Therefore, according to the scope of my research I have collected patent data to proxy for innovation at county level. (U.S. Patent and Trademark)

1.5 Identification Method

1.5.1 Explanation of using IV

Jakubicek and Woudsma (2011) conducted a study that found that proximity to transport infrastructure, particularly highways, was highly valued by logistics managers when selecting a warehouse location. Similarly, Dablanc and Rakotonarivo (2010) argued that companies within the logistics industry tend to locate their facilities as close to highway networks as possible, but also in proximity to airports. Proximity to transport infrastructure, therefore, remains a key consideration when selecting a warehouse location. Kang (2017) notes that for companies involved in international trade, it is more important to have warehouses in proximity to seaports or airports compared to distribution companies serving the local market. This is because international trade requires efficient connectivity between various modes of transport and warehouses located near seaports or airports can ensure a seamless transition between transportation modes.

By leveraging existing transportation infrastructure, Amazon can reduce its shipping costs by placing fulfillment centers near transportation hubs. This can allow Amazon to offer lower prices to customers and gain a larger market share against competitors. Locating fulfillment centers near transportation hubs can also enable Amazon to manage its inventory more efficiently. This can facilitate faster and more reliable order fulfillment, thereby providing Amazon with a competitive advantage over UPS and FedEx. Planning wise it's possible when you place your warehouse near interstate highways and airports which is also solidified with existing literature in supply chain logistics mentioned earlier. Firstly, the location of highways and airports is often influenced by physical geographies, such as mountain ranges or bodies of water, rather than economic or demographic factors. Consequently, the distance from these facilities may be unrelated to any economic outcomes. Secondly, historical factors play a role in the location of highways and airports. For instance, the placement of these facilities has been determined by the location of previous transportation routes or military bases. As a result, the distance from highways and airports may be random and unrelated to any economic outcomes. Thirdly, zoning laws and regulations often restrict where highways and airports can be located. Thus, the placement of these facilities may be random and not driven by economic factors. Fourthly, the economic factors influencing the location of highways and airports can change over time, leading to a random distribution of distances from these facilities. For instance, as cities grow and expand, new highways and airports may be constructed in locations that were previously considered rural or undeveloped. Finally, land use changes, such as the development of new residential or commercial areas, can also impact the distance from highways and airports. As a result, the distance from these facilities may be random and unrelated to any economic outcomes.

Several studies have provided evidence to support the argument that the location of highways and airports can be influenced by factors unrelated to economic outcomes. A study by Levinson and Krizek (2017) found that the location of highways in the United States was influenced by historical factors, such as the location of previous transportation routes, rather than economic or demographic factors. Furthermore, Kulkarni and Pande (2019) found that the placement of highways in India was often determined by topography and land availability rather than economic or demographic factors. These studies suggest that the location of highways and airports can be impacted by various factors, which are unrelated to economic outcomes. This finding indicates that distance from these facilities could be a useful instrumental variable for the placement of a warehouse, allowing me to control for unobserved variables that may impact the placement decision. Because of this overall random nature, the instrument (interaction of distance from cargo airport / distance from intersection of interstate highways with year dummy) should not have any effect on the economic climate of the surrounding area, apart from the effect that it has on the locations where Amazon is starting its fulfillment center.

1.5.2 Assessment and Interpretation of Baseline Results

I then proceed to carry out a number of exercises with the goal of determining which groups are subjected to the baseline treatment effects that were investigated in Section 6 how the instrument is functioning, and to what extent the geographic scale and outcome measure is contributing to the results. This is done by conducting a number of exercises with the purpose of determining which groups are subjected to the baseline treatment effects that were investigated in Section 5.

1.5.2.1 Assessing Relevance and Validity of the Instrument

To further interpret these results of the instrumental variables that are shown in Table 3.2 and Table 3.3, it is necessary to determine the extent to which the instrument—the interaction of minimum distance from cargo airport and the year dummies—satisfies the necessary assumptions for unbiased estimation. This can be done by determining the degree to which the instrument satisfies the necessary assumptions for unbiased estimation. Examining the coefficients in the "First Stage" column of Table 3.2 enables us to ascertain whether or not the instrument is important by providing us with this information. These coefficients demonstrate that the instrument, at each distance threshold, has a significant and accurately measured association with each outcome in the same region. This was determined by comparing the instrument's measurements to the outcomes. If the county is within the distance ring of 100 miles of a cargo airport with each passing year for setting up an Amazon distribution network then it is associated with a 0.4097 standard deviation (SD) increase in the probability of having an Amazon fulfillment center. The significance of the instrument in the field can be demonstrated by its highly effective first stage, which operates across multiple bandwidths. Given that standard errors are clustered, the Kleibergen-Paap F-statistic would be an appropriate indicator of relevance in this scenario. This statistic is equivalent to the F-statistic of the first stage regressions in my scenario because there is only a single endogenous variable (Amazon fulfillment center) (Kleibergen and Paap (2006)). The fact that the F-statistic is higher than the typical cutoff of 10 lends credence to the notion that the constructed instrument is relevant as a proxy for the Amazon fulfillment center.

One other method for determining whether or not my set of IVs is valid is to carry out an over-identification test using multiple instruments, focusing on the endogenous variable of whether or not there is an Amazon fulfillment facility in the county. My multiple IV does the work of conducting the Hansen J statistics so that this can be accomplished. A p-value of 0.498 was found to be associated with the Hansen J-statistic testing of over-identification when the IV specification in Column 3 Row 8 of Table 3.3 on change in total income for non-incorporated business owners caused by a presence of cargo airport within 100 miles of a county which had a subsidy from the government for setting up fulfillment center - the instrument was taken into consideration. None of the Hansen J statistics p-value is less than 0.05 in the results presented in Table 3.2 and Table 3.3. This demonstrates that I am unable to reject the null hypothesis that my instruments are producing the same results, which provides me with extra confidence that the primary instruments are credible.

The exclusion restriction is another requirement for a valid set of instrumental variables. That is, it should affect dependent variables in the second stage only through the endogenous variable. Within this framework, the created IV should only affect the outcomes at the county level through the establishment of an Amazon distribution network. In the previous part of this article, I went over the reasons why the built IV is less likely to have problems with endogeneity. However, even the constructed IV could be subject to concerns regarding the exclusion restriction if there are enough of them. In order to address such issues, I carry out a series of exercises, which are described in more detail below.

To begin, there is a possibility that one would be concerned that an increase in the number of Amazon fulfillment centers might take away a portion of the money earned by the owners of other incorporated businesses. If this were the case, then the estimated effects of the Amazon distribution network using my suggested IV method would be confounded by the effects of the revenue of the owners of the other incorporated businesses. I address this concern in a robustness check displayed in Column (6) of Table 3.2 and Table 3.3.

Specifically, I identify incorporated business owners who are mostly large or medium-level cooperating owned business or who have large franchise all over the United States, that performs similar tasks as non-incorporated business owners but on a grand scale. As a result, for them they have their own supply chain network, therefore they are not as dependent as non-incorporated business owners on an e-commerce supply chain setup. I construct a binary variable corresponding to whether a business owner is incorporated and regresses their total wage income on the IV regression setup in a specification analogous to Equation 1. Upon conducting Ordinary Least Squares (OLS) and Instrumental Variable (IV) regressions, the findings presented in Table 3.4 demonstrate a positive influence of the Amazon fulfillment center on the transition to non-incorporated business ownership at the county level, as evident in Columns 2 and 3 of Table 3.4 (coefficients of 0.011 and 0.009 respectively). Both OLS and IV results exhibit statistical significance, although the magnitude of the IV coefficients is 1.5-2 times larger than that of the OLS coefficients, indicating the presence of omitted variable bias under OLS. Conversely, my estimation fails to yield statistically significant evidence of the impact on transitioning to incorporated business ownership, despite the negative signs (-0.002, -0.003, and -0.001) observed in both OLS and IV results for Columns 4, 5, and 6 of Table 3.4. The fact that the probability of being transitioned to incorporated business owners in this reference group is not predicted by the Amazon fulfillment center at a level that is statistically significant (Column (5) and (6) of Table 3.4 suggest that I have managed to capture changes in the income of non-incorporated business owners who are specifically complementary to local business owners but who are not part of large franchise holders or large corporate companies. These findings provide evidence that supports the hypothesis that the IV can only influence the outcomes of non-incorporated business owners by acting via the medium of Amazon's fulfillment centers.

1.5.3 Basic model setup and Estimation

My instruments are the dummy year variables times the minimum distance to the nearest cargo airport/ the minimum distance to the nearest intersection of interstate, so it is designed the capture how Amazon has been trying to redesign its fulfillment center to take advantage of its geographical position close to the opportunity zone where they not only have the access to a significant portion of prime subscribers but also it helps to reduce the shipping cost to outdo other counterparts like Walmart and target. Denoting the instrument by $\theta_d(dist_d * YR_y)$, I estimate the following system of equations by two-stage least squares:

$$Amazon_{ct} = \alpha + \sum_{d=1}^{D} \beta_d dist_d + \sum_{y=1}^{Y} \beta_y YR_y + \sum_{j=1}^{J} \beta_3 X_{jict} + \sum_{d=1}^{D} \theta_d (dist_d * YR_y) + \gamma_c + \epsilon_{ict} \quad (1.1)$$

$$lnY_{ict} = \delta + \psi_q \widehat{Amazon}_{ct} + \sum_{y=1}^{Y} \psi_y YR_y + \sum_{d=1}^{D} \psi_d dist_d + \psi_r X_{jict} + \gamma_c + v_{ict} \quad (1.2)$$

where $\widehat{Amazon}_{ct}$ are the fitted values from the first stage regression 3.1. I put the IV strategy into action by segmenting the United States into 17 distance rings, each of which represents a 100-mile increase in distance from the closest cargo airport (for example, fewer than 100 miles, 100–200 miles,..., 1,600 miles or more), and developing an indicator variable for each of these distance rings.[2] The distance ring dummies are included as controls ($\beta_d dist_d$), while the interactions of the distance ring dummies with year dummies are used as instruments. I cluster standard errors by county. X is the vector of individual-level controls including age, age squared, marital status, sex, race, Hispanic, and education. The county fixed effect is shown by γ_c which helps to account for time-invariant differences across counties, and ϵ_{ict} is the error term.

This regression 3.2 differs from the regression 3.1 by replacing the (year dummy*distance to the nearest cargo airport) with the predicted values of $Amazon_{ct}$ estimated in equation 3.2 and adding the distance ring fixed effects.

Identification of ψ_q in equation 3.2 in the IV model comes from the assumption that the year dummy * distance to the nearest cargo airport interactions can be excluded from the second-state regression 3.2—that is, that these interactions are uncorrelated with changes over time in the outcome of - income and employment of worker- on the controls. By including the distance ring fixed effects in an equation 3.2, I allow for the distances to be correlated with levels of income and employment; I only need to assume that they are uncorrelated with trends.

1.5.4 Impact on income of business owners

The next specification will explore the same question, however using an instrumental variable regression. In a two-stage least squares regression, this subset of sample counties that has the amazon fulfillment centers in the sample is so-called compilers. Compilers are precisely the labor force of those counties that are induced to change their work profile due to having access to the e-commerce physical setup near their county.

I have two primary goals: the first is to show that inaccurate modeling leads to imprecise results, and the second is to investigate both the direction and size of the effect that the Amazon fulfillment center and its supply chain network has on the outcomes of entrepreneurial endeavors. The most important findings are outlined in Table 3.2. The simple OLS estimation that does not factor in any controls reveals a negative and insignificant relationship between income earned by non-incorporated business owners and the presence of an Amazon fulfillment center (Column 1).

In the second and fourth columns of estimations presented in this Table 3.2, I take endogeneity into further account using IV. Lastly, the coefficient on wage is shown to have no discernible impact using IV regression (Column 2 of Table 3.2). This occurs when a complete set of controls is considered as well as the IV strategy.

2. The 100-mile distance ring classification follows Neumark et al. (2008) and Dube et al. (2007)

The ordinary least squares method has a downward bias because of the classical measurement error in the independent variable, which the IV may attenuate. The source of endogeneity arising from the location decision of amazon fulfillment center is biasing down the estimate from OLS. According to the findings of Table 3.2, the implementation of IV does not have a discernible impact of fulfillment centers on incomes for local full time business owners. This is consistent with the results shown by OLS. The variable of working hours are unaffected as well in case of both methods- OLS and IV. F-stat stands for Kleibrgen Paap (Kleibergen and Paap (2006)) first stage F statistics. I have two set of IVs here- the first one is distance from cargo airport interacting with year dummy. The second set of IV is distance from intersection of interstate highway interacting with year dummy. After using both of set of IV, my result fail to show any impact of amazon fulfillment center on labor outcome of non-incorporated business owners in Table 3.2.

Like the prior result documented in Table 3.2 the OLS estimate is half of the LATE estimate of IV. This gap between OLS and IV estimates are arising due to misspecification or omitted variable bias in OLS which is eventually taken care of by the LATE estimate of IV method. Therefore, the LATE estimate is presenting how this amazon distribution network has higher impact on compilers whereas OLS method is associated with the total population sample.

In Table 3.3 I have presented the two set of IVs I am using as identification method in this paper. The first set of IV is the distance ring from nearest cargo airport interacting with year dummy. This IV is represented by IV A in the table. The second set of IV is distance ring from the nearest intersection of interstate highway interacting with year dummy, and this is represented by IV B. None of this IVs show any significant impact on the income of non-incorporated business owners (Column 2 and 3 of Table 3.3).

1.5.5 Likelihood of transitioning to become a business owner

In Table 3.4 I assess the situation of how one person transitions to being a full-time business owner. For linking individuals, CPS ASEC provides an indicator CPSIDP through which you can link the same person who is getting interviews at different waves of 16 months cycle (4-8-4 cycle) of CPS where they interview 60,000 households for four consecutive months (Flood et al. 2018), then give them a break for 8 months, and they take their interviews for 4 consecutive months again till they never track the person again in their further process of interview even though they participate under a different CPSIDP later stage. By using the CPSIDP I created the longitudinal transition variable who were not business owners before but 1 year later when they are being interviewed again they have become full-time business owners. Then I created additional binary variables for being transitioned to the non-incorporated full-time business owners as well as incorporated full-time business owners. After running the OLS and IV regression my result in Table 3.4

shows that the amazon fulfillment center has a positive impact (.011 and .009) on being transitioned to non-incorporated business owners at the county level (Columns 2 and 3 of Table 3.4). Though both (OLS and IV) of these results show significance, the magnitude in IV is 1.5-2 times as great as in OLS - referring to the omitted variable bias impact under OLS. On the other hand, my estimation fails to predict the impact of being transitioned to be an incorporated business owner at a statistically significant level, though the sign is negative (-.002, -.003 and -.001) in both OLS and IV results (Table 3.4 Columns 4, 5 and 6).

It has been argued by a number of economists, including David Autor, that one of the issues plaguing the economy of the United States is the disappearance of mid-wage jobs. These are jobs, particularly in the manufacturing sector, that pay well in relation to the level of education.(Autor (2019)). Until recently, there has not been a widespread consensus regarding the reasons why automation has been linked to polarization. Acemoglu and Loebbing (2022) present an alternate and complementary explanation that these tasks are the most profitable ones to automate. Cortes et al. (2020) show in their paper that despite non-routine occupational growth in aggregate, declining inflow rates to routine occupations for these groups have not been accompanied by increasing inflow rates to non-routine occupations. In spite of the growing literature on polarization, relatively little is known about where this mid-skill workers are transitioning to.

In my paper I present that in counties where there supply chain network of amazon is getting strengthened, it facilitates those mid-skill workers to be full time business owners. According to the result presented in Table 3.5 of Column 2 show that the presence of fulfillment center in a particular county is associated with the probability of mid-skill worker to be transitioned to a full-time business owner by .016 standard deviation. Additionally, the coefficient is .009 for high-skill employees but for the transition of low-skill employees there is no discernible impact of this amazon distribution network. Using the second set of IV, I get almost the similar results from Columns 4, 5 and 6 of Table 3.5

1.6 Staggered treatment effect of e-commerce on innovation

I have already seen in my paper that with the expansion of Amazon, people are more willing to be transformed into business owners, but is this also opening the door for innovation - which may result in long term sustainable growth - is another question I would like to explore here using data from the USPTO at the county level.

Although the two-way fixed effects estimator nearly always suffers from negative weighting problems (Goodman-Bacon (2021), Baker et al. (2022)) , the two-way fixed effects treatment parameter β_{TWFE} is less biased when treatment effects do not grow over time, or, in other words, when β_{TWFE} and β_{FD} are similar to each other. Using the Callaway and Sant'Anna (2021) estimator, developed for staggered implementation designs, I estimate the diffusion of e-commerce on innovation capacity of firms and individuals. Clean controls are used as never-treated groups in this estimator, which takes into account post-treatment covariates. Lastly, I compare the results of these estimators with the problematic but familiar two-way fixed effects.

Despite taking care of staggered timing and heterogeneous treatment, the endogenity problem is still a concern in my setup. Therefore Synthetic control is particularly useful as an alternative to difference-in-differences in the case of this research. One distinction is that DID makes the assumption that everyone is affected by time shocks (time FEs) in the same way and that the only thing that differentiates them is their levels (unit FEs) On the other hand, synthetic control operates under the assumption of a latent factor model, which is somewhat more comprehensive in the sense that it acknowledges the possibility of time shocks and the fact that units might be differentially exposed to shocks.

The synthetic control method (SCM) takes the results of the control groups and uses a weighted average of those results to predict the results of the adopting groups "as if" those groups had not adopted the treatment. The weights are selected to optimally match the outcomes of the adopting counties prior to the adoption, and as a result, they account for any possible trends that could influence identification without the need for an assumption of parallel trends. The estimated treatment effects from the approach are equal to the difference between the actual outcomes that occurred after adoption and the outcomes that had been projected (Abadie et al. (2011), Abadie and L'Hour (2021)). To address these challenges I utilize synthetic difference-in-differences (Synth- DiD) based on Arkhangelsky et al. (2021), which borrows strengths from the DiD method as well as the synthetic control method (Abadie et al. (2011), Abadie and L'Hour (2021)).

To reduce the mean squared error of the predicted target ATT, SynthDiD determines the best weights for control units and pre-treatment. While SCM relies on the premise of strong parallel trends for identification, SynthDiD leverages pre-treatment data of control units to generate a synthetic control for the average outcome of treated units.

Table 3.7 shows the results from different methods how innovation capacity is getting affected as amazon expands throughout all the counties over the years. The ATT of Callaway and Sant'Anna (2021) estimator reveals that the diffusion of e-commerce has a negative impact on the total patents filed in that particular county which has an Amazon fulfilment center but the impact is insignificant. Interestingly, the co-efficient is almost two-third of the coefficient of TWFE estimate which ensures the negative weighting problem TWFE suffers from. According to TWFE estimate in Column 1 of Table 3.7 e-commerce supply chain of amazon has a negative impact on user oriented and manufacture oriented innovation which is significant unlike Callaway and Sant'Anna estimate. On the other hand, point estimates for the average treatment effects using Synth-DiD suggest there is no discernible impact of e-commerce supply chain of amazon on county level of either user oriented or firm level innovation. This reflects the same impact as of Callaway and Sant'Anna estimate but the ATT of Synth-DiD is one fourth of the prior estimate.

1.7 Robustness Check

In this section, I will present and discuss the results of my baseline regression, which will measure the absolute effect of any fulfillment center exposure as well as the relative effect of more nearby cargo airports in a variety of formats In the following section, I will access these results even further and interpret them by testing the underlying assumptions, describing the groups that were most susceptible to this shock, and dissecting the contributions that scale and my outcome measure made.

The results are presented for income outcomes: a change in the income of people who are non-incorporated business owners. In Table 3.6, the results using a wider bandwidth are presented in descending order for each row (d). Using the first entry in the Column labeled "Earning(100 miles) " as an illustration, I have discovered that if a cargo airport is within 100 miles of the county then it leads to have no income change for non-incorporated business owners. This effect remains the same in magnitude as bandwidth goes down to 50 miles. Despite, as the distance of nearest cargo is within 25 miles it still shows no discernible impact on the income of non-incorprorated business owners. (Column 4-6 of Table 3.6). The effect remains null for the impact on business owners since it doesn't have any predictive power in terms of significance even when this distance to airport goes down from 100 to 25 miles (Column 1-3 of Table 3.6). Column 1-3 is the result from IV A and column 4-6 show the result for IV B in Table 3.6).

1.8 Concluding Remarks

Currently, more than 70% of American households subscribe to Prime and most Prime members head to Amazon when they want to buy something online. (Bain (2021)) The success of Amazon demonstrates the importance of raising the bar across industries while being customer focused. They started Prime Air many years ago, and now, as a result of the strain on the supply chain, they are looking into alternative delivery methods. They have taken a strategy that is diametrically opposed to that taken by Apple and NordicTrack; that is, they have implemented vertical integration and offered practically unbounded degrees of customization to their clientele. Amazon has generated revenue of $4.1 billion in the United States within the last ten years. Over the course of the past decade, Amazon has put an annual average of $200 billion dollars into the development of brand-new infrastructure. The "trickling down" effect of $4.1 billion for $200 billion in spending would be wonderful if it weren't for the fact that the $4.1 billion in question was placement money and not money to stimulate spending.

The question that needs to be answered is whether the benefits that are summarized outweigh the harm that Amazon causes to its direct competitors as well as those independent sellers on their platform beside the local business owners. The opening of an Amazon fulfillment is responsible for the demise of several local and non-local businesses, but at the same time, it opens the door for new businesses to enter the market evidenced by my quantitative experiment. The "net mortality" or "net entry" of various types of businesses in the same or adjacent locations is, therefore, an important factor to consider, and my paper shows people who were not full-time business owners before as soon as amazon strengthened its supply chain, has started to transition themselves as a full-time local business owner.

In light of the findings presented in the preceding sections, it is evident that the influence of Amazon's fulfillment centers on the incomes and labor outcomes of non-incorporated business owners is not discernible. The utilization of instrumental variable (IV) regression, which accounts for endogeneity, has not revealed any significant impact on these economic indicators, aligning with the results obtained through ordinary least squares (OLS) analysis. Turning my attention to Table 3, which outlines the two sets of instrumental variables employed in this study, it becomes apparent that neither the distance from the nearest cargo airport nor the distance from the nearest intersection of an interstate highway, both interacting with year dummies, exhibit any statistically significant impact on the incomes of non-incorporated business owners. These results further reinforce the notion that the establishment of Amazon's fulfillment centers has not translated into noticeable financial improvements for this particular group.

Despite producing a bounty of third-party sellers on their platform while reviving the hope in local areas to be an entrepreneur, Amazon still has a huge hill to climb in terms of increasing the financial well-being of these non-incorporated business owners. Despite the notable rise in third-party sellers facilitated by Amazon's platform and the revival of entrepreneurial hopes in local communities, my analysis indicates that the financial circumstances of non-incorporated business owners remain largely unaffected. These findings shed light on the complex dynamics at play within the Amazonian regime and highlight the need for further exploration into the intricate relationships between e-commerce giants, local economies, and business outcomes. The notorious culture of Walmart started to coerce the suppliers; the same aggressive trait has been shown by Amazon realizing the profit margin is eventually finite as time passes by. The sunny optimism that could be seen from the transition to being a local business owner, will slowly morph into dark pessimism if this creeping aggressiveness goes on to batter and bruise the independent sellers by pushing the price lower. The innovative flywheel of the marketplace that gave

Amazon the initial momentum to fight against the tech giants like Google and eBay, has become a platform to be the foundation of everyday lower prices at the expense of those independent sellers' financial hardship.

Chapter 2

E-commerce Expansion in US Labor Market: Blessing in Disguise?

2.1 Introduction

The history of retail in the United States is one of constant innovation and transformation. Each generation has seen a new company emerge to take up the mantle from the previous giant and reshape the economy in its image. From the pioneering A&P grocery chain of the 1920s, which introduced the concept of the supermarket, to the rise of Walmart in the latter half of the 20th century, technological innovation and the relentless drive to lower prices have been the hallmarks of success. Today, Amazon reigns supreme as the dominant force in e-commerce, leveraging the power of the internet to become the first choice for online shoppers across the country (Figure 3.1). Yet, as history has shown us, no company is invincible, and the question remains: will there be a new technology or a new player to emerge and topple Amazon from its perch, just as Walmart did to A&P? As the torch of retail dominance has been passed from one giant to another throughout history, the emergence of Amazon as the reigning champion of e-commerce begs the question of its invincibility. However, as the e-commerce giant continues to expand its physical footprint, it raises concerns about the impact on local labor markets. In this context, this paper seeks to investigate how the growth of e-commerce infrastructure affects the labor force, and whether it can compensate for negative spillover effects in other industries. This paper aims to explore the impact of e-commerce on local labor markets, specifically examining how the physical expansion of an e-commerce setup affects the labor force and compensates for negative spillover effects in other industries.

In 2021, the top three e-commerce companies in the United States accounted for 50% of the sales, and Amazon emerged as the undisputed leader, grabbing a staggering 40% of the market. Despite this, the share of e-commerce in total retail sales has not yet crossed the 10% threshold [1], highlighting the potential for further growth in this sector. According to one estimate, the last-mile delivery market [2] is expected to grow from $18.4 billion in 2018 to $55.2 billion in 2025 (Singh and Murteza (2022)). While e-commerce has led to job losses in traditional brick-and-mortar retail, it has also created new job opportunities in the transportation and warehousing sector. For example, a study by the Bureau of Labor Statistics (2021) found that the retail sector lost 778,000 jobs between March 2020 and March 2021 due to the COVID-19 pandemic and the shift towards e-commerce (U.S. Bureau of Labor Statistics). Furthermore, the establishment of Amazon fulfillment centers in the same county has been shown to lead to a reduction in retail employment and wages Chava et al. (2018). However, there are concerns that many of the jobs created by e-commerce platforms like Amazon are low-paying and offer limited opportunities for advancement. For example, a study by the Brookings Institution found that the median hourly wage for Amazon warehouse workers was $12.75 in 2018, compared to $19.46 for all occupations in the US (Bhattarai (2018)). Overall, the impact of e-commerce and Amazon's expansion on the local labor market in the US is a complex issue that requires further research and analysis. Therefore, policymakers need to address these issues to ensure that the benefits of e-commerce and Amazon's expansion are shared more evenly across different regions and sectors of the economy.

