Cross-Border Intra-Firm Trade and the Propagation of Idiosyncratic Shocks:

A New Dataset

Vanessa Alviarez
Brian Cevallos Fujiy
Tomasz Święcki

Department of Research and Chief Economist

TECHNICAL NOTE Nº IDB-TN-2398

June 2022

Cross-Border Intra-Firm Trade and the Propagation of Idiosyncratic Shocks:

A New Dataset

Vanessa Alviarez*
Brian Cevallos Fujiy**
Tomasz Święcki***

* Inter-American Development Bank
** University of Michigan
*** University of British Columbia

June 2022

Inter-American Development Bank

Cataloging-in-Publication data provided by the
Inter-American Development Bank
Felipe Herrera Library

Alviarez, Vanessa.
Cross-border intra-firm trade and the propagation of idiosyncratic shocks: a new dataset / Vanessa Alviarez, Brian Cevallos Fujiy, Tomasz Święcki.
p. cm. — (IDB Technical Note ; 2398)
Includes bibliographic references.
1. International business enterprises-Statistics-United States. 2. International business enterprises-Industrial capacity-United States. 3. Natural disasters-Economic aspects. I. Cevallos, Brian. II. Święcki, Tomasz. III. Inter-American Development Bank. Department of Research and Chief Economist. IV. Title. V. Series.
IDB-TN-2398

Abstract[*]

We study how disruptions in international production networks propagate across countries. We use comprehensive data on natural disasters around the globe over the last two decades from the EM-DAT and SHELDUS database to identify exogenous shocks to sourcing foreign inputs. We then trace out the effect of these shocks on activity of multinationals located in the United States. and their network of foreign affiliates using U.S. Bill of Lading microdata and data on domestic and international ownership linkages from Orbis. Our findings indicate that major natural disasters can have an economically significant negative impact even far from the directly affected areas. Furthermore, the strength of the propagation depends on whether the shocks led to disruptions in intra-firm or arms-length trade. This technical note provides detailed information on the construction of the novel dataset used in Alviarez et al. (2021).

JEL classifications: F23, F14, L14, L23

Keywords: Multinational firms, Production networks, Shock propagation

[*] Thanks to Keith Head, Matilde Bombardini, Natalia Ramondo and Andrei Levchenko, as well as the participants of the UBC VSE-Sauder seminar and the CEA annual meetings for very useful comments and suggestions to early versions of this work.

Email: valviarezr@iadb.org, bcfujiy@umich.edu, and tomasz.swiecki@ubc.ca.

All errors are our own.

1. Introduction

Recent events, such as the trade war between the United States and China, and the turbulent negotiations over the US-Mexico-Canada trade agreement (USMCA), replacing the North American Free Trade Agreement (NAFTA), brought to the public view risks associated with drastic trade policy changes in a world in which firms have tightly integrated operations across national borders. The current global COVID-19 pandemic has also palpably exposed the risks associated with the interconnected nature of global trade. In particular, reliance on foreign input suppliers both within and outside the boundaries of the firms can lead to a disruption of production when source countries experience negative shocks. Two opposite forces are in play. On the one hand, increases in production sharing across global value chains and the corresponding decline in the value-added share of a given country's exports can make the impact of trade disruptions and that of tariff and non-tariff shocks more severe. On the other hand, given the prominence of multinational firms in global production, it may be easier than ever to relocate production across country borders, partially offsetting the effects of idiosyncratic shocks on production and prices (Flaaen et al., 2020). Yet despite its policy relevance, academic research on the effect of trade shocks in the presence of cross-border ownership and intra-firm trade linkages is quite sparse, mostly due to the dearth of comprehensive data on cross-border activities of multinational companies. In this paper, we estimate how disruptions of within-firm trade affect the activities of domestic and foreign multinationals located in the United States. Then, we compare these effects with those of disruptions of international trade transactions among independent companies. By contrasting the impact of shocks to arms-length and intra-firm trade, we improve our understanding of the role of common ownership in the propagation of shocks within production networks across countries.

Investment-specific relationships and financial interdependencies are some of the reasons provided in the literature as to why multinationals—defined as an international network of production units under common ownership—could respond differently to economic shocks. However, whether the effect of cross-border trade disruptions on firms' performance depends on the ownership linkages between the transacting companies remains an open empirical question. We intend to answer this question by assembling a novel dataset that combine comprehensive information from U.S. international trade transactions registered in the bill of lading, and information on ownership linkages within and across borders provided by Orbis. In order to estimate the importance of multinational production networks for the transmission of trade shocks, we use major natural disasters that occurred in countries from which a multinational sources its intermediate inputs as an exogenous trade cost shock.

Our research will contribute to a broader question of how firm-level idiosyncratic shocks propagate in production networks, departing from the exiting literature in several ways. Most of the existing work focuses on the transmission of shocks among domestic production networks, regardless of their ownership linkages and their type of customer-supplier relationship (Carvalho et al., 2021; Barrot and Sauvagnat, 2016). Focusing on ownership linkages, Cravino and Levchenko (2017) study how multinational firms contribute to the transmission of shocks across countries, whereas Boehm et al. (2019) and Freund et al. (2021) focus on the supply-chain effects of the Japanese earthquake of 2011 on the U.S. affiliates of Japanese multinationals. Relative to these papers, we focus on the effects of intra-firm transactions across the worldwide network of affiliates of U.S. multinationals and on whether the effects are different when trade is carried out with unrelated parties. Since we observe the network of multinationals' affiliates across countries, including the ones in locations not affected by shocks, we can measure the extent to which new trade relationships were formed with related parties in unaffected countries, or if there was a significant increase in imports from those foreign affiliates. Another feature that distinguishes our approach is the use of a wide range of severe natural disasters across countries to construct plausible exogenous shocks.

