

**Import Competition, the Value of Time, and
Intermediation in Trade**

**A DISSERTATION
SUBMITTED TO THE FACULTY OF THE
UNIVERSITY OF MINNESOTA
BY**

Ethan S. Singer

**IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY**

THOMAS HOLMES, Advisor

June, 2016

© Ethan S. Singer 2016
ALL RIGHTS RESERVED

Acknowledgements

I am indebted to my advisor, Thomas J. Holmes and my committee members Joel Waldfogel, Jim Schmitz, and Naoki Aizawa for their advice and helpful comments. I would also like to thank Ellen McGrattan, Erzo Luttmer, Amil Petrin, Matthew H. Shapiro, Dominic Smith, Dirk van Duym, Kailin Clarke, Bitmaro Kim, Boyoung Seo, Rishabh Kirpalani, Filippo Rebessi, and members of the applied micro workshop at the University of Minnesota for helpful advice and comments.

Much of the work in the first chapter was carried out at the Minnesota Census Research Data Center. Any opinions and conclusions expressed herein are those of the author and do not necessarily represent the views of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed. Thanks to Rachelle Hill for help within the Minnesota Census Research Data Center.

Special thanks to Thomas J. Holmes who co-authored the second chapter. We thank Dominic Smith and Jonathan Williard for their research assistance on this chapter. The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Minneapolis, the Federal Reserve Board or the Federal Reserve System.

Abstract

This dissertation contains three chapters, each focusing on a different aspect of the micro structure of trade flows and shipments to and within the United States. All three chapters make use of rich micro level data on trade flows.

In the first chapter, I study imports of heavy goods and how these products compete with local production. Although imports of heavy goods are often thought to stay near the coast, these goods frequently travel much farther. I develop a structural model of demand with large choice sets where all transaction prices are observed, but offered prices are unknown. Using unique micro data linking shipments to locations, I find that the rail network plays an important role in allowing shipments to reach distant locations. Once firms pay the fixed cost of accessing this network, shipments can cheaply go far. As a result, the ability of imports to easily reach all locations disciplines prices even in areas with low import shares. Because domestic producers have access to the same transportation infrastructure, they also discipline prices in distant locations. I also find that the ability to price discriminate by location enables this type of competition over long distances.

The second chapter is coauthored with Thomas Holmes. This chapter studies the shipment of internationally-traded goods, focusing on the path the goods take to get to their final destination, and in particular taking into account the internal geography of the destination country. We use Wal-Mart's distribution network as our primary empirical example, modeling the flow of shipments from origination country to U.S. port to import distribution center and finally on to the final consumer. The paper estimates the costs incurred by Wal-Mart on account of transit time, by studying the choice behavior, as Wal-Mart trades off shorter transit times in exchange for higher freight rates. The paper assembles a variety of new data sources including bills of lading for ocean shipping transactions that have been processed to match to data on selected individual firms, and that have been merged with GPS data on ocean vessels to ascertain shipping times. As an application, the paper considers the recent labor market disturbance on the West Coast ports of the United States. The paper produces

estimates of the cost of this disruption as well as the gains to Wal-Mart of having a “four corners” distribution network.

The third chapter looks at internal shipments within the United States. Shipment distances in the Commodity Flow Survey (CFS) are disproportionately very short. This chapter looks at the distribution of shipment distances by industry using the 2012 public CFS. I use this data to support the explanation that wholesale networks result in low average distances. When measuring the total distance from manufacturer to consumer, if all layers of the wholesale network are considered then distances would be much larger.

Contents

Acknowledgements	i
Abstract	ii
List of Tables	vi
List of Figures	viii
1 Imports, Heavy Goods, and Competition Away from the Coast	1
1.1 Introduction	1
1.2 Data and motivating facts	5
1.2.1 Data	5
1.2.2 Heavy goods and distance shipped	9
1.3 Model	15
1.3.1 The Choice of Supplier	16
1.3.2 The Choice of Mode	17
1.3.3 Suppliers	17
1.4 Estimation and Results	19
1.4.1 Step 1: Estimating Mode Choice Parameters	19
1.4.2 Step 2: Demand and Cost Estimation	21
1.4.3 Counterfactual simulations	27
1.5 Conclusion	35
2 Using Four Corners to Import Goods: Estimates for Wal-Mart	37
2.1 Introduction	37

2.1.1	Some Previous Literature	41
2.2	Model	42
2.2.1	Description of the Model	42
2.2.2	Port-Choice Problem	45
2.2.3	The DC Allocation Problem with Product-Level Customization	46
2.2.4	The DC Allocation Problem without Customization	47
2.3	Data	48
2.3.1	Bill of Lading Data	48
2.3.2	Ocean Time in Transit	55
2.3.3	Ocean Freight Rates	55
2.3.4	Internal Geography and Ultimate Consumers	56
2.3.5	Domestic Freight and Time	56
2.4	Preliminary Analysis	58
2.4.1	Import Distribution Network	58
2.4.2	Order Allocation and Country of Origin	61
2.4.3	Order Allocation in Response to a Temporary Port Shock	66
2.5	Quantitative Analysis	68
2.5.1	The Port Choice Problem	68
2.5.2	DC Choice Problem	71
2.5.3	Analysis of the Estimated Model	72
3	Understatement of Shipping Distances in the United States	79
3.1	Introduction	79
3.2	Data	80
3.3	Model	85
3.4	Conclusion	87
	References	88

List of Tables

1.1	Summary Statistics, Selected Import Commodities, 2007	6
1.2	Summary Statistics, Inland Transportation Modes, 2012	7
1.3	Summary Statistics, Heavy Domestic Shipments, 2012	8
1.4	Estimates of Distribution Parameter	12
1.5	Mode Choice Parameter Estimates	20
1.6	Supplier Choice Parameter Estimates	24
1.7	Supplier Choice Parameter Values, Price Discrimination	27
2.1	Information Contained in Two Example Bill of Ladings	50
2.2	Counts of Bills of Lading and Containers in Wal-Mart Sample	52
2.3	Distribution of Container Imports by Country of Origin	53
2.4	Summary Statistics of Final Sample	54
2.5	Wal-Mart Product Categories	59
2.6	Distribution of Products by Number of DC Destinations	60
2.7	Mean Share Allocation across Destination DCs	63
2.8	Mean Share Allocation across Destination DCs	64
2.9	Port Choice Given Destination DC	65
2.10	Port Choice Estimates	71
2.11	DC Choice Estimates, General Merchandise	72
2.12	Estimated Cost Per Container	73
2.13	Gains from Customization	75
2.14	Estimated Share of China to Houston DC Shipments Going Through Port of Houston for Different Disruption Costs	76
2.15	Estimated Cost per Container of Labor Disruption Across Scenarios	77

3.1	Averages and Distribution of Shipment Distances by Establishment Industry Type	81
3.2	Averages and Distribution of Shipment Distances by NAICS Code for Wholesale and Warehousing Establishments	82
3.3	Averages and Distribution of Shipment Distances by Metro Area for Textile Shipments from Wholesale and Warehousing Establishments	84

List of Figures

1.1	Map of Counties with Cement Manufacturers, 2007	9
1.2	Distribution of Shipment Distances	10
1.3	Log of Cumulative Distribution, Heavy Good Shipments	11
1.4	Average Price by Distance: Heavy Good Shipments	13
1.5	Log of Cumulative Distribution, by Mode	14
1.6	Rail Share: Heavy Good Shipments	15
1.7	Fitted Value Import Shares by County	28
1.8	Fitted Value Import Shares from Mexico	29
1.9	Modeled Import Shares from Mexico, with Anti-Dumping Duty	30
1.10	Modeled Import Shares, High Tariff	31
1.11	Fitted Value Total Price	32
1.12	Total Price if Imports Were Prohibited	33
1.13	Modeled Import Shares Without Rail	34
2.1	Wal-Mart's Import DC System	40
2.2	Voyage Duration and Port Choice During West Coast Port Slowdown	67
2.3	Plot of Estimated DC Share Allocations by Time for Two Samples	69

Chapter 1

Imports, Heavy Goods, and Competition Away from the Coast^{*}

1.1 Introduction

The cost of moving goods by ship across oceans is generally cheap compared to the cost of moving goods over land. For this reason, imports of heavy goods, such as construction materials, tend to penetrate coastal areas of domestic markets, as compared to markets in the interior. Nevertheless, a striking fact about heavy goods is that when they do get shipped to the interior, they often go relatively far, sometimes even a thousand miles or more. This paper documents this fact about the movement of heavy goods, and provides an explanation for the pattern based on differences in fixed and marginal costs of alternative modes of inland transportation. The paper develops a model of the market for heavy goods in which both domestic firms and foreign firms compete across regional markets. The model is estimated with rich transaction-level data on both foreign and domestic shipments by origin and destination. The data also includes

^{*}Much of this work in this chapter was carried out at the Minnesota Census Research Data Center. Any opinions and conclusions expressed herein are those of the author and do not necessarily represent the views of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed.

information about mode choice of transportation. The main finding of the paper for the primary sample considered (the cement industry in 2007) is that the threat of imports disciplines pricing of domestic firms, even deep into the interior. The fundamental factor underlying this result is nonlinearity of transportation costs—once fixed costs are paid to lower marginal transportation costs, imports can go deep into the interior.

The key assumption about transportation technologies is that different modes vary in fixed and marginal costs. The two main modes of overland transportation are trucking and rail. Trucking tends to have low fixed cost, but a relatively high marginal cost per mile, compared to rail. For rail, part of the higher fixed cost is due to the issue that the final destination is generally not at a railhead, resulting in an extra cost that is incurred to switch transportation modes to complete the last mile of the journey. Trucks however are flexible and can more easily go anywhere, but are more labor and energy intensive than rail, resulting in higher costs per mile. This cost structure suggests that trucks will be used primarily for short haul shipments and rail for longer shipments and this is exactly what happens in the data. For example, as demonstrated in the data below, over 90 percent of heavy-good shipments less than 100 miles travel by truck, while for shipments above 500 miles, over 90 percent travel by rail. The break-even distance for truck and rail is estimated to be between about 350 and 425 miles, consistent with transportation studies.¹

A highlight of this paper is its use of rich transaction-level census data on both foreign and domestic shipments. Every import shipment is observed and I match these imports to the recipient establishment. Transaction data has been used in earlier work by Bernard, Jensen, Redding, and Schott (2007) and related papers where transactions are linked to firms and where trade is not disaggregated within the United States.² Rather than matching to the firm, I match to the establishment allowing for the exact location of the shipment and the distance traveled to be identified. I also use data on shipments within the United States and this is the first paper to combine transaction level trade data with domestic shipment data, allowing for a complete picture of transaction-level

¹A U.S. Department of Transportation study estimates that the rail-truck break-even distance is approximately 375 miles when diesel is \$2 per gallon. “Impact of High Oil Prices on Freight Transportation: Model Shift Potential In Five Corridors”, prepared for Maritime Administration, U.S. Department of Transportation, October 2008, www.marad.dot.gov/wp-content/uploads/pdf/Modal_Shift_Study_-_Executive_Summary.pdf. See Oum (1979) for another study on mode choice.

²See also Bernard, Jensen, and Schott (2009) and Bernard, Jensen, Redding, and Schott (2010).

trade flows. The third element of the data is the locations of all domestic manufacturers which is used to identify the geographic relationships between all plants and customers.

The structural model of supplier choice for consumers of heavy goods takes into account both geography and the different transportation technologies. Two choices are modeled. First, consumers of heavy goods choose which supplier to use, and select among domestic and foreign firms. Second, consumers choose the domestic shipping mode of the supplier. Mode choice is observed for the domestic shipments and this identifies the relative fixed and variable transportation costs of different modes (truck or train). Because distances between all suppliers and all consuming firms are known, expected transportation costs can be constructed and the mode choice decision can be integrated out of the decision problem that is estimated using the full sample. Suppliers engage in geographic price discrimination when setting prices. One issue the estimation confronts is that prices are observed for all transactions that occurred, while offered – but not accepted – prices are unobserved. The unobserved offered prices are obtained using the profit maximizing conditions of the suppliers and estimates of marginal costs based on the observed prices of the suppliers.

I document the fact that conditional on a shipment traveling beyond the coast, it tends to go very far and propose using multiple transportation technologies to explain this pattern. I entertain alternative explanations for this, such as variations in quality, and show that quality heterogeneity cannot explain the shipping patterns. The parameter estimates for the different fixed and variable costs of each mode provide a break-even distance for trucking and rail that is consistent with industry estimates. Counterfactual simulations demonstrate the importance of two central elements of the model: multiple transportation modes and the ability to price discriminate. First, when there are multiple transportation modes, changes in trade costs result in larger changes in import penetration compared to a model with a single linear-cost transportation technology. If transportation costs were assumed to be linear in distance, the model fails to explain that heavy goods are often shipped far, and a decrease in trade costs would result in a more gradual expansion of the areas exposed to trade. Second, the ability to price discriminate enables firms to compete for distant sales even when at a transportation cost disadvantage. If uniform prices were assumed, as is common in the demand estimation literature when only average prices are observed, the extent of import competition

would be underestimated and import penetration would not be as responsive to import price changes. Both non-linear transportation costs and price discrimination are needed for the model to be consistent with the occasional long-distance shipments observed in the data. Although most imports remain near the coast, a key finding is that imports are able to discipline prices in the interior because the transportation technology makes it possible for these heavy goods to travel long distances.

The paper is closely related to research in the trade literature that uses Census transaction data such as Bernard, Jensen, and Schott (2009), Bernard, Jensen, Redding, and Schott (2007), and Bernard, Jensen, Redding, and Schott (2010). By linking the trade transactions to establishments, I can examine trade transactions at a more granular geography than was previously possible. There is an emerging literature integrating international trade and regional economics. Although the international trade literature has generally viewed the U.S. as a single national market, this research, including that of Holmes and Stevens (2014) and Coşar and Fajgelbaum (forthcoming), has begun to recognize the value of viewing the U.S. as a collection of disaggregated markets. By modeling trade flows to specific locations, this paper quantifies regional differences and shows that there is substantial within country heterogeneity. The importance of fixed costs in trade has been emphasized by Melitz (2003) and others, and this paper explores fixed costs of entering domestic transportation networks instead of looking at fixed costs in the exporting decision. Hummels (2007) documents the relationship between transportation costs and the volume of international trade. This paper extends the literature on transportation costs and trade by modeling inland transportation decisions in conjunction with supply decisions.

This paper also relates to the industrial organization literature looking at spatially differentiated competition including Beckert, Smith, and Takahashi (2015) and Miller and Osborne (2014). Like those papers, I allow for price discrimination, but this paper is different in that it uses shipment level census data and considers the role of different transportation modes in determining transportation costs. Unlike much of the industrial organization literature, such as Ryan (2012) and Miller and Osborne (2014), that downplay the competitive effects of distant products, this paper allows products to compete over long distances. Salvo (2010a) and Salvo (2010b) look at the cement industry in Brazil examining the role of the threat of import entry and trade costs.

Salvo (2010a) finds that the threat of competition is important, however, because most economic activity in Brazil is near the coast, that study does not need to consider inland transportation technologies, which is an important feature of these markets in the United States. The idea that imports can influence local producers without actually shipping any product has been documented by Schmitz (2005) which shows that in response to the threat of import competition, iron ore producers exhibited substantial increases in productivity.

1.2 Data and motivating facts

This section proceeds first by describing the basic elements of the data and then by documenting the key facts in the data that motivate the specification of the model.

1.2.1 Data

Three primary confidential micro data sources are used in the analysis.³ The first source of data is shipment level import data from the Census. This data is obtained from customs declarations and contains the complete records of all imports into the United States. Shipment level imports are matched to the receiving establishment and contain information on product type, date of shipment, weight, customs value, and country of origin. The linkage to establishments extends the firm links used by Bernard, Jensen, Redding, and Schott (2007) among others and allows for the locations of the shipment destinations to be identified. Shipments are matched to establishments using an algorithm that relies on information from the customs form such as destination state and firm identifiers and that takes advantage of the fact that establishments in the same industry tend to receive shipments of the same type of commodities. There is also a public version of the import data, which is aggregated to the commodity, month, origin country, and port level and can be used to calculate aggregate statistics. Table 1.1 provides summary statistics for a number of heavy goods.⁴ From the aggregate

³Data is available at Federal Statistical Research Data Centers. All results have been reviewed to ensure that no confidential information is disclosed.

⁴This table illustrates the types of products that I classify as heavy goods and does not include all commodities that fit that classification. Although I will define heavy shipments as those with unit values of 15 cents per kilogram or less, some HS codes with average unit values above 15 cents contain shipments with values below that cutoff. For example, the 4 digit HS code for fuel wood contains

data, we can see that heavy goods are commonly imported. For example, the published census tabulations show that in 2007 the 22.7 million metric tons of cement valued at \$1.3 billion were imported into the United States in 48 thousand import transactions. The average unit value of these cement imports was 5.8 cents per kilogram.

Table 1.1: Summary Statistics, Selected Import Commodities, 2007

HS code (4 digit)	Description	Customs Value (\$ millions)	Metric Tons (millions)	Avg Unit Value (cents per kg)	Shipments (cards)
2517	Pebbles, Gravel	141.15	16.68	0.8	5,271
2520	Gypsum	111.70	9.40	1.2	1,268
2501	Salt	170.92	8.64	2.0	5,830
2807	Sulfuric Acid	119.95	2.62	4.6	26,468
2701	Coal	1,723.10	32.87	5.2	10,479
2523	Cement	1,323.87	22.73	5.8	48,402
2522	Quicklime	42.40	0.34	12.5	10,984
3102	Fertilizers, Nitrogenous	3,101.67	24.38	12.7	67,473
2306	Oilcake	206.04	1.58	13.0	24,931
4707	Waste Paper	100.59	0.69	14.5	29,476
2806	Hydrogen Chloride	26.51	0.18	14.6	3,246
3104	Fertilizers, Potassic	1,685.80	10.81	15.6	111,786
4401	Fuel Wood	145.90	0.86	16.9	30,783

Source: 2007 public use import data, selected 4-digit HS codes and abbreviated descriptions.

The second source of data is the Commodity Flow Survey (“CFS”), which is a sample of domestic shipments. In the 2007 CFS, approximately 13.5% of covered establishments were surveyed.⁵ Survey respondents reported between 20 and 40 randomly selected shipments for each of four one-week periods, one week in each quarter of the year. This shipment level data is also very detailed and includes product type, date of shipment, weight, value, and mode of transit. There is a public version of the 2012 CFS that is similar in structure to the restricted data, but lacks the detailed commodity and industry codes. The public CFS has product codes at the two-digit level and does therefore not directly enable the identification of specific commodities. I define heavy goods throughout the paper as shipments with unit values of less than 15 cents per

detailed HS codes for nonconiferous wood chips which have average unit values of 4.8 cents and also contains the HS code for wood shavings which has the average unit value of 18.3 cents.

⁵U.S. Department of Transportation and U.S. Department of Commerce, “2007 Economic Census, Transportation, 2007 Commodity Flow Survey” <http://www.census.gov/econ/cfs/2007/US%20FINAL.pdf>

kilogram. A sample of cement shipments is approximated in the public data by taking the heavy goods shipments with the 2-digit commodity code that contains cement and 3-digit industry code that contains cement manufacturers. Table 1.2 summarizes the 2012 CFS data by mode of transportation. The top panel contains summary statistics for all heavy good shipments, and the bottom panel contains summary statistics for the cement sample. Rail shipments go farther than truck shipments on average and the majority of shipments travel by truck or train. Because of the small number of shipments by other modes and since often the other mode is unknown, the analysis uses only truck and rail shipments.⁶

Table 1.2: Summary Statistics, Inland Transportation Modes, 2012

	Value (\$ millions)	Metric Tons (millions)	Observations	Avg Unit Value (cents per kg)	Avg Distance (miles)
All Heavy Goods					
Truck	95,500	2,901	260,325	3.3	45
Rail	33,319	879	9,673	3.8	499
Other modes	15,063	321	3,607	4.7	218
Total	143,882	4,101	273,605	3.5	156
Cement					
Truck	24,998	404	46,707	6.2	35
Rail	1,532	23	890	6.7	220
Other modes	443	5	118	9.5	284
Total	26,973	432	47,715	6.2	48

Source: 2012 public use CFS. Includes shipments with unit values of less than 15 cents per kilogram. Figures reflect weighting by the sampling weights. Average distances are great circle distances weighted by shipment weights. Truck includes CFS mode codes for private and for-hire trucks. Rail includes rail and truck multimodal shipments. Cement sample defined as heavy good shipments with SCTG code equal to 31 and NIACS code equal to 327.

Table 1.3 summarizes the heavy domestic rail and truck shipments by commodity. Although the two-digit commodity codes in the public data are fairly broad, there are big differences in average unit values and average shipment distances among the different classifications. Coal on average travels the farthest compared to other heavy goods, with an average distance of 487 miles, while the category with the most shipments in terms

⁶Other modes include multimodal trips that are not rail-truck, observations where mode was suppressed from the public sample, pipelines, and water.

of weight, gravel and crushed stone travels 27 miles. Cement is contained within SCTG code 31 and this is the third largest category of heavy goods by weight.

Table 1.3: Summary Statistics, Heavy Domestic Shipments, 2012

SCTG (2-digit)	Description	Metric			Avg	
		Value (\$ mil)	Tons (mil)	Obs	Unit Value (¢ per kg)	Avg Dist (miles)
12	Gravel and Crushed Stone	15,474	1,304	91,868	1.2	27
15	Coal	27,057	783	9,198	3.5	487
31	Non-Metallic Mineral Products	32,940	544	62,008	6.1	48
11	Natural Sands	5,238	389	27,423	1.3	61
19	Other Coal and Petroleum	10,029	162	13,763	6.2	26
13	Other Non-Metallic Minerals	2,887	102	9,183	2.8	115
26	Wood Products	5,239	99	12,448	5.3	96
20	Basic Chemicals	6,836	74	12,000	9.3	223
4	Animal Feed, Eggs, Honey	4,253	45	5,570	9.4	119
14	Metallic Ores and Concentrates	3,389	43	570	7.8	261
	Other Commodities	15,478	234	25,967	6.6	187
	Total	128,819	3,780	269,998	3.4	151

Source: 2012 public use CFS. Includes shipments with unit values of less than 15 cents per kilogram by truck or rail. Figures reflect weighting by the sampling weights. Average distances are great circle distances weighted by shipment weights.

The final source is the Census Bureau’s Standard Statistical Establishment List (“SSEL”) which provides the locations of domestic manufacturers. The SSEL is only used to identify the locations of suppliers in the choice set that are not in the CFS. A related public data source is the County Business Patterns (‘CBP’) data. In the 2007 CBP there are 7.7 million establishments in the United States, 283 of which are classified as cement manufacturing establishments and 6,063 of which are ready-mix concrete plants, the primary consumer of cement. The cement manufacturing establishments are relatively dispersed throughout the United States and their locations are mapped in figure 1.1.