In this study, I aim to address a series of issues related to the impact of e-commerce on labor markets. Specifically, I identify three key problems that I will investigate in detail. The first issue pertains to the challenge of accurately quantifying the consequences of e-commerce on the labor markets of individual communities. This challenge arises since e-commerce sales increase at a pace that is dictated by local economic conditions, making it difficult to differentiate between treatment and control areas based on their exposure to e-commerce. To overcome this challenge, I develop a novel instrumental variable (IV) technique that capitalizes on the radial pattern of Amazon's expansion. This allows us to account for the endogeneity of Amazon's entry, which is a potential source of contamination for estimates of the effect Amazon has on earnings. I believe that my evidence represents a significant improvement on previous research in this area, particularly due to the implementation of an identification strategy that considers the endogeneity of fulfillment center location and timing of entry, as well as their potential correlation with future changes in earnings or employment of relevant industries.

Chava et al. (2018) used a staggered roll-out of fulfillment centers (FCs) as a measure

1. https://www.bls.gov/opub/btn/volume-11/retail-trade-employment-before-during-and-after-the-pandemic.htm

2. The last-mile delivery market refers to the final stage of the delivery process, where goods are transported from a distribution center or retailer to the end consumer's location. It is often considered the most crucial and challenging part of the logistics chain, as it involves the movement of goods over relatively short distances, typically within urban or suburban areas.

for the presence of local e-commerce, but this approach may introduce an identification bias due to the retailer's location bias in choosing FC locations. To address this issue, they employed a fixed effect methodology. However, it is worth noting that the e-commerce retailer might strategically locate its FCs in areas with declining competition from brick-and-mortar stores. To mitigate this potential bias, the study considered the location choice and timing of the retailer's expansion To address this potential bias, they consider the location choice and timing of the e-commerce retailer's expansion. Building upon previous research by Basker (2005), Basker (2005), and Arcidiacono et al. (2020), my paper aims to establish a more robust causal relationship between income, employment, and the physical expansion of an e-commerce chain.

Expanding on the insights provided by Derenoncourt et al. (2021), it is important to consider that Amazon's top 25 advertised occupations on Burning Glass Technologies are primarily concentrated in the retail, transportation and logistics, and production industries (using job posting data from 2014-2019). While this observation holds merit, it is crucial to contextualize this information within a broader understanding of the labor market and the various factors that influence job opportunities.

In light of this concern, I utilize a database of Amazon fulfillment centers to examine the impact of their presence on local income and employment at the county level for these three specific industries. Importantly, I develop an instrumental variable (IV) technique leveraging Amazon's expansion to account for the potential endogeneity of their entry, which can contaminate estimates of the effect on earnings. I believe that my approach significantly enhances the existing body of research on these and related questions, particularly through the implementation of this improved methodology. The most important aspect of this improvement is the implementation of an identification strategy that considers the endogeneity of fulfillment center location and timing of entry, as well as how these factors may be correlated with future changes in earnings or employment. It has been suggested that amazon's strategy was to locate in near the bigger cities from where the tax and labor cost both are higher than average since 2011 (Houde et al. (2017)) , and it is reasonable to expect that it will be situated near the cargo airports since FedEx could achieve the 49% of express delivery (2012) in the domestic USA only by having dominance in air shipping. Furthermore, both FedEx and UPS offer next-day delivery with their air shipping option. This set of information bolsters the concept that if Amazon wants to achieve its goal of same-day shipping, setting up its distribution network near cargo airports is the only way to go further.

To account for the endogeneity of Amazon's entry, we utilize two sets of instrumental variables: IV A and IV B. IV A represents the interaction of distance from cargo airport with year dummy. The determination of the fulfillment center's location within a county is primarily influenced by the minimum distance from the closest cargo airport, while the temporal aspect of its establishment is typically determined by the utilization of year dummies. IV B corresponds to the distance from the intersection of the interstate highway, as well as the incorporation of year dummies. These instruments are carefully constructed based on the relevant economic literature, taking into account the unique characteristics of Amazon's expansion strategy and the local labor market dynamics. By employing these instruments in our 2SLS models, I can obtain unbiased estimates of the causal effects of Amazon fulfillment centers on earnings, employment, and the number of establishments.By employing a meticulous instrumental variable approach and considering the specific instruments I construct based on existing literature, my study aims to provide a detailed and nuanced understanding of how Amazon's presence affects the local labor markets in the retail, transportation, and production sectors.

Overall, the literature on Amazon's impact on the labor market is mixed, with some studies highlighting the potential benefits of the company's growth for local employment rates, while others emphasize the potential negative effects on wages, working conditions, and displacement of jobs in traditional retail. Thus, the contribution of this paper to the literature is to find a definitive answer on how a physical expansion of an ever-growing amazon distribution network directly affects local labor markets at the county level. My results indicate that as amazon sets up a fulfillment center in a county it shows the downward pressure on wages in retail and transportation industry. The magnitude of OLS coefficient is much less compared to the two sets of IV I am using for my identification method. Therefore, it shows the magnitude of omitted variable bias in OLS estimates when I am not dealing with the endogeneity issue of the location of the amazon distribution setup decision.

The remaining parts of the paper are organized as follows. The literature is discussed in the second section. In the third section, I offer some insights into why amazon has been investing in acquiring lands near the cargo airports, as well as how this will impact the foreseeable logistics industry if amazon successfully wants to keep the promise of same-day delivery. Section four outlines our data sources and shows our identification approach. In the fifth and sixth sections, I present not only descriptive statistics but also our most important empirical findings, as well as the outcomes of several different robustness and specification tests. In the final and concluding section, number six, I will go over the consequences of our findings.

2.2 Background of Amazon Fulfillment Center (FC)

The significance of Amazon's delivery capabilities has been evident in recent years, with the company investing heavily in its logistics network and expanding its delivery fleet. With Amazon's vast resources and growing delivery capabilities, the company may be able to offer more competitive pricing and faster delivery times, which could place pressure on existing shipping companies to adapt to changing market dynamics. Since the inception of Amazon's fulfillment center in

2000, the company has been committed to improving its logistics and distribution network to better serve its customers. Initially, the company relied heavily on third-party logistics providers to manage its supply chain, but over the years, Amazon has increasingly built out its own fulfillment infrastructure. The company's first fulfillment center was opened in Seattle, and by the end of 2001, Amazon had 8 fulfillment centers in operation. At the time, Amazon was primarily a bookseller, and the fulfillment center was designed to efficiently store and ship books to customers. However, as Amazon expanded its product offerings and customer base, it became apparent that a more robust fulfillment network was needed. In 2005, Amazon introduced its Prime membership program, which promised customers free two-day shipping on eligible items. This program was a major driver of growth for Amazon and helped to establish the company as a leader in e-commerce. To support the growing demand for fast and reliable shipping, Amazon continued to expand its fulfillment center network, with a particular focus on building facilities near major population centers.

Moreover, Amazon's focus on smaller, last-mile delivery facilities that sort packages for final delivery to customer homes provides the company with a competitive advantage over rivals. These smaller facilities enable Amazon to deliver products to customers more quickly and efficiently, which has helped the company become a leader in e-commerce (Figure 3.6). By contrast, Walmart and Target primarily use their physical stores to fulfill online orders, which can be a less efficient and more costly approach. Finally, Amazon's recent push into grocery delivery, with the acquisition of Whole Foods, further strengthens its position in the retail industry. With the addition of Whole Foods' physical stores and distribution network, Amazon can offer customers a more diverse and extensive range of products and services, both online and in-store. This acquisition also provides Amazon with a platform to introduce its own private-label grocery products, potentially further expanding its customer base and revenue streams.

The Chief Executive Officer of Amazon's Worldwide Consumer Business, Dave Clark, made an announcement of great significance to the shipping industry, stating that the company is on track to surpass its long-standing competitors, UPS and FedEx, and become the largest package delivery service in the United States by the beginning of 2022 (Palmer (2022)). This announcement carries important implications for both the shipping industry and consumers. Amazon has several potential strategies that it could consider enhancing its fulfillment center operations and gain an edge in the competition against UPS, FedEx, and Walmart. Firstly, Amazon could continue to invest in its own delivery infrastructure, such as by building out its network of fulfillment centers and last-mile delivery facilities. This would reduce the company's reliance on third-party delivery providers and allow it to offer faster and more reliable delivery to customers. Secondly, Amazon could place greater emphasis on expanding its network of smaller last-mile delivery facilities, which are located closer to customers. This approach can enable faster and more efficient delivery, which could help differentiate Amazon from competitors such as Walmart. Thirdly, Amazon could integrate drone delivery and other innovative technologies into its operations to further enhance speed and convenience for customers. The company has already made significant investments in robotics and artificial intelligence, which have helped improve the efficiency of its fulfillment centers. It is taking measures toward being able to serve Prime members in an even more timely manner by installing miniature fulfillment facilities in local markets (Figure 3.5).

According to Dave Clark, Amazon's Chief Executive Officer of Worldwide Consumer Business, UPS Next Day Air and FedEx Express played important roles in the company's fulfillment network by transporting less frequently purchased items to Prime customers when they were not stocked nearby Amazon's fulfillment centers (Stone (2022)). However, despite relying on UPS and FedEx for some of its delivery needs, Amazon has been steadily building up its own in-house delivery network in recent years. Dave Clark also stated that Amazon is poised to become the largest package delivery service in the US by early 2022. This is a significant milestone and would place Amazon ahead of UPS and FedEx, long-standing competitors in the shipping industry. This growth is likely due in part to Amazon's focus on building out its delivery network and fulfillment infrastructure, as well as its investments in technologies like robotics and artificial intelligence to improve the efficiency of its operations. Meanwhile, Walmart has announced that it plans to offer same-day drone delivery in five states starting in June 2022 (Clark (2022)). This move could potentially help Walmart keep up with Amazon in terms of delivery speed and convenience, particularly since Walmart already has a significant physical presence across the US, with 90% of the population living within 10 miles of a Walmart store (Stallbaumer (2022)). The strategic location of fulfillment centers near infrastructural amenities allows Amazon to respond to changing market conditions with greater flexibility. Based on these plans done by Dave Clark, it's safe to assume that the newly planned fulfillment centers are placed near infrastructural hub of a county where airport and interstate highway facilities are available to ensure fast and efficient order fulfillment, and simultaneously have one leg up against other supply chain delivery providers.

In conclusion, Amazon's extensive and growing distribution network, which includes a wide range of fulfillment centers and last-mile delivery facilities, is likely to continue to be a major factor in the company's success. By leveraging its logistical capabilities, Amazon can offer faster, more efficient delivery to customers, which has helped the company gain market share in the highly competitive retail landscape. While competitors such as Walmart and Target also have distribution centers, given in depth of the problem, my analysis says placing fulfillment centers near cargo airport and interstate highway makes the most sense. Therefore, my instrumental variable is based on these two factors.

2.3 Literature Review

Amazon's distribution network has been a major factor in its success in the retail industry, but it has also put pressure on the wages of its fulfillment center workers and retail workers. According to a report by the Economic Policy Institute, Amazon's expansion into new markets and its aggressive pursuit of market share have resulted in downward pressure on wages for both Amazon workers and workers in other companies in the retail sector. The report also notes that Amazon's use of automation in its warehouses has resulted in the displacement of workers and the erosion of job quality in the industry.

However, the impact of Amazon's distribution network on wages in the retail industry is a topic of ongoing debate among economists. While some argue that the company's use of automation and technology has led to a decline in traditional retail jobs and put pressure on wages, others point out that Amazon's growth has also created new job opportunities in areas such as transportation, logistics, and technology. Nevertheless, the extent to which Amazon's growth has affected wages in the retail industry remains a topic of ongoing research and analysis. In a 2019 study by the Economic Policy Institute, Jones and Zipperer (2018) found that Amazon's dominance in the online retail space has contributed to a decline in wages for non-managerial retail workers. The study found that for every new Amazon fulfillment center that opens in a county, there is a 30% reduction in the average wage of workers in the warehousing and storage industry in that county.

A more recent report by the Brookings Institution in 2021 found that Amazon's distribution network has enabled the company to capture a significant share of the retail market, particularly during the pandemic. The report notes that Amazon's sales increased by 38% in 2020, while traditional brick-and-mortar retailers experienced a decline in sales (Borchert et al. (2021)). The report also highlights the role of Amazon's distribution network, including its vast network of warehouses and delivery infrastructure, in allowing the company to quickly adapt to changes in consumer behavior and meet increased demand for online shopping. Jones and Zipperer (2018) notes that the impact on wages is not limited to the retail sector, as Amazon's dominance also affects transportation and production industries, which supply and move the goods sold through the company's platform. The study concludes that while Amazon's expansion has created new jobs, it has also contributed to wage stagnation and a decline in job quality for workers.

Empirical research has shown that the growth of e-commerce has led to a decline in retail employment in local areas. Chun et al. (2020) finds that a 1% increase in online spending leads to a reduction of 1.17% in local retail employment but did not account for unobserved economic conditions that may have affected both online penetration rates and employment. Similarly, Chava et al. (2018) shows that the establishment of an Amazon fulfillment center in a county leads to a reduction in local retail employment and wages. These findings suggest that Amazon's expansion may have a negative impact on the local labor market in the US. However, this approach from Chava et al. (2018) did not consider the potential spillover effects of the introduction of a fulfillment center in a neighboring county.

A study by Mitchell et al. (2020) published by Institute for Local Self-Reliance (ILSR) found that Amazon's expansion between 2007 and 2015 led to the loss of over 149,000 jobs in brick-and-mortar retail, with most of those job losses occurring in the years following the Great Recession. The authors argue that the displacement effect was strongest in counties with a high concentration of retail jobs and that the impact on employment was particularly severe for workers in low-wage, low-skilled jobs. However, other researchers have questioned the magnitude of the displacement effect, with some arguing that the jobs lost in traditional retail have been replaced by jobs in e-commerce and related industries.

In conclusion, the impact of distribution networks on the wage pressures faced by retail and fulfillment center workers is a complex and multifaceted issue. While the growth of e-commerce and the increasing use of technology and automation can improve efficiency and productivity, they also raise concerns about job quality and the displacement of human workers. To ensure that the benefits of technological progress are shared equitably, policymakers and stakeholders must address the challenges posed by distribution networks and support the development of a more inclusive and sustainable job market. Existing research suggests that Amazon's impact is complex, with both positive and negative effects on employment and wages in local labor markets. Further research is needed to understand the mechanisms underlying these effects and to inform policy decisions aimed at mitigating the potential negative impacts of Amazon's dominance in the e-commerce sector.

Using the first date of planning for a Walmart discount store as a proxy for the entry of Walmart into counties in the United States has allowed researchers to circumvent the issue of temporal endogeneity (Basker (2005), Basker (2007)). It has been argued that the opening date of a store can be manipulated to coincide with favorable conditions, whereas the date on which the store initially plans to open cannot be manipulated in this way. This is because the opening date of a store can be manipulated to coincide with favorable conditions. Using a measure that acts as a proxy for the date that a store's initial planning began, she argues that while the precise timing of a store's opening can be manipulated to coincide with favorable conditions, planning is done far enough in advance that it is not likely to be endogenous to a growth spurt — or sudden decline — exactly coinciding with Walmart's entry. This is because planning is done in sufficient advance. One of the most significant limitations of this IV strategy is the fact that the "instrument" (store planning date) is only defined for locations in which a Walmart store was eventually opened. This is one of the most significant limitations. Because of this, it is challenging to arrive at any definitive conclusions regarding the effect that the presence of a Wal-Mart store has on communities that do not have one. Both Dube et al. (2007) and Neumark et al. (2008) exploit the fact that counties enter Walmart's sphere at different times, depending on their distance from Benton County. These researchers did their research independently and published their findings in 2005.

Because it captures an "intent to treat" and therefore, in principle, allows estimating what Angrist and Imbens (1995) call a "local average treatment effect," the distance instrument is intuitively appealing because it is clearly exogenous to Walmart's entry. Additionally, it has the potential to be much more powerful than the planning date that was used by Basker (2005), Basker (2007). I am putting together an interaction variable that will tell me about the minimum distance from each county to the nearest cargo airport with year dummy which will tell me how with time Amazon kept changing it's fulfillment center location strategy. The minimum distance from the nearest cargo airport will determine the location, and year dummy, in most cases, will determine when the fulfillment center will enter the county, just as Neumark et al. (2008) and Dube et al. (2007) did for Walmart's entry. Moreover, I am using another set of instrumental variable which captures the minimum distance from each county to the nearest intersection of interstate highway. Previous later suggests that placing fulfillment center near interstate highway is a very common scenario amongst the supply chain related companies (Dablanc and Rakotonarivo (2010); Kang (2017)). Therefore, I am interacting the distance from intersection of interstate highway with year dummy as my second set of IV.

In addition, there are additional biases that are present, such as the problem that was mentioned earlier, which is that it is impossible to accurately identify the source of any changes that occur at the county level in terms of employment or wages. This is because there are a variety of factors that could be responsible for a change in employment or wage levels within a county that is unrelated to the establishment of an Amazon fulfillment center. The paper makes use of two-way fixed effects regressions, the details of which will be provided later in the paper, in order to take into account these issues.

2.4 Data and Descriptive Statistics

The dataset used in this study is the County Business Patterns (CBP) data provided by the U.S. Census Bureau, covering the period of 2000-2021 for all counties in the United States. The CBP data includes information on the number of establishments, employment, first-quarter payroll, and annual payroll for each 6-digit industry. The data has been harmonized with the North American Industry Classification System (NAICS) 2017 codes and the county codes have been made consistent throughout the panel.

The analysis examines changes in employment and wages within these industries over time. Employment is measured as the number of workers employed during the week of March in each year, while wages are measured as the annual payroll of each industry. To preserve confidentiality, the Census Bureau suppresses the number of employees for the majority of county-industry cells. However, this issue is addressed by imputing the suppressed values using the linear programming method developed by Eckert et al. (2020a). The resulting data set allows for an accurate analysis of changes in employment and wages over time.

The CBP data contains information on establishments that are classified based on the number of employees they have. The classification system used groups firms into twelve categories, ranging from those with 1 to 4 employees to those with more than 5,000 employees. This data provides a comprehensive view of the size distribution of firms within the covered labor market, which is crucial for understanding the distribution of labor market power across counties in the US. The summary statistics of the treated counties and control counties are discussed in detail in Table 3.8. This information is then merged with the county-level database of Amazon fulfillment center openings.

2.4.1 Definition of Specific industries

The covered labor market in this study refers specifically to the retail, transportation, and production industries, which are of particular interest due to their large share of employment in the US economy. This study aims to analyze the impact of labor market concentration on wages and employment in these industries at the county level. The retail industry is defined by the North American Industry Classification System (NAICS) code 44-45, which includes establishments engaged in the retail sale of merchandise. The transportation industry is defined by NAICS code 48-49, which includes establishments engaged in the transportation of goods and passengers. The production industry includes all manufacturing establishments, as defined by NAICS codes 31-33.

2.4.2 Cargo airports

The FAA annually asks airports to report their annual cargo landings. Please note that all-cargo reporting is limited to aircraft operations dedicated to the exclusive transportation of cargo. Here, to track the air cargo data I use one annual database sourced from the Federal Aviation Administration (Federal Aviation Administration) Since years 2000-2021 I have collected the data for all the cargo airports that carry the highest tonnage of cargo goods throughout the whole time. This eventually summed up to a total of 100 cargo airports situated in different counties across the united states.

2.4.3 Distance between counties

The Haversine formula is used to determine the great-circle distance between any two places inside a given county. These distances are referred to as county distances. This dataset of counties is uploaded on NBER public website which also provides the distance between counties which I eventually used to find the minimum distance from counties with cargo airports to other counties. (NBER). If a county has at least one cargo airport in it, then the distance variable for that county is zero for my estimation.

2.4.4 Distance from intersection of interstate

This study aims to measure the distance from each county in the USA to the nearest intersection of interstate highways using geospatial analysis techniques. To achieve this goal, two essential shapefiles are required: a shapefile of USA counties (Census Bureau (2023)) and a shapefile of interstate highways (Federal Highway Administration (2023)) that intersect or pass through the counties of interest. The data was obtained from the US Census Bureau. The shapefiles should be loaded into a Python geospatial library, such as GeoPandas, to facilitate the analysis. To ensure the accuracy of the analysis, it is important to ensure that both shapefiles are in the same coordinate reference system (CRS). If not, one of the shapefiles must be reprojected using the to_crs method in GeoPandas to reproject the shapefile to a different CRS. This step is crucial to guarantee that the measurements between the two shapefiles are consistent.The sjoin_nearest method in GeoPandas was used to join the county and interstate shapefiles based on the nearest intersection. This method will create a new GeoDataFrame that includes the county name and the distance to the nearest interstate intersection for each county.

2.4.5 Fulfillment centers

Fulfillment centers came from MWPVL International, a company that specializes in the supply chain, logistics, and distribution. (MWVPL). This unique source provides information on all of Amazon's fulfillment centers, including their locations, opening years, types, and sizes in square feet. In addition, the data that was collected by MWPVL International identifies the type of fulfillment centers that are used by Amazon. For example, the data shows whether the facilities are normal large sortation centers or Prime Now Hubs. With the use of these statistics on the various types of fulfillment centers, it is possible to determine whether there is a major difference between the various types of fulfillment centers in terms of the impact they have on the employment and wage situation in the surrounding area. The data collection also includes information regarding the amount of smaller fulfillment facilities that have either been shut down entirely or merged with other, larger centers.

2.4.6 Seer County Demographics Data

Data on population by age group is from the Surveillance, Epidemiology, and End Results Program of National Cancer Institute (SEER). The original data is available at the county-year level. We define the working age population as the population with age between 18 and 64 (National Cancer Institute)

2.5 Identification Method

2.5.1 Explanation of using IV

Jakubicek and Woudsma (2011) conducted a study that found that proximity to transport infrastructure, particularly highways, was highly valued by logistics managers when selecting a warehouse location. Similarly, Dablanc and Rakotonarivo (2010) argued that companies within the logistics industry tend to locate their facilities as close to highway networks as possible, but also in proximity to airports. Proximity to transport infrastructure, therefore, remains a key consideration when selecting a warehouse location. Kang (2017) notes that for companies involved in international trade, it is more important to have warehouses in proximity to seaports or airports compared to distribution companies serving the local market. This is because international trade requires efficient connectivity between various modes of transport and warehouses located near seaports or airports can ensure a seamless transition between transportation modes.

By leveraging existing transportation infrastructure, Amazon can reduce its shipping costs by placing fulfillment centers near transportation hubs. This can allow Amazon to offer lower prices to customers and gain a larger market share against competitors. Locating fulfillment centers near transportation hubs can also enable Amazon to manage its inventory more efficiently. This can facilitate faster and more reliable order fulfillment, thereby providing Amazon with a competitive advantage over UPS and FedEx. Planning wise it's possible when you place your warehouse near interstate highways and airports which is also solidified with existing literature in supply chain logistics mentioned earlier.

Firstly, the location of highways and airports is often influenced by physical geographies, such as mountain ranges or bodies of water, rather than economic or demographic factors. Consequently, the distance from these facilities may be unrelated to any economic outcomes. Secondly, historical factors play a role in the location of highways and airports. For instance, the placement of these facilities has been determined by the location of previous transportation routes or military bases. As a result, the distance from highways and airports may be random and unrelated to any economic outcomes. Thirdly, zoning laws and regulations often restrict where highways and airports can be located. Thus, the placement of these facilities may be random and not driven by economic factors. Fourthly, the economic factors influencing the location of highways and airports can change over time, leading to a random distribution of distances from these facilities. For instance, as cities grow and expand, new highways and airports may be constructed in locations that were previously considered rural or undeveloped. Finally, land use changes, such as the development of new residential or commercial areas, can also impact the distance from highways and airports. As a result, the distance from these facilities may be random and unrelated to any economic outcomes.

Several studies have provided evidence to support the argument that the location of highways and airports can be influenced by factors unrelated to economic outcomes. A study by Levinson and Krizek (2017) found that the location of highways in the United States was influenced by historical factors, such as the location of previous transportation routes, rather than economic or demographic factors. Furthermore, Kulkarni and Pande (2019) found that the placement of highways in India was often determined by topography and land availability rather than economic or demographic factors. These studies suggest that the location of highways and airports can be impacted by various factors, which are unrelated to economic outcomes. This finding indicates that distance from these facilities could be a useful instrumental variable for the placement of a warehouse, allowing me to control for unobserved variables that may impact the placement decision. Because of this overall random nature, the instrument (interaction of distance from cargo airport . and the distance from intersection of interstate highways with year dummy) should not have any effect on the economic climate of the surrounding area, apart from the effect that it has on the locations where Amazon is starting its fulfillment center.

2.5.2 Basic model setup and Estimation

Our instruments are the dummy year variables times the minimum distance to the nearest cargo airport/ the minimum distance to the nearest intersection of interstate, so it is designed the capture how Amazon has been trying to redesign its fulfillment center to take advantage of its geographical position close to the opportunity zone where they not only have the access to a significant portion of prime subscribers but also it helps to reduce the shipping cost to outdo other counterparts like Walmart and target. Denoting the instrument by $\theta_d(dist_d * YR_y)$, we estimate the following system of equations by two-stage least squares:

$$Amazon_{ct} = \alpha + \sum_{d=1}^{D} \beta_d dist_d + \sum_{y=1}^{Y} \beta_y YR_y + \sum_{j=1}^{J} \beta_3 X_{jict} + \sum_{d=1}^{D} \theta_d (dist_d \times YR_y) + \gamma_c + \epsilon_{ict} \quad (2.1)$$

$$lnY_{ict} = \delta + \psi_q \widehat{Amazon}_{ct} + \sum_{y=1}^{Y} \psi_y YR_y + \sum_{d=1}^{D} \phi_d dist_d + \psi_r X_{jict} + \gamma_c + v_{ict} \quad (2.2)$$

where $\widehat{Amazon}_{ct}$ are the fitted values from the first stage regression 3.1. I put the IV strategy into action by segmenting the United States into 17 distance rings, each of which represents a 100-mile increase in distance from the closest cargo airport (for example, fewer than 100 miles, 100–200 miles,..., 1,600 miles or more), and developing an indicator variable for each of these distance rings.[3] The distance ring dummies are included as controls ($\beta_d dist_d$), while the interactions of the distance ring dummies with year dummies are used as instruments. I cluster standard errors by county. X is the vector of individual-level controls including age, age squared, marital status, sex, race, Hispanic, and education. The county fixed effect is shown by γ_c which helps to account for time-invariant differences across counties, and ϵ_{ict} is the error term.