The rest of the technical note is organized as follows. Section 2 introduces the scope of the data assembled in Alviarez et al. (2021) and provides a detailed description of the individual databases used in the analysis. Section 3 presents summary statistics of the assembled dataset, and Section 4 concludes.

2. Data

To understand how the nature of production networks affects the propagation of shocks across countries, we start by assembling a novel dataset that combines four micro level datasets: i) the financial data of non-financial firms from Compustat; ii) the universe of U.S. maritime import transactions for the period 2007-2020, extracted from the U.S. bill of lading; iii) the ownership linkages between firms, within and across countries, extracted from Orbis; and iv) detailed information on the type, exact location and number of affected people of major natural disasters that occurred worldwide during the sample period, taken from the EM-DAT and SHELDUS database. The combined dataset allows us to have a more complete portrait of the global operation and international trade transactions carried out by US companies, including U.S. MNCs, and affiliates of foreign MNCs operating in the United States.

First, we obtain financial data for non-financial firms from the Compustat North America Fundamentals Quarterly database for the period 2007-2020. These data include information on sales, cost of goods sold and profits, as well as daily stock price data from the Center for Research and Security Prices (CRSP) and Compustat Global. Second, we merge Compustat with the U.S. import bill of lading. This allows us to identify those firms in Compustat that are also importers and the extent of their international trade transactions (number of transactions, longevity of the relationship, product and quantity transacted) with foreign suppliers.

Third, we merge the ownership linkages available in the Orbis database, which are targeted towards corporate customers that need to know the ownership structures, hierarchy, and contact information of private and public enterprises throughout the world. It has detailed information about ownership linkages between firms and across countries. Thus, the Orbis dataset allows us to identify for each company in Compustat i) which firms are headquartered in the United States, and which are affiliates of a company headquartered in a foreign country; and ii) whether the firm is part of a global corporation (i.e., a multinational), and if so, in which countries and sectors the rest of the affiliates belonging to the corporate group have operations. Orbis provides detailed information on the name, country of operations, as well as information on sales and assets of the firms belonging to the same multinational corporation. Therefore, Orbis allows us to identify which of the foreign partners, from which U.S. firms source their imported intermediate inputs, are also part of the same multinational corporation and which are independent trade partners. Importantly, Orbis also identifies foreign firms that are part of the same corporation but with which a U.S. company does not have any trade relationship. Thus, for each firm in Compustat we identify the foreign firms with which they have trade transactions, distinguish between intra-firm and arm's-length trade, as well as those foreign firms that are part of the same corporation regardless of whether there is a trade relationship between the parties.

Finally, we bring the information on the location, length and severity of the natural disasters around the world from the Emergency Events Database (EM-DAT) and the Spatial

Hazard Events and Losses Database for the United States (SHELDUS) for the period 2007-2020. After using the physical address of the firms in the sample located in the United States and overseas to provide latitude and longitude coordinates, we proceed to construct a measure of the level of firms' "disaster exposure" by calculating the distance between the firms with which U.S. firms keep either trade or ownership relationships and the region within the country that has been affected by the natural disaster.

While none of these sources is new in isolation, our innovation on the data side is to merge them. This is not a trivial task, and it involves, in particular, geocoding the location of millions of firms and thousands of natural disaster locations in order to precisely identify which firms have been hit by which natural disasters. To our knowledge, this is the first time these four datasets have been linked. In the following sections, we provide a more detailed description of these four datasets.

2.1 Firm-to-Firm Trade Transactions from Bill of Lading Records

Information on international trade transactions is from the U.S. shipment-level bill of lading (BoL) compiled and provided by S&P Panjiva.1 The original dataset contains over 155 million transaction-level records of goods traded across borders since 2007, including information on the date of the transaction, port of landing, name and address of the exporter (shipper) and the importer (consignee), the product description, and the quantity transacted. Panjiva enhances the dataset by i) assigning a Harmonized System (HS) product code to each product description included in the BoL; ii) providing a unique identifier to each importer (consignee Panjiva ID) and exporter (shipper Panjiva ID), allowing the longitudinal tracking of firms engaged in international trade; and iii) including a unique company identifier variable linking the BoL data to Capital IQ, another S&P Global dataset containing key company's financial information, such as sales, costs of goods sold, and profits. This crosswalk between Panjiva and Capital IQ is particularly useful for us since Capital IQ contains two key identifiers,

[1] A bill of lading in shipping is a record of the traded goods which have been received on board. It is a legal document that establishes an agreement between a shipper and a transportation company for the transportation of goods. Transportation Company (carrier) issues these records to the shipper. A bill of lading indicates a particular carrier through which the goods have been placed to their final destination and the conditions for transporting the shipment to its final destination. A BoL contains detail information such as both the shipper and consignee name and address, description of the goods, vessel name, transport company name, ports of loading and unloading, weight, and quantity. Panjiva acquires these data by collecting bills of lading from U.S. Customs and Border Protection (CBP).

the Committee on Uniform Securities Identification Procedures number (CUSIP), which is used to identify U.S. and Canadian registered stocks, and the Central Index Key (CIK) assigned by the Securities and Exchange Commission. These IDs are also available in the Compustat dataset, allowing us to match the Panjiva trade data with Compustat dataset, a procedure that we describe in detail below.[2]

Geocoding Firm Addresses

In order to measure the distance between the physical address of the firm and the geographical location of any natural disaster, we proceed to convert the physical address component of all U.S. importers and their corresponding foreign partners into geographical coordinates (latitude and longitude). After harmonizing the addresses of those firms with the same Panjiva ID but with multiple spelling addresses, we provide longitude and latitude coordinates for 470,435 unique physical address of U.S. importers and 995,568 unique physical address of foreign exporters.[3]

To perform the geocode for each firm we use the most complete address available in the Panjiva BoL. Our preferred firm address is the one constructed by concatenating the individual components of a firm's physical address including route, city, region/state, postal code, and country. When the only non-empty individual components of the firm address were region/state and country, or country only, we use the firm's complete or full address in the BoL. The reason why we prefer the firm address constructed from parsed components is because it is more structured and therefore easier to geocode using API web services. Full addresses originally