The analysis focuses on heavy goods because the effects of mode choice are more important for these products where transportation is a large share of the total cost and it is therefore easier to uncover the role of transportation decisions. While these same forces are at work for other more valuable goods, the effects will be less pronounced. The estimation proceeds initially using all imports of cement and the sample of domestic shipments of cement in 2007. Future versions will include estimates for other heavy good

Figure 1.1: Map of Counties with Cement Manufacturers, 2007



Source: 2007 County Business Patterns. Counties with at least one establishment with NAICS code 327310.

industries. Using these three data sources the choice sets are constructed and include all domestic suppliers and any foreign suppliers that supply any firm in the U.S. from 2006-2008.

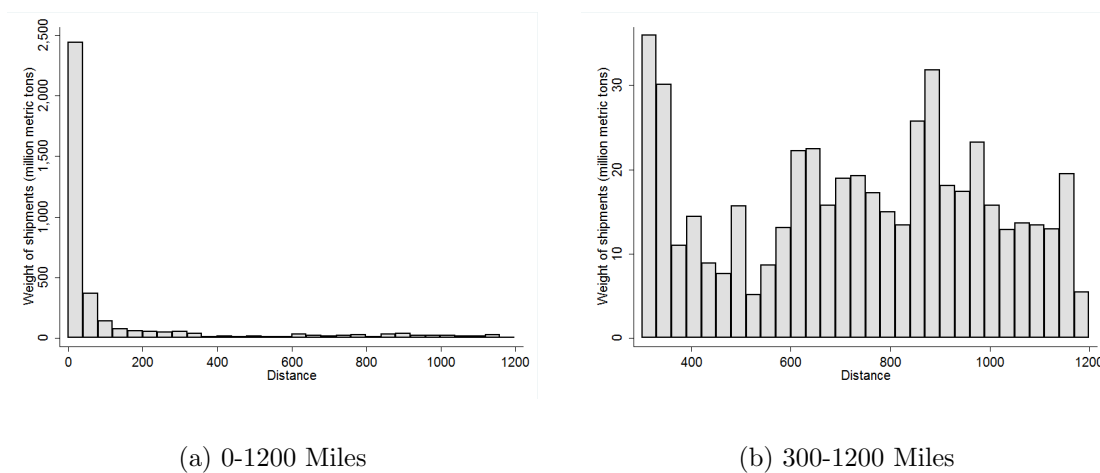
1.2.2 Heavy goods and distance shipped

I turn now to the key motivating fact highlighted in the introduction, that while imported heavy goods generally are shipped to locations near the coasts, conditional on going into the interior these goods are often shipped relatively far. There is a more general pattern that applies both to domestic shipments of heavy goods, as well as imports, namely, that while heavy good shipments tend to be short in distance, conditional on going at least a moderate distance, these shipments often go far. In this subsection, I first document the fact and then consider two simple explanations for the pattern. The first explanation is based on heterogeneity of quality. Specifically, if higher quality goods (with lower transportation costs relative to value) are mixed in with lower quality goods (with higher transportation costs relative to value), then it could simply be that

the goods being shipped far are only the higher quality goods. The second explanation is based on the existence of two modes of transportation, trucking and rail. If heavy goods travel by rail, they tend to be shipped far, regardless of quality. In this section, I provide evidence that the second explanation is the driving force underling the pattern of heavy good shipments.

I begin the analysis with a visual depiction of how shipments vary with distance.⁷ Figure 1.2 presents histograms of shipment distances for heavy goods.⁸ Panel 1.2a includes all distances and panel 1.2b condition on distances of 300 miles or more. Panel 1.2a makes clear that heavy goods tend to go short distances, with the vast majority traveling less than 200 miles. Panel 1.2b shows that for heavy goods that are shipped at least 300 miles, the shipments tend to go far, and the distribution is relatively flat in distance.

Figure 1.2: Distribution of Shipment Distances



Source: 2012 Commodity Flow Survey public use micro data. Heavy good shipments (less than 15 cents per kilogram).

It is useful to consider a summary statistic of the distribution of distance. Suppose we model the cumulative distribution function for shipment distances as follows, with

⁷Here we focus on the domestic shipment data, due to the availability of the unrestricted 2012 CFS public use micro data.

⁸Heavy goods are defined as those with a value of 15 cents per kg or less. The histograms are truncated at 1200 miles, and display 98.6% of heavy good shipments.

a lower bound on distance \underline{d} , and a shape parameter α ,

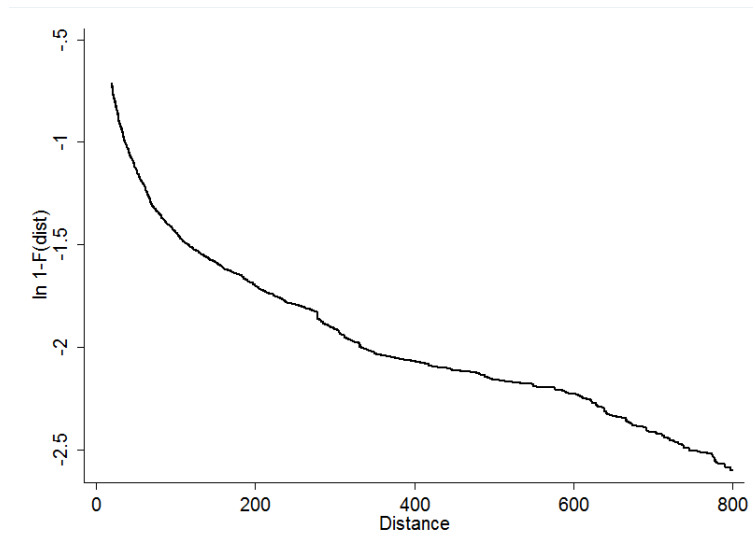
$$F(d) = Pr(\tilde{d} \leq d) = 1 - \left(\frac{\exp(\underline{d})}{\exp(d)} \right)^\alpha.$$

If the distribution were this shape, then there is a log-linear relationship between the upper tail of the distribution and distance,

$$\ln(1 - F(d)) = \alpha \underline{d} - \alpha d. \quad (1.1)$$

This is a function of distance, d , where the shape parameter α is equal to the slope. Under this assumption, the slope should be the same for both short and long distances. Figure 1.3 plots this relationship with the domestic shipments data used in Figure 1.2, and we can see that the slope isn't constant at all. At short distances, the slope is quite steep, i.e. there is a big drop-off in the distribution as mileage increases. However, at longer distances the slope is quite shallow, and the affect of distance in attenuating shipments is less.

Figure 1.3: Log of Cumulative Distribution, Heavy Good Shipments



Source: 2012 Commodity Flow Survey public use micro data, shipments between 20 and 800 miles.

The first column of Table 1.4 summarizes this pattern in the data with three slope

estimates. The estimate $\hat{\alpha}^{all}$ is from an OLS regression of (1.1) using the range of distances from 20 to 800 miles, while $\hat{\alpha}^{<300}$ uses the data for shipments below 300 miles and $\hat{\alpha}^{>300}$ uses the data for shipments farther than 300 miles. If the relationship between the distribution and distance was log-linear, then these would all be the same. However, we see that $\hat{\alpha}^{<300}$ is over four times larger than $\hat{\alpha}^{>300}$. While the distribution declines by 0.47 percent for each mile increase in distance at short distances, this becomes 0.11 percent per mile for distances greater than 300 miles. A qualitatively similar pattern emerges repeating this exercise using data on cement import shipments. Thus we conclude the same basic pattern—that at short distances, there is a big effect of distance on shipment levels, but this relationship sharply attenuates at long distances—applies for import shipments and domestic shipments alike. This is consistent with the assumption used throughout this paper that imports and domestic shipments use the same inland transportation technology.

Table 1.4: Estimates of Distribution Parameter

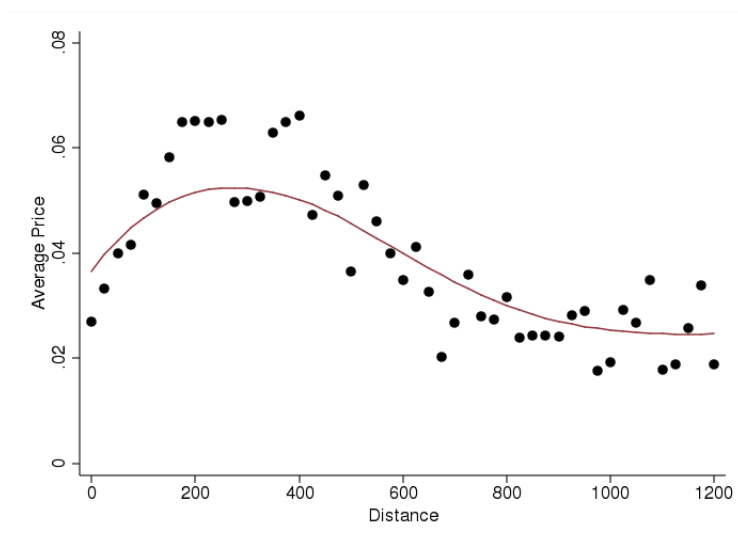
Sample:	(1)	(2)	(3)	(4)	(5)	(6)
	Commodity Flow Survey					
	Heavy goods	0 - 10 cents	10 - 15 cents	50 - 65 cents	Truck	Rail
$\hat{\alpha}^{all}$	0.27	0.27	0.38	0.30	0.69	0.13
$\hat{\alpha}^{<300}$	0.47	0.49	0.50	0.43	1.16	0.14
$\hat{\alpha}^{>300}$	0.11	0.11	0.26	0.19	0.25	0.11

Source: 2012 public use CFS. Estimates reflect percent change in cumulative distribution per mile.

As noted above, one potential explanation for the pattern, unrelated to non-linearity of transportation costs, is heterogeneity of quality. Higher quality goods (within the heavy good category), having more value relative to weight will tend to ship further than low quality goods. The goods being shipped more than 300 miles may simply all be higher quality products. Under this hypothesis, we would expect no difference in the slope estimates above and below 300 miles for goods of similar quality, but would expect a slower decline for the higher quality goods. To examine this hypothesis, we partition the heavy goods into those with unit values below 10 cents per kilogram and those with unit values between 10 and 15 cents per kilogram and repeat this exercise. Columns 2 and 3 of Table 1.4 report the results for these two categories. The same sharp pattern where $\hat{\alpha}^{<300}$ is large and $\hat{\alpha}^{>300}$ is small continues to hold. Even after controlling

for quality, at short distances, there is a big effect of distance on shipment levels, and this relationship becomes less important at longer distances. We do not see smaller estimates for the higher quality goods compared to the lower quality goods, indicating quality heterogeneity does not account for why some goods are shipping farther. We can check whether this pattern remains for higher valued goods, and as column 4 of table 1.4 shows, a similar pattern remains for goods valued at 50 to 65 cents per kilogram. Additionally, if we plot average price of heavy goods by distance shipped, as displayed in figure 1.4, we can see that it is not the higher valued goods that are traveling farther.

Figure 1.4: Average Price by Distance: Heavy Good Shipments

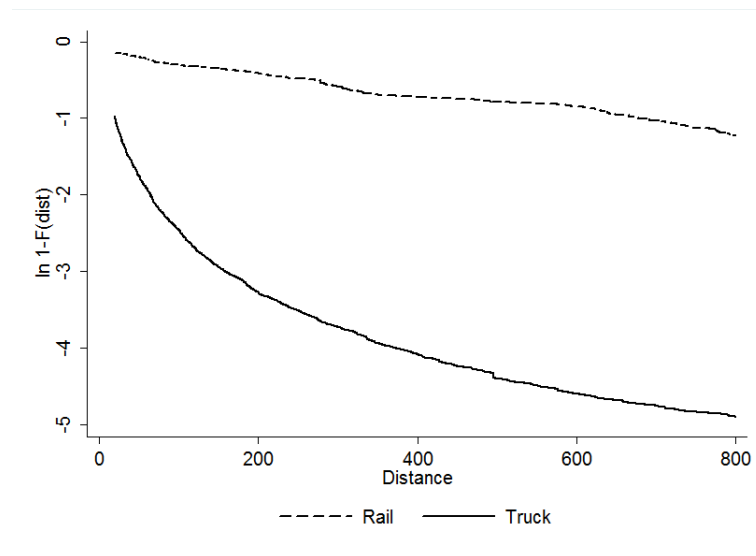


Source: 2012 Commodity Flow Survey Public use micro data. Shipment distances by 25 mile distance groups.

To explore the second explanation, that availability of multiple transportation technologies are behind the observed shipping patterns, we can repeat the above calculations separating out shipments by mode, truck versus rail. Figure 1.5 plots $\ln(1 - F(d))$ by distance for rail and truck shipments. Similar to Figure 1.3, the slope isn't constant for the truck shipments. At short distances, the slope is steep, and at longer distances the slope is shallow. Rail however, appears very different, with a constant and much flatter slope. This reflects the much fatter tail of the distribution of rail shipments. Columns 5 and 6 of Table 1.4 display the slope estimates for the truck and rail subsamples. Two

important findings emerge. First, overall, the estimates for trucks are much bigger than those for rail, indicating that trucking drops off more quickly as distances increase. Second, the estimates for rail are similar regardless of distance, suggesting that shipments decline at a steady rate as distances increase. Interestingly, trucks also have a smaller slope at longer distances, perhaps reflecting efficiencies present in longer-distance trucking.

Figure 1.5: Log of Cumulative Distribution, by Mode

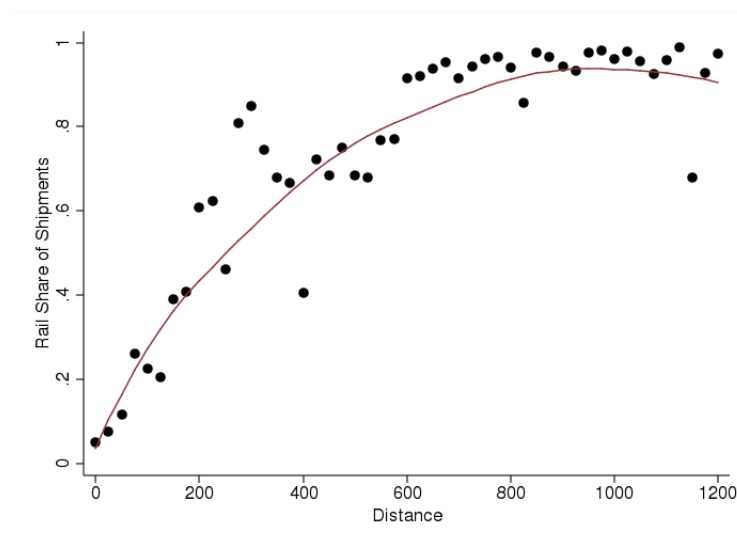


Source: 2012 Commodity Flow Survey public use micro data.

There is further supporting evidence that the flattening of the distribution beyond 300 miles is attributable to the low cost per mile of rail. Figure 1.6 shows the percent of shipments using rail by distance for heavy goods and notably rail begins to receive a larger share of the shipments at around the same distance that the overall distribution flattens. Short distance shipments almost exclusively travel by truck, with the percent of rail shipments increasing with distance. At approximately 300 miles rail accounts for about half of the shipments. Longer distance shipments are almost exclusively by rail. Over 90 percent of heavy-good shipments less than 100 miles travel by truck, while for shipments above 500 miles, over 90 percent travel by rail. Further highlighting the difference between trucking and rail, the average shipment distance for heavy goods

within the U.S. is 45 miles if the shipment goes by truck and 499 miles for heavy rail shipments.⁹ Looking only at trips longer than 100 miles, the average rail distance for these heavy goods is 666 miles. From this, we can conclude that mode choice is the key factor underlying the pattern of shipments in the data.

Figure 1.6: Rail Share: Heavy Good Shipments



Source: 2012 Commodity Flow Survey Public use micro data. Shipment distances grouped into 25 mile distance bands.

1.3 Model

There are two types of agents in this economy. The first type, indexed by $i = 1, \dots, I$ are *consumers* of heavy goods. The second type, indexed by $j = 1, \dots, J$ are *suppliers* of heavy goods. Suppliers can be classified as either domestic (i.e. located in the United States) or foreign (outside the United States). Consumers are all located in the United States. Note that since heavy goods are generally intermediated goods like building materials, the agents that I refer to as consumers are actually firms themselves, that are downstream from the upstream firms producing the heavy goods. As I will be taking

⁹Source: 2012 Commodity Flow Survey public use micro data. Distances weighted by sampling weights and shipping weight.

this downstream sector as given, it will be convenient to simply refer each to downstream firm i as consumer i .

Demand for heavy goods is modeled as a discrete choice with consumers, i , choosing which supplier to use for each transaction. A product in this environment is a shipment of the commodity from a specific supplier. We are also interested in the role of distance in product selection and the choice of domestic shipping mode, $m \in M$, which is made after the choice of supplier. It is assumed that the consumer's choice of supplier for the single input commodity is separable from other production and pricing choices, which are all taken to be exogenous. For any transaction, a consumer i faces a choice set with characteristics that are unique to that transaction.

1.3.1 The Choice of Supplier

Each consumer i has in its choice set the option to purchase from each supplier j . In this choice, consumers care about four things. The first is the price p_{ij} supplier j sets to consumer i . If supplier j is domestic, then this p_{ij} is the price at the factory gate. If supplier j is a foreign producer, then p_{ij} is the price at the United States port of entry. Assume the consumer bears the transportation cost from pickup at the factory (if the supplier is domestic) or from pickup at the port (if the supplier is foreign). This transportation cost of shipping internally within the United States is the second thing consumers care about. Let t_{ij} be expected transportation cost of choosing supplier j . The third characteristic that differentiates supplier is that they vary in quality ϕ_j . Finally, as is standard in the discrete choice literature, I allow for a random component of utility ϵ_{ij} associated with each choice j for each consumer. The resulting value for consumer i of choosing j is then

$$u_{ij} = \alpha(p_{ij} + t_{ij}) + \phi_j + \epsilon_{ij}$$

where α is the coefficient on the total cost to the consumer.

Let $y_{ij} = 1$ if j is i 's supplier choice. The probability of i choosing supplier j is:

$$s_{ij} \equiv Pr(y_{ij} = 1) = Pr(u_{ij} \geq u_{ij'} \forall j' \in J)$$

Given the logit error assumption the probability that i chooses j follows the standard form:

$$s_{ij} = \frac{\exp(\alpha(p_{ij} + t_{ij}) + \phi_j)}{\sum_{j' \in J} \exp(\alpha(p_{ij'} + t_{ij'}) + \phi_{j'})} \quad (1.2)$$

Collectively, these I probabilities for each supplier j give the demand that supplier j will face.

1.3.2 The Choice of Mode

Mode choice conditional on supplier is also modeled as a discrete choice. This is specified to allow for different fixed and variable transportation costs for each mode. The value that i receives by choosing supplier mode m conditional on supplier j is

$$u_{im|j} = \delta_0^m + \delta_1^m d_{ij} + \epsilon_{ij}^m$$

where d_{ij} is the distance between i and j . Each mode has two parameters, δ_0^m is the fixed cost per unit, while δ_1^m is the per mile variable cost of using mode m .

Assuming logit errors, the expected maximum value from this mode choice decision is $\log \sum_{m \in M} \exp(\delta_0^m + \delta_1^m d_{ij})$. Because the supplier choice is made before the mode shocks are observed, this is used by consumers as the expected transportation cost. This timing setup is used because it allows for the model to be estimated using a standard nested logit structure where mode choice is nested within the supplier choice. Additionally, this is realistic, in that it is common while shipping heavy goods for the exact transportation costs to be unknown at the time of purchase, with the final transportation costs passed through to the buyer at the time of shipment.

1.3.3 Suppliers

Suppliers are profit maximizing Bertrand competitors. This paper considers two models of supplier pricing. The first constrains each supplier to charging a single uniform price to all customers.¹⁰ The second pricing model considered, like the models used by Miller and Osborne (2014) and Beckert, Smith, and Takahashi (2015), allows suppliers to price discriminate and charge each customer a different price. In both cases, suppliers are

¹⁰Recall that these prices do not include inland transportation costs.

assumed to have constant marginal costs at any point in time. This assumption will allow marginal costs to be inferred from estimated demand elasticities for small changes in output.

Uniform pricing

In a world where transportation costs are small or non-existent, suppliers would be unable to price discriminate based on geography because consumers could simply pay the lower price at a different location. If this arbitrage were possible it would result in each supplier setting a uniform price. Although this is an unlikely scenario for heavy goods, I consider the uniform pricing assumption in order to compare to the price discrimination case. Following this approach, if it is assumed that each supplier can set just one price to charge all customers, then supplier j 's maximization problem can be used to solve for optimal prices. The maximization problem is

$$\max_{p_j} (p_j - c_j) \sum_I s_{ij}(p_j).$$

Taking first order conditions and solving for the optimal price gives,

$$p_j^* = c_j + \frac{\sum_I s_{ij}(p_j)}{-\alpha \sum_I s_{ij}(p_j)(1 - s_{ij}(p_j))}.$$

Under uniform pricing, I assume that foreign suppliers, which can ship goods through multiple U.S. ports, must choose a single price regardless of port.

Price discrimination

Given the importance of geography and transportation costs for heavy goods, a second supply specification is considered where suppliers are able to flexibly set prices based on the locations of consumers. With the ability to set different prices for each consumer, the maximization problem for supplier j becomes

$$\max_{p_{ij}} \sum_I (p_{ij} - c_j) s_{ij}(p_{ij}).$$

Taking first order conditions and solving for the optimal prices gives

$$p_{ij}^* = c_j + \frac{1}{-\alpha(1 - s_{ij}(p_{ij}))}.$$

Equilibrium prices will be different when price discrimination is allowed because suppliers are now able to respond to the probabilities that they are chosen by each consumer and adjust each single price accordingly, instead of setting a uniform price to maximize national profits.

1.4 Estimation and Results

The estimation proceeds in two steps. First, the mode choice parameters are estimated. Second, the demand parameters and costs are estimated. Counterfactual simulations are then performed using the results from the estimation.

1.4.1 Step 1: Estimating Mode Choice Parameters

In this subsection, as a first step, I estimate the parameters governing mode choice, taking as given the supplier choice of each consumer. This approach is feasible given the earlier assumption that the logit costs shocks affecting mode choice are realized after the supplier choice is made.

