Received: 22 November 2021 | Revised: 18 October 2022 | Accepted: 22 December 2022

DOI: 10.1111/roie.12657

ORIGINAL ARTICLE

WILEY

Bill of lading data in international trade research with an application to the COVID-19 pandemic

Aaron Flaaen[1] | **Flora Haberkorn**[2] | **Logan Lewis**[2] | **Anderson Monken**[2] | **Justin Pierce**[1] | **Rosemary Rhodes**[3] | **Madeleine Yi**[2]

[1]Division of Research and Statistics, Federal Reserve Board, Washington, District of Columbia, USA

[2]Division of International Finance, Federal Reserve Board, Washington, District of Columbia, USA

[3]Mathematica Inc., Washington, District of Columbia, USA

Correspondence
Logan Lewis, Federal Reserve Board, Washington, DC, USA.
Email: logantlewis@ltlewis.net

Abstract

We evaluate high-frequency bill of lading data for international trade research. These data offer some advantages over both other publicly accessible trade data and confidential datasets, but they also have drawbacks. We analyze three aspects of trade during the COVID-19 pandemic. First, we show how the high-frequency data capture the within-month collapse of trade between the United States and India that are not observable in official monthly data. Second, we demonstrate how U.S. buyers shifted their purchases across suppliers over time during the recovery. And third, we show how the data can measure vessel delivery bottlenecks in near real time.

KEYWORDS
bill of lading, COVID-19 trade, firm-level trade

JEL CLASSIFICATION
F14, F17, C81

1 | INTRODUCTION

Researchers, policymakers, and firms increasingly turn to nontraditional, administrative, or other so-called "big data" to measure economic activity. These data are often available more quickly and

The views expressed here should not be interpreted as reflecting the views of the Federal Reserve Board of Governors or any other person associated with the Federal Reserve System.

offer a finer level of disaggregation than official statistics, but they can also pose new challenges. Without a proper understanding of features such as conceptual definitions, representativeness, and reporting details, such data can result in improper inference, biased forecasts, or nonreplicable results. This article provides a detailed analysis of the utility of a major source of nontraditional administrative data related to international trade: the shipment-level bill of lading (BoL) data collected by U.S. ports. In doing so, it is intended as a guide for economists and other social scientists considering using BoL data for their research and a source of diagnostic information on the data that would be beyond the scope of most standard research papers.

BoL data have advantages and disadvantages relative to other publicly accessible official data and confidential datasets. In Section 2, we describe the data in detail, and in Section 3, we explore these advantages and disadvantages. In this article, we use S&P Panjiva as our source of BoL data, as they provide both the raw data and also a number of useful derivative variables, including identifiers that allow researchers to longitudinally track firms engaged in international trade. However, as we discuss below, BoL data are available from a variety of data vendors, with the core transaction-level data—which are obtained by those vendors via Freedom of Information Act Requests—common across all platforms.

Advantages of the BoL data include detail, timeliness, and the data's unrestricted nature. Data are available at the shipment level, often with company names for both the shipper (exporter or freight forwarder) and consignee (the importer, person, or firm taking final delivery of the merchandise). They are also available to researchers within weeks, rather than months or years in the case of some detailed confidential data. The ability to access the data outside of restricted environments allows easier merging with other datasets as well as diving into specific case studies that can help illuminate how these shipments work in practice for both researcher and reader. A list of the top 10 U.S. consignees and shippers in these data is simple and illustrative to show (see Tables 4 and 5, below), but doing the same with public U.S. data would be impossible, and with confidential U.S. data, explicitly prohibited.

As with all datasets, BoL data have disadvantages as well. U.S. law restricts public access to bill of lading records to only those shipped via vessel; some countries have broader access, but in this article, we will largely focus on the strengths and weaknesses of U.S. data. In addition, shipment values are missing from BoL data. Quantity measures and descriptions are included, but mapping these to commonly used product classifications and estimating values introduces measurement error. Companies also have the right in U.S. law to redact their name from the records, which can hamper efforts to track supply chains comprehensively.

One of the most novel aspects of these data is information on the shippers (exporters) and consignees (importers) for each shipment. This is unique among publicly available datasets, especially for the United States, where access to and disclosures from the U.S. Census Bureau's confidential Longitudinal Firm Trade Transactions Database (LFTTD) are highly restricted.

In Section 4, we dive into the characteristics of these shippers and consignees to better understand global supply chains. We show how about 60% of U.S. consignees have only one foreign shipper, but that these consignees represent less than 10% of import volumes. We also find that most shipper-consignee pairs ship in only three or fewer months in a year, with a surprisingly small number of pairs shipping every month, though these monthly shippers represent over 50% of trade. We also show how the number of shippers per consignee dropped in early 2020 but recovered in a matter of months.

Finally, we turn to ways in which these data are well-positioned to analyze the striking effects of the COVID-19 pandemic on U.S. trade. The daily frequency of the data show how exports from

India to the United States fell within just a few weeks of the start of the pandemic, and given shipping lags, how that collapse took 5–10 weeks to show up in U.S. import data.[1]

We can also analyze the margins on which imports collapsed and subsequently recovered: the intensive margin of changes for a given consignee-shipper pair, the net extensive margin of entry and exit of consignees, and the switching by a given consignee to a different shipper, a different country, or both. We begin to analyze these margins by focusing on an industry with particularly interesting trade patterns during this period: furniture. After initially plunging in the first half of 2020, demand for durable goods, such as furniture, skyrocketed, and furniture's weight and size tends to preclude shipping by air, making it an ideal case to analyze with BoL data.

We find that, during the initial collapse in trade, the intensive margin accounted for much of the plunge in trade volumes. Then, during the extraordinary rebound in U.S. imports, the intensive margin was most important for the first few quarters, with the extensive margin and switching margin growing in relative importance by the end of 2020. We find that the intensive margin is similarly important in the first few quarters of the recovery in *total* U.S. imports. These results provide important lessons on the limitations of supply chain flexibility in the very short run and the time required to source products from new shippers, or for new importers to enter the market.

The field of international trade is particularly well-situated to benefit from new sources of nontraditional data, such as the port data we examine. Since the pioneering work of Bernard et al. (1995), trade economists have focused on firm-level participation in international trade. Subsequent work by Monarch (2022) and Heise et al. (2019) has exploited information on the timing and frequency of trade transactions. This research has been conducted almost entirely using the confidential data available from the U.S. Census Bureau, which comes with strict access and disclosure restrictions. The administrative data collected at ports offer an alternative data source—albeit one with associated weaknesses—for firm- and transaction-level research, without these restrictions.

Several papers have already used the BoL data in international trade research, though their number is dwarfed by those that have employed the confidential Census Bureau data. In particular, the recent availability of this processed BoL data is enabling researchers to conduct more detailed studies of global trade flows, supply chains, and firm operations. Ganapati et al. (2021) use an extract of BoL data (also from Panjiva) and pair it with with vessel location data derived from transponders used for navigational safety purposes. They use these combined data to present new stylized facts on the shipping network of global trade flows, with corresponding implications for trade costs. Bonfiglioli et al. (2020) use BoL data from PIERS to show that richer countries have higher average sales per firm from two sources of heterogeneity. In related work, Bonfiglioli et al. (2021a) show that market concentration in international trade has fallen overall.[2] In addition, Feenstra and Weinstein (2017) use BoL data to estimate the concentration of exporters to the United States from markets outside of Canada and Mexico.

In addition to the trade literature, BoL data have been combined with financial datasets to yield new insights on the behavior of firms that operate internationally. Jain et al. (2014) construct a novel dataset by combining BoL data with publicly available country-year-level data on business regulations and firm-quarter-level accounting data to evaluate the participation of different firms and sectors in global trade. Jain and Di Andrew (2020) use BoL data to examine the sourcing of different categories of imported goods by firms with global supply chains, exploring the relationship between firms' global sourcing strategy and future profitability.[3] Bruno and Shin (2020) match BoL data from Mexico to financial data from Capital IQ (both available from S&P). They show that when the U.S. dollar appreciates, dollar wholesale-funded banks pare back credit to Mexican exporters, hampering their exports. While all of these papers provide descriptions of BoL data, our contribution is to provide an evaluation of a broader range of aspects of the data to allow prospective researchers to consider whether the data would be appropriate for their research questions.

2 | DATA DESCRIPTION

Bill of lading data from S&P Panjiva—the data provider we use for this analysis—contain over one billion transaction-level records of goods traded across borders, with information including consignees and shippers, product descriptions, quantity, and, in limited cases, estimated values of shipment transactions (in USD). The data provide trade flows across 17-country-level datasets, including Bolivia, Brazil, Chile, China, Colombia, Costa Rica, Ecuador, India, Mexico, Panama, Pakistan, Paraguay, Peru, Sri Lanka, Uruguay, the United States, and Venezuela. For each of these countries, data users are able to observe both imports and exports of goods for all trading partners.

We focus our analysis specifically on U.S. import data and, to a lesser extent, U.S. export data. Panjiva provides transactions since 2007 for imports and since 2009 for exports. U.S. import data are updated several times per week, but U.S. export data updates are typically delayed by a 23-day lag for regulatory reasons (Panjiva).

Panjiva acquires these data by collecting bills of lading from U.S. Customs and Border Protection (CBP), which are freely available under the Freedom of Information Act of 1966 (FOIA). A BoL is a legal document that serves as a record that a shipment has been transported from its origin to its final destination. It also details the contract between the shipper and consignee. Each BoL requires companies to fill out various fields, including shipper/consignee name and address, description of the goods, vessel name, transport company name, ports of lading (loading) and unlading (unloading), weight, quantity, and container information. (See Online Appendix Figures 19 and 20 for the CBP inward (import) and outward (export) cargo declaration forms.)

In addition to providing the raw information collected on BoL, Panjiva generates additional variables that may be of use to researchers. First, Panjiva imputes a standard measure of volume, twenty-foot equivalent units (TEUs), based on existing container information and other shipment characteristics. Second, while BoL forms require product descriptions, they do not collect data on Harmonized System (HS) product codes. Panjiva attempts to assign HS codes to each shipment by searching product descriptions for HS codes that may have been optionally included by shippers and by using a text processing algorithm to translate descriptions to HS codes. Third, Panjiva attempts to provide an estimate of the value of a transaction, since this information is not required in a BoL. As discussed below, these values, which are based on publicly-available average unit values, are only estimates and they are also currently unavailable for most transactions. Fourth, Panjiva also includes a unique company ID variable that can be used to link the trade transactions of some shippers and consignees to their associated companies in other S&P Global datasets, such as S&P Capital IQ. One limitation is that this company ID linking variable only exists for 10%–15% of shippers and consignees in U.S. import data at this time, so not all transactions can be linked to S&P's broader ecosystem of data.