[2] The variables added by Panjiva to those originally included in the BoL come with some caveats. First, the provided HS product code is at six digits (around 6,000 products), rather than at 10 digits (more than 10,000 products) as the one recorded by the U.S. Census Bureau Customs data. Second, a given transaction can list multiple HS6 codes, but only one shipment weight in kilograms, making it impossible to distinguish the weight associated to each HS6. Third, Panjiva uses a text processing algorithm in order to map a given company name and address to a unique numerical ID. Nonetheless, there is a fraction of cases in which the algorithm fails to recognize that two companies are the same legal entity, wrongly assigning different Panjiva ID numbers. We overcome this challenge by refining the procedure, allowing us to assign a temporal consistent ID number to a company that shows up in the data with slightly different spellings in their names and/or addresses. Spelling differences in the address component are resolved by geocoding all the addresses in the BoL dataset. For additional details about the BoL data for the United States see Flaaen et al. (2021) and Alviarez and Blyde (2021).

[3] Our data include a total of 2,586,903 unique physical address for U.S. importers and 2,782,417 unique physical address for foreign exporters. However, we prioritize the geocoding of U.S. firms, and their corresponding foreign exporters, that satisfy the following criteria: i) U.S. importers that are also exporters; ii) U.S. importers in Panjiva that are also in the Compustat dataset; and iii) U.S. importers that have more than one foreign partner. After applying these restrictions, we geocode 470,435 US address and 995,568 foreign addresses.

entered in the BoL can be difficult to parse, as they lack punctuation separating the sub-components of the firm's address, thus complicating the geocoding process.[4]

Table 1 shows that for 59 percent of U.S. importers the geocode was based on concatenated fields where all address components were available (route, city, region, postal code and country), whereas for 4.6 percent of those importers all components other than route were available and used in the geocode of the physical address of the firm. For 25.5 percent, the full address directly provided by BoL was used for geocode. For foreign exporters, we have to rely on the full address for 60 percent of firms.

Table 1. Distribution of Geocoded Firms by Address Type (%)

	US Importers	Foreign Partners	Total
Full address	25.5	37.3	33.5
Route/City/Region/Postal Code/Country	59	21.1	33.3
City/Region/Postal Code/Country	4.6	7.7	6.7
City/Region/Country	3.5	11.7	9.1
Route/City/Region/Country	2.7	6.6	5.3
Route/Region/Postal Code/country	1.8	3.2	2.8
Other	2.9	12.4	9.3

Note: This table shows the information contained in the firm's physical address used in the geocode process for U.S. importers and for foreign exporters. The first row shows the percentage of addresses that were geocoded using the full or complete address directly reported by the Panjiva BoL. The next five rows show the percentage of geocoded addresses that were constructed by concatenating the individual components of the firm's physical address including route, city, region/state, postal code, and country. The last row shows the percentage of firms that use a different combination of route, city, region/state, postal code, and country, not listed in the previous rows.

We use the Geopy python library, which is a python client for several popular geocoding web services, such as Google maps, OpenStreet, among others. In particular, we use Bing Maps Location API geocoder to obtain the latitude and longitude corresponding to each address. Importantly, Bing API provides some variables that can be used to assess the quality of the outcome it generates. These variables are i) confidence, ii) match codes, iii) inland, and iv) country. A given geocode has a medium or low confidence level when only a subset of the address components is matched (i.e., if only the postal code of the full address is matched). The

[4] The full address constructed by concatenating the individual address fields is often of better quality because it is more systematic and less subject to error.

match code can take three values: good, ambiguous, and up-hierarchy, depending on whether the location has one or multiple returned matches. To further assess the reliability of the geocoding outcome, we create an inland variable that takes a value of one when, according to the International Space Station, the returned coordinates lie inland, and zero if they are on the ocean. Second, since it is possible to retrieve the geographic variables associated with the coordinates provided by the Bing geocoding algorithm, we generate a dummy variable that compares the firm country with the country returned by Bing.

The results are displayed in Table 2. Bing API provides geocode coordinates for the vast majority of the searched addresses (99.8 percent). Out of the 469,742 geocoded US addresses, 85.2 percent show a high confidence level and 85.3 percent an unambiguous (good) match codes; 9.3 percent of the addresses are inland and 90.6 percent of the returned address report the same country as the one provided as input in the geocoding process. The quality of the geocode for U.S. foreign partners is lower (see the third column of Table 2), which is expected, given that international addresses are often more complex and less structured.

Table 2. Geocode Quality Assessment

	US Importers	US Foreign Partners	Total
Geocoded	469,742 (99.9)	993,587 (99.8)	1,463,329 (99.8)
Confidence level			
High	400,780 (85.2)	678,749 (68.2)	1,079,529 (73.7)
Medium	61,802 (13.1)	300,486 (30.2)	362,288 (24.7)
Low	7,160 (1.5)	14,352 (1.4)	21,512 (1.4)
Match Codes			
Good	387,054 (85.3)	514,962 (51.7)	902,016 (61.6)
Ambiguous	42,924 (9.1)	253,862 (25.5)	296,786 (20.2)
Up Hierarchy	39,764 (8.5)	224,763 (22.6)	264,527 (18.0)
In Land	467,359 (99.3)	979,759 (98.4)	1,447,118 (98.8)
Same Country	426,406 (90.6)	833,695 (83.7)	1,260,101 (86.1)
Total	470,435	995,568	1,466,003

Note: This table shows the quality of the geocoded procedure. The first row shows the total number of addressed for which we obtain geographical coordinates (latitude and longitude). Rows 4-7 show the number and percentage of firms with high, medium and low confidence level as reported by Bing API. Rows 7-9 shows the number and percentage of firms with different levels of ambiguity. The 10th row shows the number and fraction of returned coordinates that lie inland, and the last row reports the number and fraction of firm's addresses for which the country returned by the geocoded procedure coincides with the country listed in the address used as inputs.

