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

Elmer Zongyang Li†

November 29, 2024

Click here for the latest version

**Abstract**

E-commerce exposes consumers to a broader set of goods and retailers, and online retailing is inherently more mobile in space. This paper studies the spatial general equilibrium effects of e-commerce on regional economies, focusing on the redistribution effects. Using a panel of the universe of products and retailers on Amazon, as well as Amazon's facilities, I find that online retailers are more agglomerated in space, especially those using Amazon-fulfilled services or selling durable goods, and that both their locations and proximity to Amazon's facilities are associated with differential trade flows. By integrating consumer search and retailer location choices into a multi-sector quantitative spatial trade model with heterogeneous workers, I show that increases in online matching efficiency and reductions in shipping frictions drive greater online retailer agglomeration, as well as industrial and occupational specialization.

Quantitative analysis shows that Amazon's expansion from 2007 to 2017 increases average state-level welfare by 6.7 percent, primarily through price effects, though Midwestern states face negative income effects. This expansion reallocates workers from manufacturing and brick-and-mortar retail into the online retail sector, reducing non-employment by 0.5 percentage points overall. However, in states with low online retail density, employment shifts more substantially to the service sector and non-employment, with increases of 1.2 and 1.7 percent, respectively. Regional disparities intensify, with the Gini index on welfare rising by 20 percent and non-employment inequality increasing fourfold. Counterfactuals indicate that redistributing regional trade benefits and adjusting online market structures 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 transforms the retail sector, regions across the United States face vastly different economic outcomes. While a town in New Jersey might experience growth in warehouses and manufacturing, a town in Wyoming could primarily suffer from the decline of local brick-and-mortar stores. Existing studies have explored e-commerce's impact on demand, productivity, and markups for physical retail stores (e.g., Goldmanis et al. 2010; Pozzi 2013; Ellison and Ellison 2018) and consumer welfare (Fan et al. 2018; Dolfen et al. 2019). However, little research has addressed the regional inequality and redistribution effects of e-commerce in terms of economic activity and job opportunities. Given that regional economic divergence has significant implications for life outcomes (see Chetty and Hendren 2018; Austin et al. 2018), understanding how e-commerce influences regional inequality is crucial for policy-making.

In this paper, I use a spatial trade perspective to analyze e-commerce's impact on regional economies, incorporating consumer search, retailer location choices, and regions' comparative advantages. E-commerce is characterized by two key features: consumers search online, leading to higher match efficiency, and online retailers are not tied to customer locations, allowing greater mobility. As in Krugman (1991) and Krugman and Venables (1995), this increased mobility induces agglomeration in the online retail sector. In a setting where online retailers act as intermediaries between upstream producers and downstream consumers, the agglomeration of online retailers, combined with higher consumer match efficiency, will lead to greater specialization in both the upstream and retail sectors.

Using a panel dataset of the universe of products and retailers on Amazon, as well as Amazon's fulfillment and distribution facilities, I document five new facts that suggest online retailers are more agglomerated in space and both their locations and the location of Amazon facilities are associated with differential trade flows of tradable goods. First, online retail sales are more geographically concentrated than overall retail sector sales, particularly for items fulfilled by Amazon. Second, online sales of durable and standardized products are more concentrated than sales of non-durable and non-standardized products. Third, unlike traditional retail, online sales have a weaker relationship with corporate taxes and population but are closely aligned with regional truck volumes. Fourth, origin markets with more online retailers export more tradable goods, while destination markets with more online retailers import less tradable goods. Finally, proximity to Amazon's fulfillment facilities is associated with increases bilateral trade flows. I demonstrate that these findings align with a simple micro-structure of how online retailers engage in e-commerce.

Taking these new facts of e-commerce into account, I develop a multi-sector quantitative spatial trade framework focused on intra-regional retailing to examine the impact of e-commerce. The role of e-commerce first emerges as consumers engage in costly sequential search and matching of retailers as in Weitzman (1979), where each retailer is evaluated based on an unknown match value that reflects the efficiency of the online retail platform. This match value, alongside price, determines the utility a consumer gains from purchasing from that retailer, and consumers weigh this utility against the cost of continued search. Moreover, the model includes two types of retailers within a vertical production structure: brick-and-mortar retailers who source intermediate goods to serve only local consumers, and online retailers who choose locations that maximize cost advantages for both sourcing inputs and reaching multiple markets, giving rise to agglomeration incentives.[1] Workers are heterogeneous in productivity and choose employment sectors optimally á la Roy (1951).

Despite its rich micro-foundation, this framework aggregates to an extended gravity trade model with Constant Elasticity of Substitution (CES) demand, capturing the key features of inter-regional retail through e-commerce. The CES demand shifter reflects the match efficiency of the online platform, the measure of variety is related to the presence of online retailers, and the iceberg cost is influenced by the shipping costs of online retailers. This model's gravity equation diverges from traditional forms as it incorporates not only the retail production costs of the origin but also the probability of online retailers being present in a region and the online matching efficiency. Further, it captures the competition between local brick-and-mortar stores and online retailers, and also achieves quantitative tractability by linking directly to observable data on the locations of online retailers.

Using Amazon's rise as a prominent e-commerce shock, I quantify the model with data on Amazon retailers, sales, and facilities. I sequentially estimate two key fundamentals embodying Amazon's impact: reductions in iceberg costs and level in match efficiency. First, I specify iceberg costs as a function of shipping costs, with changes in the latter informing iceberg cost shifts. A key challenge in identifying Amazon-induced shipping cost reductions concerns their potential endogeneity to demand-side factors. To address this, I apply a spatial simulated instrumental variable strategy (Duflo and Pande 2007; Lipscomb et al. 2013; Faber 2014). Instead of relying on the actual Amazon facility locations, I simulate counterfactual distribution center locations based on exogenous

[1] This setup differs from the multinational production framework in Arkolakis et al. (2018), where firms select optimal production locations to serve specific markets; here, retailers choose locations that maximize profitability across all markets, considering both upstream and downstream production processes and costs.

geographic and climatic factors. These simulated locations help instrument the observed decline in shipping frictions and iceberg costs. Given the estimated iceberg cost, the regional expenditure share ratio between local online and brick-and-mortar retailers identifies online match efficiency. My results indicate that Amazon's growth from 2007 to 2017 led to a 3 percent reduction in iceberg cost; meanwhile, consumers in 2017 being 27 percent more likely to purchase from online retailers due to Amazon's match efficiency.

The model counterfactual indicates that the expansion of Amazon and e-commerce has had substantial effects on regional economies, influencing welfare, employment, and regional inequality. On average, state-level welfare increased by 6.7 percent, primarily driven by price reductions resulting from Amazon's expansion. Price effects alone would have increased welfare by 13.1 percent; however, income effects associated with the reallocation of economic activities offset this gain, reducing total welfare by 5.4 percent. States like New York, Massachusetts, Wisconsin, California, and Florida—those with a comparative advantage in online retail and diverse industrial structures—see positive income effects that boost overall welfare. In contrast, states in the Midwest, such as North Dakota, Montana, and Wyoming, face adverse income effects from heightened retail competition and worker shifting to lower-wage sectors, though these are offset by positive price effects from lower initial online retail spending.

Amazon's expansion also drives a significant reallocation of labor across sectors, with worker movement from traditional brick-and-mortar retail and manufacturing to the growing online retail sector. Overall, the model estimates that non-employment declined by 1.3 percent, equivalent to a reduction of 0.5 percentage points from the 2007 non-employment rate of 38.5 percent. However, employment changes varied significantly across regions. Midwestern states, which lack a comparative advantage in online retail and have less diversified industries, saw a more pronounced shift toward non-employment and service sectors. In states where online seller density is below the 50th percentile, online retail employment rose by 46.5 percentage points lower than the overall average increase. These states also experienced higher reallocation to service and non-employment sectors, with increases of 1.2 and 1.7 percent, respectively.

To capture the inequality implications of Amazon's impacts, I examine changes in the Gini index. From 2007 to 2017, Amazon's expansion caused the Gini index of welfare (or GDP per capita) to rise from 0.11 to 0.13, a 20 percent increase. Meanwhile, the Gini index of non-employment increased from 0.05 to 0.25, a fourfold increase. These results suggest that while e-commerce expansion has benefited the economy overall, it has also deepened regional inequality in welfare and employment opportunities.

The widening gaps in economic outcomes due to the rise of e-commerce underscore the need for national-level policy interventions. To address these inequalities, I propose a revenue redistribution policy aimed at equalizing welfare changes across states. This policy reallocates regional nominal incomes to achieve a common welfare level, while remaining budget-neutral by ensuring that the total redistribution equals the net impact of Amazon's expansion. Results indicate that achieving balanced welfare requires an average welfare reduction of 3 percent across states. Redistribution amounts are closely tied to income effects, with states experiencing income gains, such as New York, New Jersey, and Delaware, contributing to states with income losses, such as Wyoming and Montana. While price effects play a smaller role, with weak positive correlation with redistribution. This policy highlights how a budget-neutral approach can address regional disparities caused by e-commerce while maintaining a balanced, though slightly reduced, welfare changes nationwide.

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.

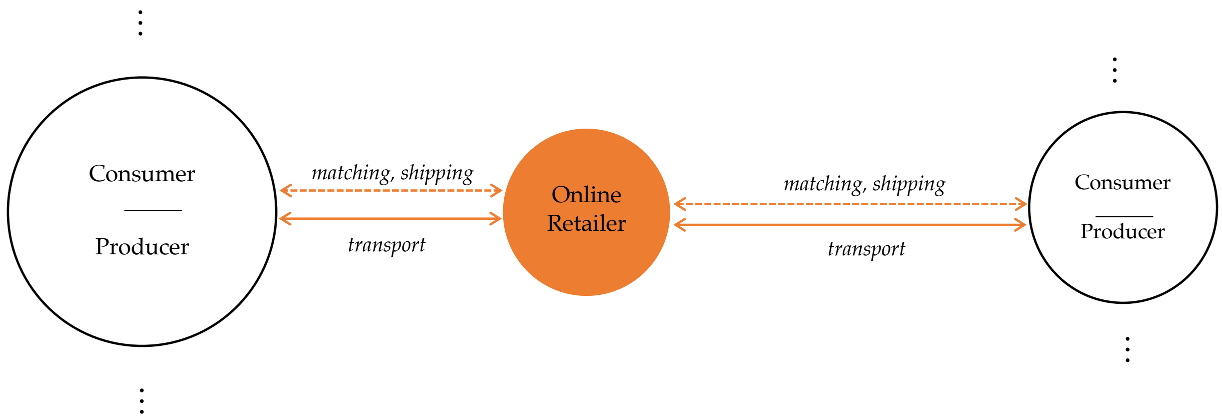
Firstly and most relevant, this work builds on the literature studying the market structure of the retail sector and the impact of e-commerce. Two important findings emerge from this literature. For the retail industry, it is found that e-commerce reduces the demand of the physical department stores, 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

Figure 1: The Online Retail Business Model




*Notes*: This figure illustrates the business model and operational structure of a typical online retailer. It demonstrates the flow from producer to retailer to consumer, with matching and shipping mechanisms. The setup emphasizes the spatial flexibility of online retailers and the separation between their location and that of consumers.