There are other sources of BoL data apart from S&P Panjiva, most notably PIERS.[4] While the raw source data in the form of BoL forms will be the same, each provider will process the data differently, resulting in distinct features and characteristics. Comparing different sources of U.S. BoL data is beyond the scope of this article.

Table 1 lists variable names and descriptions for some of the key variables contained in the Panjiva BoL data, with the top panel reporting raw BoL data variables and the bottom

TABLE 1 U.S. import data description for select variables

Raw variable	Description
arrivaldate	Arrival date of shipment
shpname[a]	Entity resolved name of the shipper
conname[a]	The party to take final delivery of the merchandise
shpmtorigin	Location from which shipment left for the U.S
portoflading	Port of lading
portofunlading	Port of unlading
weightkg	Shipment weight in kilograms
vessel	Name of the vessel that transported the goods
Imputed variable	**Description**
panjivarecordid	Unique Panjiva ID for shipment record
shppanjivaid	Unique Panjiva ID for party acting as shipper
conpanjivaid	Unique Panjiva ID for party acting as consignee
volumeteu	Volume of shipment in TEU
valueofgoodsUSD	Value of goods in USD
hscode	Harmonized Item Description and Coding System (HS)
companyid	Capital IQ company ID

[a] shpname and conname are extracted from "original format" fields that also include their addresses.

panel reporting variables that are imputed by Panjiva. We report the remaining fields in Online Appendix D, Table 7.

The massive size of these datasets combined with continuous updating makes data management a particular challenge. In Online Appendix D, we describe some key technical features of the system we have created at the Federal Reserve Board of Governors to update, store, and query the complete Panjiva BoL data files.

How bill of lading data compare to official data on trade flows

Here, we evaluate how well BoL aggregates align with official public trade aggregates, focusing on a portion of trade for which the two data sources can reasonably be compared. Figure 1 shows two measures of trade volume from BoL data: containers, measured by twenty-foot equivalent units, or TEUs (in blue) and shipments (in red), both normalized so that 2009 = 100. A shipment is the cargo, regardless of size, recorded in a single bill of lading.[5] That TEUs and shipments track one another closely implicitly highlights the stability in the average number of TEUs per shipment. In order to exclude transshipments that ultimately end up in a different country, we limit this analysis (as well as all our further analysis of U.S. import BoL data) to shipments where the consignee country is either listed as the United States or is missing. In addition, while BoL data contain noncontainerized vessel trade, in particular oil imports, these do not have corresponding TEU values and represent relatively few shipments of large value. Therefore, the most relevant publicly available measure of trade flows to compare to our BoL measures is

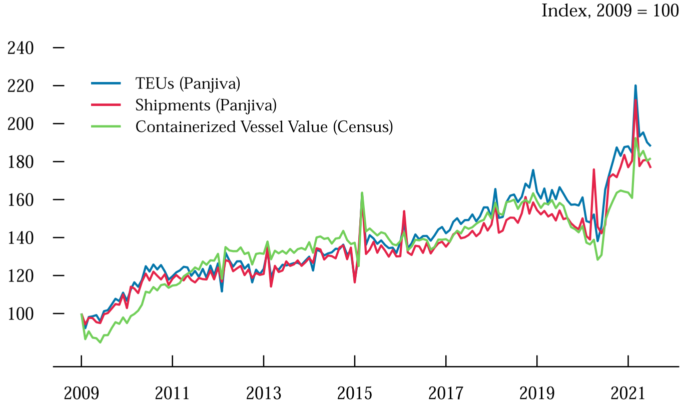

FIGURE 1 Comparison of bill of lading data and Census containerized vessel value for U.S. imports. *Source*: S&P Global Market Intelligence, U.S. Census, and authors' calculations (Seasonally adjusted) [Colour figure can be viewed at wileyonlinelibrary.com]

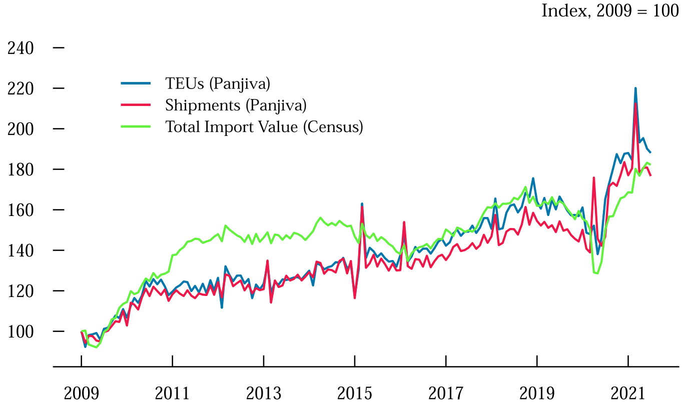

FIGURE 2 Comparison of bill of lading data and Census total value for U.S. goods imports. *Source*: S&P Global Market Intelligence, U.S. Census, and authors' calculations (Seasonally adjusted) [Colour figure can be viewed at wileyonlinelibrary.com]

the containerized vessel import value available from the Census Bureau. Importantly, Figure 1 indicates that this nominal measure aligns quite closely with the BoL volume measures over time.[6] In this sense, the BoL data seem to offer a relatively useful high-frequency indicator for the value of containerized U.S. maritime imports. Of course, the BoL data will be less useful as an indicator for total U.S. imports (or exports) given the omission of trade that occurs via land or air.

Figure 2 shows the comparison between BoL aggregates and total U.S. goods import value. Here, we see that BoL data still capture the broad pattern of trade growth, as well as the dramatic trade collapse and recovery during 2020. Relative to Figure 1, the Census total import value in Figure 2 includes nonmaritime trade as well as vessel trade not via containers, notably including oil imports, which leads to some modest differences with the Panjiva trade measures. Oil prices were elevated in 2011–2014, for example, contributing to the Census total value line being above the lines from BoL data. As we discuss in Section 3.2, the limitation to only maritime trade with U.S. BoL data should be carefully considered in the context of each research question. For example, as we examine the 2020 trade collapse and recovery, we focus on categories like furniture rather than medical equipment or semiconductors, as the latter two categories are more likely to use air shipping.

3 | ADVANTAGES AND LIMITATIONS OF BILL OF LADING DATA

3.1 | Advantages of bill of lading data

Bill of lading data have a variety of advantages relative to official trade statistics, making them a valuable resource for both researchers and policymakers.

The first benefit of these data stems from the fact that shipments are associated with specific firms on both the shipper (exporter) and consignee (importer) side of the transaction. The combination of these data allows consideration of firm characteristics such as the frequency of shipments per consignee, which provides important information about the nature of firms' procurement systems (Heise et al., 2019). We discuss interesting stylized facts based on exploiting the shipper and consignee identifiers in Section 4.

A second benefit of the data is their high frequency. Official trade data are available at a monthly frequency, but BoL data track shipments arriving or departing the U.S. at a daily frequency. This higher frequency is important in many contexts, with one prominent example being an analysis of the timing of the collapse in trade associated with COVID-19. Our examination in Section 5.1 of U.S. imports from India during the initial days of the global pandemic lockdown reveals intra-month shifts in trade that are simply not observable at the monthly level.

A third benefit of the BoL data is the timeliness of their release relative to official data. Official monthly U.S. trade data typically lag the close of a month by more than 30 days. By contrast, and as summarized in Appendix Figure A3, BoL data are updated nearly continuously; data for a particular day are reasonably complete within 10–14 days. This timeliness allows for observation of supply chain disruptions, such as those arising from COVID-19 or the blockage of the Suez Canal in essentially real-time.

A final and intriguing benefit of BoL data is the potential of combining transaction-level data from multiple trading partners. Combining data in this manner opens the possibility of linking shippers and consignees across multiple countries' trade data, allowing for a level of detail on firms' global supply chains that is not available elsewhere. Linking multiple countries' data also holds the potential of observing trade networks (see e.g., Bernard & Moxnes, 2018; Dhyne et al., 2021) and the propagation of supply chain shocks across firms and borders (Boehm et al., 2019).

3.2 | Limitations of bill of lading data

While the advantages for these data relative to publicly available sources can be substantial, there are also limitations about which researchers should be aware. These limitations include missing or redacted data, as well as a general lack of nonimputed data on transaction values.

Limitation to maritime trade (United States)

One of the key limitations of bill of lading data is its lack of information on nonmaritime trade for the United States. As indicated in Figure 3, maritime trade–that is, trade transported by vessel–is the largest mode of transport by value, accounting for nearly 50% of the value of U.S. imports and nearly 40% of the value of U.S. exports in 2019. Nonetheless, the remaining value of U.S. trade, which is split between air and land-based transport like trucks, railroads, and pipelines, is not available in U.S. bill of lading data. Moreover, as shown in Figure 4, the relevance of this exclusion has also grown somewhat over time, with land and air increasing in importance as modes of transportation.

