NBER WORKING PAPER SERIES

INDIVISIBILITIES IN DISTRIBUTION

Thomas J. Holmes
Ethan Singer

Working Paper 24525
http://www.nber.org/papers/w24525

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2018

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, the Federal Reserve System, or the National Bureau of Economic Research.

At least one co-author has disclosed a financial relationship of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w24525.ack

Indivisibilities in Distribution
Thomas J. Holmes and Ethan Singer
NBER Working Paper No. 24525
April 2018
JEL No. D24,F0,L0

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 for which it is possible to observe the contents of individual containers. In the model, a firm is able to adapt to indivisibility constraints by using consolidation strategies, and by determining how dense to make its distribution network in both space and time. Indivisibilities are found to create significant scale economies in distribution.

Thomas J. Holmes
Department of Economics
University of Minnesota
4-101 Hanson Hall
1925 Fourth Street South
Minneapolis, MN 55455
and The Federal Reserve Bank of Minneapolis
and also NBER
holmes@umn.edu

Ethan Singer
Compass Lexecon
200 State Street
9 th Floor
Boston, MA 02109
singe119@umn.edu

1 Introduction

Indivisibilities arise in the distribution sector when, for example, dividing a shipment in half does not 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. When deliveries are infrequent, loads accumulate, and this is one strategy for helping to ensure that boxes ship full. However, it can be a costly strategy: infrequent deliveries make it hard for firms to exploit new information technologies that respond to demand and infrequent deliveries force firms to carry high inventories. An alternative strategy is to expand scale to fill up the boxes, and expand the variety of goods carried that can be consolidated into the boxes. Mass retailers with broad product lines like Walmart and Target do this, and can be viewed as efficient technologies for resolving indivisibilities.

In this paper, we develop and estimate a model of indivisibilities in shipping. We use data on container imports to lay out a set of facts, including that big retailers like Walmart and Target pack shipments fuller compared to smaller firms. We estimate the cost effects of indivisibilities, including frictions that arise when multiple shipments of different goods are consolidated within the same containers. We estimate the rate at which consolidation frictions fall when the volume of operations increases. Higher scale makes it easier to get the right combination of goods that will exactly fill the boxes and makes it easier to synchronize the optimal shipping times of the various goods being consolidated. To identify scale effects on frictions, we consider both seasonal effects on volume, as well as differences in volume across import source locations. 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 to simulate the cost effects of splitting Walmart in half, and obtain relatively sharp bounds on the effects. Specifically, we estimate such dissolution would raise inbound freight costs for imports between 4.0 and 4.7 percent of expenditures on ocean freight, or, alternatively, 0.22 to 0.26 percent of import retail revenue. As Walmart's profit rate is approximately 3 percent of revenue, our estimates show such a dissolution would have a significant impact on Walmart's profit rate.

We also use the analysis to contrast the indivisibility costs of the traditional *bricks and mortar* model of retail with the Amazon *online* model of retail. In the traditional model, regardless of whether a good is low volume or high volume, a good needs to be *in the store* for a consumer to buy it. In Walmart's case, it has to figure out how to keep over 100,000 products continuously on the shelves of over 3,000 stores. In Amazon's case, goods are not on shelves for immediate purchase. To facilitate same day delivery, high volume goods tend to be placed in numerous fulfillment centers near large metro areas. Low volume goods can be held in a few places, ready to be shipped to any consumer in the country with a one or two-day lag. We show that Amazon's import strategy is indeed different and more flexible than Walmart's, and we use the estimated model to quantify what the savings in indivisibility costs would be for Walmart in an ideal case of perfect flexibility.

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, Target, and Amazon, and we are to determine how they consolidate products into containers at the level of the item numbers these firms use for their 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.

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 broad structure of the distribution system. 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 includes the choice of the number and locations of import distribution centers to use for import flows, as well as decision rules determining delivery frequency by product. Here the firm faces a trade-off where the benefits of having more import distribution centers and higher delivery frequencies are offset by higher indivisibility costs. To quantify the trade-offs, we estimate a structural model of the shipment-level decision, and in this way obtain estimates of indivisibility costs. We also obtain direct estimates of ocean and inland freight costs.

We use product-level data to estimate delivery frequency policy rules. When we put all of this together, we are able to bound the cost effects of splitting up Walmart.

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 a 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 are similar to the issues on the back end in many ways. 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) provides 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 for why mass discounters have an advantage at importing high-variety, low-value goods from China.

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. Our analysis sheds light on these shifts.

There is also a literature specifically on the emergence of containerization and its contribution to stimulating increases in globalization. (See Cosar and Demir (2017), Bernhofen, El-Sahli, and Kneller (2016) Rua (2014), Hummels (2007), and Levinson (2006).) 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. Blum et al (2018) is a related paper that incorporates "lumpiness" in shipping, and obtains this lumpiness by putting a fixed order cost in an inventory model. This is different from the mechanism for lumpiness here, which arises from an explicit model of indivisibilities, rather than a fixed cost.

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, 2011). Our modeling approach differs from what researchers do in this literature.

The rest of the paper proceeds as follows. Section 2 presents a simple version of our model to fix ideas. Section 3 describes the data and presents descriptive results. Section 4 presents the theory. Sections 5, 6, and 7 present the main estimates and dissolution counterfactual. Section 8 quantifies cost savings from flexibility. Section 9 concludes.

2 A Simple Example

Before getting into details of the data or the full model, 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) to be delivered to various locations in the home country. The home country is a circle surrounded by an ocean, as illustrated in Figure 1. Consumers are uniformly distributed throughout the circle.

There are I different products, indexed by i, and let Q_i be the total volume of product i distributed across all consumers in the home country over one year. National annual volume Q_i is exogenous.

The firm sets up import distribution centers (IDCs) at port locations on the circumference of the circle, through which imports are funneled through. Let m be the count of IDCs. Fixing m, it will be optimal to equally space the IDCs. Figure 1 illustrates example spacing for different possible counts of IDCs.

There is an internal freight cost of τ to ship one volume unit, per unit distance of inland transportation within the home country. The greater the IDC count m, the lower the average domestic distance shipped. For example, suppose $m = 1$, as in Figure 1(a). Assuming a circle of radius one, the average distance between the single IDC and every point in the circle is 1.131. The mean inland freight cost is then 1.131τ. With $m = 2$, as in Figure 1(b), average domestic freight cost drops to 0.750τ. Let $r^d(m)$ be the average inland freight cost given IDC count m. It is strictly decreasing and convex.

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

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

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. For now, assume consolidation across different goods is not possible, so each product i ships in its own containers.

To model a benefit of frequent deliveries, suppose that if there were no other considerations, it would be optimal to spread out deliveries continuously and uniformly over time. This will not be practical, so instead let frequency f_i be the integer count of deliveries of product i in a year chosen by the firm. Let $\frac{\varphi_i}{f_i}$ be the average penalty, per unit volume consumed, incurred by delivering good i at a rate other than the ideal of complete smoothing over time. We refer to φ_i as the waiting cost parameter for good i. It potentially varies across goods depending on the relative value of the goods and any issues of perishability.

We can summarize the discussion so far by writing the distribution cost of product i, given IDC count m and delivery frequency f_i, as

$$C(m, f_i, Q_i, \varphi_i) = r^d(m)Q_i + \frac{\varphi_i}{f}Q_i + n_i m f_i \kappa^\circ \tag{1}$$

where n_i is the number of containers used in a delivery, satisfying

$$n_i = ceil(\frac{Q_i}{m f_i q^\circ}). \tag{2}$$

The first term is annual domestic freight cost for good i, the second term is the waiting cost, and the third term ocean freight. Note the volume of each shipment of product i equals $x_i \equiv \frac{Q_i}{m f_i}$ (annual national volume divided by the number of deliveries in a year and the number of ports). The order volume x_i is then divided by container capacity q° and rounded up to the nearest integer (the ceiling function) to determine the count n_i of containers per shipment of good i. The firm's problem is to pick m and $(f_1, f_2, ..., f_I)$, all integers, to solve

$$\min_{m, f_1, f_2, ... f_I} \sum_{i=1}^{I} C(m, f_i, Q_i, \varphi_i). \tag{3}$$

Note the constraint that m must be the same across all goods i; in Section 8, we relax this constraint. We can rewrite the firm's problem (3) by breaking it into two stages, where the firm first chooses m, and then chooses f_i for each product i given m. Let $f^*(m, Q_i, \varphi_i)$ be the cost-minimizing delivery frequency for good i, given m. It is straightforward to prove that delivery frequency is *weakly increasing* in the annual volume Q_i of product i. It is also *weakly decreasing* in the IDC count m.

We define the fill rate of product i as follows,

$$fill\_rate_i \equiv \frac{x_i/n_i}{q^\circ},$$

which is loaded volume per container x_i/n_i divided by container capacity. The *empty rate* is one minus the fill rate. In general, the empty rate will be positive because of the indivisibility constraint.

Consider a limiting case where container capacity q° is made arbitrarily small, while the freight rate per unit volume κ°/q° remains fixed. In the limit, indivisibility constraints disappear, and the optimal plan is to deliver continuously over time and use IDCs continuously over space. Suppose instead we are not at this limit and for simplicity assume there is only one product variety, $i = 1$. If Q_1 is very small, the optimal solution is obviously to make one delivery per year to a single IDC. Moreover, with Q_1 very small it will not fill the container, so the empty rate is positive. Next consider the opposite extreme where Q_1 is arbitrarily large. It is immediate that the delivery frequency f_1 and the IDC count m must become arbitrarily large, and that the empty rate must go to zero. That is, with large Q_1, the indivisibility becomes irrelevant.

Finally, we consider the possibility of consolidation. Assume that in an earlier environment, frictions existed precluding the consolidation of different products. Suppose there is a technological advance that makes it possible to frictionlessly consolidate different products. For simplicity, assume the waiting cost is constant across types, $\varphi_i = \varphi$. It is immediate that in the new regime, the firm will ship containers that consolidate all products. The optimal IDC count m^{con} and delivery frequency f^{con} solve

$$\min_{m,f} C(m, f, \sum_{i=1}^{I} Q_i, \varphi).$$

If the total aggregate volume $\sum_{i=1}^{I} Q_i$ is large, the firm will choose high m^{con} and high f^{con}, and container empty rates will be small. As we will see, this is a good approximation 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. We then establish 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. We contrast this with consolidation by intermediaries of shipments from 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 facts about the count of import distribution centers that firms use (variable m in the simple model). Note, we will also develop facts about delivery frequency (variable f in the simple model), but it is convenient to postpone this discussion until Section 6.