After obtaining the longitude and latitude location of each firm, and after considering the shape-files from the Food and Agriculture Organization of the United Nation's Global Administrative Unit Layers (FAO GAUL), we assign an administrative level (ADM) to each firm.[5] The shape-files are available at three administrative levels: adm0 (national), adm1 (provinces), and adm2 (districts).

2.2 Financial Firm-Level Information

To measure the impact that a shock on a foreign partner has on a U.S. firm we use financial information from the Compustat North America Fundamentals Quarterly database, which is a sub-module from the CRPS/Compustat Merged dataset that focuses on balance sheet variables of firms in the United States. These variables include the following: sales, cost of goods sold, total assets, long-term debt, earnings per share, and dividends per share. We also retrieve information on return on assets (ROA), return on equity (ROE), standard deviation of ROA and ROE, ratio of non-interest income to operating income, ratio of loans to total assets, ratio of deposits to total assets, and ratio of equity to total assets. Finally, we retrieve information on firm investment, measured as the ratio of quarterly capital expenditures to the lag of quarterly property, plant and equipment; and long-term leverage, defined as the ratio of total long-term debt to total assets.[6]

Another key variable we use to measure the propagation effects of idiosyncratic shocks is the weekly firm's stock returns,[7] which are calculated using the following formula:[8]

[5] In order to obtain the shape-files we contacted the GeoNetwork@fao.org. Thanks to Nelson Rosas Ribeiro Filho for sharing the maps with us. The shape-files required some adjustments, for which we obtained the centroid of each polygon.
[6] These variables from *CRPS/Compustat Merged* dataset have been widely used in academic research, including Bhargava (2014), Li et al. (2021), and Almeida et al. (2009), among many others.
[7] We have calculated three versions of weekly returns: week1, week2, and week3. In "Week1" weekly returns are calculated using the first day of the year, which is the first day we can find for each firm. From that moment on we count 7 days each time until the end of the year. Thus, in this definition there are not necessarily 52 weeks in a year. In "Week2" we find the first Friday of each week and calculate returns with respect to Monday. In "Week3" we calculate returns from Friday to previous Friday; and whenever there is not a Friday (either today or previous Friday), we use the previous day. This latter definition is our preferred one since it is the closest to a calendar week. Moreover, under this definition we are not losing events happening during the weekends (as will happen under the "Week2" measure). "Week3" definition is also useful to calculate compounded returns at a lower frequency (e.g., monthly), since it allows to collapse data easily (this does not happen with "Week2" since we only considered from Monday to Friday).
[8] To calculate firms' weekly returns we closely follow Compustat Global WRDS documentation.

$$Returns_t = \frac{\frac{PRCCD_t}{AJEXDI_t} TRFD_t - \frac{PRCCD_{t-1}}{AJEXDI_{t-1}} TRFD_{t-1}}{\frac{PRCCD_{t-1}}{AJEXDI_{t-1}} TRFD_{t-1}} X\ 100 \tag{1}$$

where *PRCCD* are closing prices, *AJEXDI* is a daily adjustment factor, and *TRFD* is a daily total return factor. Whenever *TRFD* is missing we calculate returns as:

$$Returns_t = \frac{\frac{PRCCD_t}{AJEXDI_t} - \frac{PRCCD_{t-1}}{AJEXDI_{t-1}}}{\frac{PRCCD_{t-1}}{AJEXDI_{t-1}}} X\ 100 \tag{2}$$

Whenever *AJEXDI* is missing or zero, returns are calculated as:

$$Returns_t = \frac{PRCCD_t TRFD_t - PRCCD_{t-1} TRFD_{t-1}}{PRCCD_{t-1} TRFD_{t-1}} X\ 100 \tag{3}$$

Finally, whenever both, *TRFD* and *AJEXDI* are missing, returns are calculated as:

$$Returns_t = \frac{PRCCD_t - PRCCD_{t-1}}{TRFD_{t-1}} X\ 100 \tag{4}$$

Exports from Bill of Lading: Using the export US bill of lading we calculate, for each firm in Compustat, the total quantity exported each quarter (measured in kilograms), to each destination country and HS6 product. We use this variable as an additional outcome variable in our analysis. Notice that the US BoL data do not provided information on the value of the transaction. To estimate the value of the products exported, we use the unit values calculated from the US Census Bureau at the HS6 product-quarter-destination level, and we multiplied them by the weight in kilograms provided in the BoL.

2.3 Global Ownership Linkages

Our data on firm-to-firm ownership linkages come from ORBIS, a worldwide dataset maintained by Bureau van Dijk that provides detailed information on ownership linkages between firms and across countries, contact information of private and public enterprises throughout the world, and information on firms' revenues and assets. ORBIS includes information on both listed and unlisted firms collected from various country specific sources, such as national registries and annual reports.

The main advantage of ORBIS is the scope and accuracy of its ownership information: it details the full lists of direct and indirect subsidiaries and shareholders of each company in the dataset, along with a company's global ultimate owner and other companies in the same corporate family. This information allows us to build links between affiliates of the same MNC, including cases in which the affiliates and the parent are in different countries. We specify that a parent should own at least 50 percent of an affiliate to identify an ownership link between the two firms. Thus, ORBIS allows us to identify the U.S. firms that are part of a larger multinational operation, distinguishing whether they are majority-owned U.S. affiliates of a foreign multinational, or U.S. parent firms that have majority-owned operations overseas. This gives us a more complete characterization of the operation of the MNCs located in the United States and their worldwide network of affiliates.

Most importantly for our analysis, ORBIS provides information on both, firms' names and physical addresses, making it possible to merge it with the BoL dataset, allowing us to identify which foreign exporters listed in the US import BoL belong to the same corporate group as the US importer and which ones are independent parties.