This regression 3.2 differs from the regression 3.1 by replacing the (year dummy $\times$ distance to the nearest cargo airport) with the predicted values of $Amazon_{ct}$ estimated in equation 3.2 and adding the distance ring fixed effects.

Identification of ψ_q in equation 3.2 in the IV model comes from the assumption that the year dummy $\times$ distance to the nearest cargo airport interactions can be excluded from the second-state regression 3.2—that is, that these interactions are uncorrelated with changes over time in the outcome of - income and employment of worker-on the controls. By including the distance ring fixed effects in an equation 3.2, we allow for the distances to be correlated with levels of income and employment; we only need to assume that they are uncorrelated with trends.

2.5.3 Assessment and Interpretation of Baseline Results

I then proceed to carry out a number of exercises with the goal of determining which groups are subjected to the baseline treatment effects that were investigated in Section 6 how the instrument is functioning, and to what extent the geographic scale and outcome measure is contributing to the results. This is done by conducting a number of exercises with the purpose of determining which groups are subjected to the baseline treatment effects that will be investigated in Section 5.

2.5.3.1 Assessing Relevance and Validity of the Instrument

To further interpret these results of the instrumental variables that are shown in Table 3.9, Table 3.10 and Table 3.11, it is necessary to determine the extent to which the instrument—the interaction of minimum distance from cargo airport and the year dummies—satisfies the necessary assumptions for unbiased estimation. This can be done by determining the degree to which the instrument satisfies the necessary assumptions for unbiased estimation. Examining the coefficients in the "First Stage" column of Table 3.9 enables us to ascertain whether or not the instrument is important by providing us with this information. These coefficients demonstrate that the instrument, at each distance threshold, has a significant and accurately measured association with each outcome in the same region. This was determined by comparing the instrument's measurements to the outcomes. If the county is within the distance ring of 100 miles of a cargo airport as year progresses

3. The 100-mile distance ring classification follows Neumark et al. (2008) and Dube et al. (2007)

for setting up an Amazon distribution network then it is associated with a 0.3781 standard deviation (SD) increase in the probability of having an Amazon fulfillment center. The significance of the instrument in the field can be demonstrated by its highly effective first stage, which operates across multiple bandwidths. Given that standard errors are clustered, the Kleibergen-Paap F-statistic would be an appropriate indicator of relevance in this scenario. This statistic is equivalent to the F-statistic of the first stage regressions in my scenario because there is only a single endogenous variable (Amazon fulfillment center) (Kleibergen and Paap (2006)). The fact that the F-statistic is higher than the typical cutoff of 10 lends credence to the notion that the constructed instrument is relevant as a proxy for the Amazon fulfillment center.

One other method for determining whether or not my set of IVs is valid is to carry out an over-identification test using multiple instruments, focusing on the endogenous variable of whether or not there is an Amazon fulfillment facility in the county. My multiple IV does the work of conducting the Hansen J statistics so that this can be accomplished. None of the Hansen J statistics p-value is less than 0.05 in the results presented in Table 3.9, Table 3.10 and Table 3.11. This demonstrates that we are unable to reject the null hypothesis that our instruments are producing the same results, which provides me with extra confidence that the primary instruments are credible.

The exclusion restriction is another requirement for a valid set of instrumental variables. That is, it should affect dependent variables in the second stage only through the endogenous variable. Within this framework, the created IV should only affect the outcomes at the county level through the establishment of an Amazon distribution network. In the previous part of this article, I went over the reasons why the built IV is less likely to have problems with endogeneity. However, even the constructed IV could be subject to concerns regarding the exclusion restriction if there are enough of them. In order to address such issues, I carry out a series of exercises, which are described in more detail below.

2.5.4 Impact on income of retail, transportation and production industry workers

The next specification will explore the same question, however using an instrumental variable regression. In a two-stage least squares regression, this subset of sample counties that has the Amazon fulfillment centers in the sample is so-called compilers. Compilers are precisely the labor force of those counties that are induced to change their work profile due to having access to the e-commerce physical setup near their county.

I have two primary goals: the first is to show that inaccurate modeling leads to imprecise results, and the second is to investigate both the direction and size of the effect that the Amazon fulfillment center and its supply chain network has on the outcomes of entrepreneurial endeavors. The most important findings are outlined in Table 3.9. The simple OLS estimation that does not factor in any controls reveals a negative and insignificant relationship between income earned by non-incorporated business owners and the presence of an Amazon fulfillment center (Column 1).

2.5.4.1 Retail Industry

Table 3.9 presents the estimated impact of Amazon fulfillment centers (AFCs) on various outcomes in the retail industry, using different regression specifications. We analyze three dependent variables: logarithmic earnings (Earn), logarithmic employment (Emp), and logarithmic number of establishments (Est). For each outcome, we estimate three models: an ordinary least squares (OLS) model without controls, and two instrumental variable (IV) models using two different instruments (IV A and IV B) to address potential endogeneity issues arising from the location of Amazon FCs.

The first IV method (IV A) utilizes a binary instrument that uses the interaction from the minimum distance of each county to the nearest cargo airport interacting wuth year dummies. The second IV method (IV B) uses the interaction of minimum distance of each county to the nearest intersection of interstate highway with year dummies. These instruments were selected based on their relevance to the location decision of the Amazon fulfillment centers, as well as their exogeneity to the dependent variable.

The OLS estimates reveal a negative and statistically significant coefficient for the presence of Amazon FCs on logarithmic earnings (Table 3.9 Column 1) of retail industry workers. However, these results may be biased downwards due to potential measurement error and endogeneity issues. To address these concerns, we estimate two IV models using IV A and IV B as instruments for Amazon FCs. The results show a much larger and statistically significant negative effect of Amazon FCs on logarithmic earnings, and significant effect on logarithmic employment or establishments, when compared to the OLS estimates. In this table IVA stands for my first set of instrumental variables which is the interaction of year dummies with the distance ring dummy for the nearest cargo airport. The second IV is represented by IV B (Column 3) which is the interaction of year dummies with the distance ring dummy for the nearest intersection of interstate highway.

The OLS coefficient estimates of -0.157 suggests that a one-unit increase in the presence of Amazon FC is associated with a 15.7% decrease in earnings, holding all other variables constant. However, this OLS estimate may suffer from a downward bias due to the presence of measurement error in the independent variable (i.e., the presence of Amazon FC). To address this endogeneity issue, two instrumental variables (IV) were used in the analysis, IV A and IV B. The IV A coefficient estimate of -1.116 indicates that a one-unit increase in the presence of Amazon FC, as instrumented by IV A, is associated with a 67.2% decrease in earnings, holding all other variables constant. Similarly, the IV B coefficient estimate of -1.080 suggests that a one-unit increase in the presence of Amazon FC, as instrumented by IV B, is associated with a 66% decrease in earnings, holding all other variables constant.

The substantial decline in the percentage of wage can be attributed to the amalgamation of various contributing factors. The emergence and growth of e-commerce fulfillment centers have intensified competition within the retail industry. These centers, equipped with advanced technologies and streamlined supply chains, offer online shopping experiences that are often more convenient and cost-effective for consumers. As a result, traditional brick-and-mortar retailers face increased competition for customers and must find ways to remain relevant in this rapidly evolving landscape. In response to the competitive pressure from e-commerce fulfillment centers, traditional retailers may resort to cost-cutting measures to maintain their market share and profitability. One significant area where cost reductions can be implemented is labor costs, which often constitute a significant portion of operating expenses. Retailers may seek to lower wages to decrease overall labor costs and maintain price competitiveness with their online counterparts. Secondly, technological advancements, including automation and robotics, have revolutionized various aspects of the retail industry. E-commerce fulfillment centers leverage these technologies to streamline order fulfillment processes, such as picking, packing, and shipping. By automating these tasks, fewer workers are required, leading to job displacement or reduced demand for certain roles within the retail sector. This reduction in labor demand can contribute to downward pressure on wages. As automation and technology reshape the retail industry, workers may be required to transition to different roles that align with the evolving needs of the sector. However, these new roles may often be lower-skilled or lower-wage positions compared to the jobs that were displaced. The shift in job requirements and skill demands can result in downward wage adjustments for affected workers. Lastly, E-commerce platforms and larger retail chains often hold significant market power and influence due to their size, market reach, and dominance. This influence can extend to wage negotiations with workers in the retail industry.

These IV estimates are significantly larger in magnitude than the OLS estimate, indicating that the OLS method may have underestimated the true impact of Amazon FC on earnings due to endogeneity bias. Therefore, the IV approach is a more reliable method for estimating the causal impact of Amazon FC on earnings in the retail industry. The validity of the instruments was tested using the Hansen J-statistic, which confirms the exogeneity of the instruments in both IV A and IV B methods. Furthermore, the F-statistics for both methods indicate that the instruments are strong and that the IV methods are more efficient than the OLS method. Specifically, the F-statistic for IV A is 15.020, and for IV B, it is 21.509, which are both greater than the minimum threshold of 10.

For log value of employment outcome in Column 5 of Table 3.9 the IV A coefficient estimate of -0.365 suggests that a one-unit increase in the presence of Amazon FC, as instrumented by IV A (i.e., a one-unit increase in distance from Amazon FC), is associated with a 14.4% decrease in employment in the retail industry, holding all other variables constant. Similarly, the IV B coefficient estimate of -0.307 suggests that a one-unit increase in the presence of Amazon FC, as instrumented by IV B (i.e., the presence of a highway exit), is associated with a 12.1% decrease in employment in the retail industry, holding all other variables constant. The IV B estimate in column 6, using the presence intersection of interstate highway as the instrumental variable, also suggests a negative effect of Amazon FC on log employment, with a coefficient of -0.307 and a standard error of 0.142. This estimate is statistically significant at the 5% level, indicating that the presence of Amazon FC leads to a decrease in log employment in the retail industry. Overall, the IV estimates in columns 5 and 6 suggest that the negative relationship between Amazon FC and employment in the retail industry is stronger than what the OLS estimate in column 4 suggests. This highlights the importance of addressing potential endogeneity issues when estimating the impact of Amazon FC on employment.

The IV A coefficient estimate of -281.76 indicates that a one-unit increase in the presence of Amazon FC, as instrumented by IV A, is associated with a -281.76 unit decrease in the total number of establishments, holding all other variables constant. However, OLS and the IV B coefficients (Column 7 and 9 of Table 3.9) of estimate suggest there is no impact of amazon fulfillent center on the total number of retail establishments, holding all other variables constant. Overall, the results of this study demonstrate the importance of accounting for endogeneity in the analysis of the impact of Amazon fulfillment centers on the retail industry. The IV methods employed in this study provide more reliable estimates of the causal relationship between Amazon fulfillment centers and the outcomes of entrepreneurial endeavors in the retail industry, as compared to the OLS method. F-stat stands for Kleibergen and Paap (2006) first stage F statistics.

2.5.4.2 Transportation Industry

The OLS estimates reveal a negative and statistically significant coefficient for the presence of Amazon FCs on logarithmic earnings (Column 1 in Table 3.10). In Column 2 the coefficient of IV A -0.913 indicates that a one-unit increase in the presence of Amazon FC, as instrumented by IV A, is associated with a 60.0% decrease in earnings, holding all other variables constant. Similarly, the IV B coefficient estimate of -1.105 suggests that a one-unit increase in the presence of Amazon FC, as instrumented by IV B, is associated with a 66.9% decrease in earnings, holding all other variables constant.

Column 5 of Table 3.10 shows the IV A coefficient estimate of 0.426 which suggests that a one-unit increase in the presence of Amazon FC, as instrumented by IV A (i.e., a one-unit increase in distance from Amazon FC), is associated with a 53.2% increase in employment in the transportation and logistics industry, holding all other variables constant. Similarly, the IV B coefficient estimate of 0.415 in Column

6 of Table 3.10 suggests that a one-unit increase in the presence of Amazon FC, as instrumented by IV B (i.e., the presence of a highway exit), is associated with a 51.4% increase in employment in the transportation industry, holding all other variables constant.

2.5.4.3 Production Industry

Based on the results presented in Table 3.11, Columns 1 through 6, the impact of Amazon fulfillment center on the logarithm of earnings and employment appears to be no discernible impact. However, the analysis reveals that the presence of Amazon fulfillment centers has a statistically significant negative impact (Column 7-9 of Table 3.11) on the number of establishments in production industries, unlike transportation.

In terms of earnings, the results for production establishments in Table 3.11 show coefficients of 0.416 and 0.425 for IV A and IV B, respectively. These coefficients suggest a positive but statistically insignificant impact of AFCs on earnings. In contrast, the results for retail establishments in Table 3.9 show negative and statistically significant coefficients for earnings (-1.116 and -1.080 for IV A and IV B, respectively). This indicates that the presence of AFCs is associated with a decrease in earnings in the retail sector. Similarly, the results for transportation establishments in Table 3.10 show negative and statistically significant coefficients for earnings (-0.913 and -1.105 for IV A and IV B, respectively). These findings suggest that AFCs have a detrimental effect on earnings in both the retail and transportation sectors, while the impact on production establishments is less pronounced and statistically insignificant.

2.5.4.4 Comparison of IV results in different industries

Table 3.12 provides a comparison of instrumental variable (IV) results for different outcomes in the retail establishments. The outcomes of interest are logarithmic earnings (Earn), logarithmic employment (Emp), and the total number of establishments (Est). Two different IV models, labeled as IV A, IV B, are used to estimate the effects of Amazon fulfillment centers (AFCs) on these outcomes. The table presents the coefficient estimates along with standard errors and sample sizes for each model. For the outcome of logarithmic earnings, the IV A and IV B models show a statistically significant negative effect of AFCs. The coefficient estimates for IV A and IV B are -1.116 and -1.080, respectively, indicating that a one-unit increase in the presence of AFCs is associated with a substantial decrease in earnings in the retail establishments. Regarding logarithmic employment, both IV A and IV B models reveal a negative relationship between AFCs and employment in the retail establishments. The coefficient estimates for IV A and IV B are -0.365 and -0.307, respectively, indicating that a one-unit increase in the presence of AFCs leads to a decrease in log employment. These estimates suggest that AFCs have a significant impact on reducing employment levels in the retail industry. For the total number of establishments, only IV A shows a statistically significant effect. The coefficient estimate for IV A is -281.760, indicating that a one-unit increase in the presence of AFCs is associated with a considerable decrease in the total number of establishments in the retail sector. However, the coefficient estimate for IV B does not reach statistical significance.

The results in this table highlight the importance of using instrumental variable methods to address potential endogeneity issues when estimating the effects of AFCs on retail establishments. The IV estimates provide more reliable insights into the causal impact of AFCs on earnings, employment, and the number of establishments compared to ordinary least squares (OLS) estimates. The Hansen statistics and F-statistics reported in the table indicate the validity of the instruments and the strength of the IV models. It is important to note that the results are based on the data from the County Business Pattern for the years 2000-2021. The standard errors are clustered by county to account for potential heterogeneity across geographic regions. The inclusion of individual and county controls helps control for relevant demographic and economic factors.

Table 3.13 presents the results of the instrumental variable (IV) analysis for various outcomes in the transport establishments. Similar to Table 3.12 , the table is divided into two columns for different specifications of the IV models (IV A and IV B). The outcomes examined in this table are earnings and employment, as well as the total number of establishments in the transport industry. The IV estimates for earnings in the transport establishments reveal a negative relationship between the presence of Amazon fulfillment centers (AFCs) and earnings. Both IV A and IV B models indicate a statistically significant decrease in earnings associated with AFCs. The coefficient estimates for IV A and IV B are -0.913 and -1.105, respectively, suggesting that the presence of AFCs leads to a considerable reduction in earnings for workers in the transport industry. In terms of employment, the IV analysis shows a positive relationship between AFCs and employment levels in the transport sector. Both IV A and IV B models indicate a statistically significant increase in employment associated with the presence of AFCs. The coefficient estimates for IV A and IV B are 0.426 and 0.415, respectively, suggesting that AFCs contribute to a significant expansion of employment opportunities in the transport industry. For the total number of establishments, only IV A exhibits a statistically significant effect. The coefficient estimate for IV A is 353.198, indicating that the presence of AFCs is associated with a substantial increase in the total number of transport establishments. However, the coefficient estimate for IV B does not reach statistical significance.

The results from the IV analysis provide insights into the impact of AFCs on earnings, employment, and the number of establishments in the transport industry. The negative effect on earnings suggests potential challenges or competition faced by workers in this sector due to the presence of AFCs. On the other hand, the positive effect on employment indicates job creation and expansion opportunities facilitated by AFCs. The table presents the results of the instrumental variable (IV) analysis using two different instruments (IV A and IV B) and their impact on various outcome variables in the transport establishments. Surprisingly, despite using different instruments, the results show a similar extent of impact on all the outcome variables, suggesting a consistent pattern.

This consistency in the results across different instruments adds credibility to the findings and strengthens the argument that the observed effects are indeed driven by the presence of Amazon fulfillment centers (AFCs) rather than other confounding factors. When two independent instruments produce similar estimates, it reduces concerns about the validity of the instruments and increases confidence in the causal interpretation of the results. The similarity in the estimated impacts across different outcome variables also highlights the comprehensiveness of the analysis. It suggests that the presence of AFCs has a consistent and systematic influence on various aspects of the transport establishments, including earnings and employment. This provides a more holistic understanding of the effects of AFCs on the transport industry, as it goes beyond a single outcome variable and captures multiple dimensions of the economic impact.

Moreover, the consistent results across different outcome variables indicate that the impact of AFCs is not isolated to a specific aspect of the transport industry. Instead, it suggests a broader and more pervasive influence on different facets of the establishments' performance. For example, the negative impacts on earnings and employment imply that AFCs may lead to reduced profitability and workforce size in the transport sector. However, it is important to note that despite the similarity in the estimated impacts, the magnitude of the effects may differ across the outcome variables. The coefficients associated with each instrument (IV A and IV B) represent the estimated change in the outcome variable for a one-unit increase in the presence of AFCs, holding all other variables constant. These coefficients can be compared within each instrument to assess the relative importance of the different outcome variables. In conclusion, the consistent extent of impact observed across multiple outcome variables using different instruments provides robust evidence of the influence of AFCs on the transport industry. This consistency strengthens the validity of the findings and suggests a comprehensive effect of AFCs on various aspects of the establishments' performance.

Regarding employment, the results for production establishments in Table 3.14 show a negative coefficient of -0.477 for IV A, indicating a potential decrease in employment associated with AFCs. However, this coefficient is not statistically significant. In comparison, the results for retail establishments in Table 5 show statistically significant negative coefficients for employment (-0.365 and -0.307 for IV A and IV B, respectively). The results for transportation establishments in Table 3.13 also demonstrate statistically significant negative coefficients for employment (0.426 and 0.415 for IV A and IV B, respectively). These findings suggest that AFCs have a more consistent and significant negative impact on employment in both the retail and transportation sectors, while the impact on production establishments is inconclusive.

Overall, the comparison of results across the three sectors indicates that AFCs have a more substantial and consistent impact on earnings and employment in the retail and transportation sectors compared to the production sector. While AFCs appear to have limited direct effects on earnings and employment in production establishments, they show a more pronounced negative impact on these outcomes in retail and transportation establishments. These differences may be attributed to sector-specific factors, such as the nature of operations, supply chains, and consumer behavior. However, further analysis and robustness tests are necessary to delve deeper into the specific mechanisms and drivers of these variations.

Overall, these findings contribute to understanding the impact of AFCs on the retail industry and emphasize the need to consider endogeneity when evaluating their effects on earnings, employment, and the number of establishments.

2.6 Robustness Tests

While the results from the IV analysis provide valuable insights, it is also important to consider additional robustness tests to assess the reliability of the findings. The robustness analysis in Table 3.15 examines the impact of Amazon fulfillment centers (AFCs) on earnings in the retail industry using different instrumental variables and varying distances. The instrumental variable "Cargo Airport IV" is based on the distance of each county from cargo airports. By using this variable with varying distances of 100 miles, 50 miles, and 25 miles, the analysis explores how the proximity of AFCs to cargo airports affects earnings in the retail industry. For the outcome variable of earnings, the coefficient estimates for all three specifications using the "Cargo Airport IV" are statistically significant and negative. The coefficient estimates for the 100 miles, 50 miles, and 25 miles specifications are -1.116, -1.125, and -1.142, respectively. These results suggest that an increase in the presence of AFCs within 100 miles, 50 miles, or 25 miles is associated with a substantial decrease in earnings in the retail industry. This finding suggests that the presence of AFCs near cargo airports has a detrimental effect on earnings.

Similarly, for the "Interstate Highway IV" instrumental variable in Table 3.16, the coefficient estimates for all three distance specifications (100 miles, 50 miles, and 25 miles) are statistically significant and negative. The coefficient estimates for the 100 miles, 50 miles, and 25 miles specifications are -1.080, -1.093, and -1.107, respectively. These results suggest that an increase in the presence of AFCs along interstate highways within the specified distances is associated with a considerable decrease in earnings in the retail industry. Overall, the robustness analysis in Table 3.15 reinforces the findings from the previous tables (e.g., Table 3.12) regarding the negative impact of AFCs on earnings in the retail industry. The use of different instrumental variables and distance specifications helps examine the robustness of the results and provides additional evidence supporting the negative relationship between AFCs and earnings.

Similarly, the instrumental variable "Interstate Highway IV" captures the presence of AFCs based on the intersection of interstate highways with varying distances of 100 miles, 50 miles, and 25 miles. This variable allows for an examination of how the proximity of AFCs to interstate highways influences earnings in the retail industry. The coefficient estimates for all three distance specifications are statistically significant and negative, indicating that as the distance from interstate highways decreases, there is a significant decrease in earnings in the retail industry. This finding suggests that the presence of AFCs near interstate highways also has a negative impact on earnings.

The robustness analysis in Table 3.16 focuses on the impact of Amazon fulfillment centers (AFCs) on employment in the transportation industry. Similar to the previous table, this analysis utilizes different instrumental variables and varying distances to assess the robustness of the main results. For the "Cargo Airport IV" specifications, the coefficient estimates for all three distance specifications (100 miles, 50 miles, and 25 miles) are positive and statistically significant, indicating that as the distance from cargo airports decreases, there is a significant increase in employment in the transportation industry. Similarly, the "Interstate Highway IV" specifications also yield positive and statistically significant coefficient estimates for all three distance specifications. In summary, the robustness analysis in Table 3.9 demonstrates that the main results regarding the positive relationship between AFCs and employment in the transportation industry hold across different instrumental variables and varying distances.

2.7 Concluding Remarks

In conclusion, my analysis of the impact of Amazon's physical expansion on the US local labor market reveals some interesting findings. My analysis using instrumental variables shows that the presence of Amazon fulfillment centers can lead to a decrease in earnings and employment in the retail industry, while increasing employment in the transportation and logistics industry. However, I also find that the presence of Amazon fulfillment centers has a statistically significant positive impact on the number of establishments in production industries, such as retail and transportation. This indicates that Amazon's physical expansion may have a spillover effect on other industries in the local labor market.

With the rise of e-commerce platforms like Amazon, consumers are increasingly turning to online shopping, which has had a significant impact on brick-and-mortar retail stores. This shift in consumer behavior has led to a decrease in demand for traditional retail jobs, particularly those in small, local businesses. Although, the impact of e-commerce on the US labor market extends beyond just the retail industry. The growth of online shopping has also had a significant impact on the transportation and logistics industry. Companies like Amazon have invested heavily in their logistics operations, with the goal of increasing the speed and efficiency of their delivery systems.

Moreover, it is worth noting that Amazon's physical expansion has also been associated with a shift towards automation in its warehouses and distribution centers. While this may have negative implications for certain jobs in the short term, it could also create new opportunities for workers in the long run, particularly those with technical skills. As such, it is important for policymakers and stakeholders to consider both the short- and long-term impacts of Amazon's physical expansion on the labor market.

Overall, my analysis provides insights into the complex relationship between Amazon's physical expansion and the US local labor market. While the impact may vary across different industries, our findings suggest that Amazon's expansion has the potential to create new opportunities for workers and drive economic growth in the long term.

Chapter 3

RML and US Labor Market Dynamics: Unveiling the Untold Impact

3.1 Introduction

The legal marijuana industry in the United States has experienced remarkable expansion and remarkable financial performance. As of September 2021, the legal landscape pertaining to recreational marijuana has witnessed the enactment of laws in 22 states, thereby authorizing the possession of limited quantities for personal use Figure 3.13. Specifically, the adoption of recreational marijuana laws (RMLs) has occurred in 18 states, alongside the District of Columbia, thereby granting legality to the recreational use, possession, and sale of marijuana within these jurisdictions (Anderson and Rees (2023); NORML 2021). Using state-level panel data from 2000-2019 from NUSDUH Sabia et al. (2021) find that the enactment of an RML increased adult marijuana use by 1.6 to 3.6 percentage-points (18.7 to 42.5 percent) In the year 2021, a number of states passed legislation legalizing recreational marijuana, fearing that if they did not, neighboring states would receive a large share of the tax revenue that neither party wanted to miss. According to a by the Marijuana Policy Project (MPP), till late 2021 states that have legalized cannabis have together collected more than $10 billion in tax revenue since the first legal sales began in 2014. As reported by the Marijuana Policy Project, the taxation of recreational cannabis sales in states where it is legalized yielded an astounding $3.7 billion in revenue in 2021, marking a notable 34% surge compared to the previous year. These substantial financial figures underscore the industry's capacity to deliver significant economic advantages to the states that have opted for legalization. In 2021, the cannabis business had a $110,000 increase in income and 34% increase in employment (Barcott and Whitney (2022)).

Combined U.S. medical and recreational cannabis sales are expected to achieve a substantial milestone, reaching $33.6 billion by the conclusion of 2023 (Figure 3.14).

This notable growth can be largely attributed to the emergence of new adult-use markets, which have contributed significantly to the industry's expansion. Furthermore, an in-depth analysis conducted by the MJBiz Factbook suggests that retail cannabis sales could surpass $53.5 billion by the year 2027. This projected increase underscores the ongoing growth and potential of the cannabis market. Particularly promising is the outlook for the adult-use segment of the U.S. marijuana industry, especially in the near term. States such as New York, which possess the capacity to establish extensive cannabis markets, are in the process of finalizing the necessary arrangements to launch approved programs. This development bodes well for the overall growth and development of the industry, further solidifying its positive trajectory. It is important to note that as the cannabis industry has evolved, the transition from medical to recreational markets has been expedited. This accelerated shift demonstrates the industry's maturation and highlights the increasing significance of recreational cannabis in the market landscape. The fact that the cannabis industry has matured to the point where recreational sales are driving growth showcases the increasing acceptance and normalization of cannabis use. This suggests a positive trend in public perception and regulatory frameworks surrounding cannabis.