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

[2] We obtained raw CBP records from Ealing Market Data Engineering as well as Panjiva Inc.

[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.

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 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.

This microwave example is useful for a brief digression on the topic of freight rates for containers, and the value of the goods inside. As discussed further in Appendix A3, in our sample period it generally costs around \$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 all products imported by Walmart in 2015 arriving by ship. 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. 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.

indivisibility constraint limiting shipments to integers such as 6, 7, or 8 containers is unlikely to be a big issue.[5] 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 half of the containers from China originated in Shenzhen.) Over the 9-year sample, 87 percent of Walmart's shipments (1.7 out of 2.0 million) originated from China. Note that Hong Kong is included as part of China throughout the paper. 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).

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

[5] Only 12 of the 828 containers shipped had less than the full load of 640 microwaves..

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 row 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. Walmart completes its paperwork in a way that is convenient for our purposes. All shipments within a container of a given product (at the item-number level) are generally reported on the same bill of lading. Furthermore, each different product in the container generally gets its own bill of lading. Thus for Walmart, our bill-of-ladings based measure of consolidation captures the actual degree of product consolidation. Appendix A1 provides additional discussion of this point. We can say something about 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 use data from the first 15 months of our sample period, during which Walmart generally did not invoke the option to redact the shipper and consignee. For this sample, we estimate 34 percent of consolidated shipments, on a container-weighted basis, aggregate products from two or more distinct suppliers.

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. 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. After we estimate the model, we will have more to say about Bangladesh and India.

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 retailers for which we found 40,000 or more container imports from China in the sample, which includes Walmart, Target and four others. It is striking how similar Target is to Walmart in its consolidation behavior. 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, when we estimate the model 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 freight-forwarded containers that consolidate shipments going to different consignees is quite low, equaling only 4.8 percent. Even if we weaken the definition of consolidation 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 done by Walmart and Target. 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 more likely to ship multi-container, unconsolidated shipments, compared to single container

[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).

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. (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 a 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 2(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] Also note 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 2(b) is constructed the same way, except we use Walmart's imports from China. Note the dramatic difference in fill levels. There is hardly a hint of the use of any half-size containers. Also, 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

[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.

2(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. Unlike Bangladesh, where virtually all containers are packed in one place (the port of Chittagong), source locations in India are geographically dispersed, which limits the ability to consolidate. From the figure, we can see that Walmart sometimes uses half-size containers for imports from India, and sometimes uses full-size containers shipped with significant empty space.

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. These cross-country differences in Walmart's use of half sizes are little changed if we put in controls for product (2-digit HS level). 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. We define the empty rate for full sizes to be

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

[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.

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's 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 out of 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, on average, half-size containers ship a quarter empty. 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 3 we plot the empty rate over time for Walmart out of China. While noisy, there is a decreasing pattern from around 1.9 percent at the beginning of the sample to 1.4 percent at the end.

3.4 Evidence on Counts of Import Distribution Centers

In the simple model of indivisibilities in Section 2, the optimal number of Import Distribution Centers (IDCs) is higher, the greater the product volume. This relationship is nicely illustrated by the evolution of Walmart's import supply chain. In the 1990s, Walmart's imports from China were relatively small, and it imported goods through one IDC in Savannah. As imports grew, Walmart changed its import strategy. Walmart added Los Angeles and Norfolk in 2000, and Houston and Chicago in 2005/2006 (see MWPVL (2017)).

The expanded count of IDCs that coincided with the explosion in imports from Asia is a more general phenomenon. Target, for example, has four IDCs (Los Angeles, Seattle, Savannah, and Norfolk). In the logistics literature, the strategy of putting four IDCs at approximate "corners" of the U.S., to minimize shipment distances within the U.S., is sometimes called a "four corners" import strategy (see Leachman 2011). The "five corners" strategy of Walmart adds Houston. Note that Walmart has a larger volume than Target, and its choice of five IDCs versus Target's four is consistent with the theory.

[12] See *Cargo From China*.

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 IDCs 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 IDC. 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. In Section 2, we modeled the long-run decision about IDC count and delivery frequency, but we kept the shipment-level decision simple. Here, we develop a richer model of the shipment-level decision.

Formally, let ℓ be an index summarizing all the long-run decisions that the firm has made. For example, in the notation of Section 2, $\ell = (m, f_1, f_2, ...f_I)$, which is the count of IDCs, and the delivery frequency of each product i. We take the long-run decision ℓ as fixed throughout this section.

The choice of ℓ determines the arrival distribution of shipments that will need to go out. Let s index a particular shipment originating at j with destination k, and let x_{sjk} denote an initial targeted volume of the shipment, which is determined by issues separate from indivisibilities. Let $G_{jk}(x|\ell)$ be the cumulative distribution function (cdf) of the target volume x, for goods from origin j to destination k, given the long-run decision ℓ. To see why the cdf G_{jk} would depend on ℓ, observe that if ℓ specifies a high level of delivery frequency and a large number of IDCs, then we would expect to see small values of x, as shipments are 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:

$$c_{con}(x|j,\ell) = (1 + \eta_{j\ell})\lambda x. \tag{4}$$

Interpret parameter λ as the shipping cost in an ideal world with no indivisibilities. We will relate λ to container prices below. The parameter $\eta_{j\ell} \geq 0$ is the *consolidation friction*, which incorporates 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 $\eta_{j\ell}$ would decrease. Finally, note that we allow the friction $\eta_{j\ell}$ to depend upon the source location j as well as the long run decision ℓ. 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 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 from x to y results in a net benefit to the firm (excluding freight cost) equal to

$$b(y, x, \phi) = \alpha(y - x) - \phi x \left(\frac{y - x}{x} \right)^2. \tag{5}$$

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 $\phi \in \{\phi_1, \phi_2, ...\phi_H\}$ and let the probability the firm draws ϕ_h be ω_h.

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

$$\begin{aligned} \tilde{n}^2(y) &= floor(\frac{\tilde{n}^{only1}(y)}{2}) \\ \tilde{n}^1(y) &= \tilde{n}^{only1}(y) - 2\tilde{n}^2(y). \end{aligned} \tag{6}$$

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_h. The optimum adjustment y given the realization of x and ϕ_h solves

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

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 factor.

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

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

where dependence on the long run decision ℓ is implicit. The expected value given x, integrating over the shocks $\boldsymbol{\varepsilon}$ and the draws of ϕ_h, is

$$V^{ind}(x) = \sum_{h=1}^{H} \omega_h \zeta x \ln\left(\exp\left(\frac{-c_{con}(x)}{\zeta x}\right) + \exp\left(\frac{v(x, \phi_h)}{\zeta x}\right)\right). \tag{7}$$

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 (5) 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 the volume to fill 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 s, let yes_s^{con} be an indicator variable for whether or not the shipment is consolidated. Let z_s be the volume of a particular shipment. In the model, if $yes_s^{con} = 1$, then $z_s = x_s$; that is, the shipment that goes out is the initial target level. If instead $yes_s^{con} = 0$, then $z_s = y_s$; that is, the shipment going out is some adjustment from the target level. In cases where $yes_s^{con} = 0$, the count of containers of each type, n_s^1 and n_s^2, is determined by (6). A complete description of observation s is $\{yes_s^{con}, z_s, n_s^1, n_s^2\}$. For constructing moments, it will also be useful to define

$$\tilde{z}_s \equiv 1_{\{yes_s^{con}=1\}} z_s + 1_{\{yes_s^{con}=0\}} \left(n_1 q_1 + n_2 q_2\right),$$

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

To simplify estimation, we make parameter restrictions. First, for a given origin j and destination k, we assume the distribution $G_{jk}(x)$ of the initial target shipment level is

[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.

lognormal, with parameters μ_{jk} and σ_{jk}^2. To motivate this assumption, consider Figure 4(a), which is a histogram of log shipment volume ($\tilde{z}_s$) for Walmart out of Shenzhen, including both consolidated and unconsolidated shipments. We normalize the volume of one full-size container to be one, $q^2 = 1$. According to the model, we see the initial target x_s only in cases where $yes_s^{con} = 1$; otherwise we see y_s, 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_s 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

$$\alpha = \lambda \equiv \frac{\kappa^2}{q^2} = 1, \tag{8}$$

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,. This assumption is motivated by a consideration of the full problem where the firm picks the distribution of arrival sizes of x: If indivisibilities are negligible, then (8) must approximately hold.

Fourth, we assume a two-point distribution for the adjustment cost ϕ_h, with $\phi_2 = \infty$ and $\phi_1 = 2\alpha$. This implies growth for $h = 1$ is $\gamma_1 = 0.25$ with probability ω_1 and growth for $h = 2$ 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, up to the container capacity. 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,

[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.

having recovered the level of the friction for particular sources, in a second stage we estimate 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.	$\Pr(n_s^1 = 1) \times \Pr(z_s \leq q^2, yes_s^{con} = 0)$
2.	$\Pr(n_s^1 = 1) \times \Pr(z_s > q^2, yes_s^{con} = 0)$
3.	$E[\tilde{z}_s] \times \Pr(yes_s^{con} = 0)$
4.	$E[\tilde{z}_s^2] \times \Pr(yes_s^{con} = 0)$
5.	$E[\tilde{z}_s] \times \Pr(yes_s^{con} = 0)$
6.	$E[\tilde{z}_s^2] \times \Pr(yes_s^{con} = 0)$
7.	$E\left[\tilde{z}_s\right]$
8.	$E\left[\tilde{z}_s^2\right]$
9.	$E\left[yes_s^{con} = 0\right]$
10.	$E[yes_s^{con} = 0] \times \Pr\left(z_1 > q^2\right)$
11.	$E[empty\_rate_s] \times \Pr\left(n_s^1 = 0, n_s^2 = 1, yes_s^{con} = 0\right)$
12.	$\Pr\left\{\tilde{z}_s = q^2\right\}$
13.	$\Pr\left\{\tilde{z}_s > q^2\right\}$

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. Appendix A2 shows that the shipment volumes to the five import distribution centers are roughly the same. 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 4(b) is the fitted value of the shipment volume distribution. It closely resembles the data in Figure 4(a). (See Figure 4(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 5(a) and 5(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 5(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 slightly above half size but below full size, it gets rounded up. When x exceeds a full-size load, the policy, essentially, is to round up or round down to the closest full-size load.

Figure 5(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. It hits the minimum again at $x = 2$, at $x = 3$, and so on. The fact 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 6 plots the estimated friction for each sample against the shipment volume (in log scale). There is 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. The friction falls to less than half that level in high-volume locations.