I assume that the mode choice parameters determining internal shipments by domestic firms are the same as the mode choice parameters determining shipments of imports from their ports of entry into the United States to their ultimate destinations. This is reasonable, because imports and domestic shipments use the same internal transportation network within the United States. Recall the notation that δ_0^m and δ_1^m are the fixed and variable costs for using mode m . In this first stage estimation of mode choice, it is only possible to identify the differences in fixed costs between the modes, since if the fixed cost of both modes were increased by the same amount, the mode choice decision would be unchanged. Analogously, since we are taking supplier choice as given, and thus the distance shipped to be fixed, it is only possible to identify the difference in the marginal distance cost between modes. Let $\Delta_0 \equiv \delta_0^{rail} - \delta_0^{truck}$ and $\Delta_1 \equiv \delta_1^{rail} - \delta_1^{truck}$ denote these differences. I normalize the logit error draws to the

standard type 1 extreme value distribution and leave estimation of a scaling coefficient to convert the mode coefficients to dollars to the second estimation step below.

The parameters Δ_0 and Δ_1 are estimated using the sample of domestic shipments in the CFS, since mode choice for foreign shipments are not observed. Table 1.5 reports the results and as expected, rail has higher fixed costs ($\Delta_0 > 0$), and lower variable costs ($\Delta_1 < 0$). I estimate the parameters over 3 alternative samples. The first sample uses cement shipments from the restricted 2007 CFS, while the second uses the approximate sample of cement shipments from the 2012 public CFS. The third sample uses all shipments of heavy goods in the 2012 CFS. The parameters are qualitatively similar in each case and estimated with a high degree of precision, given the large samples.

Table 1.5: Mode Choice Parameter Estimates

	Cement 2007	Cement 2012	Heavy goods 2012
$\Delta_0 \equiv \delta_0^{rail} - \delta_0^{truck}$	3.75 (0.041)	3.60 (0.0248)	2.34 (0.0062)
$\Delta_1 \equiv \delta_1^{rail} - \delta_1^{truck}$	-0.0089 (0.00023)	-0.0084 (0.0002)	-0.0067 (0.00003)
N		41,796	268,300
$\hat{d} = \frac{\Delta_0}{-\Delta_1}$	421	428	349
$Pr(truck d = 0.75\hat{d})$	72%	71%	64%
$Pr(truck d = 600)$	17%	19%	16%
$Pr(truck d = 200)$	88%	87%	73%

Note: Standard errors in parenthesis. 2007 estimates use restricted data, 2012 estimates from public use data; 2012 Cement sample includes heavy shipments with SCTG code 31 and NAICS code 327.

It is useful to calculate the cutoff distance of indifference between the two modes, given the same logit errors, $\epsilon_{ij}^{rail} = \epsilon_{ij}^{truck}$. At distance, $\hat{d} = \frac{\Delta_0}{-\Delta_1}$, the consumer would be indifferent between the two modes. The break-even distances are reported in table 1.5. Using the cement data from the 2007 CFS, the break-even distance is 422 miles, which is very similar to that found when using the shipment data for cement from the public 2012 CFS. The third sample, that looks at all heavy shipments instead of just cement, has a break-even distance that is lower at 349 miles.

The last issue to be discussed is how much work the logit errors are doing in accounting for mode choice. At the cutoff \hat{d} , half the time truck is chosen and half the

time rail is chosen. The logit errors would be doing a lot of work if for distances well above \hat{d} trucks were often chosen, and if for distances much below \hat{d} rail was often chosen. The logit errors do play a role in the mode choice and this reflects the substantial variability in choice observed in the data. If we look at distances 25 percent below \hat{d} , the estimated probability of choosing truck ranges from 64 to 72 percent, as shown in table 1.5. For 600 mile shipments, trucks are used between 16 and 19 percent of the time and for 200 mile shipments trucks have a 73-88 percent probability. However, for very far or short distances the logit errors are having less of an effect on choice. For shipments of 1000 miles, the estimated probability of choosing a truck drops to 1 percent, while for shipments of 20 miles the probability of using a truck is 97 percent in each of the cement samples and 90 percent in the heavy goods sample.

1.4.2 Step 2: Demand and Cost Estimation

With the mode choice parameter estimates in hand, the next step is to turn to the demand and cost estimation. As mentioned in section 1.3, consumers care about four things in their supplier choice: price, transportation costs, quality, and the random component of utility. Prices are observed for all transactions that occurred, while offered – but not accepted – prices are unobserved. The demand estimation proceeds using two competing assumptions on pricing. The first assumes uniform prices must be offered and the estimation uses the average observed price for a supplier as that suppliers' offered price. The second allows for price discrimination and for that variation of the estimation, I must solve for the optimal price that each supplier offers each consumer. While each estimation strategy is discussed separately below, I first turn to elements of the estimation that are common to both the uniform pricing and price discrimination cases.

Transportation costs are partially inferred from the estimated mode choice parameters, but a scaling coefficient is needed to convert these coefficients to dollars. Although the mode choice estimation only identified the difference in the cost per mile between modes the additional cost per mile that must be paid regardless of mode is relevant for the supplier choice. Note that the fixed cost that is mode invariant does not affect the choice of supplier because it does not vary by supplier. Expected transportation cost

t_{ij} for i to use supplier j is

$$t_{ij} = \lambda \log (\exp(-\Delta_0) + \exp(\Delta_1 d_{ij})) + \delta_2 d_{ij}$$

where the parameter λ scales the expected maximum value from the mode choice estimation into a dollar cost and δ_2 is the per mile cost that does not depend on mode. Both λ and δ_2 are estimated in this second step.

One advantage of the data is that I can see all potential domestic suppliers and all foreign suppliers who have shipped to the United States. As a result, even suppliers that have very few sales are in the data. This is an important feature, because it allows for the construction of comprehensive choice sets. Consumers care about supplier quality ϕ_j . However, for the suppliers that have a small number of observed sales, individual fixed effects are not estimated. Instead an import dummy variable, M_j , with coefficient β , is added to the estimation to capture the average quality of imported goods. It may be the case that quality is unknown to some consumers and treating smaller suppliers as having average quality would then be consistent with the information available to consumers.¹¹

The next issue that must be addressed is the construction of the choice set. Since not all domestic suppliers have observed shipments in the CFS, suppliers with unobserved shipments are added to the choice set based on firm information from the SSEL. If suppliers with unobserved shipments were not added back to the choice set, all of the estimated choice probabilities would be biased upwards, since the unobserved suppliers would be forced to have a zero probability of being chosen. Because the SSEL includes all plants and their exact locations, all the variables required to add these plants to the choice set can be constructed. Note that because of the importance of location, we cannot simply weight the plants with observed shipments by their sampling weight, instead all plants in the choice set must be at their true locations.

Finally, because imports are a census and domestic shipments are a survey, the sampling weights in the CFS must be taken into account for the maximum likelihood estimation. Denote Ω_{ij} as the sampling weight for the shipment from j to i . To account for unobserved shipments, y_{ij} is weighted in the likelihood function estimation using

¹¹ Additionally, the import dummy variable, by providing the average quality of imports relative to domestic goods is a potentially interesting parameter.

Ω_{ij} .¹²

Uniform pricing

In the uniform pricing case, demand parameters $\Theta = \{\alpha, \delta_2, \lambda, \beta, \phi_j\}$ are estimated to maximize the likelihood of observing the choices in the data. Because this case assumes each supplier must offer the same price, p_j , to all consumers, the parameters are estimated using the average observed price for each supplier as p_{ij} . That is, for all i

$$p_{ij} = \bar{p}_j = \frac{\sum_I p_{ij}^{data}}{\sum_I y_{ij}}.$$

For foreign suppliers the price information comes from the trade data and are the unit values reported on the customs forms, inclusive of customs duties and freight charges associated with getting to the port. Although foreign shipments can enter through multiple ports, in this case it is assumed that the uniform price is the same across all ports. For domestic suppliers with shipments in the CFS, prices are the unit values from that source. Average prices for domestic suppliers who do not have shipments in the CFS are taken to be the average of the prices for the observed domestic suppliers.

The likelihood function is

$$L(\Theta) = \prod_{i \in I} \prod_{j \in J} s_{ij}^{y_{ij} \Omega_{ij}},$$

where the probability s_{ij} is

$$s_{ij} = \frac{\exp(\alpha(\bar{p}_j + \lambda \log(\exp(-\Delta_0) + \exp(\Delta_1 d_{ij})) + \delta_2 d_{ij}) + \beta M_j + \phi_j)}{\sum_{j' \in J} \exp(\alpha(\bar{p}_{j'} + \lambda \log(\exp(-\Delta_0) + \exp(\Delta_1 d_{ij'})) + \delta_2 d_{ij'}) + \beta M_{j'} + \phi_{j'})}.$$

Once the demand parameters are estimated, optimal pricing behavior allows us to infer the marginal cost for each supplier,

$$c_j = \bar{p}_j - \frac{\sum_I s_{ij}(\bar{p}_j)}{-\alpha \sum_I s_{ij}(\bar{p}_j)(1 - s_{ij}(\bar{p}_j))}.$$

¹²Due to the CFS design, we know that these observed shipments are a random sample. See U.S. Department of Transportation and U.S. Department of Commerce, “2007 Economic Census, Transportation, 2007 Commodity Flow Survey” <http://www.census.gov/econ/cfs/2007/US%20FINAL.pdf>, for details of the sample construction.

The estimated costs can then be adjusted in counterfactual simulations.

Table 1.6 displays the results of the supplier choice estimation under the uniform pricing case. The estimate of α shows a large negative effect of price and this gives an average consumer level price elasticity of -7.75. Imports are less likely to be chosen than domestic production of the same price and transportation cost as indicated by the large negative coefficient (β) on the import variable. The estimate for δ_2 provides a dollar cost estimate of the per mile cost of rail per kilogram, while λ is used to scale the estimates from the mode choice problem into dollar costs. This results in a per mile cost of rail that is significantly cheaper than trucking, as expected. Although the mode fixed costs should be interpreted as including both direct and indirect monetary and non-monetary costs (i.e. the opportunity costs of delay associated with rail, the cost of loading and unloading rail cars, and transportation to/from the rail-head), the implied fixed cost of rail of over \$1000 per metric ton is unrealistically high. The parameters also imply very high markups, often larger than prices, which would unrealistically indicate negative marginal costs. The unrealistic fixed costs and markups indicate that the uniform pricing assumption might not be appropriate in this industry.

Table 1.6: Supplier Choice Parameter Estimates

	Uniform Pricing
α	-6.20 (0.0243)
λ	-0.33 (0.0014)
δ_2	0.00017 (0.000007)
β	-4.11 (0.0232)
ϕ_j	Not reported
Avg consumer price elasticity	-7.75
Fixed cost: rail (\$/metric ton)	1216
Per mile cost: truck (\$/metric ton)	3.06
Per mile cost: rail (\$/metric ton)	0.17

Note: Standard errors in parenthesis.

Price discrimination

In the price discrimination case, suppliers are able to offer different prices to each customer i . As a result, to estimate the demand parameters, prices for each supplier-buyer pair are needed. However, only transacted prices are observed while offered but not accepted prices are unknown. To estimate the demand parameters in this environment, I assume that a given supplier j has a single cost, c_j , to supply every customer i . Under this assumption, optimal pricing behavior by the suppliers implies that the constant marginal cost is

$$c_j = p_{ij} - \frac{1}{-\alpha(1 - s_{ij}(p_{ij}))} \quad \forall i, j.$$

Costs and demand parameters, $\Theta = \{\alpha, \delta_2, \lambda, \beta, \phi_j\}$, are jointly estimated to maximize the likelihood of observing the choices and the prices that are in the data. Given costs and parameters, the optimal price is

$$p_{ij}^*(\mathbf{p}_i^*; \Theta, c_j) = c_j + \frac{1}{-\alpha(1 - s_{ij}(\mathbf{p}_i^*))}$$

where the optimal price offered by j to i is a function of all prices faced by the consumer, \mathbf{p}_i^* . In order to match estimated prices to the data, I assume that there is an optimization error, ν_{ij} and that the prices offered to consumers are $\tilde{p}_{ij} = p_{ij}^* + \nu_{ij}$. There are a number of reasons why this error term might exist. First, prices may be agreed upon in advance, before actual marginal costs or other supplier's prices are realized. Second, some customers may be more or less expensive to supply because of extra services provided or existing relationships that may tend to lower transaction costs and these savings or costs may be passed on to the price.

The error term, ν_{ij} , is assumed to be normal with mean 0 and variance σ_ν . For prices that are observed, ν_{ij} is equal to the difference between the observed price and the optimal price. For offered prices that did not result in transactions, ν_{ij} is simulated. Denoting $\tilde{\mathbf{p}}$ as the matrix of all offered prices, the probability of observing supplier j and price p_{ij} in the data is

$$Pr(y_{ij} = 1) = Pr(y_{ij} = 1 | \tilde{\mathbf{p}}) Pr(p_{ij} | \sigma_\nu).$$

The first term is the standard logit probability (where prices are taken as given) from equation 1.2, while the second term follows the normal probability density function for ν_{ij} ,

$$Pr(p_{ij}|\sigma_\nu) = Pr(\nu_{ij}|\sigma_\nu) = \frac{1}{\sigma_\nu\sqrt{2\pi}} \exp\left(\frac{-\nu_{ij}^2}{2\sigma_\nu^2}\right).$$

The combined likelihood function for observing both the choices and the prices is

$$L(\Theta, \sigma_\nu, \{c_j\}) = \prod_{i \in I} \prod_{j \in J} Pr(y_{ij} = 1)^{y_{ij}\Omega_{ij}} = \left(Pr(y_{ij} = 1|\tilde{\mathbf{p}}) Pr(p_{ij}|\sigma_\nu) \right)^{y_{ij}\Omega_{ij}}.$$

If all prices were observed, then there would be no need to estimate the probability of observing the prices and only the choice probability would remain. In order to illustrate the properties of the likelihood function, let's assume for simplicity that parameters are such that the error for the observed prices is zero. Then $Pr(p_{ij}|\sigma_\nu)$ will be large, but the logit probability, $Pr(y_{ij} = 1|\tilde{\mathbf{p}})$, may be small if the unobserved prices of products not chosen are similar to the prices of the chosen products. On the other extreme, if the errors for the observed prices are all towards the far left tail of the distribution, the probability of observing the prices will be smaller. However, it would be expected that the logit probability would be much larger than before because now the observed prices will be much smaller than the prices for the products not selected. This trade-off in estimating the parameters is embodied in the likelihood function.

Table 1.7 displays the results of the supplier choice estimation under the price discrimination case. The estimate of α shows a larger negative effect of price than in the uniform case and this gives an average consumer level price elasticity of -9.33. Imports are less likely to be chosen than domestic production of the same price and transportation cost as indicated by the large negative coefficient (β) on the import variable, which is similar in magnitude to the uniform price estimation. The estimate for δ_2 indicates the per mile cost of rail per metric ton is 6 cents. Using λ to scale the estimates from the mode choice problem into dollar costs, gives lower cost estimates than in the uniform price estimation. Rail is still significantly cheaper than trucking, but the estimates are now much more reasonable at \$1.20 per metric ton per mile for trucks and a fixed cost of rail of \$503 per metric ton. The variance of the pricing error, σ_ν is 0.027 indicating a moderate amount of noise in the observed pricing.

Table 1.7: Supplier Choice Parameter Values, Price Discrimination

	Price discrimination
α	-15.64 (0.0591)
λ	-0.13 (0.00057)
δ_2	0.00006 (0.0000025)
β	-4.27 (0.0233)
ϕ_j	Not reported
σ_ν	0.027
Avg c_j , imports	0.032
Avg c_j , domestic	0.039
Avg consumer price elasticity	-9.33
Fixed cost: rail (\$/metric ton)	503
Per mile cost: truck (\$/metric ton)	1.26
Per mile cost: rail (\$/metric ton)	0.06

Note: Standard errors in parenthesis.

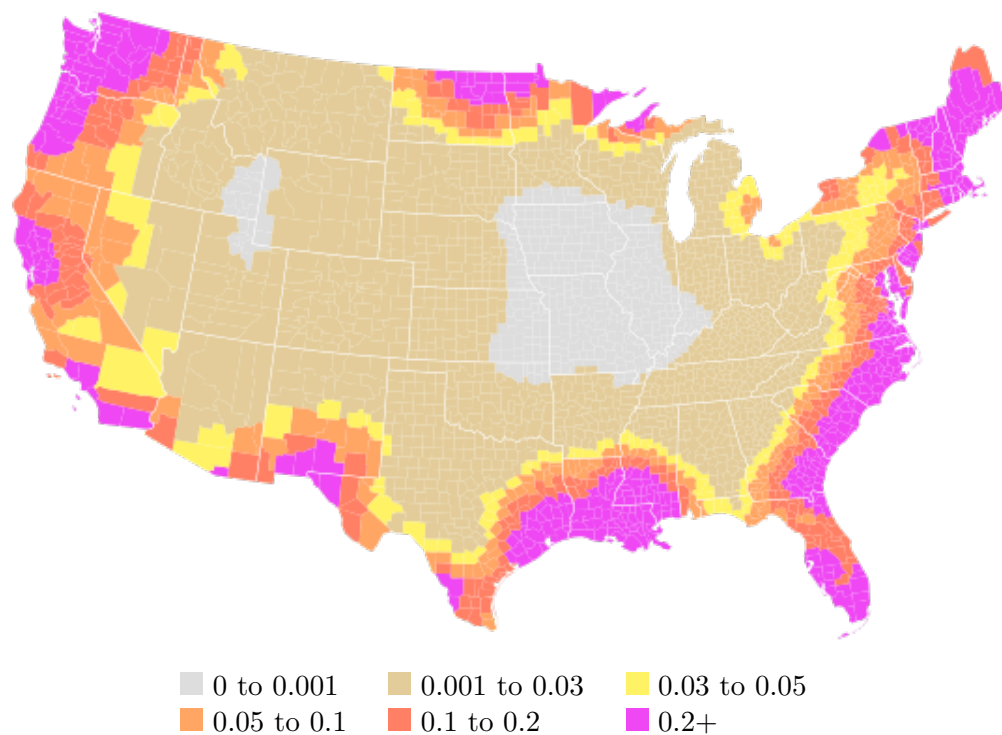
Turning now to the marginal costs estimates, we find costs that are more plausible than when uniform pricing was assumed. While the individual costs are not reported, the average cost for imports is 3.2 cents per kilogram (\$32 per metric ton) while the average marginal cost for domestic producers is estimated to be \$39 per metric ton. Using the parameter values in table 1.7, which produce more reasonable estimates of transportation costs, section 1.4.3 illustrates how changes in policy effect import flows, when firms are able to set prices by location.

1.4.3 Counterfactual simulations

With model parameters and costs estimated, changes to trade costs can be simulated. Using public data on the locations of firms and the ports used by importers, I solve for the optimal prices under price discrimination and construct fitted values for the consumer choices in each U.S. county. Figure 1.7 maps the share of imports by county,

and while we can see that the highest import concentrations are near the ports, imports reach very far into the interior and are found in most of the country.

Figure 1.7: Fitted Value Import Shares by County



Turning now to changes in trade policy, it is instructive to consider a recent policy change towards imports of cement from Mexico, before examining broader hypothetical policy changes. In 1990, the United States began charging anti-dumping tariffs on cement imports from Mexico. These punitive tariffs were removed in 2006, reducing the tariff on Mexican cement by \$23 per metric ton. Figure 1.8 displays the share of imports from Mexico in each county under current policy and figure 1.9 shows the share of imports from Mexico if a tariff of \$23 per metric ton were reinstated. From these figures we can see that areas near the boarder saw Mexican imports decrease and that the imports would not go as far into the interior with the tariff.