Recreational marijuana usage facilitates access for individuals with medical conditions that have not been previously sanctioned or for those whose ailments do not meet the criteria for medical cannabis qualification but who prefer not to enroll in a registry. Notably, the affordability of marijuana potentially plays a contributory role in its widespread availability among regular users. According to the most recent figures, Michigan will sell $1,311,951,737 worth of marijuana for adult use in 2021, while the state's medical marijuana sales would total $481,225,540 (Soloveichi (2021)). As a result, the state is collecting hundreds of millions in tax money. Considering the context, it is important to remark that the escalating revenue generated from sales is not a consequence of mounting costs. Indeed, the cost of cannabis for medical and adult use is steadily decreasing, both on a monthly and annual basis. Adult use was $350/oz and medicinal use was $265/oz in December 2020. Adult usage will cost $185/oz and medicinal use $175/oz in December 2021. Prices for recreational marijuana have increased considerably lower than the cost of raw plant material, according to figure 3.10.

Empirical evidence indicates a notable escalation in the financial burden imposed on employers when seeking and recruiting qualified employees, stemming from a surge in positive marijuana tests. The prevalence of positive marijuana tests has experienced a threefold increase subsequent to the legalization of the drug in Colorado and Washington.(Bowman (2014)). According to West Sound Workforce, a hiring agency, the number of positive marijuana tests has risen since Washington approved recreational marijuana usage in 2012 (Bowman (2014)). As a result, the organization has expanded its search for new employees and increased its screening time. West Sound Workforce believes that each of its four full-time recruiters spends an additional 100 hours a year because of the rise in marijuana-positive tests (Elliott et al. (2019)). A lot of cost difficulties can be seen in this case. Staffing agencies may charge more to find the same employees previously located by organizations employing staffing services. As a result, the cost of maintaining a drug-free workplace is rising as marijuana becomes more widely accepted in society. Finally, as the number of positive marijuana tests rises, the expenses of maintaining a drug-free workplace rise. This is because drug testing has gotten more expensive (Box (2015)). Employers must also decide how to run their businesses in this new environment, on top of all the other expenses.

Moreover, remote jobs neither need any pre-employment nor post-employment drug testing, and according to a group of economists in their paper, all the categories of remote jobs belong to high-skill service jobs [Information (NAICS 51), Finance and Insurance (NAICS 52), Professional Services (NAICS 54), and Management of Companies (NAICS 55)] (Eckert et al. (2020b)). Although some occupations, such as those that handle heavy machinery or carry large objects(construction and mining), and those that are physically vulnerable to injury or hazard do, nevertheless, necessitate drug testing. When it comes to blue-collar jobs, drug testing is more widespread than when it comes to white-collar jobs (Tunnell (2004)).

Marijuana usage has been shown to increase the chance of occupational injuries, which could have a negative impact on the workplace (Leigh (2011)). Given that on-the-job injuries cost an estimated $192 billion in the United States in 2007 (Leigh (2011)), it is perhaps surprising that little empirical evidence has been produced on the influence of RML on workplace safety. Minchin Jr et al. (2006) evaluated evidence suggesting drug usage in the construction business may have a negative impact on safety and productivity. The authors conducted a study of 34 construction companies, using Bureau of Labor Statistics data that shows the construction industry has the highest rate of workplace injuries. In addition, they examined five case studies, four of which dealt with construction enterprises. For most construction companies surveyed, drug testing significantly reduced the number of accidents while also lowering insurance premiums, enhancing productivity and quality of work, and decreasing employee turnover. This might create a difference in being productivity and income level in different skill spectrum of the economy.

Due to the exponential growth of this newly legalized sector, most studies have focused exclusively on medicinal marijuana. Medical marijuana has a tremendous effect on patients suffering from a small number of severely debilitating conditions. On the other side, adult cannabis users are more prevalent. Despite having enough research papers on how marijuana consumption might affect the labor market, there is no previous work on what factors might contribute to a differential impact by workers' skill in the economy post RML.

My paper contributes to the existing literature in the following ways. First, I would classify the job sectors based on high-skill, mid-skill, and low-skill service workers in the economy and see how each class is affected through wage, working hours, productivity. To my knowledge, no previous research has looked into this from this perspective after MML and RML took effect in different states. Therefore, my focus will only be narrowed to RML while keeping MML as the control variable.

Secondly, I use all the recently developed event study (weighted event study estimators) which considers staggered treatment timing of a policy and time-varying treatment effects. It takes a little time for the price to become affordable for a significant number of consumers because there are fewer shops and there is more competition. According to the design of the event research, therefore, I might observe a significant shift in the outcome of the labor market two to three years following legalization. A recent study conducted by Rand found that following the legalization of marijuana in California, the average price of marijuana plummeted by 80 percent. The monthly average price of marijuana in Colorado, Washington, Oregon, and California exhibited a consistent lower trend following legalization, and there was no trace of fluctuation in price over several years after legalization, according to another investigative study from Dills et al. (2021). So, with the help of staggered timing dynamic event study, I will be able to capture the labor outcome effect in those states compared to other treated states to see if in the latter states the outcome shows more considerable impact or not.

Thirdly, only a couple of papers showed how recreational marijuana had some labor outcome effects, and that's only from two states' perspectives – Washington and Colorado – focusing on agricultural jobs and retail service jobs in those states using QCEW data. In my paper, I'm looking into all the states using newly developed DiD methods with event study to show the effects on a different level of skilled workers due to three factors (income effect, vulnerable mid-skill workers due to not having far less access to public insurance compared to low-skill workers, and drug testing frequency for different occupations) mentioned earlier.

I have three principal findings. (i) Using Two-way Fixed Effect estimator, I have found that high-skill and mid-skill workers are seeing a reduction in their weekly earnings whereas low-skill workers are experiencing an increase in weekly income (ii) After using the event study method developed for staggered timing and heterogeneous treatment by Callaway and Sant'Anna (2021) and Sun and Abraham (2021) there is no sign of impact on income of either high-skill, mid-skill or low-skill workers which is contrary to the results found through TWFE (iii). After using both methods – TWFE and new DiD estimators – my result shows that RML has no speculative impact on working hours.

3.2 Literature Review

Economists have yet to agree on how easy access to medical marijuana affects domestic economic growth, even if a new booming industry may be created at an astonishing rate (Nicholas and Maclean (2019) and Sabia and Nguyen (2018)). On the one hand, the establishment of medical cannabis dispensaries has the potential to have favorable effects on employment and incomes, at least in the population over the age of 50 (Nicholas and Maclean (2019)). On the other hand, Sabia and Nguyen (2018) discover that males between the ages of 20 and 39 experience a slight drop in income. An prior study that used survey data on cannabis use rather than the introduction of legally prescribed medical cannabis found that cannabis consumption was associated with a 10% decrease in salaries (Van Ours (2007)). By instrumenting drug consumption with the average price of drugs, DeSimone (2002) found empirical evidence that illicit drug usage substantially reduces the likelihood of employment. In addition to making it simpler for people to get their hands on recreational marijuana could lead to more people relying on self-medication and becoming more dependent on it. It is possible that MMLs have a negative influence on health and well-being (Volkow et al. (2014), Van Ours et al. (2013)) and, as a result, wages and productivity for workers could be negatively impacted (Banerjee et al. (2017); Feltcher 2013). In contrast, new research show that legalizing medical marijuana could improve health and productivity (Chihuri and Li (2019); Nicholas and Maclean (2019); Ullman (2017)). According to Abouk et al. (2021), marijuana usage among older persons has increased following the implementation of RML, but not misuse. Their results also demonstrate that there is a decline in the number of prescriptions filled for drugs that are used to treat chronic pain after RML. A parallel fall in labor supply is not necessarily lowering Workers' Compensation participation, or industry composition shifts are not leading to an increase in the share of workers in safer industries, they note.

When drug testing was introduced, the question of whether or not it had a positive impact on a corporation was a major concern. Shepard and Clifton (1998) performed a survey as part of their research on the subject. It was conducted with a sample of 63 businesses across a diverse variety of SIC code classifications where most of them are from high tech companies and the employees are high-skilled workers. Evidence from their work suggests that drug testing techniques do not improve productivity. The use of pre-employment and random employee testing has been shown to have a negative impact on work satisfaction efficiency. Shepard and Clifton (1998) About 420,000 people are subjected to drug and alcohol testing by the federal government. staff in order to ensure the safety of the general public (SAMSHA 2004). In the United States, to identify a single drug user sucessfully in an employee drug testing program costs more than $75,000 based on the program's output, it appears. Over $11.7 million was spent by the federal government spent on a drug testing program in the first year of the new millennium, although just 0.5% of those tests yielded positive results (Strossen (1999)). During the year 1988, a significant proportion of companies implemented drug-testing programs, encompassing over half of all corporations. However, by 1990, this ratio experienced a notable decline, reducing to one third. Remarkably, around fifty percent of large-scale enterprises made a strategic decision to discontinue drug testing protocols, deeming them insufficiently cost-effective to warrant their continued application. Consequently, the implementation of such programs entails a substantial financial commitment (Hayghe (1991)).

he impact of recreational marijuana laws on industry composition is also noteworthy. Pacula et al. (2013) show that the passage of medical marijuana laws led to a decline in employment in the manufacturing sector, particularly in blue collar jobs. This suggests a potential shift in employment from traditional industries to the marijuana sector. Additionally, Gavrilova et al. (2019) find that recreational marijuana legalization in Washington state led to a reduction in reported workdays lost due to sickness or injury, indicating potential improvements in workplace safety.

Regarding drug testing, evidence suggests that it is more prevalent in blue collar jobs compared to white collar jobs. Blum et al. (2018) highlight that drug testing is more common in industries such as construction, manufacturing, and transportation, where safety concerns are paramount. They argue that the physical demands and safety risks associated with blue collar jobs contribute to a higher likelihood of drug testing. On the other hand, white collar jobs, which typically involve less physical labor and lower safety risks, are subject to drug testing less frequently.

This disparity in drug testing practices between blue collar and white collar jobs is supported by studies such as Frone (2003), who found that drug testing policies were more prevalent in industries with a higher proportion of blue collar workers. Moreover, Leukefeld et al. (2002) note that drug testing is more common in jobs requiring manual labor and occupations involving heavy machinery or public safety responsibilities.

In conclusion, the literature indicates that recreational marijuana laws have diverse effects on the US local labor market, including employment growth, wage increases in specific sectors, and potential shifts in industry composition. Moreover, drug testing practices are more prevalent in blue collar jobs, which are typically associated with higher safety risks and physical demands, compared to white collar jobs. These findings contribute to a deeper understanding of the labor market implications of recreational marijuana laws and the differential application of drug testing across job types.

3.3 Data and Descriptive Statistics

3.3.1 Source

My main data source for this analysis is from Basic Monthly Samples of IPUMS CPS ranging from 2009 to 2021 which contains individual records. For linking individuals basic monthly sample of CPS provides an indicator CPSIDP through which you can link the same person who is getting interviews at different waves of

16 months' cycle (4-8-4 cycle) of CPS where they interview 60,000 households for four consecutive months (Flood et al. 2018), then give them a break for 8 months, and they take their interviews for 4 consecutive months again till they never track the person again in their further process of interview even though they participate under a different CPSIDP later stage. Based on the IPUMS CPS data the official us unemployment rates are constructed, this matching procedure allows to quantify the transition of the labor market of survey respondents to be related to aggregate unemployment to three scenarios of labor market status every month – which I am going to discuss.

3.3.2 Type of workers

To examine how RML legalization is affecting the wage of workers on an hourly and weekly basis I need to collect the variables from IPUMS CPS. Due to the variety of changes of the industry and occupation classification with time, I construct cross-walks to facilitate the analysis of different occupations which belong to a certain industry across the different decades of CPS. To maintain the coding of occupation and industry on basis of standard definition, I follow the Standard Occupation Classification (SOC) for classifying occupation. At the same time to keep the standard format, I derive the classification system for industry from the North American Industrial Classification System (NAICS). To pin down the heterogeneous effect of the different skilled workers' wage, working hours, and hourly wage, I intend to classify the level of skill an employee or worker (**?**). To account for the base of classification I followed the percentage of skilled technical workers present in major occupation groups (2014 BLS Occupation Employment survey and O*NET Version 19).occupation and industry from different decades with the standard format I classify the occupations in three categories such as high-skill, mid-skill, and low-skill jobs to narrow down the impact of the RML on each category in form of outcome of wage, working hours etc.(Table 3.21)

3.3.3 Marijuana consumption for different income level

Table 3.17 provides the summary statistics of weekly earning , hourly wage , and hours worked last week for workers classfied by skill. Here, T means the states who are treated. TL means who are treated later eventually and NT means control states who are never treated. Table 3.18 provides income bins and calculated in each box the percentage of people who are taking marijuana of total people belonging in the same income bins from BRFSS data. I classified marijuana consumption days in three categories - 1 to 10 , 11 to 20 and 21 to 30 days.[1] The last two columns say what type of marijuana they are consuming (medical only or recreational only). In the fourth column, I calculated the proportion of that income bin as percentage of

1. BRFSS 2016-2019 provides the data of how much each respondent is consuming marijuana every thirty days

whole population (Like what's the percentage of people who earn less than 10,000$ is in the survey). In the fifth and sixth column, it mentions the total number of people of that income bin taking medical marijuana and recreational marijuana only.

3.3.4 Medical and Recreational marijuana legalized states

Effective dates for the two policies are summarized in Figure 3.9, with details provided in Table 3.15; blank cells indicate no enactment of the policy through the end of 2021.

3.4 Model and Estimation

3.4.1 Potential Problems with Differences-in-Differences Estimates

Although TWFE regressions similar to equation (1) are the workhorse models for staggered adoption research designs, they have been shown to deliver consistent estimates only under relatively strong assumptions about homogeneity in treatment effects (De Chaisemartin and d'Haultfoeuille (2020); Borusyak et al. (2021); Callaway and Sant'Anna (2021); Goodman-Bacon (2021); Sun and Abraham (2021).

To be more specific, as demonstrated in Goodman-Bacon (2021), the treatment effect estimate obtained from a TWFE model is a weighted average of all possible 2 times 2 difference-in-differences comparisons between groups of units treated at different points in time. If treatment effects are homogeneous across treated groups and across time, the TWFE estimator is consistent for the ATT. On the other hand, if the effects of treatment vary according on the group or over the course of time, the TWFE estimator will not produce consistent estimates for the ATT.

Because the TWFE procedure favors changes that occur midway through an analysis period rather than those that occur at its beginning or end, my estimates may be skewed if treatment effects have changed over time since policy implementation. The problem occurs because weights on individual policy changes are proportional to group sizes and variance of the treatment variables, with the latter being highest for groups treated in the middle of the panel. In that case, it is incorrect to assume that previously treated locations serve as a control for states that are later treated. Unless enough time has passed for these treatment effects to reach a steady-state, similar issues arise for locations that implement policies before the analysis period. I address issues raised regarding the reliability of the TWFE estimator by recreating my previous findings with the robust estimators shown in Because the TWFE procedure favors changes that occur midway through an analysis period rather than those that occur at its beginning or end, my estimates may be skewed if treatment effects have changed over time since policy implementation. The problem occurs because weights on individual policy changes are proportional to group sizes and variance of the treatment variables, with the latter being highest for groups treated in the middle of the panel. In that case, it is incorrect to assume that previously treated locations serve as a control for states that are later treated. Unless enough time has passed for these treatment effects to reach a steady-state, similar issues arise for locations that implement policies before the analysis period. I address issues raised regarding the reliability of the TWFE estimator by recreating my previous findings with the robust estimators shown in De Chaisemartin and d'Haultfoeuille (2020); Callaway and Sant'Anna (2021); and Sun and Abraham (2021). The robust estimators provide consistent estimates despite the presence of heterogeneous treatment effects across time and/or treated units. This is accomplished by disabling the 2 * 2 difference-in-differences comparisons that were previously mentioned between newly treated and already treated units. It is not a good idea to use some because they do not deal well with time-varying covariates, some of which may be critical in my study. Besides, only one treatment is considered in the case of their estimation method, whereas I analyze two policies (MML and RML sales) ; and Sun and Abraham (2021). The robust estimators provide consistent estimates despite the presence of heterogeneous treatment effects across time and/or treated units. This is accomplished by disabling the 2 times 2 difference-in-differences comparisons that were previously mentioned between newly treated and already treated units. It is not a good idea to use some because they do not deal well with time-varying covariates, some of which may be critical in our study. Besides, only one treatment is considered in the case of their estimation method, whereas I analyze two policies (MML and RML sales)

3.4.2 DID estimation methods

Both the DID-M estimator, which relies on the first differences created by De Chaisemartin and d'Haultfoeuille (2020), and the staggered implementation estimator, which takes into account long-term impacts and was developed by Callaway and Sant'Anna (2021), are used to estimate the relationship of RML on different labor outcomes. When compared to states without legalized recreational marijuana, I have not found any discernible impact on labor outcome susing the newly developed difference in difference estimators. Although treatment effects have remained relatively constant over time, this finding would not have been possible with the traditional OLS two-way fixed effects (TWFE) and first-differences (FD) estimators because the weighting problems with these estimators, as documented by Goodman-Bacon (2021), De Chaisemartin and d'Haultfoeuille (2020), Baker et al. (2022), and others, attenuate the effect to the point where it is not clearly detected. For the first time, my study shows that legalizing recreational marijuana does not lower earning outcome or working hours. On the other hand, according to the results using TWFE estiamtes, my Table 3.21 demonstrates negative impact on high skill worker's weekly earning when we are interacting RML with high skill workers' dummy. It also reduces the hourly wage of high skill workers and mid-skill workers though the weekly earning is showing no sign of significance in case of mid-skill workers. Interestingly enough, according to the results in Column (2) and Column(4) of Table 3.21 low skill workers has positive boost in weekly earning and hourly wage both post RML. However, all these distinct and significant estimates are absent when I have used staggered difference in difference estimators in Table 3.22 and Table 3.23. None of the estimators -Callaway and Sant'Anna (2021) De Chaisemartin and d'Haultfoeuille (2020) - showed any sign of discernible impact of RML on the labor outcomes. Previous research, which used the TWFE estimator, found a modest drop in income for males aged from 30-39 years old(Sabia and Nguyen (2018)). Therefore, my finding shows the difference of biased and misleading results one will produce if we rely on TWFE estimators.

Changes in labor outcomes- weekly earnings, hourly wages, and working hours for different skill levels in states where cannabis is legal versus those where it is not are the primary comparisons in this design. In order to estimates the causal effect of legalization on wages and working hours, I must assume that trends in labor outcome in states with and without legalization in recreational marijuana would have been similar in the absence of those changes. Figure 3.11, which depicts leads and lags, are shown in advance of legalization. We can conclude that there was little evidence of a significant shift before the change in labor outcomes. There is a noticeable difference in trends between the placebo effects for workers with mid and low skill levels, but the pretrend test of Callaway and Sant'Anna found that the p-value for pretend is jointly insignificant.[2] De Chaisemartin and d'Haultfoeuille (2020) as well as Callaway and Sant'Anna (2021) developed state-of-the-art difference-in-differences estimators, which are compared to traditional TWFE estimators in this paper. As it allows for both RML and MML treatments, the "DID-M" estimator developed by De Chaisemartin and d'Haultfoeuille (2020) is one of the better suited newly developed staggered DiD estiamtors for this situation. Since 2014, every state that has legalized RML has also legalized MML. Therefore, it would be difficult to determine the precise causal impact of RML and here "DID-M" estimator to estimate separately the effect of medical and recreational marijuana laws on labor outcome, unlike the biased estimate from regular TWFE. On the other hand, while giving the opportunity of staggered treatment with heterogeneous effect Callaway and Sant'Anna (2021) offers to include controls which De Chaisemartin and d'Haultfoeuille (2020) does not have at its' arsenal. Therefore in this paper I am going to compare the results derived by traditional TWFE to these two state-of-the-art difference-in-differences estimators. Basically, according to De Chaisemartin and d'Haultfoeuille (2020) setup, in a treatment effect parameter is derived by averaging the annual effects of legalization to arrive at an

2. Though, I have not provided the pretrend result from Callaway and Sant'Anna (2021) here, but it'll be available upon request

overall average effect. State-level percentage of black and Hispanics, a state-level percentage of young people ranging from age 18-25, a state-level unemployment rate are included in the preferred specification. First-differences estimators are not as robust to state heterogeneity as the DID-M estimates, which do not suffer from the same weighting issues that can change the magnitude of β_{FD} , or even cause it to be the opposite sign of all the average treatment effects of which it is comprised, as the traditional first differences estimator does. However, this estimator may be biased by attenuation if effects of legalization are asymmetric and grow over time.

Using the Callaway and Sant'Anna (2021) estimator, developed for staggered implementation of RML, I also estimate the legalization's effects separately. Clean controls are used as never-treated groups in this estimator, which takes into account post-treatment covariates. This estimator calculates the average treatment effects on the treated for groups of subjects treated in year g and estimated in year t using group-specific average treatment effects ($ATT_{g,t}$). While the DID-M estimator uses only first-differences to calculate the instantaneous effect, this group-time average treatment effect is numerically identified in a manner similar to that of DID-M. When calculating the ($ATT_{g,t}$), I make use of not-yet-treated states as comparison units.

Both in a balanced event study that shows the evolution of dynamic effects over time among groups treated for at least five periods before the end of the panel and in a simple summary parameter that gives a point estimate for the overall effect, I combine all $ATT_{g,t}$ from the Callaway and Sant'Anna (2021) estimator. For each group treated in year g, the event study aggregation takes the estimated $ATT_{g,t}$ and aggregates them. As a final step, Callaway and Sant'Anna (2021) propose the use of placebo estimators to test the parallel trends assumption that underlies their estimators. Their estimates of the placebo effect can withstand a wide range of effects. While Callaway and Sant'Anna (2021) rely on bootstrapping to estimate the standard errors of regression coefficients. The not-yet-treated are not suggested as controls, and neither are estimators based on the conditional parallel trends assumption. Under certain assumptions, Borusyak et al. (2021)) have proposed estimators that may be more efficient than those in Callaway and Sant'Anna (2020) and Sun and Abraham (2021). Their group and time fixed effects, and fixed effects for each treated (g, t) cell, can be used to estimate outcomes in TWFE regressions. Callaway and Sant'Anna (2021) and Sun and Abraham (2021) may be more biased than Borusyak et al. (2021), if parallel trends do not exactly hold. The estimates of Borusyak et al. (2021), Callaway and Sant'Anna (2020), and Sun and Abraham (2020) may have a bias-variance trade-off if parallel trends do not hold exactly (de Chaisemartin and D'Haultfoeuille (2022)). However, in this paper Callaway and Sant'Anna (2021) and Sun and Abraham (2021) will be used to estimate the dynamic effect of staggered treatment of recreational cannabis use. Pre-trends can be estimated using data going back as far as psoiible, despite the fact that treatment status has not been tracked before 2014.

Lastly, I compare the results of these estimators with the problematic but familiar two-way fixed effects and first differences estimators. Work by Goodman-Bacon (2021), Baker et al. (2022), Jakiela (2021) establish the properties of the two-way fixed effects estimator both within and outside of staggered implementation settings. Although the two-way fixed effects estimator nearly always suffers from negative weighting problems, the two-way fixed effects treatment parameter β_{TWFE} is less biased when treatment effects do not grow over time, or, in other words, when β_{TWFE} and β_{FD} are similar to each other.

3.4.3 Empirical Methodology

To adjust the sample drawn from CPS, I need to restrict my sample to the people who are aged between 18-65. Summary statistics for variables of interest are already shown in Table 3.17. I run a normal TWFE OLS regression with the following empirical equation in my sample:

$$
\begin{aligned}
ln(Y_{ist}) &= \alpha + \beta_1 \text{RML Sale} + \beta_2 MML + \beta_3 X_i + \gamma_s + \lambda_t + \delta_t + \epsilon_{ist}; \quad (3.1) \\
ln(Y_{ist}) &= \alpha + \psi_1 RMLSale + \psi_2 MML + \psi_3 RMLSale * HighSkill \\
&\quad + \psi_4 RMLSale * MidSkill + \psi_5 RMLSale * LowSkill \\
&\quad + \psi_6 highSkill + \psi_7 MidSkill + \psi_8 X_i + \gamma_s + \lambda_t + \delta_t + \epsilon_{ist} \quad (3.2)
\end{aligned}
$$

Where $ln(Y_{ist})$ is the log outcome of weekly earning, log hourly wage, log weeks worked last week. Here, i represents the individual, s is the respective states and t represents the time. RML Sale is the binary indicator of the policy being activated. X is the vector of individual-level controls including age, age squared, marital status, sex, race, ethnicity, and education. The state fixed effect is shown by γ_s which helps to account for time-invariant differences across states. Additionally, λ_t is the year fixed effect and δ_t is the month fixed effect – which captures the time varying shock across states and ϵ is the error term.

Here, ψ_1 , ψ_3, ψ_4 and ψ_5 are my interest of coefficients. ψ_1 estimate indicate whether with RML legalization if there is any change in labour outcome in those states compared to the states where recreational marijuana is not legalized. Estimates of ψ_3, ψ_4 , ψ_5 measure if after the period of the legalization , any labor outcome is decreasing or increasing for different skill spectrum – high-skill , mid-skill and low-skill workers.

My model also incorporate state-specific linear time trends in order to address this issue. Confounding factors that persist over time are accounted for by this specification. However, it is still vulnerable to the effects of regional differences in outcomes. The effective minimum wage, medical marijuana legalization, and the unemployment rate are just a few of the state-level policy controls we consider, following Ghimire and Maclean (2020) and Sabia and Nguyen (2018).

