E-commerce and Regional Inequality: A Trade Framework and Evidence from Amazon's Expansion*

Elmer Zongyang Li†

October, 2024

Click here for the latest version

Abstract

E-commerce exposes consumers to a broader set of goods and retailers, and online retailing, particularly through platforms like Amazon, is notably mobile in space. This paper investigates the spatial general equilibrium and redistribution effects of e-commerce on different local labor markets, focusing on the production technology changes in the retail sector. By examining a panel of the universe of products and retailers on Amazon, I find that online retailers tend to agglomerate, especially around Amazon's distribution and fulfillment centers as well as transportation hubs, and this agglomeration alters the trade flows of upstream goods. Integrating consumer search and retailer location choices into a multi-sector gravity trade model with an elastic supply of heterogeneous workers, the model predicts that increases in online shopping efficiency, online retailer agglomeration, and reduced shipping friction will drive greater industrial and occupational specialization. Quantitative analysis reveals that the expansion of Amazon from 2007 to 2017 led to an average increase in total welfare of 6.7 percent across states, primarily driven by price effects, but a negative income effect for Midwestern states. The shift also resulted in the reallocation of workers from manufacturing and brick-and-mortar towards the online retail sector, reducing overall non-employment by 0.5 percentage points, especially in states with a comparative advantage in online retail. However, the Gini index on non-employment across regions rose by 20 percent, indicating heightened regional inequalities. Counterfactual redistribution of regional trade surpluses and intervention in online market design improve spatial efficiency.

Keywords: e-commerce, trade, inequality, agglomeration
JEL Codes: F11, F16, F17, L81, O18, R11, R13

*I thank Philipp Kircher, Michael Lovenheim, and Julieta Caunedo for their guidance and continuous support. Thanks to Mathieu Taschereau-Dumouchel and Ryan Chahrour for their suggestions and support. I also want to thank Mario Crucini, Klaus Desmet, Russ Hillberry, Xian Jiang, David Kuenzel, Yuhei Miyauchi, Ezra Oberfield, and Yoto Yotov for their inspiring comments. Also thanks to numerous seminar participants. All errors are my own.

†Li: Department of Economics, Cornell University, 109 Tower Road, 404 Uris Hall, Ithaca, New York, 14853. E-mail: zl685@cornell.edu. Website: www.elmerli.net.

1 Introduction

As e-commerce is transforming the retail sector, regions across the United States face very different prospects. While a town in New Jersey might see expanding warehouses and manufacturers, another town in Wyoming may mostly suffer from the collapse of local brick-and-motor stores. Studies that examine the impact of e-commerce have noted its impact on the demand, productivity, and markup of physical retail stores (i.e., Goldmanis et al. 2010; Pozzi 2013; Ellison and Ellison 2018), as well as on consumer welfare (Fan et al. 2018; Dolfen et al. 2019). However, little work has thoroughly examined the regional inequality and redistribution effects of e-commerce in terms of economic activities and job opportunities. As the divergence in regional economies has key implications for life outcomes (see Chetty and Hendren 2018; Austin et al. 2018), understanding the consequences of e-commerce on regional inequality is important for policy making.

In this paper, I adopt a trade perspective to study e-commerce's impact on different local labor markets, taking into account trade and input-output linkages and regions' comparative advantages. A key feature of e-commerce is that online retailers don't have to be where the customers are, therefore having more mobility in their location. As in Krugman (1991) and Krugman and Venables (1995), the additional mobility will induce agglomeration in the online retail sector. In an environment where online retailers are the intermediary between the upstream producers and downstream consumers, online retailers would want to locate near the largest consumer or the cheapest producer, but also need to take into account the resulting rise in wages and land prices. The intermediary nature and agglomeration of online retailers will imply greater specialization in both the upstream and the online retail sectors.

Using a comprehensive panel dataset of products and retailers on Amazon, as well as Amazon's fulfillment and distribution facilities, I document four stylized facts that suggest online retailers are more agglomerated in space and their agglomeration is associated with greater trade flows of the upstream goods. First, online retail sales are more spatially concentrated than overall retail sector sales, and are and are less correlated with population and more correlated with manufacturing output; second, online retailers that use Amazon's fulfillment and distribution facilities more agglomerated than those that don't use the facilities; third, destination markets with more online retailers import more wholesale trade goods, whereas origin markets with more online retailers export less wholesale trade goods; fourth, regions near to Amazon's fulfillment and distribution facilities import and export less wholesale trade goods.

Taking these key features of online retailing into account, I build a multi-sector spatial trade framework of intra-regional retailing to analyze e-commerce's impact. The role of e-commerce is first reflected in that consumers have to conduct costly simultaneous search and matching of retailers as in Weitzman (1979), the efficiency of which is subject to the online retail platform. Moreover, I allow online retailers to optimally choose their locations where they import from the upstream sectors and ship to consumers, giving rise to agglomeration incentives. To better understand the impact on employment, I also let workers be heterogeneous in their productivity and optimally choose the sector of employment or to be unemployed. I show that despite the rich micro-foundation, this framework can still aggregate to a gravity trade model with CES demand, with the demand shifter reflecting online match efficiency and the iceberg cost influenced by the shipping cost of online retailers. The location probability of online retailer in a region directly scale up the gravity of trade flows in that region, highlighting the important role of online retailer agglomeration in the model.

I then estimate key fundamentals to take the model to the data, particularly the reduction in shipping friction and the increase in match efficiency related to the rise of e-commerce. I apply the datasets I obtained on Amazon retailers and sales, as well as Amazon facilities to conduct the estimation. The major challenge in identifying the impact of Amazon's expansion concerns its endogeneity to other factors, particularly from the demand side. To overcome this issue, I employ a spatial simulated instrumental variable strategy (Duflo and Pande 2007; Lipscomb et al. 2013; Faber 2014). Instead of using the actual location of Amazon's facilities to calibrate the shock, I build counterfactual distribution centers with the simulated location choices based solely on plausibly exogenous geographic and climatic factors. The shipping cost reduction due to these counterfactual centers is used to instrument the actual decline of shipping frictions and iceberg costs. Conditional on the estimated reduction in iceberg cost, the predicted changes in regional online retail expenditures identify the increase in online match efficiency. My estimation results show that Amazon's growth has led to a 3 percent decline in iceberg cost and a 29 percent increase in online matching efficiency from 2007 to 2017.

Equipped with the estimated shocks and calibrated model parameters, I evaluate Amazon's impact on regional economies in terms of total welfare and employment. My findings suggest that, on average, states see an increase in total welfare of 6.7 percent due to the Amazon shock, primarily driven by price effects. Without the mitigating impact of price declines, total welfare would have increased by 13.1 percent. However, the expansion of Amazon also leads to the reallocation of economic activities and workers, which differentially changes income levels across regions and would have decreased total welfare by 5.4 percent in the absence of compensating price changes. This results in an actual net welfare increase but also reveals significant regional disparities. States on the East and West Coasts and those with a comparative advantage in online retailing, like New York, Massachusetts, Wisconsin, California, and Florida, experience welfare gains due to positive income effects from employment increases and a diversified industrial composition. Conversely, Midwestern states such as Indiana, North Dakota, and Wyoming face welfare losses due to business displacement, despite benefiting from significant positive price effects.

Regarding employment, the Amazon shock has prompted a sectoral shift, particularly moving workers from brick-and-mortar and manufacturing to the growing online retail sector. This reallocation has resulted in a decrease in non-employment by 0.5 percentage points from a base of 38.5 percent in 2007. There is substantial regional variation in this trend, with Midwestern states, which have less comparative advantage in online retailing, experiencing higher shifts towards non-employment and service sectors. The Gini index of non-employment has increased from 0.11 to 0.13, marking a 20 percent growth and indicating a widening gap in employment opportunities across regions.

The widening gaps in economic outcomes due to the rise of e-commerce, as exemplified by Amazon, underscore the need for national-level policy interventions. To address the growing trade imbalances across regions, local governments might consider imposing domestic "tariffs" on non-local goods to recover the first-best allocation, though this could also create welfare losses for consumers (Costinot et al. 2015; Antràs et al. 2022). Given the spatial nature of the market failure, there is a compelling case for a national-level revenue reallocation. Additionally, since the main impact of the e-commerce shock is through matching and shipping frictions, direct government intervention in online retail market design could be warranted. Counterfactual analyses with these policy experiments will be conducted in the next phase.

The rest of the paper is organized as follows. The ensuing section reviews the relevant literature in more detail and highlights this paper's contributions. Section 3 presents the stylized facts on the online retailers and associated trade flow. Section 4 presents the theoretical framework and how to use it to conduct comparative statics and welfare analysis. I discuss model quantification in section 5 and the estimation of Amazon shock. Section 6 shows the results on the impacts of Amazon. In Section 7, I discuss alternative modeling of online retailers' location choices. The last Section concludes.

2 Literature Review

The rise of e-commerce presents a salient case where technology progress redistributes economic opportunities not only across sectors, but also across spaces. This paper propose using a trade framework to study e-commerce, and particularly highlighting the agglomeration of online retailers. It contributes to the literature by applying and extending a standard trade framework to study the spatial general equilibrium effects of e-commerce with new data and identification strategy. Specifically, this paper closely relates to four strands of literature.

Firstly and most relevant, this work builds on the literature studying the market structure of the retail sector and the impact of e-commerce. Two important findings emerge from this literature. For the retail industry, it is found that e-commerce reduces the demand of the physical department stores, raising their productivity but reducing the mark-up in the consumer goods sector (Stanchi 2019;Goldmanis et al. 2010). This supports the modeling of e-commerce as a productivity shock to the retail sector as adopted in this paper. For consumers, Dolfen et al. (2019) finds that e-commerce increases consumer welfare mainly through substituting to online merchants. Fan et al. (2018) shows e-commerce increases domestic trade and benefit consumers in smaller cities and markets particularly. This paper instead studies e-commerce from a general equilibrium spatial trade framework and focuses on its impact on employment and GDP growth differentials across regions. In the welfare analysis I take into account the consumption channel and evaluate the trade-offs.

The theoretical framework of this paper builds on the large literature on of international trade and spatial equilibrium models, and presents a novel application of these theories to study e-commerce. In particular, I adopt the analogy to "globalization" and model e-commerce as a trade shock; for the geographic implications, I apply a Ricardian trade framework focusing on intra-regional and sectoral reallocation taking into account comparative advantages of localities for labor market outcomes (Caliendo et al. 2018; Caliendo et al. 2019; Lee 2020; Adao et al. 2019). Theoretically, I add into a typical Eaton and Kortum (2002) framework with information frictions, transportation cost and worker sorting to more accurately depict the retail sector, as well as roles played by local and federal governments to discuss policy implications. Empirically, I use Amazon's expansion as the source of variation and present new estimation strategy that introduces simulated IV into a typical Bartick estimator.

This paper also relates to studies about the differential impact of technological changes on workers. The earlier discussion in this literature focuses on the wage premium for higher-skill workers, or "skill-biased technological change" (Autor et al. 1998; Acemoglu 2007). It is also found that starting from 1980s, workers conducting "routine" tasks are more likely to be substituted, leading to the polarization of the labor market (Autor et al. 2003; Acemoglu and Autor 2011; Autor and Dorn 2013). This paper contributes to this literature by focusing the spatial nature of a technological change (e-commerce) that has the feature of both an automation and a trade shock, and analyzes under a full general equilibrium spatial trade framework of its impact on workers across sectors and regions.

Lastly, this paper speaks to the literature that examines the differential economic opportunities across spaces. Kline and Moretti (2013) and Amior and Manning (2018) show that there is strong persistence of unemployment and labor force participation differences across regions; Amior and Manning (2018) argues that this is mainly due to the long adjustment to persistent local labor demand shocks. Also relevant is the large literature revealing the importance of neighborhood quality differences on one's life outcomes and hence place-based policies (i.e., Chetty et al. 2016). Here I analyze a particular technology-induced local labor demand shock (e-commerce) that has strong spatial redistributive effects and explores place-based taxes and subsidies. A related literature on the mismatch between workers and jobs found that mismatch across industry and three-digit occupations could explain up to a third of the increase in unemployment (Şahin et al. 2014). This paper investigates a particular cause of mismatch from the labor demand side at the intersection of location, industry and occupation, and explores the tax policy implications.

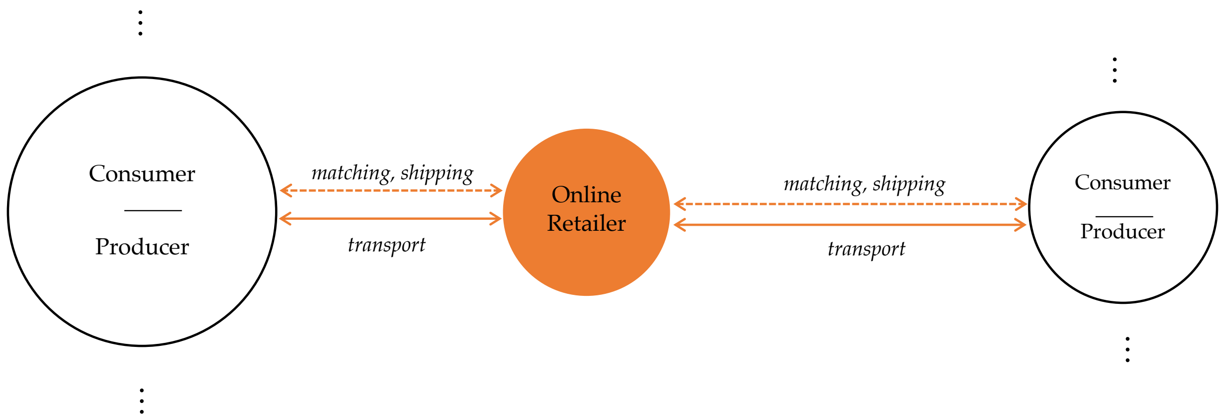
3 Evidence on Online Retail Sales

In this section, I show that the empirical data patterns are consistent with the agglomeration of online retailers and corresponding trade shock. I first lay out how online retailers engage in e-commerce based on direct and indirect industrial evidence. These observations generates implications for online sellers' locations choices, agglomeration, and for intra-regional trade flows. I then introduce the specific data regarding online sellers, products, as well as intra-regional trade. Finally, I conduct empirical analysis to test the implications.

3.1 The Online Retail Business Model

A distinguishing feature of conducting e-commerce relative to conventional retail sales is the decoupling of retailer and consumer location. Retailers don't have to physically

Figure 1: The Online Retail Business Model



present where the consumers are to sell their goods, and instead, they engage in online match making with consumers through online platforms. Figure 1 shows the e-commerce business model of a typical online retailer. Different from brick-and-mortar retailing where the consumers need to commute to the store, in e-commerce, consumers obtain their goods either directly from the online retailer or from the storage the retailer has in the fulfillment center, both incurring a shipping cost. Nonetheless, e-commerce share one common feature with brick-and-mortar retail: the retailers has to buy goods from producers in the wholesale market, and incur transport cost there.

The key assumption for the e-commerce business model in Figure 1 is that online retailers first purchase the goods and place in their locations, before shipping to consumers, either directly or through third-party fulfillment service. Despite that the academic literature has little to say about the shipping modes of online sellers, in the Amazon data that I obtained, 72 percent of Amazon sellers and 78 percent of products sold use the Amazon fulfillment service, implying the use of direct shipping from producer to consumer is not a huge part of the sample. Some indirect evidence, such as case studies of Amazon sellers also indicate that these sellers' physical location act mainly as inventory storage, acting as the relaying point between producer and consumer.

Implications: The greater flexibility of online retailers' locations creates strong incentives of agglomeration. As in Krugman (1991); Krugman and Venables (1995) and Puga (1999), the presence of both spatial frictions and input-output linkages creates pecuniary spillovers of co-location. Specifically in the setting of the e-commerce business model, faced with matching frictions and shipping cost in the downstream, as well as transport cost in the upstream, online retailers would want to locate closer to either their major consumer or producer to save the costs, the decision of which depends on the relative cost magnitude on the two sides.

Such an insight makes it clear how the drastic expansion of e-commerce affect the economy by altering the location motives of online retailers. A key feature of e-commerce platform expansion (i.e., Amazon) post-2005 is its improvement of online shopping experience and rolling out of fulfillment centers. These changes reduces the matching friction and shipping cost to the downstream consumers. While online sellers' transport cost with upstream producer is not reduced as fast.[1] Such asymmetric changes in the spatial friction should motivate online retailers to locate more agglomerated in space to major producers, and the agglomeration is likely to be stronger when the online retailer has better access to fulfillment centers, since the shipping burden would be reduced more.

The potential agglomeration of online retailers will also alter the trade flows across regions. As online retailers serve as the intermediary of selling upstream producers to the downstream consumers, their agglomeration in a region will direct more purchases of the upstream goods in that region. If the region happens to the destination market where consumers are, there will be more imports of the upstream goods into the region; if on the contrary, the region happens to be the origin market, there will be less exports since the online retailers source locally. In testing these implications, I will use intra-regional wholesale trade data from CFS to check the purchase of upstream goods by online retailers. Under a similar vein, as a region gains better access to fulfillment facilities, it relaxes the burden of online retailer to the destination market, and is likely to be associated with reductions in wholesale trade flows.