Panel B of Table 7 reports model estimates for Target for originations from Shenzhen. The estimates are remarkably close to the analogous Walmart estimates. This is consistent with our earlier finding that Walmart and Target have similar descriptive statistics. The estimates for Costco are presented right 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. Costco's log mean shipment size of 0.6 is substantially higher than for Walmart or Target, which is approximately -0.7 for both.

Earlier we noted a downward trend in Walmart's empty rate (Figure 3), which suggests the possibility that Walmart has experienced 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, mean shipment size has significantly declined, (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 examine seasonal effects, we divide the year into two-month intervals. Figure 7 plots average import volume from Shenzhen in two-month interval periods. 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 7. Note the clear countercyclical pattern of the friction level. The friction falls to its lowest point of

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

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 6) and the seasons (Figure 7), there is a negative relationship between the 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 lower the consolidation friction, then keeping everything else the same, more advantageous locations will have higher volumes. As a result, 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 $1.0016 = 1 + 0.02 \times 0.08$, which we expect would only have a small effect on sales volume. 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.

[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.

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 Delivery Frequency

The simple model developed in Section 2 allows the firm to choose both delivery frequency f and the count of IDCs m. These decisions interact. In particular, the delivery frequency policy function $f^*(m, Q_i, \varphi_i)$ for product i was found to be weakly decreasing in the IDC count m. In the counterfactual analysis to follow, we need to address how the firm's choice of delivery frequency adjusts under various scenarios. This section provides a brief treatment of the issue.

It is convenient at this point to make explicit distinctions across products. Think of product i as indexing a stock-keeping unit (SKU) or barcode. Assume products are grouped into a product type h, which will take into account various product characteristics such as the relative value of the product as well as perishability. Let h_i be the type of product i. As in Section 2, products also differ by projected annual national sales volume, which we denote Q_i. We assume products have one source location, which is generally the case in our application, and let j be the source of product i. Now write the policy function for the delivery frequency of product i as $f^*(m, Q_i, h_i, \eta_j)$. Here we are assuming that the product type h_i subsumes φ_i. Also we take into account that the consolidation friction η_j at the source location j of product i can potentially affect delivery frequency.

We assume the delivery frequency policy function is homogeneous in the following respect:

$$f^*(m, Q_i, h_i, \eta_j) = f^*(1, h_i, \frac{Q_i}{m}, h_i, \eta_j). \tag{9}$$

[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.

This says that the optimal policy when there are m IDCs and national annual product volume is Q_i is the same as when there is one IDC with a national volume of $\frac{Q^i}{m}$. In the comparison, the total volume to destination IDC is fixed. ($\frac{Q_i}{m}$ units each to m IDCs on the left-hand side, $\frac{Q_i}{m}$ to a single IDC on the right hand side.) Thus, the condition is that the firm cuts up a fixed annual delivery volume into the same number of trips, regardless of the market area served by the destination IDC, holding the product type fixed. Assumption (9) is satisfied, for example, in the simple model of Section 2.

We now turn to the data on delivery frequency. Let Walmart's 9-digit item number correspond to the product classification i in the model. The item's 6-digit HS code is taken to be the product's type h_i. We restrict attention to the primary source locations in China, India, and Bangladesh that we have geocoded. (These are the 16 locations listed in Table 7.) Of the 1.6 million containers in this set, we have item number information for the contents of 1.0 million containers. We classify shipments by item, year shipped, and destination IDC, yielding a sample with 481,665 different item/year/IDC combinations.[19] Define $f_{i,t,k}$ as the number of different days in year t that a shipment containing item i arrived at IDC k. The 50th and 75th percentiles of delivery frequency $f_{i,t,k}$ equal 2 and 3 deliveries a year. Weighted by annual container volume, the corresponding figures are 5 and 10 deliveries a year.

Table 10 reports the regression of $\log f_{i,jk}$ on log annual product volume (in container units) and a time trend. The coefficient on log annual volume is 0.20.[20] We have also run the analogous regression for Amazon, and obtain a similar coefficient on log annual volume. The coefficient 0.03 on the time trend indicates that delivery frequency for Walmart has been increasing at a remarkable annual growth rate of 3 percent. Recall our earlier result that the consolidation rate has increased over time for Walmart and that the consolidation friction has declined. It is straightforward to see with the simple model of Section 2 how increases in the ability to consolidate leads to higher delivery frequencies.

We have the 6-digit HS code for 81 percent of the observations in our sample, and in the second specification we include fixed effects for the HS code. We also include fixed effects for the source location and destination IDC. Inclusion of these fixed effects makes little difference for the volume or trend growth coefficients. In the third specification we

[19] In the final sample, we delete item/year observations with only a single destination IDC. This deletion eliminates 8.5 percent of the containers in the sample. Cases with a single IDC arise in the data both because of measurement error of the item number variable and also because for a limited set of goods (shoes in particular) Walmart uses a single IDC rather than the standard 5-IDC model.

[20] For a given i, t, k, we use as the volume measure total container volume for this i and t of other IDCs besides k. In this way, the right and left hand side are constructed from different records, eliminating a mechanical correlation. Doing it this way turns out to make virtually no different, as to a first approximation, the firm is sending similar amounts of each i and t to each IDC.

take out the source fixed effect and instead include the estimated consolidation friction η for each source, reported earlier in Table 7. As expected, the relationship is negative. We expect the ability to consolidate would be more relevant for low-volume goods compared to high-volume goods, like the microwave example of Table 1. In specification 4, we distinguish high-volume goods (more than 3 containers per year per IDC) from low-volume goods and make the parametric assumption that the consolidation friction only affects the delivery frequency of low-volume goods. The estimated coefficient is -0.593, which is much larger than in specification 3, where the effect is assumed the same across all volume sizes. To provide a sense of magnitudes, the estimated difference in friction between Mumbai and Shenzhen in Table 7 equals $0.344 = 0.470 - 0.126$, which according to specification 4 would be associated with a 20 percent ($-0.024 = 0.593 \times 0.344$) decrease in delivery frequency in Mumbai for low-volume products. Specification 5 adds location fixed effects, and the estimated coefficient is even bigger.

Below we use specification 5 as a first-order approximation to the policy function $f^*(m, h_i, Q_i, \eta_i)$, for $m = 5$.[21] Then we use (9) to predict how delivery frequency varies with counterfactual changes in m.

7 Firm Choice and Counterfactual Policy Analysis

This section provides a quantitative analysis of the firm's choice of IDC count. It also provides estimates of the effects of counterfactual policies on costs. To develop the analysis, we construct a measure of indivisibility cost and use the model estimates to determine how indivisibility cost varies as the IDC count m varies. We also construct an estimate of how freight cost varies with m. An additional component of costs, the waiting cost that depends upon delivery frequency, is not be measured in the analysis. This leads us to develop a bounds approach to quantify choice behavior and to evaluate the welfare effects of policy changes. The bounds turn out to be relatively sharp. The section begins by laying out our empirical methodology and then presents the results.

7.1 Methodology

Formally, let $C^{ind}(m, f, \eta)$ be average cost arising from indivisibility, per unit volume shipped, as a function of the IDC count m, delivery frequency f, and the consolidation friction η. We

[21] We do not use specification 5, because the fixed location effects will absorb effects of η on high volume goods.

derive an expression for this below. For now, we evaluate this for a given source location, and in the quantitative analysis take a weighted average across locations. The delivery frequency variable f is actually a vector that lists the delivery frequency for each individual product shipped from the given source. A change in m affects total volume shipped per IDC, which in turn affects the consolidation friction. Let $\hat{\eta}(m)$ denote the functional dependence of the friction on m. We use the relationship between the friction and volume estimated in Section 5 to pin this down. Optimal delivery frequency $f^*(m, \eta)$ depends upon m and η, and we use the homogeneity assumption in equation (9) and the cross-section relationship between delivery frequency in Section 6 to pin this down. We then define $\hat{C}^{ind}(m)$ to include the direct effects of m as well as the indirect effects through effects on f and η,

$$\hat{C}^{ind}(m) \equiv C^{ind}(m, f^*(m, \hat{\eta}(m)), \hat{\eta}(m)).$$

Let $C^{freight}(m)$ be average freight cost per unit volume shipped, including both ocean and inland freight, *excluding* all costs related to indivisibility, which are absorbed in $\hat{C}^{ind}(m)$. Below we use information about freight rates and store locations to directly estimate $C^{freight}(m)$.

Let $\hat{C}^{wait}(m)$ be average waiting cost per unit volume shipped, as a function of m. The choice of m affects delivery frequency, which in turn affects average waiting cost. This will remain unmeasured in our analysis. We assume the function is weakly increasing and weakly convex.

The firm chooses m to minimize

$$C(m) \equiv \hat{C}^{ind}(m) + C^{freight}(m) + \hat{C}^{wait}(m).$$

Let m^* solve this problem. Consider the pseudo-problem of minimizing just the first two terms,

$$C^{pseudo1}(m) = \hat{C}^{ind}(m) + C^{freight}(m),$$

leaving out waiting cost. This is the first pseudo problem we consider, so we call this pseudo 1. Let $\bar{m}$ minimize $C^{pseudo1}(m)$. It is clear that the solution to the first pseudo problem is an upper bound to the solution of the actual problem, $m^* \leq \bar{m}$.

To construct a lower bound, define

$$\begin{aligned} C^{pseudo2}(m) &= C(m), \text{ if } m = 1, \\ &= C^{Ind}(m, f^*(m-1, \hat{\eta}(m-1)), \hat{\eta}(m)) + C^{freight}(m) + \hat{C}^{wait}(m-1), \text{ for } m > 1 \end{aligned}$$

This is the actual cost at $m = 1$. Otherwise, for $m > 1$, this is the cost when delivery frequency is set at its level for $m - 1$, everything else the same. It is clear that $C(m) \leq C^{pseudo2}(m)$ for $m > 1$, because cost at m can only be lower when the optimal delivery frequency is used instead of what is chosen for $m - 1$. Define $\underline{m}$ as the minimum value of m such that

$$C^{pseudo2}(m + 1) - C(m) \geq 0.$$

We are able to calculate this difference because the waiting cost is the same in both terms and cancels out. It is clear that $\underline{m}$ is a lower bound to the solution of the actual problem, $m^* \geq \underline{m}$.