3.4.4 DiD estimation and results

Data from the baseline DiD regression are presented in table 3, which only includes workers between the ages of 18 and 65. Columns 1, 3, and 5 of Table 3.21 show that the legalization of RML sales over the entire economy did not cause a change in significance in earnings , hourly wage and hours worked. Based on a pooled sample of skilled population, these estimates do not take into account or classify jobs based on heterogeneity in worker skill sets. I will continue to show the regression results based on the different skill levels of workers in all industries so that I can better understand how legalization of cannabis affects each spectrum's workers. Using the 3rd row of Table 3.21, we can see how RML interacts with high skill levels. High-skilled workers have a negative impact on their weekly earnings and hourly wages in this example. The impact of RML has no discernible impact on high-skilled workers' working hours according to the result in Column 5. Mid-skilled workers' weekly earnings does not have a significant impact due to RML in the fourth row of Column 2 Table 5, but their hourly wages suffer greatly(Column 4). But for low-skilled workers, both weekly earnings and hourly wages are boosted after RML. A graphical presentation is also done in Figure 3.11 to get a comparative measure of delta coefficients to check how much worse/better off the labor force in various skill spectrum - high skill, low-skill and mid-skill workers.

Using the DID-M estimator, which is based on first differences in whether or not states have legalized recreational marijuana, the results shown in Panel A of Table 3.22 shows that legalizing RML does not have any significant impact on the earnings and hourly wages of the pooled workers. None of the outcomes - weekly earning, hourly wage and hours worked last week - are significant. Using the DID-M estimators in Table 3.22 from Panel B to Panel D, we can see how different skill levels have different labor outcomes due to RML. The DID-M estimator does not account for this potential asymmetry when averaging effects. Staggered difference-in-differences estimators developed by Callaway and Sant'Anna are used to estimate the effects of legalization separately (2020). Table 3.23 shows the aggregated results of staggered legalization, while shows the event study results in more detail. According to Table 3.23, states with legalized recreational marijuana have no distinctive effect on weekly earning, hours worked last week and hourly wages for all type of skilled workers. Even with the sample of pooled workers there is no significant impact on each labor outcome - which is showing the similar pattern we got from DiD-M estimators developed by de Chaisemartin and D'Haultfoeuille (2022).

Figure 3.11 suggests that the effect remains relatively stable over time and that there is little evidence of a clear difference in trends in the years leading up to the marijuana legalization, suggesting that states with and without legalization would have likely had similar trends in labor outcomes in the absence of the legalization.

3.4.5 Event Study and dynamic treatment effects

I estimate an event study model with state and year fixed effects in order to further test the validity of my identification strategy. Specifically, we estimate the following equation:

$$lnY_{ist} = \alpha + \beta_j \sum_{j \neq -1} \text{RML\_Sale} + \gamma_s + \lambda_t + \epsilon_{ist}; \tag{3.3}$$

This is to ensure that RML only uses one year when defining event time. β_0, for example, depicts the outcome of interest in the year that RML is implemented, while β_{-2} depicts the outcome of interest two years prior to implementation. The coefficient β_{-1}, which represents the year preceding implementation, is omitted. β_{-3} is the coefficient that represents 3 years prior to implementation in this equation.

If treatment effects are heterogeneous, there is no single estimand of interest, and there are decisions to be made about which weighted average one should be interested in. So in the last seven years, RML has been implemented on a statewide level and has been adopted by several states. Now, I have several groups that have been treated. It is still necessary to decide whether to weight each cohort equally or to give more weight to cohorts that were treated in more states if we use an "event-study" parameter as my estimand of interest, in which case we want a weighted average of the treatment effects k years after implementation. As a result, a researcher must be clear about what he is trying to estimate and why a weighted average is interesting from an economic standpoint. For my event study approach, I am using the weighted estimator from Callaway and Sant'Anna (2020) and Sun and Abraham (2021) to show the heterogeneity of treatment due to staggered timing, which addresses my current issue. Table 3.24 compares the weighted and unweighted event study estimands (Clarke and Tapia-Schythe (2021)), and I include a comparison with the unweighted estimator to demonstrate the differences. From Table 3.24, I'm showing the dynamic treatment from high-skill workers collapsed sample and none of the weighted event study estimands are showing any sign of causal impact on weekly earning after RML legalization which is also the result from ATT of Table 3.23. Though, from Table 3.21 it shows using TWFE OLS that RML has a negative impact on high-skill workers but from Table 3.24 there is no sign of causal relation , nor does it show the sign in panel (b) of Figure 3.11.

Workers from pooled workers sample (Panel (a) of Figure 3.11) , Workers from high-skill segment (Panel (b) of Figure 3.11), Workers from mid-skill component (Panel (c)), and workers from low-skill part (Panel d)) are all included in the event study model. Before implementing RML, there is no evidence to suggest that the labor supply for any of these groups was trending differently across states. After implementing RML, however, there is no firm evidence of significant adverse effects of RML on the labor supply of workers from the middle and low-skilled groups. The event study graph shows there is no signifcant impact on pooled worker and high-skilled worker sample either(panel a and panel b of Figure 3.11). This is completely contrary to the evidence of earnings after RML implementation in my difference-in-differences results for high skill workers from Table 3.5. This again assures in staggered treatment setup if there there is heterogeneity in treatment then adopting the traditional TWFE method will lead to biased and misleading results. Keeping in line with pooled workers and hih skill workers' sample, evidence from the event study graph suggests no significant impact on the mid-skill workers sample and the sample of low skilled workers' either (Panel (a) and (b) respectively from Figure 3.11) which does not reflect the result from TWFE OLS regression result of Table 3.21.

3.5 Robustness Check

In the context of this research, the synthetic control method can be employed as a robustness test to further validate the main results obtained using difference-in-differences (DID). If the results from the synthetic difference in differences analysis align with the main findings of the paper, it would provide additional support for the conclusion that legalizing RML does not have a significant impact on earnings and wages. Conversely, if the results differ or show a significant effect, it would indicate the need for further investigation or interpretation of the main findings. DID assumes that time shocks affect everyone equally, with only unit-level differences distinguishing them. On the other hand, synthetic control recognizes the possibility of differential exposure to shocks and incorporates a latent factor model.

The synthetic control method (SCM) takes the results of the control groups and uses a weighted average of those results to predict the results of the adopting groups "as if" those groups had not adopted the treatment. The weights are selected to optimally match the outcomes of the adopting counties prior to the adoption, and as a result, they account for any possible trends that could influence identification without the need for an assumption of parallel trends. The estimated treatment effects from the approach are equal to the difference between the actual outcomes that occurred after adoption and the outcomes that had been projected (Abadie et al. (2011), Abadie and L'Hour (2021)). To address these challenges I utilize synthetic difference-in-differences (Synth- DiD) based on Arkhangelsky et al. (2021), which borrows strengths from the DiD method as well as the synthetic control method (Abadie et al. (2011), Abadie and L'Hour (2021)).

The results from the Synth-DiD analysis, presented in Panel A of Table 3.27, demonstrate that the legalization of RML does not yield statistically significant impacts on the earnings and hourly wages of the pooled workers. Specifically, none of the examined outcomes, including weekly earnings, hourly wage, and hours worked last week, exhibit statistically significant differences. These findings suggest that the legalization of RML does not appear to exert discernible influences on the financial outcomes of the analyzed workers.The incorporation of the Synth-DiD methodology enhances the reliability and robustness of the empirical findings, mitigating potential biases associated with differential exposure to shocks and the assumption of parallel trends. Overall, the utilization of the Synth-DiD approach lends credibility to the conclusion that the legalization of RML does not generate statistically significant changes in the earnings and hourly wages of the pooled workers.

3.6 Conclusion

In comprehending the broader consequences of cannabis legalization. This study presents the first state-level estimates of the effects of recreational marijuana laws (RMLs) on employment and wages, shedding light on the labor market dynamics associated with these policies. The internal validity tests, including event studies and the analysis of average treatment effects, indicate a lack of impact on wages and employment for high-, mid-, and low-skilled workers within the pooled sample. This finding is supported by the absence of significant effects on these worker categories as demonstrated by both the event study graph and the estimation of the average treatment effect. While the standard time-fixed effects (TWFE) estimator from Table 5 initially suggests a significant impact on three types of skilled workers, the utilization of the dynamic treatment effect methodology proposed by Callaway and Sant'Anna (2021) reveals no such effect. This emphasizes the importance of employing sophisticated econometric techniques to obtain more accurate estimates of the labor market consequences of RMLs. To enhance the reliability of the findings, a robustness test was conducted using the Synth-DiD approach, which accounted for potential biases and addressed challenges associated with differential exposure to shocks and parallel trends assumption. The results of the robustness test further reinforced the main result, indicating that the lack of significant impact of RML legalization on workers' earnings and wages persisted.

From a broader perspective, the legalization of recreational marijuana can be seen as a positive step for US labor dynamics due to several reasons. First, it expands employment opportunities within the legal cannabis industry, creating new jobs and fostering economic growth. This emerging sector has the potential to generate employment across various skill levels, ranging from cultivation and production to retail and distribution. Second, the legalization of recreational marijuana can contribute to tax revenue and stimulate local economies. The regulated market can provide a source of funding for public services and infrastructure development, thereby supporting job creation and economic stability. Additionally, the legalization of recreational marijuana may lead to a reduction in illicit drug trade and associated criminal activities. This shift can potentially free up law enforcement resources, allowing them to focus on more pressing issues and enhancing public safety.

It is important to note that while this study focuses on the impact of RMLs on labor dynamics, other aspects such as public health, social equity, and regulatory frameworks should also be taken into consideration when evaluating the broader implications of marijuana legalization. In conclusion, based on the main result and robustness test of this paper, the legalization of recreational marijuana appears to be a step forward for US labor dynamics, as it does not significantly disrupt earnings and wages for the pooled workers. This supports the notion that the regulated cannabis market can coexist with existing labor structures, providing economic opportunities and contributing to overall labor market stability. It is hoped that this study will help policymakers better understand the impact of recreational marijuana use on state-level labor markets and that this information will be used to make better public policy decisions in the future. Further research and analysis are warranted to gain a more comprehensive understanding of the multifaceted impacts of marijuana legalization on various aspects of society and the economy.

3.7 Tables and Figures

Table 3.1 — ***Summary Statistics***

Panel A - Business Owners				
Sample	*2000-2004*		*2005-2009*	
	Wage	Hours	Wage	Hours
	(1)	(2)	(3)	(4)
Counties with AFC	918.77	44.67	1,029.04	41.34
	(5,453.14)	(12.03)	(5,682.80)	(12.70)
Counties without AFC	1,387.97	43.73	1,224.07	42.55
	(6,954.67)	(14.85)	(6,487.03)	(14.50)
Panel B- Business Owners				
Sample	*2010-2014*		*2015-2021*	
	Wage	Hours	Wage	Hours
	(1)	(2)	(3)	(4)
Counties with AFC	547.38	39.44	496.61	39.50
	(3,993.23)	(12.45)	(3,154.30)	(12.45)
Counties without AFC	997.47	41.28	1,270.45	40.83
	(7,782.56)	(14.50)	(15,969.34)	(13.63)

Note: Summarizing wage and hours in pre and post treatment. Here, in panel A we have summarized wage and working hours of full time business owners. In panel B, we have the pooled sample of business owners and gig workers. Here, wage is deflated by 1999 CPI which is provided by CPS ASEC variable cpi99.

Table 3.2 — *Impact of Amazon FC on Full-time business owners*

Sample	*Non-Incorporated Business*				*Incorporated Business*			
Outcome	Earn	Earn	Hours	Hours	Earn	Earn	Hours	Hours
Specification	(OLS)	(IV A)	(OLS)	(IV A)	(OLS)	(IV A)	(OLS)	(IV A)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	-0.203	-0.872	0.014	-0.009	-0.160	-0.235	-0.008	-0.017
	(0.196)	(0.535)	(0.017)	(0.026)	(0.108)	(0.366)	(0.020)	(0.023)
Obs	2,402	2,402	6,215	6,215	5,545	5,545	5,545	5,545
Hansen		0.498		0.405		0.417		0.351
F-stat		15.210		17.042		22.356		22.356

[*] Note: Here, I've run a regression using the CPS ASEC sample from 2000-2021 with just non-incorporated business owners in panel (a) and incorporated business owners' sample in panel(b). I have collapsed the data at county and year level for both panels. There in the first panel the number of observations are 2,402 and in the second panel the number of observations are 5,545. Since the income of non-incorprated business owners are not reported , therefore there is discrepany in the hours and wage section for non-incorprated business owners. In the first two columns, I tried to see how amazon's fulfillment center affects their income using OLS and IV regression. F-stat stands for Kleibrgen Paap First stage F statistics. Given that standard errors are clustered, the Kleibergen-Paap F-statistic would be an appropriate indicator of relevance in our setup (Kleibergen and Paap (2006)). Hansen signifies the over-identification Hansen J statistics test's p-value which ensures that my multiple IVs are valid. Here, wages and hours are in logarithmic form, and wage is weighted by aspect and deflated by 1999 CPI which is provided by CPS ASEC variable cpi99.

[*] Here, individual controls include age , experience, race, education

[*] Here, county controls include age composition, employment rate, and share of people living on welfare income

[*] Time FE and County FE have been applied

[*] Standard errors clustered by county in parentheses in each column.

[*] Source: CPS ASEC 2000-2021.

[*] * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.3 — *Impact of Amazon FC on business owners*

Sample	*Non-Incorporated (Business)*			*Incorporated (Business)*		
Outcome	Earn	Earn	Earn	Earn	Earn	Earn
Spec	(OLS)	(IV A)	(IV B)	(OLS)	(IV A)	(IV B)
	(1)	(2)	(3)	(5)	(6)	(7)
	-0.203	-0.872	-0.516	-0.160	-0.235	-0.469
	(0.196)	(0.535)	(0.403)	(0.108)	(0.366)	(0.581)
Obs	2,402	2,402	2,402	5,545	5,545	5,545
Hansen		0.498	0.310		0.417	0.366
F-stat		15.210	22.863		22.356	27.932

* Note: Here, I've run a regression using the CPS ASEC sample from 2000-2021 with just non-incorporated business owners. I have collapsed the data at county and year level for both panels. In the first two columns, I tried to see how amazon's fulfillment center affects their income using OLS and IV regression. F-stat stands for Kleibrgen Paap First stage F statistics. Given that standard errors are clustered, the Kleibergen-Paap F-statistic would be an appropriate indicator of relevance in my setup (Kleibergen and Paap (2006)). Hansen signifies the over-identification Hansen J statistics test's p-value which ensures that our multiple IVs are valid. Here, wages and hours are in logarithmic form, and wage is weighted by aspect and deflated by 1999 CPI which is provided by CPS ASEC variable cpi99.
* Here, individual controls include age , experience, race, education
* Here, county controls include age composition, employment rate, and share of people living on welfare income
* Standard errors clustered by county in parentheses in each column.
* Time FE and County FE have been applied
* Source: CPS ASEC 2000-2021.
* * p < 0.10, ** p < 0.05, *** p < 0.01

Table 3.4 — ***Impact of Amazon FC on being transitioned***

	Dependent variable: Transition					
Business	*Non-incorporated Business*			*Incorporated Business*		
Specification	(OLS)	(IV A)	(IV B)	(OLS)	(IV A)	(IV B)
	(1)	(2)	(3)	(4)	(5)	(6)
	0.006**	0.011**	0.009**	-0.002	-0.003	-0.001
	(0.003)	(0.005)	(0.004)	(0.002)	(0.008)	(0.004)
Observations	6,684	6,684	6,684	6,684	6,684	6,684
F-stat		18.496	22.581		18.496	22.581
Hansen		0.415	0.328		0.407	0.311

* Note: Here, Here, Noninc Business means non-incorporated business, and Inc Business stands for an incorporated business. I have collapsed the data at county and year level. I have run OLS and IV regressions using the CPS ASEC sample from 2000-2021 with the full sample. Here, the dependent outcome for the first two columns is the binary variable which indicates that if a person wasn't a full-time business owner 1 year before. But, when the CPS interviews him after a year the person changes his job to be a full-time business owner according to the definition given by Fairlie and Fossen (2020). The binary indicator can be created by the upside variable which helps to see the transition in the occupation of a person in the CPS ASEC sample after a year. Here, I create this transition variable for people in two categories - A person who transitioned to be a non-incorporated business owner after a year which is used as the dependent variable for the first two columns. I also used the binary indicator of being transitioned to be an incorporated business owner as dependent outcome for the last two columns.
* Here, county controls include age composition, employment rate, and share of people living on welfare income
* Standard errors clustered by county in parentheses in each column.
* Time FE and County FE have been applied
* Source: CPS ASEC 2000-2021.
* * p < 0.10, ** p < 0.05, *** p < 0.01

Table 3.5 — *Transition of different spectrum of skilled workers to business owners*

	Dependent variable: Transition					
Skill	High-Skill	Mid-Skill	Low-Skill	High-Skill	Mid-Skill	Low-Skill
Specification	(IV A)	(IV A)	(IV A)	(IV B)	(IV B)	(IV B)
	(1)	(2)	(3)	(4)	(5)	(6)
	0.009***	0.016**	0.005	0.013***	0.006**	0.004
	(0.003)	(0.007)	(0.004)	(0.005)	(0.003)	(0.006)
Observations	6,684	6,684	6,684	6,684	6,684	6,684
F-stat	18.496	18.496	18.496	22.581	22.581	22.581

* Note: Here, To account for the base of classification I followed the percentage of skilled technical workers present in major occupation groups (2014 BLS Occupation Employment survey and O*NET Version 19) occupation and industry from different decades with the standard format. I have collapsed the data at county and year level. I classify the occupations in three categories such as High-skill, Mid-skill, and low-skill jobs. David and Dorn (2013) divided occupations into 12 categories which they define as having similar properties based on characteristics including routine intensity of the work, average educational attainment of the workers, and employment dynamics
* Here, county controls include age composition, employment rate, and share of people living on welfare income
* Standard errors clustered by county in parentheses in each column.
* Time FE and County FE have been applied
* Source: CPS ASEC 2000-2021.
* * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.6 — *Impact of Amazon FC on business owners' income with different bandwidth*

Sample	*Non-Incorporated Business*			*Non-Incorporated Business*		
Outcome	Earning	Earning	Earning	Earning	Earning	Earning
Airport Distance	(100 miles)	(50 miles)	(25 miles)	(100 miles)	(50 miles)	(25 miles)
Specification	(IV A)	(IV A)	(IV A)	(IV B)	(IV B)	(IV B)
	(1)	(2)	(3)	(4)	(5)	(6)
	-.872	-.913*	-.952*	-.516	-.580	-611*
	(0.535)	(0.506)	(0.491)	(0.403)	(0.395)	(.342)
Obs	2,402	2,402	2,402	2,402	2,402	2,402

* Note: Here, I've run a regression using the CPS ASEC sample from 2000-2021 with just non-incorporated business owners in the panel. With a different threshold of distance ring from the nearest cargo airport (100, 50, and 25 miles ring respectively) I wanted to show here how the non-incorporated business owners' income changes with the presence of an Amazon fulfillment center. Hansen signifies the over-identification Hansen J statistics test's p-value which ensures that our multiple IVs are valid. Here, wages and hours are in logarithmic form, and wage is weighted by aspect and deflated by 1999 CPI which is provided by CPS ASEC variable cpi99.
* Here, individual controls include age, experience, race, education.
* Here, county controls include age composition, employment rate, and share of people living on welfare income.
* Standard errors clustered by county in parentheses in each column.
* Source: CPS ASEC 2000-2021.
* * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.7 — *Impact of Amazon FC on Innovation of incorporated and non-incorporated firms*

	Dependent variable: Total Patents				
Specification	(TWFE)†	(IV A)	(IV B) §	(CSDID)††	(Synth-DiD)‡
	(1)	(2)	(3)	(4)	(5)
	-60.414**	-21.830	-7.551	-44.494	-10.451
	(20.242)	(19.565)	(18.234)	(30.026)	(5.893)
Observations	52,661	52,661	52,661	52,661	67,474
Individual Controls	✓	✓	✓	✗	✗
County Controls	✓	✓	✓	✗	✗
Time FE	✓	✓	✓	✓	✓
County FE	✓	✓	✓	✓	✓

† Here, I've run TWFE regression using the USPTO data from 2000-2021. Here, the dependent outcome is the total patents filed by individuals and manufacturers at different years across all the counties of united states.

†† This table presents results from non-parametric event study method of Callaway and Sant'Anna (2021). It shows the average treatment effects of E-commerce diffusion on treated counties. IV in Column (2)

‡ Point estimates for the average treatment effects using Synth-DiD developed by Arkhangelsky et al. (2021) is in Columns (4).

* Here, county controls include poverty rate and total income

* Standard errors clustered by county in parentheses in each column.

* Source: USPTO data 2000-2021

* * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.8 — *Summary Statistics*

Sample	*Retail Inudstry*		*Transportation Industry*		*Production Industry*	
	Treat (1)	Control (2)	Treat (3)	Control (4)	Treat (5)	Control (6)
Earning	28,976.58	34,706.91	20,125.47	23,009.28	31,417.55	29,602.04
Employment	27,792.23	32,909.65	21,481.94	14,604.78	29,560.47	31,087.92
Establishment	1,906.97	2,115.03	1,526.58	674.69	902.41	1,480.72

* Note: This provides the summary statistics of earning , employment , total number of establishments in retail, transportation and production industry.
* Here, all dollar values of earning are deflated by CPI 2017.
* Source: County Business Pattern Data (2000-2021)

Table 3.9 — *Impact on Retail Industry*

Sample	*Retail*						*Retail*		
Outcome	Earn	Earn	Earn	Emp	Emp	Emp	Est	Est	Est
Spec	(OLS)	(IV A)	(IV B)	(OLS)	(IV A)	(IV B)	(OLS)	(IV A)	(IV B)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	-0.157***	-1.116***	-1.080***	-0.031**	-0.365**	-0.307**	29.284	-281.760**	-233.405
	(0.038)	(0.336)	(0.415)	(0.014)	(0.173)	(0.142)	(31.581)	(125.196)	(257.613)
Obs	64,218	64,218	64,218	64,218	64,218	64,218	64,218	64,218	64,218
Hansen		0.316	0.319		0.334	0.358		0.346	0.371
F-stat		15.020	21.509		15.020	21.509		15.020	21.509

* Here, individual controls include age, experience, race.
* Here, county controls include age composition, employment rate, and share of people living on welfare income.
* Here, I have collapsed the data at county and year level.
* Here, Earn means log values of earning deflated by CPI 2017. Emp means log value of employment. Est means total establishment in a county.
* IV A stands for the distance from cargo airport and year dummy interaction. IV B stands for distance from intersection of interstate highway and year dummy interaction.
* Standard errors clustered by county in parentheses in each column.
* F-stat stands for Kleibrgen Paap First stage F statistics. Given that standard errors are clustered, the Kleibergen-Paap F-statistic would be an appropriate indicator of relevance in our setup Kleibergen and Paap (2006).
* Hansen signifies the over-identification Hansen J statistics test's p-value which ensures that our multiple IVs are valid.
* Source: County Business Pattern 2000-2021.
* * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.10 — *Impact of Amazon FC on Transportation and Logistics Industry*

Sample	*Transportation and Logistics*					*Transportation and Logistics*			
Outcome	Earn	Earn	Earn	Emp	Emp	Emp	Est	Est	Est
Spec	(OLS)	(IV A)	(IV B)	(OLS)	(IV A)	(IV B)	(OLS)	(IV A)	(IV B)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	-0.149***	-0.913***	-1.105**	0.184***	0.426**	0.415***	152.279***	353.198***	318.616***
	(0.030)	(0.328)	(0.527)	(0.064)	(0.201)	(0.133)	(27.693)	(128.501)	(111.383)
Obs	56,900	56,900	56,900	56,900	56,900	56,900	56,900	56,900	56,900
Hansen		0.325	0.332		0.349	0.361		0.330	0.386
F-stat		14.162	19.708		14.162	19.708		14.162	19.708

* Here, individual controls include age, experience, race.
* Here, county controls include age composition, employment rate, and share of people living on welfare income.
* Here, I have collapsed the data at county and year level.
* Here, Earn means log values of earning deflated by CPI 2017. Emp means log value of employment. Est means total establishment in a county.
* IV A stands for the distance from cargo airport and year dummy interaction. IV B stands for distance from intersection of interstate highway and year dummy interaction.
* Standard errors clustered by county in parentheses in each column.
* F-stat stands for Kleibergen Paap First stage F statistics. Given that standard errors are clustered, the Kleibergen-Paap F-statistic would be an appropriate indicator of relevance in our setup Kleibergen and Paap (2006).
* Hansen signifies the over-identification Hansen J statistics test's p-value which ensures that our multiple IVs are valid.
* Source: County Business Pattern 2000-2021.
* * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.11 — *Impact of Amazon FC on Production and Manufacturing Industry*

Sample	*Production and Manufacturing*					*Production and Manufacturing*			
Outcome	Earn	Earn	Earn	Emp	Emp	Emp	Est	Est	Est
Spec	(OLS)	(IV A)	(IV B)	(OLS)	(IV A)	(IV B)	(OLS)	(IV A)	(IV B)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	0.021	0.416	0.425	-0.037	-0.477	-0.365	-78.035***	-328.073**	-323.457**
	(0.032)	(0.381)	(0.479)	(0.052)	(0.386)	(0.272)	(21.004)	(127.928)	(139.582)
Obs	63,262	63,262	63,262	63,262	63,262	63,262	63,262	63,262	63,262
Hansen		0.317	0.321		0.329	0.338		0.347	0.359
F-stat		14.825	21.173		14.825	21.173		14.825	21.173

* Here, individual controls include age, experience, race.
* Here, county controls include age composition, employment rate, and share of people living on welfare income.
* Here, I have collapsed the data at county and year level.
* Here, Earn means log values of earning deflated by CPI 2017. Emp means log value of employment. Est means total establishment in a county.
* IV A stands for the distance from cargo airport and year dummy interaction. IV B stands for distance from intersection of interstate highway and year dummy interaction.
* Standard errors clustered by county in parentheses in each column.
* F-stat stands for Kleibergen Paap First stage F statistics. Given that standard errors are clustered, the Kleibergen-Paap F-statistic would be an appropriate indicator of relevance in our setup Kleibergen and Paap (2006).
* Hansen signifies the over-identification Hansen J statistics test's p-value which ensures that our multiple IVs are valid.
* Source: County Business Pattern 2000-2021.
* * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.12 — *Comparison of IV results in retail establishments*

Sample	*Establishments*				*Establishemnts*	
Outcome	Earn	Earn	Emp	Emp	Est	Est
Spec	(IV A)	(IV B)	(IV A)	(IV B)	(IV A)	(IV B)
	(1)	(2)	(3)	(4)	(5)	(6)
	-1.116***	-1.080***	-0.365**	-0.307**	-281.760**	-233.405
	(0.336)	(0.415)	(0.173)	(0.142)	(125.196)	(257.613)
Obs	64,218	64,218	64,218	64,218	64,218	64,218
Hansen	0.316	0.319	0.334	0.358	0.346	0.371
F-stat	15.020	21.509	15.020	21.509	15.020	21.509