A similar exercise can be performed to examine what would happen if a tariff of similar size to the Mexican case were imposed on all imports. Figure 1.10 shows estimated

Figure 1.8: Fitted Value Import Shares from Mexico

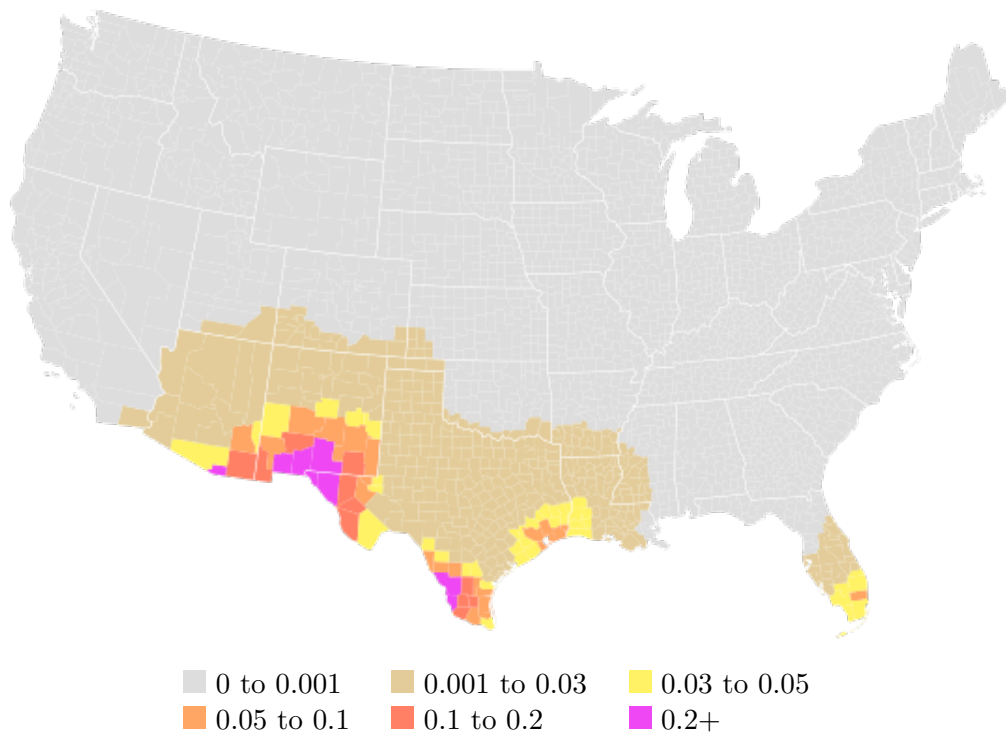
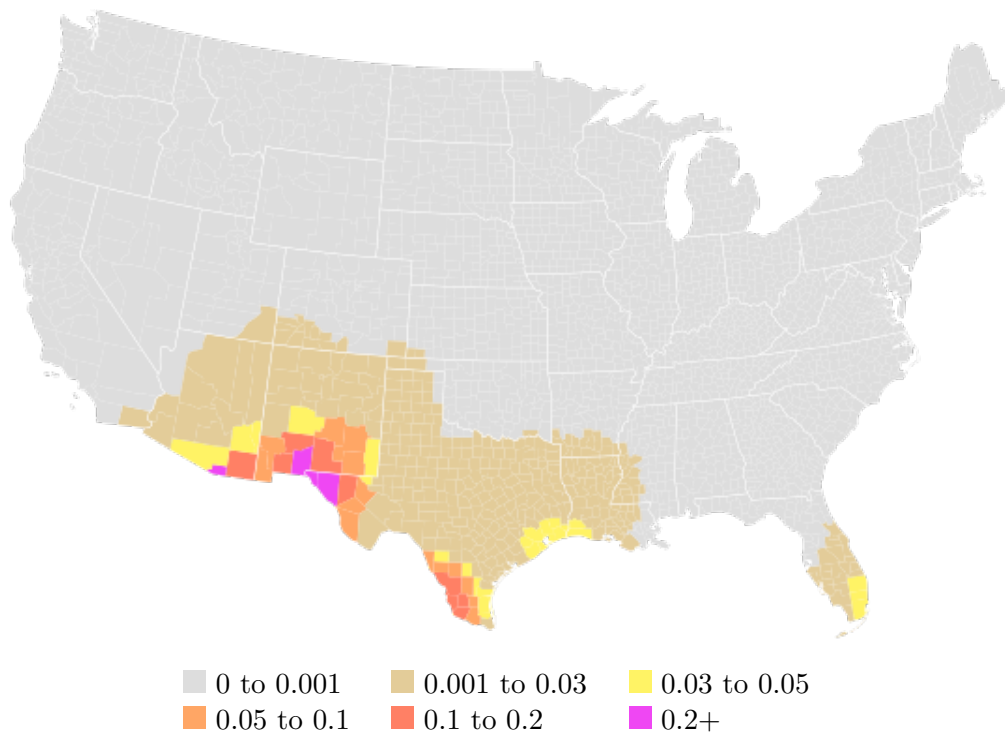
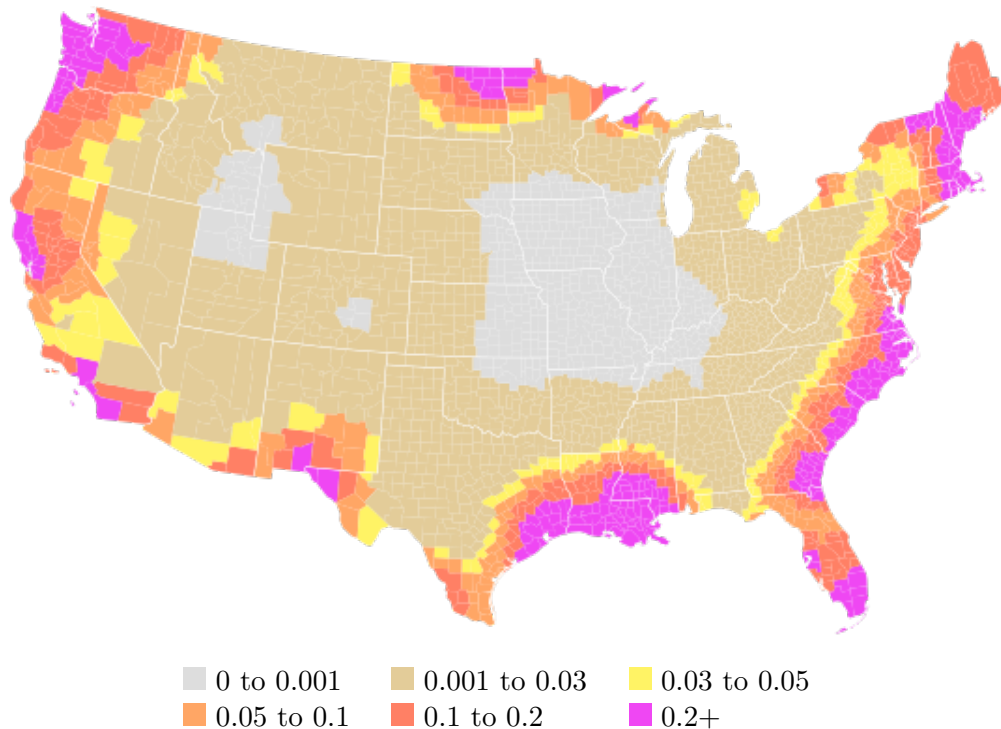


Figure 1.9: Modeled Import Shares from Mexico, with Anti-Dumping Duty



import shares in this scenario. The effect of the tariff differs by location. While areas near the coast still receive large amounts of imports, there is a general pulling back of shipment distances and a large area of the mountain west loses all access to imports.

Figure 1.10: Modeled Import Shares, High Tariff



While figure 1.10 demonstrated that import shipments were responsive to changes in trade policy, we can also look at what happens to prices. To do this, let's consider a more extreme policy that bans all imports. Figure 1.11 shows the average total price (price plus transportation costs) in each county in the baseline mode and figure 1.12 shows the prices without any import competition. Here we can see that prices go up in some coastal areas but impacts are largely absent in the interior. The most noticeable increase occurs in the New Orleans area, where domestic production is absent, as noted in figure 1.1. Because domestic producers have access to the same transportation network as importers, the reduction to foreign competition is mitigated by increases in domestic trade across regions.

Figure 1.11: Fitted Value Total Price

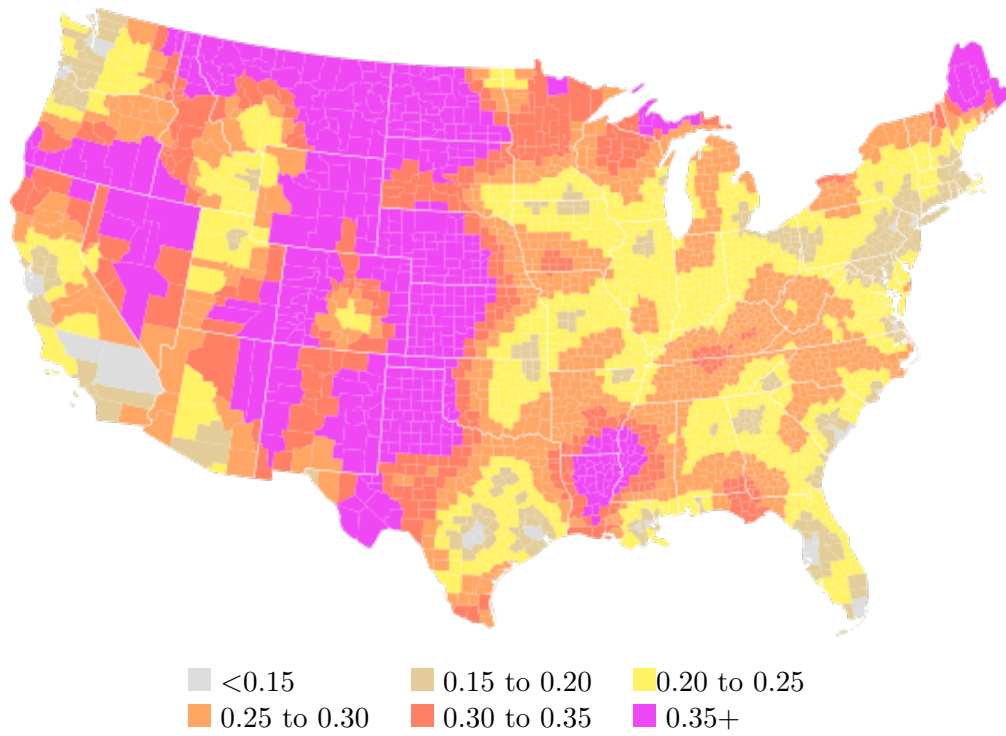
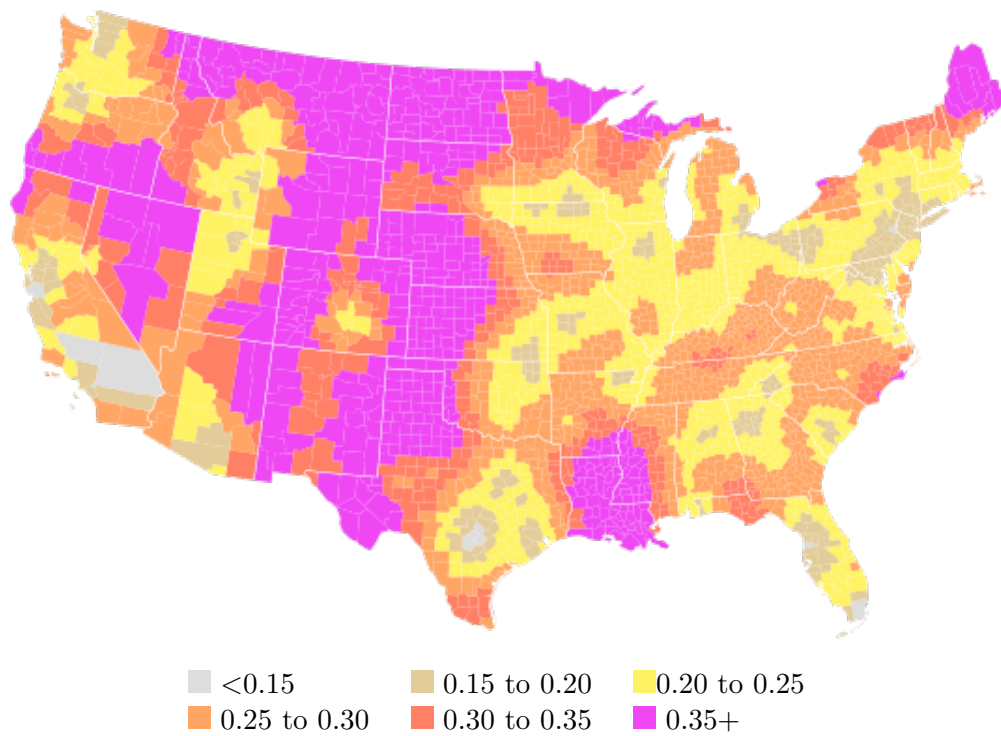
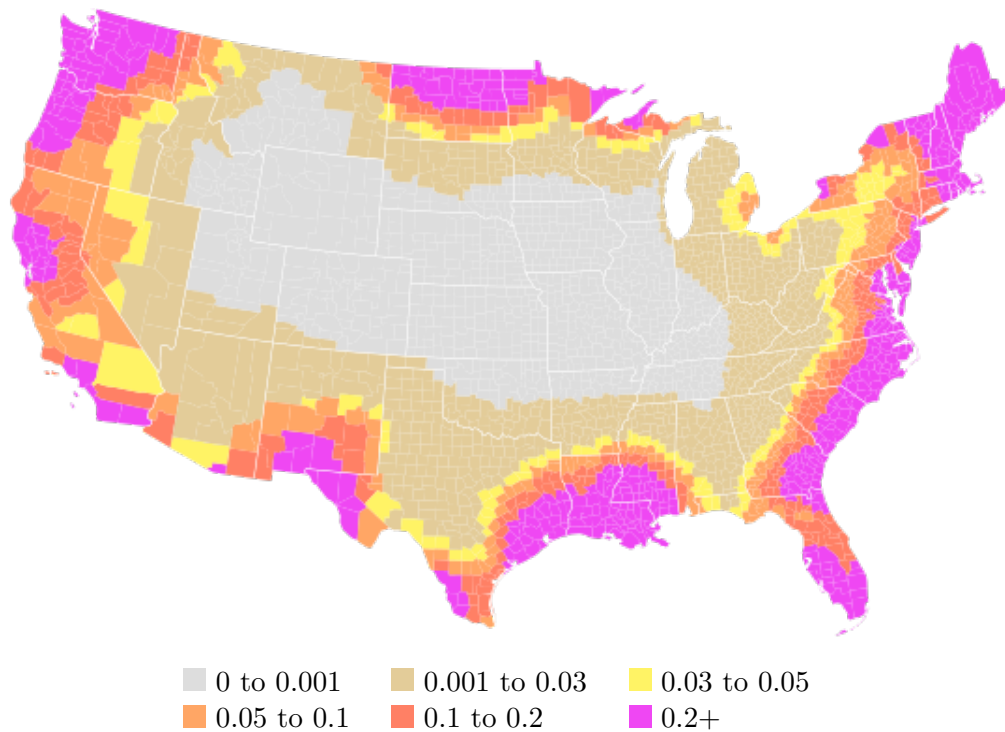


Figure 1.12: Total Price if Imports Were Prohibited



The rail transportation network plays an important role enabling imports (and domestic production) to travel long distances. As fixed costs of rail are met, imports make their way to more locations away from the ports. Figure 1.13 models import shares if all shipping were required to be done by truck. Compared to figure 1.7, there is little effect at the coasts, but a large interior area that once had a small import share now does not receive any imports. Additionally import shares *increase* near the coast, since without rail, domestic producers have more difficulty moving product to coastal regions to compete with imports. While the removal of the rail network is not a policy under consideration, this demonstrates that it is important to consider the availability of this lower marginal cost transportation mode when modeling trade flows.

Figure 1.13: Modeled Import Shares Without Rail



For a final counterfactual exercise, consider what would happen if firms were unable

to price discriminate based on location.¹³ The estimation already demonstrated unreasonably high markups (and negative costs), if uniform pricing was assumed. For this exercise we will consider what would happen if uniform pricing were mandated, given the marginal costs estimated in the price discrimination case. While looking at a map of pricing under this scenario is not too informative, an examination of the underlying data reveals that some consumers would pay more and others less under a uniform price (plus transportation cost) mandate. The consumers that are geographically close to a single manufacturer and far from competition would see total price declines on average, while consumers would be offered higher prices from more distant plants. This indicates that allowing geographic price discrimination can actually increase price competition by enabling firms to set prices to compete in markets where they would not be competitive if they were required to charge a single price.

1.5 Conclusion

Because a large share of shipments of heavy goods stay near the ports, the role of competition for these products away from the coast is often ignored. However, we do observe these shipments traveling beyond the coast, and when they do, they tend to be shipped far. By incorporating the geographic structure of firms and allowing for non-linear transportation costs and price discrimination, I am able to estimate a model that explains this observed pattern of trade.

Using this model we can see that imports and distant domestic production compete in local markets over long distances. Although imports are traveling far, the effect on prices of changes in import policy tend to be small in the interior. This is because, like imports, domestic production also can travel far, potentially mitigating some of the effects of reduced import competition. Despite the fact that most heavy imports stay near the coast, just having the option of going far provides competition in the interior. This competition is enabled by the rail transportation network. The marginal cost of an additional mile of rail transport is small. Once the fixed costs of rail transport are paid, it is relatively easy for shipments to travel far. Without rail technology, imports

¹³Recall that the only source of price discrimination in this model is geographic location and that positive transportation costs enable this form of price discrimination.

would not be as competitive in the interior and this can be seen by simulating the model with linear in distance transportation costs. Models that assume linear transportation costs, or ignore distant competition all together, risk underestimating the amount of competition that firms face.

Chapter 2

Using Four Corners to Import Goods: Estimates for Wal-Mart^{*}

2.1 Introduction

A container bearing imports from Asia, after being unloaded at a U.S. port, generally makes a first stop at a firm's import distribution center (DC), where the container is unpacked, and the goods are routed through the firm's distribution network downstream. A firm with limited scale will generally have only a single import DC. The leading retailers in the U.S., such as Wal-Mart, Target, K-Mart, and Home Depot, have enormous scale, and in recent years have adopted what is called a "four corners" strategy with at least four import DCs, spread out across the West and East coasts. This strategy lowers a firm's transportation costs compared with using a single import DC, since using both coasts makes greater use of cheap ocean transport. In addition, the strategy potentially makes a firm less vulnerable to a disruption, such as a labor union strike, at a particular port or distribution center. If a firm is dependent upon a single port and a single import DC, a disruption at either point can shut down the firm's entire operation.

In this paper, we examine a firm's use of a four corners import strategy. Our first objective is to develop and estimate a model of the import process. Second, we

^{*}This chapter is joint work with Thomas J. Holmes.

use the model to examine what the firm gains by adopting a four corners approach, both in terms of savings in transportation, and in terms of reducing its vulnerability to disruption. Third, we examine how adoption of a four corners approach affects which locations are chosen for port and distribution center activities; this is of interest because port and distribution center activity at a location potentially connect with a location's growth.

Our empirical application is the import distribution system of Wal-Mart, which by far is the largest importer of containers into the United States. We have collected a unique micro data set with a list of 1.7 million 40-foot containers imported by Wal-Mart over the period 2007 through 2015. Our sample consists of a little more than half of all Wal-Mart's ocean container imports over the period. The data includes details such as where each container originated, the ship used, the foreign and domestic ports, and the import DC receiving the container. We have information about the contents of the container, specifically for most records we observe the product item number that Wal-Mart uses to track products, as well as product classification information. Thus we are able to determine the quantity allocations across DCs at the product item level. We have merged the shipment records with GPS information on port calls by vessels, through which we calculate transit times for ocean voyages. To estimate our model, we also make use of container-level data that we have constructed on ocean freight expenditures.

In the model, the firm first chooses how many import DCs to use and where to put them, second how to allocate quantities of imports across DCs, and third which ports to use to get the desired allocation shares to each DC. We highlight two key aspects of the allocation share choice problem. The first concerns the degree to which the shares are optimized at the source country level. Retailers like Wal-Mart import goods from a variety of different countries that can vary in their relative proximity to the East and West coasts. In particular, China is relatively close to the West Coast and gets to the East coast via the Panama Canal. Indian Ocean countries are relatively close to the East Coast, and get there via the Suez Canal. We ask whether Wal-Mart customizes its DC allocation shares for the goods from individual source countries, to optimize to source country relative transportation cost. Or do the scale economies of treating the goods from each source country the same at a DC center outweigh the gains from

customization? We find the latter. DC allocation shares tend to be invariant across source country. However, the ports used to get to a particular DC can vary across source countries, and through this margin alone, we show the firm is able to get quite close to the what the optimum would be with full customization. Take, for example, the case of India. If Wal-Mart could customize DC allocation shares specifically for India, it would lower average transportation costs from India by about one percent. For China imports, it would be just a tiny fraction of that, because with China's huge share of import volume, Wal-Mart essentially optimizes its network to China.

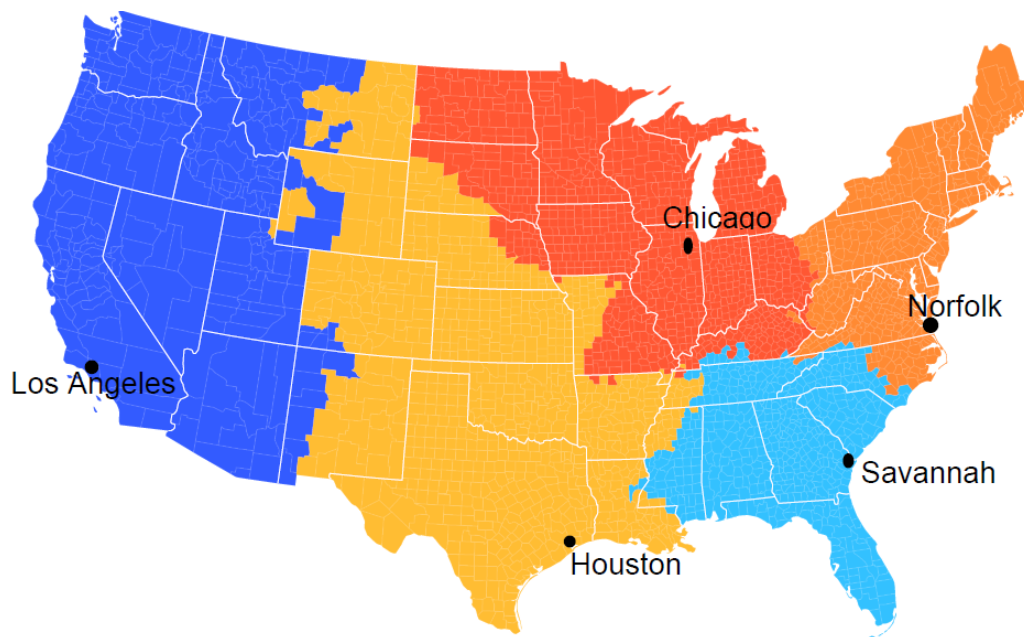
The second share allocation issue that we address concerns what happens when there is a supply disruption. In the very short run, can the firm adjust DC allocation shares? We study this issue by examining Wal-Mart's behavior during the West Coast port slowdown that occurred at the end of 2014 and beginning of 2015. We find there was no short-run response in DC allocation shares. However, there is a short-run response in port choice. That is, we get an analogous result to our finding about sensitivity to country of origin: DC allocation shares are not flexible, but port choice is.

There are good reasons to believe minimizing vulnerability to disruption is a key consideration in the design of a supply chain of a large retailer. Import channels are a choke-point, and if workers can shut down goods flow at a choke-point, everything stops all the way down the supply chain, giving the workers tremendous leverage. While as a general rule union power has declined significantly in the United States, dockworkers have maintained significant power, sitting at a choke-point on top of a distribution pyramid of billions of dollars of goods flow. In 2002, a labor disruption completely shut down West Coast ports for 11 days. Goods began to move when President Bush intervened, but goods flow was backlogged for months. Import DCs are also a choke-point, and Wal-Mart takes steps to keep unions out, including having them operated by third party firms, who in turn hire temporary workers.¹ (Even so there was a strike in one import DC and a slowdown in another, two of the few instances of work stoppages at any Wal-Mart facility.) We use the estimated model to evaluate the degree to which adoption of a four corners strategy reduces a firm's vulnerability to disruption. In

¹In contrast, Wal-Mart's regular distribution centers are all directly run by Wal-Mart with their own employees. Wal-Mart has significant redundancy in its regular distribution system (e.g. it has 42 general distribution centers) and if one is shut down because of labor unrest it can substitute other neighboring DCs.

particular, after the 2002 West Coast shutdown, Wal-Mart expanded its import DC network by adding two central DCs in Houston and Chicago, to the one previous on the West Coast (Los Angeles) and two previous on the East Coast (Savannah, Georgia, and Norfolk, Virginia). Figure 2.1 illustrates the location of the five DCs, as well as estimated market areas for each DC. We find that despite the addition of the two new DCs, Wal-Mart remains vulnerable to disruption. The problem is first, in the short run, Wal-Mart cannot quickly adjust its allocation shares across DCs. Second, it needs to get goods to the West Coast, and when the West-Coast ports go on strike, there are little in the way of practical alternatives.

Figure 2.1: Wal-Mart's Import DC System



Note: The market areas are estimated, based on model estimates in the paper.