3.2 Data

Products and Sellers on Amazon: The major data I used to test for the empirical implications and later quantitatively evaluate the model comes from Keepa (`www.Keepa.com`), an online marketing intelligence firm that serves both Amazon buyers and sellers by providing detailed information on products and sellers. Keepa started collecting data Amazon since 2011; once a product is searched by a consumer, Keepa will track it in its database. Therefore, Keepa's database includes any products that have ever been looked at by consumers, and is updated on a daily or weekly basis depending on the information. As of January 2023, Keepa's database includes more than 674 million products of 36 root categories sold on Amazon in the United States. For the purpose of my analysis, I took a

[1]The transport cost with upstream producers could also reduce in this period, due to general improvement in infrastructure and transportation and information technology. What's the key in driving the result, however, is the asymmetric changes in frictions, due to Amazon's more expansive presence on the downstream side.

1 % random sample out of each category and restrict to the period 2016-2018, which is after Amazon pick-up of e-commerce's expansion. Online Appendix Table **??** details the number of products of each category included in the analysis of this paper.

The product data I collect from Keepa contains each product's root category and brand, as well as longitudinal information such as prices, sales rank, and ratings. Several studies in the marketing literature show that a Pareto distribution fits the sales rank and quantity relationship well over e-commerce platforms. Using a combination of a book publisher's data and authors' own experiment, Chevalier and Goolsbee (2003) found that the coefficient of a regression of log sales quantities on log rankings to be around -0.76 to -1.11, while using the online sales data of 734 products of a retailer, Brynjolfsson et al. (2011) found the coefficient to be -0.88. Therefore, I convert the sales rank into quantity sold by running a similar regression and adopt an coefficient of -0.9.[2] Together with price information I then obtain the total sales revenue of a product overtime.

Moreover, I also obtain detailed seller information for the products sold on Amazon. Keepa starts to track sellers in 2016 and assign each seller with a unique identifier, which can then be linked to the seller profile on Amazon that contain information on the seller's address, fulfillment method, and whether the seller ships products from China. I retain all sellers that are located within the United States and that do not directly ship from China. Since a product can be available from multiple sellers at each point in time, I assign the seller of product to be the one that appear in the "BuyBox" , which accounts for more than 80% of sales of a product.[3]

Amazon Facilities: I obtain information on Amazon's fulfillment and distribution facilities from the supply-chain consulting firm MWPVL (`www.mwpvl.com`). The provided data contains the specific year and location a facility is built, its square footage, and detailed description of its functionality. For the purpose of my analysis, I focus on relatively bigger fulfillment and distribution centers that handle the common-sized domestic orders of non-perishable goods in typical regions. These are the facilities that most likely will lead to a decrease in shipping cost and therefore, consumers' shopping patterns and sellers' locations decisions.[4]

[2]What will also be important for the imputation is the intercept of the regression, since different product categories might have different innate level of sales quantity, despite the Pareto distribution fits well the quantity-rank relationship. I adopt Brynjolfsson et al. (2011)'s estimated intercept of 8.13 since their data cover broader product categories

[3]BuyBox is the "Add to Cart" and "Buy Now" section of the product detail page. Winners of the BuyBox are determined by Amazon algorithm that takes into account the price, product rating, delivery method of the sellers.

[4]Amazon also runs other specific centers that deal with fresh food and orders placed through Prime Now or Whole Foods, as well as centers that deal with in-bound goods and located near the airports, or deal

Figure 2: Expansion of Amazon Facilities

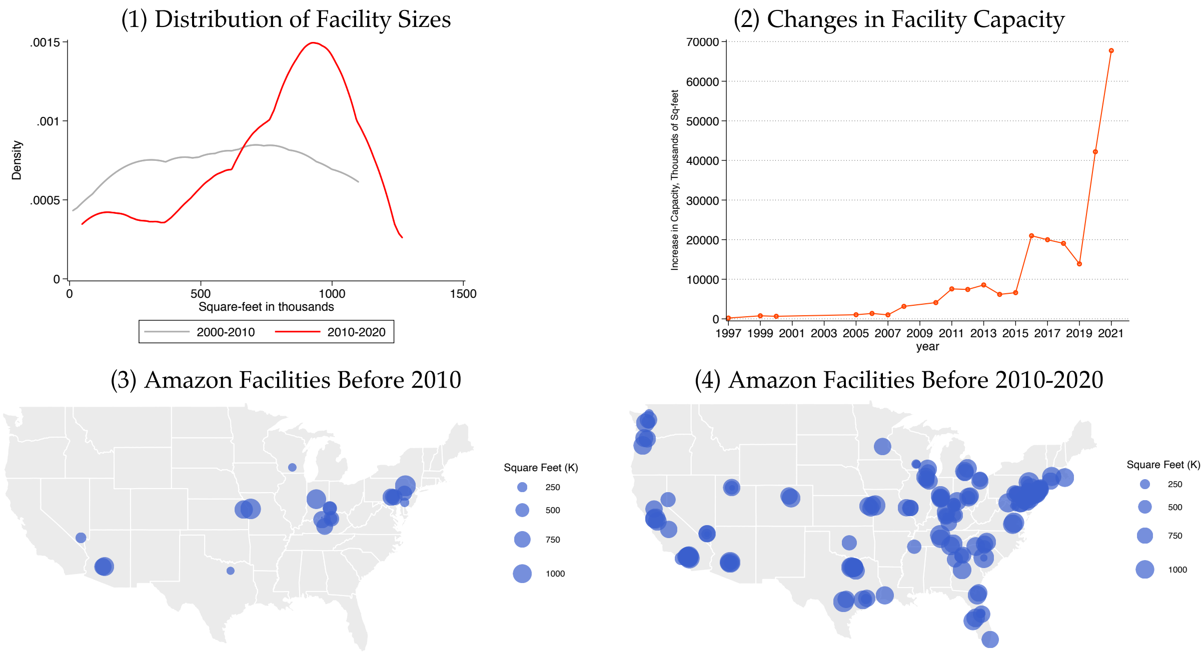


Figure 2 illustrates the capacity changes of Amazon's fulfillment and distribution facilities from 2000-2020. Panel (1) and (2) illustrate that there is a huge increase in center sizes from 2010-2020, with the majority of facilities built in this period at around 1 million square feet. Since 2015, there is a huge upsurge of 14-20 million square feet per year, leading to the height of 42 square feet built in year 2020. Panel (3) and (4) maps the locations of the centers using geo-coded address. From 2000-2010, most centers are concentrated in 3 states: New York, Kentucky, and Arizona. The geographical distribution of centers spread substantially starting 2010, covering most U.S. states with concentrations in the east and west coast.

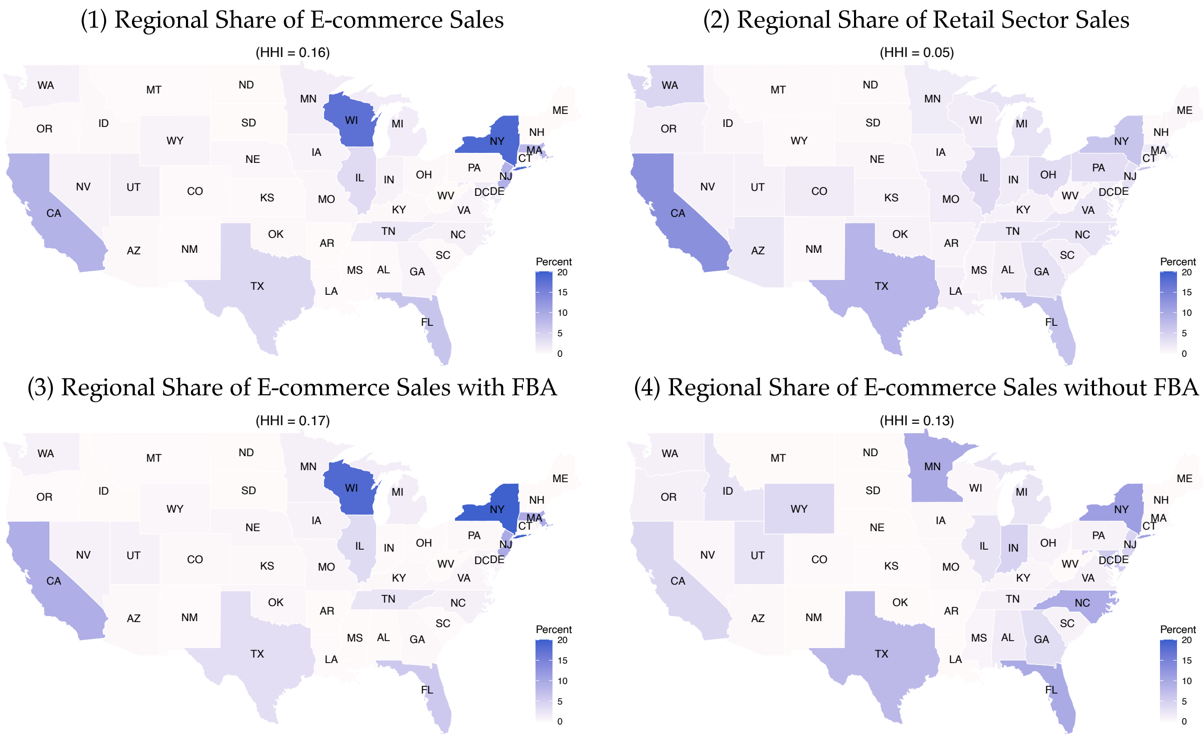
Trade Flow. For the data patterns regarding intra-regional trade flows, I rely on Commodity Flow Survey (CFS) that provides representative shipment level trade flows in value and quantity for all the 30 manufacturing and retail sectors across 50 U.S. states.

3.3 Data Patterns

In this section, I document four broad data patterns that point to the differential concentration of online sellers and implications for intra-regional trade flows.

with small packages; these facilities are excluded from my analysis. Within the fulfillment and distribution category, I don't differentiate whether the center is serving more in terms of storage or sortation, as both reduces the shipping time and cost.

Figure 3: Spatial Concentration of Online vs. Overall Retail Sales and Sellers



Pattern 1 *Online retail sales are more spatially concentrated than overall retail sector sales, particularly for those that are FBA.*

Figure 3 panels (1) and (2) depict the states' shares of total online retail sales on Amazon and their shares of overall retail sector value-added, based on average values from 2016 to 2021.[5] I assign the sales value of online sellers to states according to the sellers' addresses, and I use data from the BEA to obtain states' shares of retail sector value-added. The results show that online retail sales are more spatially concentrated than overall retail sales. Specifically, New York and Wisconsin account for 36 percent of total online retail sales, followed by California and Florida. In contrast, the distribution of overall retail sector value-added is more aligned with state population sizes. The Herfindahl-Hirschman Index (HHI) confirms this concentration, with a value of 0.16 for online retail sales and 0.05 for overall retail value-added.

Further, Amazon's fulfillment services reduce the shipping burden for online retailers, leading to greater agglomeration. Figure 3 panels (3) and (4) show the states' shares of online retail sales using Amazon's Fulfillment by Amazon (FBA) service compared to those that do not. Sales through FBA are more spatially concentrated, driving the

[5] States' shares of retail value added serve as an effective proxy for their retail sales shares under a constant factor share production function with consistent returns.

Table 1: HHI Index by Product Categories

Category name	**HHI Index**
Toys & Games	0.12
Patio, Lawn & Garden	0.12
Arts, Crafts & Sewing	0.07
Sports & Outdoors	0.14
Office Products	0.16
Grocery & Gourmet Food	0.08
Tools & Home Improvement	0.21
Movies & TV	0.08
Musical Instruments	0.10

Table 2: Online and Total Retail Sales with Population and Corporate Taxes

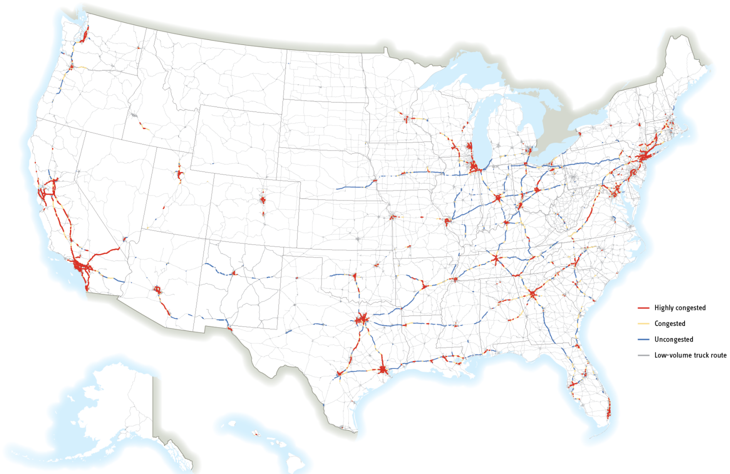
Dependent Variable (in %)	**Online Retail**	**Overall Retail**
ln (corporate tax)	-0.01	0.03*
	[1.29]	[0.02]
Population share (%)	14.54*	1.06***
	[7.92]	[0.26]
Year, State FE	X	X
Observations	230	230
R-squared	0.52	1.00

overall concentration of online retail sales. The HHI is higher for FBA sales compared to non-FBA sales (0.17 versus 0.13).

Pattern 2: *Online retail sales of durable and standardized products are more concentrated than those of non-durable and non-standardized products.*

The agglomeration patterns of online sellers vary by product groups. Standardized and durable products, which benefit from economies of scale, predictable demand, and optimized transportation and storage, tend to have more concentrated seller locations. In contrast, non-durable and non-standardized products have a more dispersed seller distribution. Table 1 shows the HHI for sales across regions for nine popular goods on Amazon. Durable goods like "Tools & Home Improvement" and "Office Products" have higher HHI indices (0.21 and 0.16), while non-durable goods like "Arts, Crafts & Sewing" and "Grocery & Gourmet Food" have lower indices (0.07 and 0.08). The concentration of online retailers of durable and standardized product provides additional confirmation of their address information since these businesses tend to have larger transactions and operate at fixed and easily verifiable locations.

Figure 4: Peak Period Congestion on the High-Volume Truck Routes in 2020



Pattern 3: *Online retail sales exhibit a weaker relationship with corporate taxes and population, but aligns closely with regional truck volumes.*

To better understand how online retail sales differ from the overall retail sector in terms of agglomeration patterns, Table 2 presents regression results analyzing the relationship between states' shares of online retail sales, overall retail sector value-added, and state percentages of population and corporate tax revenues over time, with state and time fixed effects included. The findings indicate that the overall retail sector is more strongly linked to corporate taxes and population, as shown by the higher significance of these variables. Importantly, the R-squared values for the online retail regression is significantly lower than that for the overall retail sector. While state fixed effects, time trends, corporate taxes, and population explain nearly all the variation in retail sector value-added shares, they account for only about 50 percent of the variation in online sales shares.

Moreover, regional sales shares in the online retail sales closely align with regional truck volumes. Figure 4 depicts peak period congestion on high-volume truck routes, showing that states like Wisconsin, Illinois, New York, Texas, Florida, and California experience the highest levels of truck volume congestion – these are exactly the states that have the highest concentration of online retailers. This correlation may stem from the advanced transportation and logistics infrastructure available in these areas, which supports the agglomeration of online sellers.

Pattern 4: *Destination markets with more online retailers import less tradable goods, whereas origin markets with more online retailers export more tradable goods.*

Table 3: Inter-regional Trade, Online Retailers, and Amazon Facility

Dependent Variable:	Δ ln (Shipment)	
Δ share (%) of online sellers - destination	-1.4*	
	[0.7]	
Δ share (%) of online sellers - origin	3.5***	
	[0.8]	
Δ ln (bilateral distance via Amazon facility)		4.92*
		[2.53]
Origin, destination FE	✓	
Industry FE	✓	
Observations	24,693	24,693
R-squared	0.20	0.19

As online retailers act as focal point between producers and consumer, their agglomeration in a region is closely linked to the shifts in import and export trade flows. Column (1) of Table 3 presents the relationship between the percentage change in regional online retailer shares and the log difference in wholesale trade flows. Since Keepa is available only after 2016 and CFS is conducted every five years, with the latest in 2017, I regress changes in inter-regional trade flows from 2012 to 2017 on changes in states' shares of online retailers between 2016 and 2017, controlling for fixed origin, destination, and industry characteristics. The results indicate a one percent increase in a destination state's share of online retailers is associated with a 1.4 percent decline in wholesale shipments to that state, while a one percent increase in the origin state's share of online retailers is associated with a 3.7 percent increase in wholesale shipments from that state.

Pattern 5: *Regions near to Amazon's fulfillment facilities have more bilateral tade flows*

Improved logistics infrastructure from Amazon's fulfillment services reduces shipping friction between regions, potentially attracting more online retailers and boosting bilateral trade flows. Table 3 column (2) shows the relationship between the log difference in shipment value of wholesale trade goods between an origin-destination pair and the change in log distance to the nearest Amazon fulfillment center, controlling for fixed origin, destination, and industry characteristics. Following the method in Houde et al. (2021), which shows that over 90 percent of orders are handled by the three nearest centers, I assign the closest center to both the origin and destination as the most likely to handle the shipment. Consistent with the prediction, the results indicate that a 1 percent reduction in bilateral distance due to Amazon's fulfillment expansion is associated with a 4.92 percent increase in the shipment value of wholesale trade goods between regions.