* Here, individual controls include age , experience, race
* Here, county controls include age composition, employment rate, and share of people living on welfare income
* Standard errors clustered by county in parentheses in each column.
* Source: County Business Pattern 2000-2021.
* * p < 0.10, ** p < 0.05, *** p < 0.01

Table 3.13 — *Comparison of IV results in transport establishments*

Sample	*Establishments*				*Establishemnts*	
Outcome	Earn	Earn	Emp	Emp	Est	Est
Spec	(IV A)	(IV B)	(IV A)	(IV B)	(IV A)	(IV B)
	(1)	(2)	(3)	(4)	(5)	(6)
	-0.913***	-1.105**	0.426**	0.415***	353.198***	318.616***
	(0.328)	(0.527)	(0.201)	(0.133)	(128.501)	(111.383)
Obs	56,900	56,900	56,900	56,900	56,900	56,900
Hansen	0.325	0.332	0.349	0.361	0.330	0.386
F-stat	14.162	19.708	14.162	19.708	14.162	19.708

* Here, individual controls include age , experience, race
* Here, county controls include age composition, employment rate, and share of people living on welfare income
* Standard errors clustered by county in parentheses in each column.
* Source: County Business Pattern 2000-2021.
* * p < 0.10, ** p < 0.05, *** p < 0.01

Table 3.14 — *Comparison of IV results in production establishments*

Sample	*Establishments*				*Establishemnts*	
Outcome	Earn	Earn	Emp	Emp	Est	Est
Spec	(IV A)	(IV B)	(IV A)	(IV B)	(IV A)	(IV B)
	(1)	(2)	(3)	(4)	(5)	(6)
	0.416	0.425	-0.477	-0.365	-328.073**	-323.457**
	(0.381)	(0.479)	(0.386)	(0.272)	(127.928)	(139.582)
Obs	63,262	63,262	63,262	63,262	63,262	63,262
Hansen	0.317	0.321	0.329	0.338	0.347	0.359
F-stat	14.825	21.173	14.825	21.173	14.825	21.173

[*] Here, individual controls include age , experience, race
[*] Here, county controls include age composition, employment rate, and share of people living on welfare income
[*] Standard errors clustered by county in parentheses in each column.
[*] Source: County Business Pattern 2000-2021.
[*] * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.15 — *Impact of Amazon FC on retail industry earning robustness*

Sample	*Cargo Airport IV*			*Interstate Highway IV*		
Outcome	Earning	Earning	Earning	Earning	Earning	Earning
Distance	(100 miles)	(50 miles)	(25 miles)	(100 miles)	(50 miles)	(25 miles)
Specification	(IV A)	(IV A)	(IV A)	(IV B)	(IV B)	(IV B)
	(1)	(2)	(3)	(4)	(5)	(6)
	-1.116***	-1.125***	-1.142***	-1.080***	-1.093***	-1.107***
	(0.336)	(0.331)	(0.327)	(0.415)	(0.398)	(0.391)
Obs	64,218	64,218	64,218	64,218	64,218	64,218

* Here, Earning means log values of earning deflated by CPI 2017 in a county.
* Cargo Airport IV stands for the distance from cargo airport and year dummy interaction. Interstate Highway IV stands for distance from intersection of interstate highway and year dummy interaction. By using this variable with varying distances of 100 miles, 50 miles, and 25 miles, the analysis explores how the proximity of AFCs to cargo airports or intersection of interstate highway affects earnings in the retail industry.
* Standard errors clustered by county in parentheses in each column.
* * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.16 — *Impact of Amazon FC on Transportation Industry Employment Robustness*

Sample	*Cargo Airport IV*			*Interstate Highway IV*		
Outcome	Emp	Emp	Emp	Emp	Emp	Emp
Distance	(100 miles)	(50 miles)	(25 miles)	(100 miles)	(50 miles)	(25 miles)
Specification	(IV A)	(IV A)	(IV A)	(IV B)	(IV B)	(IV B)
	(1)	(2)	(3)	(4)	(5)	(6)
	0.426**	0.434**	0.467**	0.415***	0.433***	0.452***
	(0.201)	(0.196)	(0.191)	(0.133)	(0.128)	(0.126)
Obs	56,900	56,900	56,900	56,900	56,900	56,900

* Here, Emp means log value of employment in a county.
* Cargo Airport IV stands for the distance from cargo airport and year dummy interaction. Interstate Highway IV stands for distance from intersection of interstate highway and year dummy interaction. By using this variable with varying distances of 100 miles, 50 miles, and 25 miles, the analysis explores how the proximity of AFCs to cargo airports or intersection of interstate highway affects employment in the transportation and logistics industry.
* Standard errors clustered by county in parentheses in each column.
* * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.17 — *Summary Statistics*

Sample	*Pooled*			*Low-Skill*			*Mid-Skill*			*High-Skill*		
	T	T L	NT	T	TL	NT	T	T L	NT	T	T L	NT
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Week	1,182.7	979.2	908.2	786.5	672.0	647.3	962.6	853.8	828.5	1,671.1	1,439.78	1,321.07
Hours	38.49	38.03	38.41	36.12	35.71	35.98	38.11	37.92	38.88	40.72	40.65	40.84
Wage	20.48	17.76	16.89	16.02	13.82	13.31	20.78	18.50	17.67	30.05	26.55	24.33

* Note: This provides the summary statistics of earnings, hourly wage, and weeks worked last week for workers with different sets of skills. Here, T means the states that are treated. TL means states that are treated later eventually, and NT means control states that are never treated. The middle one always represents states that were treated later.
* Columns (4) to (6) are for low-skill workers, columns (7) to (9) for mid-skill workers, and columns (10) to (12) for high-skill workers. Columns (1) to (3) represent the whole sample.
* Here, control states are the exact states that eventually implemented RML as a statutory policy, and the rest of the states are excluded.
* Source: CPS 2009-2021

Table 3.18 — ***Marijuana intake days on income level (%)***

Level of Income	Marijuana 1 to 10	Marijuana 11 to 20	Mar 21 to 30	Medical only	Recreational only
	(1)	(2)	(3)	(4)	(5)
Income < 10000	1.04	0.31	0.85	0.72	0.48
10000 < Income < 15000	0.84	0.27	0.87	0.66	0.38
15000 < Income < 20000	0.77	0.19	0.74	0.49	0.37
20000 < Income < 25000	0.71	0.19	0.74	0.44	0.38
25000 < Income < 35,000	0.67	0.19	0.67	0.37	0.40
35,000 < Income < 50,000	0.68	0.18	0.57	0.31	0.43
50,000 < Income < 75,000	0.67	0.17	0.49	0.28	0.42
Income > 75,000	0.70	0.17	0.36	0.24	0.46

* Note: Here, I created some income bins and calculated in each box the percentage of people taking marijuana out of the total people belonging to the same income bins. Marijuana consumption days are classified into three categories - 1 to 10, 11 to 20, and 21 to 30 days. The last two columns indicate the type of marijuana they are taking.
* Source: BRFSS 2016-2019

Table 3.19 — *Marijuana intake days for total observation of an income bin*

Level of Income	1 to 10 (1)	11 to 20 (2)	21 to 30 (3)	Proportion of sample (4)	Medical only (5)	Recreational only (6)
Income < 10000	737	221	605	.039	511	340
10000 < Income < 15000	629	207	657	.042	499	283
15000 < Income < 20000	829	205	803	.061	534	395
20000 < Income < 25000	948	250	989	.075	595	509
25000 < Income < 35,000	1,054	297	1,043	.088	583	618
35,000 < Income < 50,000	1,414	376	1,182	.117	653	900
50,000 < Income < 75,000	1,585	391	1,160	.134	653	1,001
Income > 75,000	3,420	775	1,771	.276	1,172	2,235

* Note: Here, I counted the total number of people taking marijuana for a different number of days of a certain income bin. In the fourth column, I calculated the proportion of that income bin as a percentage of the whole population. In the fifth and sixth columns, it mentions the total number of people of that income bin taking medical marijuana and recreational marijuana only.
* Source: BRFSS 2016-2019

Table 3.20 — *Summary statistics of dependent and control variables*

	Mean	Standard Deviation	Minimum	Maximum	Number of observations
Weekly earning	920.028	662.583	1	2884.61	1907825
Hours worked	38.417	12.369	0	99	8104004
Hourly Wage	17.03	9.875	1.01	99.99	1109662
Age	41.926	13.679	18	65	1.19e+07
Male	.486	.4998	0	1	1.19e+07
Hispanic	.135	.342	0	1	1.19e+07
Minimum wage	6.765	.987	0	10.94	1.19e+07
Employed Percentage	.005	.005	.0009	.032	1.19e+07
Less than high school	.099	.299	0	1	1.19e+07
High school	.290	.453	0	1	1.19e+07

* Note: This provides the summary statistics of earning, hourly wage, weeks worked last week for workers with different sets of skills. "Auto" stands for automated workers (routine-intensive work like manufacturing, construction, etc.), where it is very labor-intensive. We are trying to see if labor-intensive workers are more affected based on productivity compared to high-skill workers, where more mentally challenging work is involved instead of being labor-intensive. Here, control states are the exact states that eventually took RML (statutory policy), and the rest of the states are excluded.
* Source: CPS 1996-2021

Table 3.21 — *Impact of RML Sales on Weekly Earning Interaction*

	Earning	earning	Hourly wage	Hourly wage	Hours worked	Hours worked
	(1)	(2)	(3)	(4)	(5)	(6)
RML Sales	0.005		0.006		0.002	
	(0.004)		(0.004)		(0.002)	
MML	-0.008***	-0.009***	-0.003*	-0.003*	-0.003*	-0.003*
	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
RML High		-0.017***		-0.013***		-0.003
		(0.005)		(0.004)		(0.003)
RML Mid		-0.001		-0.010***		0.003
		(0.005)		(0.003)		(0.003)
RML Low		0.032***		0.026***		0.007
		(0.006)		(0.008)		(0.004)
Observations	1907825	1907825	1109662	1109662	8096204	8096204
Number of clusters	51	51	51	51	51	51
Adjusted R2	0.474	0.474	0.454	0.454	0.121	0.121

* Note: In the pooled sample, it includes all the workers and in the next three columns, it only creates a subsample regression on different levels of skilled people to see how their weekly earnings are showing variation after MML and RML legalization across different states.
* Year-month FE. State FE, Individual controls, State controls and Occupation FE were used in the regression
* Here, state controls include minimum wage in different states across different years from 1996-2021. Besides, minimum wage I control for race composition, age composition, and unemployment rate.
* Standard errors clustered by state in parentheses in each column.
* Source: CPS 1996-2021.
* * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.22 — *DID-M Estimator for different skill spectrum*

	Panel A - Pooled Workers		
	DID-M Estimator		
	Earning	Hourly wage	Hours worked
	(1)	(2)	(3)
ATT	-.005	-.002	.003
	(.004)	(.004)	(.004)
	Panel B- High-skill		
	DID-M Estimator		
	Earning	Hourly wage	Hours worked
	(1)	(2)	(3)
ATT	-.006	.001	-.002
	(.005)	(.009)	(.002)
	Panel C- Mid-skill		
	DID-M Estimator		
	Earning	Hourly wage	Hours worked
	(1)	(2)	(3)
ATT	-.010	.002	.011
	(.007)	(.511)	(.009)
	Panel D- Low-skill		
	DID-M Estimator		
	Earning	Hourly wage	Hours worked
	(1)	(2)	(3)
ATT	-.010	-.018	-.006
	(.013)	(.009)	(.006)
State FE	✓	✓	✓
Time FE	✓	✓	✓
Ind controls	✓	✓	✓
Observations	663	663	663
clusters	51	51	51

Note: Summarizing wage and hours in pre and post treatment. Here, in panel A we have summarized wage and working hours of full time business owners. In panel B, we have the pooled sample of business owners and gig workers. Here, wage is deflated by 1999 CPI which is provided by CPS ASEC variable cpi99.

Table 3.23 — *Staggered Estimator for different skill spetrum*

	Earning	Hourly wage	Hours worked
Panel A - Pooled Workers			
	Staggered Estimator		
	(1)	(2)	(3)
ATT	-.011	-.003	.002
	(.005)	(.005)	(.005)
Panel B- High-skill			
	Staggered Estimator		
	Earning	Hourly wage	Hours worked
	(1)	(2)	(3)
ATT	-.007	.001	-.003
	(.009)	(.013)	(.004)
Panel C- Mid-skill			
	Staggered Estimator		
	Earning	Hourly wage	Hours worked
	(1)	(2)	(3)
ATT	-.012	.002	.014
	(.008)	(.771)	(.009)
Panel D- Low-skill			
	Staggered Estimator		
	Earning	Hourly wage	Hours worked
	(1)	(2)	(3)
ATT	-.010	-.018	-.006
	(.013)	(.009)	(.006)
State FE	✓	✓	✓
Time FE	✓	✓	✓
Ind controls	✓	✓	✓
Observations	663	663	663
clusters	51	51	51

Notes. * $p< 0.1$; ** $p< 0.05$; *** $p< 0.01$
Here, I collapsed the data from repeated cross section to balanced panel at state and year level of 663 observations for low-skill sample workers separately.
Standard errors clustered by state in parentheses in each column.
Source: CPS 2009-2021.

Table 3.24 — *Dynamic Treatment effects for pooled workers*

	CS	**SA**	**EVENTDD**
	(1)	(2)	(3)
RML_{-3}	-.014*	-.005	-.005
	(.006)	(.008)	(.011)
RML_{-2}	-.001	-.003	-.003
	(.013)	(.004)	(.007)
RML_{0}	-.014***	-.011*	-.006
	(.004)	(.005)	(.004)
RML_{+1}	-.011*	-.006	-.005*
	(.013)	(.005)	(.002)
RML_{+2}	-.008	-.017*	-.011**
	(.013)	(.006)	(.004)
RML_{+3}	-.015	-.016*	-.010*
	(.016)	(.007)	(.004)
State FE	✓	✓	✓
Time FE	✓	✓	✓
Individual controls	✓	✓	✓
Observations	663	663	663
Number of clusters	51	51	51

Notes. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$
This table reports the dynamic effect of RML sales law on weekly earning for pooled workers sample from 2009-2021. All columns are estimated with respect to the period before the law is enacted RML_{-1}. Dynamic treatment weighted estimator developed by Callaway and Sant'Anna (2020) is shown in column 1. In columns 2, we implement the interaction weighted estimator proposed by Sun and Abraham (2021). In column3 , we report results using an unweighted-stacked event study design using eventdd. We report standard error in parentheses below the coefficients.
Standard errors clustered by state in parentheses in each column.
Source: CPS BMS 2009-2021.

Table 3.25 — *Effective Dates of Marijuana Policies*

State	Medical Marijuana	Recreational Marijuana
Alabama	.	.
Alaska	3/1999	10/2016
Arizona	4/2011	.
Arkansas	.	.
California	11/1996	01/2018
Colorado	6/2001	01/2014
Connecticut	5/2012	.
Delware	7/2011	.
DC	7/2010	.
Florida	.	.
Georgia	.	.
Hawaii	12/2000	.
Idaho	.	.
Illinois	.	01/2020
Iowa	.	.
Louisiana	.	.
Maine	12/1999	.
Maryland	.	.
Massachusetts	1/2013	11/2018
Michigan	12/2008	12/2019
Minnesota	.	.
Montana	11/2004	.
Nevada	10/2001	07/2017
New Hampshire	07/2013	.
New Jersey	10/2010	.
New Mexico	07/2007	.
New York	.	.
North Carolina	.	.
North Dakota	.	.
Oklahoma	.	.
Oregon	12/1998	10/2015
Pennsylvania	.	.
Rhode Island	01/2006	.
South Dakota	.	.
Tennessee	.	.
Texas	.	.
Vermont	07/2004	.
Virginia	.	.
Washington	11/1998	07/2014
West Virginia	.	.
Wyoming	.	.

Effective date Medical Marijuana and Recreational Marijuana Dispensary Law

Table 3.26 — Skill Classification

	Occupational Title	Skill
1	Managers	High skill
2	Professionals	High skill
3	Technicians	High skill
4	Office and admin	Mid skill
5	Mechanics and repairers and construction workers	Mid skill
6	Transportation and material moving occupations	Mid skill
7	Precision production and machine operators	Mid skill
8	Food preparation and Services	Low skill
9	Buildings and grounds, cleaning	Low skill
10	Personal care and personal services	Low skill
11	Agriculture	Low skill
12	Sales and related	Low skill

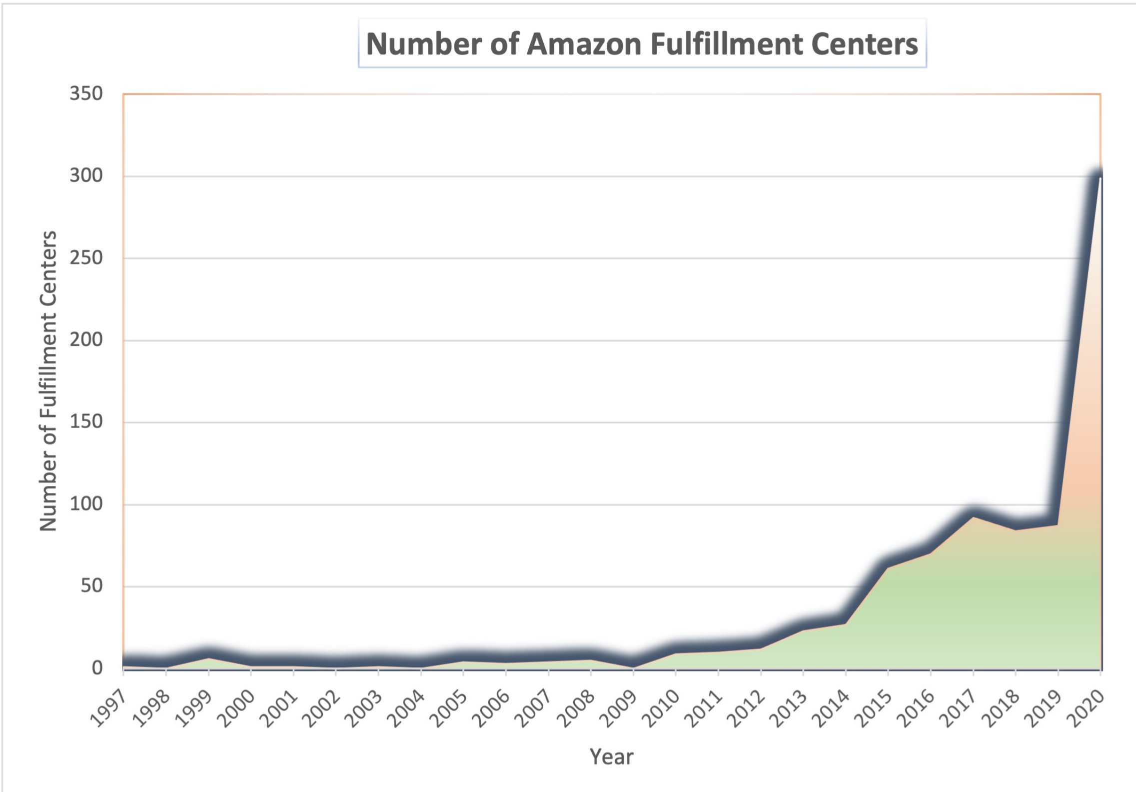
Note: Autor and Dorn (2013) divided occupations into 12 categories which they define as having similar properties based on characteristics including - routine intensity of the work, average educational attainment of the workers, and employment dynamics

Table 3.27 — *SDID for different skill spetrum*

	Panel A - Pooled Workers		
	SDID		
	Earning	Hourly wage	Hours worked
	(1)	(2)	(3)
ATT	-.008	-.002	.001
	(.005)	(.002)	(.006)
	Panel B- High-skill		
	SDID		
	Earning	Hourly wage	Hours worked
	(1)	(2)	(3)
ATT	-.006	.002	-.003
	(.007)	(.009)	(.005)
	Panel C- Mid-skill		
	SDID		
	Earning	Hourly wage	Hours worked
	(1)	(2)	(3)
ATT	-.010	.001	.006
	(.009)	(.014)	(.008)
	Panel D- Low-skill		
	SDID		
	Earning	Hourly wage	Hours worked
	(1)	(2)	(3)
ATT	-.007	-.012	-.004
	(.008)	(.014)	(.003)
State FE	✓	✓	✓
Time FE	✓	✓	✓
Ind controls	✗	✗	✗
Observations	663	663	663
clusters	51	51	51

Notes. * $p< 0.1$; ** $p< 0.05$; *** $p< 0.01$
Here, I collapsed the data from repeated cross section to balanced panel at state and year level of 663 observations for low-skill sample workers separately.
Standard errors clustered by state in parentheses in each column.
Source: CPS 2009-2021.

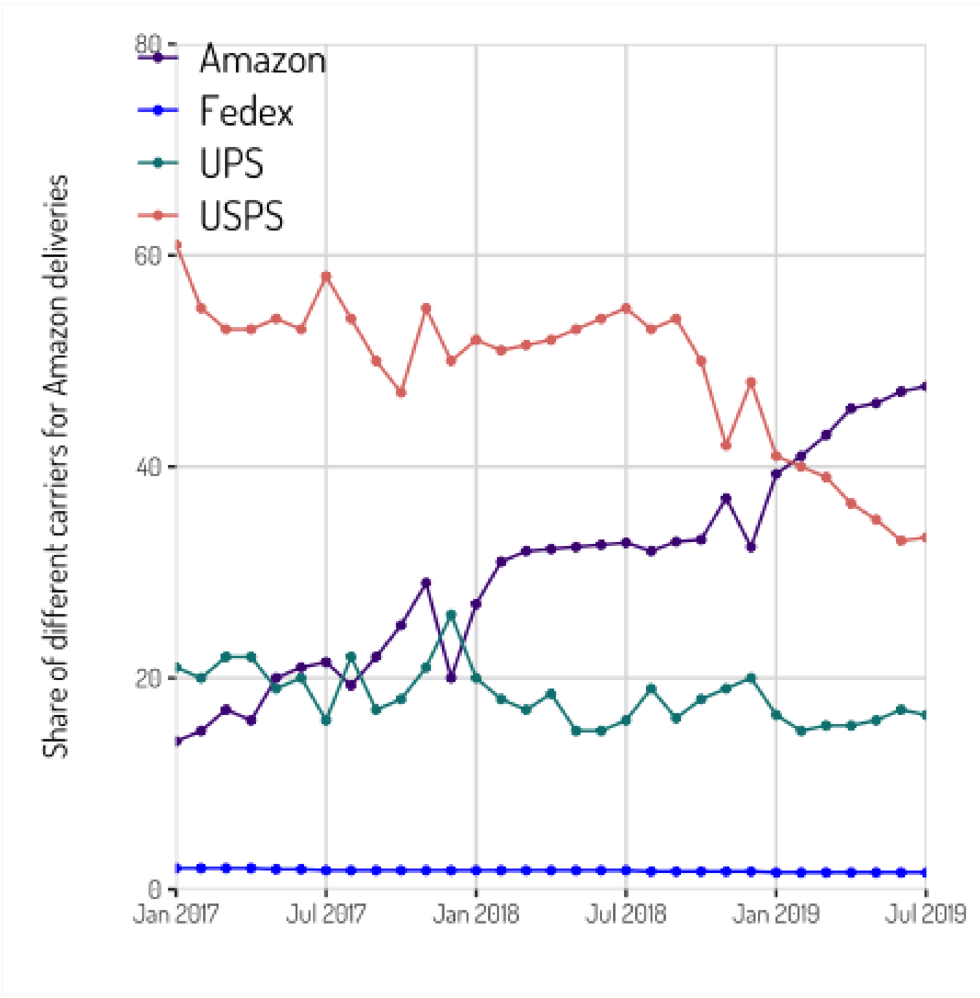
Figure 3.1 — Number of fulfillment centers across the USA from 1997-2020



Note: This figure provides a rundown, in single-year increment, of the expansion of Amazon's fulfillment centers over the course of the period beginning in 1997 and ending in 2020. From a total of one individual location in 1997, the company increased the number of fulfillment centers in its network to 320 by 2020. A densification of the network has occurred concurrently with the expansion of the number of fulfillment centers. One sign that this density is increasing is the rise in the number of states that have at least one active cluster; in 1999, there were only four of these states, but in 2018, there were thirty (Houde et al. (2017)).

