Research Division
FEDERAL RESERVE BANK of MINNEAPOLIS

Indivisibilities in Distribution

Thomas J. Holmes
University of Minnesota
and Federal Reserve Bank of Minneapolis

Ethan Singer
Compass Lexecon

Working Paper 739
March 2017

Keywords: Indivisibilities; Scale economies; Technological change; Walmart
JEL classification: L1, F14, R40

Federal Reserve Bank of Minneapolis • 90 Hennepin Avenue • Minneapolis, MN 55480-0291
https://www.minneapolisfed.org/research/

Indivisibilities in Distribution

Thomas J. Holmes* and Ethan Singer**

March 10, 2017

*University of Minnesota and Federal Reserve Bank of Minneapolis

**Compass Lexecon

Abstract

This paper develops and estimates a model of indivisibilities in shipping and economies of scale in consolidation. It uses highly detailed data on imports where it is possible to observe the contents of individual containers. In the model, firms are able to adapt to indivisibility constraints by using consolidation strategies and by making adjustments to shipment size. The firm determines the optimal number of domestic ports to use, taking into account that adding more ports lowers inland freight cost, at the expense of a higher indivisibility cost. The estimated model is able to roughly account for Walmart's port choice behavior. The model estimates are used to evaluate how mergers or dissolutions of firms or countries, and changes in variety, affect indivisibility costs and inland freight costs.

Note: We thank Dominic Smith and Jonathan Willard for their research assistance on this project. 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.

1 Introduction

Indivisibilities arise in the distribution sector when, for example, dividing a shipment in half does not necessarily divide the cost in half. Such indivisibilities are common: an ocean container shipping half empty or a truck delivering a half-empty trailer generally ship for the same price as a full load. Indivisibilities tend to be particularly relevant when the variety of products shipped is large and volume shipped of any particularly product is small, such as when it is only a tiny fraction of the available space in a container. In such a case, consolidation of many different varieties into the same shipment can ensure a full load, though potentially coordination costs and other frictions are associated with such consolidation. Indivisibilities can be expected to have greater bite for low-value goods; if expensive goods ship in a half-full container, it matters less as a share of value. Indivisibilities tend to be relevant when firms set a high level of delivery frequency or ship to a large number of downstream distribution locations. Everything else the same, dividing shipments more finely over time or space only makes shipments smaller. Into this environment, a distribution system like that of Walmart and Target provides a means of overcoming indivisibilities. These firms use advanced information processing capabilities that help minimize consolidation frictions. With their massive sales volumes, the firms are able to consolidate a wide variety of low-value goods, with shipments finely divided over time and over space, taking care that ocean containers and delivery trucks are fully utilized.

In this paper, we develop and estimate a model of indivisibilities in shipping. We use unique, highly detailed data on container shipments to lay out a set of facts, including a fact that big retailers like Walmart and Target consolidate shipments more intensively and pack shipments fuller compared to smaller firms. In particular, we show that while intermediaries do exist to provide consolidation services to small firms, the extent of this consolidation in ocean shipping is relatively small. We also examine the geographic structure of import distribution. As discussed in Leachman and Davidson (2012), in recent years many large retailers have adopted what the logistics industry calls a "four corners" import strategy, which entails using multiple ports on both the East and West coasts to minimize inland freight costs. We present facts connecting the geographic structure of imports to the indivisibility issue. In our model, we allow firms to adapt to indivisibility constraints by both consolidation and adjustments to shipment size, as well as changing the number of destination ports. We estimate the cost effects of indivisibilities, including the magnitudes of consolidation frictions, and determine how these frictions vary with volume. To identify scale effects on frictions, we consider both seasonal effects on volume, as well as cross-location differences in volume. We find that Walmart and Target face relatively low indivisibility costs for imports from China, where they enjoy massive economies of scale. From other source countries, including India, Walmart's indivisibility cost is relatively big. We also estimate the model for a sample of small importers and find that such firms generally experience significant indivisibility costs. We use the model estimates to evaluate how mergers or dissolutions of firms or countries affect indivisibility costs and inland freight costs.

A highlight of our analysis is our data set on individual shipments, based on bills of lading filed by importers. There has been a large body of work in recent years exploiting confidential transaction-level data on imports including Bernard, Jensen, and Schott (2009), and Bernard et al. (2007, 2010). The bills of lading data set is closely related to this previous data but is different in two key respects in terms of what we do. First, the data lists container identification numbers, making it possible for us to identify different shipments being consolidated into the same containers. Our data are particularly granular for Walmart, and we are to determine how it is consolidating its products into containers at the level of the item numbers Walmart uses for its internal stock-keeping. Second, confidentiality restrictions in the previous data preclude reporting any analysis at the level of a specific firm. In contrast, with the bills of lading, we are able to conduct and report the analysis at the firm level, which is essential for our project. For our results on Walmart, we use a sample of 1.8 million containers that Walmart imported into the United States over the period 2007 through 2015. We present results for other firms as well, using a second sample covering the container imports of all companies, for a selected set of 18 months, consisting of 17.0 million container imports.

We provide a few comments about our model and how we take it to the data. We break the analysis of the firm's problem into two parts: a (short-run) shipment-level decision and a (long-run) decision about the geographic structure of import distribution. At the shipment level, the firm chooses whether to make any adjustments to a given shipment size (expanding the shipment to help fill a container or perhaps rounding down). The firm can also incur a friction to consolidate the shipment with other shipments. A key feature of the data that helps identify the magnitude of the friction is how much empty space a firm will leave in a container before choosing to consolidate it with other shipments. Our measure of indivisibility cost includes three components: (1) the frictions incurred to consolidate shipments, (2) the distortions from rounding shipment quantities up or down to match standard container sizes, and (3) the cost of empty space when neither the consolidation nor rounding up strategies fill the container. The long-run decision made by the firm is the choice of the number and locations of domestic ports to use for import flows. Here the firm faces a trade-off that, through the use of more destination ports, the inland freight cost per container is reduced, but this benefit is offset by higher indivisibility costs. We develop estimates of ocean and inland transportation costs to quantify the first benefit, and we use our estimates from the shipment-level model to quantify the indivisibility cost portion of the trade-off. We then compare the predictions of the estimated model with actual choices firms make in setting up the geography of import distribution flow.

Our work is related to several literatures. One literature concerns the economics behind the phenomenon of mass discounters, a format that has come to dominate retail in recent decades, and which has had broader impact on the economy, including the labor market (see Autor et al (2017)). Holmes (2001) develops a theory about how new information technologies complement high delivery frequency, which can be more efficiently achieved if firms can consolidate a wider variety of goods into the same shipments. That paper focused on the last stage of distribution (from regional distribution center to store shelves), while the focus here is the front end (from foreign source to import distribution center, which is where imports pass through on the way to regional distribution centers). However, indivisibility issues on the front end mirror the issues on the back end and in essence are the same. The advantage of looking at imports is that we have access to detailed shipment data collected as part of customs, data generally not available for domestic shipments. Holmes (2011) provides estimates of economies of density achieved by scale economies for distribution centers. Here there is a different mechanism underlying scale economies, one based on indivisibilities. Basker and Van (2010) provide an empirical analysis connecting the emergence of mass discounters with increased imports from developing countries such as China. Our paper complements this research by fleshing out and estimating an underlying microeconomic mechanism underlying why mass discounters have a comparative advantage at importing high-variety, low-value goods from China.

This is an extensive literature on the benefits of international trade through increasing product variety (e.g., Feenstra (1994) and Broda and Weinstein (2006)). Our paper makes the point that indivisibility costs potentially limit increases in variety, depending on the extent firms are able to consolidate efficiently. Walmart does consolidate efficiently, according to our estimates, but even Walmart faces higher costs when variety levels expand. We estimate that if Walmart's product variety doubled, holding current volume and everything else fixed, distribution costs would increase about two percent on account of indivisibilities.

Another related literature attempts to integrate the analysis of international trade with intra-regional trade. Examples include Holmes and Stevens (2014), Cosar and Fajgelbaum (2016), Atkin and Donaldson (2015), and literature on the link between transportation and the spatial organization of economic activity, surveyed in Redding and Turner (2015). In our analysis, the location of within-home-country distribution services is endogenous, depending on which ports a firm chooses to channel imports. In the data there have been reallocations of import flows across ports, including a decline in Los Angeles's import share, and our analysis sheds light on these shifts. Ports are often viewed as engines of regional growth, and local policy makers often show great interest in promoting the growth of local ports.

There is also a literature specifically on the emergence of containerization and its contribution to stimulating increases in globalization (see Bernhofen, El-Sahli, and Kneller (2016) and also Rua (2014) and Hummels (2007)). What is new here is the way we highlight the economics of indivisibilities that go hand in hand with the adoption of this technology. Before containerization, cartons were packed individually in the holds of ships, so indivisibility was less of an issue. Studies of containerization (e.g., Levinson (2006)) note that containerization created advantages directly at the port (because containers could be loaded quickly on and off ships) and advantages away from the port (because goods could stay in containers en route to final destinations). As we will explain, our estimates of very high coordination costs for consolidating small firm imports is consistent with a significant role for away-from-the-port advantages of containerization.

There is also an operations research literature that aims to assist firms in how to optimize their distribution systems.[1] We utilize previous work in this literature when we estimate transportation costs and in particular appeal to studies by Leachman (2005, 2008, 2010). Our modeling approach differs from what researchers do in this literature.

2 A Simple Example

Before getting into details of the data or theory, it is useful to work through a simple example to fix ideas. Suppose a firm imports from one source location (think of it as Shenzhen) and delivers the good to domestic locations. The home country is a circle surrounded by an ocean, as illustrated in Figure 1. Assume consumers are uniformly distributed throughout the circle.

[1] For example, the literature has developed algorithms for packing containers (see e.g., Pisinger (2002)).

Let Q be the total volume of imports from the source location, to be distributed across consumers in the home location. Think of the time frame as one year, so that Q is total annual national volume of the good.

There is an internal freight cost of τ per unit distance of inland transportation within the home country. The more ports used, the lower the average domestic distance shipped. For example, suppose there is one port, as illustrated in Figure 1(a). Assuming a circle of radius one, the average distance between the single port and every point in the circle is 1.131. The mean inland freight cost is then 1.131τ. With multiple domestic ports, the inland freight cost is minimized when the ports are spaced at equal distances. If there are two ports (Figure 1(b)), the average domestic freight cost drops to 0.750τ. With three ports (Figure 1(c)), it drops further to 0.590τ. Let $r^d(m)$ be the average inland freight cost given m ports. It is strictly decreasing and convex and in the limit goes to 0.333τ as m becomes large. The choice of m determines the *geographic structure of import flow.*

Goods are shipped in containers. Suppose the volume of one container is q°, and let the ocean freight cost per container be κ°. Note we assume that ocean freight is the same regardless of destination port, while inland freight varies proportionately with distance. In the real world, the differences between ocean freight and inland freight are not this extreme, but the abstraction is useful for the illustrative results we derive here.

To model the value of frequent deliveries, suppose that if there were no other considerations, it would be optimal to spread out deliveries continuously and uniformly over time. Let f be the frequency of deliveries in a year, and let $\frac{\varphi}{f}$ be the average penalty, per unit volume consumed, incurred by delivering goods at a rate other than the ideal of complete smoothing over time. We refer to φ as the waiting cost parameter.

Finally, suppose that for every port selected, the firm must set up an import distribution center (IDC), and let $\beta \geq 0$ be the fixed cost of setting one up.

The firm's problem is to choose a distribution structure (m, f), where m and f are integers, to minimize cost, which equals

$$C(m.f) = r^d(m)Q + \frac{\varphi}{f}Q + m\beta + nmf\kappa^{\circ}$$

for

$$n = ceil(\frac{Q}{mfq^{\circ}}). \tag{1}$$

The first term is total domestic freight costs. The second term is total waiting costs. The third term is the fixed cost that must be incurred for each IDC used. The fourth term is total ocean freight. To see this, note the volume shipped for each order is $x \equiv \frac{Q}{mf}$ (annual national volume divided by the number of deliveries in a year and the number of ports). We divide the order volume x by container capacity q° and then use the ceiling function to round up to the nearest integer. Let the fill rate be

$$fill\_rate \equiv \frac{x/n}{q^{\circ}},$$

which is loaded volume per container x/n divided by container capacity. The *empty rate* is one minus the fill rate. Because of the indivisibility constraint, in general the empty rate will be positive.

Suppose there is no indivisibility constraint, in which case the rate per unit is the same regardless of how infinitesimally small an order size was. Suppose we shut down the fixed cost β per distribution center, so $\beta = 0$. Then the optimal plan is to deliver continuously over time and continuously over space. Suppose instead there is an indivisibility. If Q is very small, the optimal solution is obviously to make one delivery per year to a single port. Moreover, with Q very small it will not fill the container, so the empty rate is positive. Next consider what happens when Q is arbitrarily large, also allowing $\beta > 0$. It is immediate that delivery frequency f and port count m must become arbitrary large and that the empty rate must go to zero. That is, with very large Q, the indivisibility becomes irrelevant.

The last point to consider is what happens when there are multiple products from the same originating location. Perhaps in an earlier environment, frictions existed making it costly to consolidate. Suppose there is a technological advance making it possible to frictionlessly consolidate goods. In this case the effect is the same as increasing volume Q for a fixed variety and we get a dense geographic footprint of domestic ports used, frequent deliveries, and low container empty rates. As we will see, this is the essence of what Walmart is doing.

3 The Data and Some Descriptive Results

This section begins by providing an overview of our data, leaving details to the data appendix. The section then establishes three sets of facts that motivate our model and empirical approach. First, we document that Walmart and Target—the two largest importers by container volume—do a substantial amount of consolidation in their import operations, and we contrast this behavior with that of small firms. Second, we show that Walmart and Target use bigger, more cost efficient container sizes than small importers and have lower empty rates. Furthermore, we make an analogous comparison within Walmart, across import source locations that vary in volume, and find qualitatively similar scale effects. Third, we examine the geography of import flow and present results that are consistent with the predictions of the simple model of the previous section, where we connect indivisibility constraints and the number of domestic ports used for imports.

3.1 Data Overview

Bills of lading are receipts issued for transactions in international trade. They are filed with the U.S. Department of Customs and Border Protection (CBP) as part of customs. The CBP records around 1 million bills of lading per month, for imports that arrive by water. The CBP sells the waterborne import data to various shipping information companies, which then resell it.[2] Our data consist of the complete set of filings for a select set of 18 months over the period 2007 to 2015.[3] For Walmart, we have extracted additional data covering all months between January 2007 through December 2015, obtaining what we estimate to be a 60 percent sample of Walmart's waterborne imports over this 9-year period. We will restrict attention to imports that arrive by container, an exclusion that mainly leaves out vehicles and bulk arrivals such as oil. (Virtually all of Walmart's waterborne imports arrive in containers.) Our 18-month sample consists of 18.8 million bills of lading, covering the arrival of 17.0 million containers. Our extended Walmart sample consists of 2.0 million Walmart bills of lading, covering the arrival of 1.8 million containers. The data appendix provides details about our samples and the extensive processing we have applied to the raw data.

The information in a bill of lading is best illustrated by examples. Table 1 is a partial list of the information from a bill of lading for a shipment of a particular type of microwave (a black, 1.1 cubic feet, digital, Hamilton Beach microwave) to Walmart that arrived in the Port of Houston on January 7, 2015. The record provides specifics such as the place of receipt (Zhongshan, which is close to Shenzhen), the foreign port (Chiwan in Shenzhen), and the vessel name. A bill of lading also specifies the shipper name and consignee. However, firms

[2] We obtained raw CBP records from Ealing Market Data Engineering.

[3] The months are November, December of 2008, 2012, 2013, 2014 (8 months); January, February, March of 2013, 2014, 2015 (9 months); December 2007, for a total of 18 months.

have the option to redact these two fields from public disclosure, and the redaction option was selected in this case. As discussed further in the data appendix, redaction of shipper and consignee information is a major limitation of this data. Nevertheless, in the "Marks" field (which cannot be redacted), we can easily see the shipment is to Walmart. Also, in the products field, we see the text pattern "GLN: 0078742000008," which is a marker for Walmart. (A search on this GLN code is the source of the vast majority of the records in our Walmart sample.) There are eight containers in the shipment, and the 11-digit international container identification code is listed for each of the eight containers.[4] The record specifies the *piece count* for each container, which in this case is 640 microwaves in each container. The product field reports various details about the shipment, including the 9-digit *item number* (Walmart's internal stock-keeping number), and the 10-digit HS product code used for customs reporting.

It is useful while going through this microwave example to briefly digress on the topic of freight rates for containers and the value of the goods inside. As discussed further in the appendix, over our sample period it generally costs on average in the range of \$2,500 to \$3,000 to ship a container from Asia to a U.S. port. In our microwave example with 640 units in a container, this works out to an ocean freight cost of about \$4 per unit. We use public U.S. Census Bureau tabulations on imports from China in 2015 for this type of microwave to estimate that the wholesale cost (including freight to a U.S. port) is approximately \$42 per unit. (The wholesale cost of an entire container load would be $\$26,880$.) Thus for this good, the freight cost of delivering the container to a U.S. port represents about 10 percent of the wholesale cost of getting there. For more valuable goods, such as footwear and electronics, the share would be lower, say 5 percent. Below we assume the share is 8 percent, and we say more about this in the data appendix.