4 A Spatial Retail Trade Model

In this section, I build a multi-sector spatial retail trade model in which consumers search for retailers across regions, and online retailers choose their optimal locations. The general environment of the model contains N regions denoted by n or m, and J sectors denoted by j or k. For each region-sector pair, there is a representative manufacturer (M) and a brick-and-mortar store (B), as well as a flexible measure of online retailers (R). I show that consumers' search and shopping problem simplify to a CES demand with a demand shifter of online retailers representing the efficiency of matching. The production intermediate varieties follows multi-sector Eaton and Kortum (2002), for which regions with comparative advantages obtain higher share of the market demand subject to shipping frictions.

The distinguishing feature of the model is the incorporation of two types of retailers within a vertical production structure. While brick-and-mortar retailers gather manufactured varieties to sell exclusively to local consumers, online retailers have the flexibility to select locations that offer cost advantages for both purchasing inputs and selling to multiple markets. The location choices of online retailers then play an important role in determining these trade flows. In what follows I describe the spatial retail trade problem, explore the role of e-commerce, and then derive the comparative statistics. Quantitative evaluation of the model is presented in the next section.

4.1 Search-Based Demand Derivation

Consumer Search Micro-Foundation: There is a continuum of consumers in region n, each consuming goods from the durable and non-durable retail sectors, as well as services, with sectoral weights η^j. The retail sector operates under monopolistic competition, where each retailer sells a unique variety. There is a total measure of $1 + O$ retailers available for consumers to purchase from, where there is measure 1 of local brick-and-mortar store, and O is the normalized measure of online retailers. Consumers have Cobb-Douglas utility over sectoral goods, expressed as $u_n = \sum_j \eta^j \ln c^j_{ni}$, where i represents the chosen retailer for sector j goods. With an income y_n, a consumer's optimal consumption from the chosen retailer is $c^{j*}_{ni} = \eta^j y_n / p^j_{ni}$, where p^j_{ni} is the price to obtain goods from retailer i that includes commuting costs for local stores or shipping costs for online retailers.

Consumers engage in sequential directed searches for retailers, with the value of each potential match being unknown initially. Specifically, the indirect utility a consumer from region n gains from purchasing a good from retailer i in sector j is given by

$v^j_{ni} = \ln \eta^j y_n - \ln p^j_{ni} + \epsilon^j_{ni}$. The term ϵ^j_{ni} captures the idiosyncratic match value between a consumer and a retailer, assumed to be independently distributed according to a function $F(\epsilon)$ which is unknown to the consumer and retailer. The match value for local brick-and-mortar stores is normalized to zero ($\epsilon^j_{n0} = 0$), while the average relative match value with online retailers is $\ln(\mu)$.[6] Consumers may purchase directly from local stores or undertake sequential searches for online retailers, incurring a cost s at each step to observe the match value ϵ^j_{ni} of an online retailer.

In line with Weitzman (1979), the optimal consumer strategy is to order their search by $\epsilon^{j*}_{ni} - p^j_{ni}$, where ϵ^{j*}_{ni} represents the minimum match value that makes the consumer indifferent between continuing to search or stopping ($s = \int_{\epsilon^j_{ni}}^{\bar{\epsilon}^{j*}_{ni}} (1 - F(\epsilon)) d\epsilon$).[7] This sequential search process leads to an eventual purchase choice Choi et al. 2018; Armstrong 2017; Armstrong and Vickers 2015. Defining $\omega^j_{ni} \equiv min\{\epsilon^j_{ni}, \epsilon^{j*}_{ni}\}$, which represents the "effective match value" of a retailer, the consumer will buy from the retailer i that maximizes $\omega^j_{ni} - p^j_{ni}$.[8]

$$
\begin{aligned}
D^j_{ni} &= P(\omega^j_{ni} - \ln p^j_{ni} > max_g\ \omega^j_{ng} - \ln p^j_{ng}) \\
&= \int \Pi_{g \neq i} F_{\omega^j_{ng}}(\epsilon - \ln p^j_{ng}) f_{\omega^j_{ni}}(\epsilon - \ln p^j_{ni}) d\epsilon
\end{aligned} \tag{1}
$$

Optimal Demand Derivation: Sequential directed search by consumers leads to a CES demand representation in two steps. First, consumers' eventual purchases based on ω^j_{ni} yield a discrete choice formulation of optimal demand, as described in Anderson et al. (2022). The demand of a representative consumer in region n for retailer i, noted as D^j_{ni}, mirrors a discrete choice model when $F_{\omega^j_{ni}} = F_{\epsilon^{j,DC}_{ni}}$, where $\epsilon^{j,DC}_{ni}$ reflects random utility from ω^j_{ni}. Second, CES demand can be seen as a special case derived from discrete choice. Given that the average ϵ^j_{ni} is 0 for brick-and-mortar stores and $\ln(\mu)$ for online retailers, we express $\epsilon^{j,DC}_{ni}$ as $\ln(\mu) + \chi^j \tilde{\epsilon}^j_{ni}$, where $\tilde{\epsilon}^j_{ni}$ has a mean of 0 and unit variance, and χ^j represents the sector-specific variance of the effective match value ω^j_{ni}.[9] Assuming

[6]This relative match value $\ln(\mu)$ distinguishes the shopping experiences between physical and online stores. Taking the logarithm simplifies the representation and is without loss of generality. A value of $\mu > 1$ suggests consumers derive higher utility from online shopping, whereas $\mu < 1$ indicates the opposite.

[7]The consumer will stop and make a purchase from either a local brick-and-mortar store or an online retailer i if $max\{v^j_{n0}, -max_{g \in \bar{O}} \ln p^j_{ni} + \epsilon^j_{ni}\} > max_{g \in \bar{O}} - \ln p^j_{ng} + \epsilon^j_{ng}$, where $\bar{O}$ denotes the retailers the consumer has searched so far.

[8]As Choi et al. (2018) shows, to guarantee the existence and uniqueness of the equilibrium, one needs the density and loss function sof ω^j_{ni} to be log-concave, and the density function to be unbounded above. These are taken as assumptions for this paper.

[9]This means that the effective match value ω^j_{ni} also has a mean of $\ln(\mu)$, given the abundance of online

an extreme type I distribution of $\tilde{\epsilon}^j_{ni}$, the consumer demand in region n from retailer i becomes $D^j_{ni} = \frac{(p^j_{ni}/\mu)^{-1/\chi^j}}{\sum_{g=1}^N (p^j_{ng}/\mu)^{-1/\chi^j} + (p^j_{n0})^{-1/\chi^j}}$, representing a standard CES expenditure share.

Theorem 1 presents the final CES demand of consumers and highlights the role of search matching in shaping consumer demand. This expression clarifies the importance of μ: as online shopping and matching improve in efficiency, μ increases, leading consumers to shift more of their demand toward online retailers. The variance of consumers' effective match value, χ^j, determines the elasticity of substitution among retailers, $\sigma = \frac{1+\chi^j}{\chi^j}$. When there is less uncertainty regarding the value of goods from online retailers, χ^j becomes smaller, making retailers more substitutable. Under monopolistic competition, the markup charged by retailers is $\tilde{\sigma}^j = \frac{\sigma^j}{\sigma^j - 1}$.

Theorem 1. *A representative consumer in region n with sectoral consumption weights η^j has nest Cobb-Douglas and CES demand as below under sequential ordered search if only if the effective match value* $\min\{\epsilon^j_{ni}, r(\ln \mu)\}$ *is distributed extreme type I*

$$C_n = \Pi_{j=1}^J (C^j_n)^{\eta^j}, \quad C^j_n = [(c_{n0})^{\frac{\sigma-1}{\sigma}} + \mu \sum_{i=1}^N (c_{ni})^{\frac{\sigma-1}{\sigma}}]^{\frac{\sigma_j}{\sigma_j - 1}} \text{ for } j \geq 2 \tag{2}$$

Proof: See Appendix B.

4.2 Production

Production is organized as a multi-stage vertical process to capture the role of retailers. Retailers first gather intermediate goods from various regions, convert them into final products, and sell these to consumers in different regions, applying a markup in the process. This model thus features two layers of intra-regional trade: one for intermediate goods and another for final goods.

A key distinction in this model is the presence of two types of retailers: brick-and-mortar (B) and online (R). Brick-and-mortar retailers serve only local consumers, while online retailers can sell to all regions, strategically choosing their locations to maximize revenue while considering spatial production and cost structures. The location decisions of online retailers ultimately shape the intra-regional trade flows of final retail goods.

Intermediate Varieties. The intermediate goods market is perfectly competitive, with a representative firm in each sector j of region n producing a continuum of varieties

retailers compared to local stores. Since the cost of searching for an additional retailer s is relatively low online and decreases with ϵ^{j*}_{ni}, ω^j_{ni} approximates ϵ^j_{ni}, which has a mean of $\ln(\mu)$.

$e^j \in [0, 1]$. The production function is given by:

$$q_n^{j,M}(e^j) = a_n(e^j) \left[h_n(e^j)^{\beta_n} l_n(e^j)^{1-\beta_n} \right],$$

where $a_n(e^j)$ is the factor-neutral productivity for variety e^j, and $l_n(e^j)$ represents labor. The production function includes regional structures, $h_n(e^j)$ that complemens labor, bundled in a Cobb-Douglas form with shares controlled by β_n. All firms across regions use this constant returns to scale technology and possess no market power. Prices are set to unit costs, as in equation (3), with r_n^h as structure costs and w_n^j as wages. Intermediate goods trade involves an iceberg cost, so κ_{ni}^M units are required to ship one unit from i to n.[10] The price of variety e^j in region n ($p_{ni}^{j,M}(a(e^j))$) is the lowest effective unit cost, adjusted by the iceberg cost, also specified in equation (3).

$$c_n^{j,M} = \left[(\frac{r_n^h}{\beta_n})^{\beta_n} (\frac{w_n^j}{1-\beta_n})^{1-\beta_n} \right], \quad p_{ni}^{j,M}(a(e^j)) = \min_i \{ \kappa_{ni}^M \frac{c_n^{j,M}}{a_i(e^j)} \}. \tag{3}$$

Further parameterizing the productivity distribution as in Eaton and Kortum (2002) gives a gravity representation of trade. Specifically, let the productivity vector across regions be $a(e^j) = \{a_1(e^j), \ldots, a_N(e^j)\}$, where each $a_n(e^j)$ is a random draw from a Fréchet distribution with shape and scale parameters θ^j and $T_n^{j,M}$, respectively: $\phi_n^j(a_n(e^j)) = \exp(-T_n^{j,M} z^{-\theta^j})$. The Fréchet shape parameter θ^j determines the dispersion of productivities across regions and thus the within-sector specialization, while $T_n^{j,M}$ regulates regions' absolute production advantages and cross-sector specialization. Using the properties of the Fréchet distribution, the expenditure share of region n on region i for sector j intermediate goods, $x_{ni}^{M,j} = X_{ni}^{M,j} / X_n^{M,j}$, can then be expressed in a gravity formula:

$$x_{ni}^{j,M} = \frac{(\kappa_{ni}^M c_i^{j,M})^{-\theta^j} T_i^{j,M}}{\sum_{m=1}^N (\kappa_{nm}^M c_m^{j,M})^{-\theta^j} T_m^{j,M}}, \tag{4}$$

which reflects the likelihood that consumers in region n purchase sector j varieties from retailers in region i.[11]

Retail Sector. The retail sector features a vertical production structure that connects upstream intermediate producers with downstream consumers. In a given region and sector, both brick-and-mortar and online retailers first gather various intermediate vari-

[10] The iceberg cost satisfies standard requirement $\kappa_{ni}^M > 1$ for $i \neq n$ and $\kappa_{ni}^M = \kappa_{in}^M$.

[11] Here, θ^j can be interpreted as the trade elasticity with respect to cost. A higher θ^j indicates lower productivity dispersion across regions and greater responsiveness of import volumes to cost changes.

eties $e^j \in [0,1]$ from the lowest-cost producers. These varieties are then combined into a retail bundle $q_n^{j,R/B}$ for the production of retail goods, as described in equation (5), where α^j regulates the elasticity of substitution among the varieties in sector j.[12]

$$q_n^{j,R/B} = [\int_0^1 q_n^{j,M}(e^j)^{\frac{\alpha^j-1}{\alpha^j}} d\phi^j(a^n(e^j))]^{\frac{\alpha^j}{\alpha^j-1}} \tag{5}$$

$$Q_n^{j,R/B} = z_n^{j,R/B}\left[(h_n^{j,R/B})^{\beta_n}(l_n^{j,R/B})^{1-\beta_n}\right]^{\gamma_n^j}\left[q_n^{j,R/B}\right]^{1-\gamma_n^j} \tag{6}$$

The retailers then combine the retail good aggregate with other inputs in a nested Cobb–Douglas production function to produce the final retail good, with share of value-added given by γ_n^j as in equation (6). Both type of retailers uses labor and structure bundle with labor shares given by $1-\beta_n$, similar to that of intermediate producers. Given retail sector's production function, the unit cost of retail good is given by:

$$c_n^{j,R/B} = (\varrho_n^{j,R/B})^{\gamma_n^j}(p_n^{j,M})^{1-\gamma_n^j}, \tag{7}$$

where $p_n^{j,M} \equiv (\Gamma(\dfrac{\theta^j+1-\alpha^j}{\theta^j}))^{\frac{1}{1-\alpha^j}}(\sum_{m=1}^{N}(\kappa_{nm}^M c_m^{j,M})^{-\theta^j}T_m^{j,M})^{\frac{1}{-\theta^j}}$, $\varrho_n^{j,R/B} \equiv (\dfrac{r_n^{j,R/B}}{\beta_n})^{\beta_n}(\dfrac{w_n^{j,R/B}}{1-\beta_n})^{1-\beta_n}$.

Here, $p_n^{j,M}$ represents the price index of the aggregate intermediate varieties, derived from the properties of the Fréchet distribution applied to the productivity vector $\phi^j(a^j(e^j))$, with $\Gamma(.)$ being the gamma function evaluated at $\frac{\theta^j+1-\alpha^j}{\theta^j}$.[13] The term $\varrho_n^{j,R}$ denotes the unit cost of labor and structure in the retail sector. Given the monopolistic market structure of the retail sector, the price of retail goods shipped from market i to n is calculated as $p_{ni}^{j,R} = \tilde{\sigma}\kappa_{ni}^R c_i^{j,R}$, where $\tilde{\sigma}$ is the markup, $c_i^{j,R}$ is the unit retail cost, and κ_{ni}^R represents the iceberg cost that subsumes shipping cost and other bilateral frictions.

Online Retailer Location: The distinguishing feature of online retailers is their operational flexibility across different locations. While each brick-and-mortar store's productivity is tied to its specific location $z_n^{j,B}$, online retailers can draw a vector of productivities across

[12]This model structure differs from the input-output linkages in Costinot and Rodríguez-Clare (2014) and those in recent quantitative trade models, where intermediate goods production also requires these aggregates as inputs. In contrast, this structure more accurately reflects the retail industry, where intermediate production relies solely on primary factors, and and retail goods are intended for final consumers only.

[13]Since the vector of productivity draws for variety e^j across regions is $a(e^j) = \{a_1(e^j), \ldots, a_N(e^j)\}$, their joint distribution becomes $\phi^j(a^j(e^j)) = \exp\left\{-\sum_{n=1}^{N} T_n^{j,M}(z)^{-\theta^j}\right\}$. The parameter condition $\theta^j + 1 - \alpha^j > 0$ is assumed to ensure that the price index is well-defined.

various locations $(z_1^{j,R}, ..., z_N^{j,R})$ to establish their operations in any given region m, paying a fixed entry cost in labor units f_m. Once located in region m, they import intermediate varieties from various regions and distribute the final retail goods to consumers across different locations.[14] The optimal location choice for online retailers is then:

$$m^* = \arg\min_m \left\{ \sum_n \left(\tilde{\sigma} c_m^{j,R} \frac{\kappa_{nm}^R}{P_n^{j,R}} \right)^{\sigma^j - 1} \cdot \frac{1}{\eta^j X_n} \right\},$$

which indicates that online retailers optimally locate in region m if it minimizes the product of unit retail production costs and the weighted sum of iceberg costs to various destinations. The weights are determined by the total expenditure of each destination market, X_n, and the destination market's retail price index, $P_n^{j,R}$.