SOURCE: Author's Own Calculation

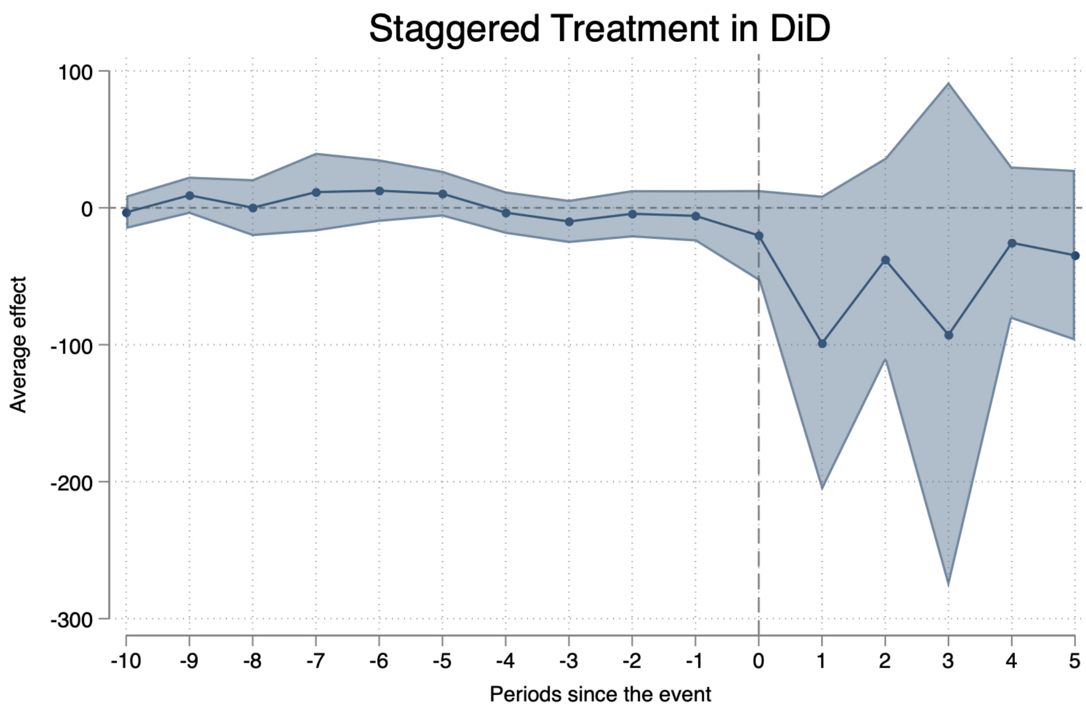
Figure 3.2 — Incremental share of amazon logistics in handling their deliveries over time



SOURCE: Author's own estimation

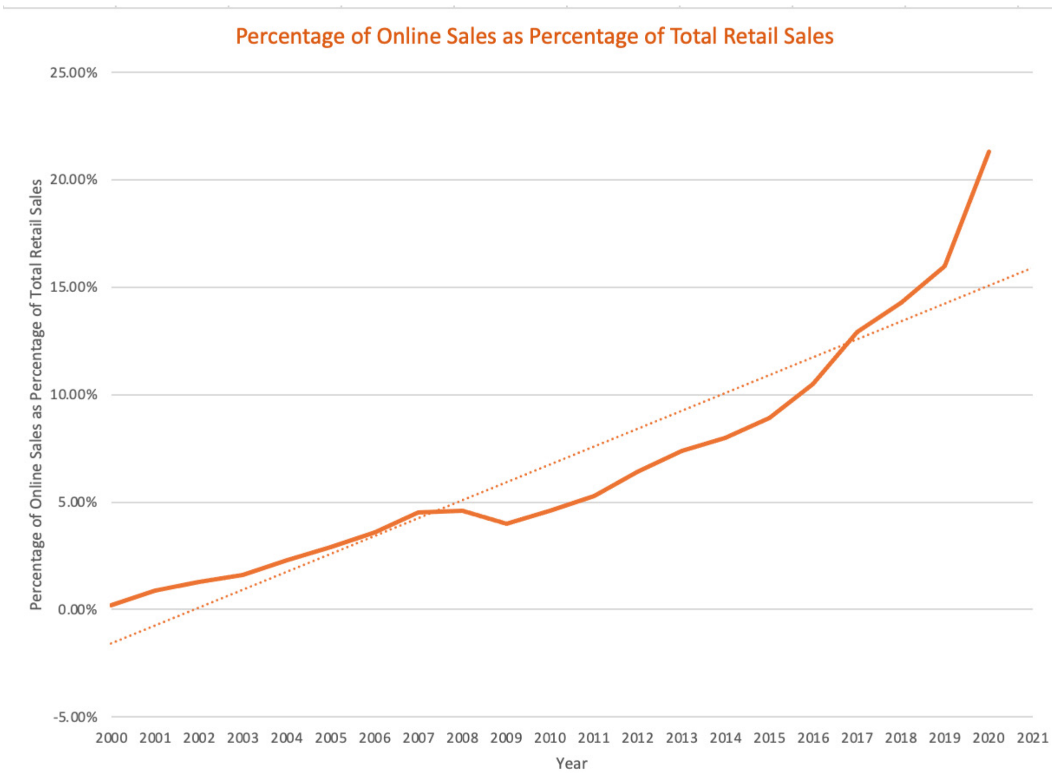
Note: Here, it shows how amazon evolved over time to take over its' own logistics from UPS and USPS. Compared to the 1.9 billion parcels it delivered in 2019, Amazon Logistics delivered almost 4.2 billion packages in 2020. It is now responsible for 21% of all parcel deliveries in the United States, placing it third in terms of volume, behind the United States Postal Service (38%) and UPS (24%), but ahead of FedEx for the first time (16%). These numbers put into perspective what many people in the business world have suspected for some time now: that Amazon has become one of the largest delivery forces in the United States.

Figure 3.3 — Dynamic Treatment Effect of E-commerce diffusion on innovation



Note: This figure presents average effects for each event time in Callaway and Sant'Anna (2021). Here, the outcome of interest is total patents filed in different counties over the years from 2000-2021. Callaway and Sant'Anna (2021) is estimated without controls. This figure suggests how with the diffusion of e-commerce at county level has had any impact on the innovation of incorporated and non-incorporated firms. Innovation is proxied here by the total number of patents filed at county level each year ranging from 2000-2021.

Figure 3.4 — Onlines sales as percentage of total retail sales in USA

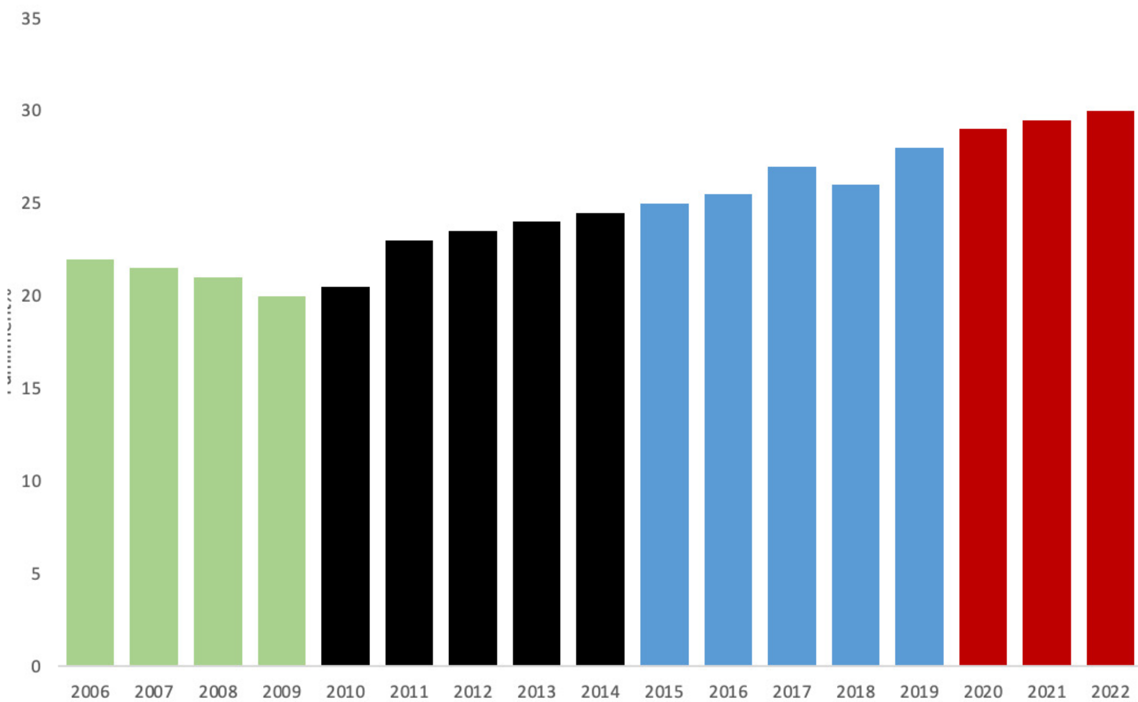


Note: According to the provided results, the percentage of online sales as a percentage of total retail sales has been increasing steadily over the years. In 2000, the percentage was only 0.20%, while in 2020, it reached 21.30%, indicating a significant increase in online sales over the past two decades. Additionally, in 2022, e-commerce retail trade sales in the United States were worth over one trillion U.S. dollars, up from 870.78 billion U.S. dollars in 2021. [a] Overall, the data suggests that online sales are becoming an increasingly important part of total retail sales in the United States, and this trend is expected to continue in the future.

SOURCE: FRED EConomic Data

a. https://www.statista.com/statistics/185283/total-and-e-commerce-us-retail-trade-sales-since-2000/

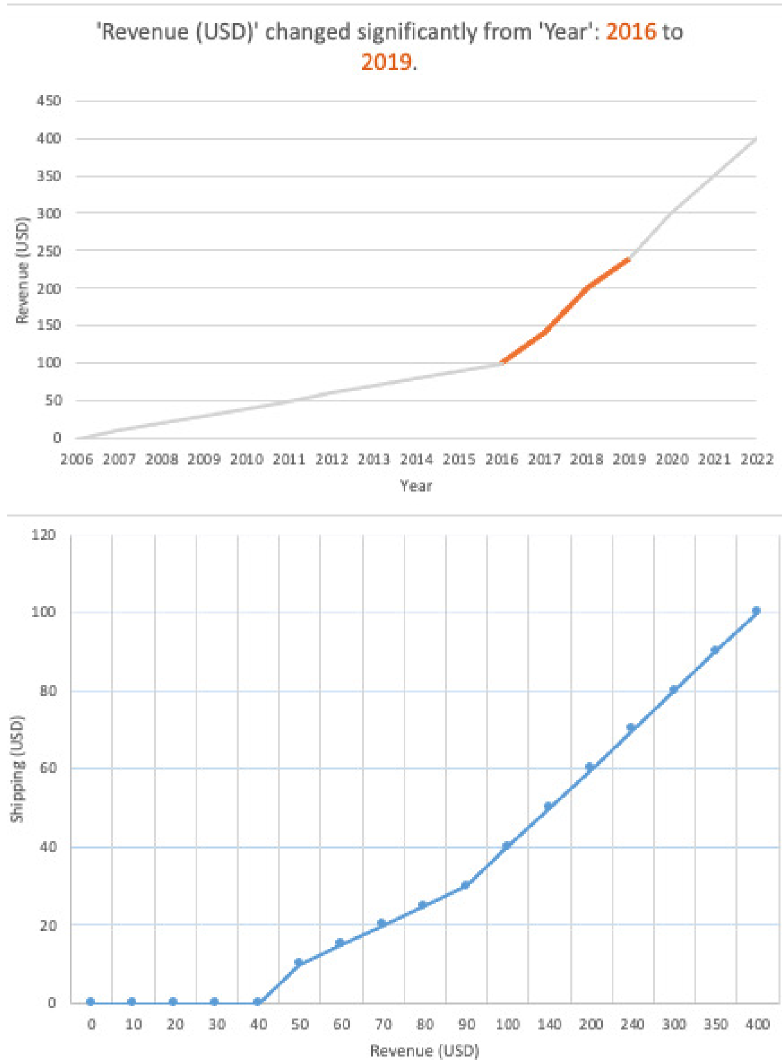
Figure 3.5 — Incremental share of amazon logistics as expenses as percentage of the totl revenue generated by amazon sales



SOURCE: Amazon’s revenue report and Statista

Note: The provided data represents Amazon’s revenue, shipping costs, and percentage of fulfillment center costs as a percentage of their total generated revenue from sales from 2006 to 2022. The percentage of fulfillment center costs as a percentage of their total generated revenue from sales increased from 22% in 2006 to 30% in 2022. The rise in increasing costs for Amazon has led to the company’s investment in automation, which can improve efficiency, reduce labor costs, and handle increased demand, ultimately leading to cost savings over time

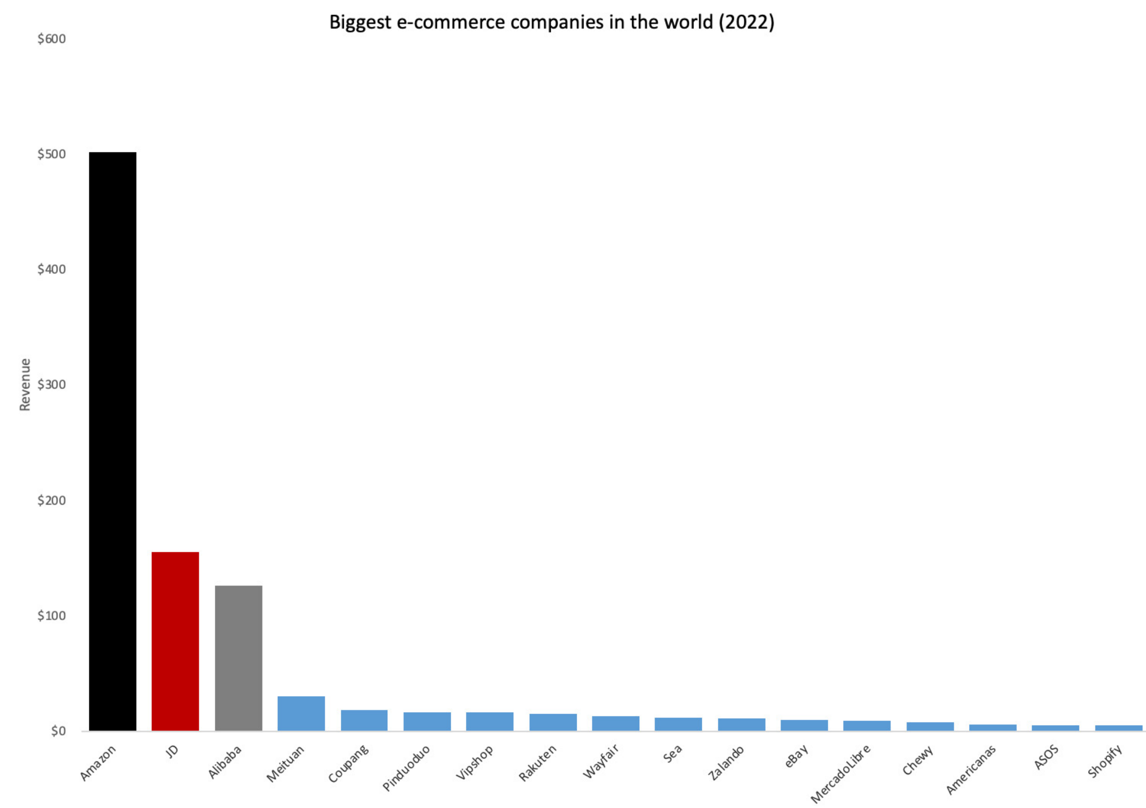
Figure 3.6 — Amazon's Revenue Over Time



Note: Amazon's revenue report and Statista.

The revenue generated by Amazon through sales increased from 0 USD in 2006 to 400 USD in 2022. Similarly, the shipping cost, which refers to the cost of shipping goods to customers, also increased from 0 USD in 2006 to 100 USD in 2022. This data suggests that the revenue generated by Amazon has been increasing steadily over the years, and so has the cost of shipping. The increase in shipping costs could be due to the growing demand for online shopping and the subsequent increase in the number of orders shipped. One way Amazon has been working to reduce shipping costs is by investing in building their distribution network through Amazon Fulfillment Centers (AFCs). By investing in their distribution network and using advanced technologies, Amazon can not only reduce their shipping costs but also offer faster and more reliable delivery to their customers. This helps them to maintain their competitive advantage in the market and enhance customer satisfaction.

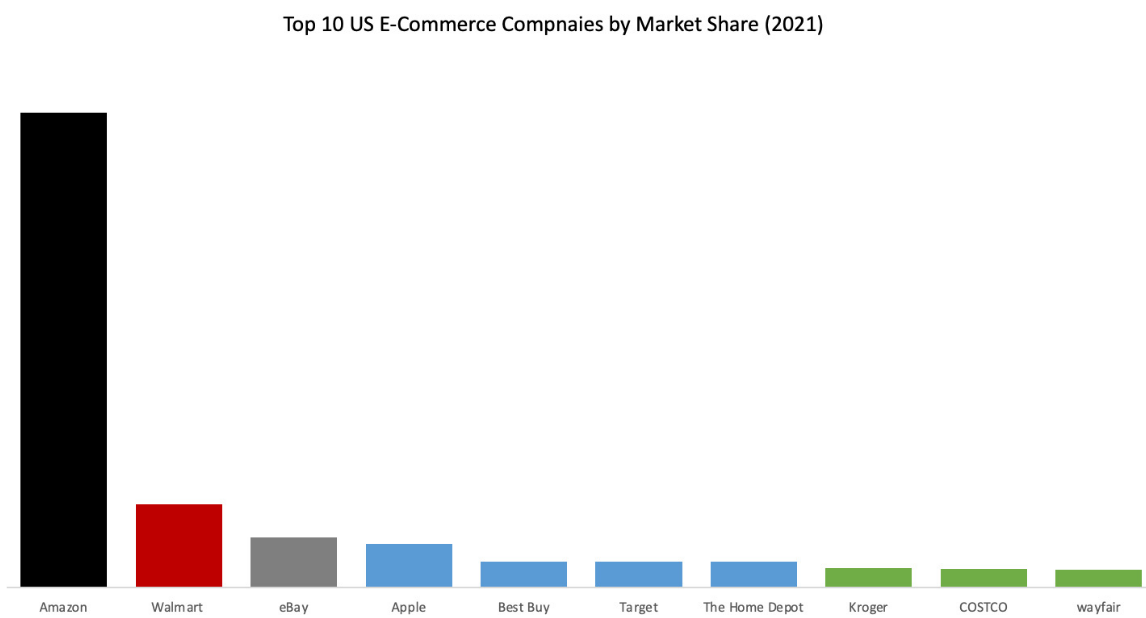
Figure 3.7 — Biggest E-commerce Companies in the World



Note: Data from CompaniesMarketcap.com

Amazon stands out as the clear leader in terms of revenue, earning a substantial $502 billion over the four quarters. This highlights Amazon's dominant position in the global e-commerce landscape. JD and Alibaba follow Amazon, although at a considerable distance, with revenues of $155 billion and $126 billion, respectively. Chewy, Americanas, ASOS, Shopify, and other companies listed exhibit relatively lower revenues ranging from $7.80 billion to $5.20 billion.

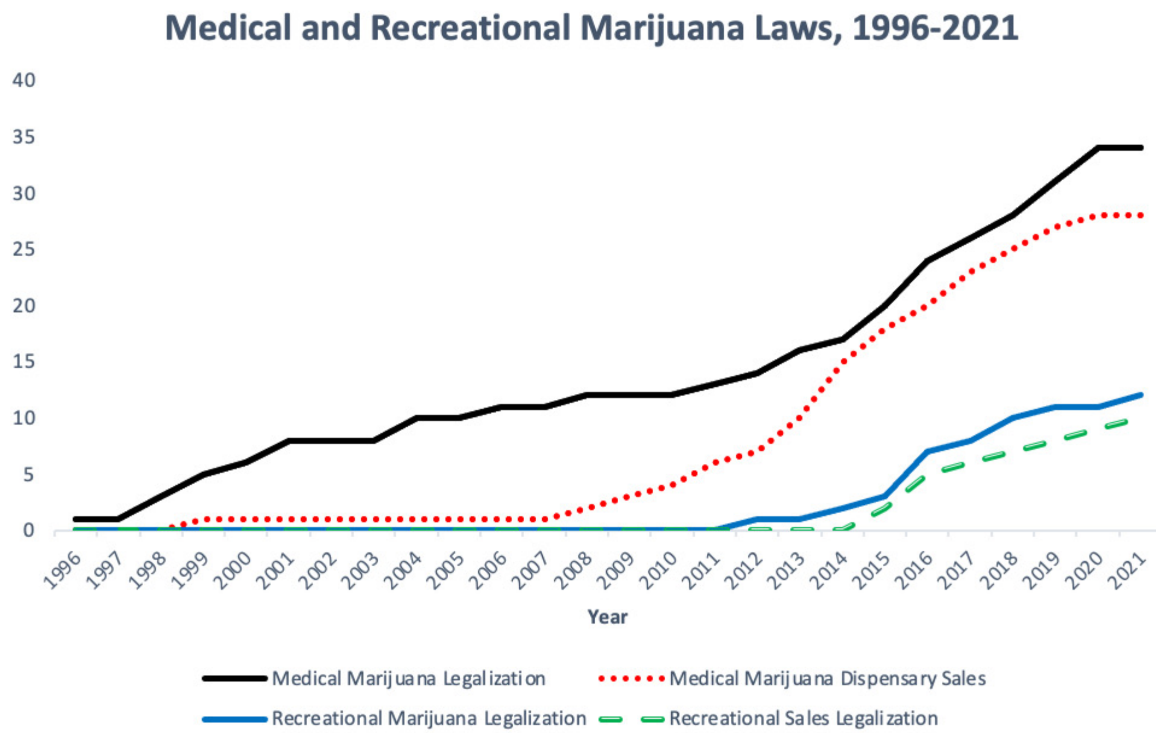
Figure 3.8 — Amazon is Undisputed Leader of E-Commerce Sales in the US



Note: Data from EMarketer

Amazon maintains a significant lead in market share with 40.4%. This reinforces Amazon's position as the market leader and exemplifies its unparalleled dominance in the e-commerce industry. While Walmart, with a market share of 7.1%, holds the second-largest position, it is still significantly behind Amazon. However, its well-established presence in the retail sector allows Walmart to maintain a substantial share of the e-commerce market. The presence of renowned retailers like Apple, Best Buy, and Target, alongside specialized platforms like Wayfair, adds diversity to the market and offers consumers a range of choices.

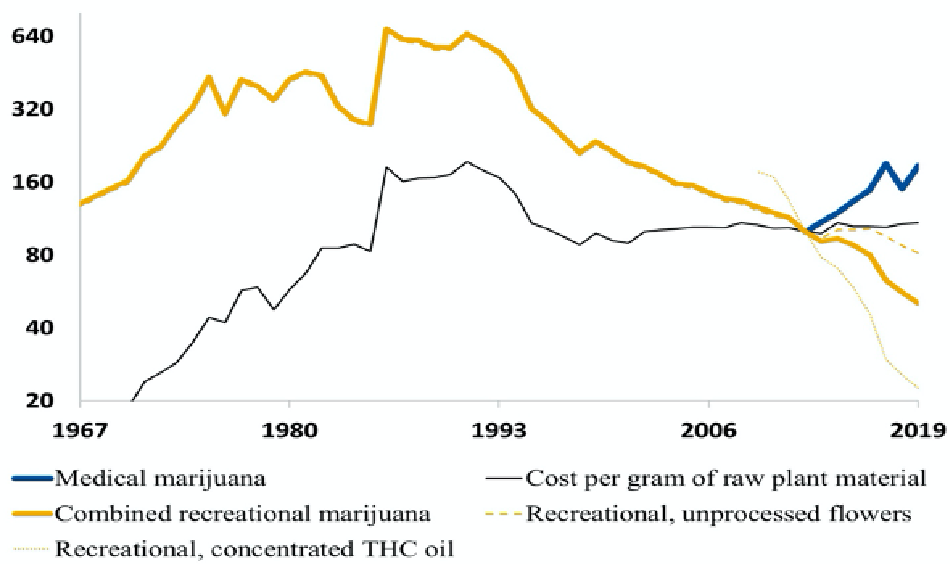
Figure 3.9 — Number of states legalization of MML and RML



NOTE: The data demonstrates the evolving landscape of marijuana legislation in the United States over the past two decades. The number of states that have legalized medical marijuana gradually increased over the years. It started with 1 state in 1996 and reached 34 states by 2021. Recreational Marijuana Legalization indicates the number of states that have legalized the recreational use of marijuana. This category initially had no legalization until 2012 when one state approved it. By 2021, a total of 12. Recreational Sales Legalization started at 0 and grew to 10 states by 2021.

SOURCE: Data is collected from Anderson and Rees (2023)

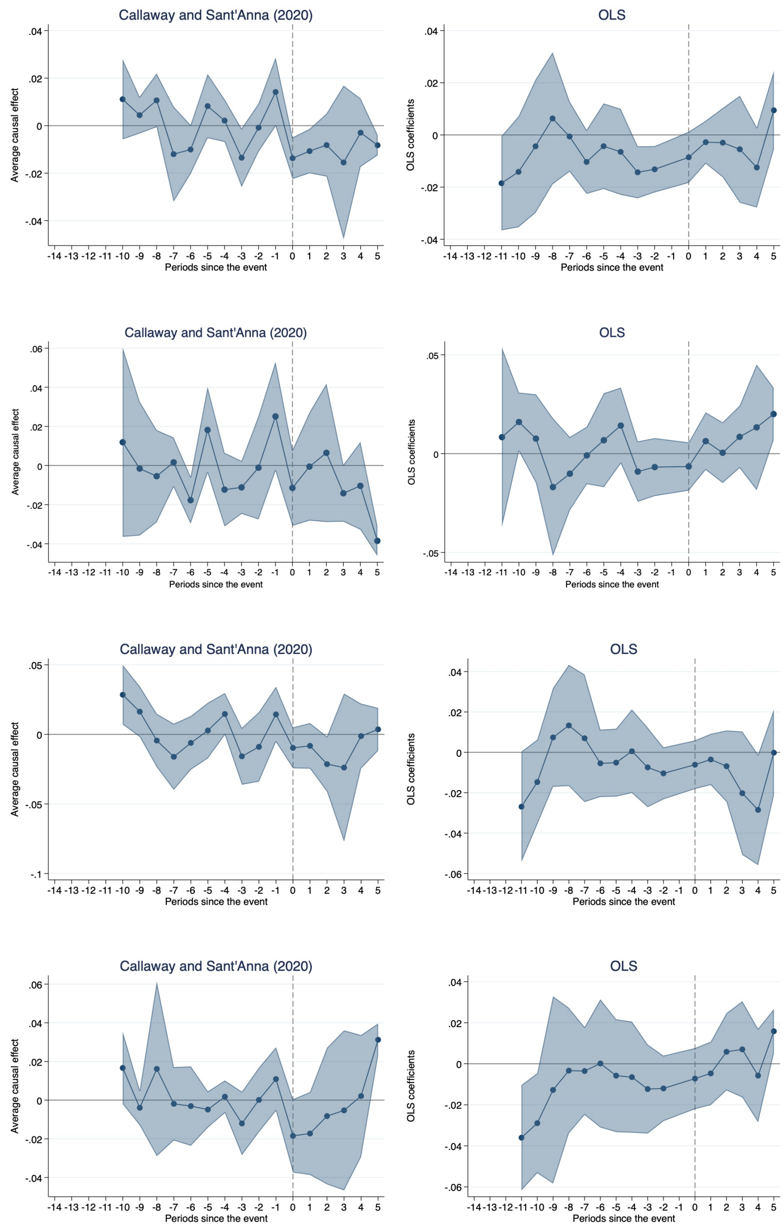
Figure 3.10 — Price of Marijuana over the years



NOTE: It becomes evident that the cost of cannabis for medical and adult use has steadily declined. This decline is observed on both a monthly and annual basis, suggesting a sustained trend rather than a temporary fluctuation. This inference is supported by the fact that the cost of recreational marijuana has experienced a considerably lower increase when compared to the cost of raw plant material This trend of decreasing prices for marijuana is noteworthy as it impacts both the medical and recreational consumption across the United States.