This particular microwave happens to be the highest volume product for 2015, at the item number level, across products imported by Walmart in 2015 over water. Walmart imported 828 containers of this product in 2015, and virtually all of them were stuffed with exactly 640 units, adding up to over half a million microwaves.[5] Delivery frequency in 2015 for this item averaged twice a month ($f = 24$ in the notation of the simple model), and Walmart uses 5 import distribution centers ($m = 5$), so the average number of containers in one shipment is approximately $n = \frac{828}{fm} = 7$. For this particular high volume product, the

[4] The registry is the Bureau International des Containers, which determines a *BIC Code* for all containers used in international trade.

[5] There are 12 containers packed with 610 microwaves.

indivisibility constraint limiting shipments to integers such as 6, 7, or 8 containers is unlikely to be a big issue. However, it is the exceptional case of the largest volume good.

In Table 2, we turn to more typical cases where shipment volumes for particular products are significantly less than what would fill a container. In these cases, Walmart typically consolidates different products within the same container. In the two examples in Table 2, each distinct product (at the 9-digit item level) is given its own bill of lading. Henceforth, we equate the terms *shipment* and *bill of lading*. Panel A is an example where a single shipper (Buzz Bee Toys) accomplishes the consolidation, combining four different products that it sells to Walmart. Panel B is an example where consolidation takes place across five different products from five different firms. These shipments arrived in 2007, a year when Walmart generally did not invoke the redaction option, and so for these records we observe shipper information.

The second to last column of Table 2 specifies the shipment volume, measured in cubic meters (cbm). The total volume of the shipments in the first container is 54 cbm and in the second is 69 cbm. We say more about container sizes below but mention here that the first total volume is just below the practical carrying capacity of a standard 40-foot container, and the second volume indicates use of the slightly larger "high-cube" container, which Walmart commonly uses. For both cases, the combined weight of the goods shipped sums to around 5,000 kg, which is about a fifth of the maximum capacity by weight. For imports of consumer goods from China, the relevant capacity limitation in a container is virtually always volume, not weight, so when we refer to the *fill rate*, volume will be the relevant measure.

The place of receipt variable tells us the source location where a container was packed. Panel A of Table 3 provides counts of shipments and containers for the 9-year Walmart sample, including counts from China as source location, and more narrowly from Shenzhen, China. Over the 9-year sample, 87 percent of Walmart's shipments (1.7 out of 2.0 million) originated from China. If we calculate the China share at the container level rather than the shipment level, the share is approximately the same, 86 percent (1.6 out of 1.8 million containers). Over half of the containers from China originated in Shenzhen.

Walmart and other large importers negotiate shipping contracts directly with shipping companies. For such transactions, the consignee in the bill of lading is referred to as the *Beneficial Cargo Owner* (BCO). In contrast, smaller importers generally work with an intermediary called a *freight forwarder*, who then negotiates with the shipping company. In these cases there are two shipping records for a given shipment, the *Master Bill of Lading*, covering the contract between the shipping company and the freight forwarder, and the *House Bill of Lading*, covering the contract between the freight forwarder and the ultimate consignee. In our analysis, we separate out the master bills of lading to avoid double counting. We will refer to shipments with house records as *Freight Forwarder (FF) Intermediated*. The last line of Table 3 reports that in the 18-month sample, just over half of all shipments are FF intermediated (7.3 out of 14.0 million), and these account for 38 percent of containers (6.5 out of 17.0 million). Overall, China accounts for 44 percent of container imports (7.4 out of 17.0 million), and Shenzhen is the source of 27 percent of Chinese imports (2.0 out of 7.4). Given the enormous role China plays in container imports, much of our analysis will focus specifically on imports from China.

3.2 Evidence on Consolidation

We define a shipment as *consolidated* if any container listed on the shipment record is also referenced by some other shipment arriving at the same time. We group shipments linked by shared containers and call any such combination a *consolidated shipment group*. For example, if shipment 1 is linked to shipment 2 through shared container A, while shipment 2 is linked to shipment 3 through shared container B, then shipments 1, 2, and 3 are all part of the same group. The shipments in Panel A of Table 2 are all in one consolidated shipment group, and the shipments in Panel B are in another group. We refer to an individual shipment such as the microwave example in Table 1 as an *unconsolidated shipment*.

We discuss consolidation by Walmart first and then turn to other firms. For the shipments illustrated in Table 2, there are no overlapping products at the 9-digit item level across shipments in the same group. The absence of such overlap is typical in the Walmart data. We estimate that 94 percent of consolidated shipments have zero overlap with any products found in other shipments of the same consolidated group. (The data appendix provides details.) Thus, in our Walmart data, when we see multiple shipment records for products in the same containers, it generally represents true consolidation of different product varieties rather than just additional paperwork.

To estimate the extent to which consolidation is taking place across different Walmart suppliers (as in Panel B of Table 2), as opposed to within a single supplier (Panel A), we utilize the observations in the first 15 months of our sample period. This subset of the data is useful because over this period, Walmart generally did not invoke the option to redact shipper and consignee information. For this sample, we estimate 34 percent of consolidated shipments, on a container-weighted basis, aggregate products from two or more distinct suppliers. We also find that 43 percent of the cross-supplier consolidation records explicitly reference a logistics firm providing consolidation services (usually Maersk Logistics), while only 12 percent of within-supplier consolidation records do this. We expect that outsourced logistics firms provide the cross-supplier consolidation services, even when not listed on the record. Such consolidation generally takes place at container yards at ports. Within-supplier consolidation can be expected to take place at source factories.

In the examples from Tables 1 and 2, each distinct shipment record lists only a single Walmart item number. While this is typical, cases where a single shipment lists multiple products do occur. We calculate that among unconsolidated shipments, 20 percent list multiple product items. Among consolidated shipments, 34 percent list multiple products. Thus, measuring consolidation across shipment records masks additional consolidation occurring within a shipment record. Despite this undercount, for our main results we will focus on consolidation measured through the shipment record information rather than the product information. The data are much cleaner to work with at the shipment level, as we are missing product information for about a third of the shipments.

In Table 4 we report the share of containers in consolidated shipments, and for unconsolidated shipments, we distinguish single container versus multi-container shipments. Panel A uses our 9-year Walmart sample to report on imports from China, as well as from the next four highest volume source countries. Consolidated shipments account for 42.0 percent of all Walmart's imports from China. Unconsolidated shipments are mainly multi-container (49.9 percent), and the remaining percent in single containers is small (8.1 percent). The next largest source country is Bangladesh. (Note the remarkable disparity in container volume between first and second highest: 1.57 million from China, 0.03 million from Bangladesh.) The consolidation rate for Bangladesh is quite high, 75.3 percent. Bangladesh specializes in clothing, a product segment where product variety is important, and this magnifies the bite of the indivisibility issue. India's rate is a little less than China's. Below when we estimate the model, we will have more to say about Bangladesh and India.

Figure 2 illustrates how consolidation has changed over time for Walmart. (The plot is for Walmart imports from China, but the pattern excluding China is similar.) There is a clear monotonic pattern of increase over time, rising from 35.1 percent in 2007, all the way to 49.6 percent in 2015. Below we will examine this pattern through the lens of the model and account for the pattern through a decrease in consolidation frictions over time, combined with a decrease in average shipment volume per product.

We process the 18-month complete sample to derive information for additional firms. For BCO records, we begin with a list of large companies and take various processing steps to identify the records for these companies. Panel B in Table 4 lists the six retailers for which we found 40,000 or more container imports from China in the sample. Walmart is the largest, of course. The consolidation rate of 46.2 percent in this sample is a little higher than the 42.0 percent from the 9-year sample, which not surprising given the upward trend in the rate and the fact that the 18-month sample is weighted toward more recent months. Next consider Target. It is striking how similar Target is to Walmart in its consolidation behavior, and even in the balance between unconsolidated single and multi-container shipments. Later, when we estimate the model and incorporate additional aspects of the data beyond Panel B, we will find that the estimated parameters for Target are remarkably similar to those for Walmart. Next, skip a few rows to Costco. The measured consolidation rate is zero. To understand why Costco is so different from Walmart or Target, we need to recognize the fundamentally different business model used by Costco. In the Costco format, there is very little product variety, and very high volumes per product, and therefore less of a need for consolidation. Now look at K-Mart, which is similar to Walmart and Target in its type of business, but the reported consolidation share is only 10.2 percent. From inspection of the records, it appears very common for K-Mart to consolidate different products into the same shipment record, and this reporting practice likely accounts for much of the discrepancy between what we find for K-Mart and our results for Walmart and Target. Finally, the hardware/building supply giants Lowes and Home Depot have very low consolidation rates both because they often sell bulky items such as patio sets that need no consolidation and because, like K-Mart, they appear to often consolidate multiple products into a single shipment record. For this data reason, in our empirical analysis of large retailers below, we will focus on Walmart, Target, and Costco.

We next turn to FF-Intermediated shipment records. One fortunate thing about these records is that the option to redact consignee is generally not used. (BCO records are quite different. All the BCO retailers in Panel B, except Costco, generally redact.) We process the consignee information to pull out address information and then link shipment records by consignee name and address, obtaining 380,176 unique consignees. (We continue to focus on imports from China.) We then classify each consignee by its total count of different consolidated shipment groups or unconsolidated shipments. Call this number the consignee's group count (think of an unconsolidated shipment as one shipment group). We will treat this count as a measure of a firm's size as an importer. For FF-Intermediated shipments, we define a shipment as consolidated if and only if the shipment shares a container with a shipment of at least one different consignee. Note that in Panels A and B of Table 4, where we are focused on BCOs, the consignee is the same (e.g. Walmart) across shipments in a group. Different products from the same or different suppliers are being aggregated, but all of the products are being shipped to the same consignee, with the ocean container going all the way to the door of the consignee's distribution center in the United States. In contrast, the consolidations in Panel C are what is known as "Less than Container Load," or LCL shipments in the trade. Freight forwarders consolidate LCL goods for different consignees into containers at foreign ports, and then after arrival at U.S. ports unpack the ocean containers and forward the individual LCL shipments to the ultimate consignees. In other words, LCL shipments miss out on the "Full Container Load" or FCL benefit of a locked, packed container delivered all the way to the door of the consignee.

The first point to make is that the share of containers that are consolidated is quite low, equaling only 4.8 percent. Again we require that there be two distinct consignees in the same container to fit the definition of consolidated, but even if we weaken the definition to only require two different shipments (and not necessarily different consignees), the overall rate is only 8.9 percent. Thus, we see that consolidation of LCL loads for different consignees is small compared to the consolidation of different products that Walmart and Target are doing. The share reported in the table is on a weighted container basis. The unweighted shipment share (not reported) that is consolidated is higher, of course, and equals 36.8 percent.[6] However, the analogous statistics for Walmart and Target are also higher (84 percent in both cases), and the wide difference persists. Our point that the LCL market is relatively small may be surprising to some because there are many well-known companies that offer LCL service, such as DHL Global Forwarding. However, DHL's public statistics indicate that LCL container volume is only on the order of a 2.5 percent share of the firm's total volume, consistent with the low share we find here.[7]

Next, examine the pattern across consignees of different size. At the bottom of Panel C, we use the number of shipment groups a consignee received to classify the consignee into one of six size categories. The very smallest consignees have the highest consolidation rate, but it is still only 9.0. The rate sharply decreases as size increases, falling to only 1.4 percent of containers for the largest consignees. Note also that larger consignees are substantially

[6] Unweighted shipment shares are reported by consignee size class in Table 9 below.

[7] See DHL, "Ports of the World" (undated), which reports FCL shipments of 1.375 million 40-foot equivalents (FEU) and LCL of 2.0 million cbm, which we convert to FEU following Table 5 (58 cbm = 1 FEU).

more likely to ship multi-container, unconsolidated shipments, compared to single container shipments.

3.3 Evidence on Container Size and Empty Rates

In this subsection, we present evidence on container sizes and empty rates. Before getting to the results, it is useful to provide background information about containers. There is some variation in container size, and Table 5 lists the main choices. A standard 40-foot container has a rated volume of 67.7 cbm, but the practical volume is considered to be about 58 cbm.[8] The half-size 20-foot container has a practical volume of 28 cbm, a little less than half of the full size. The price discount on the half size is at most only about 25 percent compared to the full size (see the data appendix for more about pricing), so price per usable volume shipped is on the order of 50 or more percent higher (=0.75/0.50) when a half size is used. There is also 40-foot version that is one foot taller than the standard container, called a *high cube* with 68 cbm of practical space. There can be a cost advantage to using a high cube instead of a standard container. However, the difference is small relative to the difference between the standard and half size. Also, while in our data we can separate out half-size containers, we cannot always tell whether a given 40-foot container is standard or the high-cube variant. For this reason, in our analysis we will generally lump together the standard and high-cube sizes and refer to both as *full-size* containers.

We use the term *fill amount* to denote the volume of goods contained in a particular imported container. We take the 18-month sample and consider the subsample for which we have volume of the contents. In Figure 3(a), we start with a histogram of fill amounts of container imports originating in China, derived from our 18-month complete sample. Notice that there is a concentration of mass just below 28 cbm (the capacity of the half size), another just below 58 (the capacity of the standard size), and another below 68 (the capacity of a high cube).[9] Notice that while there is mass near fill amounts consistent with full containers, there is also mass in size levels with significant empty space, such as at 40 cbm, a point where a container would be about a third empty.

Figure 3(b) is constructed the same way, except we use Walmart's imports from China. Note the dramatic difference in fill levels. There is only a hint of any use of half-size

[8] See *Cargo From China* in the references for a table listing practical volumes.

[9] We can see some even higher than 68. This is a combination of (1) there is yet another size, 45 foot; (2) in certain cases, firms may be able to pack goods tighter and get closer to the theoretical maximum capacity; (3) there is measurement error.

containers. There is virtually no probability weight around the 40 cbm range. We only start seeing mass at fill levels above 50 cbm. There are two peaks above 50, one corresponding to standard containers and the higher one to high-cube containers. Figure 3(c) in the series depicts Walmart's originations from India. India is the third largest source country. Nevertheless, the volume obtained from India is tiny compared to the volume from China. Also, unlike Bangladesh, where virtually all containers are packed in one place (the port of Chittagong), in India sources are spread throughout the Indian subcontinent, which limits the ability to consolidate. We can see in the figure that, for India, Walmart makes some use of half-size containers. Also, sometimes full-size containers go out with only 40 cbm.

Table 6 summarizes how much empty space there is in the various data samples considered in Table 4. The first column reports for each sample the share of containers that are half size. We have emphasized so far that for imports from China, containers tend to hit the volume limit ("cube out") before hitting the weight limit ("weigh out"). However, for dense goods such as cement, the weight limits may hit first, and for these goods it can be economical to use the half size. We have found that a good way to separate dense goods is to pull out shipments that list two or more half-size containers.[10] In the next column of Table 6, we report half-size container shares, after these dense goods are pulled out. Either way, the first point to make is that Walmart's use of half-size containers is miniscule, equal to 0.6 percent for our preferred statistic. Note this is also true for Bangladesh, which is able to achieve high levels of consolidation (recall Table 4). In contrast, Walmart makes some use of half sizes out of India (4.3 percent) and the other source countries. Next look at other large retailers out of China. Both Target and K-Mart are similar to Walmart. Next consider Costco. Since Costco is not consolidating, it sometimes needs to use the half sizes, doing so at the relatively high rate of 7.2 percent. Finally, the bottom part of the table shows how smaller importers that use freight forwarders are behaving. In the smallest size category, a third of all containers are half size.[11] The half-size rate decreases monotonically with our measure of size, falling to only 7.6 percent for the largest size category, which is about the same level as Costco.

Next we define fill and empty rates. We exclude consolidated shipments and condition on whether a half or full size is used. Let cbm_i be the cbm per container for shipment i.

[10] If the weight limit is not binding, it would be much cheaper to use one full size rather than two half sizes. We find that shipments using more than one half-size container are typically heavy and would be above the weight limit if the contents were doubled and put in full-size containers.

[11] For the statistics in Table 6 for FF-Intermediated firms, we restrict the sample to shipments that are not consolidated.

We define the empty rate for full sizes to be

$$empty\_rate_i^{full} = 1 - \frac{\min(cbm_i, 58)}{58},$$

where, again, 58 is the practical capacity of a standard container. The empty rate for half sizes is analogous, using 28 as the cutoff. Empty rates are reported in Table 6 for the various samples. For Walmart goods coming out of China and Bangladesh in half size containers, the empty rate is almost 20 percent. But recall it is very rare to use half sizes, and when it happens it is probably some unusual circumstance where consolidation is impossible. Nevertheless, note the empty rates out of China or Bangladesh are 20 points lower than for the other source countries. Looking farther down the table at the other samples, we can see lots of empty space in the half-size containers throughout all samples. In particular, half-size containers ship a quarter empty, on average. This is consistent with standard advice that if a shipment fills at least 50 percent of a half-size container, it is cheaper to send it unconsolidated rather than as a LCL shipment.[12] The last column presents empty rates for full sizes. For Walmart out of China or Bangladesh, the empty rate is only about 1.7 percent. Empty rates out of the other countries are higher by a factor of two or three. Looking at the bottom for FF-Intermediated out of China, we see for the small size class that the empty rate is 6.6 percent, which is four times higher than the empty rate out of China for Walmart.

Finally, in Figure 4 we plot the empty rate over time for Walmart out of China. While noisy, there is a pattern of a decrease from around 1.9 percent at the beginning of the sample to 1.4 percent at the end. This trend is consistent with the pattern of increasing consolidation reported in Figure 2.