We find that the benefit to Wal-Mart of expanding its distribution to five DCs (which we think of as adopting the four corner strategy) relative to its 3-DC system is a substantial reduction in per container cost, on the order of \$200 per container (with the exact number depending upon the origination country). This is approximately 5 to 10 percent of cost. Moreover, the reduction in cost in going from a 1-DC system to a 3-DC system is the same order of magnitude. Our calculations do not include the

fixed cost of setting up the additional DCs. Nevertheless, conditional on adoption, the marginal cost is the crucial cost that feeds into the volume of activity. Our calculations suggest that adoption of four corner import strategies by large retailers has played an important role in reducing marginal transportation costs.

2.1.1 Some Previous Literature

In addition to freight cost, our analysis takes into account the value of time and an important predecessor paper on time is Hummels and Schaur (2013). This previous paper uses public census tabulations on how imports enter by mode (in particular, air versus water), and by port of entry (in particular, West Coast versus East Coast). The paper controls for particular products, and examines the tendency to substitute away from water towards air, for deliveries of the particular product to the coast opposite to the originating country (i.e. the East Coast from Asia, or the West Coast from Europe). We highlight three ways our paper is different. First, while the previous paper considers the margin of air versus water, in our paper all the imports are waterborne, and the margin is which ports and DCs to use. We think this is a good margin to focus on, because it is relevant for a substantial fraction of import activity. Second, in Hummels and Schaur, the geographic structure is two points, East and West Coast, and no internal transportation cost is considered. In our paper, we consider a rich geography and jointly study domestic transportation cost as well as foreign transportation cost. Third, while we also make some use of published census tabulations of aggregated data, our main focus is on rich micro data.

Our micro data are the bills of lading filed with the U.S. Customs and Border Patrol (CBP) as part of the customs process. There is a recent international trade literature (see Bernard et al (2007, 2009, 2010)) that has linked confidential customs data to firm-level information. Unlike this earlier literature, we are interested in the specifics of the geography of where goods arrive. Moreover, by using the publicly available bills of lading data rather than the confidential data, we are able to report firm-level estimates.

Our paper is part of a recent literature integrating the analysis of international trade with intra-regional trade. See, for example, Holmes and Stevens (2014), and Cosar and Fajgelbaum (forthcoming). More broadly, it is part of an emerging literature aiming to estimate tractable, yet highly detailed models of economic geography, including

Ahlfeldt et al (2015), and Allen and Arkolakis (2014), and recent work specially aiming to quantify intraregional trade frictions, such as Atkin and Donaldson (2015).

There is extensive analysis of port choice in operations research. Leachman and Davidson (2012), for example, explicitly consider Wal-Mart’s port choice problem in a calibrated model.² A key difference here is the way we take a revealed preference approach, estimating parameters based on Wal-Mart’s behavior.

2.2 Model

2.2.1 Description of the Model

Consider the problem of a firm that imports a variety of products which it distributes across a variety of domestic locations. Starting at an originating foreign location, goods are shipped across the ocean, and unloaded at a domestic port of entry. Next goods are transferred to an import distribution center, where they are processed and then sent further down the distribution pipeline, which in general might include secondary, regional distribution centers. For simplicity, we will abstract from secondary distribution centers, and we will treat the primary distribution centers (i.e. the import distribution centers which are the first stop) as distributing goods to the ultimate consumers. This is a reasonable abstraction for Wal-Mart, because it has a large network of 42 secondary distribution centers for general merchandise, that are relatively close to the ultimate retail locations that they serve. (See Holmes (2011) for an analysis of Wal-Mart regional distribution centers.)

Formally, products are indexed by h , foreign source locations by i , domestic ports by j , distribution centers (henceforth “DCs”) by k , and ultimate consumer locations by l , and we let H, I, J, K, L denote the counts of goods, foreign locations, domestic ports, DCs, and ultimate consumer locations. We assume for each product h , the foreign source location is unique, and we write it $i(h)$. We can show that this is the usual case in our empirical application.

Firms pay freight costs to move goods from one transportation node to the next. We define the units of a product on the basis of volume, because that is how transportation

²See Bradley and Guerrero (2010) for an overview of what is called the *global replenishment problem*. See also Veldman and Bückmann (2003) and Veldman and van Drunen (2011).

is priced in the empirical application. Specifically, define one unit of a product to be one cubic meter. Let $r(a, b)$ be the freight cost in dollars to move one cubic meter from node a to node b . If goods start at foreign source i and get to ultimate location l by going through port j and DC k , then the freight cost of this journey in dollars for one unit of the good (i.e. one cubic meter) is

$$\text{total freight cost (dollars per unit)} = r(i, j) + r(j, k) + r(k, l).$$

Analogously, let $e(a, b)$ be the elapsed time taken to go from a to b , measured in units of days. Letting τ_h be the dollar cost of one extra day to deliver one cubic meter of good h , the time cost in dollars is

$$\text{total time cost (dollars per unit)} = \tau_h [e(i, j) + e(j, k) + e(k, l)].$$

There are also port and DC-specific costs, each with a deterministic and random component. To introduce the random component, let each unit of a good be indexed by g . The port-specific cost of moving unit g through port j is

$$x_j + \varepsilon_{jg},$$

while the DC-specific cost to move unit g through k is

$$y_k + \eta_{kg}.$$

The deterministic port-cost component x_j accounts for any additional direct dollar costs not included in the freight rate, such as port-specific container taxes, as well as the dollar value of any implicit costs of using a particular port, such as congestion associated with a port. The deterministic DC-specific cost y_k of using DC k accounts for wage differences across locations and concerns about unionization, and can also vary across locations to reflect different investments the firm might have made across DCs that affect DC marginal cost.

Let the random DC cost component $\eta_{g,k}$ be drawn i.i.d. across units g and DCs k , from the type-one extreme value distribution with standard deviation σ_η . For now, make a similar assumption on the random port-cost component $\varepsilon_{g,j}$, setting the standard

deviation at σ_ε . (In the estimation we will modify this assumption, to have the random draw take place at the container level rather than the cubic meter level.)

Putting all of this together, the total cost to move unit g of good h from foreign location i through port j and DC k to ultimate location l is

$$c_{ghijkl} = r(i, j) + r(j, k) + r(k, l) + \tau_h [e(i, j) + e(j, k) + e(k, l)] + x_j + y_k + \varepsilon_{gj} + \eta_{gk}.$$

We take demand at each location as given. Specifically, the firm solves a planning problem to deliver $M_{h,l}$ units of good h to location l . To allow for demand shocks across locations, assume $M_{h,l}$ is drawn from a distribution $\Phi_{h,l}(M)$, with mean $\mu_{h,l}$. We assume a flexible distribution and do not rule out correlation in the realization of these draws across nearby locations. We assume the firm observes the draw of $M_{h,l}$ before making order decisions.

In our assumptions about timing, we want to take into account that setting up a distribution network is a more longer term decision than port choice. This motivates our assumption to allow port choices to be made *after* observing the port random shock ε_{gj} , while decisions about what distribution center to use is made before this realization. Another thing we desire to capture in our model is to parameterize the degree to which a firm can customize its supply chain from the distribution center onward for particular products. In one extreme case there is no customization. In this case, a given DC ships the products it carries to the same set of downstream locations in a similar way. This will be the case when economies of scale are important, making it cost efficient to run a homogeneous distribution network. In the opposite extreme case, different products get treated differently in the same DC. There may be an incentive to do this, if scale economies are not important, and if the same distribution center carries goods from very different originating countries. For example, suppose there is one DC on the East Coast, and another on the West Coast and suppose there is one good (product “E”) from a country near the East Coast, and a second product (“W”) produced near the West Coast. If product-level customization is feasible, it might be desirable to have the East Coast DC reach further into the interior with product E compared to product W . That is, the optimal allocation shares are likely to differ.

To formalize this in the model, assume for each good h and each location l , there is a

probability λ that the firm can customize how goods leave the DC to get to the ultimate location, and with complementary probability $1 - \lambda$ the supply chain is constrained to treat goods sent to location l the same way, regardless of origination. Formally, assume in the event the firm is able to customize distribution of product h , the firm picks s_{kl}^h , the share of good h supplied by DC k to location l , $k = 1, \dots, K$. In the event the firm is constrained to treat all goods the same way, the firm selects a share s_{kl}^o from k to l that does not depend upon h .

2.2.2 Port-Choice Problem

We begin the analysis of the firm's problem with the port choice decision. The timing is such that when the port decision is made, the firm has already determined for each good h , the allocation of deliveries across DCs. Hence, in the port decision, we take as given that a particular unit g of product is being shipped to DC k . Thus costs at the DC onward can be taken as fixed. At this point, we need only to focus on the cost of getting to the DC. Consider unit g of a good that needs to be shipped to DC k . The cost of using port j is

$$\begin{aligned} c_{ghi}^{port} &= r(i, j) + r(j, k) + \gamma_h [e(i, j) + e(j, k)] + x_j + \varepsilon_{gj}. \\ &= \delta_{hijk}^{port} + \varepsilon_{gj} \end{aligned}$$

where we gather the deterministic components of cost into δ_{hijk}^{port} . For each destination DC k , we allow for the fact that some ports may not be economically feasible, and we let $\Lambda(k)$ be the relevant choice set of k . The optimal port minimizes

$$\min_{j \in \Lambda(k)} \delta_{hijk}^{port} + \varepsilon_{gj} \quad (2.1)$$

The probability port j is selected is

$$s_{hijk}^{port} = \frac{\exp\left(-\frac{\delta_{hijk}^{port}}{\sigma_\varepsilon}\right)}{\sum_{j \in \Lambda(k)} \exp\left(-\frac{\delta_{hij'k}^{port}}{\sigma_\varepsilon}\right)},$$

The expected minimized value of cost for delivering goods to DC k is

$$\tilde{c}_{hik}^{DC} = -\sigma_\varepsilon \ln \left(\sum_{j \in \Lambda(k)} \exp \left(-\frac{\delta_{hijk}^{port}}{\sigma_\varepsilon} \right) \right) - \sigma_\varepsilon \gamma, \quad (2.2)$$

where γ is Euler's constant.

2.2.3 The DC Allocation Problem with Product-Level Customization

Suppose the supply chain can be customized at the product level. When making the order decision, the firm first observes $M_{h,l}$, which is the inelastic demand for product h that must be delivered to location l . We can write the cost of choosing DC k to ship unit g as

$$\begin{aligned} c_{ghik}^{DC} &= \tilde{c}_k^{h,DC} + r(k,l) + \tau_h e(k,l) + y_k + \eta_{gk}. \\ &= \delta_{kl}^{h,DC} + \eta_{gk}. \end{aligned}$$

The cost includes the expected cost for a unit of h to get to DC k , plus the freight and time costs of getting from k to l , plus the DC cost y_k , plus a random component of cost. As before, we collapse the deterministic components of cost into $\delta_{kl}^{h,DC}$. Note this will depend upon h for two reasons. First, different products may originate from different countries that are at different distances from a given distribution center. Second, products may vary in the value of time parameter τ_h . The probability that DC k is selected is

$$s_{kl}^h = \frac{\exp\left(-\frac{\delta_{kl}^{h,DC}}{\sigma_\eta}\right)}{\sum_{k'=1}^K \exp\left(-\frac{\delta_{k'l}^{h,DC}}{\sigma_\eta}\right)}.$$

The minimized cost is

$$\tilde{c}_l^h = -\sigma_\eta \ln \left(\sum_{k=1}^K \exp \left(-\frac{\delta_{kl}^{h,DC}}{\sigma_\eta} \right) \right) - \sigma_\eta \gamma.$$

The expected share of product h going through k to all destinations, conditional on customization, is

$$s_k^h = E \left[\frac{\sum M_{h,l} s_{kl}^h}{\sum M_{h,l}} \right],$$

where the expectation is taken with respect to the random realization of demand $M_{h,l}$.

2.2.4 The DC Allocation Problem without Customization

Now consider the case where the firm is limited in its ability to customize the distribution network at the product level. In particular, suppose the firm is constrained to select share s_{kl}° to supply goods at l from k , which note does not depend upon h .

The expected total volume of goods being shipped to l across all goods h is

$$Q_l = E \left[\sum_{h=1}^H M_{h,l} \right]$$

Define

$$\delta_{kl}^{\circ,DC} = \frac{E \left[M_{h,l} \delta_{kl}^{h,DC} \right]}{Q_l} \quad (2.3)$$

This is the weighted average of the deterministic component of cost. The optimal DC shares to supply location l are then

$$s_{kl}^\circ = \frac{\exp\left(-\frac{\delta_{kl}^{\circ,DC}}{\sigma_\eta}\right)}{\sum_{k'=1}^K \exp\left(-\frac{\delta_{k'l}^{\circ,DC}}{\sigma_\eta}\right)}, \quad (2.4)$$

and minimized cost is

$$\tilde{c}_l^{*,l} = \sigma_\eta \ln \left(\sum_{k=1}^K \exp \left(\delta_{kl}^{*,DC} \right) \right) + \sigma_\eta \gamma. \quad (2.5)$$

The expected overall share of DC k , across all goods, is

$$s_k^\circ = \frac{\sum_{l=1}^L Q_l s_{kl}^\circ}{\sum_{l=1}^L Q_l} \quad (2.6)$$

Having calculated the value with and without customization, the expected minimized

cost is then

$$\tilde{c}_l^{overall} = \lambda E \left[\frac{\sum_{h=1}^H M_{h,l} \tilde{c}_l^{h,l}}{Q_l} \right] + (1 - \lambda) \tilde{c}_l^{\circ,l}$$

It is immediate that customization is beneficial (i.e. $\tilde{c}_{kl}^{overall}$ strictly decreases in λ).

Total distribution cost across all locations equals

$$\tilde{c} = \sum_{l=1}^L Q_l \tilde{c}_l^{overall}.$$

The expected share overall volume share of DC k is

$$s_k = \lambda E \left[\frac{\sum_{l=1}^L \sum_{h=1}^H M_{h,l} s_{kl}^h}{\sum_{l=1}^L \sum_{h=1}^H M_{h,l}} \right] + (1 - \lambda) s_k^{\circ}.$$

If $\lambda > 0$ so at least some customization is possible, then in general goods that have different origination countries will have different average shares. In contrast, if $\lambda = 0$, then under regularity conditions about the distribution of demand shocks and transportation costs, expected shares will be the same across originating countries. In the empirical application, we will find that the shares are close to invariate, consistent with $\lambda = 0$.

With the model in place, we now turn to a description of the data.

2.3 Data

2.3.1 Bill of Lading Data

The U.S. Department of Customs and Border Protection (CBP) releases detailed information about bills of lading of waterborne imports. A bill of lading is a document issued by a carrier that provides details of the shipment. The CBP sells the raw data to various shipping information companies who then resell the data. We have accessed this data through a subscription to Ealing Market Data Engineering. Over the period 2007 through 2015 that we consider, there are typically a little more than one million bills of lading each month.

Table 2.1 illustrates information from two example bills of lading, one for a shipment

of children’s car seats, another for a shipment of 50-inch HD TVs. The place of receipt is listed, as well as the foreign port where the shipment was laded onto the vessel. These happen to be the same for the two examples, but will differ when the place of receipt is inland, or when there is transshipment at a hub port.³ For example, goods originating in Bangladesh, destined for West Coast U.S. ports, are frequently transhipped in Singapore, which is then listed as the foreign port. The record specifies the name of the vessel carrying the shipment to the U.S., as well as the arrival date. The *port of unloading* is the location where the shipment was unloaded off the vessel and the *port of entry* is where the shipment clears customs. These generally differ for intermodal shipments. For example, the car seat record indicates that the shipment was unladed in Los Angeles, and was then shipped (presumably by rail) to Houston, where it went through customs. For the TVs, the ports of unloading and entry are both Norfolk.

In the car seat example, the shipper is listed as the Evenflo Company (a manufacturer of car seats), and the consignee is listed as Wal-Mart. Shippers and consignees may request that both the shipper and consignee data fields be suppressed. The first record is among of the rare exceptions where the confidentially option was not chosen by Wal-Mart. Information in the remaining fields can potentially be used to identify the shipper or consignee. Notice, in particular, in the products field for both records, you can see the code “GLN# 0078742000008.” This is the global location number that Wal-Mart uses for shipments with destinations in the United States. A search for the text pattern “0078742000008” in either the products field, or in a related field called “marks,” results in 1.5 million bills of lading observations for the period 2007 to 2015. That these are indeed Wal-Mart observations can be corroborated from other information on the records. Using various other strategies described in an online appendix, we are able uncover an additional 0.4 million records for a total sample of approximately 1.9 million Wal-Mart bills of lading.⁴

All internationally shipped containers have a unique 11 digit identification number

³It is important for record keeping that the place of receipt be accurate, because the actual country where the good originates may affect customs duties, and because transportation costs in moving inland freight is deductible for customs duties.

⁴Note we exclude from this sample 90,000 bills of lading for shipments going to Canada and Mexico that go through US ports.

Table 2.1: Information Contained in Two Example Bill of Ladings

Variable Name	Car Seat Example (Consignee and Shipper shown)	HD TV Example (Consignee and Shipper withheld)
Bill of Lading Number	MAEU582050599	MSCUY7523668
Place of Receipt	Qingdao	Yantian
Foreign Port	57047 - Qingdao, China	57078 - Yantian, China
Vessel Name	Cma Cgm Vivaldi	Msc Maria Elena
Arrival Date	5/5/2012	8/10/2014
US Port (Unlading)	2704 - Los Angeles, California	1401 - Norfolk, Virginia
US Port (Entry)	5301 - Houston, Texas	1401 - Norfolk, Virginia
Shipper	Evenflo Company Inc. 4 A Hillwood Road,Tst,Hongkong China	
Consignee	Wal-Mart Stores, Inc. 601 N Walton Blvd Bentonville, AR 72716	
Count of Containers	2	417
Container BIC Code	POCU1011444,MRKU0190542	TCLU9487515,MSCU9800119, etc.
Container Piece Counts	1056,1056	297, 297, (rest are the same)
Volume (Cubic Meters)	126 (63 per container)	123849 (65.7 per container)
Weight (Kilograms)	8131 (4066 per container)	2278821 (5,265 per container)
Products	Evenflo Big Kid Booster Po#7 902676111 Item#550117639 GLN #0078742000008 Po Type#45 Dept#79 Shipper Provided H S Code:HS Code:9401806020	"50" Full Hd Led Tvstock No. Lf501em5f Assortment/Item No. 552794071 Po No. 9352749202 Funai Corp P/O No. A33909 Funai Corp P/O No. A33940 Funai Corp P/O No. A33941 Funai Corp P/O No. A33965 Gln: 0078742000008 Department No.: 00072 Hts:8528726400 Po Type:0045

(BIC code) that is maintained by an international registry. Each bill of lading lists the BIC codes of all the containers that are part of the shipment. The car seat shipment consists of two containers. There are 417 containers for the TV shipment (this is the maximum container count across all Wal-Mart shipments in our sample). For each

container, the record lists the piece count, which for Wal-Mart is generally the count of cartons in the container. For the car seat example, each carton contains a single car seat, and there are 1056 units in each container. Similarly, for the TVs, each carton is a single TV and there are exactly 297 TVs in each of the 417 containers. Some web searches show that the original Wal-Mart prices for these products were \$50 for the car seat and \$498 for the TVs, which, after multiplying by the piece counts result in container-level retail values of \$53,000 and \$148,000 for the car seats and TVs, and the total retail values of the orders are \$103,000 and \$62 million.

Volumes per container equal 63 and 65.7 cubic meters in the two examples, which are close to the maximum usable space of 67.7 cubic meters for a 40ft standard container.⁵ Weights per container in the two examples are 4,066, and 5,265 Kg, significantly less than the maximum container weight of around 25,000 Kg. This illustrates an important point that for the types of goods Wal-Mart sells, containers hit capacity first in volume instead of weight. Since container shipping prices do not depend upon weight, transportation cost considerations are all about the volume of the goods, not the weight. (See also Leachman (2005).) Since the cubic meter volume measure is missing for many records in the sample, we will use counts of containers, and allocated portions of containers, as our unit of measure.

We henceforth refer to each individual bill of lading as a *shipment*. Table 2.2 tabulates the distribution of shipment counts and container counts across the nine years in our sample. One thing to note is that it is common for there to be multiple shipments corresponding to the same arriving container. This happens when a particular order does not fill up a container. Wal-Mart will then generally combine orders with different bills of lading into the same container, to efficiently stuff containers full. Fortunately, we have the BIC container codes for each shipment, so we can keep track of when Wal-Mart does this. The total count of unique containers in our sample is 1.7 million, a little less than the bills of lading count of 1.9 million.⁶ To get a sense of the coverage of our sample, we compare our counts with statistics on company-level aggregate annual container imports, published by PIERS.⁷ Overall, the container count in our sample is

⁵See Leachman (2005, p 77).

⁶Note the same container may be used on multiple ocean voyages. Throughout we will refer to the same containers used on different voyages as different containers.

⁷To produce these estimates, PIERS uses the same bill of lading data from CBP that we have, plus

53 percent of the aggregates reported by PIERS.⁸ One potential source of a difference is mistakes in the way the paperwork is filled out, e.g. the reported GLN may be off by one digit and we miss this. In our statistical analysis will attempt to take this into account by focusing on products where we think the measurement error is likely to be small. Another possibility is that different departments might have different paperwork protocols, e.g., that do not include the GLN, and there are likely certain categories of goods for which we are missing all the records for all the products. For much of what we do, this will not be a problem, because we mostly will be conditioning on particular products, and as we will see, we will have plenty of them.

Table 2.2: Counts of Bills of Lading and Containers in Wal-Mart Sample

	Wal-Mart Sample		PIERS Published Aggregates	Sample Share Relative to PIERS (Percent)
	Count of Shipments (1000)	Count Containers (1000)	Count Containers (1,000 FEU*)	
All Years	1,892	1,733	3,267	53.1
By Year				
2007	222	261	360	72.5
2008	191	208	351**	59.4
2009	165	160	342	46.8
2010	179	167	348	48.0
2011	179	154	355	43.4
2012	186	149	360	41.4
2013	238	198	366	54.2
2014	259	212	388	54.7
2015	273	224	398†	56.3

*FEU is Forty Foot Equivalent. PIERS reports units in TEU (Twenty Foot), so figures for PIERS are 0.5 x published figure.

**The 2008 PIERS figure is based on interpolating 2007 and 2009.

†The 2015 PIERS figure is not yet available, and is estimated based on trend growth.

Table 2.3 displays the distribution of Wal-Mart container imports across country of origin. (We use the place of receipt variable to determine this.) We list every origin country with 10,000 containers or more. China completely dominates the list, it uses additional information it obtains directly from shippers.