The exclusion of air and land-based trade also leads to substantial differences in coverage across major U.S. trading partners. As shown in Table 2, trade with Mexico and Canada–two of the largest trading partners of the United States–is conducted almost entirely via land-based modes of transportation. Bilateral U.S. trade with those countries, therefore, is largely excluded from BoL data. However, the vessel share of trade is, unsurprisingly, much higher for other important U.S. trading partners outside North America. Trade by vessel accounts for 64% of the value of

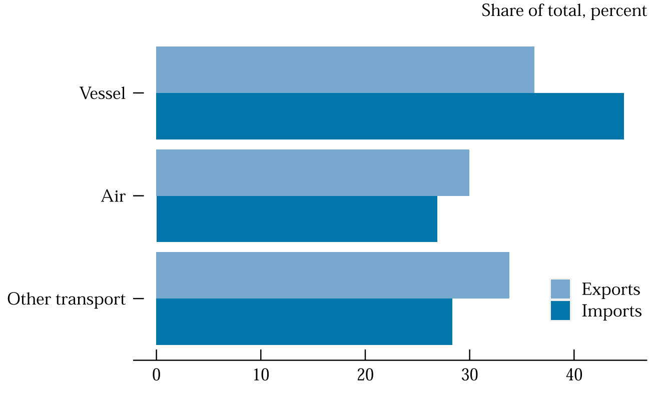

FIGURE 3 U.S. trade shares by mode of transport, 2019. *Source*: U.S. Census and authors' calculations (Other includes rail, vehicle, pipeline and so forth) [Colour figure can be viewed at wileyonlinelibrary.com]

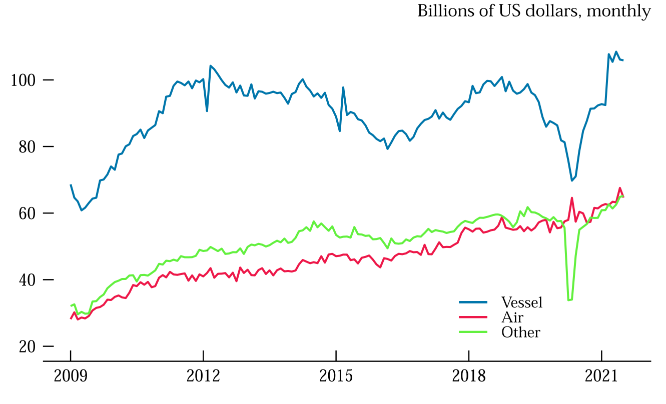

FIGURE 4 U.S. imports by mode of transport. *Source*: U.S. Census and authors' calculations (Other includes rail, vehicle, pipeline and so forth. Seasonally adjusted) [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 2 Trade shares by mode of transport, 2019

	Imports			Exports			
	Vessel	Air	Other	Vessel	Air	Other	Total value[a]
Mexico	9.39	1.98	88.63	12.37	3.50	84.13	608.43
Canada	5.34	4.61	90.05	4.79	6.25	88.96	607.85
China	63.61	28.96	7.43	49.24	43.00	7.76	557.81
Japan	71.80	24.57	3.63	51.79	40.09	8.13	217.77
Germany	51.62	40.39	7.99	33.94	57.75	8.32	187.00

[a]In billions.

Note: Includes top 5 U.S. trading partners by value.

Source: U.S. Census and authors' calculations.

U.S. trade with China, 72% of the value of trade with Japan, and 52% of the value of trade with Germany.

Missing data

Most big data sources suffer from missing information in some observations, and BoL data from Panjiva is no exception. There are two primary sources of missing data in Panjiva: fields for which a firm requests that the U.S. Customs and Border Protection (CBP) redact their identity in the shipper or consignee field, and fields like TEU, HS code, and value that Panjiva imputes from other information that is not always available. Fields that are directly filled in on CBP form 1302 (see Online Appendix Figure 19) are generally available for the vast majority of observations, with the important exception of redacted shippers and consignees.

Table 3 reports the share of U.S. import observations for which particular key variables are missing. As shown in the table, the variables for the shipper/consignee IDs and value have the highest probability of being missing, while the HS code and twenty-foot equivalent unit (TEU) fields are missing in a much lower share of observations.[7] A few key variables, such as weight and shipment origin country, are not included in the table since they have nearly zero missing observations in U.S. import data. Importantly, the share of observations with missing data for particular variables can vary fairly substantially over years. For example, across the years 2007 to 2021, the share of observations with missing shipper (consignee) IDs ranges from 19.9% (9.8%) to 36.1% (36.6%).

Firms' requests for redactions of shipper and consignee information contribute to variation in the share of missing data over time. After a firm requests redaction, this request is fulfilled for two years before requiring renewal. When a request expires, a firm's transactions from that point forward are no longer redacted. These redaction requests must be made for a specific firm name, so firms that use multiple names on bills of lading must submit a request for each entity. Given that one feature Panjiva adds to the raw data is the matching of firm names (including likely typos) to a corporate entity in their overall data framework, this can lead to firms having some but not all of their shipments represented in the database.

One important illustration of this phenomenon is Walmart, Inc., which appears to redact its information incompletely. Like many other companies, Walmart is associated with multiple consignee names on bill of lading forms (e.g., "Walmart", "Wal Mart Stores, Inc.", etc.). As

TABLE 3 Missing U.S. import data by variable (percent)

	Shipper ID	Consignee ID	HS Code	TEU	Value
2007	19.9	16.7	5.0	3.6	100.0
2008	23.7	24.3	3.5	3.8	94.5
2009	29.0	29.1	3.4	3.4	64.9
2010	32.5	33.0	38.7[a]	3.1	64.7
2011	36.1	36.6	3.8	3.3	64.5
2012	33.4	32.6	3.3	3.4	64.6
2013	26.5	10.3	2.6	3.1	64.3
2014	25.0	9.8	2.7	3.0	63.9
2015	25.0	9.9	2.8	3.0	63.7
2016	29.7	12.3	2.8	2.8	64.2
2017	33.4	14.5	2.9	2.8	65.0
2018	33.1	15.1	2.9	2.7	66.1
2019	34.7	15.4	2.8	2.6	64.9
2020	32.9	19.9	3.5	1.8	66.4
2021	34.1	19.3	3.6	1.9	67.5

[a]For HS codes in 2010, Panjiva is aware of a new issue with the data feed.
Source: S&P Global Market Intelligence and authors' calculations.

shown in Figure 5, Walmart's monthly shipments generally hover around zero (as mentioned above, consignee names that should be redacted are sometimes included in the data if they are misspelled on the original bill of lading). Walmart's shipments spike briefly in 2007 and 2012, which suggests that some of Walmart's redaction requests may have briefly expired before they were subsequently renewed. In 2018, Walmart's shipments spike again due to the introduction of the new consignee name "Walmart Inc. Bentonville." Shipments associated with this name then fall to near-zero in mid-2019, suggesting that Walmart made an additional request to redact this version of its name. Due to these redaction requests, users may not be able to track particular companies, which may hinder efforts to track supply chains. In addition, data users should be aware that multiple string names can be associated with a single firm, which adds complications to string-matching or other exercises.

Product descriptions versus product codes

As described above, CBP forms require shippers to report product descriptions, but not HS product codes. The HS codes provided in the data, therefore, are not official HS codes. Rather, in the case of BoL data accessed through Panjiva, HS codes are scraped from product descriptions, when available, and are otherwise assigned based on Panjiva's proprietary algorithm.[8] The assignment is actually quite comprehensive: as indicated in Table 3, the imputed HS code variable is generally well-populated, with 5% or fewer of observations missing for this variable in most years.

Moreover, researchers could consider ways to implement their own HS code assignment algorithm, which could fill in HS codes for some of the remaining shipments with missing values,

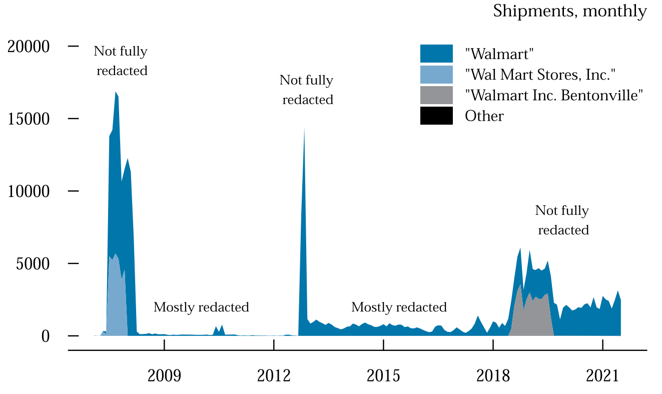

FIGURE 5 Walmart, Inc.'s monthly shipments. *Source*: S&P Global Market Intelligence and authors' calculations [Colour figure can be viewed at wileyonlinelibrary.com]

improve assignments for other observations, and provide a replicable and transparent means of assigning codes. Recent improvements in natural language processing techniques may allow better matches when algorithms are trained using the product descriptions from the U.S. Hamonized Tariff Schedule. Moreover, for assignment of six-digit HS codes, algorithms could employ information from the English-language versions of *other countries'* harmonized schedules.

Even with these potential improvements in HS code assignments, it is important to emphasize that BoL records are based on shipments, and therefore an individual record (and hence unit of quantity) could be comprised of more than one (and often many) individual products. This feature can make disaggregation by product an imperfect exercise.

Limited data on trade values

While BoL data consistently report transaction weights, they typically lack data on the value of trade associated with each transaction, as these data are not collected on U.S. BoL forms. For BoL data accessed via Panjiva, a small share of transactions include data on trade value pulled from the transaction description. However, for the majority of other observations the shipment value is either missing or imputed by applying average unit values from public trade data to the BoL weights. As a result, over 60% of observations have missing data for shipment value.

Researchers may be able to improve on this imputation using information from other datasets. One possibility would be to merge in firm characteristics of shippers or consignees from other datasets (e.g., Compustat), and then consider the extent to which characteristics such as size, ownership, or multinational status are determinants of shipment unit values and values.

3.3 | Comparison of BoL data and confidential census bureau data

For decades, researchers have used the transaction-level LFTTD data from the Census Bureau to examine research questions requiring U.S. firm- or transaction-level international trade data. It is important to note that many of the weaknesses associated with BoL data described above are not present in the Census data, which indicates that for many research applications, Census data will dominate BoL data. In addition to covering all modes of transportation, the LFTTD includes the universe of U.S. international trade transactions, with no ability of firms to request redaction of data. Moreover, the LFTTD contains high-quality longitudinal firm identifiers—which can be linked to other Census Bureau datasets—as well as transaction values and well-populated Harmonized System product codes.

Despite these clear advantages of the LFTTD, BoL data can be useful to researchers examining certain research questions. As mentioned above, BoL data are far more accessible than the LFTTD, which requires a lengthy project application and approval process with the Census Bureau. Similarly, while results using LFTTD data must undergo a disclosure review process to ensure that no confidential information is revealed, users of the BoL data are free to report results involving specific firms. The timeliness of BoL data—which are available within weeks of goods passing through U.S. ports—is even more impressive when compared against the LFTTD, where data become available with a lag of nearly two years. In addition, the ability to track supply chains across multiple countries is possible in BoL data, but not the LFTTD, which only contains bilateral U.S. trade transactions. Lastly, the BoL data contain more complete information on the foreign shipper involved in U.S. import transactions than the LFTTD. In particular, while the LFTTD only includes an alphanumeric code composed of the first few letters of a firm's name and address (see Kamal & Monarch, 2018), the BoL data contain full names and addresses, which can facilitate tracking foreign shippers consistently over time.