2.4 Natural Disasters

To identify firm-level idiosyncratic shocks, we consider major natural disasters occurring around the word during the sample period 2007-2020. This information is retrieved from two main datasets.

1. EM-DAT (Emergency Events Database): For disasters taking place outside the United States we use the Emergency Events Database (EM-DAT) compiled by the Centre for Research on the Epidemiology of Disasters (CRED), which contains essential core data on the occurrence and the effects of over 22,000 mass disasters around the world. Critical for our research, EM-DAT lists all locations and number of people affected by the disaster and the time period in which it took place. The locations are identified by ADMs and

then we aggregate the data at the ADM/time level, where time are weeks, months, quarters, or years.

2. SHELDUS (Spatial Hazard Events and Losses Database for the United States): For each natural disaster occurring in the United States, the SHELDUS database provides information on the start date, the end date, and the Federal Information Processing Standards (FIPS) code of all affected counties. Through GIS, an ADM code–our geographic unit of analysis—is assigned to each FIPS code. We then aggregate the natural disasters data at the ADM/time level, where time is measure in weeks, months, quarters, or years.

We restrict the sample to major natural disasters. Specifically, we only consider those disasters-ADM pairs above the 90th percentile of the distribution of i) the number of deaths, for disasters occurring outside the United States, and ii) of the damage to property (measured in constant U.S. dollars), for disasters occurring in the United States. Throughout the analysis we use two measures of the impact of disasters on firms. First, we construct a dummy indicating whether a firm is located in the same administrative level as the major natural disaster. Second, we calculate the Euclidean distance, in miles, from the firm's geolocation to the centroid of the administrative entity where the firm resides, to capture the notion that firms closer to the centroid might be more severely affected by the natural disaster.[9]

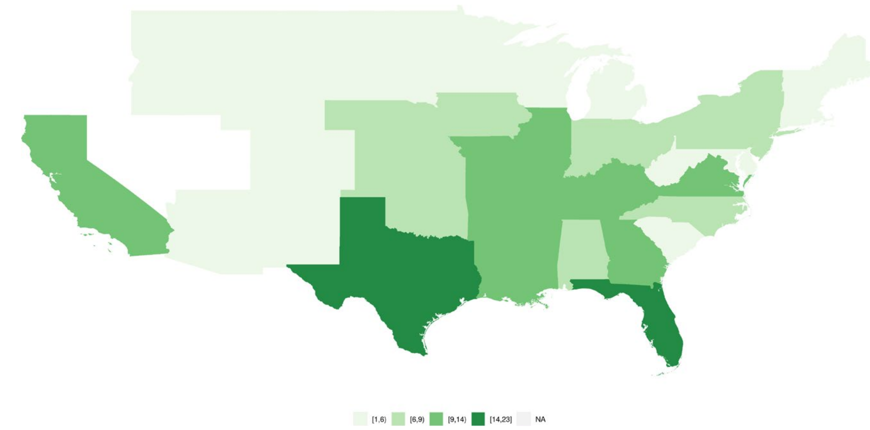
Figure 1 shows the geographical distribution of the number of natural disasters in the United States (1a) and in the rest of the word (1b) used in our sample. Similarly, Figure 2 and Figure 3 show the spatial distribution of the number of two types of natural disasters, floods and storms respectively, for the United States and for the rest of the world.

---

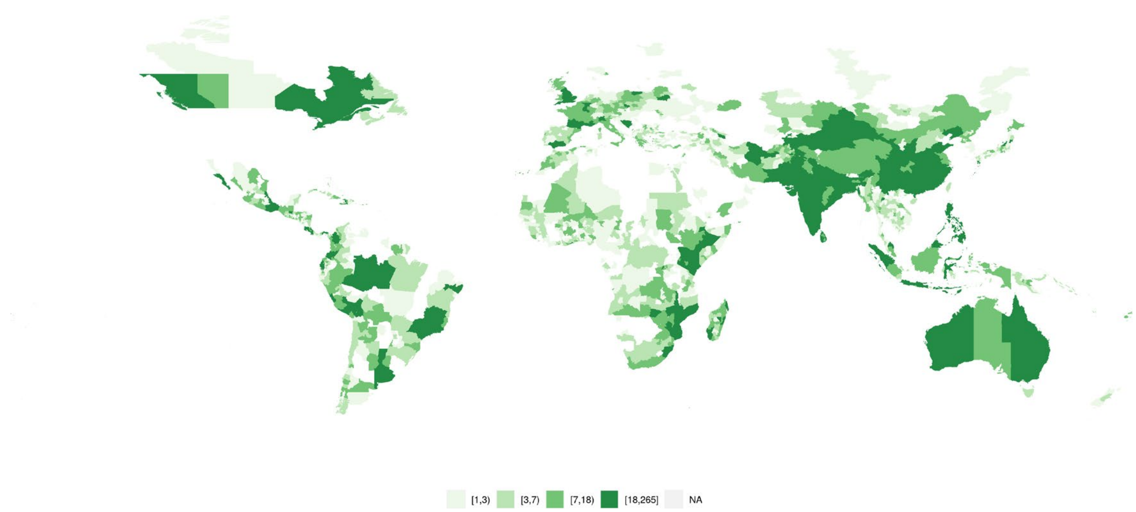
[9] The exact procedure we use to measure the impact of disasters on firms is as follows. First, we upload GAUL-FAO shape-files at administrative levels 1 and 2 and the geocoded information of all firms in our sample into QGIS (Quantum Geographic Information System), a software that supports the analysis of geospatial data. In QGIS, we assign a province (admin level 1) or a district (admin level2) to each firm, according to the administrative boundary within which the company is located. Additionally, QGIS provides the geographic coordinates (longitude and latitude) of the centroid of each administrative entity. Similarly, our data on natural disasters (EM-DAT and SHELDUS) contain detailed information on the location affected by each natural disaster at the admin level 1 mostly, or admin level 2 whenever available. Then we proceeded to construct measures of the impact of disasters on firms: First, we constructed a dummy indicating whether a firm falls in the same administrative level as the natural disaster. Second, using the information on the longitude and latitude of each firm in the sample, we calculated the Euclidean distance, in miles, from the firm's geolocation to the centroid of the administrative entity where the firm resides. This measure attempts to account of the fact that firms closer to the centroid are more likely to be more severely affected by the natural disaster.