This setup highlights the agglomeration and dispersion forces in the model, in the spirit of Krugman (1991). Online retailers prefer to locate where iceberg costs to large markets are the lowest (HME) or where imported goods are the cheapest. However, agglomeration also leads to higher wages and land prices, raising retail production costs. Further, online retailers will enter a market only if the total revenue across destinations exceeds the entry costs $\sum_n (\frac{p_{nm}^{j,R}/\mu}{P_n^{j,R}})^{1-\sigma^j} \eta^j X_n \geq \sigma^j w_m^{j,R} f_m$. This entry condition determines the threshold unit cost for online retailers to enter a market:

$$\bar{c}_m^{j,R} = \frac{\mu}{\tilde{\sigma}^j} \left[\frac{\sigma^j}{\eta^j} \frac{w_m^{j,R} f_m}{\sum_n (\kappa_{nm}^R / P_n^{j,R})^{\sigma^j - 1} X_n^{-1}} \right]^{\frac{1}{1-\sigma^j}} \tag{8}$$

To gain tractability and derive closed form solution for online retailers' locations, I follow the multinational production literature (Arkolakis et al. 2017, 2018) to assume that the productivity vectors of online retailers are randomly drawn from a multi-variate Pareto distribution $P(Z_1^{j,R} < z_1, ..., Z_N^{j,R} < z_N) = 1 - (\sum_{m=1}^N [T_m^{j,R} z_m^{-\tau}]^{\frac{1}{1-\rho}})^{1-\rho}$.[15] The scale parameter $T_m^{j,R}$ measures the absolute advantage of region m in producing sector j goods, whereas τ controls the degree of heterogeneity across different vectors, and ρ controls the degree of heterogeneity within a single vector of different realizations. Define $\xi_m^j \equiv \sum_n (\frac{c_m^{j,R} \kappa_{nm}^R}{P_n^{j,R}})^{\sigma^j - 1} \frac{1}{X_n}$, so $m^* = \arg\min_m \{ \frac{\xi_m^j}{z_m^{j,R}} \}$, the probability of a sector j retailer to

[14] This setup departs from the framework typically seen in the multinational production literature, such as in Arkolakis et al. (2018), where a firm selects the optimal production location specifically to serve a targeted destination. Instead, in this model, retailers select a location that maximizes profitability across all markets, considering upstream and downstream production processes and cost structures.

[15] The support of this distribution requires $z_m \geq (\sum_{m=1}^N (A_m^j)^{\frac{1}{1-\rho}})^{1-\rho}$ and $\rho \in [0, 1)$.

locate in m can then be expressed as

$$\Psi_m^j = P(m = \arg\min_m \{\xi_m^j / z_m^{j,R}\} \cap c_m^{j,R} < \bar{c}_m^{j,R}) = \psi_m^j (\bar{c}_m^{j,R})^\tau, \tag{9}$$

where $\psi_m^j = T_m^{j,R}(\xi_m^j)^{\frac{-\tau}{1-\rho}} / \sum_{m=1}^N [T_m^{j,R}(\xi_m^j)^{-\tau}]^{\frac{-\rho}{1-\rho}}$.[16] This equation presents a probabilistic formulation of online retailers' location choices, accounting for the elasticity of substitution across regional productivities in retail production, parameterized by τ and ρ.

The location of online retailers plays an important role in determining the intra-regional aggregate trade flows. With a total of measure O of online retailers, the measure of online retailers in location m is $O_m = O\Psi_m^j$. Therefore, the total sales from region m to n is the product of sales per firm and the measure of firms: $(\frac{p_{nm}^{j,R}/\mu}{P_n^{j,R}})^{1-\sigma^j}\eta^j X_n O_m$. We can then obtain the bilateral online retail expenditure share $x_{nm}^{j,R}$ as in equation (10), which represents an extended gravity equation of Chaney (2008)'s version of the Melitz model.

$$x_{nm}^{j,R} = \frac{\Psi_m^j(\kappa_{nm}^R c_{nm}^{j,R}/\mu)^{1-\sigma}}{\sum_h \Psi_h^j(\kappa_{nh}^R c_{nh}^{j,R}/\mu)^{1-\sigma} + \frac{1}{O}(c_{n0}^{j,R})^{1-\sigma}} \tag{10}$$

Unlike standard gravity equation of trade, the numerator ("bilateral resistance") depends not only on the retail production cost of the origin, but on the probability of online retailers locating in that region, as well as the online matching efficiency; the denominator ("multi-lateral resistance") includes both the sum of bilateral resistance as well as the cost of local brick-and-mortar store divided the measure of available online retailers. Further, this model allows substantial quantitative tractability because the location probability of online retailers can be directly observed from the data.

4.3 Labor Supply

To characterize workers' sorting and heterogeneous labor supply across sectors, I adopt a Roy (1951) framework with probabilistic productivities (Lagakos and Waugh 2013; Hsieh et al. 2019; Galle et al. 2022; Lee 2020). In each region, workers obtains a vector of region-sector specific productivities $z_n = \{z_n^0, z_n^S, z_n^{1,M}, z_n^{1,R}, z_n^{1,B}, ..., z_n^{j,M}, z_n^{j,R}, z_n^{j,B}\}$ for each unit of its labor provided, for which sector 0 is treated as non-employment as Dvorkin (2014) and Caliendo et al. (2019).[17] The productivities are drawn independently

[16] Note that $\tilde{\sigma}$ and η^j do not appear in the definition of ξ_m^j and Ψ_m^j since they are constant within a sector.

[17] Non-employment is treated as a sector that workers can allocate their labor into, with a wage of w_n^0 per efficiency unit of labor that can be understood as the marginal return for home production, and households' consumption when non-employed depend on the labor units they withdraw from the employment sectors.

from a Fréchet distribution $\psi_n^{j,K}(z_n^{j,K})$ with shape parameter ν_n and scale parameter $A_n^{j,K}$, $K = \{M, R, B, \emptyset\}$. The scale parameter $A_n^{j,K}$ gives the absolute advantage while the shape parameter ν_n regulates the comparative advantage of workers, jointly determining the sorting pattern on the labor market.

From properties of Fréchet distribution, the joint distribution of productivities draws follows another Fréchet distribution $\psi_n(z_n) = \sum_{j=0}^{J} \sum_{K=M,R} A_n^{j,K} z^{-\nu_n}$. Taking account the idiosyncratic productivity, workers' wage per unit of labor supply is $w_n^{j,K} z_n^{j,K}$, which workers seek to maximize by choosing sector (j, K) optimally. Define the optimum choice set for a sector (j, K) by $\Lambda_n^{j,K} \equiv \{z_n^{j,K} st. z_n^{j,K} > z_n^{H,k} \forall (H, k)\}$, then a worker will choose to be employed in (j, K) if the obtained vector draw of productivities is in this set. Applying the properties of the joint Fréchet distribution for the productivity draws $\psi_n(z_n)$, we can drive the probability of non-employment, as well as the employment in sector (j, K) as:

$$\pi_n^0 = \frac{A_n^0 (w_n^0)^{\nu_n}}{\Phi_n}, \; \pi_n^{j,K} = \frac{A_n^{j,K} (w_n^{j,K})^{\nu_n}}{\Phi_n},$$
$$\text{where } \Phi_n = \sum_{j=1}^{J} \sum_{K=M,R} A_n^{j,K} (w_n^{j,K})^{\nu_n} + A_n^0 (w_n^0)^{\nu_n}. \tag{11}$$

The probability of being non-employed or employed in a certain sector is shown to be proportional to the return of home production or sectoral wage relative to the total returns of being employed and non-employed, scaled by the Fréchet parameter ν_n that plays the role of the elasticity of labor adjustment.[18] Therefore, as labor demand changes affect wages, they also alter households employment decisions. Another tractability gained from the Fréchet distribution is that the efficiency units of labor supply can be conveniently derived; specifically, for a sector (j, K):

$$l_n^{j,K} \equiv \Gamma(\frac{\nu_n - 1}{\nu_n}) \frac{\Phi_n^{1/\nu_n}}{w_n^{j,K}} \pi_n^{j,K} L_n \tag{12}$$

where $\Gamma(.)$ denotes a gamma function. Workers' income as well as firms' production depend on this efficiency units of labor provided, and the wage return for workers in sector $w_n^{j,K} l_n^{j,K}$ becomes $\Gamma(\frac{\nu_n - 1}{\nu_n}) \Phi_n^{1/\nu_n} \pi_n^{j,K} L_n$.

[18] As discussed in Galle et al. (2022), if $\nu_n \to \infty$, the households become homogeneous in employment choices and $\nu_n \to 1$ delivers the same comparative statics as sectoral specific labor supply.

4.4 Market Clearing and Competitive Equilibrium

On the goods market there exist two types of expenditures: consumers purchase retail goods across retailers, and retailers acquire intermediate varieties from different regions. In equilibrium, both of these markets need to be cleared:

$$X_n^{j,R} = \sum_{i=1}^{N} x_{in}^{j,R}(I_i L_i), \text{ where } I_i L_i = \sum_{k=0}^{J} \sum_{K=M,R} (r_i^{h,k} h_i^{K,k} + w_i^k l_i^{K,k}) - \Omega_i, \tag{13}$$

$$X_n^{j,M} = \sum_{i=1}^{N} (1-\gamma_i^j) x_{in}^{j,M} X_i^{j,R}. \tag{14}$$

The total expenditure or demand of sector j retail goods sold from region n denoted by $X_n^{j,R}$ has to equal to the product of the retail expenditure share on region n's retail goods $x_{in}^{j,R}$ and total income $I_i L_i$ across regions. In the benchmark model, households' total income comes from their wage earnings and ownership of land, minus a region's trade deficit denoted by Ω_i that is assumed to be exogenous.[19] On the other hand, the total demand for sector j intermediate goods from region n, denoted by $X_n^{j,M}$, equates the expenditure share on region n's intermediate goods $x_{in}^{j,M}$ times the portion of retail sector's spending on intermediate varieties $(1-\gamma_i^j)X_i^{j,R}$ summed across regions. Accounting for regional trade deficits leads to the balance of trade equation:

$$\sum_{j=0}^{J} \sum_{i=1}^{N} (x_{ni}^{j,M} X_n^{j,M} + x_{ni}^{j,R} X_n^{j,R}) + \Omega_n = \sum_{j=0}^{J} \sum_{i=1}^{N} (x_{in}^{j,M} X_i^{j,M} + x_{in}^{j,R} X_i^{j,R}). \tag{15}$$

The clearing of the markets for primary factors including labor and structures follows the same manner that each of their return needs to equal to the portion of value-added. However, since these factors are used in the production of both intermediate and retail goods that are subject to different production functions, the market clearing conditions differ for intermediate and retail sectors. Specifically, for the labor market:

$$\begin{gathered} w_n^{j,M} l_n^{j,M} = w_n^{j,M} \int_0^1 h_n(e^j) d\phi_n^j(a_n(e^j)) = \beta_n X_n^{j,M}, \\ w_n^{j,R} l_n^{j,R} = \gamma_n^j \beta_n X_n^{j,R}, \quad w_n^{j,B} l_n^{j,B} = \gamma_n^j \beta_n X_n^{j,B}, \end{gathered} \tag{16}$$

[19] In Section V's discussion of policy interventions, households' total income will also depend on the "tariff" that a local region imposes on others, and an endogenous deficit that is affected by revenue reallocation.

Model Equilibrium and Comparative Statics. To characterize the competitive equilibrium for this interregional retail trade framework, we need to specify the economy's fundamentals and model parameters. The fundamentals of the model economy include the sector-region productivities in producing intermediate goods as well as retail goods $(T^M, T^R) = \{T_n^{j,M}, T_n^{j,R}\}_{n=1,j=1}^{N,J}$, workers' productivities in different sectors $A^K = \{A_n^{j,K}\}_{n=1,j=1}^{N,J}, K = \{M, R, B, \emptyset\}$, the demand shifters for retail goods across regions μ, the iceberg trade costs of manufacturing and retail goods $(\kappa^M, \kappa^R) = \{\kappa_{ni}^M, \kappa_{ni}^R\}_{n=1,i=1}^{N,N}$, the stock of structures across markets $(h^M, h^R) = \{h_n^{j,M}, h_n^{j,B}\}_{n=1,j=1}^{N,J}$, and the exogeneous trade deficits of different places $\Omega = \{\Omega_n\}_{n=1}^{N}$. For clarity, here I denote these fundamentals by $\Psi \equiv \{T^M, T^R, A^K, \mu, \kappa^M, \kappa^R, h^M, h^R, \Omega\}$.

The model parameters include sector consumption shares (η_n^j), elasticity of substitution among retailers $(\frac{1}{1-\sigma^j})$, elasticity of substitution among intermediate varieties $(\frac{1}{1-\alpha^j})$, the Fréchet shapes of worker (v^n) and sector productivities (θ^j), the labor share of value-added $(1 - \beta_n^j)$, and the value-added share of retail goods (γ_n^j), all of which are assumed to be constant. The model's endogenous variables consist of labor allocation $\{L_n^{j,K}\}_{n=1,j=1}^{N,J}$, and the allocation of trade and expenditures for manufacturing $(x_{ni}^{j,M}, X_n^{j,M})$ and retail goods $(x_{ni}^{j,R}, X_n^{j,R})$. All prices can be expressed with respect to wages. The equilibrium can then be defined as a vector of wages that the optimality conditons and market clearing conditions are satisfied, resulting in the endogenous allocations. Appendix C provides the formal definition of this competitive equilibrium.

When studying an e-commerce shock that impacts certain economic fundamentals $\Psi' \subseteq \Psi$, determining the equilibrium necessitates solving a complex system of nonlinear equations that depend on all other fundamentals and parameters. This task is particularly challenging in models with extensive spatial dimensions. Following Dekle et al. (2008) and as detailed in Appendix C, it is possible to compute comparative statics for all endogenous allocations in terms of changes, thereby eliminating the need to specify every fundamental and parameter. Specifically, let $\hat{x}$ represent the proportional change in any vector from its original value x to a counterfactual value x'. Appendix C explains that, by conditioning on initial allocations, one can solve for new equilibrium allocations in response to changes in a subset of fundamentals Ψ'. This method does not require knowledge of the initial levels of the rest of the fundamentals, as the initial allocations inherently contain this information. [20]

[20] This method essentially represents a targeted moments comparative statics exercise. This method of conducting comparative statics in ratios also reduces the burden of calibrating the elasticity of substitution across intermediate varieties $(\frac{1}{1-\alpha^j})$.

4.5 E-commerce and Equilibrium Outcomes

E-commerce Shock. Applying the theoretical framework, I intend to answer the question: what are the equilibrium implications of an e-commerce shock on the economy, particularly those related to the dispersion of economic outcomes across different regions? As shown in Definition (1), the model equilibrium is conditional on fundamentals, hence addressing this question requires to specify how the economic fundamentals might be affected the e-commerce. In light of the model in this paper, there are three channels through which e-commerce is likely to bear an impact. Firstly, as online shopping eases consumers' search frictions (Goldmanis et al. 2010; Dinerstein et al. 2018), it may alter the across-region demand shifter μ_{ni}^{j} such that online retailers seize a higher demand. Secondly, the rolling-out of fulfillment and distribution facilities of e-commerce giants such as Amazon significantly reduces the shipping costs of consumer goods (Houde et al. 2021), lowering κ_{ni}^{R}.

Welfare Analysis. The general equilibrium effects of an e-commerce shock on welfare across different regions can also be conveniently analyzed in proportional changes. Define the welfare of a region by its real income per capita $W_n = \frac{Y_n/L_n}{P_n}$, where $Y_n = I_n L_n + \Omega_n$ is the total income in a region including trade deficit. Y_n can be further simplified into $Y_n = (\frac{1}{1-\beta_n})\Gamma(\frac{\nu_n-1}{\nu_n})\Phi_n^{1/\nu_n}L_n$. The changes in welfare can then be expressed as $\hat{W}_n = \hat{\Phi}_n^{1/\nu_n}\Pi_{j=1}^{J}(\hat{P}_n^{j,R})^{-\eta_j}$. Using labor market allocation, we can get $\hat{\Phi}_n^{1/\nu_n} = \hat{w}_n^0(\hat{\pi}_n^0)^{\frac{-1}{\nu_n}}$, while expression of retail trade share in equation (26) leads to that $\Pi_{j=1}^{J}(\hat{P}_n^{j,R})^{-\eta_j} = \Pi_{j=1}^{J}(\hat{x}_{nn}^{j,R})^{\frac{-\eta_j}{\sigma^j-1}}(\frac{\hat{c}_n^{j,R}}{\hat{\mu}})^{-\eta_j}$. Taken together, the counterfactual changes in welfare is:

$$\hat{w}_n^0(\hat{\pi}_n^0)^{\frac{-1}{\nu_n}}\Pi_{j=1}^{J}(\hat{x}_{nn}^{j,R})^{\frac{-\eta_j}{\sigma^j-1}}(\frac{\hat{c}_n^{j,R}}{\hat{\mu}})^{-\eta_j} \tag{17}$$

The above expression of welfare changes highlights several general equilibrium channels that e-commerce could affect an economy with inter-related regions and sectors as well as elastically supplied labor. The term $\Pi_{j=1}^{J}(\hat{x}_{nn}^{j,R})^{\frac{-\eta_j}{\sigma^j-1}}(\frac{\hat{c}_n^{j,R}}{\hat{\mu}_{nn}^{j}})^{-\eta_j}$ comes from the changes in consumer retail good price index aggregated across sectors $\Pi_{j=1}^{J}(\hat{P}_n^{j,R})^{-\eta_j}$ and captures the price effects of a shock. Such effects depend on the consumer expenditure share on a region's local goods $\hat{x}_{nn}^{j,R}$, and a negative power term that comprises the elasticity across retailers σ^j as well as consumers' expenditure shares η^j, both varying at the sector level. A region's expenditure share of its own good and the trade elasticity represent the sufficient statistics for welfare change in a wide variety of trade models, as discussed in Arkolakis et al. (2012). By shifting demand towards non-local retailers and reducing transportation friction, the rise of e-commerce will increase welfare through this price channel; adding to that, sectoral heterogeneity in trade elasticities and consumer's expenditure share also matters for welfare in this model.