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 Retailer Agglomeration and Trade

In this section, I document several new facts on the agglomeration of online retailers and examine how their locations, along with Amazon facility locations, link to inter-regional trade flows. I begin by outlining a simple micro-structure on how online retailers engage in e-commerce, which has implications for their location choices, agglomeration patterns, and intra-regional retail trade flows. I then introduce the datasets covering the universe of Amazon sellers, products, and facilities, as well as data on U.S. intra-regional trade, before presenting the empirical patterns.

## 3.1 The Online Retail Business Model

A key feature of e-commerce, in contrast to brick-and-mortar retail, is the separation of retailer and consumer locations. Retailers do not need a physical presence near consumers to sell goods; instead, they connect with consumers through online search and matching. Figure 1 illustrates the business model of a typical online retailer. Online retailers encounter both matching and shipping frictions when reaching consumers across markets of varying sizes, and may ship goods directly or from storage facilities in fulfillment centers. They also face transport costs when sourcing from upstream producers in different regions.[2]

The location flexibility of online retailers creates strong incentives for spatial agglomeration. As in Krugman (1991); Krugman and Venables (1995) and Puga (1999), spatial frictions combined with input-output linkages generate pecuniary spillovers that favor co-location. This insight clarifies how the expansion of e-commerce influences the economy by changing the locational incentives for online retailers. A key aspect of e-commerce platform expansion, especially by firms like Amazon post-2005, is the enhancement of the online shopping experience and the establishment of fulfillment centers, which reduce both matching frictions and shipping costs to downstream consumers. These reductions in spatial friction encourage online retailers to concentrate their locations, especially when they have improved access to fulfillment centers, transportation infrastructure, or focus on standardized, durable products.

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 retail sales from that region to other regions. Therefore, states with higher shares of online retailers tend to have higher outgoing shipments and lower incoming tradable goods shipments. Under a similar vein, as a region gains better access to fulfillment facilities, it reduces shipping friction between regions, potentially attracting more online retailers and boosting bilateral 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.

[2]The e-commerce model in Figure 1 assumes that online retailers first purchase and store goods before shipping to consumers, either directly or via third-party fulfillment. In Amazon data, 72 percent of sellers and 78 percent of products use Amazon fulfillment, suggesting direct shipping from producers to consumers is minimal. Case studies also indicate sellers' locations mainly serve as inventory storage, acting as a relay point between producers and consumers.

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 % 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.[3] Together with price information I then obtain the total sales revenue of a product overtime.

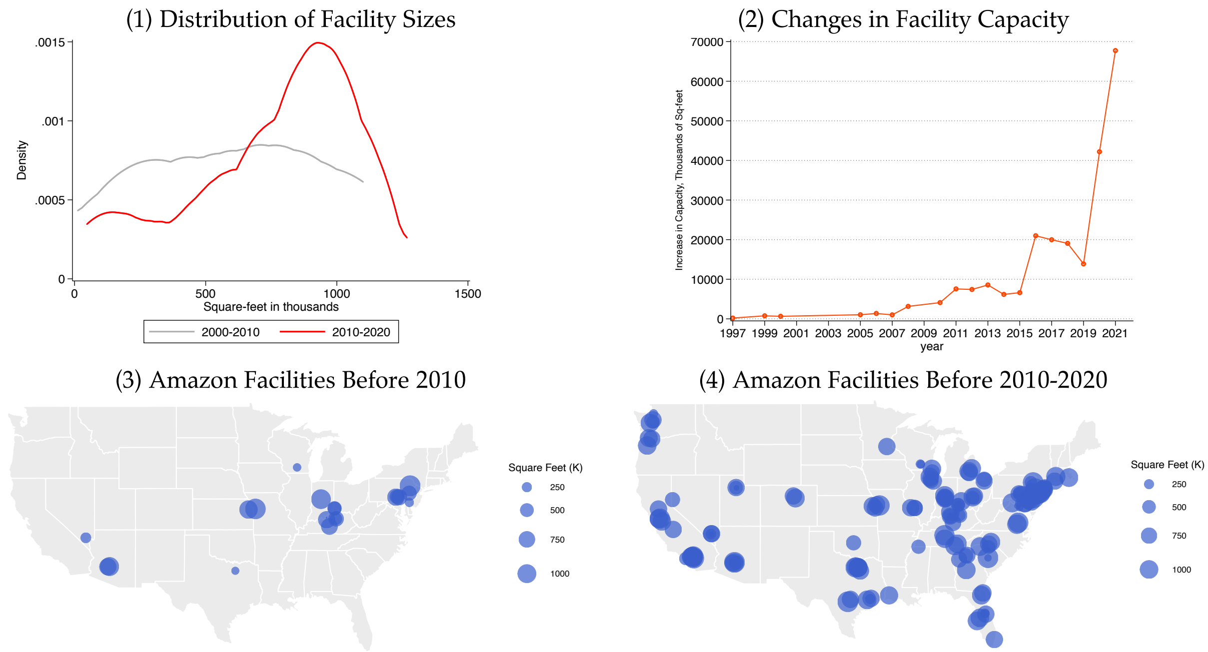
Moreover, I also obtain detailed seller information for the products sold on Amazon. Keepa starts to track sellers in 2016 and assign each seller with a unique identifier, which can then be linked to the seller profile on Amazon that contain information on the seller's address, fulfillment method, and 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.[4]

**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'

[3]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

[4]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.

Figure 2: Expansion of Amazon Facilities




*Notes*: This figure presents Amazon's facility growth, differentiated by facility sizes and distribution across regions. The data comes from MWPV, a supply chain consulting company (www.mwpvl.com). The focus is on large fulfillment centers that handle non-perishable goods, which are more likely to influence retail patterns and regional economies.

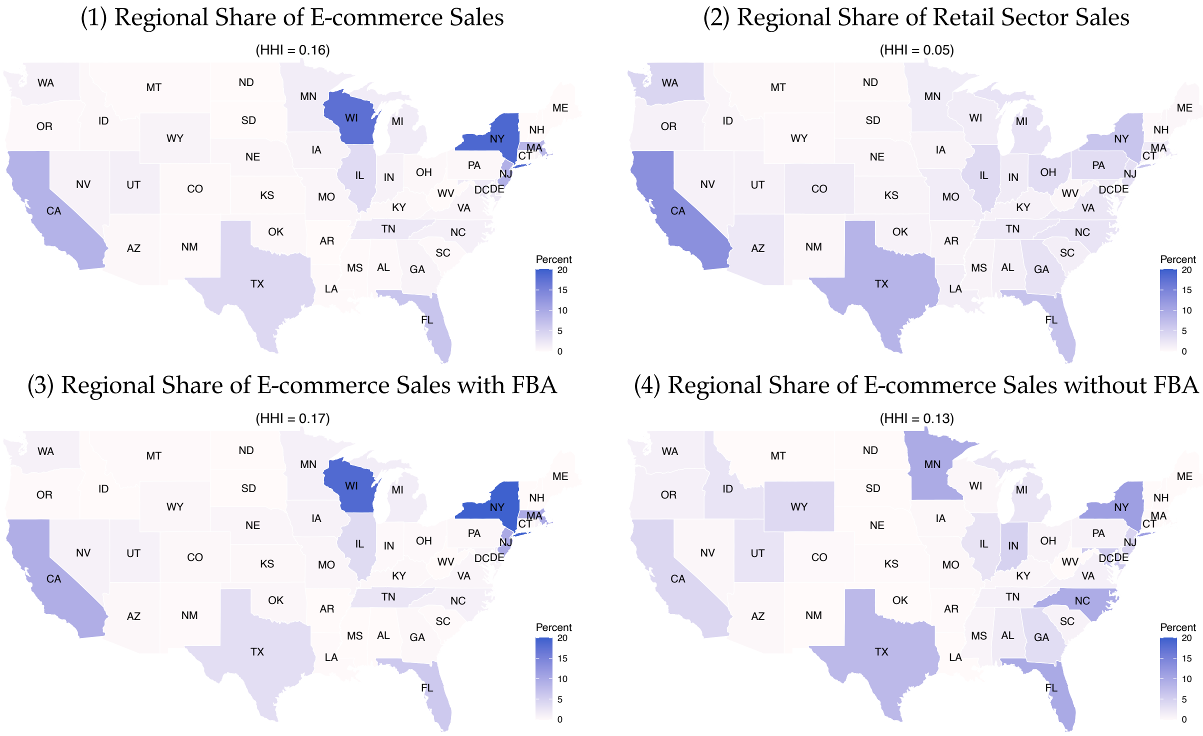
locations decisions.[5]

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

[5]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 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




*Notes*: These density plots show the share of e-commerce and overall retail value added across U.S. states from 2016 to 2021. The sales value of online sellers is allocated to each state according to the seller's address, and the data comes from Keepa, which also contains information on whether the retailer uses Amazon's fullfillment services. The data on the value-added share of the retail sector in each state comes from BEA.

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 agglomeration of online sellers and implications for intra-regional trade flows.

**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.[6] 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

[6]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 |

*Notes*: This table reports the Herfindahl-Hirschman Index (HHI) measuring the concentration of online retail sales across regions for nine product categories, differentiating between durable and non-durable goods. The data is sourced from Keepa, which tracks sales by online sellers, with sales allocated to states based on the sellers' addresses.

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 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

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 |

*Notes*: This table shows the regression results examining how state population shares and corporate tax rates relate to online and overall retail sales, conditional on state and year fixed effects. The source of online retail data is based on the data analysis of Keepa, and the data on the value-added share of the overall retail sector comes from BEA.

of their address information since these businesses tend to have larger transactions and operate at fixed and easily verifiable locations.

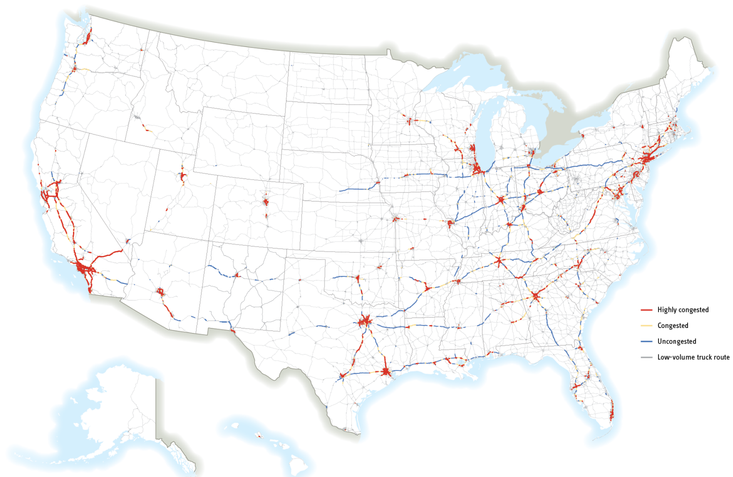
**Pattern 3:** *Online retail sales exhibit a weaker relationship with corporate taxes and population, but aligns closely with regional truck volumes.*

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*

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




*Notes*: High-volume truck portions of the National Highway System carry more than 8,500 trucks per day, including freight-hauling long-distance trucks, freight-hauling local trucks, and other trucks with six or more tires. Highly congested segments are stop-and-go conditions with volume/service flow ratios greater than 0.95. Congested segments have reduced traffic speeds with volume/service flow ratios between 0.75 and 0.95. Data from U.S. Department of Transportation, Federal Highway Administration, Office of Highway Policy Information, Highway Performance Monitoring System, and Office of Freight Management and Operations, Freight Analysis Framework, version 3.4, 2012.