We now turn to the details of how $C^{ind}(m, f, \eta)$, $f^*(m, \eta)$, and $\hat{\eta}(m)$ are calculated. Define the *unit indivisibility cost* for a given shipment target level x to be

$$C^{ind}(x) = \frac{-V^{ind}(x)}{-V^{no\_ind}(x)} - 1, \tag{10}$$

where $V^{ind}(x)$ is the maximized value at x, given indivisibilities (see equation (7)), 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,

$$C^{ind}_{avg} = \frac{\int_0^\infty C^{ind}(x) x dG(x)}{\int_0^\infty x dG(x)}. \tag{11}$$

Let m° denote the choice the firm actually makes, e.g. $m^\circ = 5$ for Walmart, and $m^\circ = 4$ for Target. Let f° be the corresponding delivery frequency policy from a given source location and let η° be the resulting consolidation friction at the source. When we take the parameter estimates of the model for a particular firm and source locations, and use them to calculate C^{ind}_{avg} with equation (11), we obtain an estimate of indivisibility cost at the observed choice, $C^{ind}(m^\circ, f^\circ, \eta^\circ) = C^{ind}_{avg}$.

To determine $C^{ind}(m, f, \eta)$ at alternative evaluation points, we first show how $C^{ind}(m, f, \eta)$

varies with m, holding f and η fixed. We assume throughout the analysis that the firm divides imports exactly m ways across the m IDCs. 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 at this particular time is then xm°. If instead, the firm were to divide shipments m' ways, holding delivery frequency the same, the target shipment for each IDC would equal $x' = x\frac{m^{\circ}}{m'}$, a proportionate shift in the distribution. The assumption that x is log normal is convenient because the mean of $\log(x')$ equals

$$\mu' = \mu^{\circ} + \ln(m^{\circ}) - \ln(m'), \tag{12}$$

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

Next note that if the vector f scales by a proportionate factor, the effect on the mean of $\log x$ is the same as a proportionate increase in m, and again the standard deviation is unchanged. We assume the following functional form for the delivery frequency policy function at level of an individual product i,

$$f_i^*(m,\eta) = c_1(h_i)c_2(\eta)\left(\frac{Q_i}{m}\right)^{\nu},$$

for functions $c_1(h)$ and $c_2(\eta)$, where the scaling is indeed proportional. Here we use the homogeneity assumption (9), and the first-order log-linear approximation of the policy function estimated in Section 6, which gives us an output elasticity estimate equal to $\hat{\nu} = 0.22$. This specification allows for product-type effects (through $c_1(h_i)$), but these drop out through log differencing. We make the functional form assumption $c_2(\eta) = \omega_0 e^{\omega_1 \eta}$, which corresponds to specification 3 in Table 10.[22]

The last step is $\hat{\eta}(m)$. We use the semi-elasticity 0.064 from the cross-sectional relationship in Table 8, and we can write it as

$$\hat{\eta}(m') = \hat{\eta}(m^{\circ}) + 0.064(\ln(m') - \ln(m^{\circ})). \tag{13}$$

[22]The estimate of ω_1 in column (3) of table 10 is -0.133. However, to be conservative we use the higher absolute value of -0.593 from column (4),.which assumes a great response and yields a more conservative bound. The difference is negligible, however.

7.2 Indivisibility Cost Estimates

We use the model estimates in Tables 7 and 9 and the above methodology to estimate unit indivisibility costs and how they vary with m. These estimates correspond to $\hat{C}^{ind}(m)$, defined in the methodological section. Table 11 displays the results. The shaded cells are the values obtained for the baseline IDC count the firm actually uses (i.e., where $m' = m^{\circ}$). All figures are reported as a percent of ocean freight.

The first row is for Walmart out of Shenzhen. At the IDC count of $m = 5$ that Walmart actually uses, estimated unit indivisibility cost is 10.3 percent. 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, which in this case equals 12.6 percent. Second, if the firm does not consolidate and sends a partially loaded container or uses a half size, the unit freight cost is higher than it would be for a completely-loaded, full-size container. 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. Note that the estimated indivisibility cost 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 and distortions in shipment size for unconsolidated shipments.

As we move across the columns, we vary the IDC count m. If Walmart were to use only one IDC for imports out of Shenzhen, the indivisibility cost would fall to only 2.7 percent. If Walmart doubled the count to $m = 10$, the cost would rise to 14.4 percent. Note the calculation takes into account if Walmart were to double m to $m = 10$, it would endogenously reduce delivery frequency, which attenuates the increase in indivisibility cost, compared to what it would be if delivery frequency were held fixed. The calculation also takes into account that doubling the IDCs would half volumes going to any one IDC, which eliminates some of the scale benefits in consolidation.

The next row contains the estimates for Walmart out of Mumbai. The indivisibility cost in the baseline is quite high, 25.3 percent. Given our earlier discussion of the relatively high empty rates out of India and the high use of half-size containers, we expect to find high indivisibility costs from this source.

We assume the firm picks the same m to process imports from all originating locations. To minimize overall costs, the firm must take the volume-weighted average across all source locations. The third row of Table 11 presents the volume-weighted average across all source locations in Asia.[23] The difference between the weighted average case and Shenzhen is quite small. Shenzhen has a large weight in the average, and other leading sources such as Xiamen

[23] 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.

and Chittagong also have very low frictions, which offset the high indivisibility costs from places such as Mumbai.

Next we report the estimates for Target out of Shenzhen. Indivisibility cost is higher for Target than Walmart. At Target's choice of $m = 4$, its indivisibility cost 12.0 percent compared to Walmart's level of 10.3 percent at its choice of $m = 5$. Walmart's container import volume is about 50 percent higher than Target's, and these estimated differences are consistent with Walmart being able to enjoy larger scale economies than Target.

The last set of estimates are for freight-forwarded intermediated shipments, based on the model estimates in Table 9. By definition of the sample, consolidation takes place external to the firm. 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. We continue to allow delivery frequency to vary with volume, using the same elasticity estimated for Walmart. At the bottom of Table 11, 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. If a firm were to choose to increase m above 1, indivisibility cost would become quite high.

7.3 Freight Costs

We now explain our estimates of how freight costs vary with IDC count m. This is the function $C^{freight}(m)$, in the notation from above. We estimate this relationship for Walmart. To determine the ultimate destination of imports, we take data on Walmart store locations and sales, to estimate Walmart's sales share across locations. We 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.

To facilitate comparisons with Table 11, 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 \$2,500 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.

Walmart currently has five IDCs and is planning a sixth at Mobile, Alabama. We constrain Walmart to chose locations among these six sites. We sequence the openings as follows. The cost minimizing location with only one IDC is Los Angeles, and we let this be the first one. The freight-cost minimizing location if we add a second, conditioned on one already at Los Angeles, is Norfolk. Continuing on, the m-th IDC minimizes freight cost given m IDCS and given the locations of the first $m-1$ IDCs. This yields a sequence order: Los Angeles, Norfolk, Savanna, Houston, Chicago, Mobile.

The first row of Table 12 reports the results.[24] Using various data sets on freight rates, and making various assumptions discussed in the appendix, we calculate that with a single IDC at Los Angeles, the total freight, as a percentage of \$2500, equals 155.4 percent. When the second IDC opens at Norfolk, the freight rate drops to 144.6, and the rate falls even more if additional IDCs are opened. We note that by shipping containers to the East Coast instead of the West Coast, additional ocean freight expenses are incurred. These incremental ocean costs are included in our calculations, as explained in the appendix.

7.4 Bounding Optimal Choice

We now take our estimates of indivisibility cost and freight cost and use the methodology developed above to estimate bounds on the firm's choice of the IDC count m. We begin by taking the weighted Asia average indivisibility cost in Table 11 and pasting it in Table 12 directly under $C^{freight}$. We label it $\hat{C}^{ind}$, using the earlier notation. The next row contains the sum $C^{freight} + \hat{C}^{ind}$. The IDC count minimizing this sum is $\bar{m} = 5$. This solves the first pseudo problem that ignores changes in waiting costs caused by changing m (and delivery frequency) and the solution provides an upper bound for the firm's choice, as explained above. Next we use the methodology to calculate a lower bound for m^*, determined by holding fixed waiting times, when evaluating an incremental increase in m. The estimated lower bound is $\underline{m}= 5$. Hence, our bounds are sharp, and the point estimate of the optimal choice is $m^* = 5$, which is consistent with Walmart's behavior.

Walmart uses the same import distribution system across source countries and for this reason we have used the weighted-average-from-Asia indivisibility cost measure. Suppose instead that the firm could customize its system for different sources. In the second to last row of Table 12 we report indivisibility cost from Mumbai and then in the bottom row add in freight. We can see that the upper bound number of IDCs if the firm could customize is only 2 for Mumbai, where the minimized cost level is 161.1. This is 2.8 percentage points lower than the actual level of 163.9.[25] Thus imports from low volume countries are disadvantaged

[24] A technical issue that we do not address is that in estimating indivisibility costs in Table 11, under counterfactual numbers of IDCs, we assume equal divisions across destinations. But in calculating freight costs in Table 12, we do not enforce equal shares across destinations.

[25] Note this is a lower bound for the cost difference because if $m = 2$ were set instead of $m = 5$, delivery frequency would be higher, and therefore waiting cost would be lower.

as the firm designs its distribution system around what works best for imports from China, its high volume source.

We note that various factors have been left out in the analysis. The analysis does not include a fixed cost per IDC. Adding fixed costs into the analysis would tend to decrease the optimal number of IDCs. The analysis does not include the possibility of disruptions, including labor strife at ports or at the distribution centers themselves. Adding the possibility of disruption would tend to increase the optimal number of IDCs.

7.5 The Cost Effect of Dissolution

In this subsection, we determine the cost effect of dissolution. We take as our baseline the weighted-average-from-Asia case, where the firm optimally chooses $m^* = 5$, as just explained. Suppose 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 $x^{dis} = x^{merge}/N$.

We are able to calculate lower and upper bounds for the cost increases resulting from the dissolution. To determine the lower bound, we allow the dissolved firms to respond both in the choice of m and in delivery frequency. We calculate the effect on the sum of indivisibility and freight, $\hat{C}^{ind} + C^{freight}$. The effect on this sum is a lower bound on the cost impact because the dissolved firms also suffer increased waiting costs from less frequent deliveries, which is not included in the calculation. To determine the upper bound, we hold fixed delivery frequency, letting the dissolved firms optimize on m. The cost effect calculated this way is an upper bound on the actual cost effect because dissolved firms can always choose to leave delivery frequency fixed, but instead will optimally adjust the frequency to attenuate the negative cost consequences of the dissolution.