SOURCE: Soloveichi (2021)

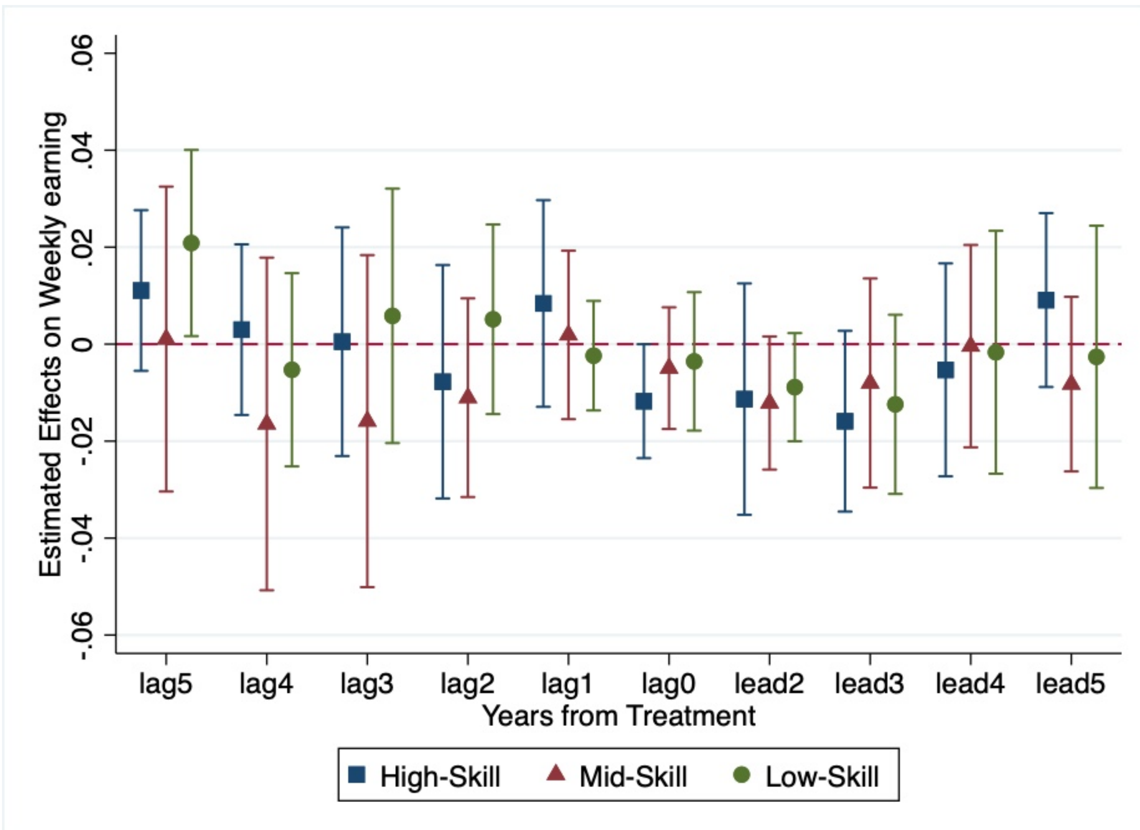
Figure 3.11 — Comparing event study between Callaway and Sant'Anna and TWFE OLS



NOTE: Here, I collapsed the data from repeated cross section to balanced panel at state and year level of 663 observations for pooled sample , high skill workers sample , mid skill sample and low-skill sample separately. Each sample is shown respectively in panel a , panel b , panel c and panel d. The change in weekly earning is illustrated for each sample and it is clear how biased TWFE is compared to Callaway and Sant'Anna with staggered timing and heterogeneous treatment

SOURCE: CPS Basic Monthly Sample 2009-2021

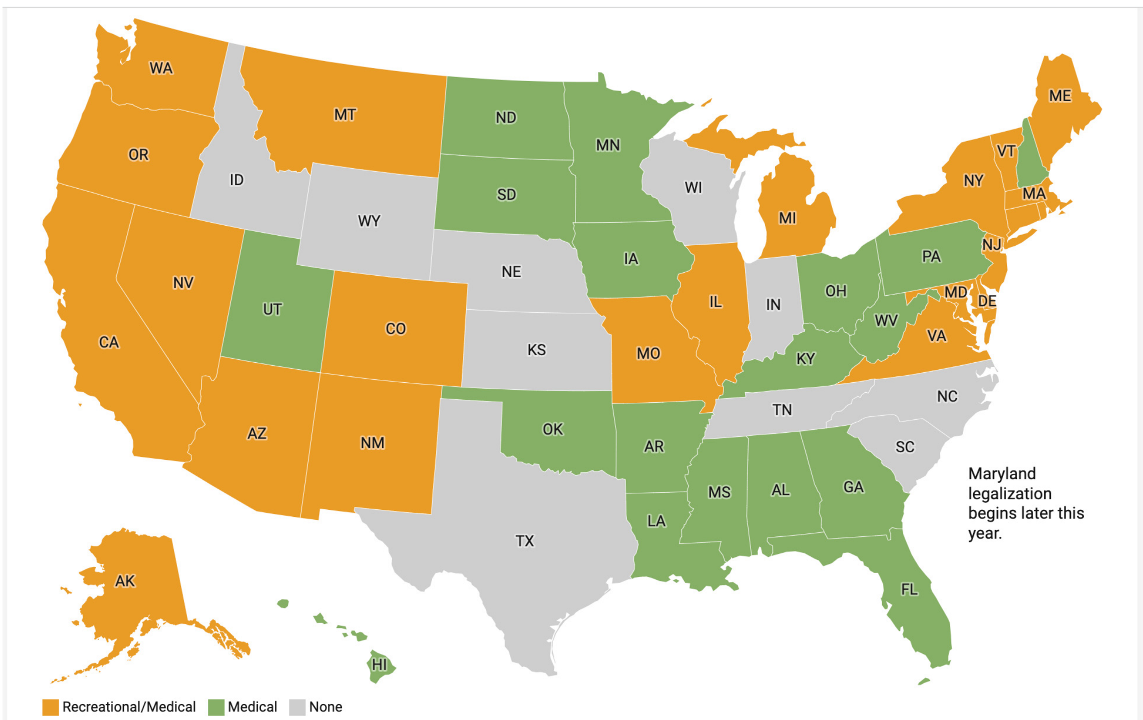
Figure 3.12 — Event study for three type of skilled workers using EVENTDD



NOTE: Here, I collapsed the data from repeated cross section to balanced panel at state and year level of 663 observations for high skill workers sample , mid skill sample and low-skill sample separately. The change in weekly earning after RML is illustrated for each sample using unweighted event study estimator eventdd

SOURCE: CPS Basic Monthly Sample 2009-2021

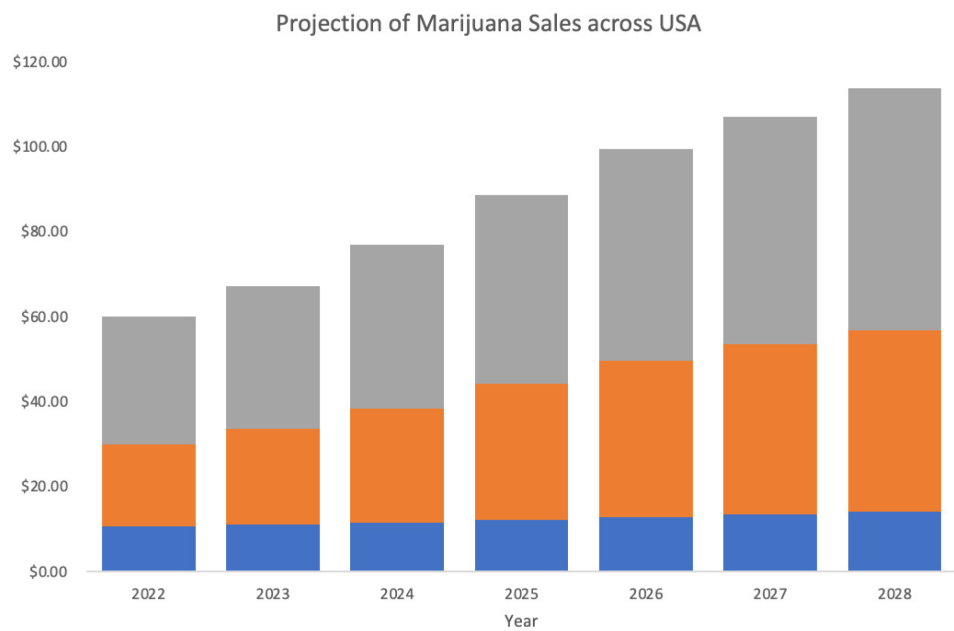
Figure 3.13 — Map of states legaization of MML and RML



NOTE: California holds the distinction of being the pioneering state to enact legislation legalizing the medical utilization of marijuana in 1996. Subsequently, a notable progression has taken place, leading to the legalization of medical cannabis employment in 40 states, alongside the District of Columbia. Moreover, with regard to the recreational or adult use of cannabis, the jurisdictions of Washington, D.C., along with 22 states, have granted their approval for its lawful implementation.

SOURCE: MJBiz Factbook

Figure 3.14 — Sales projection of medical and recreational marijuana across USA



NOTE: Combined U.S. medical and recreational cannabis sales are expected to achieve a substantial milestone, reaching $33.6 billion by the conclusion of 2023. The projected sales figures indicate a substantial expansion of the cannabis market. The estimated sales of both medical and recreational cannabis demonstrate a positive outlook for the industry. The opening of new adult-use markets is identified as a key driver behind the projected increase in cannabis sales. This suggests that the expanding recreational segment is contributing significantly to the overall growth of the industry. The accelerated transition from medical to recreational markets indicates a shift in consumer preferences and market dynamics. This transition reflects the evolving landscape of the cannabis industry, where recreational use is gaining prominence and becoming increasingly significant.

SOURCE: Data is collected from MJBiz Factbook

Bibliography

Alberto Abadie and Jérémy L'Hour. A penalized synthetic control estimator for disaggregated data. *Journal of the American Statistical Association*, 116(536):1817–1834, 2021.

Alberto Abadie, Alexis Diamond, and Jens Hainmueller. Synth: An r package for synthetic control methods in comparative case studies. *Journal of Statistical Software*, 42(13), 2011.

Rahi Abouk, Keshar M Ghimire, Johanna Catherine Maclean, and David Powell. Does marijuana legalization affect work capacity? evidence from workers' compensation benefits. Technical report, National Bureau of Economic Research, 2021.

Daron Acemoglu and Jonas Loebbing. Automation and polarization. Technical report, National Bureau of Economic Research, 2022.

D Mark Anderson and Daniel I Rees. The public health effects of legalizing marijuana. *Journal of Economic Literature*, 61(1):86–143, 2023.

Joshua D Angrist and Guido W Imbens. Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American statistical Association*, 90(430):431–442, 1995.

Peter Arcidiacono, Paul B Ellickson, Carl F Mela, and John D Singleton. The competitive effects of entry: Evidence from supercenter expansion. *American Economic Journal: Applied Economics*, 12(3):175–206, 2020.

Dmitry Arkhangelsky, Susan Athey, David A Hirshberg, Guido W Imbens, and Stefan Wager. Synthetic difference-in-differences. *American Economic Review*, 111 (12):4088–4118, 2021.

David Autor. Work of the past, work of the future. Technical report, National Bureau of Economic Research, 2019.

David Autor, David Dorn, Lawrence F Katz, Christina Patterson, and John Van Reenen. The fall of the labor share and the rise of superstar firms. *The Quarterly Journal of Economics*, 135(2):645–709, 2020.

Marc Bain. Prime Has Never Been More Important to Amazon. *Quartz*, May 3 2021. [Online]. Available: https://qz.com/2006443/prime-has-never-been-more-important-to-amazon/.

Andrew C Baker, David F Larcker, and Charles CY Wang. How much should we trust staggered difference-in-differences estimates? *Journal of Financial Economics*, 144(2):370–395, 2022.

Abhijit V Banerjee, Rema Hanna, Gabriel E Kreindler, and Benjamin A Olken. Debunking the stereotype of the lazy welfare recipient: Evidence from cash transfer programs. *The World Bank Research Observer*, 32(2):155–184, 2017.

Bruce Barcott and Bruce Whitney. The us cannabis industry now supports 428,059 jobs perspectives on facility location. 2022. [Online]. Available: https://www.leafly.com/news/industry/cannabis-jobs-report.

Emek Basker. Job creation or destruction? labor market effects of wal-mart expansion. *Review of Economics and Statistics*, 87(1):174–183, 2005.

Emek Basker. The causes and consequences of wal-mart's growth. *Journal of Economic Perspectives*, 21(3):177–198, 2007.

Brian Baugh, Itzhak Ben-David, and Hoonsuk Park. Can taxes shape an industry? evidence from the implementation of the "amazon tax". *The Journal of Finance*, 73 (4):1819–1855, 2018.

Bezos. 2015 annual report," amazon; "letter to shareholders,", 2006. [Online]. Available: https://s2.q4cdn.com/299287126/files/doc_financials/annual/2015-Letter-to-Shareholders.PDF.

Abha Bhattarai. Amazon announces it will boost minimum wage for all workers after facing criticism. *The Washington Post*, October 2018. Accessed: May 10, 2023.

Kenneth Blum, David Han, Edward J Modestino, Scott Saunders, A Kennison Roy III, W Jacobs, Darryl S Inaba, David Baron, Marlene Oscar-Berman, Mary Hauser, et al. A systematic, intensive statistical investigation of data from the comprehensive analysis of reported drugs (card) for compliance and illicit opioid abstinence in substance addiction treatment with buprenorphine/naloxone. *Substance use & misuse*, 53(2):220–229, 2018.

Ingo Borchert, Charles Garroway, Isabel Bessudo, Leandro Coutinho, and Ana Gallegos. The emerging markets e-commerce opportunity. *Brookings Institution*, March 2021. Accessed: May 10, 2023.

Kirill Borusyak, Xavier Jaravel, and Jann Spiess. Revisiting event study designs: Robust and efficient estimation. *arXiv preprint arXiv:2108.12419*, 2021.

Lisa Bowman. Legally high, legally fired for pot use. *Scripps News*, March 24 2014.

URL http://archive.naplesnews.com/news/national/legally-high-yet-legally-fired-for-pot-use-ep-459587537-34135999.Lhtml/.

Richard Box. Safely through the haze of legalized marijuana. April 8 2015. URL https://www.linkedin.com/pulse/safety-through-haze-legalized-marijuana-rboert-box.

]bls2022 Bureau of Labor Statistics. Bureau of Labor Statistics, [2022]. Available at https://www.bls.gov/.

Brantly Callaway and Pedro HC Sant'Anna. Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2):200–230, 2021.

Census Bureau. US County Shapefiles, 2023. [Online]. Available: https://public.opendatasoft.com/explore/dataset/us-county-boundaries/information/.

Sudheer Chava, Alexander Oettl, Manpreet Singh, and Linghang Zeng. Impact of e-commerce on employees at brick-and-mortar retailers. *Georgia Tech Scheller College of Business Research Paper*, (18-23), 2018.

Stanford Chihuri and Guohua Li. State marijuana laws and opioid overdose mortality. *Injury epidemiology*, 6(1):1–12, 2019.

Hyunbae Chun, Hailey Hayeon Joo, Jisoo Kang, Yoonsoo Lee, et al. Diffusion of e-commerce and retail job apocalypse: Evidence from credit card data on online spending. Technical report, Center for Economic Institutions, Institute of Economic Research . . . , 2020.

Civic Economics. Amazon and Empty Storefronts, 2015. [Online]. Available: https://cameonetwork.org/resource/amazon-and-empty-storefronts/.

Mitchell Clark. Walmart expands drone delivery service to six states in partnership with droneup. *The Verge*, May 2022. Accessed: May 10, 2023.

Damian Clarke and Kathya Tapia-Schythe. Implementing the panel event study. *The Stata Journal*, 21(4):853–884, 2021.

Guido Matias Cortes, Nir Jaimovich, Christopher J Nekarda, and Henry E Siu. The dynamics of disappearing routine jobs: A flows approach. *Labour Economics*, 65: 101823, 2020.

Tyler Cowen. *The great stagnation: How America ate all the low-hanging fruit of modern history, got sick, and will (eventually) feel better: A Penguin eSpecial from Dutton*. Penguin, 2011.

Laetitia Dablanc and Dina Rakotonarivo. The impacts of logistics sprawl: How does the location of parcel transport terminals affect the energy efficiency of goods' movements in paris and what can we do about it? *Procedia-Social and Behavioral Sciences*, 2(3):6087–6096, 2010.

H David and David Dorn. The growth of low-skill service jobs and the polarization of the us labor market. *American economic review*, 103(5):1553–97, 2013.

Clément De Chaisemartin and Xavier d'Haultfoeuille. Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, 110(9): 2964–96, 2020.

Clément de Chaisemartin and Xavier D'Haultfoeuille. Two-way fixed effects and differences-in-differences with heterogeneous treatment effects: A survey. Technical report, National Bureau of Economic Research, 2022.

Ellora Derenoncourt, Clemens Noelke, David Weil, and Bledi Taska. Spillover effects from voluntary employer minimum wages. Technical report, National Bureau of Economic Research, 2021.

Jeff DeSimone. Illegal drug use and employment. *Journal of Labor Economics*, 20(4): 952–977, 2002.

Angela Dills, Sarah Goffard, Jeffrey Miron, and Emily Partin. The effect of state marijuana legalization: 2021 update, 2021.

Arindrajit Dube, T William Lester, and Barry Eidlin. Firm entry and wages: impact of wal-mart growth on earnings throughout the retail sector. *Institute of Industrial Relations Working Paper No. iirwps-126-05*, 2007.

Fabian Eckert, Teresa C Fort, Peter K Schott, and Natalie J Yang. Imputing missing values in the us census bureau's county business patterns. Technical report, National Bureau of Economic Research, 2020a.

Florian Eckert, Shriram Ganapati, and Chris Walsh. Skilled Scalable Services: The New Urban Bias in Economic Growth, 2020b. Available at SSRN 3439118.

Paul B Ellickson and Paul LE Grieco. Wal-mart and the geography of grocery retailing. *Journal of Urban Economics*, 75:1–14, 2013.

Teressa L Elliott, Lee Kersting, and Robert Salver. The legalization of marijuana: Tax issues and costs to employers. *Journal of Management Policy and Practice*, 20(1): 15–25, 2019.

Robert W Fairlie and Frank M Fossen. Defining opportunity versus necessity entrepreneurship: Two components of business creation. In *Change at home, in the labor market, and on the job*. Emerald Publishing Limited, 2020.

Federal Aviation Administration. U.S. Department of Transportation, Federal Aviation Administration. [Online]. Available: https://www.faa.gov/airports/planning_capacity/passenger_allcargo_stats/.

Federal Highway Administration. US Primary Roads Shapefiles, 2023. [Online]. Available: https://catalog.data.gov/dataset/tiger-line-shapefile-2016-nationu-s-primary-roads-national-shapefile/resource/d983eeef-21d9-4367-9d42-27a131ee72b8.

Michael R Frone. Predictors of overall and on-the-job substance use among young workers. *Journal of occupational health psychology*, 8(1):39, 2003.

Evelina Gavrilova, Takuma Kamada, and Floris Zoutman. Is legal pot crippling mexican drug trafficking organisations? the effect of medical marijuana laws on us crime. *The Economic Journal*, 129(617):375–407, 2019.

Maris Goldmanis, Ali Hortaçsu, Chad Syverson, and Önsel Emre. E-commerce and the market structure of retail industries. *The Economic Journal*, 120(545):651–682, 2010.

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

Robert J Gordon. The rise and fall of american growth. In *The Rise and Fall of American Growth*. Princeton University Press, 2016.

Jorge Guzman and Scott Stern. The state of american entrepreneurship: New estimates of the quantity and quality of entrepreneurship for 32 us states, 1988–2014. *American Economic Journal: Economic Policy*, 12(4):212–43, 2020.

Howard V Hayghe. Anti-drug programs in the workplace: Are they here to stay. *Monthly Lab. Rev.*, 114:26, 1991.

Jean-François Houde, Peter Newberry, and Katja Seim. Economies of density in e-commerce: A study of amazon's fulfillment center network. *NBER working paper*, (w23361), 2017.

Paul Jakubicek and Clarence Woudsma. Proximity, land, labor and planning? logistics industry perspectives on facility location. *Transportation Letters*, 3(3): 161–173, 2011.

Panle Jia. What happens when wal-mart comes to town: An empirical analysis of the discount retailing industry. *Econometrica*, 76(6):1263–1316, 2008.

Janelle Jones and Ben Zipperer. Unfulfilled promises: Amazon fulfillment centers do not generate broad-based employment growth. 2018.

Sanggyun Kang. *Unraveling Decentralization of Warehousing and Distribution Centers: Three Essays*. PhD thesis, University of Southern California, 2017.

Frank Kleibergen and Richard Paap. Generalized reduced rank tests using the singular value decomposition. *Journal of econometrics*, 133(1):97–126, 2006.

J Paul Leigh. Economic burden of occupational injury and illness in the united states. *The Milbank Quarterly*, 89(4):728–772, 2011.

David M Levinson and Kevin J Krizek. *The End of Traffic and the Future of Access: A Roadmap to the New Transport Landscape*. 2017.

R Edward Minchin Jr, Charles R Glagola, Kelu Guo, and Jennifer L Languell. Case for drug testing of construction workers. *Journal of Management in Engineering*, 22 (1):43–50, 2006.

Stacy Mitchell. Amazon's toll road. 2021.

Stacy Mitchell, Ron Knox, and Zach Freed. Amazon's monopoly tollbooth. *Institute for Local Self-Reliance. URL: https://ilsr. org/amazons_tollbooth/, Zugriff*, 19:2021, 2020.

MWVPL. Geography of Amazon Fulfillment Centers. [Online]. Available: https://www.mwpvl.com/html/amazon_com.html.

National Cancer Institute. Seer county demographics data, 2000-2021. https://seer.cancer.gov/popdata/download.html. Accessed: May 10, 2023.

NBER. NBER US County Distance Database. [Online]. Available: https://data.gov/counties/.

David Neumark, Junfu Zhang, and Stephen Ciccarella. The effects of wal-mart on local labor markets. *Journal of Urban Economics*, 63(2):405–430, 2008.

Lauren Hersch Nicholas and Johanna Catherine Maclean. The effect of medical marijuana laws on the health and labor supply of older adults: Evidence from the health and retirement study. *Journal of Policy Analysis and Management*, 38(2): 455–480, 2019.

Rosalie Liccardo Pacula, David Powell, Paul Heaton, and Eric L Sevigny. Assessing the effects of medical marijuana laws on marijuana and alcohol use: The devil is in the details. Technical report, National Bureau of Economic Research, 2013.

Annie Palmer. Amazon poised to pass UPS and FedEx to become largest U.S. delivery service, 2022. [Online]. Available: https://www.cnbc.com/2021/11/29/amazon-on-track-to-be-largest-us-delivery-service-by-2022-exec-says.html.

Srikanth Paruchuri, Joel AC Baum, and David Potere. The wal-mart effect: Wave of destruction or creative destruction? *Economic Geography*, 85(2):209–236, 2009.

D Keith Robbins, Louis J Pantuosco, Darrell F Parker, and Barbara K Fuller. An ampirical assessment of the contribution of small business employment to us state economic performance. *Small Business Economics*, 15(4):293–302, 2000.

Joseph J Sabia and Thanh Tam Nguyen. The effect of medical marijuana laws on labor market outcomes. *The Journal of Law and Economics*, 61(3):361–396, 2018.

Joseph J Sabia, Dhaval M Dave, Fawaz Alotaibi, and Daniel I Rees. Is recreational marijuana a gateway to harder drug use and crime? Technical report, National Bureau of Economic Research, 2021.

Edward Shepard and Thomas Clifton. Drug testing and labor productivity: Estimates applying a production function model. In *PROCEEDINGS OF THE ANNUAL MEETING-INDUSTRIAL RELATIONS RESEARCH ASSOCIATION*, volume 1, pages 372–372. INDUSTRIAL RELATIONS RESEARCH ASSOCIATION, 1998.

Abhay Singh and Sonia Murteza. Last mile delivery market, 2022. [Online]. Available: https://www.alliedmarketresearch.com/last-mile-delivery-market.

Rachel Soloveichi. *Tracking Marijuana in the National Accounts*. US Department of Commerce, Bureau of Economic Analysis, 2021.

TJ Stallbaumer. Working as fulfillment centers, walmart stores are the star of the last mile. https://corporate.walmart.com/newsroom/2022/02/28/working-as-fulfillment-centers-walmart-stores-are-the-star-of-the-last-mile, February 2022. Accessed: May 10, 2023.

Brad Stone. *Amazon unbound: Jeff Bezos and the invention of a global empire*. Simon and Schuster, 2022.

Nadine Strossen. Keeping the constitution inside the schoolhouse gate-students' rights thirty years after tinker v. des moines independent community school district. *Drake L. Rev.*, 48:445, 1999.

Liyang Sun and Sarah Abraham. Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, 225(2): 175–199, 2021.

Ken D Tunnell. *Pissing on demand: Workplace drug testing and the rise of the detox industry*, volume 18. NYU Press, 2004.

Darin F Ullman. The effect of medical marijuana on sickness absence. *Health economics*, 26(10):1322–1327, 2017.

U.S. Bureau of Labor Statistics. Current employment statistics highlights. https://www.bls.gov/web/empsit/ceshighlights.pdf. Accessed: May 10, 2023.

US Census Bureau. U.S. Economic Census, 1997–2012. [Online]. Available: https://www.census.gov/programs-surveys/economic-census/year.html.

U.S. Patent and Trademark. U.S. Patent and Trademark Data, 2000-2021. [Online]. Available: https://www.uspto.gov/patents.

Jan C Van Ours. Cannabis use when it's legal. *Addictive behaviors*, 32(7):1441–1450, 2007.

Jan C Van Ours, Jenny Williams, David Fergusson, and L John Horwood. Cannabis use and suicidal ideation. *Journal of Health Economics*, 32(3):524–537, 2013.

Nora D Volkow, Ruben D Baler, Wilson M Compton, and Susan RB Weiss. Adverse health effects of marijuana use. *New England Journal of Medicine*, 370(23):2219–2227, 2014.

David Waite. The economic significance of small firms. *The Journal of Industrial Economics*, pages 154–166, 1973.