*origin markets with more online retailers export more tradable goods.*

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

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

| Dependent Variable: | $\Delta$ ln (Shipment) | |
|---|---|---|
| $\Delta$ share (%) of online sellers - origin | 3.5*** | |
| | [0.8] | |
| $\Delta$ share (%) of online sellers - destination | -1.4* | |
| | [0.7] | |
| $\Delta$ 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 |

*Notes*: This table presents how changes in the share of online sellers within origin and destination regions influence inter-regional shipment volumes. The data on changes in the share of regional online retailers comes from Keepa. The wholesale trade flow data comes from the Commodity Flow Survey (CFS) from 2012 to 2017. The Amazon distribution center location data comes from MWPV. The table includes fixed effects for origin, destination, and industry.

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 develop a multi-sector spatial retail trade model to examine the rise of e-commerce and its effects on regional economies. The model has two key features. First, consumers search for retailers across regions under imperfect information, which leads to a CES demand system with a demand shifter for online retailers that reflects the efficiency of matching. Second, there are two types of retailers within a vertical production structure. Brick-and-mortar retailers source intermediate varieties to sell exclusively to local consumers, while online retailers can optimally choose locations based on cost advantages for both sourcing inputs and serving multiple markets. The efficiency of online matching and the location choices of online retailers shapes inter-regional retail trade flows. The model shows how the rise of e-commerce, reflected in the rise in online match efficiency and reduction in shipping frictions affect regional outcomes.

The general environment of the model contains $N$ regions, denoted by $n$ or $m$, and encompasses a total of $J = 3$ sectors: one non-tradable service sector and two tradable goods sectors (durable and non-durable). Each tradable sector consists of three subsectors: intermediate manufacturing ($M$), online retail ($R$), and brick-and-mortar retail ($B$). Workers are heterogeneous in their productivities and make optimal decisions about which sector to work in.

In the following, I describe the consumer search process, vertical production structure, and the spatial retail trade problem. I also highlight the role of e-commerce in this model, followed by a welfare analysis based on the model's framework.

## 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 $\eta^j$. The retail sector operates under monopolistic competition, where each retailer sells a unique variety. Consumers have access to a total measure of $1 + O^j$ retailers for sector $j$ goods, consisting of a measure 1 of local brick-and-mortar stores and $O^j$, the normalized measure of online retailers. Among these, $O^j_m$ represents the online retailers originating from region $m$ ($O^j = \sum_m O^j_m$). Under Cobb-Douglas utility and given income $y_n$, a consumer's optimal consumption from a chosen retailer is $c^{j,K}_{nm} = \eta^j y_n / p^{j,K}_{nm}$, where $p^{j,K}_{nm}$ is the price of goods from retailers in origin $m$, and $K = \{B, R\}$ distinguishes between brick-and-mortar and online retailers.

Consumers face unknown match value with retailers and resolve this uncertainty by engaging in sequential ordered searches. Specifically, the indirect utility a consumer from region $n$ derives from purchasing from a retailer in $m$ is expressed as $v^j_{nm} = \ln \eta^j y_n - \ln p^{j,K}_{nm} + \epsilon^{j,K}_{nm}$. Here, $\epsilon^{j,K}_{nm}$ represents the unknown, idiosyncratic match value between the consumer and the retailer, which is independently distributed according to the function $F(\epsilon)$. For local brick-and-mortar stores, the match value is normalized to zero ($E(\epsilon^{j,B}_{nn}) = 0$), while the average relative match value for online retailers is given by $\ln(\mu)$ ($E(\epsilon^{j,R}_{nm}) = \ln(\mu)$).[7] Consumers incur a utility cost $s$ for each sequential search to reveal the match value $\epsilon^{j,K}_{nm}$ of a retailer, deciding after each revelation whether to stop or continue searching.

[7]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.

In line with Weitzman (1979), the optimal consumer strategy is to direct their search by the sequence $\bar{\epsilon}_{nm}^{j,K} - p_{nm}^{j,K}$, where $\bar{\epsilon}_{nm}^{j,K}$ represents the minimum match value that makes the consumer indifferent between continuing to search or stopping ($s = \int_{\epsilon_{nm}^{j,K}}^{\bar{\epsilon}_{nm}^{j,K}} (1 - F(\epsilon))d\epsilon$).[8] This sequential search process leads to an eventual purchase choice (Choi et al. 2018; Armstrong 2017; Armstrong and Vickers 2015). Defining $\omega_{nm}^{j,K} \equiv min\{\epsilon_{nm}^{j,K}, \bar{\epsilon}_{nm}^{j,K}\}$, which represents the "effective match value" of a retailer, the consumer will buy from the retailer from $m$ if it maximizes $\omega_{nm}^{j,K} - p_{nm}^{j,K}$.[9]

$$\begin{aligned} D_{nm}^{j,K} &= P(\omega_{nm}^{j,K} - \ln p_{nm}^{j,K} > max_g\ \omega_{ng}^{j,K} - \ln p_{ng}^{j,K}) \\ &= \int \Pi_{g \neq i} F_{\omega_{ng}^{j,K}}(\epsilon - \ln p_{ng}^{j,K}) f_{\omega_{nm}^{j,K}}(\epsilon - \ln p_{nm}^{j,K}) d\epsilon \end{aligned} \tag{1}$$

**Optimal Demand Derivation**: Sequential ordered search by consumers results in a CES demand framework in two steps. First, consumers' purchase decisions based on $\omega_{nm}^{j,K}$ align with a discrete choice formulation, as outlined in Anderson et al. (2022). The demand from a representative consumer in region $n$ for a retailer in region $m$, denoted as $D_{nm}^{j,K}$, follows a discrete choice model when $F_{\omega_{nm}^{j,K}} = F_{\varepsilon_{nm}^{j,K}}$, where $\varepsilon_{nm}^{j,K}$ captures the random utility component of $\omega_{nm}^{j,K}$. Second, CES demand emerges as a special case of this discrete choice framework. Assuming $E(\epsilon_n^{j,B}) = 0$ and $E(\epsilon_n^{j,R}) = \ln(\mu)$, we can express $\varepsilon_{nm}^{j,K}$ as $\ln(\mu) + \chi^j \tilde{\varepsilon}_{nm}^{j,K}$, where $\tilde{\varepsilon}_{nm}^{j,K}$ has a mean of 0 and unit variance, and $\chi^j$ represents the sector-specific variance of the effective match value $\omega_{nm}^{j,K}$.[10] Under the assumption of an extreme type I distribution for $\tilde{\varepsilon}_{nm}^{j,K}$, the demand function simplifies to $D_{nm}^{j,K} = \frac{(p_{nm}^{j,K}/\mu)^{-1/\chi^j}}{\sum_{g=1}^{N} (p_{ng}^{j,R}/\mu)^{-1/\chi^j} + (p_{nn}^{j,B})^{-1/\chi^j}}$, representing a standard CES expenditure share.

Theorem 1 presents the final CES demand function for consumers and highlights the role of search and matching in shaping consumer behavior. The parameter $\mu$ plays a central role: as online shopping becomes more efficient and matching improves, $\mu$ increases, shifting consumers' demand towards online retailers. The measure of non-local online retailers, $O_m^j$, determines the variety of goods available for consumption, while

[8] The consumer will stop and make a purchase from either a local brick-and-mortar store or an online retailer in $m$ if $max\{v_{nm}^{j,K}, -max_{g \in \bar{O}} \ln p_{nm}^{j,K} + \epsilon_{nm}^{j,K}\} > max_{g \in \bar{O}} - \ln p_{ng}^{j,K} + \epsilon_{ng}^{j,K}$, where $\bar{O}$ denotes the retailers the consumer has searched so far.

[9] As Choi et al. (2018) shows, to guarantee the existence and uniqueness of the equilibrium, one needs the density and loss functions of $\omega_{nm}^{j,K}$ to be log-concave, and the density function to be unbounded above. These are taken as assumptions for this paper.

[10] This implies that $\omega_{nm}^{j,K}$ has a mean of $\ln(\mu)$, reflecting the abundance of online retailers relative to local stores. Given the lower cost of searching additional retailers online, $\omega_{nm}^{j,K}$ closely approximates $\epsilon_{nm}^{j,K}$, with a mean of $\ln(\mu)$.

the variance of consumers' effective match value, $\chi^j$, shapes the elasticity of substitution among retailers, given by $\sigma^j = \frac{1+\chi^j}{\chi^j}$. Lower uncertainty about the value of goods from online retailers reduces $\chi^j$, making retailers more substitutable. Under monopolistic competition, this elasticity determines the markup charged by retailers, $\tilde{\sigma}^j = \frac{\sigma^j}{\sigma^j - 1}$.

**Theorem 1.** *A representative consumer in region n with sectoral consumption weights $\eta^j$ has nest Cobb-Douglas and CES demand as below under sequential ordered search if only if the effective match value $\omega_{nm}^{j,K} = min\{\epsilon_{nm}^{j,K}, \bar{\epsilon}_{nm}^{j,K}\}$ is distributed extreme type I.*

$$C_n = \Pi_{j=1}^{J} (C_n^j)^{\eta^j}, \quad C_n^j = [(c_{nn}^B)^{\frac{\sigma^j - 1}{\sigma^j}} + \mu \sum_{m=1}^{N} \int_0^{O_m^j} (c_{nm}^R(i))^{\frac{\sigma^j - 1}{\sigma^j}} di]^{\frac{\sigma^j}{\sigma^j - 1}} \tag{2}$$

*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 $e^j \in [0, 1]$. 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 $\beta_n$. 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 $\kappa_{ni}^{M}$ units are required to ship one unit from $i$ to $n$.[11] 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 $\theta^j$ 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 $\theta^j$ 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 $i$.[12]

**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 varieties $e^j \in [0,1]$ 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

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

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

$\alpha^j$ regulates the elasticity of substitution among the varieties in sector $j$.[13]

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

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

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 $\gamma_n^j$ 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 $\frac{\theta^j+1-\alpha^j}{\theta^j}$.[14] 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 $\kappa_{ni}^R$ 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 various locations $(z_1^{j,R}, ..., z_N^{j,R})$ to establish their operations in any given region $m$, paying

[13]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. Further, not only the intermediate varieties are tradable, the final retail goods are also tradable to capture e-commerce.

