ONLINE APPENDIX

for

New Measurement of Export Participation in U.S. Manufacturing

Christoph E. Boehm, Aaron Flaaen, and Nitya Pandalai Nayar

Appendix A: Sources of U.S. Data on Firm and Establishment Level Exports

The original source for firm and establishment level export data from the U.S. Census Bureau comes from the quinquennial Census of Manufacturers (as well as the annual supplement, the Annual Survey of Manufacturers (ASM)). This survey asks establishments to report the dollar value of their shipments that are destined for foreign countries. The advantage of this survey-based question is the tight mapping between export values and a particular manufacturing plant ostensibly involved with the actual production of that exported product. There are numerous disadvantages of this source, however. There is no product or destination-level detail, it is only an annual measure of total exports, there are concerns of reliability due to the survey basis of the reporting, there is no information on the import side, and the longitudinal nature of the data is limited.

Detailed information on trade transactions are compiled by U.S. Customs for purposes of enforcing trade laws, documenting trade flows, and monitoring border security. The administrative documentation attached to these transactions offers rich detail, including the value, quantity, detailed product, country of export/import, port location, type of transaction, and more. Beginning with Bernard, Jensen and Schott (2009), the U.S. firm associated with these trade transactions was linked to other Census datasets, thus improving the information on firm-level trade for study by economists. The principal identifier for linkage is the employer identification number (EIN), as it was recorded by Customs Bureau documents accompanying individual shipments and also as part of the establishment/firm register in the Census Bureau.[12] The resulting Linked/Longitudinal Foreign Trade Transactions (LFTTD) database has been a very useful resource for trade economists studying import/export patterns by U.S. firms. One disadvantage, however, is that the EIN-based matching does not allow exports or imports to be assigned to individual establishments (plants) of the firm. The drawbacks of this limitation are discussed further below.

The final major resources for firm-level trade are the surveys of multinational firms collected by the Bureau of Economic Analysis. The level of aggregation is a mix between the firm and establishment (affiliate), depending on the direction of trade, size of the firm, or whether a particular year falls under the BEA's benchmark survey period. Like the survey-based information from the Census Bureau, the BEA data do not have extensive information on products or high-frequency detail of shipments. And while the focus on multinational firms provides for useful splits between arms-length and related-party trade, the BEA data do not have any information on non-multinational firms.

Appendix B: Details on Establishment-Level Allocation Methodology

This Appendix provides a shortened summary of the method developed by Boehm et al. (2021) on the construction of a dataset that assigns the firm-level LFTTD export transactions to individual establishments. Please refer to Boehm et al. (2021) for further details on their approach.

Their method utilizes information on products and industries, on the geography of where export transactions originate, and (to a limited extent) the CM and ASM. We describe each of these sources below, and then give a rough summary of how they combine this information to find the best establishment within a firm for each export transaction.

[12]On the export side, the EIN is listed on the "Shippers Export Declaration. The exceptions are shipments to Canada, which do not contain EINs but rather a field listing the firm name. On the import side, the Customs Forms (7501 and 7503) record the EIN representing the "ultimate consignee" of the imported goods.

B1. Industry-Based Matching

The goal of this industry-based measure is to establish a mapping between a particular exported product and a set of likely industries that could have produced that product. It relies on specific supplementary survey data from the Census of Manufacturers, as described below.

Every five years as part of the Census of Manufacturers, the Census Bureau surveys establishments on their total shipments broken down into a set of NAICS-based (6 digit) product categories.[13] Each establishment is given a form—specific to its industry—with a list of pre-specified products. There is also additional space to record other product shipments not included in the form. The resulting product trailer file to the CM allows a researcher to construct the set of industries that are primary producers of a given product.

There are several data issues that must be addressed before using the CM-Products file to infer information about the relative value of product-level shipments by a particular firm. First, the trailer file contains product-codes that are used to "balance" the aggregated product-level value of shipments with the total value of shipments reported on the base CM survey form. We drop these product codes from the dataset. Second, there are often codes that do not correspond to any official 7-digit product code identified by Census. (These are typically products that are self-identified by the firm but do not match any of the pre-specified products identified for that industry by Census.) Rather than ignoring the value of shipments corresponding to these codes, Boehm et al. (2021) attempt to match at a more aggregated level. Through an iterated process they try to find a product code match at the 6, 5, and 4 digit product code level, and use the existing set of 7-digit matches as weights to allocate the product value among the 7-digit product codes encompassed by the more aggregated level.

Finally, the link between the Harmonized Commodity Description and Coding System (or Harmonized System, HS) codes and Standard Industrial Classification System (SIC) and North American Industrial Classification System (NAICS) product codes is referred to as a SIC base or NAICS base, depending on which CM year is being used. These basecodes are up to 8 alphanumeric characters long, with shorter basecodes representing more highly aggregated products. Given linkage between either SIC or NAICS, the first four to six digits of the basecodes are called the baseroot. Each HS code has a single baseroot, while a baseroot might be associated with multiple HS codes. They use the NAICS (or SIC) to HS concordance from Pierce and Schott (2012), to map the information from the CM-Products file to the LFTTD trade data.

We now describe how Boehm et al. (2021) construct the set of "Production-Associated Industries (PAIs)" associated with a given product. Formally, let x_{pij} denote the value of shipments of product p by establishment i in industry j during a census year. Then the total output of product p in industry j can be written as:

$$X_{pj} = \sum_{i=1}^{I_j} x_{pij},$$

where I_{jp} is the number of establishments producing p in industry j. Total output of product p is then:

$$X_p = \sum_{j=1}^{I_{jp}} X_{pj}.$$

The share of product output accounted for by a given industry j is therefore:

$$S_{pj} = \frac{X_{pj}}{X_p}.$$

Because of reporting errors and aggregation of products, Boehm et al. (2021) designate an industry as a PAI of product p provided that its share S_{pj} passes a certain threshold – which they set at 5 percent.[14]

[13] The 1992 version of the CM used SIC-based product codes.
[14] They note that they have varied this threshold without affecting their results.