A natural question for researchers is whether they should use the BoL data or LFTTD data for their particular research question. This decision will largely be made based on the strengths and weaknesses of each dataset described above, but we offer some additional considerations here. First, researchers should consider the ways in which specific weaknesses of the BoL data might introduce bias into their analysis. For example, company redaction requests might be more likely to focus on new relationships—which might convey trade secrets—rather than long-established relationships that are well-known to competitors. In this sense, the BoL data might yield biased estimates of the extent or timing of the formation of new relationships. One potential way for researchers to evaluate the extent of this bias would be to compare information on relationship formation in BoL data to publicly available data on firm entry or firm participation in international trade available from the Census Bureau. A second potential source of bias could be driven by the lack of data on nonmaritime modes of shipping. If a particular shock causes trade to shift from maritime to air or land transportation—or especially if it causes this shift to occur for particular firms or products—BoL data may yield biased information on the extent of the shift. Researchers could assess the extent of this potential bias by comparing trends in maritime trade with those for other modes of transportation in official U.S. trade data. However, for instances in which researchers or policymakers require more timely data, or data covering trade linkages across multiple countries, LFTTD data will not be useful, and BoL will be the preferred data source. As indicated above, researchers should be aware of the limitations of each dataset and consider how those limitations might affect their specific research question when choosing an appropriate dataset.

4 | CHARACTERISTICS OF SHIPPERS AND CONSIGNEES

One of the most novel aspects of BoL data is the detailed, shipment-level information on shippers and consignees. Subject to the firm-level redactions described above, researchers can track company-specific details over time, including a company's trading partners, its frequency and weight of shipments, its ports of lading and unlading, and even its contact information. In

TABLE 4 Top consignees by total TEU, 2020

Consignee name	Total TEU	TEU (%)	Shipments (%)
Expeditors International	1,145,543	5.37	7.13
Ups Supply Chain Solutions	778,857	3.65	3.17
Dole Fresh Fruit Co.	236,310	1.11	0.59
Chiquita Fresh North America LLC	171,171	0.80	0.12
Maersk Line	170,565	0.80	0.01
Samsung Electronics	167,780	0.79	0.56
Fedex Trade Networks Transport	164,186	0.77	1.00
Seaboard Marine	138,048	0.65	0.02
Geodis USA Inc.	123,523	0.58	0.41
Yusen Logistics (Americas) Inc.	118,793	0.56	0.58

Source: S&P Global Market Intelligence and authors' calculations.

addition, Panjiva assigns unique ID codes to all shippers and consignees after collecting and parsing firm names from bill of lading data, which makes it easier for users to identify and track specific companies as well as merge BoL data with other datasets.

With these data, researchers can analyze certain industries or countries by reporting top suppliers and buyers. Tables 4 and 5, for example, report the top 10 U.S. consignees and foreign shippers, respectively, in U.S. import data. Table 4 reveals that 7 of the top 10 consignees are freight and logistics companies, highlighting the importance of intermediaries in the actual execution of international trade. Table 5 shows that the top 10 foreign shippers to the United States are a mixture of these transportation companies, electronics and agricultural producers, and, improbably, Red Bull. As users of confidential Census Bureau data are well aware, revealing this type of information with those datasets is impossible.

Bill of lading data offer further information on firm-level trade that are unobservable in public official data. As shown in the left panel of Figure 6, the majority of U.S. importers have a single foreign trading partner, but these firms account for a disproportionately small share of total U.S. imports by TEU. By contrast, only a small handful of U.S. importers have many trading partners (over 1000 partners, in some cases), but this small number of firms accounts for a disproportionately large share of imports by TEU. Moreover, the number of shippers and total TEU per consignee are positively correlated. These patterns are largely the same when we switch attention to the number of U.S. consignees per foreign shipper (left panel of Figure 6), and taken together, they highlight the significance of large firms in international trade.[9] In addition, the majority of shipper-consignee pairs interact infrequently in a given year, which emphasizes the lumpiness of trade by pair. For example, in 2019, only about 5% of all long-term shipper-consignee pairs traded at least once each month, while about 50% of all pairs only traded in one or two months of the year (Figure 7).

This shipper-consignee data can also be used to track how disruptions such as recent COVID-related lockdowns affect these relationships. In Figure 8, we plot monthly data on the percent change in the number of shippers per consignee relative to the previous year. As shown in the figure, the number of shippers per U.S. consignee dropped by about 10% in April 2020 relative to April 2019, returning back to pre-COVID levels by later that year. So despite the dramatic

TABLE 5 Top shippers by total TEU, 2020

Shipper name	Country	TEU	TEU (%)	Shipments (%)
Thor Joergensen A S	Denmark	170,414	1.08	0.02
Chiquita Brands International SARL	Switzerland	153,919	0.98	0.12
LG Electronics Inc.	South Korea	82,011	0.52	0.31
Samsung Electronics Co., Ltd.	South Korea	55,531	0.35	0.37
Thai Samsung Electronics Co., Ltd.	Thailand	54,222	0.34	0.21
Samsung Electronics Digital	Mexico	43,335	0.28	0.15
Red Bull GmbH	Austria	42,895	0.27	0.04
Union De Bananeros Ecuatorianos S.A.	Ecuador	36,198	0.23	0.24
Seadom Units	Dominican Republic	35,772	0.23	0.00
Century Distribution Systems	China	35,271	0.22	0.12

Source: S&P Global Market Intelligence and authors' calculations.

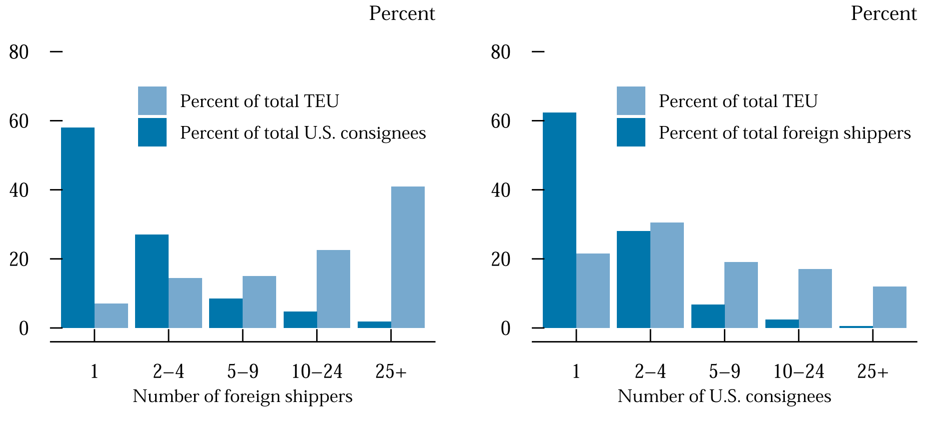

FIGURE 6 Shippers and consignees by TEU, 2019. *Source*: S&P Global Market Intelligence and authors' calculations [Colour figure can be viewed at wileyonlinelibrary.com]

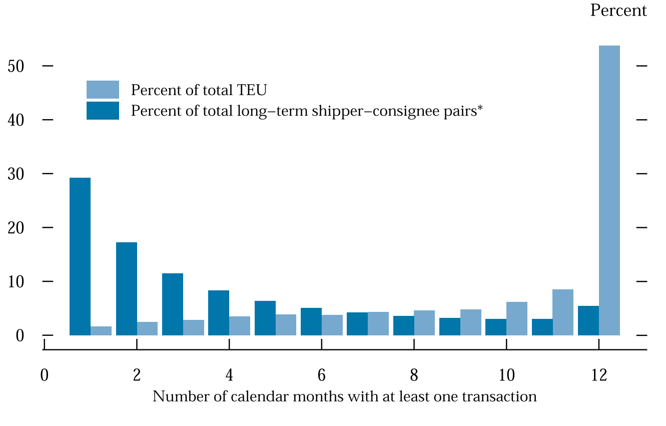

FIGURE 7 Frequency of transactions by shipper-consignee pair, 2019. *Source*: S&P Global Market Intelligence and authors' calculations. *Includes shipper-consignee pairs that traded at least once in the previous year (2018) [Colour figure can be viewed at wileyonlinelibrary.com]

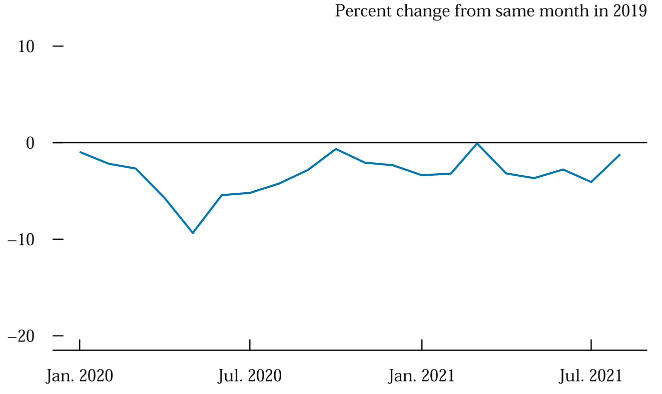

FIGURE 8 Change in shippers per consignee. *Source*: S&P Global Market Intelligence and authors' calculations [Colour figure can be viewed at wileyonlinelibrary.com]

collapse in the volume of trade, the COVID-19 episode did not leave a persistent effect on the total number of firm-firm linkages by this measure.[10]

5 | TRADE AND THE COVID-19 PANDEMIC

As mentioned, the timeliness and granularity of BoL data are especially valuable in understanding the enormous changes to international trade patterns brought on by the COVID-19 pandemic. This section details several insights from these data about the collapse and resurgence of trade during 2020–2021.

5.1 | The precise timing and effects of country-level lockdowns

Unlike official statistics, the daily frequency of the BoL transaction-level data allow the observation of intra-month patterns of trade. This feature is particularly useful in evaluating the impact of shocks to trade, with perhaps the largest and most abrupt in the modern era coming from the various country-level lockdowns associated with the early stages of COVID-19. We leverage the multiple sources of information coming from BoL data to highlight the transmission of the trade shock from the March 2020 national lockdown in India to U.S. imports.