[14]Since the vector of productivity draws for variety $e^j$ across regions is $a(e^j) = \{a_1(e^j), \dots, a_N(e^j)\}$, 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.

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.[15] The optimal location choice for online retailers is then:

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

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:

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

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 $P(Z_1^{j,R} < z_1, ..., Z_N^{j,R} < z_N) = 1 - (\sum_{m=1}^N [T_m^{j,R} z_m^{-\tau}]^{\frac{1}{1-\rho}})^{1-\rho}$.[16] The scale parameter $T_m^{j,R}$ measures the absolute advantage of region $m$ in producing sector $j$ goods, whereas $\tau$ controls the degree of heterogeneity across different vectors, and $\rho$ 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

[15]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.

[16]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 $\psi_m^j = T_m^{j,R}(\xi_m^j)^{\frac{-\tau}{1-\rho}} / \sum_{m=1}^N [T_m^{j,R}(\xi_m^j)^{-\tau}]^{\frac{-\rho}{1-\rho}}$.[17] 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 $\tau$ and $\rho$.

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.

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

$$x_n^{j,B} = \frac{\frac{1}{O}(c_n^{j,B})^{1-\sigma^j}}{\sum_h \Psi_h^j(\kappa_{nh}^R c_h^{j,R}/\mu)^{1-\sigma^j} + \frac{1}{O}(c_n^{j,B})^{1-\sigma^j}} \tag{11}$$

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 ($\Psi_m^j$) can be directly observed from the data.

The model also characterizes the regional expenditure share on local brick-and-mortar stores, $x_n^{j,B}$, in addition to the inter-regional expenditure share on online retailers. Here, bilateral trade resistance is replaced by the unit cost of local brick-and-mortar stores divided by the measure of online retailers. Since $O$ represents the relative measure of online retailers to brick-and-mortar stores and $O < 1$, local brick-and-mortar expenditure is higher than what would be expected based on cost alone. Notably, the expenditure share on local brick-and-mortar stores ($x_n^{j,B}$) differs from that on online retailers ($x_{nn}^{j,R}$) since the model allows a separate characterization for each. Section 7 introduces an

[17] Note that $\tilde{\sigma}$ and $\eta^j$ do not appear in the definition of $\xi_m^j$ and $\Psi_m^j$ since they are constant within a sector.

alternative model where brick-and-mortar stores can choose to sell online, letting retailers operate both formats, which determines $O$ endogenously.

### 4.3 Labor Supply

To analyze workers' sorting and heterogeneous labor supply across sectors, I use a Roy (1951) framework with probabilistic productivities (Lagakos and Waugh 2013; Hsieh et al. 2019; Galle et al. 2022; Lee 2020). In each region $n$, workers receive a vector of region-sector-specific productivities $z_n^{j,K}$ for each unit of labor they provide, where $j$ denotes the sector and $K = \{M, R, B, \emptyset\}$ represents the subsectors for tradable goods.[18] Non-employment is treated as sector 0, where workers can allocate their labor, earning a wage $w_n^0$ per efficiency unit of labor, which reflects the marginal return for home production (see Dvorkin (2014) and Caliendo et al. (2019) for other examples).

Worker productivities across different sectors are drawn independently from a Fréchet distribution $\psi_n^{j,K}(z_n^{j,K})$, with shape parameter $\nu_n$ and scale parameter $A_n^{j,K}$. The scale parameter $A_n^{j,K}$ determines the absolute advantage, while $\nu_n$ regulates workers' comparative advantage. The joint distribution of productivity draws follows another Fréchet distribution, $\psi_n(z_n) = \sum_{j=0}^{J} \sum_{K=M,R} A_n^{j,K} z^{-\nu_n}$. Workers maximize their wages per unit of labor supplied $w_n^{j,K} z_n^{j,K}$ by optimally choosing sectors $(j, K)$.[19] Using the properties of the joint Fréchet distribution for productivity draws $\psi_n(z_n)$, we can derive the probability of employment in sector $(j, K)$ as:

$$\pi_n^{j,K} = \frac{A_n^{j,K}(w_n^{j,K})^{\nu_n}}{\Phi_n}, \text{ where } \Phi_n = \sum_{j=1}^{J} \sum_{K=\{M,R,B,\emptyset\}} A_n^{j,K}(w_n^{j,K})^{\nu_n} + A_n^0(w_n^0)^{\nu_n}. \tag{12}$$

The probability of workers in a given sector is determined by the sector's wage return relative to total returns from both employment and non-employment, scaled by the Fréchet parameter $\nu_n$, which governs the elasticity of labor adjustment across sectors.[20] Therefore, as e-commerce expansion shifts labor demand and impacts wages, households' employment decisions are also altered. A further advantage of using the Fréchet distribution is it simplifies the derivation of labor supply in efficiency units.

[18] Specifically, $z_n^{j,K}$ has six dimensions: $z_n = \{z_n^0, z_n^1, z_n^{2,M}, z_n^{2,R}, z_n^{2,B}, z_n^{3,M}, z_n^{3,R}, z_n^{3,B}\}$, where sectors 0 and 1 represent non-employment and services, and sectors 2 and 3 represent durable and non-durable goods. Subsections $M, R, B$ correspond to intermediate producers, online retailers, and brick-and-mortar retailers.

[19] The optimal choice set for a sector $(j, K)$ is defined as $\Lambda_n^{j,K} \equiv \{z_n^{j,K} \mid z_n^{j,K} > z_n^{H,k} \, \forall \, (H, k)\}$, indicating a worker will choose to work in $(j, K)$ if the drawn productivity vector falls within this set.

[20] 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.

Specifically, for a given sector $(j, K)$, the efficiency units of labor provided becomes:

$$l_n^{j,K} \equiv \Gamma(\frac{\nu_n - 1}{\nu_n}) \frac{\Phi_n^{1/\nu_n}}{w_n^{j,K}} \pi_n^{j,K} L_n, \tag{13}$$

which affects workers' income as well as firms' output. The wage return for workers in sector $(j, K)$ simplifies to $w_n^{j,K} l_n^{j,K} = \Gamma(\frac{\nu_n - 1}{\nu_n}) \Phi_n^{1/\nu_n} \pi_n^{j,K} L_n$.

## 4.4 Market Clearing and Competitive Equilibrium

In the goods market, expenditure clears in two ways: consumers buy retail goods from various retailers, while retailers obtain intermediate varieties from different producers:

$$X_n^{j,R} = \sum_{i=1}^{N} x_{in}^{j,R} (I_i L_i), \text{ where } I_i L_i = \sum_{k=0}^{J} \sum_{K=\{M,R,B,\emptyset\}} (r_i^{h,k} h_i^{K,k} + w_i^k l_i^{K,k}) - \Omega_i, \tag{14}$$

$$X_n^{j,M} = \sum_{i=1}^{N} (1 - \gamma_i^j) x_{in}^{j,M} X_i^{j,R}. \tag{15}$$

The total expenditure on sector $j$ retail goods sold from region $n$, $X_n^{j,R}$, equals the product of the retail expenditure share $x_{in}^{j,R}$ and total income $I_i L_i$ across regions. In the benchmark model, households' total income is derived from wage earnings and land ownership, minus the region's trade deficit $\Omega_i$ that is assumed to be exogenous.[21] Meanwhile, the total demand for sector $j$ intermediate goods from region $n$, $X_n^{j,M}$, equals the expenditure share on intermediate goods $x_{in}^{j,M}$ multiplied by the portion of retail sector spending on intermediate varieties $(1 - \gamma_i^j) X_i^{j,R}$, summed across regions. Accounting for regional trade deficits, the balance of trade requires:

$$\sum_{j=0}^{J} \sum_{i=1}^{N} (x_{ni}^{j,M} X_n^{j,M} + x_{ni}^{j,R} X_n^{j,R}) + \Omega_n = \sum_{j=0}^{J} \sum_{i=1}^{N} (x_{in}^{j,M} X_i^{j,M} + x_{in}^{j,R} X_i^{j,R}). \tag{16}$$

The market clearing for primary factors including as labor and structures requires that their returns equal the corresponding portion of value-added. However, since these factors are used in both intermediate and retail production—with different production functions—the market clearing conditions differ between sectors. Specifically, for the

[21] In the discussion of policy interventions, households' total income will also depend on tariffs imposed by local regions and an endogenous deficit affected by revenue reallocation.

labor market.

$$w_n^{j,M} l_n^{j,M} = \beta_n X_n^{j,M}, \quad w_n^{j,R} l_n^{j,R} = \gamma_n^j \beta_n X_n^{j,R}, \quad w_n^{j,B} l_n^{j,B} = \gamma_n^j \beta_n X_n^{j,B}, \tag{17}$$

**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 $\mu$, 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 $(\eta_n^j)$, elasticity of substitution among retailers $(\frac{1}{1-\sigma^j})$ and among intermediate varieties $(\frac{1}{1-\alpha^j})$, Fréchet shapes of worker $(v^n)$ and sector productivities $(\theta^j)$, labor share of value-added $(1-\beta_n^j)$, and value-added share of retail goods $(\gamma_n^j)$, all assumed constant. The endogenous variables are labor allocation $L_n^{j,K} n = 1, j = 1^{N,J}$ and trade and expenditure allocation for intermediate $(xni^{j,M}, X_n^{j,M})$ and retail goods $(x_{ni}^{j,R}, X_n^{j,R})$. All prices can be expressed relative to wages, and equilibrium is defined as a vector of wages that satisfies optimality and market clearing conditions, resulting in endogenous allocations. Appendix C provides the formal definition of this competitive equilibrium.

When an e-commerce shock affects certain economic fundamentals $\Psi' \subseteq \Psi$, determining the equilibrium requires solving a system of nonlinear equations that depend on all other fundamentals and parameters, which is particularly challenging in spatial models. Following Dekle et al. (2008) and as detailed in Appendix C, I derive comparative statics in proportional changes, or "hat algebra", without specifying all fundamentals and parameters. Specifically, let $\hat{x}$ represent the proportional change in any variable from its original value $x$ to a counterfactual value $x'$. Appendix C shows the the proportional change in equilibrium outcomes in response to changes in $\Psi'$ condional on initial allocations does not require the initial levels of all other fundamentals, as the initial allocations inherently contain this information.[22]

[22]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.** In the model, e-commerce as exemplified the rise of Amazon, affects economic fundamentals through two key channels. First, as online shopping reduces consumer search frictions and enhances the overall online shopping experience (Goldmanis et al. 2010; Dinerstein et al. 2018), it improves online match efficiency, increasing $\mu$ and shifting consumer demand towards online retailers. Second, the expansion of fulfillment and distribution centers by companies like Amazon lowers shipping costs (Houde et al. 2021), reducing iceberg costs $\kappa_{ni}^{R}$. As shown in equation (10), both channels directly influence inter-regional retail trade. Moreover, because the location of online retailers is endogenously determined by online match efficiency and iceberg costs, this agglomeration effect further amplifies the initial impact on trade flows.