3.4 Geography of Import Flow

In this last subsection, we present facts about the geography of import flow. Since 2000, discount retailers have been using four-corner import strategies, as noted in the introduction. Currently Walmart bring imports through five IDCs: Los Angeles, Chicago, Houston, Savannah, and Norfolk (MWPVL (2017)). Shipments are roughly evenly divided across the five IDCs. Specifically, in the appendix we report the following estimated shares: 0.20, 0.16, 0.23, 0.23, 0.18. Target has four IDCs (Los Angeles, Seattle, Savannah, and Norfolk). As

[12] See *Cargo From China*.

illustrated in the simple model, given indivisibilities, four corner strategies are more likely to be selected when volume is greater. This will also be the case if improvements in information technology allow firms to more efficiently consolidate goods. In the 1990s, Walmart's imports from China were relatively small, and its information technologies were less developed. During that period, Walmart imported goods through one import distribution center in Savannah. Walmart added Los Angeles and Norfolk in 2000. In 2005 and 2006, it added Houston and Chicago. This is a time period in the U.S. economy when imports from China began to explode, and Walmart was an important contributor to this trend. We note that Walmart is currently in the process of adding another IDC in Mobile, Alabama, filling the gap between Houston and Savannah.

The change over time in the geography of import flow can be seen in the published trade statistics. We selected the top 10 broad product categories (at the two-digit HS code) in our Walmart sample. (These account for 88 percent of Walmart container inputs.) For each product, we calculate the share of the good being imported through each port, take the Herfindahl index, and plot this in Figure 5. (We use both value and weight to calculate shares, and it does not make a difference.) We see in the figure that the Herfindahl index decreases substantially over the period, from 0.50 to 0.31. The emergence of four-corner strategies for importing goods leads to a spreading similar to what we see in the data. Whereas in the past, Los Angeles would be the primary port and a company might bring everything through there, the adoption of four-corner strategies has led to a decline in the Los Angeles share. Over this period, the Los Angeles share has declined on average for these goods from 70 percent to 52 percent, and ports such as Houston have grown. We also note that the real value of imports from China for these 10 product segments has increased by a factor of 2.3 in the 15-year period 2000-2015.

Costco actually uses 11 import distribution centers, more than double what Walmart currently has. This may seem surprising, since Costco has much smaller volume overall compared to Walmart. However, the key thing to note again is the difference in the business model, as Costco has substantially fewer varieties and, in particular, does not face the issue of dealing with very small volumes of particular goods. In this way, the indivisibility constraint bites less for Costco, and it chooses finer geographic resolution in its import strategy. We discuss this case further below in the context of the model.

4 Theory

Consider a firm that imports goods from foreign destinations. We model two aspects of the firm's decision making. The first is a long-run decision about the structure of import distribution. This aspect is broad, including the choice of how many different domestic ports to use in the supply chain, as well as delivery frequency. The result of this decision determines the rate at which a particular-sized shipment will appear that will need to be sent from a particular source location to a particular destination port. The arrival of a particular shipment is the occasion for the second aspect of decision making. At each such point, the firm has an opportunity to adjust the size of the particular shipment, up or down, to address indivisibilities in shipping. The firm also has the option to consolidate the newly arriving shipment with other shipments.

Formally, let the various decisions about the broad structure be indexed by $s \in S$, where S is a set of possible decisions. Take as given that there are J different source locations, indexed by j. Also, we leave exogenous Q_j, a fixed volume of imports from j. Let $C_j^{ocean}(s)$ be the expected ocean freight cost per unit volume from source j to the shipment's destination import distribution center, given distribution structure s. The expected cost given s is

$$C(s) = \sum_{j=1}^{J} Q_j C_j^{ocean}(s) + C^{rest}(s),$$

where $C^{rest}(s)$ includes remaining aspects of the firm's distribution costs, excluding ocean freight. In particular, $C^{rest}(s)$ takes into account the inland transportation costs of shipping the good into the interior of the home country. It also includes returns related to the delivery frequency. Let s^* be the optimal distribution structure minimizing $C(s)$.

We can illustrate the notation in terms of the simple example presented in Section 2. There, the choice of structure is $s = (m, f)$, for an integer count m of domestic ports used, and an integer count f of annual delivery frequency per port. Expected ocean freight is

$$C_j^{ocean}(m.f) = n_j m f \kappa^{\circ},$$

where n_j is the count of containers sent per delivery (equation (1)) and κ° is the freight cost per container. The remaining distribution cost is

$$C^{rest}(m.f) = r^d(m)Q + \frac{\varphi}{f}Q + m\beta,$$

the sum of inland freight, the waiting cost, and the IDC fixed cost.

In the simple model, a shipment arrives, and the only thing that needs to be done is to count the number of containers needed to hold the shipment. We now develop a richer model of decision making at the shipment level.

4.1 The Shipment-Level Decision

The choice of distribution structure s determines the arrival distribution of shipments that will need to go out. Let i indicate a particular shipment originating at j with destination k, and let x_{ijk} denote an initial targeted volume of the shipment, which is determined by issues separate from indivisibilities. Let $G_{jk}(x|s)$ be the cumulative distribution function (c.d.f.) of the target volume x, for goods from i to j, given distribution system s. To see why the c.d.f. G_{jk} would depend on s, observe that if the structure s specifies a high level of delivery frequency and a large number of domestic ports, then we would expect to see smaller values of x, as shipments are more finely divided up over time and space.

The firm chooses whether to send a given shipment *consolidated* or *unconsolidated.* For simplicity, we assume costs for consolidated shipments are proportional to volume according to the following specification:

$$c_{con}(x|s) = (1 + \eta_s)\lambda x. \tag{2}$$

The parameter λ has the interpretation of what the shipping cost would be in an ideal world with no indivisibilities. We will relate λ to container prices below. The parameter $\eta_s \geq 0$ is the *consolidation friction*, and we intend it to incorporate a variety of different frictions. It includes any distortions in the timing of shipments that are incurred as part of consolidation. The friction may be geographic in nature, when it is necessary to combine different goods from different originating factories or goods meant for different downstream destination warehouses. The parameter also includes the coordination cost. If there are computer advances that make it easier to keep track of and coordinate processing across different products, then this would lower η_s. Finally, note that we allow the friction η_s to depend on the distribution structure s. We expect there to be economies of scale in consolidation. If a distribution system is chosen with a large number of destination IDCs, this reduces volume to any particular IDC, and we expect that consolidation will be more difficult. Below we will explicitly parameterize how the friction depends upon volume.

If the firm chooses not to consolidate, then indivisibility issues must be confronted. In this case, we allow the firm to make an adjustment to the order size as a way to minimize indivisibility costs. In particular, it can round up to fill up empty space or round down to eliminate a partially filled container. Let y be shipment volume after adjustment. Assume a change in the shipment size to y from x results in a net benefit to the firm (excluding freight cost) equal to

$$b(y,x,\phi) = \alpha(y-x) - \phi x \left(\frac{y-x}{x}\right)^2. \tag{3}$$

The parameter α specifies a linear shadow benefit of the additional volume squeezed into a container (if $y > x$) or the lost benefit from a smaller order (if $y < x$). The last term is a quadratic adjustment cost from distorting the volume choice y from the initial target. Note the cost of a given percentage deviation increases proportionately with the initial target size x. The parameter ϕ governs the magnitude of adjustment costs. We assume ϕ is a random variable drawn from a discrete distribution $\phi \in \{\phi_1, \phi_2, ...\phi_L\}$ and let the probability the firm draws ϕ_ℓ be ω_k.

We allow for two container types: type 1 (half size) and type 2 (full size), with capacities $q^1 = \frac{1}{2}q^2$. Assume the freight charge for the half size satisfies $\frac{\kappa^2}{2} < \kappa^1 < \kappa^2$, so the full size is cheaper per unit volume when shipped full. We now define the parameter λ introduced earlier as cost per unit capacity of a full-size container,

$$\lambda \equiv \frac{\kappa_2}{q^2}.$$

In our estimation, we also allow an additional charge $\kappa^{mix} \leq (\kappa^2 - \kappa^1)$ if the firm sends a shipment with a mix of type 1 and type 2 containers. Let $\tilde{n}^{only1}(y) = ceil(\frac{y}{q^1})$ be the number of half-size containers that would be needed to ship the entire load of volume y. The firm would always use one full size instead of two half sizes, so the count of containers of each type given y is

$$\begin{aligned} \tilde{n}^2(y) &= floor(\frac{\tilde{n}^{only1}(y)}{2}) \\ \tilde{n}^1(y) &= \tilde{n}^{only1}(y) - 2\tilde{n}^2(y). \end{aligned} \tag{4}$$

The freight cost to ship y is

$$c_{un}(y) = \kappa^1 \tilde{n}^1(y) + \kappa^2 \tilde{n}^2(y) + \kappa^{mix} 1_{\{\tilde{n}^1(y)>0 \text{ and } \tilde{n}^2(y)>0\}}.$$

The firm chooses y after observing the realization of adjustment cost ϕ_ℓ. The optimum adjustment y given the realization of x and ϕ_ℓ solves

$$v(x, \phi_\ell) = \max_{y>0} \alpha y - \phi_\ell x \left(\frac{y-x}{x}\right)^2 - c_{un}(y).$$

Note we require that $y > 0$, that is, there is no option for the firm to round down and simply not have any shipment go out.

We incorporate one last ingredient to the shipment-level decision, which adds shocks ε_{con} and ε_{un} to the firm's profit conditioned on whether or not the firm consolidates. Assume these random shocks are drawn i.i.d. from the type 1 extreme value distribution with standard deviation $x\zeta$. We make the distribution proportionate to the initial target shipment size x to ensure that our setup has constant returns throughout, except what happens through the indivisibility issues.

In summary, given the random realization of the initial target size x, the adjustment cost parameter ϕ_ℓ, and the shocks $\boldsymbol{\varepsilon} = (\varepsilon_{con}, \varepsilon_{un})$, the firm chooses whether or not to consolidate to maximize

$$v^*(x, \phi_\ell, \varepsilon) = \max \left\{-c_{con}(x) + \varepsilon_{con}, v(x, \phi_\ell) + \varepsilon_{un}\right\},$$

where dependence on the distribution structure s is left implicitly. The expected value given x, integrating over the shocks $\boldsymbol{\varepsilon}$ and the draws of ϕ_ℓ, equals

$$V^{ind}(x) = \sum_{\ell=1}^{L} \omega_\ell \zeta x \ln \left(\exp\left(\frac{-c_{con}(x)}{\zeta x}\right) + \exp\left(\frac{v(x, \phi_\ell)}{\zeta x}\right)\right). \tag{5}$$

To understand how this model works, suppose that if the firm were to ship the original target value x, there would be empty space in a full container. Suppose we solve the first-order condition for the adjusted level y to maximize the benefit (3) at zero marginal cost of additional freight. In this case, the optimum would adjust the shipment size upward at a rate

$$\gamma = \frac{\alpha}{2\phi}$$

yielding a choice

$$y = \min\left\{x\left(1+\gamma\right), \tilde{n}^2(x)q^2\right\},$$

which takes into account that the solution could be at the corner where the container is filled. It is easy to see that as x becomes very large, the percentage increase needed to round up volume to fill up empty space gets small. That is, for large x we will necessarily be at the corner. Note that we will not necessarily round up: there is also the option to round down. Either way, for large x, the firm will choose to adjust the shipment y so that the firm ships only full-size containers that are exactly filled. In short, the indivisibility issue becomes inconsequential when shipment sizes are large.

5 Estimating the Shipment-Decision Model

In this section we produce estimates of the shipment-decision model. The model is a data-generating process for observations on shipments. For each shipment i, let yes_i^{con} be an indicator variable for whether or not the shipment is consolidated. Let z_i be the volume of a particular shipment. In the model, if $yes_i^{con} = 1$, then $z_i = x_i$; that is, the shipment that goes out is the initial target level. If instead $yes_i^{con} = 0$, then $z_i = y_i$; that is, the shipment going out is some adjustment from the target level. In cases where $yes_i^{con} = 0$, the count of containers of each type, n_i^1 and n_i^2, is determined by (4). A complete description of observation i is $\{yes_i^{con}, z_i, n_i^1, n_i^2\}$. For constructing moments, it will also be useful to define

$$\tilde{z}_i \equiv 1_{\left\{yes_i^{con}=1\right\}} z_i + 1_{\left\{yes_i^{con}=0\right\}}\left(n_1 q_1 + n_2 q_2\right),$$

which equals shipment volume in the event of consolidation and otherwise, if unconsolidated, equals the volume of the containers used.[13]

We make parameter restrictions to simplify estimation. First, for a given origin j and destination k, we assume the distribution $G_{jk}(x)$ of the initial target shipment level is log-normal, with parameters μ_{jk} and σ_{jk}^2. To motivate this assumption, consider Figure 6(a), which is a histogram of log shipment volume ($\tilde{z}_i$) for Walmart out of Shenzhen, including both consolidated and unconsolidated shipments. We normalize the volume of one full-size

[13] Under the assumption that there is no empty space in consolidated shipments, then the sum across $\tilde{z}_i$ will add up to all the containers used in the data. In the data appendix we explain how we use piece-count information for each shipment and container to estimate $\tilde{z}_i$ for all observations. We observe z_i for some observations, but not all.

container to be one, $q^2 = 1$. According to the model, we see the initial target x_i only in cases where $yes_i^{con} = 1$; otherwise we see y_i, the volume after adjustments to respond to the indivisibility. In the figure, unconsolidated shipments are at the mass points ln(1), ln(2), and so on. Excluding these mass points, the distribution of z_i appears to be approximately log normal and motivates our choice of functional form.[14]

Second, we assume for our baseline estimates that the cost for a half-size container is 75 percent of that of a full size, $\kappa^1 = 0.75\kappa^2$. The data appendix discusses evidence on container pricing motivating this assumption. We normalize $\kappa^2 = 1$, so dollar units are in terms of the price of shipping one full-size container.

Third, we assume

$$\alpha = \lambda \equiv \frac{\kappa^2}{q^2} = 1, \tag{6}$$

that is, the marginal value of one unit of empty space in a container exactly equals price per unit volume in an idealized world with no indivisibility issues, normalized to one,. Our motivation is that if we were to step back and consider the full problem, where the firm picks the distribution of arrival sizes of x, and if in that problem indivisibilities were negligible, then (6) must approximately hold.

Fourth, we assume a two-point distribution for the adjustment cost ϕ_ℓ, with $\phi_2 = \infty$ and $\phi_1 = 2\alpha$. This implies growth for $\ell = 1$ is $\gamma_1 = 0.25$ with probability ω_1 and growth for $\ell = 2$ is $\gamma_2 = 0$ with probability $\omega_2 = 1 - \omega_1$. This boils down the analysis to a single parameter ω_1 governing the firm's ability to adjust. The extreme case of $\omega_1 = 0$ shuts down any possibility of adjustment. At the other extreme of $\omega_1 = 1$, the firm is always willing to make a 25 percent upward adjustment. This range of ω_1 allows for wide variation in the ability to adjust.

Finally, we set the mixing cost $\kappa^{mix} = \kappa^2 - \kappa^1$ so that the cost of combining one half size and one full size in the same shipment equals the cost of two full sizes. In the data, cases where one full and one half size go out in the same shipment are relatively rare. Adding this additional parameter is a shortcut for allowing for the model to fit this particular fact.

The list of parameters to be estimated is $\theta = (\eta, \omega_1, \zeta, \mu, \sigma)$, where we leave the indices for origination and destination implicit. Note that in the first stage when we estimate the shipment model for a particular source location, we take the friction η as fixed. Then, having recovered the level of the friction for particular sources, in a second stage we estimate

[14] Of course, even away from these mass points, the distribution of z in the figure does not coincide with the distribution of the original x because some of these observations have been adjusted.

parameters governing the level of η.

We match the model to the 13 statistics listed below, using generalized method of moments (GMM).[15] The moments capture features such as the size distribution of shipment volumes, whether or not consolidation is taking place, and the amount of empty space in unconsolidated containers.

Statistic	Description
1.	$\Pr(n_i^1 = 1) \times \Pr(z_i \leq q^2, yes_i^{con} = 0)$
2.	$\Pr(n_i^1 = 1) \times \Pr(z_i > q^2, yes_i^{con} = 0)$
3.	$E[\tilde{z}_i] \times \Pr(yes_i^{con} = 0)$
4.	$E[\tilde{z}_i^2] \times \Pr(yes_i^{con} = 0)$
5.	$E[\tilde{z}_i] \times \Pr(yes_i^{con} = 0)$
6.	$E[\tilde{z}_i^2] \times \Pr(yes_i^{con} = 0)$
7.	$E\left[\tilde{z}_i\right]$
8.	$E\left[\tilde{z}_i^2\right]$
9.	$E\left[yes_i^{con} = 0\right]$
10.	$E[yes_i^{con} = 0] \times \Pr\left(z_1 > q^2\right)$
11.	$E[empty\_rate_i] \times \Pr\left(n_i^1 = 0, n_i^2 = 1, yes_i^{con} = 0\right)$
12.	$\Pr\left\{\tilde{z}_i = q^2\right\}$
13.	$\Pr\left\{\tilde{z}_i > q^2\right\}$

We begin by estimating the model for Walmart, producing separate estimates for the various leading source locations. We include the 10 largest source locations from China. For Bangladesh, there is only one source, Chittagong. For India, we use the top five locations. Earlier we explained that Walmart's shipment volumes to its five import distribution centers are the same order of magnitude. For simplicity we will assume the shipment distributions are identically the same across all the destinations and estimate the model with pooled data across destinations. Table 7 reports the point estimates of the model coefficients. Given the large number of shipment observations from each location, the estimated standard errors

[15] We use a two-stage procedure to derive the weighting matrix. We use only the diagonal terms of the weighting matrix. The moments below are mean zero after we difference the expected value in the model, which we calculate through simulation.

are quite low, and are reported separately in the data appendix, to allow Table 7 to be more readable.