Two additional terms appear in the composition of the price effects. First, the change in unit cost of local retail good production $\hat{c}_n^{j,R}$ affects the local retail price positively conditional on changes in trade share of a region's own goods. Hence, consumers benefit from reduction in the price of local retail goods if it doesn't alter the trade share of local goods, and note from equation (25) that such effect also depends on the input-output linkages. As the price of intermediate inputs decreases, the price of retail goods will also drop depending on the value-added share γ_n^j. Second, the increase in preference for local goods $\hat{\mu}_{nn}^j$ reduces local retail prices conditional its effects on expenditure share on local goods, but since tastes for local goods also affect the trade share $\hat{x}_{nn}^{j,R}$, the total effect on welfare depends on the magnitude of their changes. For both $\hat{c}_n^{j,R}$ and $\hat{\mu}_{nn}^j$, their effects on welfare changes and on $\hat{\mu}_{nn}^j$ are negatively correlated, hence counterbalances the local expenditure share in determining welfare variation.

With worker heterogeneity in labor supply and imperfect mobility across regions, employment rate across sectors will also affect households' welfare. The term $\hat{w}_n^0(\hat{\pi}_n^0)^{\frac{-1}{\nu_n}}$ represents the income effects on welfare conditional on price changes, and indicates that as non-employment rate decreases or wage return for non-employment rises, welfare will tend to increase. Since the change in total income can be is positively correlated with the change in wage and negatively correlated with the change employment of any sector $\hat{\Phi}_n^{1/\nu_n} = \hat{w}_n^{j,K}(\hat{\pi}_n^{j,K})^{\frac{-1}{\nu_n}}, \forall(j,K)$, as shown in Galle et al. (2022), this implies that welfare will increase with the degree of specialization of workers. Therefore, regions with workers that have a comparative advantage in the sectors exporting more due to the e-commerce shock will see increases in welfare, while regions that loose jobs due to competition from elsewhere will see reductions in welfare. Taking stock, by explicitly capturing heterogeneous labor supply, and demand shift related to search transportation friction, the model delivers comparative statistics regarding welfare that are comprehensive of the general equilibrium mechanisms through which e-commerce affect different regional economies.

5 Model Quantification and Counterfactual Analysis

In this section, I discuss the quantification of the model to evaluate the impact of e-commerce on regional economies. I first explain the data and measurement with respect to the general economic environment, specifically the fundamentals and parameters necessary to bring the model to the data. I then consider the rise of Amazon as a salient case of e-commerce shock, and discuss how to quantify its impacts on the fundamentals of the model. Counterfactual analysis on regional economic outcomes are presented afterwards.

5.1 General Environment

To study the impact of e-commerce on regional economies, I consider 2007 as the baseline economy since only after then the online sales of Amazon started to pick up, and I consider 2017 as the post-Amazon shock equilibrium economy. The model is fit to the data and variables on 50 U.S. states and 2 tradable good sectors (durable, non-durable), service sector, as well as a non-employment sector. For each of the tradable sector, there are three subsectors: manufacturing, online retail, and brick-and-mortar.[21] In the model, a labor market is a region and sector pair, which implies that there are 400 markets in the quantification. Table 4 lists for each model section, the parameters, fundamentals and shocks that need to be calibrated or estimated, and the sources of information, which I discuss below. Appendix A provides further details for the calibration of some of the parameters.

Consumption I calibrate the expenditure shares for durable, non-durable, and service sector goods η^j using regional consumption data from the Bureau of Economic Analysis (BEA). An essential parameter in this calibration is the elasticity of substitution between different retailers (σ^j). Based on the gravity trade flows described in equation (10), I derive a reduced-form equation in log differences as in (18). This equation captures the changes in bilateral trade shares as a function of changes in origination prices and iceberg costs. Origination prices are obtained from the Commodity Flow Survey (CFS) using the per unit price, calculated as shipment value divided by shipment weight, for shipments within the region, thus excluding bilateral iceberg costs. Additionally, I adjust for changes

[21] In online Appendix Table 1, I show the allocation of 3-digit manufacturing sectors according to the North American Industry Classification System (NAICS) into durable and non-durable sectors. The breakdown by durability of online retail and brick-and-mortar sectors is discussed as below.

Table 4: Parameters, Fundamentals and Shocks for Model Quantification

Section	Param.	Description	Estimation/Caliberation
Consumption	η_n^j	Sector share of consumption	CFS 2007
	σ^j	Elasticity of subs. across retailers	Keepa + IV
Labor Supply	π_n^j	Share of empployment	CBP, ACS
	v^n	Fréchet shape of worker product.	Galle et al. (2022)
Production	β_n^j	Share of structures	BEA, Greenwood et al. (1997)
	θ^j	Fréchet shape of sector product.	Caliendo and Parro (2015)
	γ_n^j	Value-added share of retail goods	BEA, CFS
Expenditure	$x_{ni}^{M,j}$	Interm. expenditure share	CFS 2007
	$x_{n0}^{R,j}$	Brick-and-motar expenditure share	CFS 2007, E-Stats
	$x_{ni,i\geq 1}^{R,j}$	E-commerce expenditure share	CFS 2007, E-Stats
	p_{n0}^j	Brick-and-motar price index	CFS 2007, E-Stats, CES
Amazon Shock	$\hat{\kappa}_{nm}^R$	Iceberg cost change	Amazon data + CFS 2007 + IV
	μ	Matching efficiency	E-stats + CES
	Ψ_m^j	Online retailer location probability	Keepa
	O	Measure of online retailers	E-stats
	T_n^j	Fréchet scale of sectoral product.	Assume constant
	A_n^j	Fréchet scale of labor product.	Assume constant

in shipment distance, where the variation in iceberg cost is dependent on these distances.

$$\Delta \ln(x_{nm}^{j,R}) = \delta + (1-\sigma)\Delta \ln(c_n^{j,R}) + \Delta \ln(\kappa_{nm}^R) + \epsilon_{nm}^{j,R} \tag{18}$$

The major challenge in using this equation to estimate the elasticity σ^j concerns the endogeneity of prices, as they are influenced by other demand-side factors that also affect quantity. To address this endogeneity, I employ a standard Hausman instrument to isolate prices from region-specific demand shocks, with detailed results in Appendix A. The estimated elasticity of substitution among regional retail goods ranges between 1.8 and 5.0. This range is modestly lower than the elasticity found between brick-and-mortar and online retailers, which is 4.3 as documented by Dolfen et al. (2019), and the elasticity among US commuting zones, which is 5.5 according to Gervais and Jensen (2019).[22]

Labor Supply. On the worker side, the Census County Business Patterns (CBP) data pro-

[22] Hottman (2017) estimates that the elasticity of substitution among different stores within a county is 1.5 using OLS regression and 4.5 once instrumented. Since σ^j represents elasticity of substitution among different regions in bilateral trade, such elasticity is naturally expected to be lower.

vide the employment share by region and sector, denoted as π_n^j. To analyze e-commerce's impact, it is necessary to distinguish between online retail and brick-and-mortar employment, but CBP does not separate them. Therefore, I use E-Commerce Statistics (E-stats) for the sector breakdown of e-commerce versus total retail sales. As E-stats data are national, I first use 2007 Commodity Flow Survey (CFS) data to attribute inter-regional trade flows to the origin of sellers, allowing me to estimate each state's proportion of e-commerce activity. Additionally, I use BEA's value-added data for the entire retail sector to separate total retail sales into regional retail outputs. Combining these datasets, I calculate the ratio of e-commerce retail output to total retail output for each region, which I then use to divide the employment share of the entire retail sector into e-commerce and brick-and-mortar. par For workers' labor supply elasticity, v^n, I adapt the value estimated by Galle et al. (2022), which presents a multi-sector Ricardian model with Roy (1951) type sorting of heterogeneous workers whose productivities similarly characterized by joint Fréchet distributions.[23] Here I specify v^n equal to 1.5, which is the value from their preferred specification.

Production With regard to production, the share of structures in the structure-labor bundle β_n^j can be identified from the value-added share of labor over structure, which equals to $\frac{\beta_n^j}{1-\beta_n^j}$. BEA provides value-added and labor compensation, while Caliendo et al. (2018) derived value-added share of structures to be consistent with the share of capital estimates in Greenwood et al. (1997). I obtain the productivity dispersion parameter θ^j of different sectors directly from corresponding ones in Caliendo and Parro (2015), which used a multi-sector gravity equation to identify the values. For the value-added share of retail goods, γ_n^j, BEA provides the value-added for each sector, which divided by gross-output gives the share value.

Expenditure and Prices. To take the model to the data, I need three expenditure shares: the inter-regional expenditure share on intermediate goods $x_{ni}^{M,j}$, the regional expenditure share on brick-and-mortar retail goods $x_{n0}^{R,j}$, and the inter-regional e-commerce expenditure shares $x_{ni,i\geq 1}^{R,j}$. I obtain $x_{ni}^{M,j}$ directly from the 2007 CFS for durable and non-durable manufacturing goods. I calculate $x_{n0}^{R,j}$ using E-stats data. I first allocate national

[23]In their model, worker differ not only by region and sector, but also by groups that can be categorized by education level and demographics etc., leading to a more nuanced picture of welfare. Bringing the model to data on U.S. commuting zones and other countries for 13 manufacturing and a nonmanufacturing sector and using a model implied Bartick type identification, they estimate the labor supply elasticity (analogous to v^n) to range from 1.42 to 2.79, which are close to the across occupation elasticities estimated in Burstein et al. (2019) and Hsieh et al. (2019) ranging from 1.2 to 3.44.

e-commerce sales in E-stats to states based on their expenditure share of inter-regional trade in the CFS 2007 data. I then distribute total retail sales to states by using state expenditure shares from the Consumer Expenditure Survey (CES) data for each sector. This method gives me the total retail sales and their division into brick-and-mortar and e-commerce sales for each state, allowing me to calculate of $x^{R,j}_{n0}$. Finally, I determine $x^{R,j}_{ni,i\geq 1}$ by distributing the remaining share of retail sales from e-commerce across different origins using the CFS 2007 data on inter-regional wholesale sales trade of durable and non-durable goods.

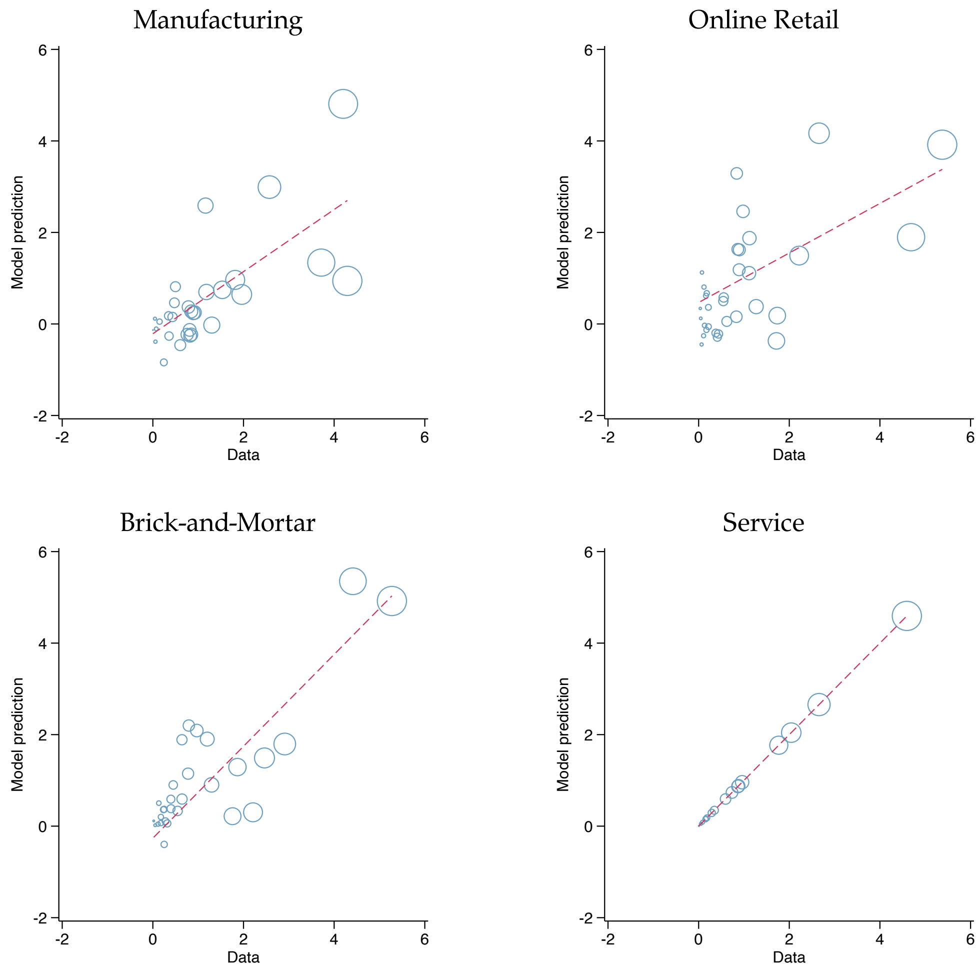
I calibrate the price levels of regional retail goods for 2007 to be used in the estimation inter-regional trade flows post the e-commerce shock, as in equation (10). The Consumer Expenditure Survey (CES) provides the regional price index for various goods across all retail sectors. To specifically determine the price index for brick-and-mortar sellers, I use the ratio of the total e-commerce expenditure share to the brick-and-mortar share as calculated above. This ratio helps isolate the brick-and-mortar price index p^j_{n0} from the overall price index P^j_n in the CES data. Specifically, I calculate $\frac{\sum_i x^j_{ni}}{x^j_{n0}} = \left[\frac{(P^j_n)^{1-\sigma}}{(p^j_{n0})^{1-\sigma}} - 1\right]$.[24] This method imputes the brick-and-mortar price indices to be consistent with their proportionate shares relative to e-commerce within the retail trade model.

Non-targeted Moments. I now report on the baseline equilibrium features of 2007 and show how the model aligns with the U.S. economy for non-targeted moments in the estimation process. The model is constructed to precisely match regional expenditure shares across sectors and between brick-and-mortar versus online retailers, as well as the value-added share in production. Given an initial guess, I apply the market clearing condition (16) to calculate the regional labor value added for each sector. Figure 5 illustrates the spatial variation in labor value added as predicted by the model (y-axis) and as observed in the BEA data (x-axis), with the size of each circle representing the observed value added.[25] As shown, the model effectively captures the regional value added based on the calibrated expenditure and production parameters. For the service sector, which is modeled as a fixed share of total income and directly taken from BEA data, the model matches the data perfectly.

[24]Specifically, this relationship is expanded as $\frac{\sum_i x^j_{ni}}{x^j_{n0}} = \frac{\sum_{i=1}^{N} (\frac{p^j_{ni}}{\mu})^{1-\sigma}}{(p^j_{n0})^{1-\sigma}} = \frac{(P^j_n)^{1-\sigma} - (p^j_{n0})^{1-\sigma}}{(p^j_{n0})^{1-\sigma}} = \left[\frac{(P^j_n)^{1-\sigma}}{(p^j_{n0})^{1-\sigma}}\right] - 1.$

[25]Since the BEA data does not segregate retail value added between online and brick-and-mortar retailers, I employ a similar method used in estimating the ratio of e-commerce retail output to total retail output to impute regional employment share π^j_n. This output share is used to seperate the BEA retail sector value added and serves as the initial guess for the model's imputation.

Figure 5: Predicted and Observed Sectoral Value Added in 2007



5.2 Sequential Estimation of the Amazon Shock

In this section, I discuss the sequential estimation of the Amazon shock to be integrated into the model quantification. The rise of Amazon influences two key model fundamentals from the consumers' perspective: the iceberg cost in the retail goods sector ($\kappa_{ni}^{j,R}$) and the match efficiency between consumers and online retailers (μ). For the iceberg cost, I estimate it as a function of distance, which allows me to compute changes in iceberg costs by estimating reductions in shipping distances due to Amazon's expansion. To isolate changes in bilateral shipping distances resulting from Amazon's expansion from other demand factors, I use exogenous geographic factors as instruments for the observed shipping distances. Lastly, with the estimated iceberg costs, I impute the change in search efficiency based on the observed relative expenditure shares between e-commerce and brick-and-mortar sellers.