**Welfare Analysis.** I now define and derive the welfare changes and analyze the key channels. Welfare for a region is defined as real income per capita, $W_n = \frac{Y_n/L_n}{P_n}$, where $Y_n = I_n L_n + \Omega_n$ represents total income. Using the derived efficiency units of labor from equation (13), we can simplify $Y_n$ as $Y_n = (\frac{1}{1-\beta_n})\Gamma(\frac{\nu_n-1}{\nu_n})\Phi_n^{1/\nu_n} L_n$. Next, using proportional changes or "hat algebra," welfare changes are expressed as $\hat{W}_n = \hat{\Phi}_n^{1/\nu_n}\Pi_{j=1}^{J}(\hat{P}_n^{j,R})^{-\eta_j}$. Labor market allocation gives $\hat{\Phi}_n^{1/\nu_n} = \hat{w}_n^{j,K}(\hat{\pi}_n^{j,K})^{\frac{-1}{\nu_n}}$ for any sector $(j,K)$. For simplicity, I use the non-employment sector since it acts as an outside option: $\hat{\Phi}_n^{1/\nu_n} = \hat{w}_n^{0}(\hat{\pi}_n^{0})^{\frac{-1}{\nu_n}}$. Finally, applying the retail trade share expression from equation (27) simplifies $\Pi_{j=1}^{J}(\hat{P}_n^{j,R})^{-\eta_j}$ to $\Pi_{j=1}^{J}(\hat{x}_{nn}^{j,R})^{\frac{-\eta_j}{\sigma^j-1}}(\hat{c}_n^{j,R})^{-\eta_j}$. These elements together yield the counterfactual welfare changes:

$$\hat{W}_n = \underbrace{\hat{w}_n^{0}(\hat{\pi}_n^{0})^{\frac{-1}{\nu_n}}}_{\text{income effect}} \underbrace{\Pi_{j=1}^{J}(\hat{x}_{nn}^{j,R})^{\frac{-\eta_j}{\sigma^j-1}}(\hat{c}_n^{j,R/B})^{-\eta_j}}_{\text{price effect}}. \tag{18}$$

The expression for welfare changes highlights the general equilibrium channels through which e-commerce can affect an economy with interconnected regions and sectors. The term $\Pi_{j=1}^{J}(\hat{x}_{nn}^{j,R})^{\frac{-\eta_j}{\sigma^j-1}}(\hat{c}_n^{j,R})^{-\eta_j}$ reflects the price effects, derived from changes in the aggregated consumer retail price index. These effects depend on the region's share of local goods in consumer retail expenditure $\hat{x}_{nn}^{j,R}$ and are influenced by the demand elasticity $\sigma^j$ and expenditure shares $\eta^j$ at the sector level. A region's own-good expenditure share and demand elasticity serve as sufficient statistics for welfare change in many trade models, as discussed in Arkolakis et al. (2012). As e-commerce shifts demand toward non-local retailers, it increases welfare through this price channel. Sectoral differences in elasticities and expenditure shares further adjusts the welfare impact.

An additional term affecting price effects is the change in the unit cost of local retail production, $\hat{c}_n^{j,R}$, which reflects input-output linkages. This change affects both local retail prices and the expenditure share of a region's own retail goods, with consumers benefiting from a lower $\hat{c}_n^{j,R}$. As shown in equation (7), this effect is closely tied to input-output linkages. As the prices of intermediate goods adjust to wage changes, the price of local retail goods will shift based on $\gamma_n^j$, the value-added share of intermediate goods.

The income effects on welfare capture the forces of comparative advantage, under heterogeneous worker productivity and imperfect mobility across regions. The term $\hat{w}_n^0(\hat{\pi}_n^0)^{\frac{-1}{\nu_n}}$ suggests that as the non-employment rate decreases or returns for home production rise, welfare tends to increase. Although I focus on the non-employment sector to illustrate income effects, the change in total income is positively correlated with wage changes and negatively correlated with employment changes in 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 indicates that welfare increases with the degree of worker specialization. Therefore, regions with a comparative advantage in sectors boosted by an e-commerce shock will see welfare gains, while those losing jobs to external competition will face welfare declines.

Taking stock, by explicitly accounting for demand shifts driven by consumer search, retailer location, shipping frictions, and heterogeneous labor supply, the model provides comparative statics that comprehensively reflect the general equilibrium mechanisms through which e-commerce impacts regional economic outcomes and welfare changes.

# 5 Model Quantification

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 quantify the impact of e-commerce on regional economies, I use 2007 as the baseline economy, as Amazon's online sales began to grow significantly only after this point, and

Table 4: Parameters, Fundamentals and Shocks for Model Quantification

| Section | Param. | Description | Estimation/Caliberation |
|---|---|---|---|
| Consumption | $\eta_n^j$ | Sector share of consumption | CFS 2007 |
| | $\sigma^j$ | Elasticity of subs. across retailers | Keepa + IV |
| Labor Supply | $\pi_n^j$ | Share of employment | CBP, ACS |
| | $v^n$ | Fréchet shape of worker product. | Galle et al. (2022) |
| Production | $\beta_n^j$ | Share of structures | BEA, Greenwood et al. (1997) |
| | $\theta^j$ | Fréchet shape of sector product. | Caliendo and Parro (2015) |
| | $\gamma_n^j$ | Value-added share of retail goods | BEA, CFS |
| Expenditure | $x_{ni}^{j,M}$ | Interm. expenditure share | CFS 2007 |
| | $x_n^{j,B}$ | Brick-and-motar expenditure share | CFS 2007, E-Stats |
| | $x_{nm}^{j,R}$ | E-commerce expenditure share | CFS 2007, E-Stats |
| | $p_n^{j,B}$ | Brick-and-motar price index | CFS 2007, E-Stats, CES |
| Amazon Shock | $\hat{\kappa}_{nm}^R$ | Iceberg cost change | Amazon data + CFS 2007 + IV |
| | $\mu$ | Matching efficiency | E-stats + CES |
| | $\Psi_m^j$ | Online retailer location probability | Keepa |
| | $O_m^j$ | 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 |

*Notes*: This table presents the model's calibration and details the sources of information for each parameter or fundamental.

2017 represents the post-Amazon shock equilibrium. The model is calibrated using data and variables from all 50 U.S. states. In this model, each labor market is defined as a region-sector pair, resulting in 400 markets in the quantification.[23] Table 4 outlines the parameters, fundamentals, and shocks across model sections that require calibration or estimation along with their sources of information, which I discuss below.

**Consumption** On the consumption side, I calibrate the expenditure shares for durable, non-durable, and service sector goods,$\eta^j$, using regional consumption data from the Bureau of Economic Analysis (BEA).

$$\Delta \ln(x_{nm}^{j,R}) = \delta + (1-\sigma)\Delta \ln(c_m^{j,R}) + \Delta \ln(\kappa_{nm}^R) + \epsilon_{nm}^{j,R} \tag{19}$$

[23] As outlined in Section 4, the model includes eight sectors: two tradable goods sectors (durable and non-durable), a service sector, and a non-employment sector. Each tradable sector contains three subsectors: manufacturing, online retail, and brick-and-mortar. Since the main dataset, CFS, uses the 3-digit North American Industry Classification System (NAICS), online Appendix Table 1 details the allocation of NAICS sectors into durable and non-durable categories, while the breakdown of online retail and brick-and-mortar sectors is discussed below.

For the elasticity of substitution between different retailers ($\sigma^j$), I calibrate using gravity trade flow equation (10), resulting in a log differences equation as in (19). This equation relates changes in bilateral retail expenditure shares to shifts in origination prices and iceberg costs. Origination prices are calculated from the CFS using per-unit prices—shipment value divided by shipment weight—for intra-regional shipments. Assuming iceberg costs are a flexible function of shipment distance, I then also control for changes in distances. To address potential price endogeneity, I apply a standard Hausman instrument to isolate prices from region-specific demand shocks, with results detailed in Appendix Table A2. The estimated elasticity of substitution is 1.8 for durable goods and 5.0 for non-durable goods. The non-durable elasticity aligns with existing estimates, such as 4.3 for brick-and-mortar versus online retailers from Dolfen et al. (2019) and 5.5 across U.S. commuting zones from Gervais and Jensen (2019); as expected, the elasticity for durable goods is lower.[24]

**Labor Supply.** On the worker side, the Census County Business Patterns (CBP) data provide employment shares by region and sector, $\pi_n^j$, but do not separate online retail from brick-and-mortar employment. To distinguish these subsectors, I impute their regional output shares. First, I compute regional e-commerce output using E-Commerce Statistics (E-stats) for national e-commerce sales and the CFS to allocate trade flows based on seller origin, estimating each state's e-commerce output. Second, I compute regional total retail output by combining E-stats' national retail sales data with BEA value-added data to allocate output regionally. Finally, I calculate brick-and-mortar output as the difference between total retail output and regional e-commerce output, which allows me to divide the overall retail employment share accordingly between e-commerce and brick-and-mortar subsectors.

Regarding 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.[25] Here I specify $v^n$ equal to 1.5, which is the value from their preferred specification.

[24]Additionally, Hottman (2017) estimates an elasticity of substitution of 4.5 among stores within a county. Naturally, the elasticity of substitution among stores across different states is lower.

[25]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.

**Production** With regard to production, the share of structures in the structure-labor bundle $\beta_n^j$ 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 $\theta^j$ 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, $\gamma_n^j$, BEA provides the value-added for each sector, which divided by gross-output gives the share value.

**Expenditure and Prices.** To solve for changes in equilibrium economic variables, I calibrate three expenditure shares: the inter-regional intermediate expenditure share $x_{nm}^{j,M}$, the regional expenditure brick-and-mortar share $x_n^{j,B}$, and the inter-regional e-commerce expenditure share $x_{nm}^{j,R}$. I obtain $x_{ni}^{M,j}$ directly from the 2007 CFS data on durable and non-durable manufacturing goods. I calculate $x_n^{j,B}$ and $x_{nm}^{j,R}$ using data on total retail sales, e-commerce sales, and inter-regional retail trade. First, I distribute national e-commerce sales from E-stats to states based on inter-regional wholesale trade expenditure shares from the 2007 CFS data. Then, I distribute total retail sales from E-stats to states using state expenditure shares from the Consumer Expenditure Survey (CES), providing $x_n^{j,B}$. Lastly, I calculate $x_{nm}^{j,R}$ by allocating the remaining e-commerce retail share to various origins based on the 2007 CFS data.