We begin by discussing the estimates for Shenzhen-sourced imports. Our estimate for the friction is $\eta = 0.126$. This means the full cost of shipping through consolidation is 12.6 percent more than a completely-loaded full-size container, on a per volume basis. We think of this as a fairly low friction, especially in relation to what we will see in other samples. Two features of the Walmart data tell us the friction cannot be exactly zero. First, while Walmart's empty rates out of Shenzhen are quite low, they are not zero, and the fact that unconsolidated containers go out even partially empty is evidence of some friction. Second, while consolidation is common, many shipments are unconsolidated. If the friction were zero, virtually all shipments would be consolidated because the chance that a randomly selected shipment size would exactly fit in a container would be negligible. Next note the estimate $\hat{\omega} = 0.785$, which is the probability that the firm draws a low cost of adjustment. This indicates that the degree to which the firm is able to make adjustments in shipping size is significant. Thus, we see that out of Shenzhen the firm is able to respond to the indivisibility constraint on two margins: consolidation and shipment-size adjustment.

The model fits the data relatively well. Figure 6(b) is the fitted value of the shipment volume distribution. It looks like the data in Figure 6(a). (See Figure 6(c) for a plot of the model and data together.) The main difference is that the right side of the distribution is smoother in the data than in the model. Our modeling assumption that the cost distribution has only two points is likely a contributing factor.

Figures 7(a) and 7(b) illustrate how underlying firm behavior varies with the initial target size x (where again, units are defined in terms of a full-size container). Figure 7(a) plots what the adjustment would be conditional on shipping an unconsolidated load and drawing the low adjustment cost. Note in the figure that if a shipment is close to a half-size load, it gets rounded to exactly a half-size load. If it is at least a little above half size but below full size, it gets rounded up. When x exceeds a full-size load, essentially the policy is to round up or round down to the closest full-size load.

Figure 7(b) plots the probability of consolidation, given the initial target x and a low adjustment cost draw. If x is less than around 60 percent of a full container, the probability of consolidation is virtually one. For x just above 60 percent, the probability of consolidation drops sharply and attains its minimum of 0.23 at $x = 1$, where the target shipment level exactly matches the indivisibility constraints. And it hits the minimum again at $x = 2$, $x = 3$, and so on. That consolidation may still take place for such x is a consequence of the random shocks to the profitabilities of choosing either consolidation or unconsolidation. These shocks are governed by the parameter $\zeta = 0.103$, which is small but not negligible. (The units are relative to the price of shipping a full-size container, again normalized to one,) Note the local maximums at $x = 1.5$, $x = 2.5$, and so on. At these points, the indivisibility problem is at its worse, from the perspective of filling up a full-size container. (Use of a half size exactly fits at these points, but half sizes are relatively expensive.)

Next consider the estimates from the other source locations. There is substantial variation across locations in shipment volume. Figure 8 plots the estimated friction for each sample against the shipment volume (in log scale). There a clear tendency for high-volume locations to have a lower friction. In the low-volume source locations, the friction is on the order of 0.30 or more. It falls to less than half that level in high-volume locations.

We noted that Target looks very similar to Walmart in our earlier descriptive statistics. In Panel B of Table 7, we report the estimate of the model for Target for originations from Shenzhen. The estimates are remarkably close to what we get for Walmart, both in terms of the estimated friction parameters and the shipment size distribution parameters. The estimates for Costco are just below those for Target. As discussed earlier, Costco's business model is to ship relatively few high-volume goods, where consolidation is not an issue. For the Costco case, we assume consolidation is infeasible, implicitly setting the friction to $\eta = \infty$.[16] For Costco, we estimate that $\omega = 1$, which is the maximum degree of adjustability. Note the very large difference in estimates of mean shipment size between Costco and Walmart or Target (the mean of $\ln x$ equals 0.6 for Costco compared to only -0.7 for Walmart and Target).

We noted earlier two descriptive statistics about changes over time in Walmart's behavior during the sample period: an increasing consolidation rate (Figure 2) and a decreasing empty rate (Figure 4). Both are consistent with Walmart enjoying technological improvement over time. In Panel C of Table 7, we use the model to investigate this issue, estimating the model separately for the earliest year (2007) and the latest year (2015) in the sample. The estimated friction fell by a third over the time period (η fell from 0.148 to 0.095). The ability to adjust also improved (ω_1 increased from 0.754 to 0.857). Finally, the distribution of shipment sizes shifted substantially to the left (μ fell from -0.388 to -0.809), which could be driven by increases in both variety and delivery frequency.

As one might expect, there is a significant seasonal pattern in Walmart's business. To

[16]The ζ parameter is irrelevant here. In estimation we eliminate moments that involve conditioning on $yes^{con} = 1$.

examine seasonal effects, we divide the year into two-month intervals. Figure 9 plots average import volume from Shenzhen by two-month interval period. We normalize average container imports relative to the peak period, September/October, which is when huge volumes of goods are shipped in advance of the holiday shopping season. At the low point in May/June, volume is only 40 percent as high as the peak. We estimate the model separately for each two-month period and plot the estimated friction in blue in Figure 9. Note the clear countercyclical pattern of the friction level. The friction falls to its lowest point of 0.085 at the peak and is highest at 0.167 in March/April and in May/June, the low-volume period in the months after Christmas.

For both the cross section of locations (Figure 8) and the seasons (Figure 9), there is a negative relationship between shipment volume and the estimated friction. Table 8 reports results of semi-log regressions for both the cross-sectional and seasonal cases, where the horizontal axis is log container volume in each case. The semi-elasticities are similar, equaling -0.064 in the cross-sectional case and -0.079 in the seasonal case. Using the cross-sectional estimate, a 1 percent increase in volume is associated with a decrease in the friction of 6 basis points.

One concern about giving a structural interpretation to the semi-elasticities in Table 8 involves the potential endogeneity of volume levels. In particular, if there are unobserved location factors that make the consolidation friction low at some locations, then everything else the same, more advantageous source locations will have higher volumes, and ordinary least squares regression estimates will tend to bias upward the magnitude of the treatment effect of volume in reducing frictions. However, we expect any bias to be minor because ocean freight is only one part of overall wholesale cost and because η is only one part of the cost of ocean freight. This point can be illustrated by an example. Take the baseline case of Walmart out of Shenzhen, where the estimated friction is $\eta = 0.126$. Now consider another location with $\eta = 0.170$, and assume everything else is the same as for the Shenzhen estimates. The difference in η is substantial, and if the true structural semi-elasticity is -0.64, container volume would have to be cut in half to achieve this increase in η. We can use the model estimates for Shenzhen to calculate that raising η from 0.126 to 0.170, everything else the same, raises the price index for ocean freight by 2 percent. As noted earlier, for retailers such as Walmart, ocean freight is on the order of 8 percent of the wholesale cost to import goods from Asia to U.S. ports. Thus, the wholesale price in the second location is higher by a factor $1.0016 = 1 + 0.02 \times 0.08$, a trivial difference. We will take the differences in volumes across locations as exogenous and going forward will treat the slope -0.64 as the structural semi-elasticity. It is encouraging that the estimate using seasonal variation is roughly similar because the assumption that the seasonal pattern is exogenous can be easily motivated.

We conclude this section by estimating the model for freight-forwarder-intermediated shipments. For such shipments from China, Los Angeles is the primary destination. We condition on shipments originating in Shenzhen destined for Los Angeles.[17] We break down the sample by the consignee size measure we used previously. Table 9 presents the estimates. One issue is that the model does not fit as well for these samples, as compared to our earlier estimates (compare the GMM criterion). The last two columns report the share of shipments that are consolidated in the model and the data. Note that before, in Table 4, consolidation rates were reported as weighted by containers. Here we report the unweighted versions, which are obviously higher, since there are relatively many small shipments that get consolidated. The model matches the overall qualitative pattern that larger consignees choose to consolidate at a lower rate. This pattern is mainly driven by the larger shipment sizes of big firms; observe the sharp increase in μ as we go down the table.[18] Notice also that the consolidation friction sharply decreases with firm size.

6 Analyzing the Estimated Model

The firm faces a trade-off: as it expands its count of IDCs to reduce inland freight, the gains may be offset by higher indivisibility costs. This section uses the model estimates to quantity the terms of the tradeoff. Part 1 of this section reports our estimates of indivisibility costs. Part 2 considers inland freight costs, which make up the other side of the trade-off, and then both sides of the trade-off are used to determine firm choice. Part 3 puts the results to work to analyze the effects of mergers and dissolutions. Part 4 examines the effects of variety.

[17] We include all shipments to the Los Angeles customs district, which includes Long Beach. We require that the shipments clear customs in the district in addition to being unladen in the district.

[18] Note we are defining consignee size categories by the count of shipment groups imported, not by the volume of particular shipments (i.e., by the extensive margin, not the intensive margin). Our finding that consignees that are bigger on the extensive margin are also larger on the intensive margin is not necessarily true by definition.

6.1 Unit Indivisibility Cost

We define the *unit indivisibility cost* for a given shipment target level x to be

$$C^{ind}(x) = \frac{-V^{ind}(x)}{-V^{no\_ind}(x)} - 1, \tag{7}$$

where $V^{ind}(x)$ is the maximized value at x, given indivisibilities (see equation (5)), and $V^{no\_ind}(x)$ is the analogous maximum in the *No Indivisibilities* case, which we define as the case where (1) there is a continuum of container sizes with price per unit equal to $\lambda = \kappa_2/q_2$ and (2) the consolidation friction is zero, $\eta = 0$. Note that under (1) and (2), the shipping cost for unconsolidated and consolidated is identical, and the decision will depend upon the random profit terms associated with each decision. The minus signs are included to flip maximized profits to minimized costs. We difference the statistic from one to turn it into a rate. If $\eta = 0$ and if x is a multiple of the full-size capacity level q_2, the indivisibility cost is zero (i.e., $C^{ind}(q_2) = 0$, $C^{ind}(2q_2) = 0$, and so on). Otherwise it is strictly positive. We define the average unit indivisibility cost as the weighted mean over x,

$$\bar{C}^{ind} = \frac{\int_0^\infty C^{ind}(x) x dG(x)}{\int_0^\infty x dG(x)}.$$

Note there are three sources of indivisibility cost in the model embodied in this statistic. First, if the firm consolidates, it incurs the friction η. Second, if the firm does not consolidate and sends a partially loaded container or uses a half size, the unit freight cost will be higher than when a full-size container with no empty space is sent out. Third, the firm may distort the volume levels of unconsolidated shipments up or down relative to the desired target level, and these adjustments yield losses relative to the ideal with no indivisibilities.

We noted that Walmart divides its imports across five IDCs with roughly 20 percent going to each. We assume now the firm under consideration divides imports exactly m° ways across destination IDCs, and we evaluate how unit indivisibility costs change when the firm alters the count to m' IDCs but leaves other aspects of its distribution structure the same. In particular, recall in the simple model the firm jointly picks IDC count m and delivery frequency f. Here we leave f *fixed*. This is a limitation of our analysis that needs to be highlighted, and we will come back to it in the policy analysis below.

In the baseline model with m° destination IDCs, we can assume that for each shipment arrival of size x, there are actually m° such arrivals of size x, one for each IDC. The total national shipment volume of the particular product is xm°. If instead the firm were to divide shipments m' ways, the target shipment for each IDC would equal xm°/m'. If the parameters of the lognormal distribution governing the baseline case are given by μ° and σ°, the mean of the arrival distribution in the new case is given by

$$\mu' = \mu^\circ + \ln(m^\circ) - \ln(m'), \tag{8}$$

and the standard deviation stays the same, $\sigma' = \sigma^\circ$.

A change in m' will change the volume going to each destination. As discussed above, a change in volume affects the consolidation friction. Let $\tilde{\eta}(m')$ denote the friction as a function of m', through the choice of m''s effect on volume. We use the semi-elasticity 0.064 from the cross-sectional relationship in Table 8, and we can write it as

$$\tilde{\eta}(m') = \tilde{\eta}(m^\circ) + 0.064(\ln(m') - \ln(m^\circ)). \tag{9}$$

We use our model estimates for the various samples in Tables 7 and 9 to estimate unit indivisibility costs. We also report the effects on the indivisibility cost of various counterfactual levels of m'. Table 10 displays the results. The shaded cells are the values obtained for the baseline IDC count the firm actually uses (i.e., where $m' = m^\circ$). Let us begin by looking at the column where the IDC count is $m = 5$, which is what Walmart actually does in our data. Our estimate of the unit indivisibility cost out of Shenzhen is 10.3 percent. This is well above the empty rate of containers out of Shenzhen and reflects the fact that in order to keep the empty rate low, Walmart incurs both consolidation frictions, as well as distortions in shipment size to round volumes to match the indivisibilities.

Next we vary m'. For illustration, in the first row we begin with an example where η does not vary, even though a change in m changes volumes to each destination. Right below this we report the estimates when η varies according to (9). At the baseline where $m = 5$, the costs are the same. But notice the substantial difference when we vary m. For example, when η is fixed, if the firm lowers the destination count to $m = 1$, indivisibility cost falls from 10.3 percent at the baseline to 6.6 percent. However, when we take into account the effect that higher volumes have scale benefits for consolidation, the indivisibility cost at $m = 1$ is 2.3 percent. Analogously, if we raise m above 5, shipment volume per destination declines, and indivisibility costs rise faster when we take this into account.

Next consider the estimate for Mumbai-sourced imports. The indivisibility cost in the baseline is quite high, 25.3 percent. Given our earlier discussion of the relatively low volumes from India, and high empty rates and high use of half-size containers, we expect to find high indivisibility costs from this source. Note also the big difference in the indivisibility cost, 8.8 percent instead of 25.3 percent, if one IDC is used instead of five. We will see below that indivisibility costs out of Mumbai are such that if the firm could customize its import network for each source country, it would not choose to use five IDCs to import goods from India.

In the fourth row of Table 10 we report an estimate of the weighted average across all source locations in Asia.[19] The difference between the weighed average in Asia and Shenzhen is quite small. Shenzhen has a huge weight in the average, and other leading sources such as Xiamen and Chittagong also have very low frictions, which offset the high indivisibility costs from places such as Mumbai. Below, when considering the firm's choice problem for how to set m, we assume the firm is constrained to use the same import network for all source locations. The weighted average from Asia case will be the relevant row used in the network choice problem of the firm.

Next consider Target. Target has four IDCs, so we evaluate the cost at $m^{\circ} = 4$. The indivisibility cost is 12.0 percent at this choice, slightly higher than the 10.3 percent figure for Walmart at its choice $m^{\circ} = 5$.

When we estimated the model for Costco, we assumed consolidation is not feasible, and the only adjustment margin is to round volumes up or down. Table 10 presents our estimate of indivisibility costs for Costco, for different counts of IDCs, where the actual is taken to be $m^{\circ} = 10$. Since no consolidation is taking place, the issue of how the friction changes with m is not relevant here. The first thing to note is that there is actually a range (see $m = 6$ to 8) where the indivisibility cost *decreases* in the count of IDCs. To see why such decreases can happen, suppose at the initial situation, all shipments that arrive happen to exactly equal what would fit in a single container. In that case, cutting up shipments differently can only make things worse. Note $m' = 9$ is actually a local minimum.[20] For the sake of illustration, in the row seventh row, below the Costco numbers, we report the exercise with the Walmart Shenzhen parameter estimates, with the difference that we make consolidation infeasible. Note how indivisibility costs explode as the count m of IDCs increases, which is a very different pattern from that for Costco. The bottom line is that Costco has low

[19] We have geocoded source locations in China, Bangladesh, and India, but not other Asian locations. We use our estimates for India as our estimate for the other Asian countries.

[20] One complicating factor here that we do not address is that Costco's shipments to its IDCs are not roughly equal as they are for Walmart.

variety and large volumes of particular products, enabling it to avoid consolidation. In this model, using eight or nine destinations is not that different from using five. Walmart has low product volumes and needs to consolidate because otherwise indivisibility costs explode.

The last case we consider includes the estimates of freight-forwarded intermediated shipments from Table 8. By definition of the sample, these firms are doing consolidation externally. We expect any one firm to be a small portion of any one freight forwarder's business. For this reason, we take the consolidation friction as fixed when we vary a particular firm's import strategy. At the bottom of Table 10, we report three of the cases: the smallest consignee size class (1), the largest (251 and up), and a middle case (21-100). We take as a baseline that the firm is using a single IDC. Note the sharp decrease in the indivisibility cost with firm size, going from 40.8 percent for the smallest category to 18.5 percent for the medium and 14.3 percent for the largest firms. Note that for the largest, this is just four points higher than for Walmart out of Shenzhen. But remember, Walmart attains this with five IDCs. If the largest firm tries this, the indivisibility cost blows up to 38.8 percent. If the smallest firms try this, the indivisibility cost increases to over 103.0 percent.