⁸The exception of 72 percent in 2007 is a result of the fact for that year, there were many observations where Wal-Mart failed to suppress the consignee field.

accounting for 85.7 percent of all imports. The next highest is Bangladesh with 1.9 percent. Guatemala and Costa Rica make the list, mostly because of bananas. The rest of the leading countries are from Asia.

Table 2.3: Distribution of Container Imports by Country of Origin

Wal-Mart Sample (2007-2015)
Source Countries with 10,000 or More Containers

Origin Country	Count of Container Imports (1000)	Percent
All Countries	1,733	100.0
China	1,486	85.7
Bangladesh	32	1.9
India	29	1.7
Thailand	27	1.6
Vietnam	27	1.5
Indonesia	20	1.2
Pakistan	15	0.9
Taiwan	15	0.8
Guatemala	12	0.7
Malaysia	12	0.7
Costa Rica	11	0.6
Rest of World	47	2.7

In the example product fields in Table 2.1, you can see a nine digit “item number.” The *Wal-Mart item number* is an internal code the company uses as a stock-keeping unit to keep track of products. There products field in the examples also contain a ten digit HS product code that is used as part of customs filing. Finally there is a two-digit code specifying the department making the order. We have parsed the product field (and the related field called “marks”) to pull out these codes. For some records there are multiple item numbers, and an individual item number may appear multiple times in a record. In such cases, we allocate the piece count of a bill of lading record across item numbers, in proportion to the number of times the item number appears in the record. In cases where the same container shipment appears across multiple bills of lading, we allocate the volume of the container in proportion to piece counts. In the end, we produce a data set disaggregated to the level of an item and shipment (bill of

lading), where we associate each observation with an allocated portion of capacity. A typical item number is only in use a single calendar year, as products come and go. In the rare cases where the same item number appears across multiple years, we treat the different years as different items. Throughout, when we use the term product, we will be referring to an item number.

We obtain item numbers for the products in 1.33 million containers (76.6 percent of the original sample). However, some observations do not have the number of digits that we expect, which are 9 digits in the later years and 7 digits in the earlier years. (This may reflect drops of leading zeros, and in the future it may be possible to use additional text parsing techniques to clean this up.) To focus on the data we think is the best quality, we restrict attention to records with the expected number of digits. This results in a sample of 1.15 million containers (66.5 percent of the original). From this sample we delete shipments to Puerto Rico, and imports of fresh fruit from Guatemala and Costa Rica (which have a special distribution system) and do other minor cleaning described in the appendix, to obtain a final sample of 1.11 million containers. There are 252,674 unique item numbers in the final sample. Table 2.4 reports summary statistics at the item level. The mean item has a total shipped volume of 4.4 containers and is reported on 7.8 different bills of lading (i.e. 7.8 shipments). The distribution of container counts is quite skewed, with a median of 0.7 containers. (Recall we allocated containers both within a bill of lading when multiple items are listed, and across bills of lading sharing the same container.) The product with the maximum container count (3,136 containers) is the HD TV example of Table 2.1. The item with the maximum count of shipments (219) is a DVD player from 2009. There are 60,600 different items for which there is only a single bill of lading. These account for 24 percent of all products, but only 2.3 percent of containers.

Table 2.4: Summary Statistics of Final Sample

Product (Item) Level

	Count of Unique Products	Mean	Quartiles			Maximum
			25th	50th	75th	
Total Containers	252,674	4.4	0.1	0.7	2.5	3,136
Total Shipments	252,674	7.8	2.0	4.0	8.0	219

2.3.2 Ocean Time in Transit

To determine ocean time in transit, we use data on port call histories of vessels we obtained through a subscription product offered by a maritime tracking service named Fleetmon. The data are based on GPS readings of ship positions. This port call data is for the period October 2013 through the end of 2015. For our Wal-Mart shipments that originate during this period, we successfully match the port call data, including departure time from foreign port and arrival time at U.S. port, for 85 percent of the sample. Wal-Mart has numerous shipments from China, resulting in a large sample size for estimating ocean times from China. As noted earlier (see Table 2.3), the sample of shipments from other countries is relatively small. To estimate travel times to these other countries, we expand our sample to include shipments for other firms besides Wal-Mart. Specifically, for the months Nov. 2013 through March 2014, and again Nov 2014 through March 2015, we use the complete set of all bills of lading. There are 10.8 bills of lading in this sample of ten months of data. We match the port call time of departure and arrival for approximately 75 percent of these observations.

2.3.3 Ocean Freight Rates

We require estimates of ocean freight rates. There exist spot market indices (e.g. the Drewry World Container Index), but there are two problems using these indices. First, the indices report spot rates. However, most transactions occur based on contract rates. In particular, we expect Wal-Mart's long-run decision making to be based on contract rates. Second, the available indices are for a limited set of port combinations, and we need freight rates for a wide set of possible combinations of foreign and domestic ports. A second resource for freight rates is the published tabulations from the U.S. Census of freight charges based on customs records. Expenditures on freight are deductible for customs duties, so freight is collected as part of the administrative procedure. The Census publishes detailed tabulations by narrow product type, month, and port of unlading and port of entry, and originating country. In these tabulations, the freight charges, the merchandise values, and the weight (kilograms) are reported for each of these cells, even if there is only one record (the data also includes a count of records). While this data is extremely rich, as noted earlier, for the kinds of goods Wal-Mart sells,

freight rates are based on volume not weight.

Our strategy to address this issue is to match the container information in the public bills of lading, with the freight charge information in the published Census tabulations. The strategy has several parts, and we explain the details in the appendix. But one of the main parts of the strategy is linking cases in the published tabulations where it is indicated that there is only a single transaction in that cell. Since the Census tabulation reports weight, we look for a matching observation in the bills of lading with the same weight in the same cell (e.g. country of origination, port of unloading, and port of entry). For 18 months, we have a complete set of bills of lading (17.8 million records) that we use for the matching exercise. We obtain 140,000 matches of Census data for which we can merge in container information.

2.3.4 Internal Geography and Ultimate Consumers

For our internal geography, we partition the U.S. according to core-based statistical areas (CBSAs). For those rural areas outside of CBSA, we allocate them according to BEA Economic Areas. After we eliminate Alaska, Hawaii, and Puerto Rico, we are left with a partition of the forty eight contiguous states into 1095 geographic units, that correspond to the locations indexed by l in the model.

Our analysis requires an estimate of the inelastic demand at each domestic location. We use estimates of Wal-Mart's store level sales for 2006 from data in Holmes (2011), and make adjustments for later dates using the distribution of stores in a particular year.

2.3.5 Domestic Freight and Time

In our estimation we make use of the rich micro data that is available on transportation times and costs within the U.S. from a number of different sources. As the basis of our inland transportation costs and times, we rely on Leachman (2005). Leachman (2005) contains information on shipping costs and times for both rail and truck on intermodal routes from eight U.S. ports (Charleston, Houston, LA-Long Beach, New York, Norfolk, Oakland, Savannah, and Seattle) and 21 destinations within the U.S.⁹

⁹See Leachman (2005), Tables 14 and 18.

In order to have a more complete set of costs and times for the entire U.S. ground transportation network, we supplement the Leachman data with information from other publicly available sources. For domestic freight rates (for both rail and truck) and for truck shipping times, we appeal to *worldfreightrates*, where we collected costs and times for 714 domestic shipping routes. To get the total inland transit time for truck shipments, we extend the truck times obtained from *worldfreightrates* (which are drive times from google maps) to account for driver rest requirements that mandate 10 hour breaks for 11 hours of driving. For transit times on rail, we use published intermodal rail schedules from the major rail carriers serving ocean ports (Union Pacific, BNSF, CSX, CN, and Norfolk Southern). Published rail transit times were collected for 556 routes between major ports and destinations. On routes with multiple trips of different times, the average transit time was used.

Although the Leachman data and the supplemental data are highly correlated for routes in both samples, we regress each of the Leachman measures (truck and rail times and costs) on each of the supplemental measures to have more comparable estimates to combine into a single set of times and costs. As expected the estimated coefficients from these regressions confirm the consistency of the various data sources. For rail time, where we regress the time from Leachman on the time from the rail schedules and an intercept, the estimated intercept is 1.08 and the estimated slope is 1.04 indicating that we need to add approximately one day to the rail schedules to account for drayage and loading/unloading times that are incorporated into Leachman's estimates. For the truck time regression we again obtain an estimated intercept near one (1.06) but a larger slope of 1.7, implying that the *worldfreightrates*/google maps time estimates are faster than the typical shipping times. We again transform the supplemental data using the regression coefficients so that it conforms with the Leachman data. Similar exercises are performed for rail and truck costs.¹⁰ With the combined data, we have information for 740 domestic shipping routes. While this larger data set contains information for the major intermodal routes, we need estimates of times and costs between the distribution centers and every final destination within the United States. To fill in these unobserved

¹⁰There are larger differences between Leachman and the supplemental data for rail costs. In this case, we obtain an intercept of 621 and a slope of 0.35. Truck costs are closer to the original pattern with an intercept of -67 and a slope of 1.03.

times and costs we rely on the large geographic scope of the observed data and interpolate the times and costs giving more weight to observed costs/times that are closer a destination with unobserved times/costs.

2.4 Preliminary Analysis

This section establishes facts about Wal-Mart’s behavior. These facts play a role in the quantitative analysis with the model in the next section.

2.4.1 Import Distribution Network

Wal-Mart’s strategy for general merchandise imports of using five import distribution centers is described in MWPVL, and in Leachman. With our data it is easy to see Wal-Mart using the strategy, as many records list the destination DC in the products or marks field. Also, the data include port of entry and port of unloading. For example, it is straightforward to see in the data that containers going through Seattle are virtually all going to the Chicago DC because (1) most of these records list Chicago as port of entry, and (2) in small percentage of cases where district of entry is left blank, there typically is indication in the product field or marks field that the shipment is going to Chicago (or actually Elwood, which is the town in the Chicago area where the import DC is located.)

Wal-Mart employs different import distribution systems for other classes of products. In particular, there is a separate system for footwear and clothing (the fashion distribution system). In addition, there is a system for the Sam’s Club warehouse store format. The Sam’s club records are identified by the first two digits of the item number being “63.” Fashion goods are identified by two-digit HS codes 61, 62, and 64, and additional text information in the products field for when the HS code is missing. General merchandise is the residual. Table 2.5 reports the distribution of product counts, shipment counts, and container counts across these 3 broad groupings of products.

We illustrate the different systems with the following exercise. We start by taking each product and count the total number of different shipments of the product. Next we calculate the total number of different destination DCs used for each product. The maximum is five (5-DC), the minimum is one (1-DC) Table 2.6 reports the distributions

Table 2.5: Wal-Mart Product Categories

Product Categories	Count of Unique Products	Count of Shipments, (1000)	Count of Containers, (1000)
All Products	252,674	1,965	1,113
General Merchandise	135,768	1,460	836
Fashion	111,270	440	128
Sams Club	5,636	65	148

across the count of DCs reached, and count of product shipments, for both general merchandise and fashion goods. For goods with only one shipment, 100.0 percent are necessarily 1-DC. With two shipments, some are 1-DC and others are 2-DC. Inspection of Table 2.6 reveals clear differences between general merchandise and fashion. Consider first goods with 5 or more shipments (except for the case of 6 shipments which we come back to). For General Merchandise, the mode is 5-DC. In fact, for the 7,858 products with 41 or more shipments, a 0.97 share hit all the DCs at least once. In contrast, for fashion, the mode outcome is 1-DC. This is true even for goods with 21-40 shipments. Typically, shipments all go to same distribution center. The exception for fashion is the category of 41 or more shipments, which only has a few cases, and these tend to go to all five DCs. Manual inspection of the 5-DC fashion cases reveals that these are footwear products like “98 cent flipflops” that are being run through the general merchandise system. Flipflops lack the variety of different sizes that are typical for footwear, and are therefore more like general merchandise.

Table 2.6: Distribution of Products by Number of DC Destinations

General Merchandise and Fashion

Count of Product Shipments	General Merchandise						Fashion					
	Count Products	1 DC	2 DC	3 DC	4 DC	5 DC	Count Products	1 DC	2 DC	3 DC	4 DC	5 DC
1	23,569	1.00					36,044	1.00				
2	13,192	0.35	0.65				21,457	0.80	0.20			
3	11,253	0.12	0.17	0.71			12,020	0.68	0.15	0.16		
4	14,490	0.04	0.09	0.11	0.75		10,044	0.49	0.24	0.09	0.18	
5	15,318	0.02	0.04	0.09	0.10	0.75	7,518	0.50	0.13	0.08	0.12	0.17
6	3,971	0.05	0.10	0.25	0.33	0.27	5,265	0.51	0.17	0.08	0.12	0.12
7-10	18,421	0.02	0.04	0.06	0.27	0.61	11,275	0.49	0.11	0.05	0.15	0.20
11-20	17,245	0.02	0.02	0.02	0.15	0.79	6,439	0.49	0.14	0.02	0.10	0.26
21-40	10,461	0.01	0.01	0.02	0.06	0.90	1,153	0.47	0.11	0.01	0.02	0.39
41 and above	7,848	0.00	0.00	0.01	0.02	0.97	55	0.13	0.02	0.00	0.07	0.78

Next look at general merchandise products with exactly 5 shipments. For a 0.75 share, exactly one shipment goes to each DC. There are is also a 0.10 share where 4 DCs are reached, i.e. one is missed, and another gets two. From manual inspection, we believe that many of the latter are reporting errors. For example, we see cases where there are two similar size shipments to Los Angeles, and none to Chicago; we believe that in many of these cases, one of the two is destined for Chicago, but this was not noted in the paperwork. Next, observe that cases with 4 product shipments, a 0.75 share hit 4 DCs. In the appendix (to be completed) we present an argument that this is likely to be measurement error where we miss a shipment, rather than Wal-Mart actually sending shipments to only 4 DCs, and then reshuffling to get product to the territory covered by the missing DC. Also, many products we are classifying as general merchandise that go to only one DC may actually be fashion goods. For these reasons, in our analysis of general merchandise, we will condition on products with shipments to all 5 DCs, where we think the data is the best, and this subsample accounts for 75 percent of the containers in our initial general merchandise sample. We also place special focus on the case where there are exactly 5 shipments, one to each DC. We regard this as a clean case. These are instances of "one-off" orders where Wal-Mart puts in an order with no subsequent order for replenishment. As we will show in a later version of the paper, these 5 shipments tend to arrive at approximately the same day to their destination DCs. One-off situations are relatively common. Notice that the frequency count for 6 shipment products drops off significantly (There 15,318 with 5 and 3,971 with 6), and moreover these don't fit the pattern of a mode point at 5 DCs. Some of these are likely misclassified fashion items.

2.4.2 Order Allocation and Country of Origin

In this subsection we examine how order allocations vary across country of origination, for destination DCs, and the ports used to get there. For much of the analysis, we group countries into three geographic categories. The first is China plus Taiwan. To get to East Coast ports from China or Taiwan, ships generally sail east through the Panama Canal. The next category includes the remaining countries in lower part of the South China Sea. To get to the East Coast from here, ships generally sail west, through the Suez Canal. The third group of countries have ports in the Indian Ocean. For

these, sailing time to the East Coast is less than to the West Coast.

We start by looking at fashion products that are delivered to a single DC. This selection excludes, for example, the 98 cent flip flops that have the footwear HS code, but are run through the general merchandise DC system. Table 2.7 reports the DC shares, by country grouping, and by individual countries of origin for the entire sample of 79,805 products. The bottom panel considers the subsample of 15,508 products with five or more shipments. Consider first the bottom panel, where the pattern is crystal clear. For China and Taiwan, there is roughly a 50/50 split of Houston or Chicago as the DC. For the South China Sea category, Chicago's share drops close to zero, and Houston's share rises all the way to 0.80. These countries are further south than China, making Chicago less desirable. Also, Houston is relatively closer, and the option value of being able to get to Houston through LA from the west, or from the east through the Suez canal is valuable. And notice that Savannah's share creeps up to 0.16, which is the East Coast option, as opposed to the more central Houston option. Finally, for the Indian Ocean countries, which are relatively close to the Suez Canal, virtually all (0.98 share) are shipped to the East Coast DC at Savannah. When we look at the top panel that includes products with less than 5 shipments, and the breakdown by individual countries, there is noise as some of the smaller cases with few shipments results in some spreading across the DC, i.e. the Savannah share for Indian Ocean countries is 0.81 rather than 0.98. Nonetheless, the overall pattern remains clear. The allocation choice of the single destination DC for handling a particular fashion good depends significantly on the location of the originating good.

Next in Table 2.8, we report DC share allocation for general merchandise products. We condition on products with deliveries to all 5 DCs, for which the data quality is best. The table reports mean share allocations by country. The mean shares are remarkably similar across countries. In the bottom panel, we consider the "one-off" subsample, where there is a single shipment for each product to each DC. Above we argued that this subsample is likely to be particularly clean. We can see that the pattern of allocation shares being constant across the geographic groups is particularly sharp. When we go to our quantitative analysis in the next section, we take Table 2.8 as evidence consistent with the customization ability parameter in our model being zero.

Table 2.7: Mean Share Allocation across Destination DCs

By Country of Origin
Fashion single-DC Sample

All Products							
		Count of Products	Los Angeles	Houston	Chicago	Savannah	Norfolk
All Asia		79,805	0.02	0.29	0.35	0.25	0.08
By Asia Region							
China/Taiwan		62,490	0.02	0.33	0.43	0.14	0.07
South China Sea		9,471	0.04	0.23	0.09	0.51	0.13
Indian Ocean		7,844	0.03	0.06	0.02	0.81	0.08
Selected Individual Countries Region							
China/ Taiwan	China	62,455	0.02	0.33	0.43	0.14	0.07
	Taiwan	35	0.00	0.00	0.89	0.00	0.11
South China Sea	Vietnam	1,971	0.08	0.42	0.27	0.17	0.05
	Cambodia	4,197	0.01	0.29	0.03	0.57	0.11
	Singapore	133	0.21	0.11	0.27	0.26	0.15
	Indonesia	3,105	0.05	0.04	0.03	0.66	0.22
Indian Ocean	Bangladesh	6,620	0.01	0.05	0.01	0.85	0.07
	India	1,102	0.14	0.08	0.06	0.62	0.10
	Sri Lanka	60	0.08	0.07	0.15	0.55	0.15
	Pakistan	62	0.05	0.24	0.29	0.39	0.03

Subsample of Products with 5 or more Shipments

		Count of Products	Los Angeles	Houston	Chicago	Savannah	Norfolk
China/Taiwan		14,188	0.01	0.51	0.48	0.00	0.00
South China Sea		518	0.00	0.80	0.03	0.16	0.00
Indian Ocean		802	0.02	0.00	0.00	0.98	0.00

So far we have looked at allocation of deliveries across DCs. Next we turn to the port choice decision. The Los Angeles, Savannah, and Norfolk DCs obtain virtually all of their deliveries from their nearby port. The Houston DC is at a port on the Gulf of Mexico, and obtains direct shipments that way, but also being central (compared to Savannah and Norfolk which are east), it is can also be reach by shipments to Los Angeles and rail for the rest of the way, and so a port choice needs to be made. For the

Chicago DC there is no ocean port, and shipments go through Seattle or Los Angeles from the west, or generally New York from the east.¹¹

Table 2.8: Mean Share Allocation across Destination DCs

By Country of Origin
General Merchandise 5-DC Sample

	Count of Products	Los Angeles	Houston	Chicago	Savannah	Norfolk	
All Asia	54,433	0.17	0.24	0.18	0.21	0.20	
By Asia Region							
China/Taiwan	51,688	0.17	0.24	0.18	0.21	0.20	
South China Sea	1,166	0.18	0.23	0.19	0.21	0.19	
Indian Ocean	1,579	0.17	0.23	0.19	0.20	0.21	
Selected Individual Countries Region							
China/ Taiwan	China	51,082	0.17	0.24	0.18	0.21	0.20
	Taiwan	606	0.19	0.23	0.17	0.23	0.18
South China Sea	Vietnam	214	0.19	0.23	0.18	0.20	0.19
	Cambodia	263	0.18	0.23	0.19	0.22	0.18
	Thailand	164	0.16	0.24	0.18	0.21	0.21
	Indonesia	383	0.18	0.23	0.20	0.20	0.19
Indian Ocean	Bangladesh	269	0.18	0.20	0.19	0.20	0.24
	India	1,004	0.16	0.24	0.19	0.20	0.21
	Pakistan	305	0.19	0.21	0.18	0.20	0.21

“One-Off” General Merchandise 5-DC Subsample (Exactly 1 Shipment to Each DC)

	Count of Products	Los Angeles	Houston	Chicago	Savannah	Norfolk
All Asia	11,499	0.18	0.23	0.19	0.21	0.19
China/Taiwan	11,134	0.18	0.23	0.19	0.21	0.19
South China Sea	200	0.17	0.23	0.20	0.20	0.20
Indian Ocean	165	0.18	0.21	0.20	0.20	0.21

Table 2.9 reports what ports are used to get to Houston and Chicago for the two classes of products, by source. For fashion to Houston from China, and the South China Sea there is roughly a half and half split between Los Angeles and Houston. And to

¹¹Shipments to Chicago from the east sometimes also go through Norfolk, but for simplicity we are going to treat these as going through New York.

Chicago virtually all of the goods go through Seattle. As noted above, fashion goods from the Indian Ocean generally not go to either Chicago or Houston (Savannah of the usual destination). For the few exception, the choice for Houston is to go their directly rather than through Los Angeles, and to get to Chicago, Los Angeles is the typical choice.

Table 2.9: Port Choice Given Destination DC

Footwear/Clothing 1-DC Sample

	Destination is Houston DC			Destination is Chicago DC			
	Count Shipment	Los Angeles Port Share	Houston Port Share	Count Shipment	Los Angeles Port Share	Seattle Port Share	New York Port Share
China/Taiwan	87,862	0.42	0.58	102,566	0.00	1.00	0.00
South China Sea	6,074	0.62	0.38	1,561	0.10	0.90	0.00
Indian Ocean	496	0.10	0.90	205	0.58	0.20	0.22

General Merchandise 5-DC Sample

	Destination is Houston DC			Destination is Chicago DC			
	Count Shipment	Los Angeles Port Share	Houston Port Share	Count Shipment	Los Angeles Port Share	Seattle Port Share	New York Port Share
China/Taiwan	223,120	0.16	0.84	197,069	0.57	0.42	0.00
South China Sea	4,250	0.57	0.43	3,566	0.80	0.20	0.00
Indian Ocean	8,043	0.08	0.92	8,498	0.17	0.02	0.81

Turning to port choice for general merchandise, again we see that port choice is highly sensitive to origination. In particular, from the Indian Ocean, East Coast ports are relatively closer than the West Coast, so the Port of Los Angeles is less likely to be chosen compared to port choice with shipments from China or the South China Sea.