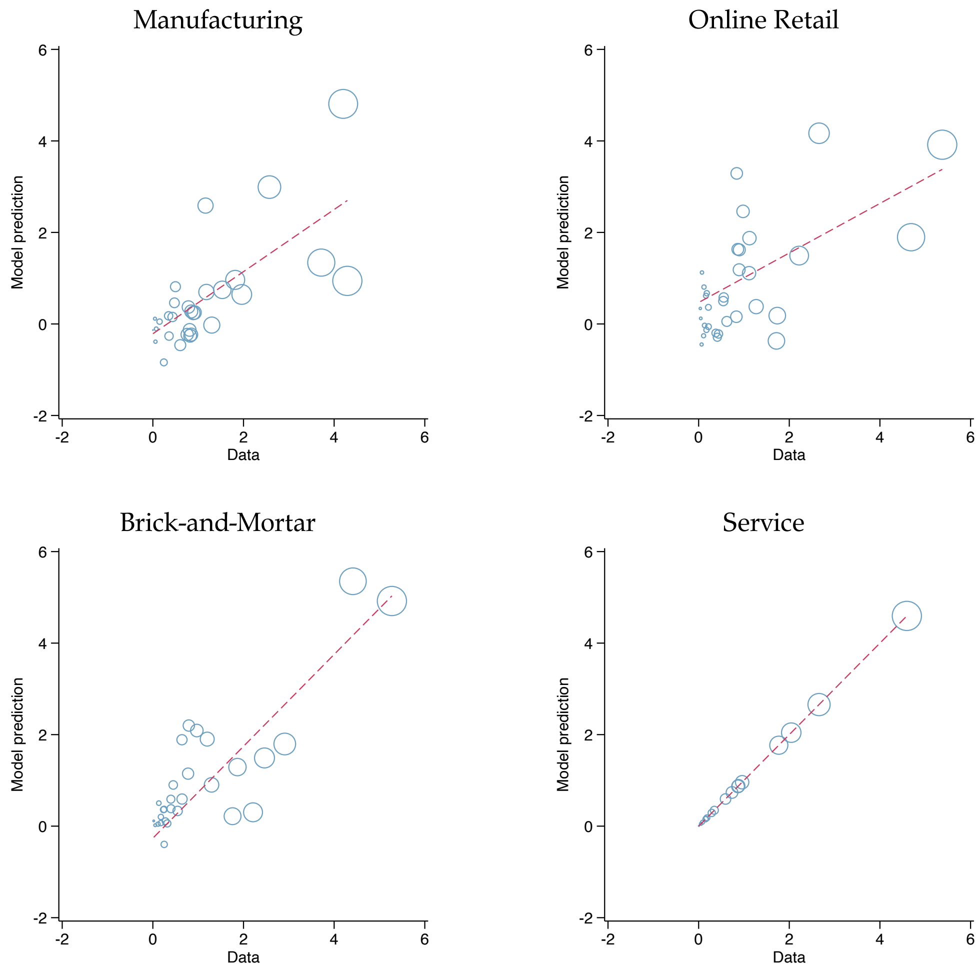
Further, I calibrate the price index of regional brick-and-mortar retail goods, $p_n^{j,B}$, for 2007 to use in solving counterfactual inter-regional trade flows, as in equation (10).[26] The CES data provide the regional price index for all retail goods. To isolate the brick-and-mortar price index, I use the regional ratio of e-commerce expenditure share to brick-and-mortar share, as derived in above steps. This ratio helps separates the brick-and-mortar price index $p_n^{j,B}$ from the overall price index $P_n^j$ in the CES data. Specifically, I calculate $\frac{\sum_i x_{ni}^j}{x_n^{j,B}} = \left[\frac{(P_n^j)^{1-\sigma}}{(p_n^{j,B})^{1-\sigma}} - 1\right]$.[27] This approach imputes the model-consistent brick-and-mortar price indices by aligning them with the observed brick-and-mortar to e-commerce expenditure shares in the data.

**Non-targeted Moments.** I now present the baseline equilibrium quantified through the model and compare it with data for non-targeted moments. As shown in Table 4, the

[26]The inter-regional e-commerce price index, $p_{nm}^{j,R}$, is also needed and is calculated as $p_m^{j,B} \cdot \kappa_{mn}^{j,R}$.

[27]Expanded as $\frac{\sum_i x_{ni}^j}{x_n^{j,B}} = \frac{\sum_{i=1}^N (\frac{p_{ni}^j}{\mu})^{1-\sigma}}{(p_n^{j,B})^{1-\sigma}} = \frac{(P_n^j)^{1-\sigma} - (p_n^{j,B})^{1-\sigma}}{(p_n^{j,B})^{1-\sigma}} = \left[\frac{(P_n^j)^{1-\sigma}}{(p_n^{j,B})^{1-\sigma}}\right] - 1$.

Figure 5: Predicted and Observed Sectoral Value Added in 2007




*Notes*: This figure compares model predictions against observed BEA data for value added in various sectors. The data of the regional expenditure change (y-axis) predicted by the model comes from the model calculation, which is obtained by applying market clearance conditions and other calculations. The observed data (x-axis) comes from BEA directly. This figure evaluates model accuracy regarding sectoral income distribution within regional economies.

model is calibrated to precisely match inter-regional trade shares across sectors as well as the value-added share in production. Starting with an initial guess, I apply the market-clearing conditions in equations (14) and (15) to calculate model-predicted regional expenditures by sector, which are non-targeted.[28] Figure 5 shows the model-predicted regional variations in expenditures (y-axis) alongside observed BEA data (x-axis), both normalized by mean and standard deviation, with circle sizes representing observed data.[29] As shown, the model closely aligns with regional expenditure patterns across

[28]Market clearing conditions described in equations (14) and (15) present a fixed-point problem. To address this, I use functional iteration to determine the equilibrium expenditures predicted by the model.

[29]Since the BEA data does not segregate retail value added between online and brick-and-mortar retailers,

sectors based on the data. 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.

## 5.2 Sequential Estimation of the Amazon Shock

In this section, I outline the sequential estimation of two key model fundamentals impacted by the Amazon expansion: the iceberg cost in the retail sector ($\kappa_{nm}^{j,R}$) and the match efficiency between consumers and online retailers ($\mu$). First, I estimate $\kappa_{nm}^{j,R}$ as a function of distance, allowing me to compute its changes based on reductions in shipping distances due to Amazon's expansion. Then, to isolate changes in bilateral shipping distances driven by Amazon's expansion from other endogenous demand factors, I use exogenous geographic factors to build counterfactual Amazon facility locations and use the corresponding changes in shipping distances as instruments. Lastly, I impute the change in search efficiency, $\mu$, based on observed relative expenditure shares between local e-commerce and brick-and-mortar retailers and the estimated iceberg costs.

**Step 1: Relating Iceberg Costs to Shipping Distance.** I begin by estimating the empirical relationship between iceberg costs for retail goods, $\kappa_{nm}^{j,R}$, and bilateral shipping distances. Equation (20) specifies iceberg costs as a function of the shipping distance between origin $m$ and destination $n$, accounting for fixed origin and destination effects along with other bilateral characteristics $X'_{nm}$. I use the CFS's great circle distance between shipment origin and destination as the main independent variable. For the dependent variable, I use the ratio of region $n$'s expenditure share of region $m$'s online retail goods over region $n$'s expenditure share of its own online retail goods, $\frac{x_{nm}^{j,R}}{x_{nn}^{j,R}}$. This ratio reflects bilateral iceberg costs, the retail goods costs in both regions, the measure of online sellers in both regions, and the elasticity of substitution. By taking log, these components become additively separable, isolating the bilateral iceberg cost $\kappa_{nm}^{j,R}$ once origin and destination fixed effects are included. The resulting coefficient divided by consumption elasticity yields the elasticity of iceberg cost with respect to shipping distance, $\frac{\hat{\delta}^j}{\sigma}$. Table 7 shows an elasticity of 1.5 for durable goods and 2.1 for non-durable goods.

$$ln(\kappa_{nm}^{j,R}) = \delta^j \text{Distance}_{nm} + X'_{nm}\theta + \delta_n^j + \delta_m^j + \epsilon_{nm}^j. \tag{20}$$

**Step 2: Shipping Distance Reduction.** Above, I estimate iceberg costs as a function

I employ a similar method used in estimating the ratio of e-commerce retail output to total retail output to impute regional employment share $\pi_n^j$. This output share is used to separate the BEA retail sector value added and serves as the initial guess for the model's imputation.

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 |

*Notes*: The table shows the actual and counterfactual changes in shipping distances resulting from Amazon's facility expansion between 2007 and 2017. Actual shipping distance data is sourced from MWPV, while counterfactual distances are calculated based on facility locations constructed using geographic factors such as land altitude and climate, obtained from Open Topography Global Datasets and the National Center for Environmental Information (NCEI). The reduction in shipping distance is determined using the three nearest Amazon facilities to each destination, selecting the facility closest to the origin, following Houde et al. (2021).

of shipping distance. I now further estimate reductions in shipping distance due to Amazon's facility expansion by imposing detailed structures on fulfillment order flows. Following Houde et al. (2021), which finds that over 90 percent of orders are fulfilled by the three closest centers to the destination, I further specify that the nearest center to the origin processes each order.[30] Table 5 Panel A presents the shipping distance reductions from Amazon's facility expansion: in 2007, an order traveled an average of 490 kilometers; by 2017, this decreased to 288 kilometers—an average reduction of 202 kilometers, or 0.5 in log units.

*Identification Strategy:* Calibrating the shipping distance reduction based on Amazon's actual facility rollout faces key endogeneity issues, as new facility locations often correlate with GDP growth, population changes, and other demand-side factors that could directly affect outcomes. To address this, I simulate counterfactual distribution centers, assigning locations based solely on exogenous geographic cost factors independent of demand (Duflo and Pande 2007; Lipscomb et al. 2013).[31] For the simulation of counterfactual

[30] 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.

[31] 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

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 | |

*Notes*: This table shows regression results identifying how geographic cost factors like temperature, elevation, and precipitation influence the likelihood of Amazon facility locations. Geographical factor data (temperature, precipitation, tornadoes, etc.) come from Open Topography Global Datasets and the National Center for Environmental Information (NCEI). The actual expansion data of Amazon facilities comes from MWPV, which is used as dependent variable. The table confirms that geographic and climatic suitability significantly shape Amazon's expansion, with pseudo R-squared values indicating robust explanatory power.