Step 1: Relating Iceberg Costs to Shipping Distance. I begin by estimating the empirical relationship between iceberg costs and bilateral shipping distances. Equation (19) presents the reduced-form specification of iceberg costs for retail goods, $\kappa_{nm}^{j,R}$, and the shipping distance between origin m and destination n, alongside fixed origin and destination characteristics and other bilateral characteristics X'_{nm}. To compute iceberg costs, I use the ratio of trade share of destination n with origin m relative to region n's brick-and-mortar consumption share $\frac{x_{nm}^{j,R}}{x_{n0}^{j,R}}$. This ratio is affected by the measure of online sellers in origin m, the cost of retail goods in both m and n, the bilateral iceberg costs, and the elasticity of substitution. By taking the logarithm of this ratio, these components become additively separable. In the reduced-form specification of equation (19), with origin and destination fixed effects included, the only residual variation on the left-hand side is the bilateral iceberg cost. Table 7 presents the results on the elasticity of iceberg cost to shipping distance δ^j from equation (19), showing an elasticity of 1.5 for durable goods and 2.1 for non-durable goods.

$$ln(\kappa_{nm}^{j,R}) = \delta^j \text{Distance}_{nm} + X'_{nm}\theta + \delta_n^j + \delta_m^j + \epsilon_{nm}^j. \tag{19}$$

Step 2: Shipping Distance Reduction. I estimate the reduction in shipping costs by analyzing the expansion of Amazon's facilities and imposing detailed micro-structures on the fulfillment order flows. Following Houde et al. (2021), which shows that over 90 percent of orders are fulfilled by the three nearest centers to the destination, I further specify that the nearest center to the origin handles the order.[26] Table 5 Panel A presents the reduction in shipping distances resulting from the rollout of Amazon's fulfillment and distribution facilities. In 2007, an order handled by Amazon facilities traveled an average of 490 kilometers. By 2017, this average distance decreased to 288 kilometers, indicating a reduction of 202 kilometers on average, or 0.5 in log units. *Identification Strategy:* Using the actual roll-out of Amazon's facilities to calibrate the shock is subject to key endogeneity issues, specifically, the location of that the new facilities expand to are correlated with GDP growth, population changes and other demand side factors that could directly affect the outcomes of interest. To overcome the endogeneity issue in Amazon's expansion, I build counterfactual distribution centers with simulated location choices based solely on

[26]Houde et al. (2021) applies a probit model of order assignment as $\tau_{ni,f} = \Phi(\alpha_1 d_{fn} + \alpha_2 d_{fi} + \alpha_3 k_f)$. The probability that a facility f processes an order from region i to n, $\tau_{ni,f}$, depends on three factors: the distance from the facility to i and n as well as the capacity of facility f. Therefore, for any order that originates in i and ends up in n, a vector of probabilities represents the chances that it is handled by each of Amazon's facilities. The parameters are then estimated by specifying the labor demand of the facility and matching it to the data.

Table 5: Transportation Cost Reduction via Amazon Facilities

	Mean	Std. Dev.	P25	P75	Corr.
Panel A. Actual Amazon Facility					
2007	490.2	376.3	234.9	739.0	–
2017	287.9	225.6	124.7	409.0	–
Log Diff.	−0.5	0.6	−0.9	0.0	–
Panel B. Counterfactual Amazon Facility					
2007	623.4	400.3	349.6	897.4	0.10
2017	335.2	278.4	143.9	412.1	0.58
Log Diff.	−0.7	0.8	−1.1	0.0	−0.02

Table 6: Probability of Amazon Facilities on Geographic Cost Factors

		Dependent: 1{AMZ Center}
Temperature (Lag)	Mean	-0.011
	Minimum	-0.002
	Maximum	0.046***
Precipitation (Lag)	Mean	-0.032
	Minimum	0.043
	Maximum	-0.015
Elevation	Mean	-0.001***
	Minimum	0.000
	Maximum	0.001***
Tornado	Magnitude	-0.051
	Injuries	-0.110
Year FE		X
Observations		55,259
Psudo R-squared		0.1663

plausibly exogeneous geographic cost factors that are orthogonal to demand-side factors (Duflo and Pande 2007; Lipscomb et al. 2013).[27] To implement such identification strategy, I obtain county-level geographic characteristics on land elevation and climate changes from Open Topography Global Datasets, as well as National Centers for Environmental Information (NCEI).

For the construction of counterfactual amazon facilities, a "budget" for Amazon of

[27] Alternatively, I may leverage the changes in sales tax collection on Amazon, the so called "Amazon tax", or the nexus tax laws imposed by different states requiring sales tax collection where Amazon maintains a physical presence to identify the impact of e-commerce (Baugh et al. 2018; Houde et al. 2021). The major challenge for this kind of identification strategy is whether those places that passed these laws are plausibly comparable to those that did not.

a particular year is determined based on the observed number of new facilities built in that year. Then each U.S. county is ranked based exclusively on topographic factors with respect to land elevation, as well as climatic factors that include temperature, precipitation and number of extreme weathers. The counties that rank the highest according to these factors will be assigned distribution facilities first depending on the budget for each year. Table 6 shows this cross-sectional probit regression of an indicator whether a county was assigned an Amazon facility on different geographic factors. The observed facility expansion pattern prefers locations that are warmer and locate on a slightly lower-elevated land. Precipitation and number of tornados are negatively correlated with facility construction, though not significantly.[28]

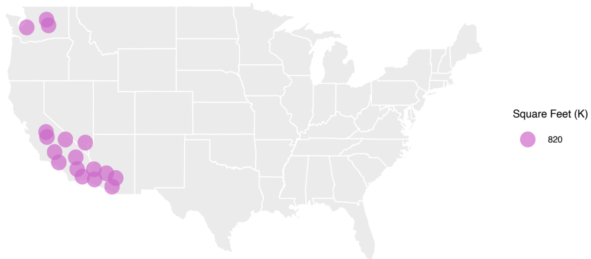
Figure 6 shows the counterfactual centers based solely on geographic factors and Amazon's budget, which exhibit both similarities and divergences in geographic patterns compared to actual facility locations. In earlier years, counterfactual analysis shows a preference for locations in California over those actually chosen; from 2010 to 2020, the counterfactual and actual locations align more closely, although states like California, Texas, Arkansas, and North Carolina are still more preferred based on geographic factors alone. Table 5 reveals that the average distance an order travels between regions was 623 kilometers in 2007, reducing to 338 kilometers in 2017—a decrease of 288 kilometers, or 0.7 in log units.[29] A major concern with using these geographic cost factors to predict facility location is their relevance. Appendix Table A1 displays first-stage regression results of actual shipping distances on predicted ones, indicating a strong correlation with relatively large F-statistics. Moreover, I show The correlation between the counterfactual shipping distance and lagged GDP and GDP growth is either weak or negative, supporting the robustness of the instrument against demand-side factors.

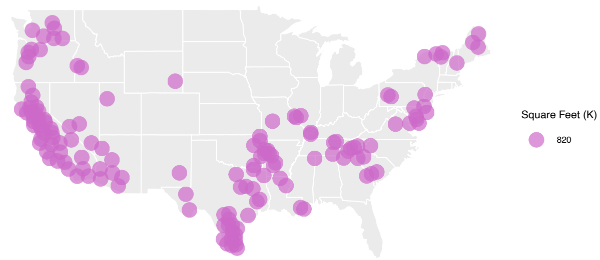
Step 3: Demand Shift. Another key aspect that Amazon affects the economy is through increased online matching efficiency, hence turning the demand more towards non-local retailers. As shown in the theoretical section, the match efficiency channel is reflected as a demand shifter μ in the nested Cobb-Douglas and CES consumption function. Taking the first order condition of the consumption function, log-linearize and take the difference between the initial value and its change due to Amazon, I obtain equation (20) that relate

[28]As a robustness check, the bottom of the table shows that the spearman rank correlation between the suitability index of distribution facility location and GDP growth is significantly negative, corroborating that exogeneity of the instrument with respect to demand-side factors related to economic growth.

[29]The greater reduction in shipping distance via counterfactual centers compared to actual ones over the same period can be attributed to the more densely distributed shipping centers in earlier years and more dispersed locations later on.

Figure 6: Location of Counterfactual Fulfillment and Distribution Centers





the changes in retail expenditure share $x_{ni}^{k,R}$ to the changes in cost of retail goods $c_i^{k,R}$ and in transportation cost as well as demand alteration, κ_{ni}^R and μ,

$$\Delta ln(x_{ni}^{Amz,k}) = (1-\sigma^j)\Delta ln(c_i^{Amz,k}) + (1-\sigma_k)[\Delta ln(\kappa_{ni}^{Amz}) + \Delta ln(\mu)] + \delta_n^k + v_{ni}^k. \quad (20)$$

This equation has an intuitive interpretation – it shows that conditional on cost of goods changes, the changes in consumer expenditure share within a region is either due to transportation cost variation or shifts of demands towards different retailers. Since we've estimated the transportation cost change from the last step, and the cost of good change can be directly obtained from CFS data, it is straight forward that the changes in match efficiency induced demand shift $\Delta ln(\mu)$ can be recovered as long as the changes in consumer expenditure shares due to Amazon $\Delta ln(x_{ni}^{k,R})$ is known. To measure the expenditure share change, the exogeneous changes in transportation costs through the constructed counterfactual centers appear to be useful. On the one hand, the transportation cost serves as an exposure measure of online sales and credible predictor of consumer expenditure shifts due to Amazon; on the other, the counterfactual variation in shipping distance is orthogonal to other demand factors that potentially affect other outcome variables.

Equation (20) displays the predictive regression where c_i^k is the csot of sector k goods produced in i, $\hat{\kappa}_{ni}^{Amz}$ is the instrumented transportation reduction induced by Amazon's facility expansion, $Am\hat{z}Sale_n^k$ is the Amazon region-sector level sales, and Z_{ni} is the average demographics for pairs of regions. Since Amazon's regional sales could also be endogenous to other factors affecting the outcome, it is instrumented in a Hausemen approach. Then consumers' predicted expenditure share variation due to Amazon can then be recovered from the estimated coefficients as $\Delta x_{ni}^{Amz,k} = \delta\Delta\hat{\kappa}_{ni}^{Amz} + b_2\Delta(\hat{\kappa}_{ni}^{Amz} Am\hat{z}Sale_n^k)$. Prediction results show that Amazon's expansion in 2007-2017 predicts a 4.8 percent growth in consumers' expenditure share on non-local goods, which

Table 7: Estimates of Iceberg Cost Change and Demand Shift

δ^{dur}	δ^{nondur}	$\hat{\kappa}$	μ
1.5	2.1	0.97	1.27
[0.2]	[0.6]	[0.15]	[1.46]

compared the actual expenditure share change of 16 percent on average, indicates that Amazon alone accounts for about 30 percent of the total increase in the purchase of non-local goods. Using $\Delta x_{ni}^{Amz,k}$ and $\Delta \hat{\kappa}_{ni}^{Amz}$, the estimate of demand shift is shown in Table 7. On average, consumers become about 27 percent more like to purchase from online retailer due to the growth of Amazon during 2007-2017.

6 The Impact of Amazon on Regional Economies

In this section, I evaluate the impact of Amazon's expansion on the aggregate and regional economies. The counterfactual question that I ask is that starting from the initial equilibrium, only the Amazon shock as embodied in iceberg cost change, match efficiency increase, as well as online retailer location probability, keeping all other fundamentals constant, what are the impacts on aggregate and regional welfare and employment? To answer this question, I take the calibrated parameters and fundamentals as well as estimated Amazon shocks to the model to conduct counterfactual analysis. I also decompose different channels and compare which margin accounts more for the total effects.

Welfare: Starting with the changes in welfare induced by the Amazon shock. On average, states see an increase in total welfare of 6.7 percent. The growth in total welfare on average is driven by the price effects, while the income effect has a negative impact on welfare. Leaving only to price effect, which is a result of price decline due to the Amazon expansion, total welfare would have increased by 13.1 percent. Mitigating the consumption benefit is the fact that Amazon's expansion also leads to the reallocation of economic activities, as well as of workers, changing the income level differentially across regions. The effect on welfare due to income changes would have decreased total welfare by 5.4 percent without the compensating price changes.

Underlying the overall welfare changes is a significant dispersion across regions, as depicted in Figure 7, which illustrates state-level changes in total welfare broken down into price and income effects. Generally, states on the East and West Coasts experience larger increases in welfare, while Midwestern states see smaller gains. States with a

Figure 7: Welfare, Employment Changes and Decompositions

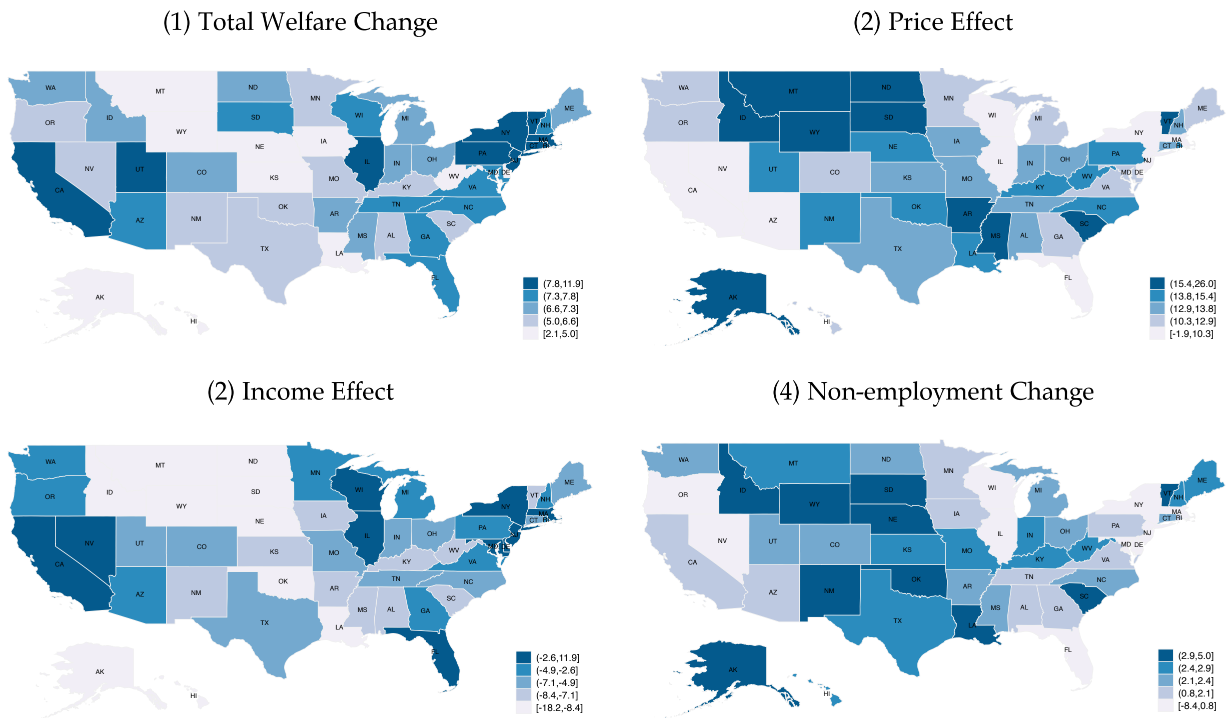


Table 8: Employment Changes by Sector and State Groups

	All States		Below 75th Percentile Online Sales Density		Below 50th Percentile Online Sales Density	
Sector	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Manufacturing	-4.3	(7.6)	-2.9	(6.0)	-1.8	(1.1)
Online Retail	109.8	(97.8)	87.6	(73.8)	63.3	(64.8)
Brick-and-Mortar	-11.1	(8.0)	-9.8	(6.2)	-8.6	(1.2)
Service	-1.6	(7.9)	-0.1	(6.2)	1.2	(1.2)
Non-Employment	-1.3	(8.1)	0.3	(6.3)	1.7	(0.8)

comparative advantage in online retailing, such as New York, Massachusetts, Wisconsin, California, and Florida, benefit from positive income effects due to employment gains, which drive their overall welfare increases. These states also have a more diversified industrial composition, which contributes to wage increases across all sectors. Conversely, Midwestern states like North Dakota, Montana, and Wyoming face negative income effects from business displacement. However, these regions start with lower consumer expenditures on online retail goods, which allows them to benefit from significant positive price effects that help offset the losses in income.

Employment: I now turn to discuss the employment changes implied by the Amazon shock and the model. Table 8 illustrates the average sectoral employment changes due to the Amazon shock in percentages. As can be seen from the table, the overall picture of employment changes due to Amazon is characterized by reallocation from all other sectors to the growing online retail sector, particularly from brick-and-motor and manufacturing. Non-employment has also declined by 1.3 percent. Since in 2007 the average non-employment rate was 38.5 percentage points, which implies that non-employment has declined by 0.5 percentage points due to the Amazon shock.

Beneath the overall increase in non-employment, there is substantial regional variation. States in the Midwest, which have less comparative advantage in online retailing and a less diversified industrial structure, exhibit a higher shift towards non-employment and service sectors. The last two columns of the data indicate that in states where online seller density falls below the 50th percentile, online retail employment constitutes about 60 percent of the average rate, and there has been an increase in service sector and non-employment rates. The Gini index of non-employment has increased from 0.11 to 0.13, a 20 percent growth. This implies that the gap in employment opportunities has become wider due to Amazon.