2.4.3 Order Allocation in Response to a Temporary Port Shock

Next we consider how order allocations respond to a port shock, in terms of destination DCs and ports that are chosen. In particular, we examine the West Coast dock slowdown that occurred between 2014 and early 2015. The contract expired July 1, 2014 and the workers continued to work without a contract. Late in 2014 slowdowns began to occur. There were arguments between management and labor as to why these occurred. Work slowed down dramatically in early 2015, and the labor issues were eventually settled on February 20, but there were several months of congestion after the settlement to work through the backlog.¹²

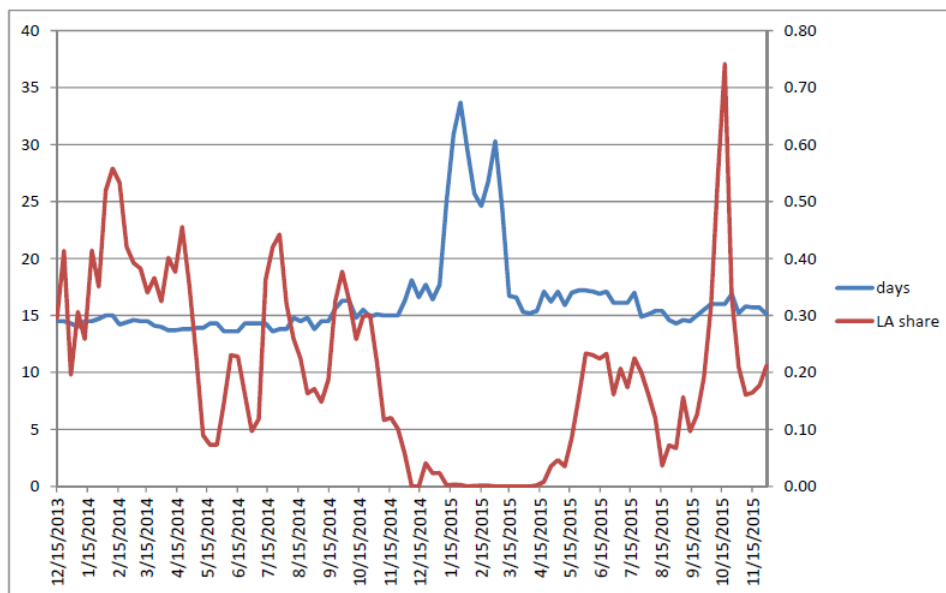
We can use our matched shipment/port-of-call data to measure the extent of the slowdown. We take our Wal-Mart data set, and work with the data at the container level. For each container, we know the vessel carrying it, and with the port-of-call data we determine the time of departure from the foreign port and the time of arrival at the domestic port. We take the difference and call this the voyage duration. Figure 2.2 plots median duration across containers to get from Shenzhen or Shanghai to Los Angeles. We aggregate the shipments to the weekly level, based on the week the container departed from China. We use a one week ahead and one week behind rolling window, and take the median. We can see the median duration was between 14 and 15 days throughout most of 2014, until late in the year. The duration goes up in December, and then explodes to over 30 days in January and February. It then fell, but remained around 17 days, until August.

Next we consider the containers that Wal-Mart shipped from China (we focus on Shenzhen and Shanghai originations) to the Houston DC. As noted above, Wal-Mart chooses between direct shipment to the port of Houston and intermodal shipment through Los Angeles. The red line in the figure is a moving average of the share of containers going through Los Angeles. (Again, week is defined by when the container leaves China.) The share moves around, but averages about 30 percent. Observe that at the end of October, the share begins to drop, and falls to zero by the end of December. The share did not go back up again, until May, when things settled down. The episode clearly illustrates that in a case like Houston for which there is close substitute

¹²For details, see “Economy Still Reeling from West Coast Slowdown” US NEWS, By Andrew Soergel — Staff Writer Feb. 23, 2016.

Figure 2.2: Voyage Duration and Port Choice During West Coast Port Slowdown

Median Time from Shenzhen or Shanghai to Los Angeles (by week of departure)
Share of Containers Destined for Houston Going Through Los Angeles



port not affected by the slowdown, Wal-Mart can be very flexible.

The substitution possibilities for getting goods from China to the Los Angeles or Chicago DCs are more limited. In the height of the stoppage, Wal-Mart actually sent 3 containers to Chicago via the Houston port, but this is negligible. Wal-Mart also sent 553 containers to Los Angeles via Oakland, a port Wal-Mart otherwise rarely uses. Oakland is covered by the same union contract as the other west coast ports, but there was variation in the slowdown at different ports at different times, allowing for some substitution possibilities.¹³

Next we consider flexibility of the allocation shares across destination DCs for general merchandise. Let $s_j(t)$ be the mean allocation share at DC j across goods delivered at time t . For this exercise, we define the delivery date as the average arrival date of the first shipments to Norfolk and Savannah (we pick the two Atlantic ports because the delivery dates for these are less likely to be connected to the port disruption on the

¹³During the stoppage some shippers reportedly rerouted containers through Canadian ports and then used rail to get them to their U.S. destinations. These shipments are not in our sample.

Pacific). We focus on imports from China. Figure 2.3 reports nonparametric estimates of the mean allocation shares for the 5-shipment (“one-off”) sample, the bottom panel reports the same thing for products with seven or more shipments.¹⁴ The red vertical lines indicate January 1 of each year, beginning with January 2007. The plots for the two samples are quite similar and the most noticeable pattern is the sharp initial increase in the Chicago share. This DC opened June 2006, and the initial upward trend is consistent with Wal-Mart gradually ramping up its use of the new DC. The main takeaway from these plots is that the relationships are relatively smooth functions of time. In particular, we don’t see a collapse of shares for Los Angeles and Chicago, at the end of 2014, which is what we would expect to see if Wal-Mart could quickly reoptimize its share allocations in response to the port disruption. It is possible there might be a slight kink near January 1 in the top panel, with of Houston trending slightly more upward and Los Angeles trending down, but the magnitudes are quite small. We will see in the quantitative model that if Wal-Mart could instantaneously adjust its DC shares, the response would be quite large. We interpret Figure 2.3 as strong evidence that Wal-Mart cannot adjust DC allocation shares in the very short run.

2.5 Quantitative Analysis

Here we undertake our quantitative analysis. Part 1 of this section estimates the parameters of the port choice problem. Part 2 turns to the DC choice problem. Part 3 puts the estimated model to work.

2.5.1 The Port Choice Problem

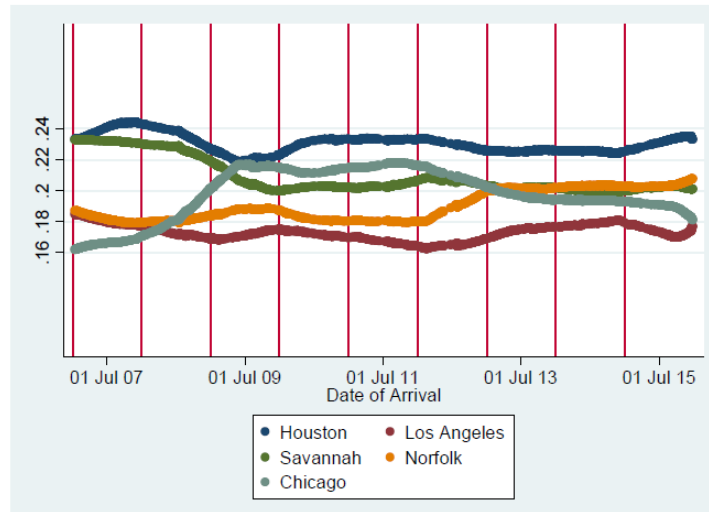
Consider the port choice problem as specified in (2.1). Recall the time cost parameter τ_h (dollars per day per container) in this problem is defined as specific to the particular product. Here we will assume that τ_h is constant τ_{gen} across all general merchandise goods, and is a constant $\tau_{fashion}$ across fashion goods.

In the analysis of the port problem, we will take as known the ocean trip durations and freight rates. Besides the time cost, the parameters to be estimated are the σ_ε and

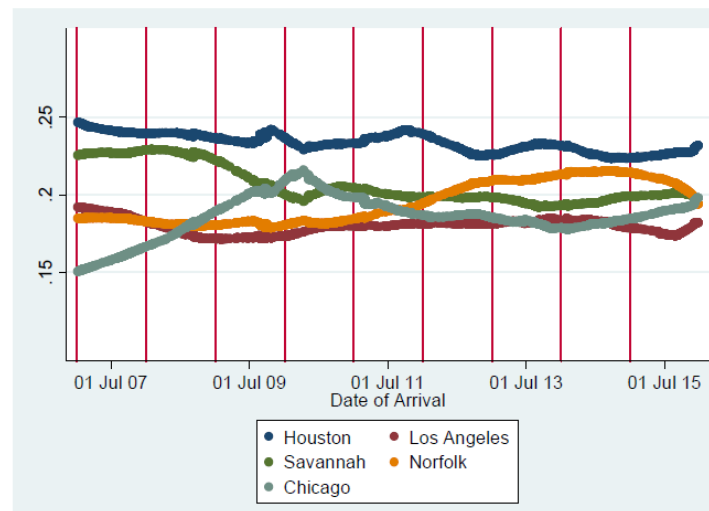
¹⁴We also condition on the shares at each DC exceeding 0.10, and total annual shipment be the equivalent of at least one full container. For the 7 plus sample, we require the first and last shipment be within 60 days.

Figure 2.3: Plot of Estimated DC Share Allocations by Time for Two Samples

LOWESS Smoother (Bandwidth = .2)



(a) Sample with Exactly 5 Shipments



(b) Sample: Condition with 7 or more Shipments

the port specific costs x_j . There is also a freight and time cost to get from the port to the destination DC . It is convenient at this point to specify port specific cost as x_{jk}

and let this absorb the DC-specific cost, as well as all costs to be transferred from port to the DC, including the cost of rail.

The relevant port choices in our application are for the Chicago and Houston DCs. Goods destined for Chicago are routed either through Seattle, Los Angeles or New York. For goods from China, Los Angeles and Seattle are relatively close substitutes, and New York is largely irrelevant. As we move away from China towards the Indian Ocean and get closer to the Suez canal, getting to Chicago through the East Coast, the New York option becomes relevant. Goods destined for the Houston DC are routed either directly to the port of Houston or through Los Angeles and then use rail the rest of the way. The intermodal option through Los Angeles saves time but is more expensive.

We comment on identification. The trade-off between time and money in the choice of whether to use the port of Los Angeles or the port of Houston provides information about the value of time τ . A confounding issue is that the full dollar cost of each option is not directly observed, because of the port/DC specific dollar costs x_{jk} that we treat as unobserved here. It is key for our identification strategy that different origination countries differ in freight and time costs to get to each port. This variation across source countries allows us to effectively difference out the x_{jk} , as we compare the differential choice behavior from source countries with different geography. Finally, the parameter σ_ε governs the degree of differentiation of ports that are otherwise similar in freight and time characteristic. The degree to which the firm substitutes ports (for example between Los Angeles and Seattle to get to Chicago) provides information about the σ_ε parameter.

We estimate the port choice model (2.1) for several samples: all Wal-Mart shipments, general merchandise and fashion. The samples and estimates are listed in Table 2.10. Note, it is necessary to make a normalization in each choice set, and we set x_{jk} equal to zero for when Los Angeles is used to reach the distribution center. The cost units for the x_{jk} estimates are on a container basis and the value of time estimates are on a per container per day basis. The value of time for general merchandise, at \$12 per container per day, is much lower than the value of time for fashion goods which is estimated at \$92 per container per day. For the general merchandise sample, the x_{jk} estimate for $x_{HOU,HOU}$ is \$814 less $x_{LAX,HOU}$ which represents the increased cost of using intermodal transportation from Los Angeles compared to going through the

Panama Canal. In addition to intermodal cost differences, the x_{jk} parameters also include any differences in port costs. Turning to the port estimates for shipments going to the Chicago DC, the port of New York is \$256 less costly than Los Angeles and the port of Seattle is \$1 more expensive than sending the shipment through Los Angeles. Later we can use our micro data (matched container to Census) to say a little more.

Table 2.10: Port Choice Estimates

	Value of time	σ_ε	$X_{NY,CHI} - X_{LA,CHI}$	$X_{SEA,CHI} - X_{LA,CHI}$	$X_{HOU,HOU} - X_{LA,HOU}$
All items	\$18	113	-273	-58	-852
General Merchandise	\$12	96	-256	1	-814
Fashion (Footwear and apparel)	\$92	357	-379	-889	-1370

With our parameter estimates we can use formula (2.2) to calculate the expected minimum cost to get to DC k from each country. We use these estimates as an input to the DC choice problem.

2.5.2 DC Choice Problem

Now we turn to the DC choice problem developed in subsections 2.3 and 2.4. The problem is to solve the optimal allocation of shares across DCs, for each ultimate destination. A key parameter in this problem is λ , which governs the extent the firm can customize allocation shares to individual originating countries. Based on our finding that Wal-Mart sets the same shares independent of origin, we take $\lambda = 0$ as our estimate for this parameter.

The only parameters left to be determined are the DC k specific costs y_k , and the differentiation parameter σ_η . We don't have any direct information on σ_η so for our initial analysis we treat it as equal to the differentiation parameter σ_ε for ports, and for robustness we consider alternative values.

We estimate the model for general merchandise. As before, we need to make a normalization and we set $y_{LAX} = 0$. If we take as given values for the remaining parameters, $\mathbf{y} = (y_{HOU}, y_{CHI}, y_{SAV}, y_{NOR})$, then we can plug in data on inland freight times and freight costs, as well as port costs obtained from the previous section,

to obtain the DC/ultimate location level level $\delta_{kl}^{\circ,DC}$ given in (2.3). These are used to obtain DC shares at the local level (equation (2.4)) which are then aggregated to the predicted national shares in (2.6) s_k of DC k . We estimate \mathbf{y} , by taking the observed shares and selecting \mathbf{y} to exactly fit the data.

Table 2.11 presents our estimates of \mathbf{y} for general merchandise. The baseline case is where $\sigma_\eta = \sigma_\varepsilon$, but we also consider variations where σ_η is half σ_ε and where it is twice the level. In the baseline case, y_{CHI} is \$141 more expensive than y_{LAX} and this reflects the net effect of the extra intermodal costs of getting to the Chicago DC relative to Los Angeles and the efficiency differences of the two DCs. The three east cost DCs – Norfolk, Savannah, and Houston – are all less costly compared to Los Angeles, with Norfolk and Savannah \$628 and \$918 cheaper while Houston is closer to Los Angeles at an \$86 discount. The estimates of \mathbf{y} are qualitatively similar for the other assumptions on σ_η .

Table 2.11: DC Choice Estimates, General Merchandise

	$\sigma_\eta = \sigma_\varepsilon$	$\sigma_\eta = 0.5\sigma_\varepsilon$	$\sigma_\eta = 2\sigma_\varepsilon$
Y_{CHI}	141	165	124
Y_{NOR}	-628	-614	-625
Y_{SAV}	-918	-946	-860
Y_{LAX}	0	0	0
Y_{HOU}	-86	-87	-67

2.5.3 Analysis of the Estimated Model

The Gains from a Network of DCs.

We consider how Wal-Mart gains from having a network of DCs, by evaluating its additional cost when the network is taken away.

We report estimates for general merchandise. There are normalizations we have to make, and we have to pick a benchmark to report relative costs. Our benchmark is the distribution cost of a container from China, taking the average across destinations, and using the 5-DC distribution system. The results are reported in Table 2.12. The first column reports the average distribution cost per container, relative to the cost for China, and using the 5-DC system in each case. Notice that Taiwan, as well as

countries in the South China Sea have slightly higher but similar costs compared to China. Moving west through the Indian Ocean, costs decrease relative to China with shipments from Bangladesh costing on average \$68 less, shipments from India costing on average \$367 less, and shipments from Pakistan costing on average \$601 less.

Table 2.12: Estimated Cost Per Container

By Origination and by Distribution System
(All relative to Origination from China with the 5-DC System)

Country	5-DC	3-DC	1-DC CHI	1-DC HOU	1-DC SAV	1-DC NOF	1-DC LAX
China	0	228	741	525	508	748	974
Taiwan	80	331	741	615	645	885	974
South China Sea:							
Vietnam	47	287	741	581	586	826	974
Cambodia	33	247	749	614	535	771	974
Singapore	5	210	749	587	486	722	974
Indonesia	73	304	749	645	611	847	974
Indian Ocean							
Bangladesh	-68	112	747	517	336	616	974
India	-367	-204	645	94	-60	182	974
Sri Lanka	-456	-294	575	2	-199	81	974
Pakistan	-601	-416	411	-174	-333	-103	974

Note: 3-DC excludes CHI and HOU. Cost to Los Angeles normalized to zero.

The next column simulates the effect on costs if we remove the Chicago and Houston DCs from the choice set, leaving everything else the same. The Houston DC was added in June 2005, and Chicago was added in June 2006, so our sample period 2007-2015 is just after Wal-Mart made the switch from a 3-DC system with nothing in the middle of the country. For each country we see that the costs are higher with in the 3-DC system. For China, this would raise costs by \$228 for the average container. For India, costs increase by \$164 moving from the 5-DC system to a 3-DC system. In every country the costs increase by between \$161 and \$251 per container, indicating a substantial gain from having the 5-DC system.

The last five columns simulate the costs where there is only 1 DC to choose from. For each country, any 1-DC system is more expensive than the 3-DC system. For China,

the lowest cost 1-DC system is Savannah, which is slightly less expensive than Houston. Other than for shipments from Taiwan and Vietnam, Savannah is the cheapest 1-DC in the other countries. For shipments from Taiwan and Vietnam, Houston is the cheapest, with Savannah a close second. Comparing the lowest cost 1-DC shipment to the 3-DC system, there is an incremental increase of \$281 for shipments from China, which reflects large gains for Wal-Mart of moving to a 3-DC system from a 1-DC system. For Taiwan and the South China Sea countries, the gains of a 3-DC system relative to a 1-DC system are similar in magnitude between \$276 and \$306. Savings of a 3-DC system is lower for the Indian Ocean countries with a \$224 saving per container for Bangladesh. The savings continue to decline for countries farther west, with Pakistan having a savings of \$82 of moving from a 1-DC system at Savannah to a 3-DC system.

Gains from Customization

In our benchmark model, the allocation shares across DCs for delivery to a given ultimate destination are invariant across origination countries. Since different countries vary in freight and time, the firm is better off it is can customize the mix, for each country. And if products differ in time costs, then there will be customization for each product as well as origination country.

Table 2.13 reports the results of allowing customization in both the 5-DC system and the 3-DC system. Overall, the gains from customization are slight compared to the gains of having a larger DC network. For China, the gains from customization are \$0.28 per container in the 5-DC case and only \$0.12 in the 3-DC case. There are two piece of intuition that can be offered to explain why these are low. First, shipments from China account for about 90 percent of imports, so when Wal-Mart optimizes to the average shipment, they are essentially optimizing to China. Second, recall that if we start at an optimum, and make small changes, there is no first-order effect on the objective function. Since the outcome is only slightly distorted from what would be optimal for China, the distortion is second order small.

Next note that for Taiwan and countries in the South China Sea, the geography is relatively similar to China, so the distortions are small there as well. However, as we move west though the Indian Ocean, the gains from customization increase, because the allocation shares being used are essentially optimized for China, and the optimal ones

Table 2.13: Gains from Customization

(On a Dollars per container basis from Given Country of Origin)

Country	5-DC	3-DC
China	0.28	0.12
Taiwan	3.88	5.63
South China Sea:		
Vietnam	1.68	2.31
Cambodia	0.31	0.07
Singapore	0.10	0.05
Indonesia	2.62	2.97
Indian Ocean		
Bangladesh	1.85	8.40
India	29.30	25.21
Sri Lanka	38.58	44.73
Pakistan	41.31	51.81

Note: 3-DC excludes CHI and HOU.

for Indian Ocean countries are quite different in some cases. .

Costs of Port Disruption

This subsection examines the cost of port disruption to the firm.

To model disruption, we will think of it as equivalent to a tax, a change in the port specific cost. Let x_{LAX} and x_{SEA} be the original port specific costs. Assume there is an additional cost ξ_{LAX} and ξ_{SEA} . We will consider two cases, one where $\xi_{LAX} = \xi_{SEA}$ and the second where $\xi_{SEA} = 0$ and the disruption only occurs in Los Angeles.

There was an extremely disruptive west coast port shut down in 2002. In the business press it is argued that after this experience, Wal-Mart and other firms were motivated to develop a distribution system that would be less vulnerable to labor unrest. Adding Houston as a DC is consider to be a key part of this strategy.

We use the observed port choice change during the 2015 disruption to estimate the disruption cost. During the peak of the disruption, all shipments from China destined to the Houston DC traveled though the port of Houston instead of using the port of Los Angeles. Table 2.14 shows the estimated share of general merchandise shipments from China to the Houston DC that would travel though the port of Houston under various

disruption costs. We see that with a zero disruption cost 80% of the shipments use Houston, corresponding to the average over the entire sample. As the disruption cost increases the share using Houston increases. The marginal change in share decreases with larger disruption costs as the share approaches 100%. Based on this we use $\xi = 300$ to approximate the disruption cost.

Table 2.14: Estimated Share of China to Houston DC Shipments Going Through Port of Houston for Different Disruption Costs

ξ_{LAX}	$share_{HOU,HOU}$
\$0	80.3%
\$50	87.3%
\$100	92.1%
\$150	95.2%
\$200	97.1%
\$300	99.0%
\$500	99.9%

We calculate an estimate of the dollar cost to Wal-Mart of the strike, assuming it is equivalent to a \$300 container tax, for three months of the year. We do this under several scenarios. First we look at the cost if Wal-Mart is unable to respond and uses the same ports and DCs as before the disruption. Second, the firm is free to pick the port in response to the disruption but uses the same DC structure as before. We showed that Wal-Mart can react fast like this. Third, we allow the firm to flexibly adjust both ports and DC shares. Table 2.15 shows the per container disruption costs under these three response scenarios for both the 5-DC system and the 3-DC system. For the 5-DC system, we consider a disruption to both west coast ports and a disruption to LA only (in the 3-DC system, the only west coast port is Los Angeles). In the 5-DC system, the disruption cost per container is \$115 if neither ports nor DCs can be adjusted and \$105 if Wal-Mart can only change ports, but cannot adjust the use of DCs. Multiplying this by 87,500, the approximate number of Wal-Mart containers in a 3 month period, puts the total cost of the disruption at approximately \$10.1 million in the inflexible case and \$9.2 million with flexible ports. The average disruption cost is less than the \$300 container tax for two reasons. First not all of the shipments use the west coast and second, Wal-Mart can avoid the tax by routing more of the Houston DC containers

through the port of Houston. Gaining the flexibility to adjust use of the DC network lowers the average disruption cost to \$88 and the aggregate disruption cost to \$7.7 million.