Amazon facilities, a yearly "budget" is set based on the observed number of new facilities. U.S. counties are ranked solely on topographic and climatic factors, such as land elevation, temperature, precipitation, and extreme weather events.[32] Facilities are then assigned to the highest-ranked counties according to each year's budget. Table 6 shows a cross-sectional probit regression of facility assignments based on these factors. The results indicate that Amazon's expansion favors warmer and lower-elevation locations, while precipitation and tornado frequency are negatively correlated with facility construction, though not significantly.[33]

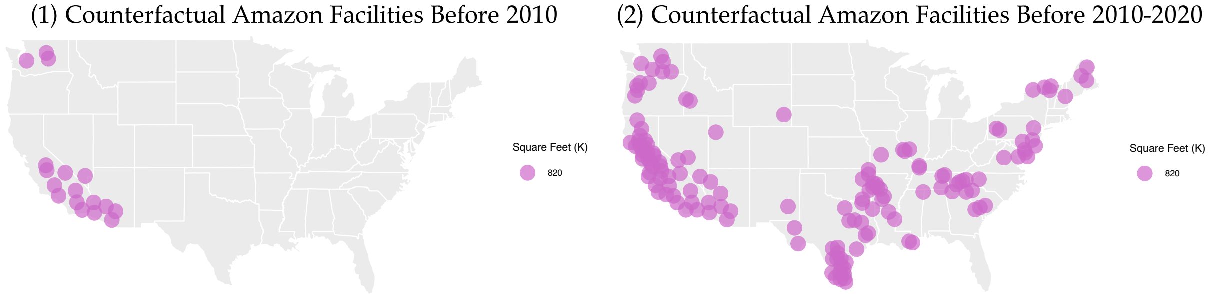
Figure 6 shows the counterfactual centers based solely on geographic factors and Amazon's budget, indicating both similarities and differences with actual facility locations. In earlier years, counterfactual locations favor West Coast more than those Amazon selected; from 2010 to 2020, however, the alignment between actual and counterfactual

comparable to those that did not.

[32]The data sources are Open Topography Global Datasets and the National Centers for Environmental Information (NCEI)

[33]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.

Figure 6: Location of Counterfactual Fulfillment and Distribution Centers




*Notes*: This figure shows counterfactual Amazon facility locations based on geographic suitability factors. County-level geographical feature data (land elevation, climate change) come from Open Topography Global Datasets and the National Center for Environmental Information (NCEI). This simulation assesses Amazon's optimal locations from a cost and logistics perspective.

sites improves, though states like California, Texas, Arkansas, and North Carolina remain more favorable based on geographic factors alone. Table 5 shows that the average order travel distance between regions decreased from 623 kilometers in 2007 to 338 kilometers in 2017—a reduction of 288 kilometers, or 0.7 in log units.[34] A potential concern is the relevance of geographic cost factors in predicting facility locations. Appendix Table A1 displays first-stage regression results of actual shipping distances on predicted values, showing strong correlation with high F-statistics. Further, correlations between counterfactual shipping distances and lagged GDP and GDP growth are weak or negative, supporting the instrument's robustness against demand-side factors.

**Step 3: Demand Shift.** In the model section, matching efficiency between consumers and online retailers is represented as a demand shifter, $\mu$, in a CES consumption function. To isolate $\mu$, I use the ratio expenditure share on local online retailors to brick-and-mortar retailors, which can be expanded and simplified as shown in equation (21):

$$\frac{x_{nn}^{j,R}}{x_n^{j,B}} = \frac{\Psi_n^j (c_n^{j,R}/\mu)^{1-\sigma}}{\frac{1}{O}(c_n^{j,B})^{1-\sigma}} = \mu^{1-\sigma} \cdot \psi_n^j (\frac{w_n^{j,R}}{w_n^{j,B}})^{1-\sigma}. \tag{21}$$

This equation indicates that through the lens of the model, given unit costs for both online and brick-and-mortar retail and the probability of online retailer locations, the regional ratio of e-commerce to brick-and-mortar expenditure reflects only online match efficiency. With all variables in equation (21) calibrated previously, except for the regional wage ratio

[34]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.

Table 7: Estimates of Iceberg Cost Change and Demand Shift

| $\delta^{\text{dur}}$ | $\delta^{\text{nondur}}$ | $\hat{\kappa}$ | $\mu$ |
|---|---|---|---|
| 1.5 | 2.1 | 0.97 | 1.27 |
| [0.2] | [0.6] | [0.15] | [1.46] |

between e-commerce and brick-and-mortar retail, which can be obtained from CBP data, the demand shift parameter $\mu$ can then be estimated. Table 7 shows the estimate of $\mu$, indicating that on average, consumers in 2017 were 27 percent more likely to purchase from online retailers due to Amazon's match efficiency.

# 6 The Impact of Amazon on Regional Economies

What are the equilibrium impacts of an e-commerce shock on the economy, especially regarding regional disparities in economic outcomes? In this section, I assess the effects of Amazon's expansion on both the aggregate and regional economies, examining impacts on welfare, employment, and inequality, with a focus on regional variation in outcomes. I then identify the channels through which the Amazon shock influences these economic measures to clarify the underlying mechanisms. Lastly, I implement a simple revenue redistribution scheme equalizing the welfare changes across regions while remaining budget neutral.

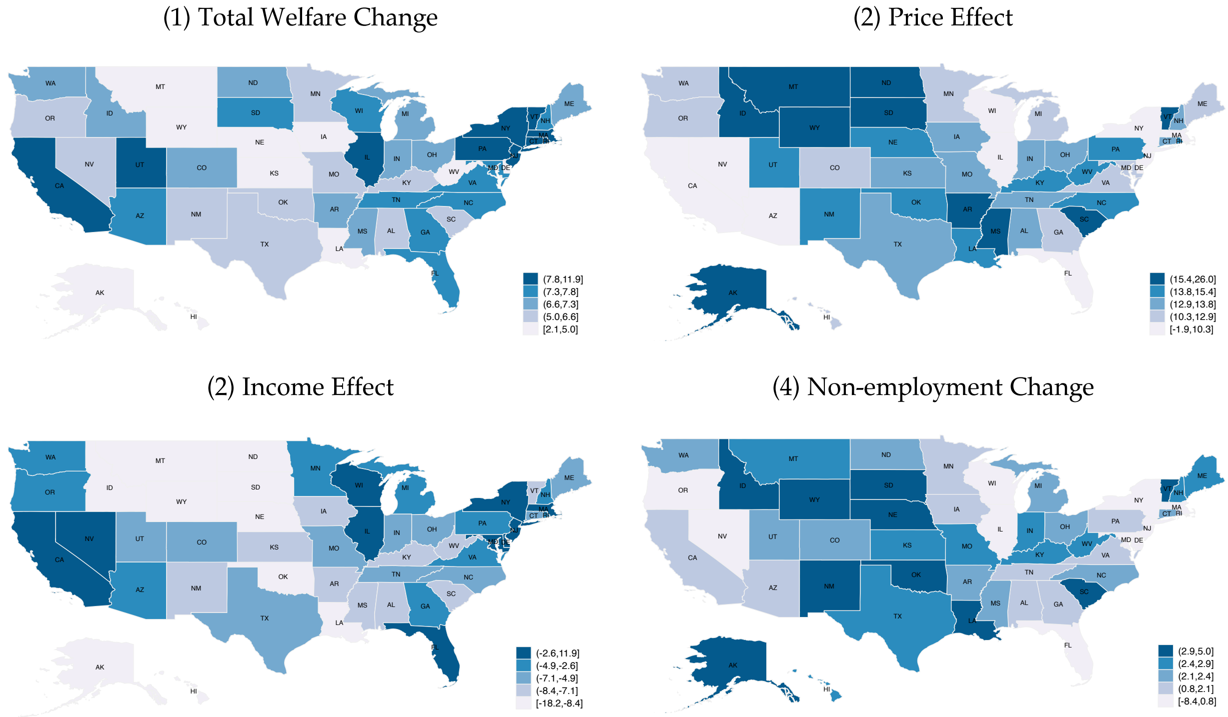
The counterfactual analysis begins with the 2007 initial equilibrium and introduces the Amazon shock, represented by changes in the iceberg cost ($\hat{\kappa}^{R}_{nm}$), online match efficiency ($\mu$), and the measure of online retailers ($\Psi^{j}_{m}$ and $O$), while holding all other fundamentals constant.

## 6.1 Welfare and Employment Outcomes

**Welfare**: Starting with welfare changes from the Amazon shock, Figure 7 panels (A) to (C) show state-level changes in total welfare, with separate decompositions into price and income effects. On average, total welfare across states increases by 6.7 percent–this is primarily due to price effects, while income effects have a negative impact on welfare. Specifically, price reductions from Amazon's expansion alone would have led to a 13.1 percent welfare increase. However, Amazon's growth also reallocates economic activities and labor, resulting in differential income changes across regions. This income effect alone would have reduced total welfare by 5.4 percent.

The aggregate welfare changes and their components reveal a significant variation

Figure 7: Welfare, Employment Changes and Decompositions




*Notes*: This figure shows state-level changes in welfare, its decomposition into price effects and income effects, as well as non-employment due to Amazon's expansion. Welfare changes and employment changes are calculated from model counterfactual analysis using hat algebra.

across regions. As shown in Figure 7, states on the East and West Coasts generally experience larger welfare gains, while Midwestern states see smaller increases. Breaking down to the sources, we observe that income effects exhibit greater dispersion than price effects, indicating that income changes drive much of the regional heterogeneity in welfare outcomes.

Focusing more specifically on the underlying mechanisms, states with a comparative advantage in online retailing—such as New York, Massachusetts, Wisconsin, California, and Florida—see positive income effects due to employment gains, which boost their overall welfare. These states also benefit from a diverse industrial composition, supporting wage growth across all sectors. In contrast, Midwestern states like North Dakota, Montana, and Wyoming face negative income effects from retail sector competition and worker shifts to lower-wage sectors. However, as these regions initially have lower consumer spending on online retail goods, they experience significant positive price effects, which help to offset the income-related welfare losses.

**Employment**: I now turn to discuss the employment changes implied by the Amazon

Table 8: Employment Changes by Sector and State Groups

| | All States | | Below 50th Percentile Online Sales Density | |
|---|---|---|---|---|
| Sector | Mean | Std. Dev. | Mean | Std. Dev. |
| Manufacturing | -4.3 | (7.6) | -1.8 | (1.1) |
| Online Retail | 109.8 | (97.8) | 63.3 | (64.8) |
| Brick-and-Mortar | -11.1 | (8.0) | -8.6 | (1.2) |
| Service | -1.6 | (7.9) | 1.2 | (1.2) |
| Non-Employment | -1.3 | (8.1) | 1.7 | (0.8) |

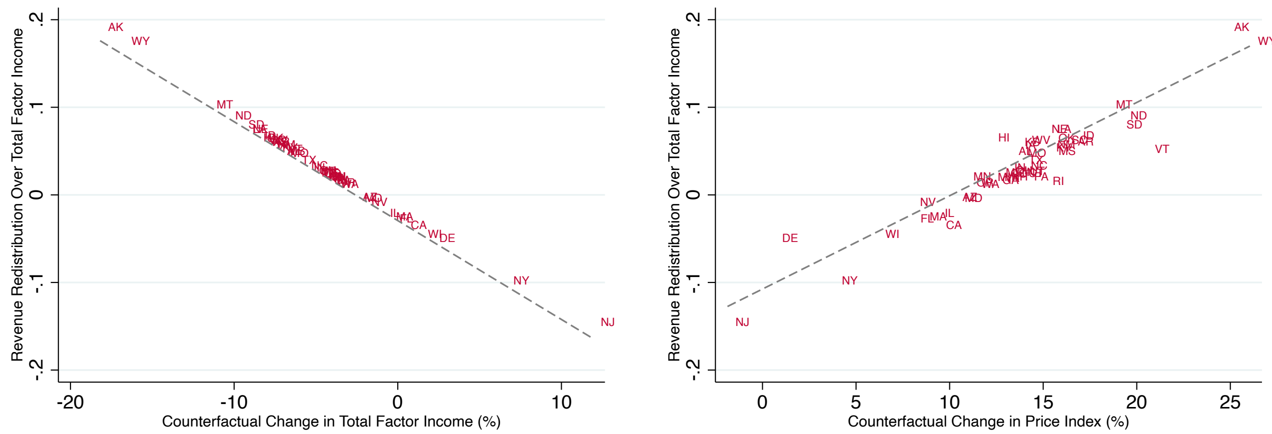
*Notes*: This table summarizes the sectoral employment shifts in response to Amazon's expansion based on the model counterfactual analysis. The data on online sales density comes from Keepa. The units are ratio relative to the 2007 baseline economy.