7 Alternative Modeling of Online Retail Location

An alternative modeling approach is to follow Chaney (2008) to understand the dynamics of retailer entry and trade. We begin by considering the productivity distribution of retailers, represented by a Pareto distribution: $P\left(Z^j < z\right) = G^j(z) = 1 - z^{-\rho}$. Retailers decide to enter the market based on a profitability condition: the expected revenue must be greater than or equal to the costs of entry. This condition is given by $\sum_n \left(\frac{p_{nm}^{j,R}/\mu}{P_n^{R,j}}\right)^{1-\sigma} \eta^j Y_n \geq \sigma w_m^{j,R} f_m$. The threshold for entry denoted as $\bar{c}_m^j$ is then given by [30]

$$\bar{c}_m^j = \frac{\mu}{\widetilde{\sigma}} \left(\frac{\sigma}{\eta^j}\right)^{\frac{1}{1-\sigma}} \left[\frac{w_n^{j,R} f_m}{\sum_n \left(\kappa_{nm}^R / P_n^{R,j}\right)^{1-\sigma} Y_n}\right]^{\frac{1}{1-\sigma}}. \tag{21}$$

The trade flow equation, can then be derived to link to the relative productivity and

[30]Since $\bar{c}_m^j = \frac{1}{z_m^j}\left(w_m^{j,R}\right)^{\gamma^j}\left(P_m^{j,M}\right)^{\left(1-\gamma^j\right)}$, we can also drive the threthold productivity $\bar{z}_m^j = \left(w_m^{j,R}\right)^{\nu_i^j}\left(P_m^{j,M}\right)^{\left(1-\gamma_i^j\right)} \frac{\tilde{\sigma}}{\tilde{\mu}}\left(\frac{\sigma}{\eta^j}\right)^{\frac{1}{\sigma-1}}\left[\frac{w_m^{j,R} f_m}{\sum_n \left(\kappa_{nm}^R P_n^{R,j}\right)^{1-\sigma} Y_n}\right]^{\frac{1}{\sigma-1}}$

cost structures of the trading regions. The bilateral export $X_{nm}^{j,R}$ from region m to n is a function of wage rates, productivity, and the relative costs of retailing and manufacturing, as below.[31] This equation below suggests that an increase in the productivity or a decrease in the wage rate of the exporting region (region m) would lead to an increase in exports $X_{nm}^{j,R}$ to region n, all else being equal. Similarly, an improvement in the transportation technology (represented by κ_{nm}^{R}) would increase the trade flow.

$$X_{nm}^{j,R} = \lambda w_m^{j,B} l_m^{j,B} \left(\left(w_m^{j,R} \right)^{\gamma^j} \left(P_m^{j,M} \right)^{\left(1-\gamma^j\right)} \frac{\kappa_{nm}^R}{\mu} \right)^{-\rho} \times \left[\frac{w_m^{j,R} f_m}{\sum_n \left(\frac{\kappa_{nm}^R}{P_n^{R,j}} \right)^{1-\sigma} Y_n} \right]^{\frac{\sigma-\rho-1}{1-\sigma}} \eta^j Y_n \left(P_n^j \right)^{\sigma-1}. \tag{22}$$

Furthermore, local brick-and-mortar (BM) sales $X_{nn}^{j,B}$ in region n are also modeled, capturing the local market dynamics. This equation considers the local wage rates and productivity, and the price index P_n^j:

$$X_{nn}^{j,B} = \left(\left(\omega_n^{j,B} \right)^{\gamma^j} \left(P_n^{j,M} \right)^{\left(1-\gamma^j\right)} \right)^{1-\sigma} \eta^j Y_n \left(P_n^j \right)^{\sigma-1} \tag{23}$$

Finally, the expenditure share of region m in n, both for retail and BM sales, is derived from these trade flow equations, as well as the price index P_n^j. These shares reflects the distribution of consumption across different regions and sectors. For a more detailed mathematical exposition of the derivation, please refer to the Appendix.

The alternative modeling approach presented here offers a different view of the retail market dynamics. Rather than thinking of online retailing location as choice of onlinre retailers based on multivariate Pareto distribution, the entry model of Chaney (2008) represents online retail entry based on regional conditions. As detailed in the Appendix, the comparative statics based on this alternative model is also different and requires additional calibration of the ice-berg cost in the first period, price index of entire retail sector, and the change in online retail efficiency. Despite these differences, quantitative results based on this Alternative model are consistent with the main model.

[31] In this equation $\lambda \equiv \tilde{\sigma}^{-\rho} \left(\frac{\sigma}{\eta^j \mu^{1-\sigma}} \right)^{\frac{\sigma-\rho-1}{1-\sigma}} \frac{-\rho}{\sigma-\rho-1}$

8 Conclusion

The rapid expansion of e-commerce, as exemplified by Amazon's growth, has brought about significant shifts in regional economies and labor markets. This paper's findings show the noticeable effects of online retailing on spatial economic disparities. In particular, while e-commerce has led to a general decline in retail prices, benefiting consumers, there has been a noticeable worker reallocation away from manufacturing sectors, contributing to a 1 percent average decrease in welfare. While some regions have reaped the benefits of increased trade and economic activity, others have faced challenges. This results in an overall increase in regional inequality, and indicates important redistribution effects of e-commerce. This paper's results imply that the growth of e-commerce, despite its benefits in terms of efficiency and consumer choice, requires careful policy considerations to reduce adverse impacts on less advantaged regions and sectors.

References

Acemoglu, D. (2007). Equilibrium bias of technology. *Econometrica*, 75(5):1371–1409.

Acemoglu, D. and Autor, D. (2011). Skills, tasks and technologies: Implications for employment and earnings. In *Handbook of labor economics*, volume 4, pages 1043–1171. Elsevier.

Adao, R., Arkolakis, C., and Esposito, F. (2019). General equilibrium effects in space: Theory and measurement. Technical report, National Bureau of Economic Research.

Amior, M. and Manning, A. (2018). The persistence of local joblessness. *American Economic Review*, 108(7):1942–70.

Anderson, S., Engers, M., and Savelle, D. (2022). An equilibrium analysis of ordered search. *Working Paper*.

Antràs, P., Fort, T. C., Gutiérrez, A., and Tintelnot, F. (2022). Trade policy and global sourcing: An efficiency rationale for tariff escalation. Technical report, National Bureau of Economic Research.

Arkolakis, C., Costinot, A., and Rodríguez-Clare, A. (2012). New trade models, same old gains? *American Economic Review*, 102(1):94–130.

Arkolakis, C., Ramondo, N., Rodríguez-Clare, A., and Yeaple, S. (2018). Innovation and production in the global economy. *American Economic Review*, 108(8):2128–2173.

Arkolakis, C., Rodrıguez-Clare, A., and Su, J.-H. (2017). A multivariate distribution with pareto tails and pareto maxima. *Working paper*.

Armstrong, M. (2017). Ordered consumer search. *Journal of the European Economic Association*, 15(5):989–1024.

Armstrong, M. and Vickers, J. (2015). Which demand systems can be generated by discrete choice? *Journal of Economic Theory*, 158:293–307.

Austin, B., Glaeser, E., and Summers, L. (2018). Jobs for the heartland: Place-based policies in 21st-century america. *Brookings Papers on Economic Activity*, page 151.

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

Autor, D. H., Katz, L. F., and Krueger, A. B. (1998). Computing inequality: have computers changed the labor market? *The Quarterly journal of economics*, 113(4):1169–1213.

Autor, D. H., Levy, F., and Murnane, R. J. (2003). The skill content of recent technological change: An empirical exploration. *The Quarterly journal of economics*, 118(4):1279–1333.

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

Brynjolfsson, E., Hu, Y., and Simester, D. (2011). Goodbye pareto principle, hello long tail: The effect of search costs on the concentration of product sales. *Management science*, 57(8):1373–1386.

Burstein, A., Morales, E., and Vogel, J. (2019). Changes in between-group inequality: computers, occupations, and international trade. *American Economic Journal: Macroeconomics*, 11(2):348–400.

Caliendo, L., Dvorkin, M., and Parro, F. (2019). Trade and labor market dynamics: General equilibrium analysis of the china trade shock. *Econometrica*, 87(3):741–835.

Caliendo, L. and Parro, F. (2015). Estimates of the trade and welfare effects of nafta. *The Review of Economic Studies*, 82(1):1–44.

Caliendo, L., Parro, F., Rossi-Hansberg, E., and Sarte, P.-D. (2018). The impact of regional and sectoral productivity changes on the us economy. *The Review of economic studies*, 85(4):2042–2096.

Chaney, T. (2008). Distorted gravity: the intensive and extensive margins of international trade. *American Economic Review*, 98(4):1707–1721.

Chetty, R. and Hendren, N. (2018). The impacts of neighborhoods on intergenerational mobility ii: County-level estimates. *The Quarterly Journal of Economics*, 133(3):1163–1228.

Chetty, R., Hendren, N., and Katz, L. F. (2016). The effects of exposure to better neighborhoods on children: New evidence from the moving to opportunity experiment. *American Economic Review*, 106(4):855–902.

Chevalier, J. and Goolsbee, A. (2003). Measuring prices and price competition online: Amazon.com and barnesandnoble.com. *Quantitative marketing and economics*, 203.

Choi, M., Dai, A. Y., and Kim, K. (2018). Consumer search and price competition. *Econometrica*, 86(4):1257–1281.

Costinot, A., Donaldson, D., Vogel, J., and Werning, I. (2015). Comparative advantage and optimal trade policy. *The Quarterly Journal of Economics*, 130(2):659–702.

Costinot, A. and Rodríguez-Clare, A. (2014). Trade theory with numbers: Quantifying the consequences of globalization. In *Handbook of international economics*, volume 4, pages 197–261. Elsevier.

Dekle, R., Eaton, J., and Kortum, S. (2008). Global rebalancing with gravity: Measuring the burden of adjustment. *IMF Staff Papers*, 55(3):511–540.

Dinerstein, M., Einav, L., Levin, J., and Sundaresan, N. (2018). Consumer price search and platform design in internet commerce. *American Economic Review*, 108(7):1820–59.

Dolfen, P., Einav, L., Klenow, P. J., Klopack, B., Levin, J. D., Levin, L., and Best, W. (2019). Assessing the gains from e-commerce. Technical report, National Bureau of Economic Research.

Duflo, E. and Pande, R. (2007). Dams. *The Quarterly Journal of Economics*, 122(2):601–646.

Dvorkin, M. (2014). Sectoral shocks, reallocation and unemployment in competitive labor markets. Technical report, Yale University.

Eaton, J. and Kortum, S. (2002). Technology, geography, and trade. *Econometrica*, 70(5):1741–1779.

Ellison, G. and Ellison, S. F. (2018). Match quality, search, and the internet market for used books. Technical report, National Bureau of Economic Research.

Faber, B. (2014). Trade integration, market size, and industrialization: evidence from china's national trunk highway system. *Review of Economic Studies*, 81(3):1046–1070.

Fan, J., Tang, L., Zhu, W., and Zou, B. (2018). The alibaba effect: Spatial consumption inequality and the welfare gains from e-commerce. *Journal of International Economics*, 114:203–220.

Galle, S., Rodríguez-Clare, A., and Yi, M. (2022). Slicing the Pie: Quantifying the Aggregate and Distributional Effects of Trade. *The Review of Economic Studies*.

Gervais, A. and Jensen, J. B. (2019). The tradability of services: Geographic concentration and trade costs. *Journal of International Economics*, 118:331–350.

Goldmanis, M., Hortaçsu, A., Syverson, C., and Emre, Ö. (2010). E-commerce and the market structure of retail industries. *The Economic Journal*, 120(545):651–682.

Greenwood, J., Hercowitz, Z., and Krusell, P. (1997). Long-run implications of investment-specific technological change. *The American economic review*, pages 342–362.

Hottman, C. J. (2017). Retail markups, misallocation, and store variety across us cities. *Board of Governors of the Federal Reserve System, Working Paper*, 63.

Houde, J.-F., Newberry, P., and Seim, K. (2021). Economies of density in e-commerce: A study of amazon's fulfillment center network. Technical report, National Bureau of Economic Research.

Hsieh, C.-T., Hurst, E., Jones, C. I., and Klenow, P. J. (2019). The allocation of talent and us economic growth. *Econometrica*, 87(5):1439–1474.

Kline, P. and Moretti, E. (2013). Place based policies with unemployment. *American Economic Review*, 103(3):238–43.

Krugman, P. (1991). Increasing returns and economic geography. *Journal of political economy*, 99(3):483–499.

Krugman, P. and Venables, A. J. (1995). Globalization and the inequality of nations. *The quarterly journal of economics*, 110(4):857–880.

Lagakos, D. and Waugh, M. E. (2013). Selection, agriculture, and cross-country productivity differences. *American Economic Review*, 103(2):948–80.

Lee, E. (2020). Trade, inequality, and the endogenous sorting of heterogeneous workers. *Journal of International Economics*, 125:103310.

Lipscomb, M., Mobarak, A. M., and Barham, T. (2013). Development effects of electrification: Evidence from the topographic placement of hydropower plants in brazil. *American Economic Journal: Applied Economics*, 5(2):200–231.

Pozzi, A. (2013). The effect of internet distribution on brick-and-mortar sales. *The RAND Journal of Economics*, 44(3):569–583.

Puga, D. (1999). The rise and fall of regional inequalities. *European economic review*, 43(2):303–334.

Roy, A. D. (1951). Some thoughts on the distribution of earnings. *Oxford economic papers*, 3(2):135–146.

Şahin, A., Song, J., Topa, G., and Violante, G. L. (2014). Mismatch unemployment. *American Economic Review*, 104(11):3529–64.

Stanchi, F. (2019). Creative destruction? the effects of e-commerce on demand and productivity for department stores. *Working Paper*.

Weitzman, M. L. (1979). Optimal search for the best alternative. *Econometrica: Journal of the Econometric Society*, pages 641–654.

APPENDICES

A Additional Empirical Results

Table A1: First Stage Results and Robustness

	Dependent Variables (in Log)	
	Actual distance	Counterfactual distance
First Stage Results		
Counterfactual log distance	0.40***	
	[0.02]	
F-Stats	670	
Robustness		
Avg. lag GDP		0.00
		[0.00]
Avg. GDP growth		-0.00***
		[0.00]
Observations	4,704	2,352
R-squared	0.12	0.04

Notes: This table displays the first-stage regression results analyzing the relationship between actual log shipping distances and counterfactual log shipping distances, as well as how counterfactual log shipping distances are affected by lagged GDP and GDP growth from 2007 to 2017. The location data for actual Amazon distribution and fulfillment facilities are obtained from MWPVL. The counterfactual log shipping distances are derived solely from topographic and climatic factors. The calculation of shipping distances between an origin and a destination state involves computing the distance from the destination state's centroid to the three closest centers, selecting the one nearest to the origin state's centroid, and then adding the distance from this center to the origin state.

Table A2: OLS and IV Estimates for σ

	OLS - Dur (1)	**OLS - Non-dur** (2)	**IV - Dur** (3)	**IV - Non-dur** (4)
$1 - \sigma$	0.05***	-0.02	-0.76**	-4.05**
	[0.02]	[0.03]	[0.36]	[1.69]
Distance control	X	X	X	X
F-Stats			62.0	4.3
Observations	15,183	10,889	15,183	10,889
R-squared	0.00	0.00	-0.10	-2.06

Notes: This table displays regression results for the logarithmic difference in a destination state's expenditure share across various origin states from 2007 to 2017, based on the logarithmic difference in prices, and accounting for changes in shipping distances using CFS data. The regression coefficient for log prices indicates the value of $1 - \sigma$ as per equation 18. For the instrumental variable (IV) specifications, I employ the Hausman instrument, which uses contemporaneous price changes of the same industrial goods in different geographic markets (states) to account for price changes.

B Derivation of Demand Function

Proof of Theorem 1: In a sequential ordered search model, consumers in region n optimally choose or purchase a good from sector j at retailer i where $\omega^j_{ni} - p^j_{ni}$ is maximized. Denoting this demand as D^j_{ni}, it can be expressed as $D^j_{ni} = P(\omega^j_{ni} - \ln p^j_{ni} > \max_g \ \omega^j_{ng} - \ln p^j_{ng}) = \int \Pi_{g\neq i} F_{\omega^j_{ng}}(\epsilon - \ln p^j_{ng}) f_{\omega^j_{ni}}(\epsilon - \ln p^j_{ni}) d\epsilon$. This demand D^j_{ni} equates to a discrete choice model with indirect utility $v^j_{ni} = -\ln p^j_{ni} + \epsilon^{j,DC}_{ni}$ if $F_{\omega^j_{ni}} = F_{\epsilon^{j,DC}_{ni}}$, where $\epsilon^{j,DC}_{ni}$ is the random utility a consumer derives from the retailer.

To transition from a discrete choice model to CES demand, we note that the average ϵ^j_{ni} is zero for brick-and-mortar stores and $\ln(\mu)$ for online retailers. Therefore, we can express $\epsilon^{j,DC}_{ni}$ as $\ln(\mu) + \chi^j \tilde{\epsilon}^j_{ni}$ where $\tilde{\epsilon}^j_{ni}$ has mean zero and unit variance, and χ^j is the variance of the effective match value ω^j_{ni}, assumed to vary across sectors but not regions. The demand then becomes $D^j_{ni} = \int \Pi_{g\neq i} F_{\epsilon^{j,DC}_{ng}}(\epsilon - \ln p^j_{ng}) f_{\epsilon^{j,DC}_{ni}}(\epsilon - \ln p^j_{ni}) d\epsilon$.

Assuming $F_{\omega^j_{ni}} = F_{\epsilon^{j,DC}_{ni}}$ follows an extreme type I distribution, the demand for retailer i if i is an online retailer becomes

$$D^j_{ni} = \frac{(p^j_{ni}/\mu)^{\frac{-1}{\chi^j}}}{\sum_{g=1}^N (p^j_{ng}/\mu)^{\frac{-1}{\chi^j}} + (p^j_{n0})^{\frac{-1}{\chi^j}}}.$$

If i is brick-and-mortar, then

$$D^j_{ni} = \frac{p^j_{n0})^{\frac{-1}{\chi^j}}}{\sum_{g=1}^N (p^j_{ng}/\mu)^{\frac{-1}{\chi^j}} + (p^j_{n0})^{\frac{-1}{\chi^j}}}.$$

Denote the elasticity of substitution among retailers by σ_j, then $\sigma_j = \frac{1+\chi^j}{\chi^j}$. This demand function leads to sector j's demand as $C^j_n = \left[(c_{n0})^{\frac{\sigma-1}{\sigma}} + \mu \sum_{i=1}^N (c_{ni})^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma_j}{\sigma_j - 1}}$. Given that the consumer's expenditure share is controlled by η^j in a Cobb-Douglas manner, the final demand function is $C_n = \Pi_{j=1}^J (C^j_n)^{\eta^j}$.