6.2 The Other Side of the Trade-off: Inland Freight

We turn to the other side of the trade-off, which is how a change in the number of destination IDCs affects inland freight. With more IDCs, the firm is able to substitute away from costly freight miles on land to relatively cheap freight miles on water, as illustrated with the simple model in Figure 1.

We include both Walmart and Costco in the exercise and we use data sets with the locations of all the stores of each chain. We allocate sales proportionately across store locations and assume shipments are sent directly from import distribution centers to stores. For Walmart, shipments actually first pass through regional distribution centers. However, there are a large number of these (42), allowing for a relatively direct flow of goods. For Costco, goods are shipped to stores from the IDCs (called "depots" by Costco). The data appendix details the various assumptions we make to price out costs under the various scenarios. We include freight costs, but not time costs.

We impose some constraints on the geographic composition of the IDC networks in order to match the observed networks. For the Walmart estimates, we assume that the 5-IDC network matches the locations of Walmart's actual network. The planned sixth IDC is Mobile, Alabama. We have freight data out of New Orleans, and we substitute New Orleans as a proxy for Mobile. For Costco, we assume that the 10-IDC network matches the locations of Costco's actual network. Other than these constraints, as we vary the count m, we choose the location of the m-th IDC to minimize freight costs, taking as given the location of the $m-1$ previously opened IDCs. For example, with just one IDC, Los Angeles is the lowest freight choice for Walmart and Costco. Taking Los Angeles as given, when a second IDC is added for Walmart, Norfolk is selected. For Costco, New York is selected. The notes to Table 11 reports the full sequence orders used in the exercise.

Table 11 displays the average freight cost of shipping the contents of one ocean container inland as a function of the number of IDCs. To facilitate comparison with Table 10, we report the cost as a percentage of a rough estimate of the ocean freight for shipping one standard container from Asia to a U.S. port. We use $3,000 as our ocean freight estimate, which is approximately the average freight rate (c.i.f.) in a sample of shipments based on public Census tabulations that we discuss in the data appendix. With only one IDC, inland freight for Walmart is estimated to average 46.2 percent of ocean freight. In Table 10, the corresponding indivisibility cost is 3.1 percent. (We use the weighted average case in Asia for this exercise.) As the number of IDCs increases, inland freight decreases, while the indivisibility cost increases. At the 5-IDC network that Walmart actually uses, inland freight is 33.5 percent and the indivisibility cost 10.3 percent, which, compared to a 1-IDC network, yields inland freight savings of $12.7 = 46.2 - 33.5$ percentage points, more than offsetting the $7.2 = 10.3 - 3.1$ increase in indivisibility costs. Next note that moving from five to six IDCs is something the firm will not want to do, because the freight cost falls by 0.6 percentage points, while the indivisibility cost increases by 1.2 percentage points.[21]

Note that out of Mumbai, the increase in the indivisibility cost of going from one to five IDCs equals 16.5 percent, which is above the 12.7 percent savings in inland freight. If Mumbai were the only source, it would not make sense for Walmart to use so many IDCs. However, since imports from China make up such a large share of Walmart's imports, the network is being optimized to what works best from China, to the disadvantage of imports from India.

The cost-benefit analysis here is rough, as we are leaving out other considerations.[22] If eight IDCs were selected instead of five, the freight cost would fall by 3.6 percent, compared

[21] Note, however, that the freight savings from the sixth IDC would likely be more if it were put in Mobile (where Walmart is putting it) instead of New Orleans (which we selected because of freight data availability) because Mobile is closer to the center of its existing operations.

[22] A technical issue that we do not address is that in estimating indivisibility costs in Table 10, under counterfactual numbers of IDCs, we assume equal divisions across destinations. But in calculating freight costs in Table 11, we do not enforce equal shares across destinations.

to a 1.6 percent increase in the indivisibility cost. However, we are not incorporating any fixed cost for IDCs, and that would work against such an expansion. Also, there are additional location factors not included in the analysis, such as concerns about unionization. The seventh and eighth IDCs are in Seattle and Philadelphia, which historically have been areas with strong unions. In our analysis, we have not modeled a role for an inland IDC. If a container goes through the port of Los Angeles, and then inland to an IDC in Chicago, it could just as easily stop at the IDC in Los Angeles, and be unpacked there. For this reason, and as shown in Table 11, there is no saving in freight cost when Chicago is added as the fifth IDC, but there would be if we expanded the model to incorporate a benefit from ocean/rail intermodal transport. With the model as specified, Walmart four IDCs is better than five. Finally, we note that with our estimates, the firm is indifferent between three and four IDCs, achieving the minimized total cost of 42.6 either way ($7.7+34.9$ with $m=3$, $9.1+33.5$ with $m=4$). To break the tie, in discussions below, we assume the firm chooses $m=4$.

Turning now to Costco, the number minimizing the inland freight plus indivisibility cost is four IDCs, but a choice of nine IDCs only adds 0.5 percent to the combined cost, and we are not taking into account benefits of inland IDCs, of which Costco makes extensive use. Recall that in the calculations for Costco's indivisibility cost, we assume consolidation is infeasible. If we shut down consolidation for Walmart and hold fixed everything else, particularly the distribution of desired shipment sizes, then Walmart would choose a single IDC to minimize the inland freight plus indivisibility cost. Product shipment volumes are much smaller for Walmart than Costco, and Walmart's ability to consolidate is crucial for its ability to distribute shipments across multiple IDCs.

6.3 Merger and Dissolutions of Firms and Countries

We now discuss three different types of merger/dissolution scenarios and use the estimated model to examine effects on indivisibility costs in each case. We take as our baseline the weighted average from Asia case in Table 10. As just noted, given these parameters and the inland freight cost from Table 11, the firm will choose four IDCs, and the resulting indivisibility cost and inland freight cost are, respectively, 9.1 percent and 33.5 percent. It is convenient for the exposition to refer to the baseline as the *merged* case, and then consider varies possible dissolution scenarios relative to the merged case.

We begin with a discussion of horizontal merger/dissolution. Suppose that relative to the baseline merged case, the firm is dissolved into N equal-sized symmetric firms, all selling the same identical set of products, to the same identical set of domestic destinations. In particular, assume that for every target shipment x^{merge} that would arrive for the merged firm, under dissolution the same shipment is divided exactly N ways, so the target shipment level for each firm is $x^{dis} = x^{merge}/N$. In this scenario, the effect of dissolution on indivisibility cost is the same as it would be from dividing up shipments to more destination IDCs. In particular, suppose that under dissolution the individual firms each use m^{dis} IDCs. Then the resulting indivisibility cost would be the same as what the merged firm would incur if it uses $m^{merge} = N \times m^{dis}$ IDCs. In Table 12 we consider three possible dissolution levels, $N = 2$, 10, or 100, and report the indivisibility cost for each N as a function of the number of IDCs used. For convenience, at the top of the table we report the analogous figures for the baseline merged case, which come directly from Table 10 in the row for the weighted average case from Asia. We can see that under dissolution with two firms and one IDC, the indivisibility cost is that same as with one firm with two IDCs. And with two firms and two IDCs, the indivisibility cost is that same as with one firm and four IDCs.

It is easy to calculate that under all three dissolution possibilities, the dissolved firms would choose $m^{dis} = 2$, as compared to the baseline merged firm, which chooses $m^{merge} = 4$. Dissolution raises overall cost, through a combination of higher indivisibility cost and higher inland freight. A dissolution to $N = 100$ is extreme, of course, but we include it because the difference between Walmart and the retail system it replaced is extreme. For the $N = 100$ case, dissolution raises the total of indivisibility plus inland freight costs by 39.4 percentage points, where again the percentage points are defined in terms of the cost of ocean freight. We reported a rough estimate earlier of the ocean freight share of the wholesale cost to U.S. port equal to 0.08, which implies that the effect on cost as a percent of the wholesale price equals $3.2 = 0.8 \times 39.4$ percent.

It is important to emphasize that we abstract from indivisibility costs that might occur downstream in the flow between regional distribution centers and retail stores. If this factor were incorporated, there would be an analogous downstream cost difference between the merged and dissolution cases. In this way, the exercise *understates* the cost differences between merger and dissolution.

It is also important to emphasize that we hold fixed delivery frequency in the exercise. After dissolution, we expect firms to reduce delivery frequency to mitigate indivisibility costs, analogous to the way the firm responded in the above analysis to reduce network size m. Therefore, in this respect, the exercise *overstates* the cost differences between merger and dissolution.

With horizontal dissolution, the effects on indivisibility cost operate in two ways, the *shipment-size channel* (the individual shipments that arrive are smaller) and the *consolidation-scale* channel (a firm's overall volume to a particular destination declines, making consolidation more costly). Next we consider a type of dissolution where the consolidation-scale channel operates, but the shipment-size channel is shut down. Suppose in dissolution, the N different firms are allocated different products in a manner that maintains symmetry across the firms. In other words, rather than divide up a given shipment N ways, we can think of this scenario as each firm taking turns getting the entirety of particular shipments. The effects on indivisibility cost are presented in Table 12, where we can see that the effects are much attenuated compared to the horizontal dissolution case, a result of the fact that the shipment-quantity channel is no longer operative. The firms respond to the dissolution by reducing import network size. However, the response is attenuated compared to the horizontal dissolution case. (Here, for all three values of N, firms respond by reducing m to $m^{dis} = 3$. For horizontal dissolution, the reduction is to $m^{dis} = 2$.)

In the last scenario, we consider mergers or dissolutions of countries. Because of customs processing issues, distribution systems for different countries are typically operated separately. For example, Walmart sells products in the United States, Canada, and Mexico, and there is a distinct import distribution system for each. Suppose we take as our baseline case that Walmart is serving a single country through four distribution centers. Suppose through a Brexit-like event, there is a dissolution of the country into four separate markets, where the new boundaries happen to exactly coincide with the boundaries of the market areas of the four distribution areas. In this extreme case, the dissolution has no effect on indivisibility costs, as the firm can run the same operation as before the dissolution. If instead the country were dissolved into eight equal-sized locations, there would be an increase in the indivisibility cost of 4.1 percentage points (using the change from 9.1 to 13.5 percent for the weighted average case from Asia).

6.4 Variety

Take our model and assume initially all goods shipped are a single color—white. Suppose next that there is technological change where the goods now come off the assembly line in N colors, with a fixed fraction $1/N$ of each shipment being each color. If there is no attempt downstream to maintain inventory levels of particular colors, and if containers are packed exactly like before when everything was white, this particular way of increasing variety would have no effect on the indivisibility cost.

Suppose instead that variety is increased in a more substantive way. In particular, fixing overall volume, suppose the firm scales up variety by a factor N, and that new varieties arrive for shipment independently of each other, perhaps even from different suppliers. Here, an increase in variety N raises indivisibility cost. Dividing shipments into N varieties has the same effect on shipment quantities as a horizontal dissolution into N firms, i.e. the shipment-size channel is operative here. However, the consolidation-scale channel does not enter here. As shipments are divided into more varieties, the total volume being sent by a particular firm to a particular destination remains fixed, so the scale levels that pin down consolidation efficiency remain unchanged. Table 12 presents indivisibility cost for different values of variety expansion N and choice of m. It is easy to calculate that under all three variety expansion scenarios, the firm contracts its import network size to $m = 2$. The total of indivisibility cost plus inland freight rise by 1.9, 6.6, or 12.1 percentage points depending on whether variety expands to $N = 2$, 10, or 100. For the $N = 100$ case this increase works out to 1.0 percent of wholesale cost (assuming an ocean freight share of 0.08), which can be compared to the 3.2 percent from a horizontal dissolution of the same magnitude. The key difference is that with variety expansion, the firm retains its scale economies of consolidation. Given the firm's efficiency in consolidation, it is relatively cheap for the firm to offer extensive variety. If instead the firm is broken up into small "mom and pops," it forgoes scale economies in consolidation, making it relatively expensive to offer high variety.

We make one last comment about the prediction of this model that everything else the same, when variety is increased the firm will chose a smaller import network m. This is consistent with what Walmart does with clothing and footwear, as compared to general merchandise. For clothing and footwear, which feature extensive varieties and low shipment volumes per variety, Walmart actually uses only a single IDC, instead of using five IDCs, as it does for general merchandise.

7 Concluding Remarks

This paper develops a model of indivisibility costs that incorporates three components: (1) the cost of unused container space, (2) frictions in consolidation, and (3) the cost to distort shipment sizes up or down to conform to lumpy container sizes. The model allows the friction to depend upon shipment volume. We use unique data on the contents of imported containers to estimate the model. We find significant scale economies in consolidation. We examine the trade-off between indivisibility costs and inland freight costs in the choice of how many domestic ports to use to bring in imports, and the estimated model does a reasonably good job in accounting for the behavior of Walmart and Costco.

The model has a number of limitations that can potentially be addressed in future work. The shipment-level model that we estimate is reduced-form in nature. In particular, in exercises where we vary the number of domestic ports, we leave delivery frequency fixed. In future work, it may be possible to incorporate the issue of delivery frequency directly into the analysis, taking advantage of the rich information in the data regarding the timing of deliveries. More generally, the role of timing is not developed in this analysis but is obviously an important part of the story.

Data Appendix

A1. Source and Processing of Bill of Lading Data

We have a complete set of bills of lading for the 18 months listed in Table A1, for a total of 18,809,816 records. As discussed in the text, we classify records as BCO, house, or master: Table A1 shows the percentage distribution across types. There is also a category for shipments either not using containers or using empty containers, representing 3 percent of the sample that we exclude.

Using the 18-month sample to begin with, we developed procedures for finding Walmart records, and then we applied the search procedures over the entire period, January 2007 through December 2015. We did text searches for Walmart across the various fields, including looking for Walmart's name as well as the GLN code, discussed in the text. We extracted 1,963,866 records, and of these, 82 percent include the GLN code. Table A2 presents the distribution of our Walmart records across years, for both counts of shipments and counts of containers. When counting containers, we use the container ID variable in each shipment record to ensure that we do not double-count containers. 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.[23] Overall, the container count in our sample is 56 percent of the aggregates reported by PIERS. Our coverage is highest in 2007 and in the first quarter of 2008, when Walmart was not redacting the consignee field. For the Walmart records, we processed the products field to pull out the item number of the records as well as the HS code.

We did extensive manual processing of the *place of receipt* field in order to identify the origin of shipments, and we geocoded shipments from China to the level of prefecture or county, and for India to the level of district or subdistrict. We processed the container information in the shipment records to create a data set at the shipment/container level. We developed an algorithm that seems to work well for determining whether a container is a 20-foot container or a 40-foot container, based on the first five characters of the container ID variable. As discussed in the text, we linked shipment records that reference the same container ID and the same shipment arrival date into consolidated shipment groups.

To allocate volume in shipments that are part of consolidated shipment groups, we use the information about piece count we have for each container used in each shipment, and allocate capacity proportionate to piece count. We also have a cbm measure, but this is

[23] To produce these estimates, PIERS uses the same bill of lading data from CBP that we have; it also uses additional information it obtains directly from shippers.

missing for many records and is at the shipment-level. In contrast, the piece count variable is available for all records and has the breakdwon by container for multi-container shipments.

For house bills of lading, we processed the text information in the consignee field, pulling out the zip code and state. We linked consignee shipments by linking on location of origination (prefecture level from China, subdistrict level from India), consignee zip code and state, and consignee name (first four characters).

A2. Walmart Import Distribution Shares

In the text, we note that Walmart's five IDCs have import shares that are roughly equal. Table A3 reports estimates of the import shares using the 9-year Walmart sample. We exclude apparel and footwear (i.e., goods with HS2 codes 61, 62, 63, and 64), since fashion goods are often distributed through a single IDC. The share of shipments is remarkably constant across the five IDCs. There is more variation in container shares, ranging from 15.6 percent at the low end to 23.4 percent at the high end.

A3. Freight Rates

The U.S. Census Bureau publishes tabulations on imports at narrow detail including: month, country of origination, commodity (10-digit HS 10), and ports of unlading and entry (U.S. Bureau of the Census, 2007-2015). The statistics reported at this narrow detail include freight charges (c.i.f.) associated with the import to deliver the good to the port of entry, as well as the entered value of the good, which includes the freight. Other statistics include weight and quantity (where the units depend on the type of commodity, e.g. number of microwaves). The Census Bureau reports these detailed statistics even if there is only a single shipment in a particular cell, so it is actually a transaction-level observation. Given all the detailed information in the Census Bureau tabulations and the bills of lading, it is possible to find some links between transactions-level observations in the published tabulations and the bills of lading. For Walmart, we have obtained 483 such links. From the matched bill of lading, we observe the number of containers in the shipment, and we can divide the freight charge by the number of containers to calculate freight on a per-container basis. Table A4 reports that the median freight is about $3,100 in this sample. We also report statistics of freight as a percentage of entered value, and the median is 7.6 percent.

As this is a small sample, we also consider a second strategy, with broader coverage. Approximately two-thirds of Walmart bills of lading contain the HS product code used for customs filings. We first use the public Census records from 2011 to 2015 to calculate the average freight as a percentage of entered value at the level of six-digit HS codes, for imports from China. These aggregate data combine transactions of Walmart with transactions of other firms. Then we merge these six-digit-level data into our Walmart records. The median freight percentage of value is 7.4 percent, across the 1.1 million Walmart records with six-digit HS coverage.[24]

Now we turn to the issue of pricing for half-size containers. In the analysis, we set the price of half-size containers equal to 75 percent of the full size. Here we motivate this assumption in two ways. First, shipping companies publish tariffs, and this is the approximate discount in the tariffs we examined. For Hapag Lloyd from China and India to U.S. ports, the published tariff for the half-size rate is approximately 75 percent of the full-size rate. OOCL, another leading shipping company, posts half-size rates that are 80 percent of full-size rates.