Table 13 reports the results, for three different levels of dissolution, to 2, 10, or 100 firms. Cutting the firm in half is roughly on the order of making Walmart the size of Target in terms of import volume. While dividing the firm one hundred-fold is extreme, the size differences between mass discounters like Walmart, Targets, Home Depot, etc., are extreme compared to the firms they replaced. The table reports upper bounds for the dissolved firms' choices of IDC count m, as well as lower and upper bounds on the cost effects. The estimated bounds are relatively tight. If the firm is cut in half, the range of the estimated bound for the cost increase is between 4.0 and 4.7 percent of ocean freight, which strikes us as large, considering the thin profit margins in the discount retail industry.

8 Indivisibility, Flexibility, and Online Retail

Earlier we calculated a gain to a type of flexibility where Walmart is able to customize its distribution system to the particulars of source locations, i.e., do it one way in Shenzhen, and another in Mumbai. Here we calculate gains from an additional degree of flexibility, where the firm customizes its system at the *individual-product level*, in addition to the source-location level. In Mumbai, for example, such flexibility turns out to be quite valuable in resolving indivisibility issues. As some motivation for the exercise, we provide evidence that Amazon operates at this high level of flexibility, and argue online retail is conducive to such flexibility. To be clear, we are not formally modeling online retail, and in particular, don't say anything about a key indivisibility in online retail related to the last mile of getting goods to the consumer's doorstep.[26] (A truck making a delivery might just as easily deliver to the neighbor.) Rather, we are making a point that costs to the doorstep may be partially offset by advantages of online retail in dealing with indivisibilities at the front end of the distribution system.

Walmart uses a relatively rigid supply chain for general merchandise imports. Imports flowing to stores in western states go through the Los Angeles IDC; imports flowing to mid-west stores go through the Chicago IDC, and so on.[27] The path to a particular store is the same, regardless of the characteristics of the imported good, and regardless of whether the good comes from China India, Bangladesh, etc.[28] In contrast, Amazon's import supply chain is flexible, and different goods, in general, take different paths to the same ultimate destination. Traditional and online retail differ in a fundamental way. Traditional retail imposes a certain symmetry across goods, in that—regardless of whether goods are high volume or low volume—they all need to be stocked on store shelves, ready for consumers to buy them. However, in online retail, high volume and low volume goods can be treated asymmetrically, with high volume goods stocked in a large number of local fulfillment centers, ready for quick delivery, and low volume goods held in a small number of warehouses, available for delivery with a longer lag. Differences in the number of places where goods are stocked naturally relates to indivisibilities, as such differences affect how many ways shipments need to be divided up.

A labor market slowdown at West Coast ports during our sample period provides striking evidence of differences in flexibility between Walmart's and Amazon's import supply chains. Figure 8 plots the average number of days to complete a container ship delivery from China to

[26] See Houde, Newberry, and Seim (2017) for an analysis of this part of Amazon's supply chain.

[27] MWPVL (2009, page 32) maps the territory covered by each IDC.

[28] Again, the supply chain we are describing is for general merchandise imports. Other kinds of imports, such as bananas and footwear, have a different supply chain.

Los Angeles, by month of delivery for 2014 and 2015.[29] The contract covering all West Coast dockworkers expired in July of 2014, and the figure shows delivery times started to rise after that, culminating in a major disturbance in early 2015 when average delivery time spiked to 35 days—compared to the usual average of 17 days. Figure 8 also plots monthly averages of shares of container imports from China that go through Pacific ports, for both Walmart and Amazon. The Walmart calculation excludes containers destined for the Houston IDC, which was set up to be able to alternate between getting goods via the Pacific (through Los Angeles plus rail), or through the Panama Canal to the Gulf of Mexico. The calculation retains goods destined for the Los Angeles and Chicago IDCs, set up to get Pacific deliveries (plus rail to get to Chicago), and goods destined for the Norfolk and Savannah IDCs, set up to get Atlantic deliveries. The figure shows that before the disturbance there is about a 50/50 split between Pacific and Atlantic deliveries. When the disturbance hits, there is very little change in the split. In other words, to get goods to stores in the West and Midwest, Walmart remained committed to the strategy of routing goods through the Los Angeles and Chicago IDCs, tolerating the delays at Pacific ports. This contrasts with Amazon's behavior, which showed great flexibility in avoiding the delays. Before the slowdown, all Amazon imports flowed through Pacific ports. After the contract expiration, Amazon immediately began shifting away from Pacific ports. By the height of the crisis, virtually nothing was coming in through the Pacific. When the crisis was over, it immediately switched back.

Analogous to the way Amazon is flexible over time in response to shocks, it is flexible across goods at a point in time. To show this, we focus on a 24-month period after the slowdown (June 1, 2015 through May 31, 2017), and restrict attention to bills of lading referencing one or more Amazon Standard Identification Numbers (ASINs). (ASINs are the unique product codes used by Amazon.) In the sample, there are 10,600 different ASINs, variously consolidated into 28,544 imported containers. Amazon posts web pages for individual ASINs with information about retail price, weight, and sales rank. We downloaded this information, obtaining matched data for 7,499 ASINs. Some products are imported through only one coast, others come through both. Define yes_both_i to be an indicator variable for whether or not both coasts were used for deliveries in at least one sample month. We regress this on product characteristics in a linear probability model, and

[29] We used the bill of ladings data to select container ships making deliveries to U.S. ports and then merged in GPS-based information from Fleetmon, a vessel tracking service, about exact times of port departure. The sample averages reported are for voyages leaving Shenzhen, Hong Kong, or Shanghai going directly to Los Angeles. Trip length is diffence in time of departure out of Los Angeles and China.

obtain the following result:

$$\begin{array}{ccccccccc} yes\_both_i & = & 0.714 & + & 0.021 & \ln\left(\frac{1}{salesrank}\right) & + & 0.086 & \ln\left(\frac{pounds}{price}\right), \\ & & (0.017) & & (0.002) & & & (0.005) & \end{array}$$

with $R^2 = 0.07$, $N = 7,499$, and standard errors in parentheses. The first regressor is a measure of sales volume. The second, weight per dollar, is a proxy for transportation cost. Increases in both variables are associated with a higher likelihood of using both coasts. The magnitudes of the estimated effects are significant. Evaluated at the mean characteristics, the fitted probability 33.8 percent. A one standard deviation increase in the log inverse sales rank is associated with a 7.0 percentage point increase in the probability. Sales rank is a crude measure for volume, so the estimated relationship likely understates what we would get with using more accurate sales data. A one standard deviation increase in the transportation cost proxy is associated with an increase of 9.5 percentage points. These empirical findings are consistent with the simple model of indivisibility introduced in Section 2. Higher transportation cost goods tend to be imported through both coasts to keep inland transportation cost down. Higher volume goods make it easier to divide up shipments between both coasts without creating indivisibility problems. Finally, note that while by the end of 2015, Amazon was using two coasts to import a third of its products, previously (i.e., early 2014), all goods were imported through only one coast. Amazon grew dramatically in this period (we estimate container import volume grew 72 percent between 2014 and 2015), which the theory suggests would be a reason to start using both coasts for some goods.

Motivated by these findings, we now return to our estimated model of Walmart, and calculate the cost savings if, for each product, the firm can flexibly choose whether to use one port, two ports, etc., rather than have the firm be constrained to make this constant across all goods. Formally, in our earlier analysis we assume that in a first stage, the firm picks a count of import distribution centers m that will apply for all goods, and that results in average indivisibility cost $C^{ind}(m)$, and average domestic freight cost $C^{freight}(m)$. Now we change the model to allow the firm to choose m conditional on the desired shipment size x.[30] Let $C^{ind}(m, x)$ be the average indivisibility cost. The average cost savings from flexibility choosing m for a given size x, compared to fixing $m = 5$ for all x (the optimum

[30] Note that for a given shipment of desired size x, we assume the firm picks m before the realization of the proft shocks ε_{con} and ε_{un} and the adjustment cost shock ϕ.

in the constrained case), is given as follows

$$\Delta^{flex} = \frac{\int \left\{ \min_{m \leq 5} \left[C^{ind}(m, \frac{5}{m}x) + C^{freight}(m) \right] - \left[C^{ind}(5, x) + C^{freight}(5) \right] \right\} x dG(x)}{\int_0^\infty x dG(x)}. \tag{14}$$

Note we are defining x to be the desired shipping size given $m = 5$. If instead $m = 1$, desired shipping size would be $5x$. For simplicity, we hold fixed the delivery frequency for each good. In this way, the above calculation understates the gains from flexibility, because in general, a change in m would induce the firm to choose a complementary change in delivery frequency.[31]

We use the estimated model parameters to calculate Δ^{flex}. For Walmart out of Shenzhen, flexibility reduces distribution cost by 2.3 percent, where again we measure cost changes as a percent of ocean freight. For Walmart out of Mumbai, flexibility leads to a cost reduction of 12.5 percent. This is a surprisingly large effect. The best way to understand this is to recall that in the baseline constrained case, shipments are being divided up five ways, causing big indivisibility problems because shipments divided this way generally don't fill up the boxes, and because consolidation frictions are large. Gaining the flexibility to instead divide things up four, three, two or even one way, goes a long way towards ensuring boxes ship out full.

9 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 good job of fitting observed choices. We also show that supply chain flexibility can lead to substantial cost savings.

[31] The calculation also abstracts from how changes in flexibility affect the consolidation friction, which can go either way.

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, that we obtained from Ealing Market Data Engineering. 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.[32] 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 also developed procedures to find Amazon records and applied the search procedures to extract these records over the period November 2013 through May 2017, excluding April 2014. Records from May 2014 through May 2017 were obtained from Panjiva, Inc.

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

[32] 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.

allocate capacity proportionate to piece count. We also have a cbm measure, but this is missing for many records and is at the shipment-level. In contrast, the piece count variable is available for all records and has the breakdown 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).

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 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.

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.[33]

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).[34] 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

[33] 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.

[34] 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.

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.[35] The implied freight cost of a 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. Inland Freight

Destination locations are defined as the core-based statistical areas (CBSAs) used in the 2010 Census. Rural counties are not included in CBSAs and for these we aggregate using Economic Areas (EAs) as defined by the BEA. (In cases were EAs cross state lines, we break them up to the level of state EAs.) To calculate Walmart sales shares across locations, we begin with the 2015 list of Walmart stores obtained from their web site. We use store-level sales data from Holmes (2006) to estimate a regression model of how location-level sales vary with population and Walmart store counts, and then take fitted values using 2015 store counts.