Figure 1. Spatial Distribution of the Number of Natural Disasters

(a) United States



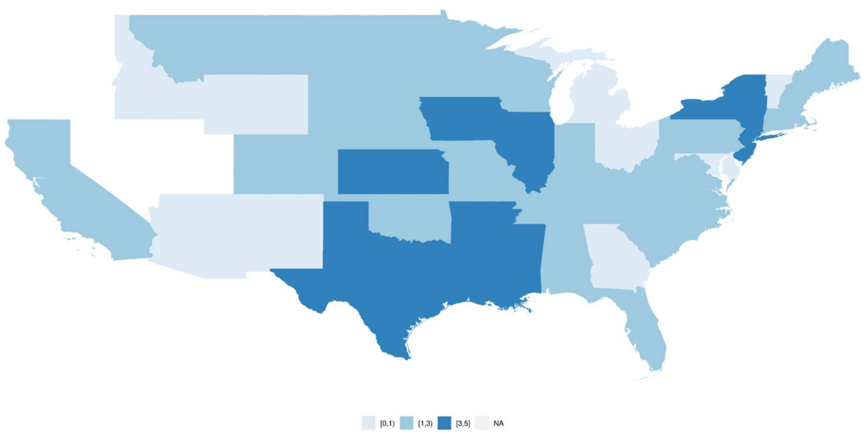
(b) Rest of the world



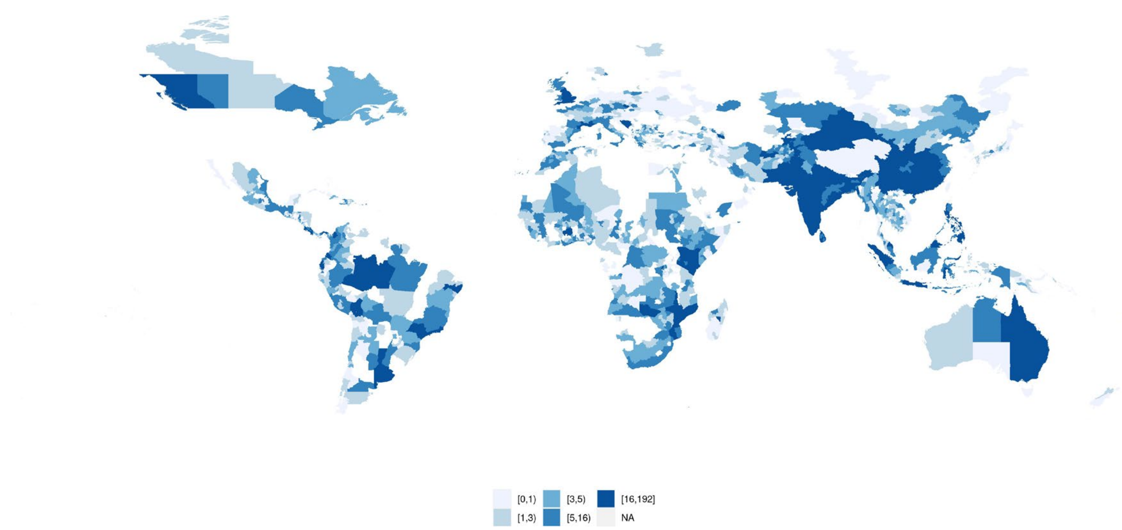
Notes: This figure shows the geographical distribution of the number of all natural disasters in the United States occurring during the period 2000-2019 (top panel), and in the rest of the world for the period 2000-2020 (bottom panel). Maps are from the FAO GAUL and use first-level administrative units. Information on natural disasters in the United States come from SHELDUS, and from EM-DAT for the rest of the world.

Figure 2. Spatial Distribution of the Number of Floods

(a) U.S.