The second way we motivate the assumption is to appeal to published Census tabulations. As just noted, some observations published in the Census tabulations correspond to a single transaction. Unfortunately, the Census does not publish the number of containers used in a transaction and whether or not the containers are half size. However, linking in bills of lading, we have been able to determine this information for two samples of the published Census records. To construct the first sample, we first link observations across different rows of the Census tabulations to select out example records that are full-container loads, and where this same type of shipment is being sent at least twice (i.e., two different months or the same month to different ports).[25] We then search for matching records in the bill of lading data. In the second sample, we directly merge the two data sets on the various common variables, such as origination, destination ports, weight, and HS code (where available in the bill of lading data). Sample 1, where the matches involve multiple shipments of the item, is more heavily weighted toward BCO transactions (68 percent), as compared to sample 2 (36 percent). For the two samples, we add the information about the quantity of containers and the indicator for half-size usage, and then regress the log of freight charge per container on the various shipment characteristics listed in Table A5.[26] The implied freight cost of a

[24] This is consistent with calculations in Leachman (2010). He assumes Walmart's imports have an average value of $14 per cubic foot. He assumes freight to the West coast is $0.96 per cubic foot and to the East coast $1.47, for an average of $1.22 per cubic foot, which is 8.7 percent of the assumed import value and is about $2,500 after converting units to a full-size container.

[25] We look across different months and different destination ports for cases where for the same commodity, there are exact matches in value per unit and weight per unit. We figure out what the piece count is for a single container load and then verify that all the different records are either a factor one times the single container piece count, a factor two times this, and so on.

[26] We restrict both samples to shipments from China and Europe. We select shipments to the following West Coast customs districts: Los Angeles, San Francisco, Portland, and Seattle, and the following East

half size relative to a full size equals 76.5 percent for the first sample and 93.7 percent for the second. The remaining coefficients are roughly similar for the two samples. The result from the second sample might understate the discount for the half size. This will be true if small firms are likely to use half sizes (which we have already documented) and if small firms face higher freight rates (which is consistent with the negative coefficient on BCO in both regressions). The bias should be less for the first sample because selecting only firms importing multiple-full-container load shipments sweeps out the smallest firms. In any case, both estimates imply significant indivisibility, with the rate for the half size being well above 50 percent.

A4. Standard Errors of Model Estimates

Simulated standard errors for the model estimates across the various samples are reported in Table A6. The sample sizes are large, and the level of precision is high.

Coast districts: New York, Savannah, Norfolk, and Charleston.

References

Autor, David, David Dorn, Lawrence Katz, Christina Patterson, and John Van Reenen, (2017) "Concentrating on the Falling Labor Share," *American Economic Review Papers & Proceedings*, forthcoming.

Basker, Emek, and Pham Hoang Van (2010), "Imports "R" Us: Retail Chains as Platforms for Developing-Country Imports." *American Economic Review,* 100 (2), 414–418.

Bernhofen, Daniel M., Zouheir El-Sahli, and Richard Kneller (2016), "Estimating the Effects of the Container Revolution on World Trade," *Journal of International Economics* 98, 36-50.

Atkin, David and Dave Donaldson (2015), "Who's Getting Globalized? The Size and Implications of Intra-national Trade Costs, unpublished manuscript.

Bernard, Andrew B., J. Bradford Jensen and Peter 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. Timothy Dunne, J. Bradford Jensen, and Mark J. Roberts, 513-552. Chicago: University of Chicago Press.

Bernard, Andrew B., J. Bradford Jensen, Stephen J. Redding, and Peter K. Schott (2007), "Firms in International Trade," *Journal of Economic Perspectives* 21(3) 105–130.

Bernard, Andrew B., J. Bradford Jensen, Stephen J. Redding, and Peter K. Schott (2010), "Wholesalers and Retailers in US Trade," *American Economic Review: Papers and Proceedings* 100(2), 408-413.

Broda, Christian and David E. Weinstein (2006), "Globalization and the Gains from Variety," *Quarterly Journal of Economics*, 121(2), 541-585.

Cargo From China, "Ocean Freight Shipping from China: FCL or LCL?", https://cargofromchina.com/fcl-lcl/, accessed March 5, 2017.

Cosar, A. Kerem., and Pablo D. Fajgelbaum (2016), "Internal Geography, International Trade, and Regional Specialization," *American Economic Journal: Microeconomics* 8(1), 24-56.

DHL Global Forwarding Ocean Freight, "Ports of the World," undated pamphlet.

Feenstra, Robert C. (1994), “New Product Varieties and the Measurement of International Prices,” *American Economic Review*, 84(1), March, 157-177.

Holmes, Thomas J., and John J Stevens (2014), “An Alternative Theory of the Plant Size Distribution, with Geography and Intra- and International Trade”. *Journal of Political Economy* 122(2), 369-421.

Holmes, Thomas J. (2001), “Bar Codes Lead to Frequent Deliveries and Superstores.” *Rand Journal of Economics*, 32(4), 708–725.

Holmes, Thomas J. (2011), “The Diffusion of Wal-Mart and Economies of Density,” *Econometrica*, 79(1), 253-302.

Hummels, David L. (2007), “Transportation Costs and International Trade in the Second Era of Globalization,” *Journal of Economic Perspectives* 21(3), 131–154.

Leachman, Robert C. (2008), “Port and Modal Allocation of Waterborne Containerized Imports from Asia to the United States,” *Transportation Research Part E: Logistics and Transportation Review* 44(2), 313-331.

Leachman, Robert C. (2005), “Final Report: Port and Modal Elasticity Study,” unpublished manuscript.

Leachman, Robert C. (2010), “Final Report: Port and Modal Elasticity Study, Phase II,” unpublished manuscript.

Leachman, Robert C. and Evan T. Davidson (2012), “Policy Analysis of Supply Chains for Asia - USA Containerized Imports," Proceedings of the 2012 Industry Studies Association Annual Conference.

Levinson, Marc (2006), *The Box: How the Shipping Container Made the World Smaller and the World Economy Bigger.* Princeton, NJ: Princeton University Press.

MWPVL (2017), “The Walmart Distribution Center Network in the United States,” MWPVL White Paper, http://www.mwpvl.com/html/walmart.html. accessed February 12, 2017.

Pisinger, David (2002),"Heuristics for the container loading problem." *European Journal of Operational Research* 141(2), 382-392.

Redding, Stephen J. and Matthew Turner (2015), "Transportation Costs and the Spatial Organization of Economic Activity" in (ed.) Gilles Duranton, J. Vernon Henderson, and William C. Strange, *Handbook of Urban and Regional Economics*, 1339-1398. Amsterdam: Elsevier.

Rua, Gisela (2014), "Diffusion of Containerization," Finance and Economics Discussion Series 2014-88, Board of Governors of the Federal Resesve System.

U.S. Bureau of the Census (2007-2015), U.S. Imports of Merchandise, Statistical Month - January 2007 through December 2015 (DVD-ROM).

Figure 1: Simple Example of Alternative Distribution Systems

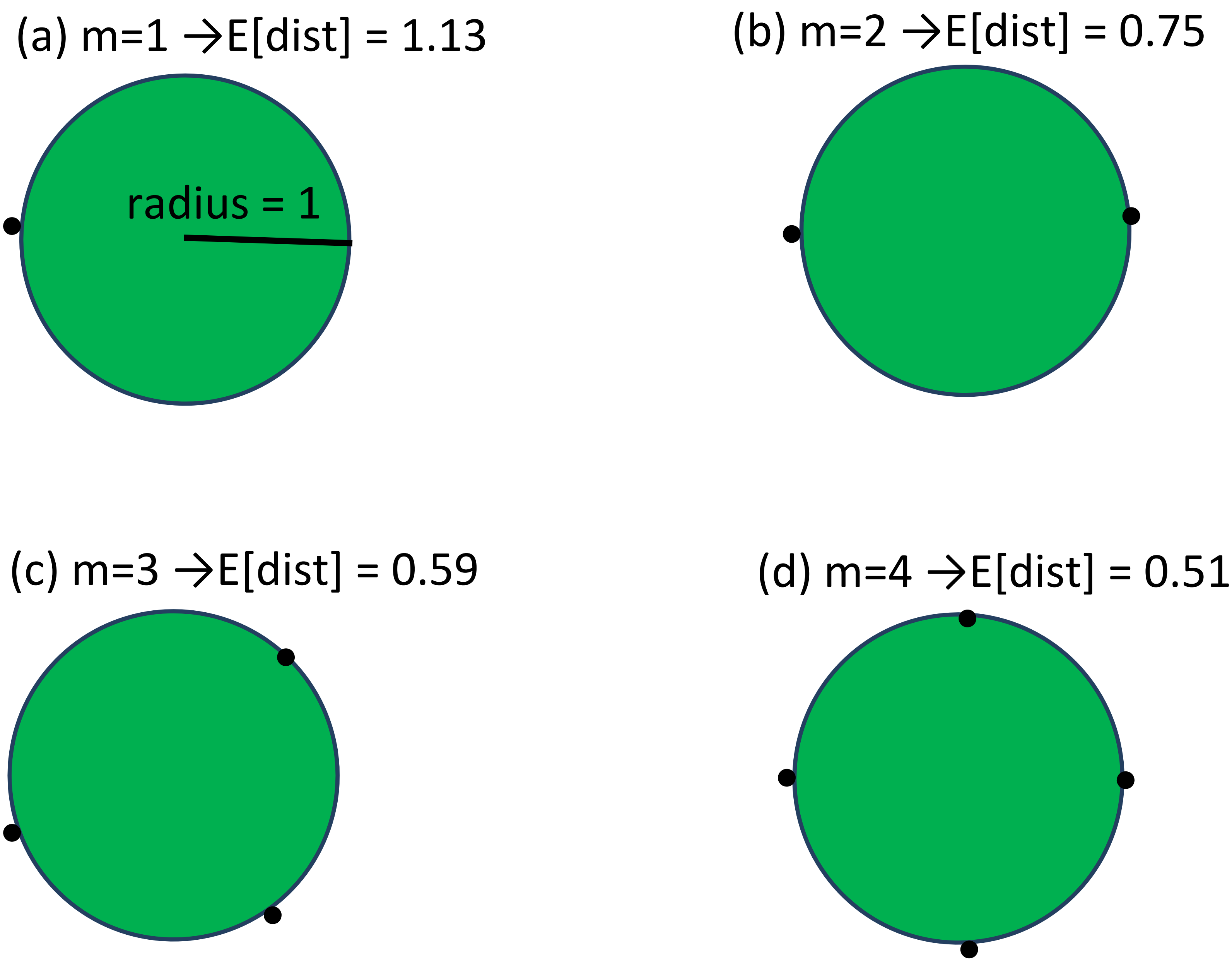


Figure 2
Consolidation Trends for Walmart 2007-2015
Percentage Share of Walmart-Imported Containers from China in Three Categories

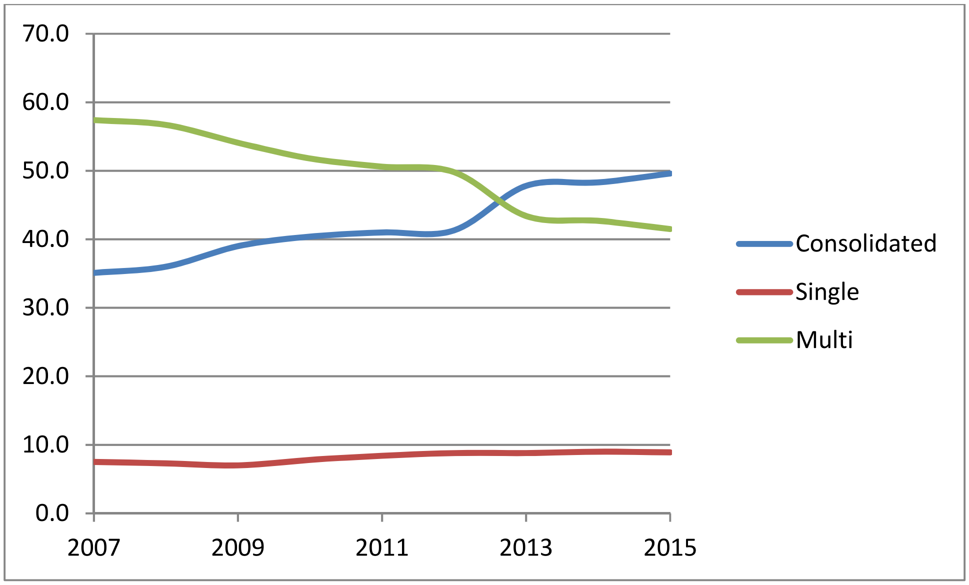
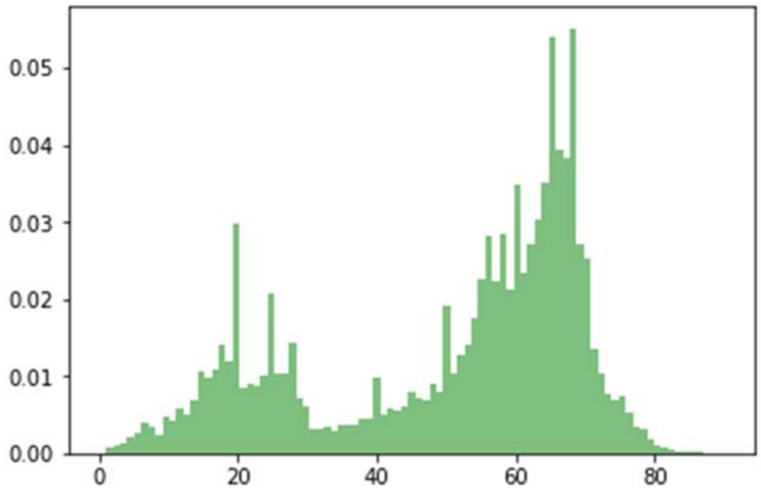
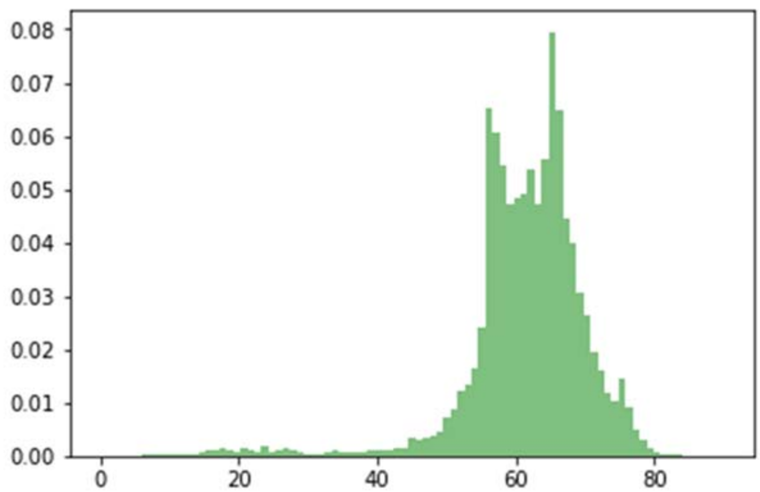


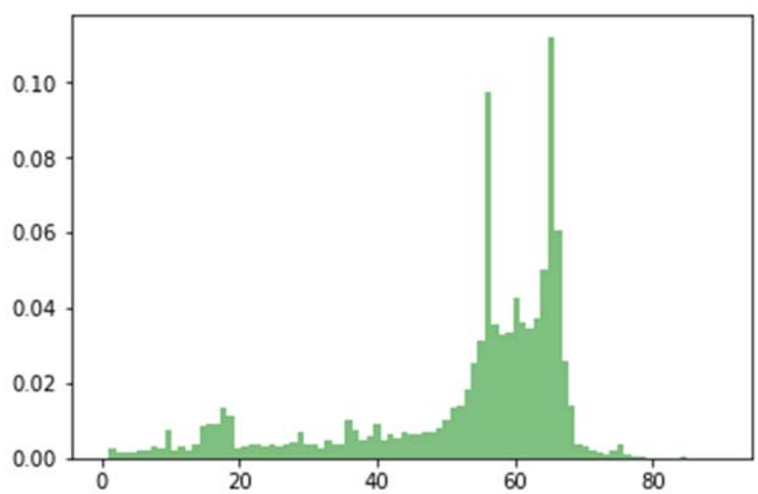
Figure 3. Histograms of Container Fill Levels (Cubic Meters) for Three Samples