As the basis of our inland transportation costs, we rely on Leachman (2005), which contains information on shipping costs for both rail and truck on intermodal routes from eight U.S. ports (Charleston, Houston, LA-Long Beach, New York, Norfolk, Oakland, Savannah, and Seattle) to 21 destinations within the U.S.[36]

In order to have a more complete set of costs for the entire U.S. ground transportation network, we supplement the Leachman data with information from other publicly available sources. For domestic freight rates (for both rail and truck), we appeal to worldfreightrates, where we collected costs for 769 domestic shipping routes.

[35] 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 Coast districts: New York, Savannah, Norfolk, and Charleston.

[36] See Leachman (2005), Tables 14 and 18.

Although the Leachman data and the supplemental data are highly correlated for routes in both samples, we regress each of the Leachman measures (truck and rail costs) on each of the supplemental measures to have more comparable estimates to combine into a single set of costs. For the truck cost regression we obtain an estimated intercept of -67 and a slope near one (1.03), implying that the worldfreightrates cost estimates are slightly more expensive than the Leachman estimations. We transform the supplemental data using the regression coefficients so that it conforms with the Leachman data. Similarly, for rail costs, we regress the cost from Leachman on the cost from worldfrieghtrates and an intercept. The estimated intercept is 621 and the estimated slope is 0.35 indicating that the supplemental data has costs that are lower than Leachman's estimates for small costs and larger than Leachman's estimates for higher cost routes. With the combined data, we have information for nearly 800 domestic shipping routes. While this larger data set contains information for the major intermodal routes, we need estimates of costs between the distribution centers and every final destination within the United States. To fill in these unobserved costs we rely on the large geographic scope of the observed data and interpolate the costs giving more weight to observed costs that are closer a destination with unobserved costs.

A5. 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.

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*, 107 (5): 180-85..

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

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.

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.

Blum, Bernardo S, Sebastian Claro, Kunal Dasgupta, and Ignatius Horstmann, (2018), "Inventory Management, Product Quality, and Cross-Country Income Difference," University of Toronto Working Paper.

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

Cosar, A. Kerem and Banu Demir, (2017) "Shipping inside the Box : Containerization and Trade," University of Virginia manuscript.

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.

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. (2011), "Evolving Strategies for Asia—USA Containerized Supply Chains: Implications and Policy Recommendations," Proceedings of the 2011 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 (2009), "Supply Chain Strategy: What are the Leaders Doing," Conference Presentation Slides, http://www.mwpvl.com/MWPVL_International_-_Supply_Chain_Strategy_-_What_Are_the_Leaders_Doing.pdf, accessed April 7, 2018.

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.

Katja Seim, Peter Newberry, Jean-François Houde (2017), Economies of Density in E-Commerce: A Study of Amazon's Fulfillment Center Network, manuscript.

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

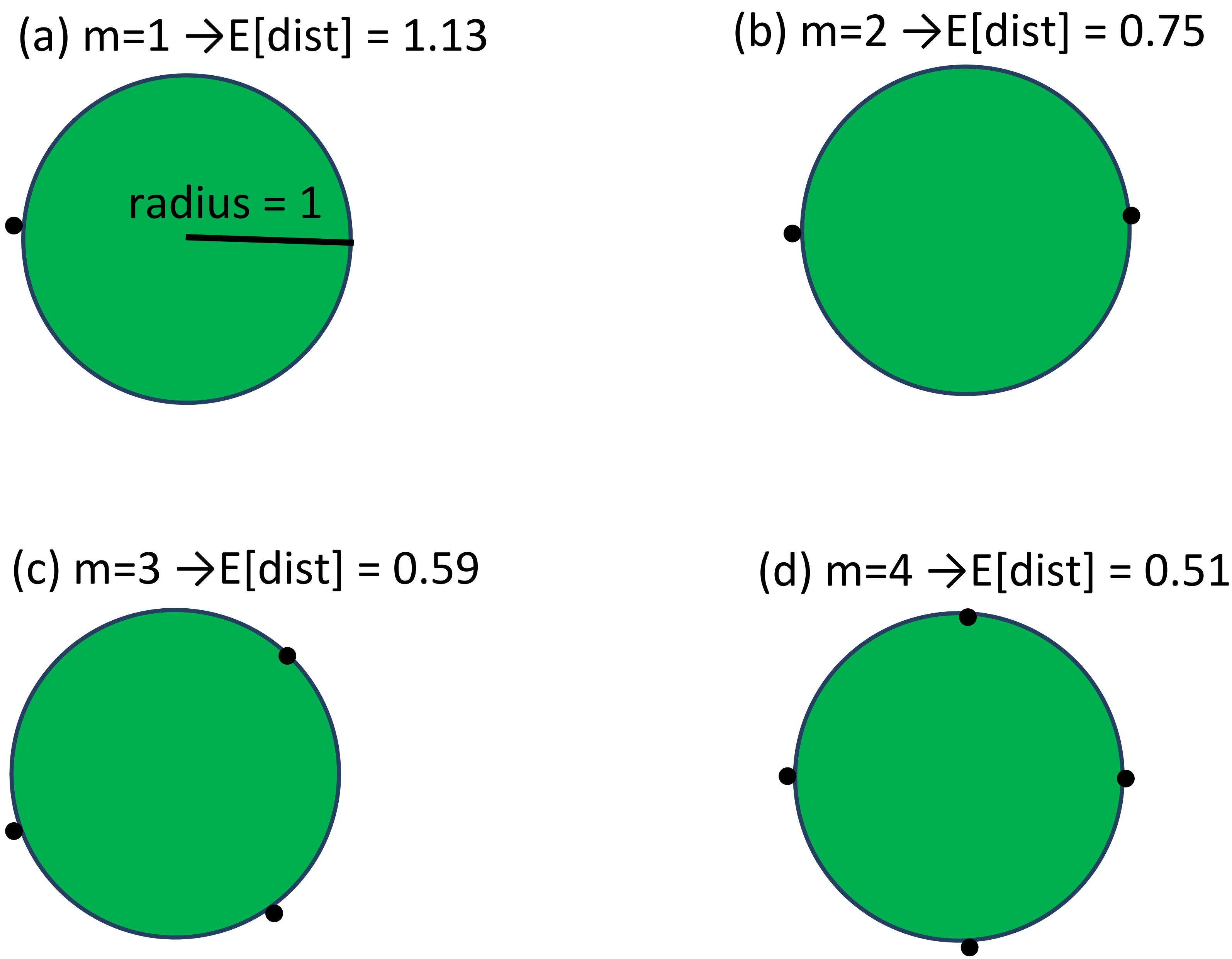
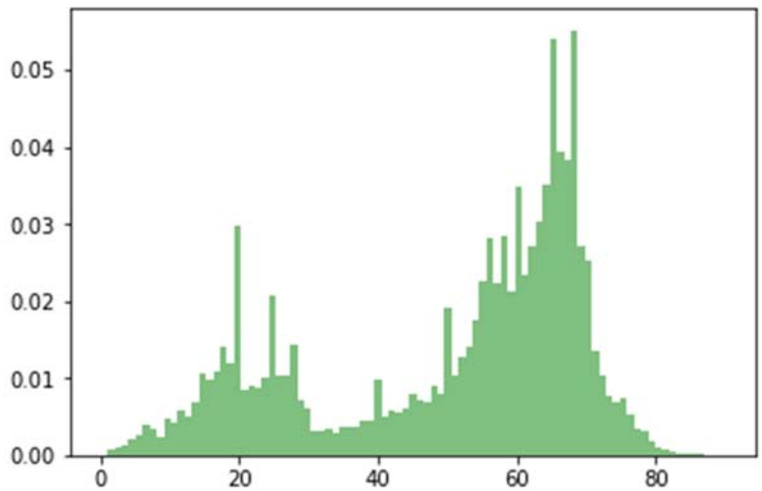
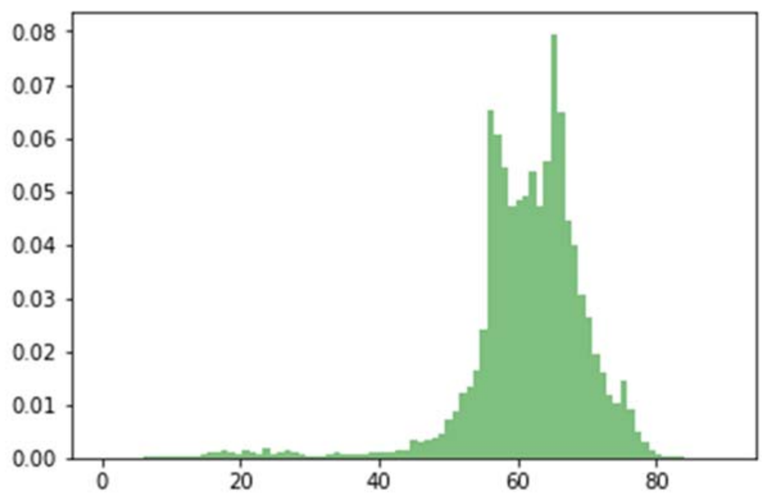


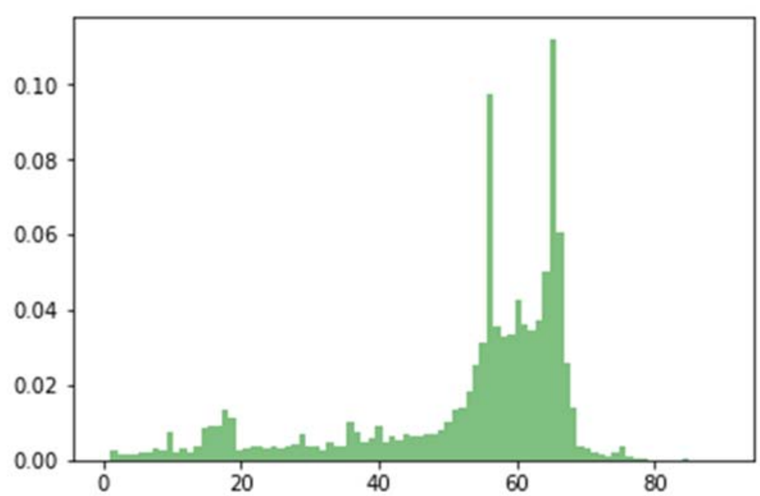
Figure 2. Histograms of Container Fill Levels (Cubic Meters) for Three Samples



(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 3
Empty Rate out of China for Walmart by Year, 2007-2015