We focus on the specific case of India because that country instituted a particularly strict COVID-19 lockdown, because pandemic-era U.S.-India trade has been relatively unstudied, and because bill of lading data are available for Indian exports to the U.S.[11] As shown in Figure 9, the national lockdown announced by the Indian government on March 24, 2020 is evident in the immediate decrease in India's exports to the United States and then subsequently in the delayed drop in U.S. imports from India several weeks later. The high-frequency BoL data reveal a much sharper drop in Indian exports to the U.S. than would be visible with monthly-frequency publicly available data. Moreover, the patterns in Figure 9 reveal important information on the translation of this shock into U.S. imports: The drop in U.S. imports from India is considerably less steep than the drop in Indian exports and lagged by 4 weeks. More broadly, Figure 9 indicates that BoL data can help researchers learn how the timing of such transmission of trade shocks varies across

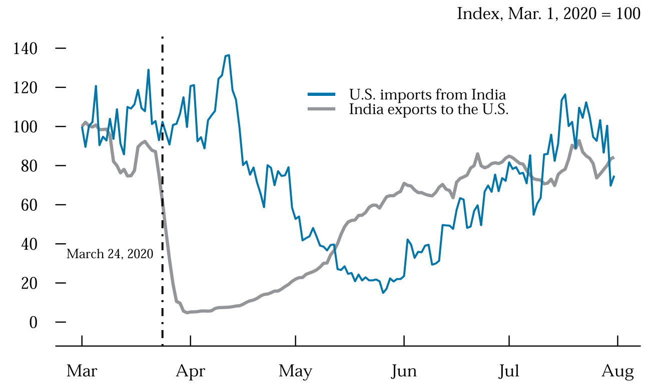

FIGURE 9 U.S.-India shipments during the COVID-19 lockdowns. *Source*: S&P Global Market Intelligence and authors' calculations (This figure plots the 7-day moving average of shipments of U.S. imports from India and India exports to the United States, with each indexed to equal 100 on March 1, 2020) [Colour figure can be viewed at wileyonlinelibrary.com]

trading partner based on distance, shipping routes (such as the use of entrepôt trading hubs), and the particular characteristics of the shock.

Such a stark episode illustrates lessons for other episodes. For example, China implemented partial lockdowns and mobility restrictions in 2021 and 2022, and the analysis above suggests that the spillovers into global supply chains takes substantial time and are smoothed out relative to the effects seen in China itself. The daily frequency also potentially allows for better causal inference for shocks like these. Rather than a largely unanticipated shock, consider a highly anticipated one, like the implementation of a large tariff change or other policy. Daily data allow the researcher to look for front-running, where firms hurry up to make shipments right before a change (or delay shipments until right after a change). In the case of the India lockdown, the daily data suggest that it was not significantly anticipated.

5.2 | Decomposing the collapse and subsequent surge in U.S. imports

The enormous drop in trade in the first quarter of 2020 was followed by a remarkable recovery, such that U.S. import volumes surpassed typical levels by the middle of 2020. Given the surprising speed of the resurgence in trade, a natural question is how importers and exporters managed to increase shipments so dramatically. For one useful perspective on both the collapse and subsequent surge in U.S. imports, we decompose the import changes based on the following margins at a quarterly frequency:

- Entry/Exit of Consignees Margin: The changes in imports due to the net entry and exit of consignees across two quarters.[12]
- Add/Drop Shipper or Country Margin: The changes in imports across two quarters from a given consignee that changes either the shipper or the country associated with the import transaction.
- Intensive Margin: The changes in imports from a given consignee–shipper–country pair across two quarters.

- Redacted: The changes in imports coming from changes in the pool of redacted consignees across two quarters.

Apart from the complicating feature of redactions, the decomposition outlined above is similar in spirit to the work of Bernard et al. (2009), which uses confidential, firm-level Census data. By contrast, with official public data, researchers are forced to define the extensive margin as something like an HS10 code coming from a particular country. That level of aggregation, however, would not capture the changes in relationships associated with entry/exit of consignees or switching among suppliers by continuing consignees. BoL data allow for the ability to track relationships defined at the consignee × shipper × country level.

To focus attention on the dynamics introduced by COVID-19, we fix the baseline period to be the fourth quarter of 2019, and then track the change along each margin in subsequent quarters. We begin with a decomposition of furniture imports (Chapter 94 in the HS classification system) due to the dramatic changes in demand experienced by this product group during our period of study. In addition, unlike some other categories, furniture is unlikely to be moved by air. Finally, to focus attention on the impacts of COVID-19 on imports along these margins, we make an adjustment to net out the effects of seasonality and trends in the margins. Specifically, we calculate the identical decomposition for each of the previous three years (baseline quarters of 2016Q4, 2017Q4, and 2018Q4), and for each margin of adjustment, and then subtract out the average change across each time horizon from the COVID-19 period.[13] The results are displayed in panel (a) of Figure 10.

The black line shows the overall change in U.S. furniture imports, relative to 2019Q4. Imports fell modestly in the first quarter of 2020 and then more significantly in the second quarter.[14] The surge in imports for product categories such as furniture is evident in subsequent quarters, with imports up over 35% (seasonally adjusted) by 2021Q1 from prepandemic levels.

We derive several useful lessons from decomposing these overall changes into the margins of adjustment outlined above, which are illustrated by the colored bars in Figure 10a. First, the drop in U.S. imports during the initial lockdowns of COVID-19 in 2020Q1 were driven largely by the intensive margin (the light blue bars), a feature that continued into the second quarter of 2020.

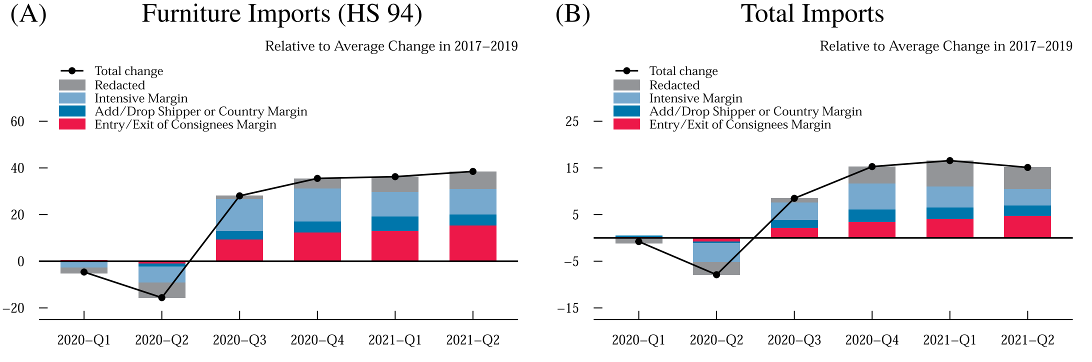

FIGURE 10 Decomposing percent change in imports (by TEU) relative to 2019Q4. *Source*: S&P Global Market Intelligence and authors' calculations (This figure plots the quarterly change in U.S. imports (by TEU) relative to 2019Q4 along four margins described in the text. The quarterly change for each margin is net of the average change during the equivalent quarter during 2017-2019 to account for seasonal variation and trend growth. Panel (a) restricts to imports of furniture (HS Chapter 94) whereas Panel (b) reports the decomposition for total imports) [Colour figure can be viewed at wileyonlinelibrary.com]

Second, although the intensive margin accounts for the largest individual share of the increase at the end of our sample period, when we combine the two extensive margins—i.e. the net consignee exit (in red) and add/drop shipper or country margin (in dark blue)—their contribution is larger than the intensive margin.[15] Hence, by 2021Q2, roughly half of the growth in furniture imports (a nearly 20 percentage point increase relative to 2019Q4) came from trading relationships that did not exist in 2019Q4. Finally, increases in consignee redactions (in gray) are also an important component in the overall increase in imports; without the consignee redaction, we would have been able to allocate these transactions into one of the other margins of adjustment.

Panel (b) of Figure 10 decomposes the growth in overall BoL imports during this time period. The most obvious difference relative to the decomposition for furniture imports in Figure 10a is the smaller and more gradual increase following the 2020Q2 nadir: overall imports were up 15 percent in early 2021 relative to the baseline compared with the roughly 40% increase for furniture imports. Qualitatively, the decomposition is similar to that for furniture.[16]

In summary, the BoL data allow researchers to understand the mechanisms underlying the extraordinary growth in imports in the months following the onset of the COVID-19 pandemic. These decompositions would have been invisible using traditional, publicly available datasets.

5.3 | Real-time measures of shipping bottlenecks during the COVID-19 trade recovery

The dramatic resurgence of trade in the second half of 2020 led to some much-discussed bottlenecks across many transportation modes. In this section, we show how the BoL data can be used to examine characteristics of vessel shipping that shed light on the prevalence and effects of bottlenecks in oceanic vessel shipping in nearly real time.

The use of BoL data to study the shipping network is highlighted by Ganapati et al. (2021) when used in conjunction with newly available vessel transponder data (otherwise known as Automatic Identification System (AIS) data) that tracks vessel ship movements.[17] While BoL data alone can identify the presence of indirect shipping—the primary topic of interest in Ganapati et al. (2021)—based on shipments coming from many different ports of lading on a given vessel-port of unlading combination, the key drawback is a lack of data associated with a shipment's foreign departure. The time stamp on AIS vessel movements enable researchers to track the precise route of a vessel through multiple ports of call. However, the key limitation of AIS data is a lack of any easily quantifiable measure of trade volume associated with each vessel. Therefore, the combination of BoL and AIS data—which would typically be accomplished through vessel name/identifiers and approximate dates—may be a fruitful application in future studies.

The analysis below leverages the vessel and ports of unlading variables that are typically reported in the BoL data, and focuses attention on the vessel congestion centered in the Ports of Los Angeles/Long Beach in late 2020 and into 2021. We take several steps to convert the raw BoL data into a dataset useful for tracking vessel arrivals at U.S. ports. First, we clean and standardize vessel name and a corresponding vessel identifier to account for inconsistencies in these variables.[18] Second, for many analyses at a vessel-port level, it is helpful to restrict attention to container vessels. While external lists can identify vessels based on vessel type, for our purposes, we classified container vessels based on a measure of observed capacity: whether the maximum observed TEUs unloaded at a particular point of time for a vessel surpassed a threshold.[19]

Third, we must identify a specific date for a vessel unloading cargo at a U.S. port. The difficulty here lies in the fact that the "arrival date" associated with BoL records typically reflect when individual shipments clear customs. Generally speaking, a large majority of BoL import shipment records from a container vessel at a port of unlading are listed as arriving within a 1 or 2 day period. However, there are frequent exceptions in which a vessel's shipments are reported as arriving over more extended periods of time, which could lead to an incorrect inference for a vessel arrival date. These records could reflect delays in clearing customs, typos in arrival date, or differences in identifying arrival date by exporters or importers. To account for these concerns, we take our baseline dataset of daily vessel-port observations and then eliminate a daily record if that day's shipments from a particular vessel were a very low share of the vessel's (observed) maximum capacity. Finally, we consolidate a vessel's arrival date into a single day if substantial shipments occur over a period of less than 5 days.