(a) Sample 1: All Containers Originating in China



(b) Sample 2: Walmart Containers Originating in China



(c) Sample 3: Walmart Containers Originating in India

Figure 4
Empty Rate out of China for Walmart by Year, 2007-2015

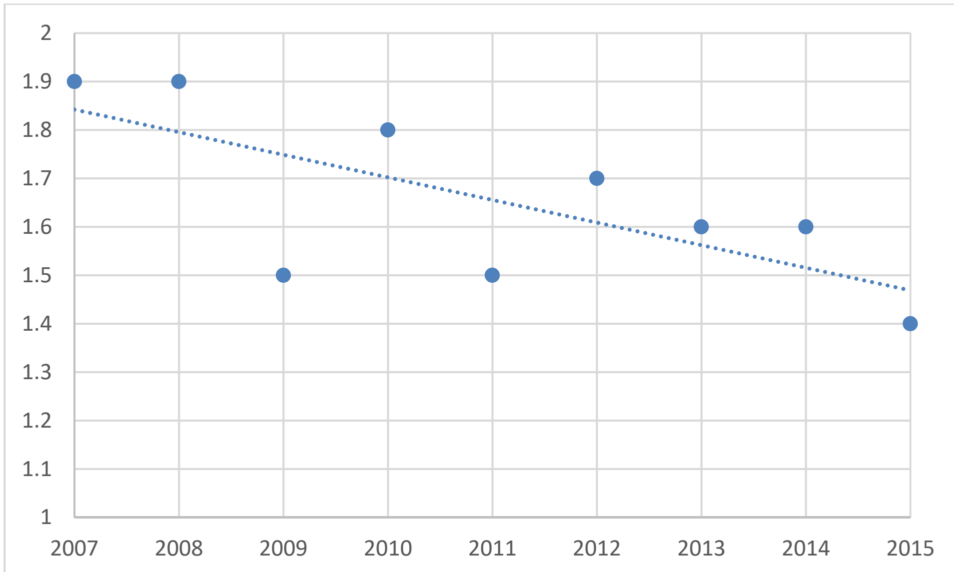
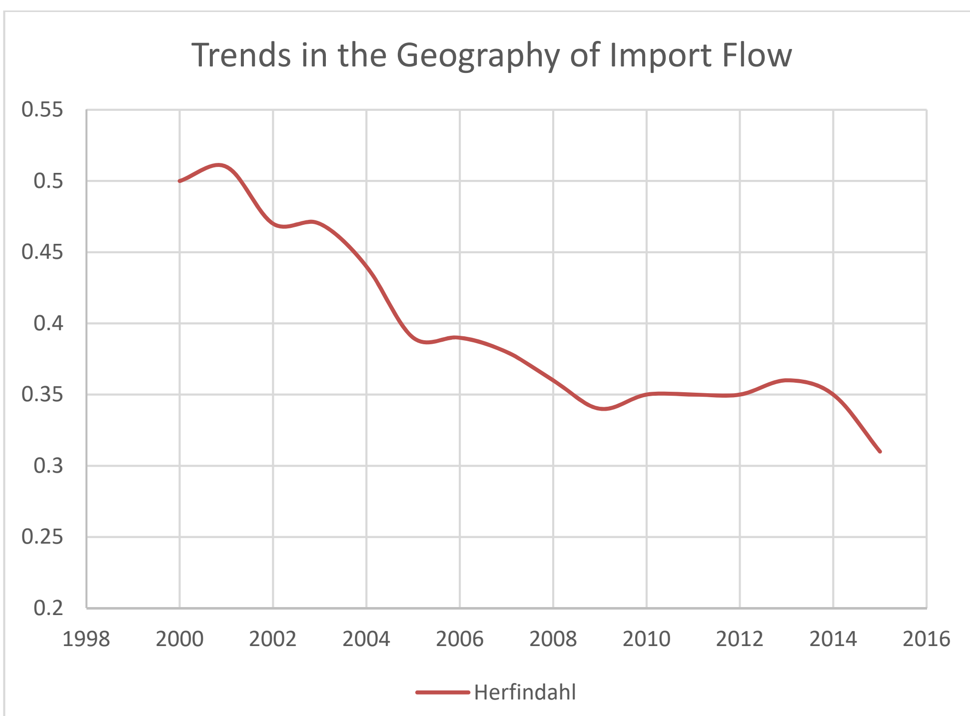


Figure 5
Herfindahl Index of Import Shares across Port Locations, 2000-2015



‘’

Figure 6(a)
Shenzhen Histogram of Log Walmart Shipment Volumes: Data (Green)

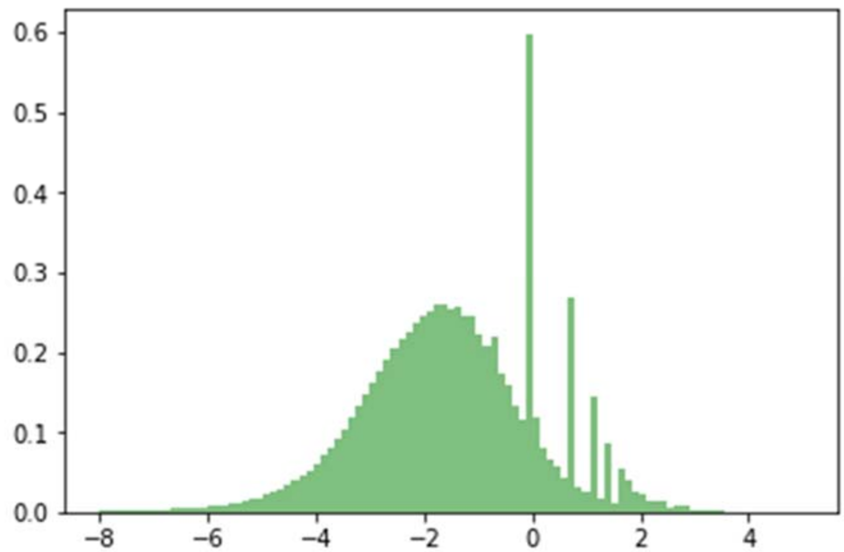


Figure 6(b)
Shenzhen Histogram of Log Walmart Shipment Volumes: Model (Blue)

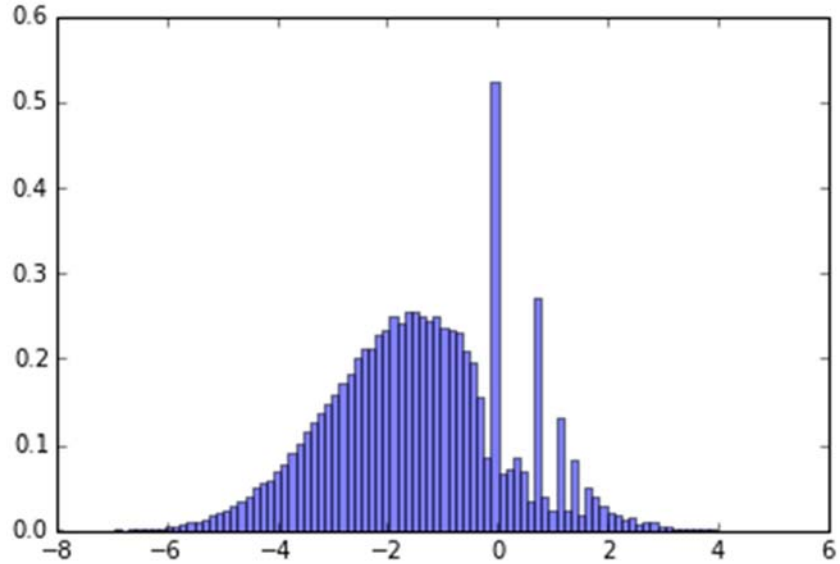


Figure 6(c)
Shenzhen Histogram of Log Walmart Shipment Volumes: Model (Blue), Data (Green)

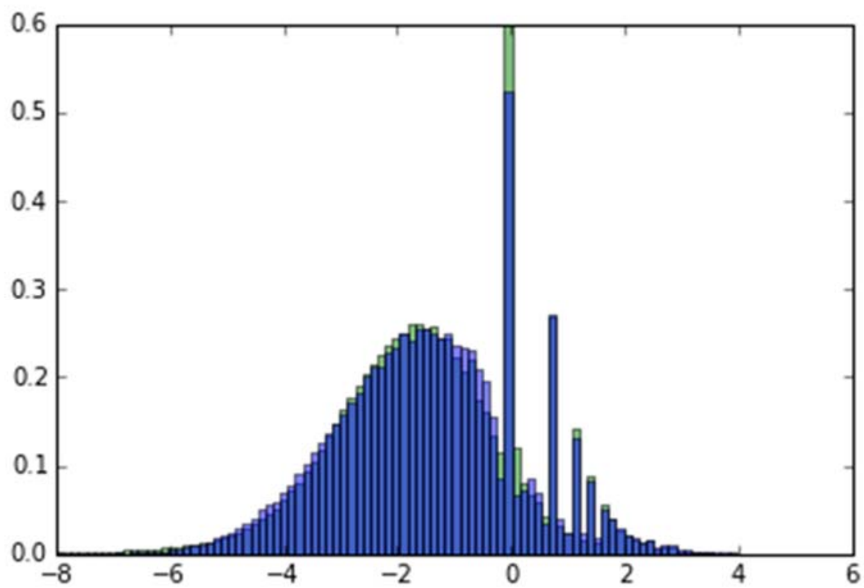


Figure 7(a): Plot of Optimal Adjusted Shipment
Given Target x and Low Adjustment Cost Draw

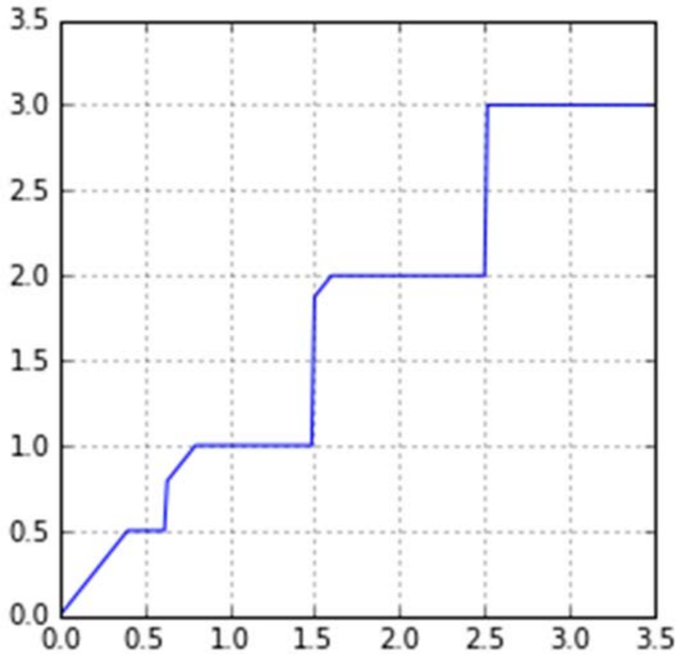


Figure 7(b): Probability of Shipment Consolidation
Given Target Size x and Low Adjustment Cost Draw

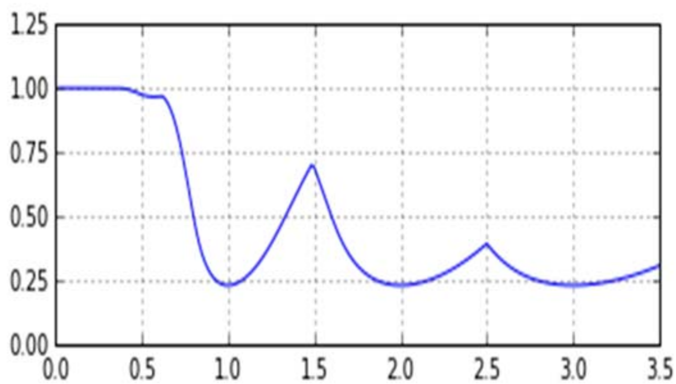


Figure 8
Consolidation Frictions and Market Size
(Horizontal Axis Is Log Container Quantity)

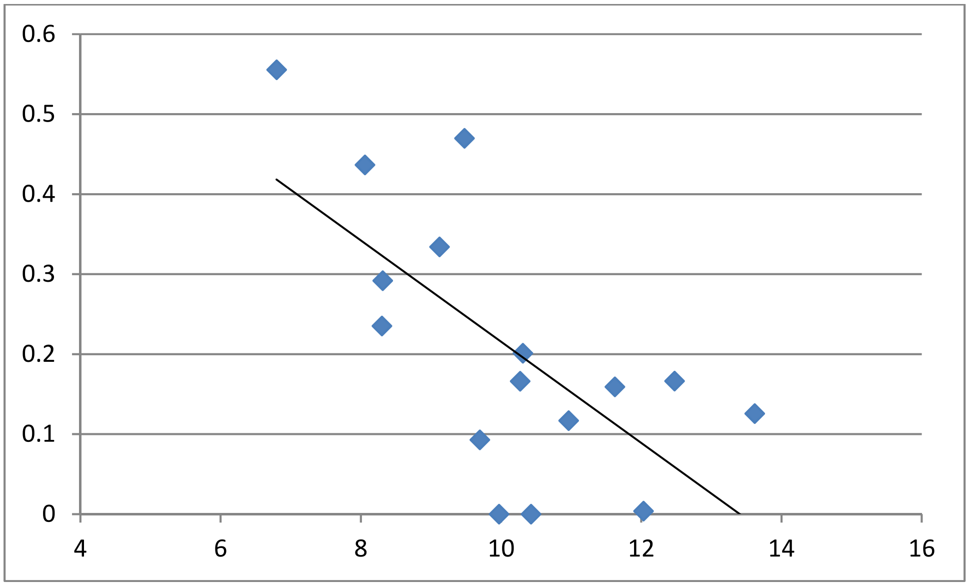


Figure 9
Walmart Seasonal Pattern out of Shenzhen and Estimated Consolidation Friction

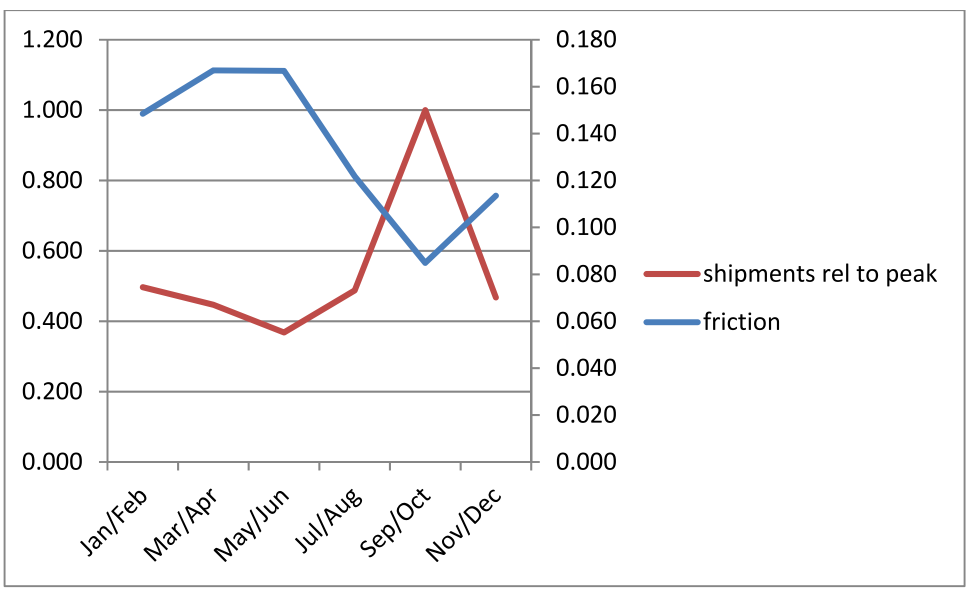


Table 1
Example Bill of Lading

Field Name	Value of Record
Bill of Lading Number	CMDUUH2053195
Shipper	**redacted**
Consignee	**redacted**
Notify Party	Schneider Logistics Attn: Peter Beth 3101 S Packerland Dr Green Bay,WI 54313 Phone: 800- 525-9358 X2244 Fax: 920-403-8627 Tewalmartdray Schneider.Com
Vessel Name	Felixstowe Bridge
Arrival Date	2015-01-07
Place of Receipt	Zhongshan,
Foreign Port	57067 - Chiwan, China
US Port	5301 - Houston, Texas
Container ID Number	CMAU5601550, CMAU4618671, …
Piece_Count	640, 640, …(each container)
Products	5120 Pcs Hb 1.1 **Cu.Ft. Digital Mwo Blk(Microwave Oven)** Purchase Order Number 0254059971 **ITEM No:550099354** This Shipment Contains No Regulated Wood Packaging Materials Freight Collect Load Type:Cy **GLN: 0078742000008** Department No.: 00014 **HTS:8516500060** …
Marks	To:Walmart Case Identification Number Us Dept 00014 (5 Digits-Counting Leading Zeros) Po 0254059971 Item 550099354 Supplier Stk P100n30als3b

Table 2
Examples of Consolidated Shipments
(Walmart Is Consignee in Each Case)

Panel A: Shipments in Container FCIU8099760 arriving 2007-01-05

Shipment	Shipper	Product Item Number	Piece Count	Volume (Cubic Meters)	Weight (Kg)
1	Buzz Bee Toys	000750151 (Toys: 3 Ball Sport Packt	1214	44	3350
2	Buzz Bee Toys	000722571 (Toys: The 5th Dimension)	62	3	272
3	Buzz Bee Toys	000760687 (Toys: Water Warrier Gremlin)	119	4	428
4	Buzz Bee Toys	000722564 (Kwik Grip XL Blasters.)	77	3	358
Total	4 Shipments	All Toys	1472	54	4408

Panel B: Shipments in Container UGMU8950592 arriving 2007-04-19

Shipment	Shipper	Product Item Number	Piece Count	Volume (Cubic Meters)	Weight (Kg)
1	Hasbro Toy Group	000780246 (Blaster Toy)	262	14	1362
2	Hong Kong City Toys	000755447 (Stuffed Doll Toy)	274	30	2219
3	Cepia LLC	0251741746 (Speed Shark Toy)	665	16	1330
4	Reeves Intl INC	000727763 (Toys: All-American...)	202	7	626
5	Zizzle (HK)	000719000 (Battle Playset)	61	2	238
Total	5 Shipment	All Toys	1464	69	5775