C Comparative Statics in Hat Algebra

Definition 1 (Competitive Equilibrium). *Given the fundamentals Ψ and labor supply L_n, a competitive equilibrium for this economy is a vector of wages $\mathbf{w} = \{w^j_n\}_{n=1,j=0}^{N,J}$ such that the optimality conditions are satisfied and all markets clear – equations (10), (11), (3), (4), as well as (13) to (16) hold.*

Comparative Statics. Computing the equilibrium outcomes out of the model requires solving a system of nonlinear equations (10), (11), (3), (4), and (13) to (16), which requires pinning down the levels of a large number of fundamentals and parameters. To ease the comparative statics analysis, I adopt the "exact hat algebra" method (Dekle et al. 2008) to characterize the equilibrium variables and solve for the economy in proportional changes, which greatly reduces the number of fundamentals and parameters to identify. Specifically, define $\hat{x} \equiv x'/x$ the relative change of any variable from its original to counterfactual equilibrium values, x and x' respectively. Since e-commerce shocks function in three channels relating to search and transportation frictions and capital capacity, proportional changes in these fundamentals can be expressed as $\hat{\mu}^j_{ni}$, $\hat{\kappa}^R_{ni}$, and $\hat{\rho}^j_n$. The equilibrium in relative changes under the e-commerce shock can be characterized by the following equations.

The share of labor in different sectors is given by:

$$\hat{\pi}^0_n = \frac{\hat{A}^0_n (\hat{w}^0_n)^{\nu_n}}{\hat{\Phi}_n}, \ \hat{\pi}^{j,K}_n = \frac{\hat{A}^{j,K}_n (\hat{w}^{j,K}_n)^{\nu_n}}{\hat{\Phi}_n}, \text{ where } \hat{\Phi}_n = \sum_{h=0}^{J} \sum_{K=M,R} \pi^{K,h}_n \hat{A}^{K,h}_n (\hat{w}^{K,h}_n)^{\nu_n}. \tag{24}$$

The input costs are given by:

$$\hat{c}^{j,M}_n = \hat{\omega}^{j,M}_n, \quad \hat{c}^{j,R}_n = (\hat{\rho}^{j,R}_n \hat{\omega}^{j,R}_n)^{\gamma^j_n} (\hat{P}^{j,M}_n)^{1-\gamma^j_n}, \tag{25}$$

$$\text{where } \hat{\omega}^{j,K}_n = \hat{w}^{j,K}_n (\hat{l}^{j,K}_n)^{\beta_n} = (\hat{w}^{j,K}_n)^{1+\beta_n} (\hat{\pi}^{j,K}_n)^{\frac{(\nu_n - 1)\beta_n}{\nu_n}},$$

$$\text{and } \hat{P}^{j,M}_n = \left(\sum_{i=1}^{N} x^{j,M}_{ni} (\hat{\kappa}^M_{ni} \hat{c}^{j,M}_i)^{-\theta^j} \hat{T}^j_i \right)^{\frac{-1}{\theta^j}}$$

The trade shares are given by:

$$x'^{j,M}_{ni} = x^{j,M}_{ni} \left(\frac{\hat{\kappa}^M_{ni} \hat{c}^{j,M}_i}{\hat{P}^{j,R}_n}\right)^{-\theta_j} \hat{T}^j_i, \quad x'^{j,R}_{ni} = x^{j,R}_{ni} \left(\frac{\hat{\kappa}^R_{ni} \hat{c}^{j,R}_i}{\hat{\mu}^j_{ni} \hat{P}^{j,R}_n}\right)^{1-\sigma^j}, \tag{26}$$

$$\text{where } \hat{P}^{j,R}_n = \left(\sum_{i=1}^{N} x^{j,R}_{ni} \left(\frac{\hat{\kappa}^R_{ni} \hat{c}^{j,R}_i}{\hat{\mu}^j_{ni}}\right) \right)^{\frac{1}{1-\sigma^j}}.$$

Market clearing conditions now become:

$$X_n^{\prime j,R} = \sum_{i=1}^{N} x_{in}^{\prime j,R} \eta^j \left[\sum_{k=0}^{J} \sum_{K=M,R} (\frac{1}{1-\beta_i}) \hat{\rho}_i^{K,k} \hat{w}_i^{K,k} \hat{l}_i^{K,k} \rho_i^{K,k} w_i^{K,k} L_i^{K,k} - \Omega_i \right], \tag{27}$$

$$X_n^{\prime j,M} = \sum_{i=1}^{N} (1-\gamma_i^j) x_{ni}^{\prime j,M} X_n^{\prime j,R}, \tag{28}$$

$$\hat{w}_n^{j,M} \hat{l}_n^{j,M} w_n^{j,M} L_n^{j,M} = \beta_n \hat{X}_n^{j,M}, \quad \hat{w}_n^{j,R} \hat{l}_n^{j,R} w_n^{j,R} L_n^{j,R} = \frac{1}{\hat{\rho}_i^{j,R}} \gamma_n^j \beta_n \hat{X}_n^{j,R} \tag{29}$$

Equations (24)-(27) from above illustrate that given the e-commerce shock $(\hat{\mu}_{ni}^j, \hat{\kappa}_{ni}^R, \hat{\rho}_n^j)$, solving for the equilibrium in proportional changes only requires information on initial allocations ($x_{ni}^{j,K}, X_{ni}^{j,K}, K = \{M, R\}$), value-added and capital capacities ($w_n^{j,K}, L_n^{j,K}, \rho_n^{j,K}, K = \{M, R\}$), exogenous trade deficits (Ω_n), as well as parameters with respect to value-added shares (β_n and γ_n^j), consumption shares (η_n^j), and trade elasticities (σ^j and θ^j). All other equilibrium variables, economic fundamentals, and parameters turn out to be irrelevant for computing real wage changes – this significantly reduces the estimation burden of conducting counterfactual analysis of the e-commerce shock.

D Alternative Modeling Details

The price index P_n^j for sector j in region n is a function of the aggregated price levels of imports from other regions and the local price level for brick-and-mortar (BM) stores. It integrates over all possible productivity levels z above a certain threshold $\bar{z}_m^j$, weighted by the productivity distribution $G(z)$, and sums up contributions from all other regions m to region n. The equation is expressed as:

$$
P_n^j = \left[\sum_{m=1}^{N} Y_m \int_{\bar{z}_m^j} \left(\tilde{\sigma} \frac{\left(w_m^{j,R}\right)^{\gamma^j} \left(P_m^{j,M}\right)^{\left(1-\gamma^j\right)} k_{nm}^R}{\mu z_m^j} \right)^{1-\sigma} dG(z) + \left(\left(\omega_n^{j,B}\right)^{\gamma^j} \left(P_n^{j,M}\right)^{\left(1-\gamma^j\right)} \right)^{1-\sigma} \right]^{\frac{1}{1-\sigma}}
$$

$$
= \left[\sum_{m=1}^{N} Y_m \left(\tilde{\sigma} \left(w_m^{j,R}\right)^{\gamma^j} \left(P_m^{j,M}\right)^{\left(1-\gamma^j\right)} \frac{\kappa_{nm}^R}{\mu} \right)^{1-\sigma} \frac{-\rho}{\sigma-\rho-1} \bar{z}_m^{j\ \sigma-\rho-1} + \left(\left(\omega_n^{j,B}\right)^{\gamma^j} \left(P_n^{j,M}\right)^{\left(1-\gamma^j\right)} \right)^{1-\sigma} \right]^{\frac{1}{1-\sigma}}
$$

$$
= \left[\tilde{\sigma}^{1-\sigma} \frac{-\rho}{\sigma-\rho-1} \sum_{m=1}^{N} Y_m \left(\left(w_m^{j,R}\right)^{\gamma^j} \left(P_m^{j,M}\right)^{\left(1-\gamma^j\right)} \frac{\kappa_n^R}{\mu} \right)^{1-\sigma} \left[\left(w_m^{j,R}\right)^{\gamma^j} \left(P_m^{j,M}\right)^{\left(1-\gamma^j\right)} \frac{\tilde{\sigma}}{\mu} \left(\frac{\sigma}{\eta^j}\right)^{\frac{1}{\sigma-1}} \left[\frac{w_m^{j,R} f_m}{\sum_n \left(\frac{R_{nm}^R}{P_n^{R_j}} \right)^{1-\sigma} Y_n} \right]^{\frac{1}{\sigma-1}} \right]^{\sigma-\rho} \right.
$$

$$
= \left[\lambda \sum_{m=1}^{N} Y_m \left(\left(w_m^{j,R}\right)^{\gamma^j} \left(P_m^{j,M}\right)^{\left(1-\gamma^j\right)} \frac{\left(\kappa_{nm}^R\right)^{\frac{\sigma-1}{\rho}}}{\mu} \right)^{-\rho} \left[\frac{w_m^{j,R} f_m}{\sum_n \left(\frac{\kappa_{nm}^R}{P_n^{R,j}} \right)^{1-\sigma} Y_n} \right]^{\frac{\sigma-\rho-1}{\sigma-1}} + \left(\left(\omega_n^{j,B}\right)^{\gamma^j} \left(P_n^{j,M}\right)^{\left(1-\gamma^j\right)} \right)^{1-\sigma} \right]^{\frac{1}{1-\sigma}}
$$

$$
= \left[\lambda \sum_{m=1}^{N} Y_m \left(\left(w_m^{j,R}\right)^{\gamma^j} \left(P_m^{j,M}\right)^{\left(1-\gamma^j\right)} \frac{\left(k_{nm}^R\right)^{\frac{\sigma-1}{\rho}}}{\mu} \right)^{-\rho} \left[\frac{w_m^{j,R} f_m}{\theta_m^j} \right]^{\frac{\sigma-\rho-1}{\sigma-1}} + \left(\left(\omega_n^{j,B}\right)^{\gamma^j} \left(P_n^{j,M}\right)^{\left(1-\gamma^j\right)} \right)^{1-\sigma} \right]^{\frac{1}{1-\sigma}}
$$

The second part of the model deals with the total exports from region m to n, denoted as $X_{nm}^{j,R}$. This equation calculates the aggregate value of goods from sector j that are exported from region m to region n. The exports are determined by the productivity threshold, wage rates, prices, and sectoral income levels in both the exporting and importing regions:

$$X_{nm}^{j,R} = \int_{\bar{z}_m^j} w_m^{j,B} l_m^{j,B} X_{nm}(\phi) dG(\phi) = \int_{\bar{z}_m^j} Y_m \left(\frac{p_{nm}^j(\phi)}{P_n^j} \right)^{1-\sigma} \eta^j Y_n dG(\phi)$$

$$= \int_{\bar{z}_m^j} Y_m \left(\tilde{\sigma} \frac{\left(w_m^{j,R}\right)^{\gamma^j} \left(P_m^{j,M}\right)^{\left(1-\gamma^j\right)} \kappa_{nm}^R}{z_m^j \mu P_n^j} \right)^{1-\sigma} \eta^j Y_n dG(\phi)$$

$$= Y_m \left(\tilde{\sigma} \frac{\left(w_m^{j,R}\right)^{\gamma^j} \left(P_m^{j,M}\right)^{\left(1-\gamma^j\right)} \kappa_{nm}^R}{\mu P_n^j} \right)^{1-\sigma} \eta^j Y_n \frac{-\rho}{\sigma - \rho - 1} \bar{z}_m^{j\;\sigma-\gamma-1}$$

$$= Y_m \left(\tilde{\sigma} \frac{\left(w_m^{j,R}\right)^{\gamma^j} \left(P_m^{j,M}\right)^{\left(1-\gamma^j\right)} \kappa_{nm}^R}{\mu P_n^j} \right)^{1-\sigma} \eta^j Y_n \frac{-\rho}{\sigma - \rho - 1} \left[\left(w_m^{j,R}\right)^{\gamma^j} \left(P_m^{j,M}\right)^{\left(1-\gamma^j\right)} \frac{\tilde{\sigma}}{\mu} \left(\frac{\sigma}{\eta^j}\right)^{\frac{1}{\sigma-1}} \left[\frac{w_m^{j,R} f_m}{\sum_n \left(\frac{\kappa_{nm}^R m}{P_n^{R,j}}\right)^{1-\sigma} Y_n} \right]^{\frac{1}{\sigma-1}} \right]^{\sigma-\rho-1}$$

$$= \tilde{\sigma}^{-\rho} \left(\frac{\sigma}{\eta^j}\right)^{\frac{\sigma-\rho-1}{1-\sigma}} \frac{-\rho}{\sigma-\rho-1} Y_m \left(\left(w_m^{j,R}\right)^{\gamma^j} \left(P_m^{j,M}\right)^{\left(1-\gamma^j\right)} \frac{\left(\kappa_{nm}^R\right)^{\frac{\sigma-1}{\rho}}}{\mu} \right)^{-\rho} \left[\frac{w_m^{j,R} f_m}{\sum_n \left(\frac{\kappa_{nm}^R m}{P_n^{R,j}}\right)^{1-\sigma} Y_n} \right]^{\frac{\sigma-\rho-1}{\sigma-1}} \eta^j Y_n \left(P_n^j\right)^{\sigma-1}$$

$$= \lambda_2 Y_m \left(\left(w_m^{j,R}\right)^{\gamma^j} \left(P_m^{j,M}\right)^{\left(1-\gamma^j\right)} \frac{\left(\kappa_{nm}^R\right)^{\frac{\sigma-1}{\rho}}}{\mu} \right)^{-\rho} \left[\frac{w_m^{j,R} f_m}{\sum_n \left(\frac{\kappa_{nm}^R m}{P_n^{R,j}}\right)^{1-\sigma} Y_n} \right]^{\frac{\sigma-\rho-1}{\sigma-1}} \eta^j Y_n \left(P_n^j\right)^{\sigma-1}$$

The total BM sales in region n, $X_{nn}^{j,B}$ can then be expressed as:

$$X_{nn}^{j,B} = \left(\frac{p_{nn}^{j,B}}{P_n^j} \right)^{1-\sigma} \eta^j Y_n = \left(\left(\omega_n^{j,B}\right)^{\gamma^j} \left(P_n^{j,M}\right)^{\left(1-\gamma^j\right)} \right)^{1-\sigma} \eta^j Y_n \left(P_n^j\right)^{\sigma-1}$$

Finally, the model considers the expenditure share of region m in region n and how it changes over time that reflects how shifts in variables like wages, prices, and productivity can impact the flow of goods and services between regions:

$$x_{nm}^{j,R} = \frac{\lambda Y_m \left(\left(w_m^{j,R}\right)^{\gamma^j} \left(P_m^{j,M}\right)^{\left(1-\gamma^j\right)} \frac{\left(\kappa_{nm}^R\right)^{\frac{\sigma-1}{\rho}}}{\mu} \right)^{-\rho} \left[\frac{w_m^{j,R} f_m}{\sum_n \left(\frac{\kappa_{nm}^R}{P_n^{R,j}}\right)^{1-\sigma} Y_n} \right]^{\frac{\sigma-\rho-1}{\sigma-1}}}{\sum_h \lambda Y_h \left(\left(w_h^{j,R}\right)^{\gamma^j} \left(P_h^{j,M}\right)^{\left(1-\gamma^j\right)} \frac{\left(\kappa_{nm}^R\right)^{\frac{\sigma-1}{\rho}}}{\mu} \right)^{-\rho} \left[\frac{w_h^{j,R} f_h}{\sum_n \left(\frac{\kappa_{nh}^R}{P_n^{R,j}}\right)^{1-\sigma} Y_n} \right]^{\frac{\sigma-\rho-1}{\sigma-1}} + \left(\left(\omega_n^{j,B}\right)^{\gamma^j} \left(P_n^{j,M}\right)^{\left(1-\gamma^j\right)} \right)^{1-\sigma}}$$

$$x_{nn}^{j,B} = \frac{\left(\left(\omega_n^{j,B}\right)^{\gamma^j} \left(P_n^{j,M}\right)^{\left(1-\gamma^j\right)} \right)^{1-\sigma}}{\sum_h \lambda Y_h \left(\left(w_h^{j,R}\right)^{\gamma^j} \left(P_h^{j,M}\right)^{\left(1-\gamma^j\right)} \frac{\left(\kappa_{nm}^R\right)^{\frac{\sigma-1}{\rho}}}{\mu} \right)^{-\rho} \left[\frac{w_h^{j,R} f_h}{\sum_n \left(\frac{\kappa_{nh}^R}{P_n^{R,j}}\right)^{1-\sigma} Y_n} \right]^{\frac{\sigma-\rho-1}{\sigma-1}} + \left(\left(\omega_n^{j,B}\right)^{\gamma^j} \left(P_n^{j,M}\right)^{\left(1-\gamma^j\right)} \right)^{1-\sigma}}$$