For a first look at the insights from this new dataset, we quantify the delays in vessel movements brought on by the shipping congestion experienced in 2020 and 2021. To measure the typical transit times for container vessels at a given port, we calculate the number of days between return arrivals of a given vessel and calculate the monthly median value for a given port. Panel (a) of Figure 11 indicates that a typical vessel would unload cargo at the Ports of Los Angeles/Long Beach (LA/LB) about every 43 days during normal times (2013–2017). This value was relatively stable in 2018 and 2019, but spiked in early 2020 following country-level lockdowns and the more general slowdown in trade during the early period of COVID-19. Round-trip transit times normalized in the third quarter of 2020 but subsequently increased in late 2020 and early 2021 due to the congestion at the Ports of LA/LB. The median number of days in between port visits of 52 during 2021Q1 and 2021Q2 reflects an increase of roughly 8 days from typical levels.

Panel (b) of Figure 11 shows that there was no such systematic delays in ship processing at an average of major U.S. East Coast ports during this period. Panel (b) also shows the longer average round-trip transit time of East Coast ports, a fact which reflects the increased prevalence of multistage trips common for vessels servicing these ports.[20]

Given reports that the congestion at the Ports of LA/LB resulted in vessels being rerouted to unload at other ports on the U.S. West Coast, we next attempt to quantify this degree of rerouting

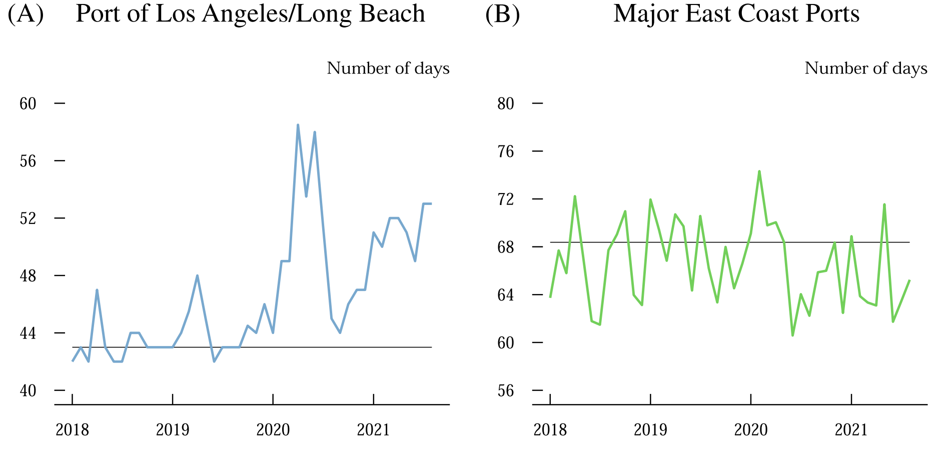

FIGURE 11 Median number of days between vessel visits at port. *Source*: S&P Global Market Intelligence and authors' calculations (This figure plots, for a given month, the median number of days since a vessel last visited the port. The black line in each figure represents the average number of days during the period 2013-2017. Major East Coast ports include the ports of Charleston, Newark/NY, Norfolk, and Savannah) [Colour figure can be viewed at wileyonlinelibrary.com]

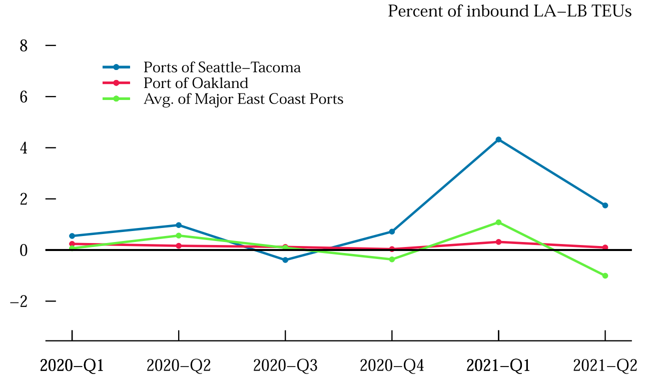

FIGURE 12 Percent of inbound Los Angeles / Long Beach activity rerouted to other ports. *Source*: S&P Global Market Intelligence and authors' calculations (This figure plots the percent of quarterly inbound LA/LB TEU imports that are identified as being rerouted to other ports. These values are net of the average observed percent rerouted to these ports from baseline periods beginning 2016, 2017, and 2018. Major East Coast ports include the ports of Charleston, Newark/NY, Norfolk, and Savannah) [Colour figure can be viewed at wileyonlinelibrary.com]

from our BoL-based dataset of vessel-port traffic. We first identify the sample of vessels that visited the Ports of LA/LB on a consistent basis in a pre-COVID period, that is, in both Q3 and Q4 of 2019. In the subsequent six quarters (2020Q1 to 2021Q2) we identify the potential set of vessel reroutings as those vessels that are not observed visiting the Ports of LA/LB but are observed visiting a different U.S. port. We measure the magnitude of these reroutings as the number of TEUs unloaded at alternative ports in a given quarter, which are then displayed as a fraction of total inbound TEUs at the Ports of LA/LB in that quarter. Finally, because what we define as rerouting may occur even in normal times, we calculate identical statistics from baseline periods in each of 2016-2018 and subtract the average of these "normal" vessel reroutings from the period of study.

The result is plotted in Figure 12 for three likely destinations of reroutings from the Ports of LA/LB: Seattle-Tacoma, Oakland, and an aggregate of four major East Coast ports. Figure 12 reveals that vessel reroutings from LA/LB to Seattle-Tacoma spiked in the first quarter of 2021 (following the onset of port congestion in late 2020) to an amount equal to roughly 8% of inbound TEUs at the Ports of LA/LB. This rerouting declines somewhat in the second quarter of 2021 but remains elevated relative to normal levels. While some reroutings were documented in press reports to the Port of Oakland, our data indicate that these did not constitute a significant fraction of inbound TEUs from LA/LB. Similarly, the data also confirm that few, if any, vessels were rerouted on net from LA/LB to the East Coast of the United States during this period.

In summary, the unique features of BoL data, together with timely access, provide both researchers and policymakers with a useful tool to analyze disruptions to trade such as those accompanying COVID-19.

6 | CONCLUSION

This article provides the first detailed analysis of the utility of data from bills of lading for international trade research, specifically the information available on U.S. imports via Panjiva. These data provide a near real-time, firm-level dataset useful for addressing a variety of economic questions that cannot be addressed with other data. Furthermore, some of the limitations of U.S. import data—including a general lack of trade values, redaction of some firm names, and being restricted to vessel shipping—do not apply to the same data available for other countries.

We use the unique elements of the data to analyze international trade relationships. About 60% of consignees (importers) have only one foreign shipper (exporter), but these consignees represent less than 10% of import volumes. Most shipper-consignee pairs ship in three or fewer months per year, though the surprisingly small number of pairs that ship every month account for about half of U.S. imports by TEU. In the COVID-induced trade collapse in 2020, the number of shippers per consignee dropped notably but recovered fairly quickly.

Finally, we explore other aspects of international trade during the COVID-19 crisis. The daily frequency shows how quickly exports from India to the United States fell following lockdowns in March 2020. Furthermore, the resulting drop in U.S. imports weeks later demonstrates clearly how international shipping lags transmit these shocks with a delay. Following the collapse, U.S. goods demand recovered briskly, and these data demonstrate the margins on which imports can rise. In the very short run, within a few months, higher imports were mostly achieved within existing shipper-consignee pairs. Over subsequent quarters, however, imports rose by consignees switching shippers or source countries, and also by the entry of new consignees.

Our work and the recent literature demonstrate that bill of lading data remains underutilized in international trade. With some caveats, these data provide a useful complementary dataset to disaggregated official public data and confidential datasets. Moreover, the ability to see most firm names of shippers and consignees opens the possibility of merging BoL data with other firm-level datasets.

DATA AVAILABILITY STATEMENT

Data used in this article come from three data sources: Panjiva bill of lading from S&P Global Market Intelligence, official trade data from the U.S. Census Bureau, and port statistics via Haver Analytics. Panjiva is a subscription service subject to third party restrictions on its redistribution. Data are available via https://panjiva.com/. Code for processing the data and producing the figures in the text will be posted to GitHub: https://github.com/maddieky/panjiva-code. U.S. Census Bureau data are available from several sources, including a subscription service and USA Trade Online (see https://www.census.gov/foreign-trade/index.html for details). Code for processing the data and producing the figures will be posted to GitHub (see above), and underlying data for the figures is available upon request. Historical port data via Haver are available via subscription service (http://www.haver.com/our_data.html). Data for many ports are available directly from port websites without a subscription.

ORCID

Logan Lewis https://orcid.org/0000-0002-4414-1919

ENDNOTES

[1] We examine exports from India to the U.S. because India instituted a particularly stringent lockdown at the onset of the COVID-19 pandemic, and because China—another natural candidate country—stopped making its BoL data available as tensions with the U.S. rose in 2018 and 2019.

[2] Bonfiglioli et al. (2021b) review the literature on heterogeneous firms in trade with additional results derived from BoL data.

[3] They supplement the Panjiva data with estimates from PIERS to fill out the dollar value of imported goods, as this variable is largely missing via Panjiva. These papers note the particular challenges that comes with working with BoL, specifically widespread spelling inconsistencies, as well as the various use of trade names and subsidiaries.

[4] PIERS was managed by IHS Markit until their merger with S&P Global in 2022. It is unclear whether these two sources will remain distinct going forward.

[5] Around 4% of shipments in the raw U.S. import data from U.S. Customs and Border Protection share a BoL number with at least one other shipment. These shipments may represent duplicate observations, though in at least some cases the arrival date is different for the two shipments while other fields are the same. Researchers should be aware of these potential duplicate shipments and consider whether their research questions require further actions to address them.