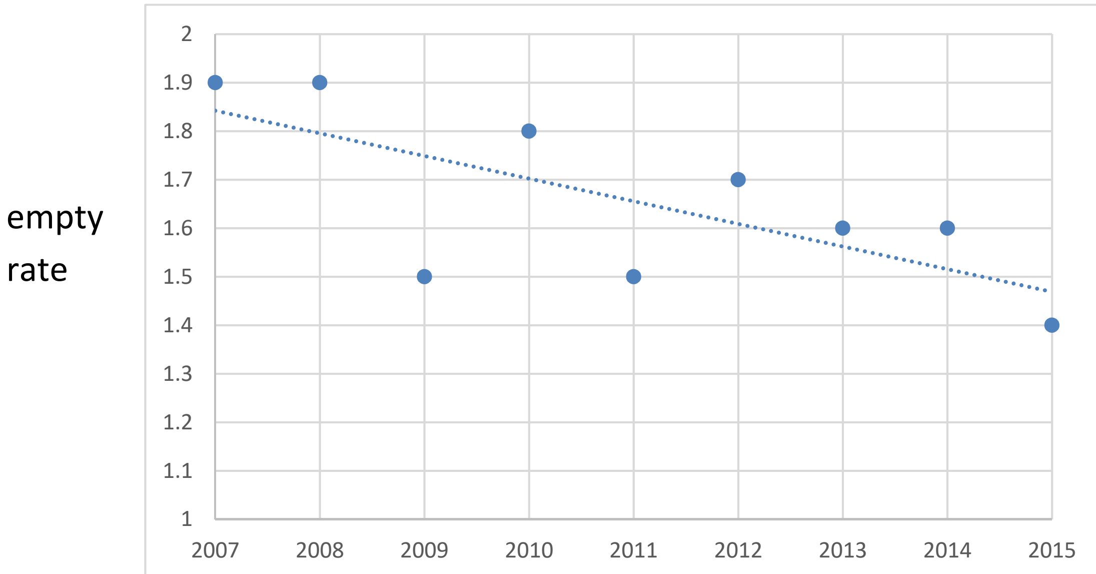


Figure 4(a)
Shenzhen Histogram of Log Walmart Shipment Volumes: Data (Green)

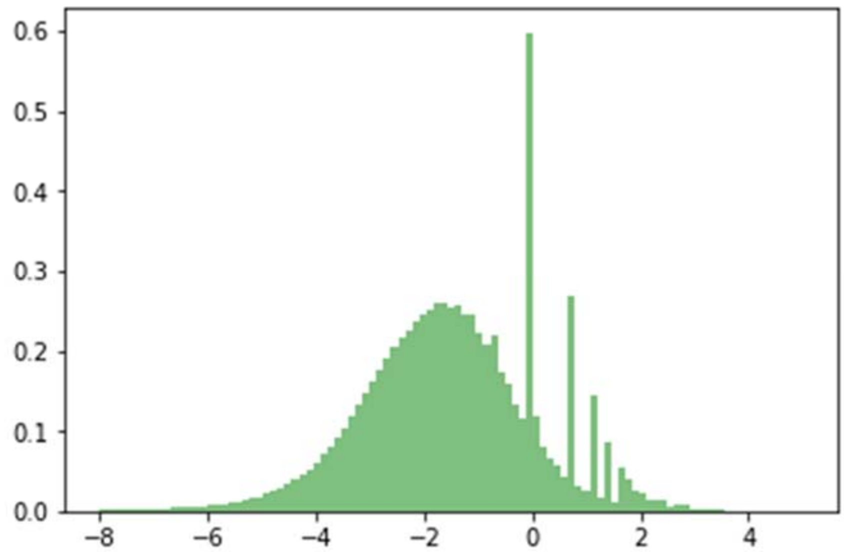


Figure 4(b)
Shenzhen Histogram of Log Walmart Shipment Volumes: Model (Blue)

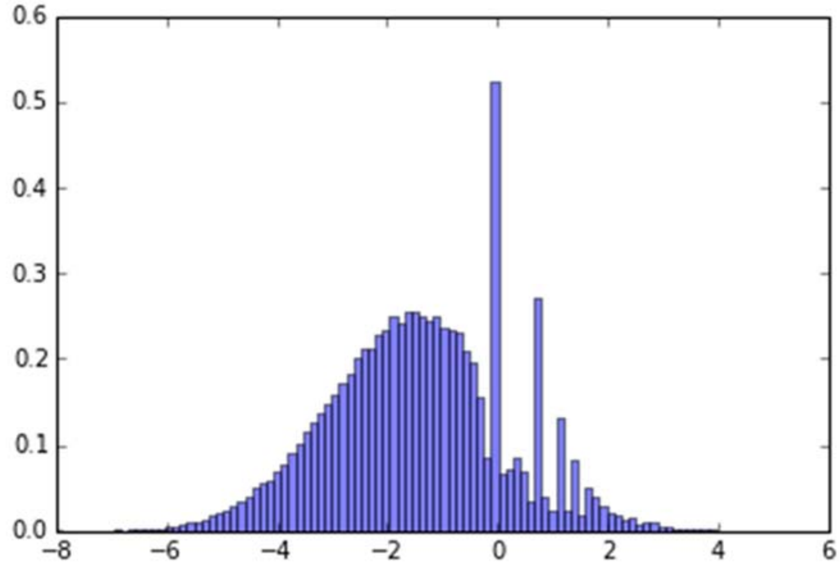


Figure 4(c)
Shenzhen Histogram of Log Walmart Shipment Volumes: Model (Blue), Data (Green)

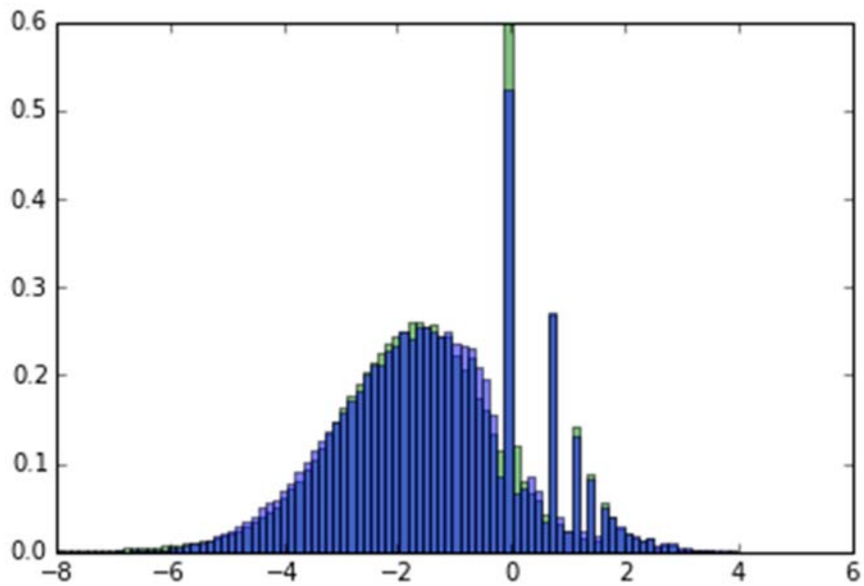


Figure 5(a): Plot of Optimal Adjusted Shipment y
Given Target x and Low Adjustment Cost Draw

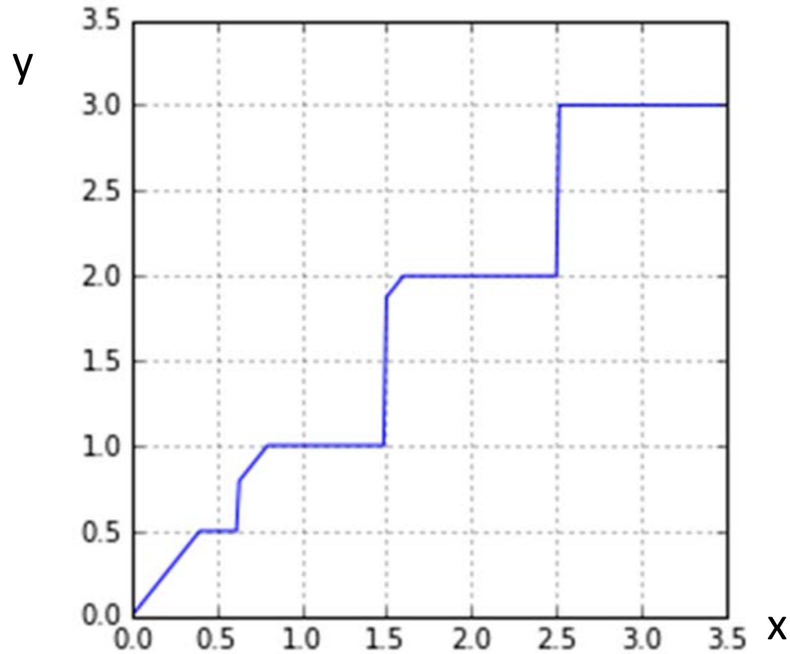


Figure 5(b): Probability of Shipment Consolidation
Given Target Size x and Low Adjustment Cost Draw

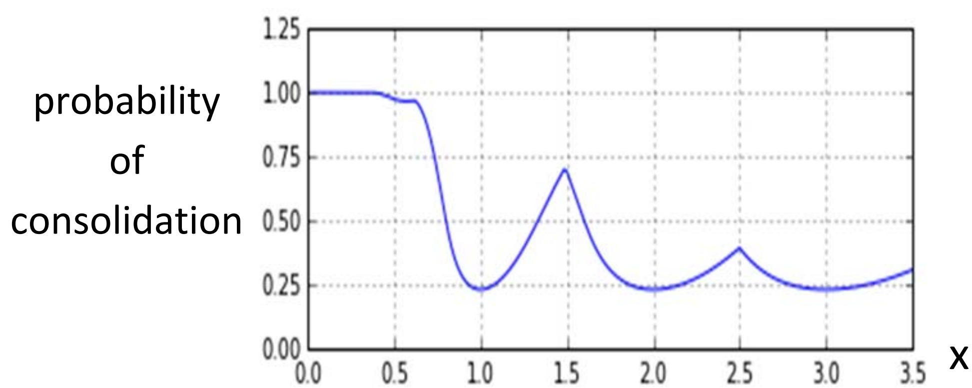


Figure 6
Consolidation Frictions and Market Size
(Horizontal Axis Is Log Container Quantity)