Table 2.15: Estimated Cost per Container of Labor Disruption Across Scenarios

Response of DC Share Allocations to Disruption	5-DC System (LA + SEA Disruption)	3-DC system	5-DC System (LA Disruption)
Inflexible	\$115	\$76	\$95
Flexible Ports	\$105	\$76	\$69
Perfectly Flexible	\$88	\$66	\$63

In the 3-DC system, the cost per container is \$76 both when Wal-Mart cannot adjust ports or DCs and when they can only change ports, but cannot adjust the use of DCs. In this case, the aggregate disruption cost is \$6.7 million. Note that in the 3-DC system, Wal-Mart is unable to avoid the tax by re-routing to different ports, since each of the three DCs use a single port. This is lower than the disruption cost with 5 DCs because without the Chicago and Houston DCs, fewer containers are using the west coast in the first place. Although the disruption cost is larger with the 5-DC system, the difference in disruption cost between the two systems is much less than the gain from moving to a 5-DC system, as demonstrated in table 2.12. Gaining the flexibility to adjust use of the 3-DC network lowers the average disruption cost to \$66 and the aggregate disruption cost to \$5.87 million. Importantly, adding the two more DCs to get from 3-DC to 5-DC does not lower the cost of disruption, it actually increases it. This follows because of the addition of the Chicago DC, which is sourced only by West Coast ports. The Chicago DC grabs market share from what would otherwise have gone to Savannah or Norfolk under a 3-DC system.

The third column of Table 2.15 displays the disruption costs of a tax only at the port of Los Angeles. When both ports and DCs are inflexible, the cost is \$95 per container. With flexible ports this drops to \$69 per container and with flexible ports and DCs it drops an additional \$6 to \$63 per container. Having port flexibility is nearly as good as full flexibility because goods going to Chicago can easily be rerouted through Seattle instead of Los Angeles to avoid the tax.

In summary, while we estimate the change from 3-DC to 5-DC led to substantial

decreases in the marginal expected cost of shipping an additional container, it didn't make Wal-Mart less vulnerable to a West Coast supply disruption.

Chapter 3

Understatement of Shipping Distances in the United States

3.1 Introduction

Researchers using data on shipments within the U.S. have commonly observed that there are a disproportionate amount of short distance shipments. Hillberry and Hummels (2003) document that wholesalers ship shorter distances than manufacturers and suggest that this explains why state borders appear to restrict trade. They posit that manufacturers may ship over long distances to wholesalers and then the wholesalers will serve local markets. This explanation however, does not account for the fact that some wholesalers do ship very far. Hillberry and Hummels (2008) look at shipments by manufacturers and note that these shipments also tend to be very local, which they find can be partially attributed to localized demand for intermediate goods. Under this explanation, intermediate good manufacturers strategically co-locate near the downstream demand for their products.

Holmes and Stevens (2012) find that larger plants tend to ship farther than smaller plants and acknowledge that part of this difference could be due to differences in use of the wholesale sector. If plants use wholesalers as intermediaries, the shipment distance will potentially be understated. Holmes and Stevens (2014) also note the concern that local shipments are over-represented in the CFS and that these shipments are excessive relative to local demand. They posit that transit through warehouses may account

for this discrepancy and document that approximately one quarter of ton miles are through the wholesale sector.¹ In this paper, I build on this explanation and explore *how* the presence of wholesale establishments and *wholesale networks* can contribute to the systematic understatement of shipment distances in the CFS.

This paper also builds on the literature that looks at intermediation in trade, including Bernard, Jensen, Redding, and Schott (2010).

3.2 Data

This paper relies on the 2012 Commodity Flow Survey’s (“CFS”) public use micro data. Although many previous studies have used earlier versions of the CFS (including Hillberry and Hummels (2003), Hillberry and Hummels (2003), Holmes and Stevens (2012), and Holmes and Stevens (2014)), the 2012 public use file is unique in that for the first time the public data is at the micro shipment level.² The CFS is a sample of within U.S. shipments.

Table 3.1 shows the average shipment distance in the 2012 CFS as well distances at various parts of the distribution. Looking at the full sample of the CFS the average distance, weighted by shipping weight, is 196 miles, and the median distance is 34 miles. There is a large amount of variation in shipment distances for different types of establishments. Focusing on the two largest categories of establishments in the CFS, manufacturers and wholesalers, the average distance for manufacturers is 237 miles, 43 percent farther than the 166 average distance for wholesalers. This fact is consistent with the earlier work that showed wholesale establishments ship shorter distances than manufacturers.

¹Holmes and Stevens (2014) look at diffuse demand industries to account for the Hillberry and Hummels (2008) intermediate good explanation.

²In the past shipment level data was only available in the confidential version of the CFS. Earlier public versions of the CFS were highly summarized. Although 2012 public micro data is at the lowest level of observation, in order to preserve anonymity a number of details have been removed, for example locations are listed at the MSA level, and commodities are reported at the level of 2-digit SCTG codes. While the confidential CFS has been used in previous studies, having the micro data available to the public increases access and does not require ex-post approval of results or descriptive statistics.

Table 3.1: Averages and Distribution of Shipment Distances by Establishment Industry Type

Establishment industry type	Distance Shipped					
	Mean	10th pct	25th pct	Median	75th pct	90th pct
Full sample	196	4	10	34	202	674
Direct selling establishments	16	4	6	11	18	30
Wholesale	144	4	8	25	105	456
Publishers	149	1	5	15	50	511
Warehousing	166	10	28	79	186	378
Mining	206	6	10	29	244	802
Manufacturing	237	5	11	54	288	749
Corporate	259	3	7	40	230	730
E-commerce	649	9	53	340	989	1891

Source: 2012 public use CFS. Note: Distances weighted by shipping weight. Industry type defined using the abbreviated NAICS codes in the CFS: Mining (212), Manufacturing (311-339), Wholesale (4231-4249), E-commerce (4541), Warehousing (4931 - includes 484), Publishers (5111), Direct selling establishments (45431), Corporate (551114).

Table 3.2: Averages and Distribution of Shipment Distances by NAICS Code for Wholesale and Warehousing Establishments

Wholesaler type	NAICS		Distance Shipped					
	Code	Share*	Mean	10th pct	25th pct	Median	75th pct	90th pct
Motor vehicle and parts	4231	1.6%	392	8	30	140	541	1183
Furniture and home furnishing	4232	0.4%	456	5	19	137	699	1524
Lumber and construction materials	4233	8.1%	76	4	8	16	46	166
Commercial equip.	4234	0.5%	452	12	33	181	676	1353
Metal and mineral	4235	3.7%	265	10	26	103	294	912
Electrical and electronic goods	4236	0.6%	379	5	14	71	561	1223
Hardware and plumbing	4237	0.5%	221	4	12	63	236	611
Machinery, equipment, and supplies	4238	1.6%	185	7	15	35	160	583
Miscellaneous durable goods	4239	6.7%	175	5	22	51	190	513
Paper and paper product	4241	1.0%	158	5	12	30	136	500
Drugs and druggists' sundries	4242	0.3%	440	18	56	158	570	1316
Apparel, piece goods, and notions	4243	0.2%	707	6	68	417	1041	2255
Grocery	4244	8.3%	161	4	11	41	136	370
Farm product raw material	4245	15.8%	270	9	18	63	469	814
Chemical and allied products	4246	2.6%	168	7	16	60	184	480
Petroleum and petroleum products	4247	34.3%	45	4	5	12	34	99
Beer, wine, and spirits	4248	1.4%	43	4	7	13	28	64
Miscellaneous nondurable goods	4249	5.2%	153	8	14	27	125	482
Warehousing and storage (incl 484)	4931	7.2%	166	10	28	79	186	378
All Wholesale and Warehousing		100.0%	146	4	9	28	114	444
-Excluding petroleum and farm products		49.9%	176	5	13	42	161	495

Source: 2012 public use CFS. Note: Distances and shares weighted by shipping weight. Includes abbreviated NAICS codes for Wholesale (4231-4249) and Warehousing (4931) establishments. *Share is the share of total wholesale and warehousing shipments accounted for by the NAICS code.

In addition to there being a large amount of variation in shipment distances across broad groups of establishment types, there is also significant variation within the wholesale sector. Table 3.2 shows the average distances for the specific wholesaling and warehousing NAICS codes contained in the CFS. In the public CFS, NAICS codes are reported at the 4-digit level for wholesale establishments. Petroleum and petroleum product wholesalers have by far the largest share of shipments (weighted by shipping weight), making up 34 percent of the sample. Petroleum wholesalers tend to ship much shorter distances than other wholesalers with an average distance of just 45 miles, a median shipment distance of 12 miles, and 90 percent of shipments below 99 miles. Although the total distance from oil well to refiner to the wholesale network to gasoline distributor to gas station may be many thousands of miles, shipments within this stage of the supply chain are very localized. This discrepancy highlights the challenge of taking the CFS data at face value. Within the CFS we observe single shipments that are part of a much longer supply chain. When the supply chain involves wholesale networks, and a product travels through multiple wholesalers or warehouses then using this data to estimate trade frictions may greatly understate the amount of trade that occurs over long distances.

In contrast to petroleum wholesalers, other wholesalers ship much farther distances on average. Electrical goods wholesalers, motor vehicle wholesalers, and apparel wholesalers ship 379 miles, 392 miles, and 707 miles on average respectively. The median distance for apparel wholesalers is 417 miles, with the 75th percentile distance of 1041 miles. Some of the differences across types of wholesalers can be explained by product differentiation and differences in value. Lumber wholesalers may sell less differentiated products than apparel wholesalers keeping shipment distances shorter. Shipping costs as a share of value are also lower for apparel relative to lumber, resulting in farther shipments on average. Even within apparel wholesalers there may be substantial variation, with some shipping to local markets and others selling nationally.

Table 3.3: Averages and Distribution of Shipment Distances by Metro Area for Textile Shipments from Wholesale and Warehousing Establishments

CFS origin metro area	Share*	Distance Shipped					
		Mean	10th pct	25th pct	Median	75th pct	90th pct
St. Louis-St. Charles-Farmington, MO-IL	1.4%	924	581	597	597	1709	1709
New York-Newark, NY-NJ-CT-PA	19.0%	801	7	70	460	1217	2438
Los Angeles-Long Beach, CA	10.4%	793	3	10	167	1670	2251
Boston-Worcester-Providence, MA-RI-NH-CT	3.2%	731	38	151	216	1204	1965
Greensboro-Winston-Salem-High Point, NC	1.2%	685	59	284	455	778	2123
Philadelphia-Reading-Camden, PA-NJ-DE-MD	2.5%	621	37	103	431	722	2089
Atlanta-Athens-Clarke County-Sandy Springs, GA	1.7%	584	42	127	477	708	1562
Miami-Fort Lauderdale-Port St. Lucie, FL	1.1%	515	3	6	170	898	1217
Nashville-Davidson-Murfreesboro, TN	1.5%	426	130	201	312	609	816
Charlotte-Concord, NC-SC	1.7%	414	42	91	403	606	809
Chicago-Naperville, IL-IN-WI	1.1%	373	18	43	267	587	805
San Jose-San Francisco-Oakland, CA	1.3%	323	14	20	45	68	1657
Salt Lake City-Provo-Orem, UT	2.1%	308	44	222	346	377	435
Baltimore-Columbia-Towson, MD	2.8%	300	23	40	67	333	1017
Las Vegas-Henderson, NV-AZ	5.1%	292	12	103	334	369	489
Kansas City-Overland Park-Kansas City, MO-KS	2.1%	250	56	146	173	238	397
Houston-The Woodlands, TX	2.9%	215	2	3	6	45	954
Dallas-Fort Worth, TX-OK	2.5%	200	13	19	132	271	380
Seattle-Tacoma, WA	1.1%	168	13	13	38	127	233
Non-metro areas	22.8%	628	41	162	547	715	1690
Other metro areas	12.5%						
Total	100.0%	580	9	63	335	715	1878

Source: 2012 public use CFS. Note: Distances and shares weighted by shipping weight. Includes abbreviated NAICS codes for Wholesale (4231-4249) and Warehousing (4931) establishments, excluding petroleum and farm product wholesalers (4245 and 4247). Shipments with SCTG code equal to 30. *Share is the share of textile shipments from wholesale and warehousing establishments originating in each CFS metro area. Other metro areas are CFS areas with less than 1 percent of the textile shipments from wholesale and warehousing establishments.

In order to bypass differences across types of goods, we can look at individual commodity codes. Table 3.3 looks at shipment distances by location of the shipping establishment for textile shipments from wholesaling and warehousing establishments. Here the average distance is 580 miles with the median shipment distance of 335 miles. 25 percent of shipments are less than 63 miles and 25 percent are more than 715 miles. This table reports the distances for all CFS metro areas with at least 1 percent of textile shipments. Unsurprisingly, New York and Los Angeles are the largest locations for wholesale shipments of textile products. Both of these metro areas have wholesalers that ship farther than average and have 75th and 90th percentile shipment distances that are farther than the overall 75th and 90th percentile. Why might New York and LA be different from Las Vegas? One possible explanation is that the wholesalers in New York and LA are playing a different role in the national wholesale network than wholesalers in Las Vegas. New York and LA are large port cities that both receive a large amount of imports. Wholesalers in New York and LA may have the scale to distribute nationally and may send long distances to smaller wholesalers that serve localized markets. The next section of the paper will explore this explanation with a stylized model.

3.3 Model

Consider a model of the wholesale sector where the purpose of wholesalers is to facilitate the distribution of finished goods from manufacturers to retailers. Let there be a single retailer at each location of final demand, with L retail nodes. Manufacturers of final goods are exogenously located at nodes M . Demand at each location L for each manufacturer's product is exogenous and proportional to population. Let $q_\ell(m)$ be the demand for manufactured product $m \in M$ at location $\ell \in L$.

Wholesale nodes, W , are located each location of final demand but all might not be used. W is indexed by w and let $w = 0$ be the option of not using a wholesaler (i.e. direct shipment). Each manufacturer must supply the desired quantity to location L and can choose whether to supply directly or through the wholesale network. Manufacturers seek to minimize total distribution costs and choose how many and which wholesale nodes to use. Manufacturers may send a single shipment through multiple tiers of wholesalers before reaching the retailer.

We turn now to an individual manufacturer's decision of how to distribute their product. If the manufacturer chooses to avoid wholesalers and sell directly to retailers, the total distribution costs are,

$$c_0(m) = \sum_{\ell \in L} q_\ell(m) t_{m\ell}(q_{m\ell})$$

where $t_{m\ell}(q_{m\ell})$ is the per unit cost of moving the product from the initial location at m to location ℓ , which is a function of the amount of the product shipped between nodes. As a simple form of the transportation cost function, we will let per unit transportation costs be proportional to distance with a fixed cost of linking the nodes,

$$t_{m\ell}(q_{m\ell}) = \alpha d_{m\ell} + \frac{\phi}{q_{m\ell}}.$$

If the manufacturer chooses to use a single tier wholesale network and have the option to ship to wholesalers before retailers, the total distribution costs are,

$$c_1(m) = \min_{W(m) \in \mathcal{W}} \gamma(N^{W(m)} - 1) + \sum_{\ell \in L} q_\ell(m) \min_{w \in W(m)} [t_{mw}(q_{mw}) + t_{w\ell}(q_{w\ell})],$$

where γ is the fixed cost of adding a wholesale node to the network and $N^{W(m)}$ are the number of wholesale nodes used by m .³ This nests the direct to retailer case if $W(m) = \{0\}$. The fixed cost of using a shipping route, ϕ , from w to ℓ must be paid once by the wholesaler, even if multiple manufacturers use the same wholesaler. This allows wholesalers to gain scale by aggregating shipments from many sources. The notation can be extended to allow for multiple tiers of the wholesale network – shipping first to one wholesaler who then ships to another.

A number of factors are important in the manufacturer's decision. First, if α is large relative to ϕ then firms will bypass the wholesale network and ship directly to consumers as total shipment distance matters more than the fixed cost of setting up a shipping route. Second, as ϕ gets larger than the cost of direct connections increase, firms will increasingly rely on the wholesale network in order to take advantage of the economies of scale associated with aggregating shipments from multiple manufacturers.

³Recall that wholesale node 0 is direct shipment so this node is 'free'.

As γ increases there will be fewer wholesale locations, while smaller values of γ would allow for more regional wholesalers.

One explanation for the patterns of shipments observed in the data could be that it is too costly for foreign manufacturers to ship to each location, so they send their products to wholesalers in New York or LA (implying a large cost of setting up a shipping route relative to distance cost). Wholesalers in New York and LA, then aggregate shipments from many manufactures and ship long distances to regional wholesalers throughout the country. These wholesalers then ship directly to retailers. Under this scenario, a product traveling from a manufacturer to a retailer may involve three (or more) shipments, with two of the shipments being from wholesale establishments.

In the extreme case where every product that reaches the wholesale network goes to two wholesalers the average total wholesale distance would be twice the average reported in the data. Once we consider the wholesale system as a network, potentially involving multiple shipments, the short distances in the CFS seem more reasonable.

3.4 Conclusion

Shipment distances in the CFS are disproportionately very short, with wholesale shipments shorter than shipments from other establishment types. There is however, substantial variation within the different classification of wholesale establishments. This paper looks at the distribution of shipment distances by industry using the CFS 2012 public micro data. Wholesalers in New York and LA, large cities with lots of import distribution ship farther than the average wholesaler. When looking at total distance from manufacturer to consumer, considering all layers of the wholesale network, distances would be much larger.

This chapter presents a model of the wholesale sector and this framework can be used in future work that estimates the model to match the summarized moments observed in the public CFS data.

References

- AHLFELDT, G. M., S. J. REDDING, D. M. STURM, AND N. WOLF (2015): “The economics of density: Evidence from the Berlin Wall,” *Econometrica*, 83(6), 2127–2189.
- ALLEN, T., AND C. ARKOLAKIS (2014): “Trade and the Topography of the Spatial Economy,” *The Quarterly Journal of Economics*, 129(3), 1085–1140.
- ATKIN, D., AND D. DONALDSON (2015): “Who’s Getting Globalized? The Size and Implications of Intra-national Trade Costs,” Discussion paper, manuscript.
- BECKERT, W., H. SMITH, AND Y. TAKAHASHI (2015): “Competitive Price Discrimination in a Spatially Differentiated Intermediate Goods Market,” Unpublished working paper.
- BERNARD, A. B., J. B. JENSEN, S. J. REDDING, AND P. K. SCHOTT (2007): “Firms in International Trade,” *Journal of Economic Perspectives*, 21(3), 105–130.
- (2010): “Wholesalers and Retailers in US Trade,” *American Economic Review: Papers & Proceedings*, 100, 408–413.
- BERNARD, A. B., J. B. JENSEN, AND P. K. SCHOTT (2009): “Importers, Exporters and Multinationals: A Portrait of Firms in the U.S. that Trade Goods,” in *Producer Dynamics: New Evidence from Micro Data*, ed. by J. B. J. Timothy Dunne, and M. J. Roberts, pp. 133–163. University of Chicago Press.
- BRADLEY, J. R., AND H. H. GUERRERO (2010): “The Global Replenishment Problem,” *Wiley Encyclopedia of Operations Research and Management Science*.

- COŞAR, A. K., AND P. D. FAJGELBAUM (forthcoming): “Internal Geography, International Trade, and Regional Specialization,” *American Economic Journal: Micro*.
- HILLBERRY, R., AND D. HUMMELS (2003): “Intranational home bias: some explanations,” *Review of Economics and Statistics*, 85(4), 1089–1092.
- (2008): “Trade responses to geographic frictions: A decomposition using micro-data,” *European Economic Review*, 52(3), 527–550.
- HOLMES, T. J. (2011): “The Diffusion of Wal-Mart and Economies of Density,” *Econometrica*, 79(1), 253–302.
- HOLMES, T. J., AND J. J. STEVENS (2012): “Exports, borders, distance, and plant size,” *Journal of International Economics*, 88(1), 91–103.
- (2014): “An alternative theory of the plant size distribution, with geography and intra-and international trade,” *Journal of Political Economy*, 122(2), 369–421.
- HUMMELS, D. (2007): “Transportation costs and international trade in the second era of globalization,” *The Journal of Economic Perspectives*, 21(3), 131–154.
- HUMMELS, D. L., AND G. SCHAUR (2013): “Time as a trade barrier,” *The American Economic Review*, 103(7), 2935–2959.
- LEACHMAN, R. C., AND E. T. DAVIDSON (2012): “Policy Analysis of Supply Chains for Asia–USA Containerized Imports,” *Proceedings of the 2012 Industry Studies Association Annual Conference*.
- LEACHMAN, R. C., T. PRINCE, T. R. BROWN, AND G. R. FETTY (2005): “Final report: Port and modal elasticity study,” *Prepared for Southern California Association of Governments*.
- MELITZ, M. J. (2003): “The impact of trade on intra-industry reallocations and aggregate industry productivity,” *Econometrica*, 71(6), 1695–1725.
- MILLER, N. H., AND M. OSBORNE (2014): “Spatial differentiation and price discrimination in the cement industry: evidence from a structural model,” *The RAND Journal of Economics*, 45(2), 221–247.

- OUM, T. H. (1979): “A cross sectional study of freight transport demand and rail-truck competition in Canada,” *The Bell Journal of Economics*, 10(2), 463–482.
- RYAN, S. P. (2012): “The costs of environmental regulation in a concentrated industry,” *Econometrica*, 80(3), 1019–1061.
- SALVO, A. (2010a): “Inferring market power under the threat of entry: The case of the Brazilian cement industry,” *The RAND Journal of Economics*, 41(2), 326–350.
- (2010b): “Trade flows in a spatial oligopoly: gravity fits well, but what does it explain?,” *Canadian Journal of Economics*, 43(1), 63–96.
- SCHMITZ, J. A. (2005): “What Determines Productivity? Lessons from the Dramatic Recovery of the U.S. and Canadian Iron Ore Industries Following their Early 1980s Crisis,” *Journal of Political Economy*, 113(3), 582–625.
- VELDMAN, S., AND E. H. BÜCKMANN (2003): “A model on container port competition: an application for the West European container hub-ports,” *Maritime Economics & Logistics*, 5(1), 3–22.
- VELDMAN, S., AND E. VAN DRUNEN (2011): “Measuring Competition between Ports,” in *International Handbook of Maritime Economics*.