Table 3
Sample Statistics
(All statistics in millions)

	Count of Shipments (millions)			Count of Containers (millions)		
	All Sources	From China	From Shenzhen	All Sources	From China	From Shenzhen
9−Year Walmart Sample	2.0	1.7	1.0	1.8	1.6	0.8
18−Month Sample	14.0	6.3	1.6	17.0	7.4	2.0
Beneficial Cargo Owners (BCO)	6.7	2.7	0.9	10.5	3.9	1.2
FF Intermediated (HOUSE)	7.3	3.6	0.7	6.5	3.4	0.8

Table 4
Distribution of Shipments by Consolidated, Single, or Multi for Various Samples

Panel A :Walmart 9-Year Sample for Selected Source Countries

Source Country	Container Imports (millions)	Consolidated Shipment (Percent)	Single Container Shipment (percent)	Multi-Container Shipment (percent)
China	1.57	42.0	8.1	49.9
Bangladesh	0.03	75.3	5.5	19.2
India	0.03	38.2	18.9	42.9
Thailand	0.03	15.5	26.0	58.5
Vietnam	0.03	39.8	13.2	47.0
Rest of World	0.14	30.5	23.7	45.8

Panel B: Selected Large BCO Retailers in 18-Month Sample (China is Source)

Company	Container Imports (millions)	Consolidated Shipment (Percent)	Single Container Shipment (percent)	Multi-Container Shipment (percent)
Walmart	230.5	46.2	8.6	45.2
Target	135.7	40.3	11.2	48.5
K-Mart	61.0	10.2	15.9	73.8
Lowes	61.0	0.0	56.3	43.6
Costco	57.6	0.0	100.0	0.0
Home Depot	44.4	1.3	68.9	29.8

Panel C: FF Intermediated Imports with China Source, by Importing Firm Size Category
(Consolidation Defined as Across Firm)

Size Category	Container Imports (millions)	Consolidated Shipment (Percent)	Single Container Shipment (percent)	Multi-Container Shipment (percent)
All Sizes	2,435.7	4.8	47.7	47.5
By Count of Linked Shipments				
1	103.6	9.0	61.4	29.7
2-4	196.5	7.0	59.0	33.9
5-20	570.1	5.9	53.1	40.9
21-100	927.6	4.7	45.5	49.8
101-250	398.5	2.9	40.7	56.4
251 and above	239.4	1.4	40.1	58.5

Table 5
Alternative Container Sizes and Capacities

Container Size	Maximum Theoretical Volume (cbm)	Maximum Practical Volume (cbm)	Maximum Weight (kg)
20 Foot Half Size	33.2	28	28,200
40 Foot Standard Size	67.7	58	26,200
40 Foot High Cube	76.3	68	26,580

Table 6
Half-Size Shares and Empty Rates for Various Samples of Unconsolidated Shipments

Sample	Half Size Share (percent)	Half-Size Share (percent) (multi half size excluded)	Empty Rate Half Size (percent)	Empty Rate Full Size (percent)
Walmart by Source Country				
China	1.3	0.6	19.6	1.7
Bangladesh	1.0	0.9	18.3	1.6
India	5.0	4.3	39.9	3.6
Thailand	2.2	2.2	41.8	4.5
Vietnam	4.8	4.3	43.6	3.1
Rest of World	5.7	3.9	39.4	5.8
BCO (18 month, China, selected firms)				
Walmart	1.1	0.6	21.2	1.7
Target	0.8	0.3	20.3	0.9
K-Mart	2.0	1.1	25.4	1.4
Lowes	13.6	10.1	14.5	3.2
Costco	7.2	7.2	16.4	2.2
Home Depot	10.1	2.2	33.4	2.5
FF Intermediated, China, by Size Class				
1	39.6	33.2	22.4	6.6
2-4	35.5	27.6	22.3	5.9
5-20	30.9	22.0	23.1	5.3
21-100	24.6	15.9	24.9	5.5
101-250	18.8	10.4	23.5	4.8
251 and above	13.8	7.6	28.1	4.6

Table 7

Estimates of Shipment-Level Model for Various Samples

Panel A: Cross Section of Walmart Source Locations, 2007-2015

Sample	Shipment Count (1,000)	eta	omega1	zeta	mu	sigma	GMM criterion
China							
Shenzhen	1,049	0.126	0.785	0.103	-0.771	1.598	0.006
Shanghai	219	0.166	0.886	0.133	-0.548	1.840	0.027
Xiamen	155	0.004	0.897	0.148	-0.648	2.031	0.044
Ningbo	136	0.159	0.856	0.110	-0.799	1.702	0.016
Qingdao	46	0.117	0.870	0.157	-0.273	1.756	0.044
Hong Kong	43	0.166	0.719	0.124	-0.772	1.511	0.004
Fuzhou	24	-0.157	0.796	0.316	-1.101	2.209	0.021
Tianjin	17	0.201	0.889	0.191	0.192	1.638	0.019
Dalian	6	0.334	0.494	0.161	0.011	1.365	0.071
Foshan	3	0.093	1.000	0.000	1.896	0.992	0.052
Bangladesh							
Chittagong	50	-0.107	0.731	0.189	-0.879	1.681	0.020
India							
Mumbai	20	0.470	0.408	0.007	-0.828	1.328	0.102
Tuticorin	6	0.437	0.680	0.348	-1.285	1.640	0.072
Mundra	2	0.235	0.888	0.135	0.694	1.140	0.257
Ludhiana	2	0.292	0.881	0.137	0.873	1.296	0.145
Chennai	2	0.556	0.624	0.112	-0.755	1.272	0.082

Panel B: Comparison of Target and Costco with 18-Month Sample

18-Month Sample	Shipment Count (1,000)	eta	omega1	zeta	mu	sigma	GMM criterion
Target/Shenzhen	90	0.154	0.942	0.094	-0.715	1.422	0.016
Costco/Shenzhen	19	∞*	1.000	n.a.	0.554	0.202	0.000

*For Costco we take as given that consolidation is infeasible

Panel C: Variation over Time from Fixed Source

Walmart/Shenzen over time	Shipment Count (1,000)	eta	omega1	zeta	mu	sigma	GMM criterion
2007	117	0.148	0.754	0.108	-0.388	1.605	0.003
2015	128	0.095	0.857	0.101	-0.809	1.530	0.014

Table 8
Regression Results: Consolidation Friction for Walmart and Shipping Volume

Parameter	Sample 1 Cross Section of Locations	Sample 2 Average Seasonal (Bimonthly) Shenzhen
Constant	0.838 (0.245)	1.060 (0.318)
Log(Count of Containers)	−0.064 (0.024)	−0.079 (0.027)
R^2	0.337	0.679
N	16	6

Table 9
Estimates for FF Intermediated Shipments
Shenzhen to Los Angeles
By Consignee Firm Size Category

Consignee Size Category (Count of shipment groups)	Total Shipments by Size Category (1,000)	eta	omega	zeta	mu	sigma	GMM criterion	Share $Yes^{col} = 1$	
								Model	Data
1	18	2.15	0.78	0.04	-1.60	1.80	0.24	0.68	0.62
2-4	29	1.92	0.72	0.09	-1.12	1.57	0.39	0.61	0.52
5-20	77	1.56	0.76	0.29	-0.66	1.44	0.49	0.53	0.42
21-100	110	1.01	0.84	0.01	-0.10	1.28	0.45	0.42	0.31
201-250	46	0.91	0.84	0.52	0.23	1.18	0.44	0.33	0.23
251 or more	26	0.69	0.86	0.57	0.18	1.38	0.57	0.41	0.27

Table 10
Estimated Unit Indivisibility Costs by Count of IDCs
(Cost Is Percentage of Ocean Freight)

	Indivisibility Cost (Percent) by Number of IDCs										
	1	2	3	4	5	6	7	8	9	10	20
Shenzhen (fixed η)	6.6	7.9	8.9	9.7	10.3	10.9	11.4	11.9	12.3	12.6	15.1
Shenzhen	2.3	5.4	7.4	9.0	10.3	11.5	12.6	13.5	14.4	15.2	21.2
Mumbai	8.8	14.4	18.8	22.2	25.3	27.8	30.0	32.0	33.8	35.4	45.7
Weighted Averege from Asia	3.1	5.8	7.7	9.1	10.3	11.5	12.5	13.5	14.3	15.1	20.9
Target Shenzhen	4.5	7.7	10.1	12.0	13.8	15.2	16.6	17.8	18.9	19.9	27.1
Costco	0.2	0.9	2.0	3.3	6.3	9.1	8.5	6.9	7.2	10.3	66.2
Walmart No Consolidation	6.4	16.2	27.7	40.1	53.2	66.6	80.4	94.4	108.5	122.6	263.6
Freight Forward Intermediated By Count of Linked Shipments											
1	40.8	64.0	80.3	93.0	103.0	111.4	118.5	124.6	130.0	134.9	164.6
21-100	18.5	33.3	44.1	52.3	58.6	63.7	67.9	71.3	74.2	76.8	90.1
251 and up	14.3	22.9	29.5	34.6	38.8	42.3	45.2	47.7	49.8	51.7	62.7

The cells shaded in gray are the baseline cases that we use for the count *m* of IDCs generating the data. The remaining cells are counterfactual levels of *m*.

Table 11
Estimated Average Inland and Intermodal Transportation Costs by Count of IDCs

	Inland Freight Cost (Percent) by Number of IDCs											
	1	2	3	4	5	6	7	8	9	10	11	12
Walmart	46.2	37.2	34.9	33.5	33.5	32.9	31.2	29.9				
Costco	38.0	29.1	25.9	23.9	22.4	21.5	21.2	21.2	20.5	20.3	20.1	20.1

Notes: the table show estimated average inland freight as a percentage of $3,000, which we take as an approximation of average ocean freight from Asia to a U.S. port. The data appendix provides more details about ocean freight.

Calculations for Walmart assume the following location sequence of IDCs: Los Angeles, Norfolk, Savannah, Houston, Chicago, New Orleans, Philadelphia, Seattle.

Calculations for Costco assume the following location sequence of IDCs: Los Angeles, New York City, Seattle, Savannah, Oakland, Miami, Baltimore, Chicago, Houston, Philadelphia, New Orleans, Norfolk.

Table 12
Unit Indivisibility Costs under Alternative Dissolution and Variety Scenarios
(Based on Weighted Average from Asia Estimates)
(Optimal IDC Choice Highlighted in Bold)

Case	Number of IDCs				
	1	2	3	4	5
Baseline Case	3.1	5.8	7.7	**9.1**	10.3
Horizontal Dissolution					
Into 2 firms	5.8	**9.1**	11.5	13.5	15.1
Into 10 firms	15.1	**20.9**	24.8	27.8	30.1
Into 100 firms	37.5	**44.8**	48.8	51.6	53.7
No-Product Overlap Dissolution					
Into 2 firms	4.8	7.3	**9.2**	10.8	12.1
Into 10 firms	7.7	10.3	**12.3**	14.0	15.5
Into 100 firms	10.2	13.0	**15.3**	17.3	19.1
Expansion of Varieties					
Into 2 varieties	4.0	**7.3**	9.6	11.3	12.8
Into 10 varieties	7.3	**12.0**	15.0	17.2	19.0
Into 100 varieties	12.6	**17.5**	20.3	22.3	23.8

Table A1

Counts for 18-Month Bill of Lading Sample and Distribution by Type

Month	All	Percent Distribution by Type of Bill of Lading			
		BCO	House	Master	No Container or Empty
2007/12	1,020,091	36.8	38.5	21.3	3.4
2008/11	978,676	35.9	39.2	21.7	3.2
2008/12	895,200	36.1	39.0	21.2	3.6
2012/11	1,018,936	35.4	39.1	22.3	3.2
2012/12	1,071,193	34.6	39.9	22.6	2.8
2013/01	1,065,879	36.0	38.5	22.7	2.8
2013/02	1,027,326	35.6	39.0	22.4	2.9
2013/03	903,288	37.9	36.9	21.5	3.7
2013/11	1,085,137	35.3	39.1	22.7	2.8
2013/12	1,043,369	34.2	40.1	23.0	2.7
2014/01	1,154,470	35.3	39.0	23.1	2.6
2014/02	975,268	36.1	38.5	22.4	3.0
2014/03	1,056,633	36.9	37.7	22.2	3.2
2014/11	1,101,479	35.7	39.1	22.5	2.7
2014/12	1,148,196	35.2	39.0	22.6	3.2
2015/01	989,477	35.7	38.7	22.4	3.2
2015/02	999,214	35.7	39.5	21.6	3.1
2015/03	1,275,984	36.1	37.0	24.1	2.8
All Months	18,809,816	35.8	38.8	22.4	3.0

Table A2
Counts of Bills of Lading and Containers in Wal-Mart Sample

	Walmart Sample		PIERS Published Aggregates	
	Count of Shipments (1,000)	Count Containers (1,000)	Count Containers (1,000 FEU*)	Sample Share Relative to PIERS (Percent)
All Years	1,964	1,820	3,267	55.7
By Year				
2007	235	274	360	76.1
2008	202	220	†351	62.6
2009	172	170	342	49.8
2010	185	176	348	50.7
2011	188	167	355	46.9
2012	194	159	360	44.1
2013	245	206	366	56.2
2014	264	219	388	56.5
2015	278	230	‡398	57.7

*FEU is Forty-†Foot Equivalent. PIERS reports units in TEU (Twenty Foot), so figures for PIERS are 0.5 times the published figure.
†The 2008 PIERS figure is based on interpolating 2007 and 2009.
‡The 2015 PIERS figure is not available and is estimated based on trend growth.

Table A3
Import Destination Shares of Walmart's Five Import Distribution Centers
(Apparel and Footwear Excluded)

Import Distribution Center	Share of Shipments	Share of Containers
Los Angeles	20.8	19.5
Chicago	17.9	15.6
Houston	20.4	23.4
Savannah	20.7	23.1
Norfolk	20.3	18.4

Table A4
Freight Charge Statistics
Walmart Shipments from China Matched to Public Census Tabulations

	mean	Percentile p25	p50	p75	N
Walmart Shipments with Exact Match to Public Tabulations					
Freight Charge Per Container	3360.8	2656.0	3121.6	3648.2	483
As Percentage of Entered Value	9.1	6.6	7.6	10.2	483
Walmart Shipments merged on HS6					
Freight Charge as Percentage of Entered Value	7.9	5.8	7.5	10.1	1,134,785

Table A5
Freight Charge Regressions with 18-Month Sample Linked to Census Tabulations
(Left-hand Side Variable is Log Freight Per Container)

	Linked Sample 1		Linked Sample 2	
Parameter	coefficient	s.e.	coefficent	s.e.
Intercept	7.376	0.097	7.120	0.025
Half Size Discount	-0.268	0.031	-0.065	0.004
China to East Coast	0.173	0.014	0.242	0.006
Europe to East Coast	0.117	0.023	0.144	0.006
Europe to West Coast	0.090	0.036	0.211	0.008
Years since Jan 2007	-0.026	0.010	-0.007	0.005
(Years since Jan 2007)^2	0.003	0.001	0.001	0.001
Entered Value (per container)	0.064	0.009	0.063	0.002
BCO	-0.159	0.014	-0.072	0.004
R^2	0.145		0.100	
N	2,507		33,084	

Table A6
Estimated Standard Errors of Model Estimates for Various Samples
(Estimated by Simulation)

Sample	Parameter				
	eta	$omega_1$	zeta	mu	sigma
CN_Shenzhen	0.001	0.006	0.001	0.002	0.002
CN_Shanghai	0.001	0.004	0.002	0.007	0.003
CN_Xiamen	0.002	0.005	0.003	0.006	0.004
CN_Ningbo	0.002	0.007	0.001	0.005	0.006
CN_Qingdao	0.003	0.007	0.003	0.006	0.008
CN_HongKong	0.005	0.012	0.002	0.008	0.009
CN_Fuzhou	0.014	0.015	0.016	0.011	0.011
CN_Tianjin	0.005	0.009	0.006	0.012	0.010
CN_Dalian	0.012	0.020	0.007	0.015	0.015
BD_Bangladesh	0.008	0.015	0.006	0.006	0.007
IN_NhavaSheva	0.001	0.013	0.000	0.010	0.005
IN_Thooth	0.018	0.026	0.027	0.020	0.015
IN_Kachchh	0.010	0.027	0.011	0.024	0.020
IN_Ludhiana	0.016	0.021	0.010	0.022	0.020
IN_Chennai	0.023	0.053	0.031	0.028	0.030
Target_Shenzhen	0.003	0.011	0.002	0.004	0.005
Walmart_Shenzhen_2007	0.002	0.009	0.002	0.007	0.004
Walmart_Shenzhen_2015	0.002	0.008	0.001	0.006	0.005
House Size 1	0.035	0.012	0.001	0.021	0.012
House_Size 2-4	0.026	0.008	0.002	0.010	0.006
House_Size 4-20	0.011	0.007	0.007	0.006	0.004
House_Size 21-100	0.007	0.003	0.000	0.003	0.003
House_Size 101-250	0.008	0.004	0.003	0.007	0.003
House_Size 251 plus	0.007	0.006	0.002	0.008	0.005