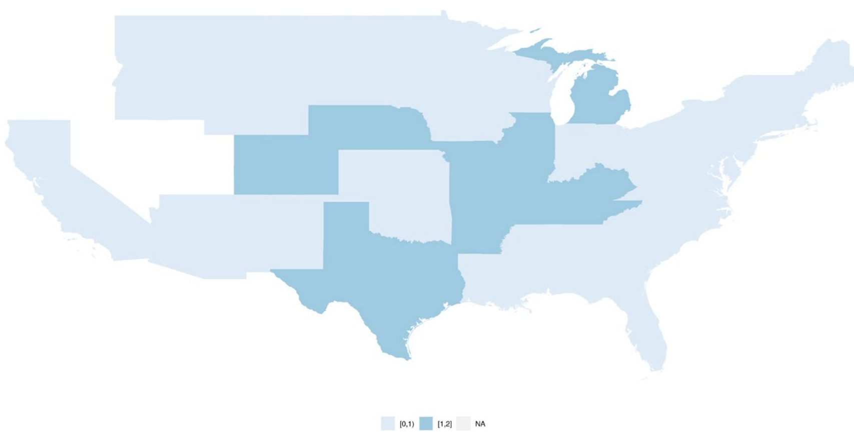
(b) Rest of the world



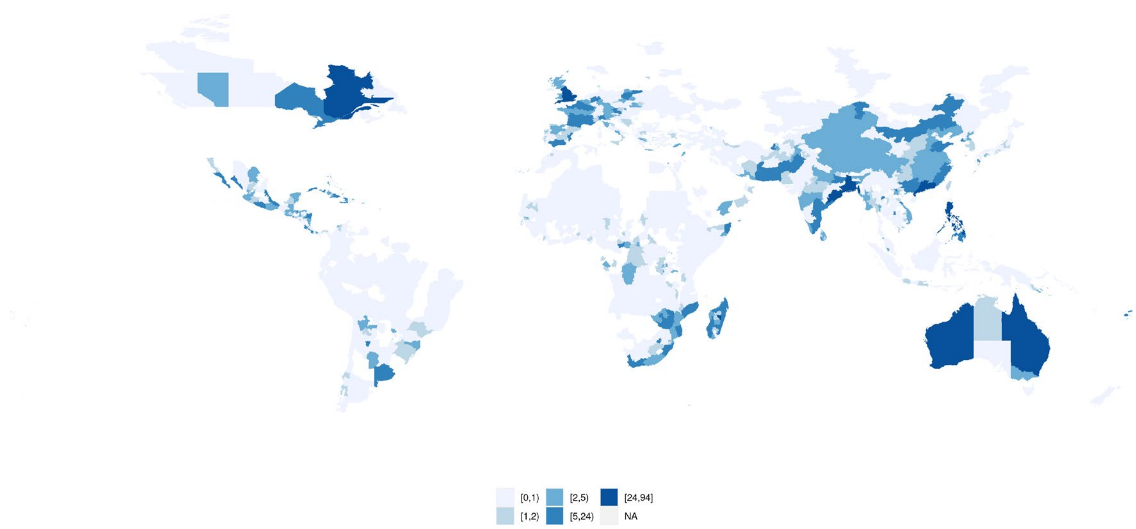
Notes: This figure shows the geographical distribution of the number of floods in the United States occurring during the period 2000-2019 (top panel), and in the rest of the world for the period 2000-2020 (bottom panel). Maps are from the FAO GAUL and use first-level administrative units. Information on natural disasters in the United States come from SHELDUS, and from EM-DAT for the rest of the world.

Figure 3. Spatial Distribution of the Number of Storms

(a) United States



(b) Rest of the world



Notes: This figure shows the geographical distribution of the number of storms in the United States occurring during the period 2000-2019 (top panel), and in the rest of the world for the period 2000-2020 (bottom panel). Maps are from the FAO GAUL and use first-level administrative units. Information on natural disasters in the United S come from SHELDUS, and from EM-DAT for the rest of the world.

Table 3 shows, for different types of natural disasters, the mean, standard deviation and key percentiles of the distribution of damage (measured in millions of constant US dollars) per affected county. Hurricanes, tornadoes, and floods are the most common natural disasters resulting in severe damage in the United States. Table 4 presents a similar breakdown for the rest of the world, with the severity of disasters measured by total deaths in the affected country. Worldwide, earthquakes and storms are among the most deadly natural disasters.

Table 3. Natural Disasters in the United States by Type (damage to property)

	Mean	SD	p25	p50	p90	p95	p99	Max	Count
Flooding	144.6	647.6	8.6	18.5	166.4	582.8	2,218.0	8,688.9	20
Hail	771.6	1,001.2	98.6	288.8	2,300.0	3,215.1	3,215.1	3,215.1	5
Hurricane/Tropical Storm	376.6	1,563.9	9.6	29.8	700.0	1,165.4	7,643.7	20,000.0	26
Severe Storm/Thunderstorm	39.1	64.4	7.3	8.1	182.7	208.1	208.1	208.1	4
Tornado	90.1	299.7	8.2	16.3	160.6	300.0	1,679.3	3,104.9	24
Wildfire	392.4	1,595.3	7.7	27.8	674.7	1,484.7	11,257.8	11,257.8	9
Wind	60.4	201.0	10.0	10.0	30.0	348.9	1,217.5	1,217.5	4
Winter Weather	22.2	25.7	5.6	9.4	70.0	70.0	70.0	70.0	2
Total	282.6	1,277.7	9.4	23.1	500.0	981.1	5,976.1	20,000.0	94

Note: The table shows, for different types of natural disasters in the United States, the mean, standard deviation and the 25th, 50th, 90th, 95th and 99th percentiles of damage to property per affected county in adjusted million U.S. dollars, base year 2017, between 2000 and 2019. The last two columns report the maximum damage per county and the total number of natural disasters in the United States for each type.

Table 4. Natural Disasters Worldwide by Type (total deaths)

	Mean	SD	p25	p50	p90	p95	p99	Max	Count
Earthquake	25,169	50,760	593	1,449	87,476	166,623	222,570	222,570	13
Extreme temperature	4,358	9,154	500	1,039	9,355	26,077	26,077	56,988	63
Flood	1,362	2,201	293	688	3,116	4,212	13,836	14,788	74
Landslide	399	267	255	388	472	1,102	1,765	1,765	2
Storm	8,276	31,042	260	502	3,300	138,366	138,366	138,366	26
Total	6,788	23,302	369	801	14,788	35,399	138,366	222,570	180

Note: The table shows, for different types of natural disasters, the mean, standard deviation and the 25th, 50th, 90th, 95th and 99th percentiles of the number of deaths of natural disaster worldwide. The last two columns report the maximum number of deaths and the total number of natural disasters worldwide for each type.

3. Summary Statistics

In this section we provide some key summary statistics of the combined dataset. Tables 5 and 6 are based on the linked Compustat, bill of landing, and natural disaster data. Table 5 reports the main facts of interest about U.S. importers among public firms (Compustat). In that group we have more than 220,000 importer-year observations. On average, 1.6 percent of importers were hit by a major U.S. natural disaster in any given time period. The chance that any of the importer's foreign suppliers was hit by a natural disaster is considerably higher, at 6.6 percent on average. The number of foreign suppliers per U.S. buyer follows a very skewed distribution, as is typical in firm-to-firm transaction data. The median U.S. buyer purchases from a single foreign seller in a given year, but a small set of buyers purchases from a large number of sellers.