[6] Import price inflation in these goods is near zero: for nonpetroleum goods, BEA national accounts data show annualized import price inflation of −0.35% over the 11-year span of 2010Q1 to 2021Q1.

[7] Missing TEU values can simply reflect shipments that are not containerized, such as oil imports.

[8] Panjiva BoL data include indicators for when HS codes have been imputed, versus being scraped from product descriptions.

[9] For example, the largest consignee, Expeditors International, accounted for 5.4% of total U.S. imports by TEU and 7.1% of total shipments in 2020.

[10] In a robustness exercise, we excluded shippers and consignees with a two step process: We removed any firms we could link to S&P Capital IQ data classified as "Air Freight and Logistics", "Marine", "Trading Companies and Distributors" or "Trucking". Then, we manually checked the top 100 consignees and shippers and removed any others that appeared to be logistics companies or similar types of intermediaries. This process removes roughly 22% of TEUs and 16% of shipments. Figures 6 through 8 are nearly identical, demonstrating that these intermediaries are not driving our results on firm linkages. In addition, Figure 10 from Section 5.2 remains similar as well. Results available upon request.

[11] BoL data on exports from China to the U.S. are not available after March 2018.

[12] A consignee is considered to have exited in a particular quarter if it has no imports during that quarter. A consignee is considered an entrant in a particular quarter if it had imports during that quarter but had no imports in 2019Q4. We emphasize that these entry and exit distinctions are defined only for importing activity, and only for the period between the quarter of study and the baseline quarter (2019Q4). Hence, this margin will include consignees who do not import anything in a particular quarter (an exit) but will subsequently import in some future quarter (one of the other margins).

[13] Further discussion of this adjustment can be found in Online Appendix B, along with the unadjusted decomposition results.

[14] Imports fell considerably more in the first quarter of 2020 on a nonseasonally adjusted basis. However, furniture imports tend to peak in the fourth quarter each year, and then fall substantially in the first quarter.

[15] As first discussed in Bernard et al. (2009) the extensive margin becomes more important as the horizon lengthens. In our case, the switching product/country margin of adjustment is predominantly composed of cases where the consignee switches suppliers but maintains the same source country.

[16] In an earlier version of this article, the extensive margin contributed a larger share to the collapse in imports and there were notably distinctions between the furniture and total import recovery decomposition. As Panjiva updated its consignee and shipper IDs, these margins changed. This suggests some caution in interpreting results stemming from these IDs in real time.

[17] See Heiland et al. (2021), Cerdeiro et al. (2020) and Cerdeiro and Komaromi (2020) for examples of recent papers using AIS transponder data.

[18] We provide detail for this process in Online Appendix E.

[19] For the discussion below, we set this threshold at a relatively low value of 200, though for other purposes researchers may want to focus on vessels with larger capacity.

[20] Median time between port visits also tends to be noisier for East Coast ports because West Coast ports have more dedicated port-to-port vessel routes, which tend to run on more predictable schedules.

REFERENCES

Bernard, A. B., & Moxnes, A. (2018). Networks and trade. *Annual Review of Economics*, *10*(1), 65–85. https://doi.org/10.1146/annurev-economics-080217-053506

Bernard, A. B., Andreas, M., Bradford Jensen, J., & Lawrence, R. Z. (1995). *Exporters, jobs, and wages in U.S. manufacturing: 1976–1987*. In *Brookings Papers on Economic Activity. Microeconomics* (pp. 67–119). Brookings Institution Press.

Bernard, A. B., Andreas Moxnes, J., Jensen, B., Lawrence, R. Z., Redding, S. J., & Schott, P. K. (2009). *The Margins of U.S. Trade (Long Version), Working Paper 14662*. National Bureau of Economic Research.

Boehm, C. E., Flaaen, A., & Pandalai-Nayar, N. (2019). Input linkages and the transmission of shocks: Firm-level evidence from the 2011 Tōhoku earthquake. *The Review of Economics and Statistics*, *101*(1), 60–75.

Bonfiglioli, A., Crinò, R., & Gancia, G. (2020). Firms and economic performance: A view from trade.

Bonfiglioli, A., Crinò, R., & Gancia, G. (2021a). Concentration in international markets: Evidence from US imports. *Journal of Monetary Economics*, *121*, 19–39.

Bonfiglioli, A., Crinò, R., & Gancia, G. (2021b). *International trade with heterogeneous firms: Theory and evidence*. In *CEPR Working Paper* (Vol. 16249). Centre for Economic Policy Research. https://cepr.org/publications/dp16249

Bruno, V., & Shin, H. S. (2020). Dollar and exports.

Cerdeiro, D. A., & Komaromi, A. (2020). Supply spillovers during the pandemic: Evidence from high-frequency shipping data. *IMF Working Papers*, (WP/20/284). https://www.imf.org/en/Publications/Search?#q=cerdeiro&sort=relevancy&f:series=[WRKNGPPRS]

Cerdeiro, D. A., Komaromi, A., Liu, Y., & Saeed, M. (2020). World seaborne trade in real time: A proof of concept for building AIS-based Nowcasts from scratch. *IMF Working Papers*, (WP/20/57). https://www.imf.org/en/Publications/Search?#q=cerdeiro&sort=relevancy&f:series=[WRKNGPPRS]

Dhyne, E., Kikkawa, A. K., Mogstad, M., & Tintelnot, F. (2021). Trade and domestic production networks. *The Review of Economic Studies*, *88*(2), 643–668.

Feenstra, R. C., & Weinstein, D. E. (2017). Globalization, markups, and US welfare. *Journal of Political Economy*, *125*(4), 35.

Ganapati, S., Wong, W. F., & Ziv, O. (2021). Entrepôt: Hubs, scale, and trade costs.

Haver Analytics. *Haver Analytics* http://www.haver.com/our_data.html.

Heiland, I., Moxnes, A., Ulltveit-Moe, K. H., & Zi, Y. (2021). Trade from space: Shipping networks and the global implications of local shocks.

Heise, S., Schaur, G., Pierce, J. R., & Schott, P. K. (2019). Tariff rate uncertainty and the structure of supply chains.

Jain, N., & Di (Andrew), W. (2020). *Can global sourcing strategy predict stock returns?* In *SSRN Scholarly Paper ID 3606884*. Social Science Research Network.

Jain, N., Girotra, K., & Netessine, S. (2014). Managing global sourcing: inventory performance. *Management Science*, *60*(5), 1202–1222.

Kamal, F., & Monarch, R. (2018). Identifying foreign suppliers in U.S. import data. *Review of International Economics*, *26*(1), 117–139.

Monarch, R. (2022). "It's not you, it's me": Prices, quality, and switching in U.S.-China trade relationships. *The Review of Economics and Statistics*, *104*(5), 909–928.

Panjiva S&P global market intelligence, Panjiva supply chain intelligence platform & data feed. https://www.spglobal.com/marketintelligence/en/solutions/panjiva-supply-chaipply-chain-intelligence

U.S. Census Bureau Merchandise Trade. https://www.census.gov/foreign-trade/index.html

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Flaaen, A., Haberkorn, F., Lewis, L., Monken, A., Pierce, J., Rhodes, R., & Yi, M. (2023). Bill of lading data in international trade research with an application to the COVID-19 pandemic. *Review of International Economics*, *31*(3), 1146–1172. https://doi.org/10.1111/roie.12657

TABLE A1 Comparison of Panjiva and Official Statistics, 2019

	Panjiva weight[a]	Census weight[a]	Panjiva weight share[a]	Census weight share[a]	Panjiva TEU[a]	Official port TEU[a]
Houston, TX	46,594	55,630	0.08	0.09	1.20	1.24
Los Angeles, CA	41,595	25,190	0.07	0.04	4.61	4.71
Long Beach, CA	35,208	54,085	0.06	0.09	3.70	3.76
Newark, NJ	58,108	56,963	0.10	0.09	3.75	3.77
Savannah, GA	20,661	20,049	0.04	0.03	2.19	2.22
Seattle/Tacoma, WA	14,489	14,398	0.02	0.02	1.45	1.37

[a]In millions.

Note: Panjiva aggregates for all ports except Newark exclude shipments where the consignee country is not the United States. Seattle/Tacoma and Houston aggregates also exclude shipments where the consignee country is also missing.

Source: S&P Global Market Intelligence, Haver Analytics, U.S. Census, and authors' calculations.

APPENDIX. COMPARING PANJIVA, U.S. CENSUS, AND PORT DATA

In this section of the appendix, we compare aggregated shipping volume BoL data to data directly reported by ports themselves and also to official U.S. Census Bureau data. These checks provide information to researchers considering the representativeness of the BoL data. We focus on comparing volume measures in BoL (weight and TEU), as they are more comprehensively available than imputed values, which suffer from both missing observations and extensive measurement error. Generally speaking, weight and TEU are very similar volume measures over time and could substitute for one another given most questions. In short, we find that BoL data closely track official port data in TEUs.

A.1 Comparing statistics by port

Table A1 compares measures of trade weight and number of TEUs by port, as reported in BoL data and by Census. BoL weight measures tend to exceed Census measures somewhat. Still, as the columns labeled "share" demonstrate, the proportion of imports going to each port is similar between Census and BoL, with the notable exception of Los Angeles and Long Beach: Here, the sum of the two ports is more comparable than their individual identification. This adds to the list of reasons why it is best practice to treat LA/LB as a single economic entity for most questions with these data.

The right two columns of Table A1 provide the total TEU count in 2019 by port for BoL data and data provided by the ports themselves. In most cases, these correspond remarkably closely.

A.2 Comparing containers by port over time

Next, we compare the number of imported TEUs reported by Panjiva to the volumes reported by ports. In particular, Figure A1 displays monthly Panjiva and official imported TEU volumes for the top six U.S. ports. Importantly, both sources tend to give similar signals for the level and changes in trade from month to month.