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 rise in non-employment, there is significant regional variation. As illustrated in Figure 7, Midwestern states, which lack a comparative advantage in online retailing and have less industrial diversity, show a stronger shift toward non-employment and service sectors. The last two columns of Table 8 reveal that in states where online seller density is below the 50th percentile, online retail employment has grown by about 63.3 percent, which is 46.5 percentage points lower than the overall increase. At the same time, these states have seen higher reallocation to service and non-employment sectors, with rates increasing by 1.2 and 1.7 percent, respectively.

**Implications for Inequality**: As discussed above, the impacts of an Amazon shock carry significant distributional implications. Although total welfare has increased and non-employment has decreased, substantial disparities exist across states. To quantify this dispersion, I examine changes in the Gini index. Between 2007 and 2017, Amazon's expansion led the Gini index of welfare or GDP per capita to rise from 0.11 to 0.13, a 20 percent increase. Meanwhile, the Gini index for non-employment grew from 0.05 to 0.25, a fourfold increase. These results indicate a widening gap in both welfare levels and employment opportunities across regions due to Amazon's influence.

Figure 8: Redistribution Amount and Relation to Income and Price Effects




## 6.2 Revenue Redistribution Policy

I now propose a simple ex-post revenue redistribution policy to address the inequality caused by Amazon's expansion. The government aims to equalize welfare changes across all 50 states ($\hat{W}_n = \frac{\hat{Y}_n}{\hat{P}_n}$) at a common level $k$. This is achieved by reallocating regional nominal incomes from $Y_n'$ to $\tilde{Y}_n'$ while ensuring budget neutrality. Specifically, the total nominal income changes based reallocation equal the net impact of Amazon's expansion, denoted as $B$, therefore the government incur no costs: $\sum_{n=1}^{50}(\tilde{Y}_n' - Y_n') = B = \sum_{n=1}^{50}(Y_n' - Y_n)$. Combining these conditions, the scaling factor $k$ is determined as $k = \frac{B+\sum_{n=1}^{50} Y_n}{\sum_{n=1}^{50} Y_n \cdot \frac{\tilde{P}_n}{P_n}}$; the redistributed amount for each region is calculated as $(\tilde{Y}_n' - Y_n') = Y_n \cdot k \cdot \frac{\tilde{P}_n}{P_n} - Y_n'$.

The common welfare change $k$ is calculated to be 0.97, indicating that achieving equal welfare changes across states requires an average welfare reduction of 3 percent. Figure 8 shows the redistribution adjustments needed to achieve this balanced outcome. Since the variation in welfare changes is primarily driven by income effects, the redistribution amounts are strongly negatively correlated with these effects. States that experience increases in total factor income, such as New York, New Jersey, and Delaware, redistribute to states with declining total factor income, such as Wyoming and Montana. However, because income effects and price effects (changes in the price index) are negatively correlated and welfare changes are less strongly linked to price effects, there is a weak positive correlation between price effects and redistribution amounts. In sum, the budget-neutral redistribution is primarily driven by income effects and achieves a uniform, though slightly negative, welfare change.

# 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 [35]

$$\bar{c}_m^j = \frac{\mu}{\tilde{\sigma}} \left(\frac{\sigma}{\eta^j}\right)^{\frac{1}{1-\sigma}} \left[\frac{w_n^{j,R} f_m}{\sum_n \left(\kappa_{nm}^R / P_n^{R,j}\right)^{1-\sigma} Y_n}\right]^{\frac{1}{1-\sigma}}. \tag{22}$$

The trade flow equation, can then be derived to link to the relative productivity and 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.[36] 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 $\kappa_{nm}^R$) would increase the trade flow.

$$X_{nm}^{j,R} = \lambda w_m^{j,B} l_m^{j,B} \left(\left(w_m^{j,R}\right)^{\gamma^j} \left(P_m^{j,M}\right)^{\left(1-\gamma^j\right)} \frac{\kappa_{nm}^R}{\mu}\right)^{-\rho} \times \left[\frac{w_m^{j,R} f_m}{\sum_n \left(\frac{\kappa_{nm}^R}{P_n^{R,j}}\right)^{1-\sigma} Y_n}\right]^{\frac{\sigma-\rho-1}{1-\sigma}} \eta^j Y_n \left(P_n^j\right)^{\sigma-1}. \tag{23}$$

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 $P_n^j$:

[35] 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 $\bar{z}_m^j = \left(w_m^{j,R}\right)^{\nu_i^j}\left(P_m^{j,M}\right)^{\left(1-\gamma_i^j\right)} \frac{\tilde{\sigma}}{\tilde{\mu}} \left(\frac{\sigma}{\eta^j}\right)^{\frac{1}{\sigma-1}} \left[\frac{w_m^{j,R} f_m}{\sum_n \left(\kappa_{nm}^R P_n^{R,j}\right)^{1-\sigma} Y_n}\right]^{\frac{1}{\sigma-1}}$

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

$$X_{nn}^{j,B} = \left( \left( \omega_n^{j,B} \right)^{\gamma^j} \left( P_n^{j,M} \right)^{\left(1-\gamma^j\right)} \right)^{1-\sigma} \eta^j Y_n \left( P_n^j \right)^{\sigma-1} \tag{24}$$

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.

# 8 Conclusion

The expansion of e-commerce, exemplified by Amazon's growth, has significantly reshaped regional economies, creating both opportunities and challenges. This paper leverages data on the universe of products and retailers on Amazon, combined with detailed information on Amazon's fulfillment and distribution facilities, to provide new empirical insights into the spatial dynamics of e-commerce. The data reveal five key patterns pointing to differential spatial agglomeration of online retailers associated trade flows. These findings, which are new to the literature, highlight how e-commerce influences trade and regional economic structures.

Using these empirical insights, the paper develops a multi-sector spatial trade framework incorporating consumer search, retailer location decisions, and regional comparative advantages. Quantitative counterfactual results demonstrates that while e-commerce has driven price reductions, more consumption varieties, and improved overall welfare, it has also deepened regional disparities and altered labor market dynamics. States with a comparative advantage in online retailing and diverse industrial structures, such as New York and California, experienced welfare gains due to positive income and price effects. In contrast, Midwestern states, like Wyoming and Montana, faced income losses and increased reliance on lower-wage sectors, despite benefiting from price reductions.

Employment patterns also shifted, with workers moving away from brick-and-mortar retail and manufacturing toward online retail, contributing to a 1.3 percent decline in non-employment nationally. However, less advantaged regions saw higher shifts to non-employment and service sectors. These dynamics exacerbated inequality, with the Gini index of welfare and non-employment increasing significantly. To address these disparities, I introduce a revenue redistribution policy, reallocating regional incomes to equalize welfare changes while remaining budget-neutral.

# 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*.

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. 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 $\sigma$

| | **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 19. 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 $\epsilon^j_{ni}$ 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 $\chi^j$ is the variance of the effective match value $\omega^j_{ni}$, 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 $\sigma_j$, 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 $\eta^j$ 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 $\Psi$ 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), (12), (3), (4), as well as (**??**) to (17) hold.*

**Comparative Statics.** Computing the equilibrium outcomes out of the model requires solving a system of nonlinear equations (10), (12), (3), (4), and (**??**) to (17), 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:

$$\hat{\pi}^0_n = \frac{\hat{A}^0_n(\hat{w}^0_n)^{\nu_n}}{\hat{\Phi}_n},\ \hat{\pi}^{j,K}_n = \frac{\hat{A}^{j,K}_n(\hat{w}^{j,K}_n)^{\nu_n}}{\hat{\Phi}_n},\ \text{where } \hat{\Phi}_n = \sum_{h=0}^{J}\sum_{K=M,R}\pi^{K,h}_n\hat{A}^{K,h}_n(\hat{w}^{K,h}_n)^{\nu_n}. \tag{25}$$

The input costs are given by:

$$\hat{c}^{j,M}_n = \hat{\omega}^{j,M}_n, \quad \hat{c}^{j,R}_n = (\hat{\rho}^{j,R}_n\hat{\omega}^{j,R}_n)^{\gamma^j_n}(\hat{P}^{j,M}_n)^{1-\gamma^j_n}, \tag{26}$$

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

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

The trade shares are given by:

$$x'^{j,M}_{ni} = x^{j,M}_{ni}(\frac{\hat{\kappa}^M_{ni}\hat{c}^{j,M}_i}{\hat{P}^{j,R}_n})^{-\theta_j}\hat{T}^j_i, \quad x'^{j,R}_{ni} = x^{j,R}_{ni}(\frac{\hat{\kappa}^R_{ni}\hat{c}^{j,R}_i}{\hat{\mu}^j_{ni}\hat{P}^{j,R}_n})^{1-\sigma^j}, \tag{27}$$

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

Market clearing conditions now become:

$$X_n'^{j,R} = \sum_{i=1}^{N} x_{in}'^{j,R} \eta^j \left[ \sum_{k=0}^{J} \sum_{K=M,R} (\frac{1}{1-\beta_i}) \hat{\rho}_i^{K,k} \hat{w}_i^{K,k} \hat{l}_i^{K,k} \rho_i^{K,k} w_i^{K,k} L_i^{K,k} - \Omega_i \right], \tag{28}$$

$$X_n'^{j,M} = \sum_{i=1}^{N} (1-\gamma_i^j) x_{ni}'^{j,M} X_n'^{j,R}, \tag{29}$$

$$\hat{w}_n^{j,M} \hat{l}_n^{j,M} w_n^{j,M} L_n^{j,M} = \beta_n \hat{X}_n^{j,M}, \quad \hat{w}_n^{j,R} \hat{l}_n^{j,R} w_n^{j,R} L_n^{j,R} = \frac{1}{\hat{\rho}_i^{j,R}} \gamma_n^j \beta_n \hat{X}_n^{j,R} \tag{30}$$

Equations (25)-(28) 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 $(\Omega_n)$, as well as parameters with respect to value-added shares ($\beta_n$ and $\gamma_n^j$), consumption shares ($\eta_n^j$), and trade elasticities ($\sigma^j$ and $\theta^j$). 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:

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

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