Table 6 presents similar statistics, but from a point of view of the foreign sellers who have a public U.S. company as one of their customers. We have more than half a million of seller-year observations for such firms. The median foreign firm sells to a single U.S. importer, but again there is considerable skewness in the distribution of the number of U.S. customers. On average, there is a 14.3 percent chance that the foreign seller is affected by a natural disaster in its country. The probability that one of the seller's U.S. customers is affected by a major U.S. disaster is lower at 6.9 percent.

Overall, these numbers indicate that being hit by a major natural disaster—either directly or indirectly through customer and supplier relationships—is an event that is rare but not negligibly so. There is also considerable heterogeneity in the exposure of individual firms to natural disasters. These features suggest that natural disasters provide a good source of variation for studying the propagation of idiosyncratic shocks through global production networks.

Table 5. Descriptive Statistics, US Buyers

	Obs.	Mean	St. Dev.	p1	p50	p99
Disaster hits buyer$_t$	222,345	0.016	0.128	0.000	0.000	1.000
Disaster hits supplier$_t$	222,345	0.066	0.249	0.000	0.000	1.000
Number of suppliers$_t$	222,345	2.569	4.327	1.000	1.000	20.000

Note: In this table we present descriptive statistics about U.S. buyers from foreign partners. The sample is comprised of U.S. buyers that purchased from foreign partners abroad and for which we also matched with Compustat. The first row indicates whether the U.S. buyer was hit by a natural disaster. The second row indicates whether some supplier of a U.S. buyer was hit by a natural disaster. The third row denotes the number of suppliers for a given U.S. buyer.

Table 6. Descriptive Statistics, Foreign Sellers

	Obs.	Mean	St. Dev.	p1	p50	p99
Disaster hits seller$_t$	550,974	0.143	0.350	0.000	0.000	1.000
Disaster hits buyer$_t$	550,974	0.069	0.253	0.000	0.000	1.000
Number of buyers$_t$	550,974	1.504	1.713	1.000	1.000	7.000

Note: In this table we present descriptive statistics about foreign sellers to U.S. firms. The sample is comprised of foreign sellers that sold to U.S. firms and for which we were able to match U.S. buyers to Compustat. The first row indicates whether the foreign seller was hit by a natural disaster. The second row indicates whether a U.S. buyer of a foreign seller was hit by a natural disaster. The third row denotes the number of buyers for a given foreign seller.

4. Conclusions

In this technical note we offer a detailed explanation of the construction of the dataset used in Alviarez et al. (2021) to study how disruptions in international production networks propagate across sectors and countries.

To understand how the nature of production networks affects the propagation of shocks across countries, Alviarez et al. (2021) assemble a novel dataset that combines four micro-level datasets: i) the financial data of non-financial firms from Compustat; ii) the universe of U.S. maritime import transactions for the period 2007-2020, extracted from the U.S. bill of lading; iii) the ownership linkages between firms, within and across countries, extracted from Orbis; and iv) detailed information on the type, exact location and number of affected people of major natural disasters occurring worldwide during the sample period.

The combined dataset allows us to have a more complete portrait of the global operation and international trade transactions carried out by U.S. companies, including U.S. MNCs, and affiliates of foreign MNCs operating in the United States While none of these sources are new in isolation, our innovation on the data side is to merge them. To our knowledge, this is the first time these four datasets have been linked, enhancing our understanding of how disruptions in international production networks propagate across borders.

References

Almeida, H. et al. 2009. "Corporate Debt Maturity and the Real Effects of the 2007 Credit Crisis." NBER Working Paper 14990. Cambridge, United States: National Bureau of Economic Research.

Alviarez, V., and J. Blyde. 2021. "Firm-to-Firm Trade Networks, Supply Chains, and Access to Markets." Washington, DC, United States: Inter-American Development Bank. Unpublished working paper.

Alviarez, V., B. Cevallos, and T. Święcki. 2021. Cross-Border Intra-Firm Trade and the Propagation of Idiosyncratic Shocks." Unpublished working paper.

Barrot, J-N., and J. Sauvagnat. 2016. "Input Specificity and the Propagation of Idiosyncratic Shocks in Production Networks." *Quarterly Journal of Economics* 131(3): 1543–1592.

Bhargava, A. 2014. "Firms' Fundamentals, Macroeconomic Variables and Quarterly Stock Prices in the US." *Journal of Econometrics* 183(2): 241–250.

Boehm, C.E., A. Flaaen, and N. Pandalai-Nayar. "Input Linkages and the Transmission of Shocks: Firm-Level Evidence from the 2011 Tōhoku Earthquake." *Review of Economics and Statistics* 101(1): 60–75.

Carvalho, V.M. et al. 2021. "Supply Chain Disruptions: Evidence from the Great East Japan Earthquake." *Quarterly Journal of Economics* 136(2): 1255–1321.

Cravino, J., and A. Levchenko. 2017. "Multinational Firms and International Business Cycle Transmission." *Quarterly Journal of Economics* 132(2): 921–962.

Flaaen, A., A. Hortacsu and F. Tintelnot. 2020. "The Production Relocation and Price Effects of Us Trade Policy: The Case of Washing Machines." *American Economic Review* 110(7): 2103-27.

Flaaen, A. et al. 2021. "Bill of Lading Data in International Trade Research with an Application to the Covid-19 Pandemic." Finance and Economics Discussion Series 2021-066." Washington, DC, United States: Board of Governors of the Federal Reserve System.

Freund, C. et al. 2021. "Natural Disasters and the Reshaping of Global Value Chains." Policy Research Working Paper 9719. Washington, DC, United States: World Bank.

Li, X. et al. 2021. "The Effect of Revenue Diversification on Bank Profitability and Risk during the Covid-19 Pandemic." *Finance Research Letters* 43: 101957.