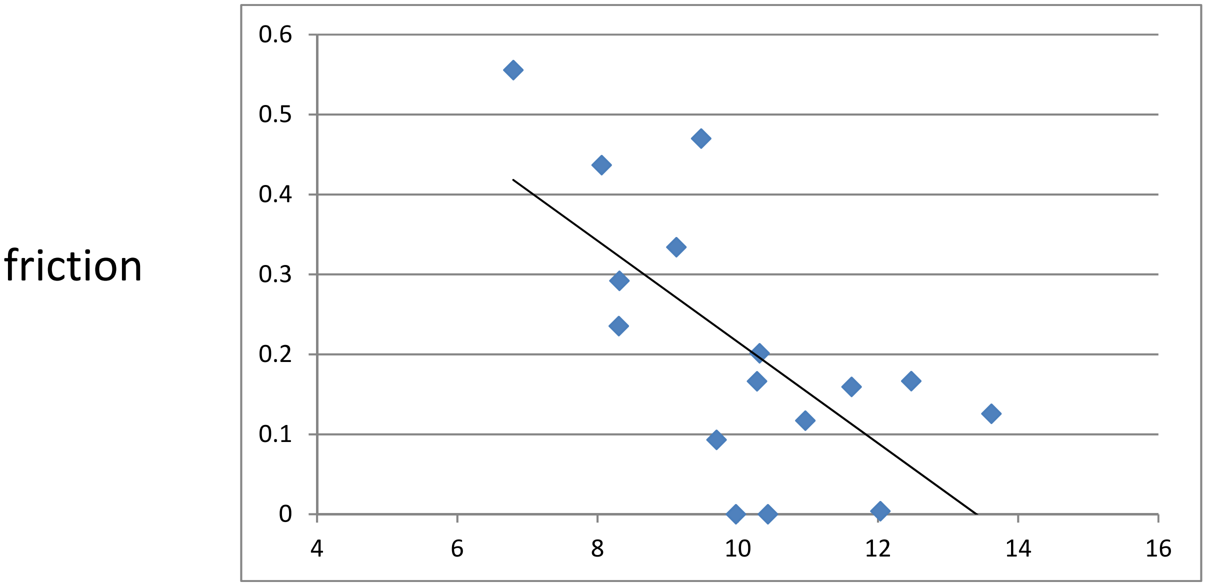


Figure 7
Walmart Seasonal Pattern out of Shenzhen and Estimated Consolidation Friction

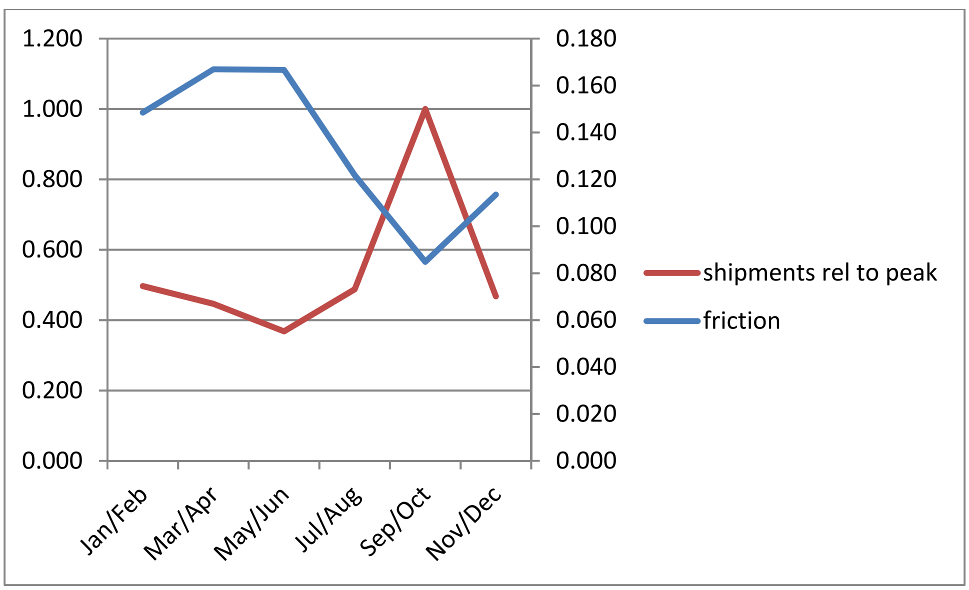
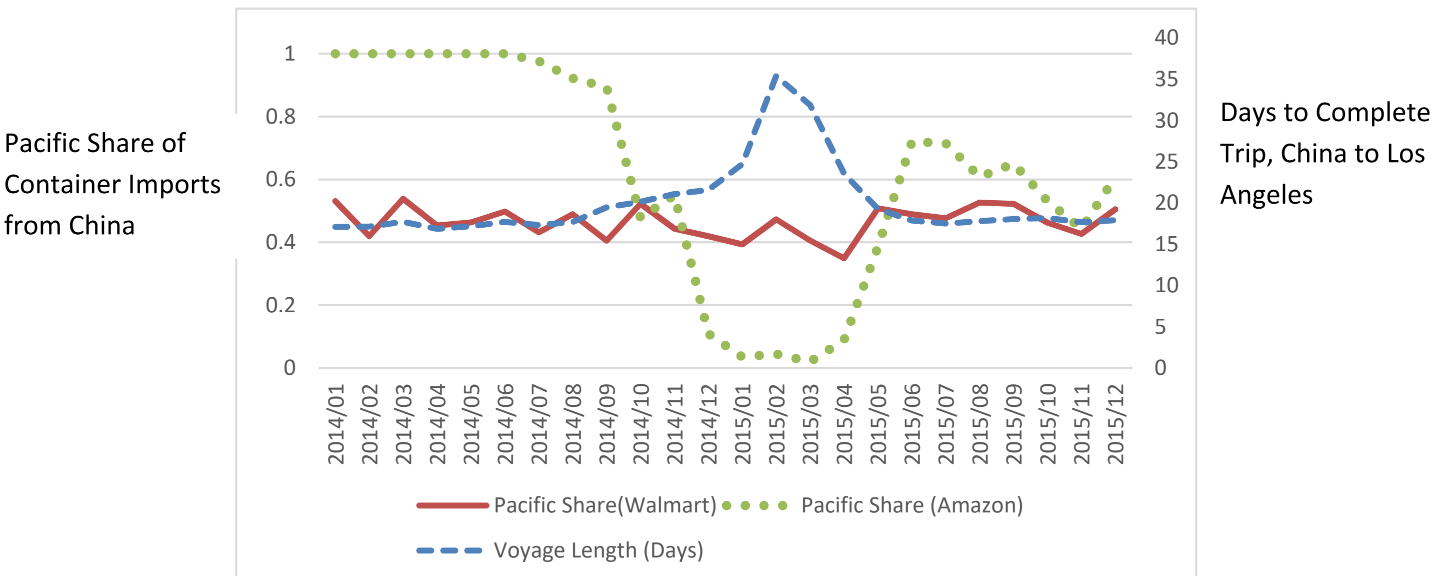


Figure 8

A Labor Market Slowdown at Pacific Ports and Walmart's and Amazon's Responses



Notes to figure: Pacific Share is the share of container imports from China unladed at Pacific ports. For Walmart, containers destined to the Houston IDC are excluded. Voyage Length is average number of days for container ship voyages leaving Shenzhen, Hong Kong, or Shanghai ports and arriving at Los Angeles/Long Beach in the specified month.

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 Shipments	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	1.060
	(0.245)	(0.318)
Log(Count of Containers)	−0.064	−0.079
	(0.024)	(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
The Relationship between Delivery Frequency and Volume Shipped
(Left-hand Side of Regression is Log Delivery Frequency at Item/Year IDC Level)

Variable	Walmart 9-Year Sample				
	1	2	3	4	5
Log Q (Annual Volume)	0.201 (0.001)	0.220 (0.001)	0.218 (0.001)	0.212 (0.001)	0.207 (0.001)
Years Since 2007	0.031 (0.000)	0.032 (0.000)	0.034 (0.000)	0.034 (0.000)	0.031 (0.000)
Location Friction η			−0.133 (0.027)		
Interaction Small Volume and Friction $(Q < 3) \times \eta$				−0.593 (0.022)	−1.149 (0.032)
Source Location FE	no	yes	no	no	yes
Product (HS6) FE	no	yes	yes	yes	yes
Destination IDC FE	no	yes	yes	yes	yes
Amazon Indicator	n.a.	n.a.	n.a.	n.a.	n.a.
R^2	0.191	0.336	0.332	0.333	0.334
N	481,665	391,597	391,597	391,597	391,597

Table 11
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
Walmart Shenzhen	2.7	5.8	7.7	9.2	10.3	11.3	12.2	13.1	13.8	14.4
Walmart Mumbai	11.5	16.5	20.1	22.9	25.3	27.2	29.0	30.5	31.9	33.3
Walmart Weighted Averege from Asia	3.6	6.3	8.0	9.3	10.3	11.3	12.2	13.0	13.7	14.3
Target Shenzhen	5.0	8.2	10.4	12.0	13.5	14.7	15.8	16.7	17.7	18.5
Freight Forward Intermediated By Count of Linked Shipments										
1	40.8	58.3	70.4	79.6	87.2	93.5	98.8	103.6	107.8	111.5
21-100	18.5	29.6	37.5	43.6	48.5	52.6	56.1	59.0	61.5	63.8
251 and up	14.3	20.7	25.4	29.2	32.2	34.8	37.1	39.1	40.8	42.4

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 12

Freight Cost and Indivisibilty Cost by Count of IDCS

	Count of IDCs					
Cost Measure	1	2	3	4	5	6
$C^{freight}$	155.4	144.6	141.9	140.1	138.6	137.8
C^{ind}(weighted Asia)	3.6	6.3	8.0	9.3	10.3	11.3
$C^{freight\,+}\,C^{ind}$(weighted Asia)	159.0	150.9	149.9	149.4	**148.9**	149.1
C^{ind}(Mumbai)	11.5	16.5	20.1	22.9	25.3	27.2
$C^{freight\,+}\,C^{ind}$(Mumbai)	166.9	**161.1**	162.0	163.0	163.9	165.0

Notes: $C^{freight}$ is estimated average inland freight as a percentage of $2,500, 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.

Table 13

Estimated Effects of Dissolution

(Based on Weighted Average from Asia Estimates)

	Upper Bound	Effect on Total Cost (Percent of Ocean Freight)	
Type of Change	*m*	Lower Bound	Upper Bound
Dissolution 2 firms	5	4.0	4.7
Dissolution 10 firms	2	14.7	17.0
Dissolution 100 firms	2	35.2	40.8

Total cost includes indivisibilty cost, inland freight, and waiting cost from infrequent delivery.

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