In terms of timeliness of data reporting, the official data on container volumes by port are available from Haver with a lag of about 3 weeks on average, while data are available from Panjiva with a lag of only about 7–14 days. While this improvement in timeliness from Panjiva data

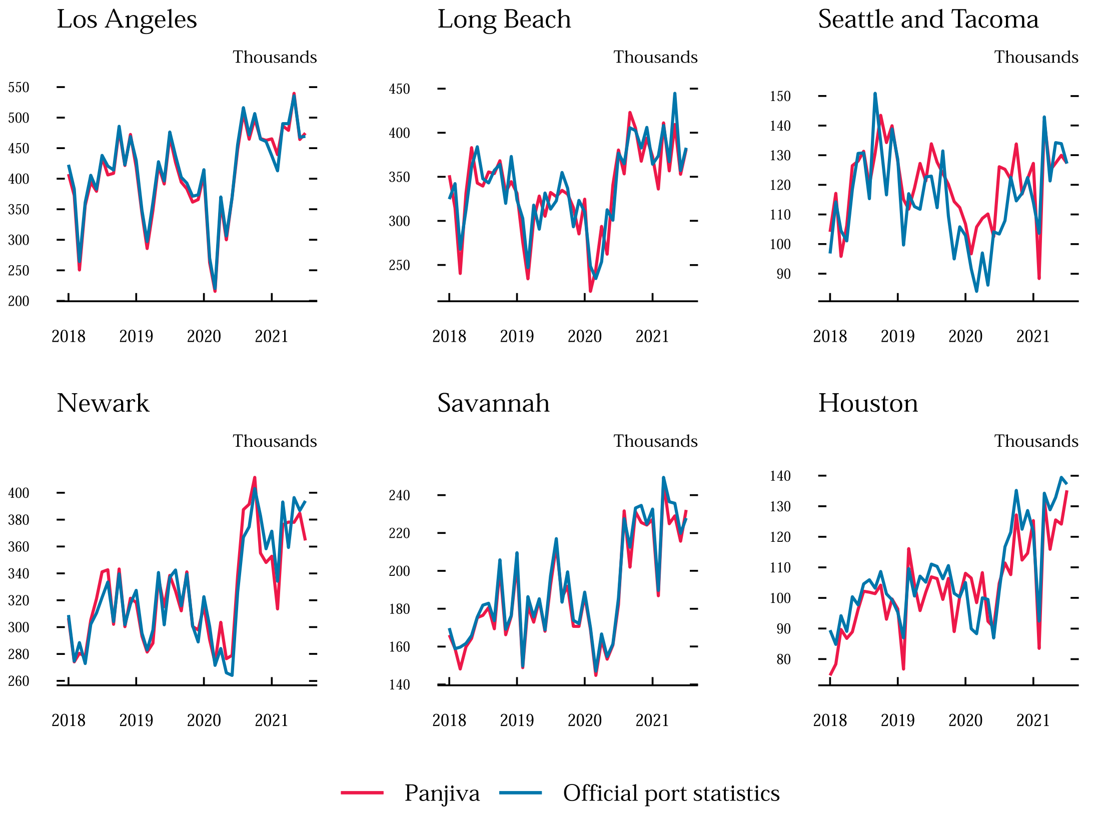

FIGURE A1 Comparison of Panjiva data and official port statistics by port. *Source*: S&P Global Market Intelligence, individual ports via Haver Analytics, and authors' calculations (Panjiva aggregates for all ports except Newark exclude shipments where the consignee country is not the United States. Seattle/Tacoma and Houston aggregates also exclude shipments where the consignee country is missing) [Colour figure can be viewed at wileyonlinelibrary.com]

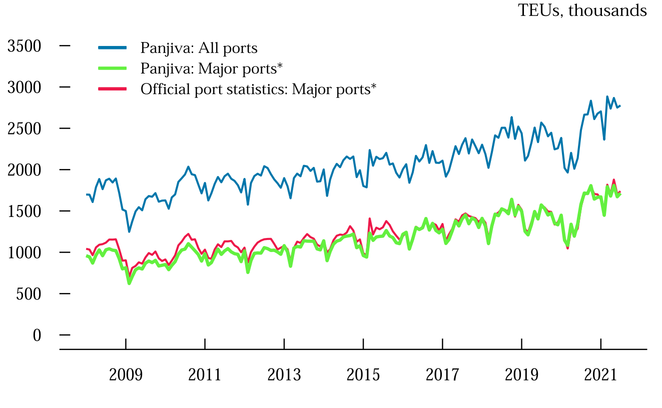

FIGURE A2 Comparison of Panjiva data and official port statistics. *Source*: S&P Global Market Intelligence, individual ports via Haver Analytics, and authors' calculations. *Major ports include Houston, Long Beach, Los Angeles, Savannah, Seattle/Tacoma, New York, and Newark [Colour figure can be viewed at wileyonlinelibrary.com]

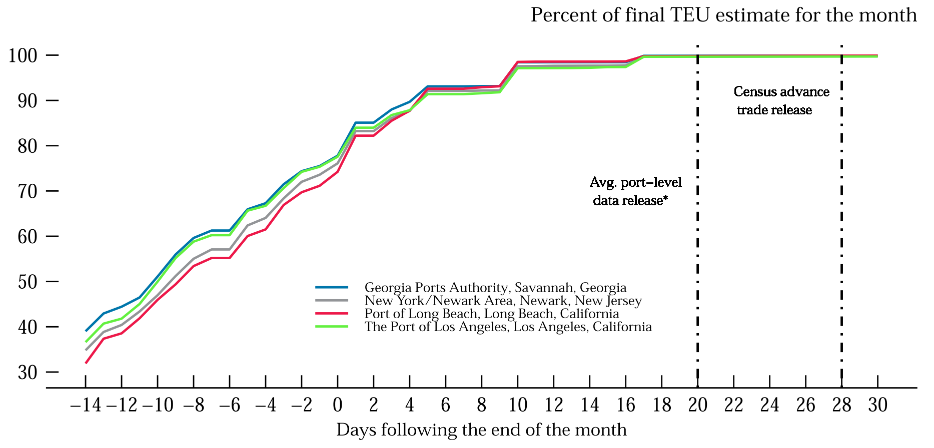

FIGURE A3 Panjiva BoL data completeness. *Source*: S&P Global Market Intelligence, individual ports via Haver Analytics, Census Bureau, and authors' calculations (100% reflects the "final" level of TEUs estimated for a given month) [Colour figure can be viewed at wileyonlinelibrary.com]

is relatively modest, it may nonetheless be valuable during times when shipping is being interrupted, such as during the COVID-19-related plunge and the backups at West Coast ports during the subsequent recovery.

Figure A2 compares the aggregate of major ports in Panjiva against official port level statistics over a longer time period, and also compares these major ports to all ports.

A.3 Comparing lags in data reporting

Figure A3 illustrates the timeliness of when data are available for a given month. As shown in the figure, data are updated continuously with roughly three-quarters of a given month's final TEU value in by the end of the month. The data then reach close to 100% of the final monthly value by around 7 to 14 days after the end of the month. This reporting is sooner than the port-level reporting and significantly sooner than the Advance Economic Indicators trade report released by the U.S. Census Bureau.

A.4 Comparing firm-level trading information

As discussed above, one of the key benefits of the BoL data, relative to public data sources, is the availability of firm identifiers for most transactions. Comparing firm-level information from BoL data to similar information in other datasets, such as the Census Bureau's Longitudinal Foreign Trade Transaction Database (LFTTD), is difficult given the confidentiality associated with official statistical datasets. Nonetheless, the Census Bureau does publish some information on characteristics of firms engaged in international trade, which can be compared to BoL sources.

One piece of information about trading firms that the Census Bureau reports is a histogram of the value of trade by the number of destination countries for each exporting firm (See top chart on page 3, Census Bureau 2020). In Figures A4 and A5, we display similar figures based on Panjiva data for both exporters and importers, respectively, though our histograms are in terms of the number of TEUs and shipments. Our figures include all firms and are therefore most comparable to the blue bars in the histogram provided by the Census Bureau.

Figure A4, for exports, shows a rightward skew of the distributions for TEUs and shipments based on BoL data, indicating the importance of firms that export to many countries in overall trade volumes. This rightward skew is consistent with, but actually somewhat less pronounced than that reported for the value of exports in Census Bureau (2020), which is reproduced in the gray bars of the Figure.

Figure A5 indicates that, in contrast to exports, firms that import from a small number of destinations account fora relatively larger share of U.S. import volumes. This difference may be indicative of smaller fixed costs associated with importing, relative to exporting.

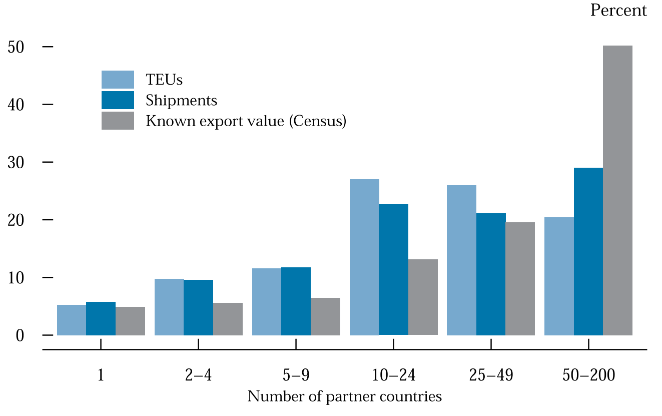

FIGURE A4 Percent of TEUs, shipments, and value by number of partner countries for U.S. exports, 2018. *Source*: S&P Global Market Intelligence, U.S. Census, and authors' calculations [Colour figure can be viewed at wileyonlinelibrary.com]

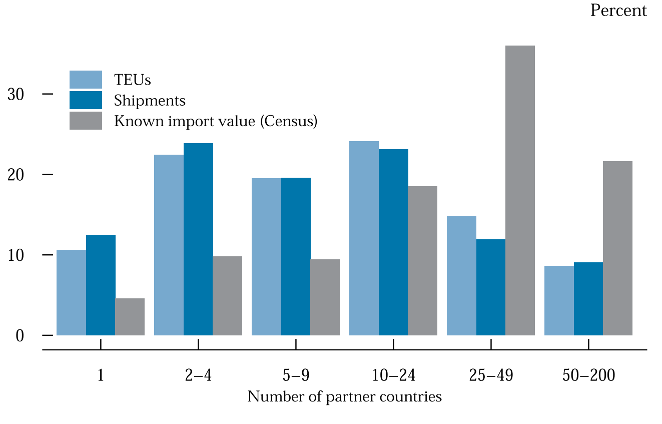

FIGURE A5 Percent of TEUs, shipments, and value by number of partner countries for U.S. imports, 201 *Source*: S&P Global Market Intelligence, U.S. Census, and authors' calculations [Colour figure can be viewed at wileyonlinelibrary.com]