That is, they define the set of industries for product p for which $S_{pj} > 0.05$ as J_p. Boehm et al. (2021) match individual years of the LFTTD data to the closest available Census year.

To summarize, this procedure allows one to associate a set of industries with each exported product from the LFTTD; the industry identifiers of establishments in the LBD can then be used to match products to establishments.

B2. Location-Based Matching

Along with industry-based information on production, Boehm et al. (2021) also use geographic information to narrow the set of potential establishments involved with a particular trade transaction. Part of the shipper's export declaration form (now electronically administered via the Automated Exporter System) asks for address information on the USPPI where "goods begin their journey to the Port of Export." Both the zipcode and state information from this entry are included in the LFTTD microdata.

Although uncommon, some trade transactions in the LFTTD record a missing or incomplete zipcode. For these observations, Boehm et al. (2021) fill in missing zip-codes iteratively by replacing missing values with the largest zip-code value within the firm, country, month, HS code and baseroot observation. By attaching a U.S. location to an export transaction, this method may also assist in identifying the relevant establishment of export, though subject to a variety of limitations that are described in greater detail in Boehm et al. (2021). In short, one must be careful because the zipcode could be associated with an export transaction is the location of production or the location of export processing.

B3. Survey-Based Information

A final set of information used to assign firm-level exports to individual establishments is the export variable included in the CM and ASM. Although the CM should be comprehensive across all manufacturing establishments, one must be more careful in ASM years given that not all of a firm's manufacturing plants may be included in the survey sample. For this reason Boehm et al. (2021) use the export indicator from the CM/ASM as a way of distinguishing between a firm's plants only when the industry/location information specified above yields multiple plants associated with a given transaction.

B4. Assignment Procedure

The paragraphs below broadly outline how Boehm et al. (2021) combine these sources of information to assign all firm-level LFTTD exports to the most likely establishment associated with that export. This assignment will not always identify the establishment of export manufacture for several reasons. First, the PPI identified in the export declaration may be a non-manufacturing firm entirely that is solely involved in the export of the good; in such cases it is not possible to identify the establishment of production. Second, if the PPI is the firm of manufacture but processes shipments for export in a separate establishment, then our location information will point to a non-production establishment. Hence, while the assignment procedure described below attempts to prioritize establishments of production over non-manufacturing establishments, the data will often only identify the establishment involved with the export process. It is worth emphasizing that both production and non-production establishments involved with exporting activity will be impacted by trade and export markets.

To retain as much detail as possible, Boehm et al. (2021) take the raw LFTTD export data and aggregate only up to the firm, product, country, month, zipcode, port, and export-method (rail, air, etc) level. Next they make copies of each of these export observations and merge them to all of the firm's establishments. Because the LBD only registers a firm if it existed on March 12th of a particular year, some firms could be trading in the LFTTD but not exist in the LBD. To remedy this issue, Boehm et al. (2021) match the trade data not found in the LBD for that period with samples from the year prior and the year following. Using this large dataset, they retain the most likely establishment for each trade observation according to an iterative set of rules, decreasing in the degree of confidence in the establishment match.

- **Case 1: Single unit firm.** The assignment is a trivial exercise for those firms having only one establishment. Boehm et al. (2021) first remove these transactions, but flag whether these establishments are manufacturing or non-manufacturing, and whether the establishment records positive export shipments in the ASM/CM.
- **Case 2: Unique zip code match: manufacturing.** If a single zip code matches to a unique establishment, then the relevant trade is assigned to that establishment.
- **Case 3: Non-unique zip code match to PAI establishment** If there is only one establishment matching the zip code that also matches based on their PAI criteria, then they allocate all trade to the zipcode match that also aligns with the appropriate industry. Given that an HS code is assigned to multiple PAIs, it is possible for there to be several establishments matching this case. Absent any distinguishing information on export activity from the ASM/CM, they use employment weights to allocate the exports across establishments.
- At this stage for any unassigned export transactions, Boehm et al. (2021) loop back through cases 3 through 6 but looking for PAI establishments at the NAICS-5, and then NAICS-4 industry basis.
- **Case 4: No zip code match but unique PAI establishment.** In this case, they simply assign all trade to the unique PAI establishment.
- **Case 5: No zip code match but non-unique PAI establishments.** For multiple establishments matching a PAI, they use employment weights to allocate the trade across matching establishments (absent distinguishing information on export status from the ASM/CM).
- **Case 6: Non-unique zip code match, and no PAI establishments.** They split these cases into those matching manufacturing establishments with those matching non-manufacturing establishments. If there are multiple establishments with the same zip code (which they mention is rare), they continue to use employment shares as weights.
- **Case 7: No zip code match and no PAI establishments.** For this final case, they first assign the export transaction to all manufacturing establishments of the associated firm (using either ASM/CM export share weights or employment weights). They also prioritize establishments in the wholesale (NAICS 42) and transportation/warehousing (NAICS 48) industries, based on the distribution of exports from prior matches. If there are no establishments in any of these industries, then the final step assigns the export transaction to all other establishments based on employment shares.[15]

B5. Characteristics of the Allocation

Figure B1 documents the share of overall U.S. exports that is assigned according to the hierarchy of cases as described above.

Appendix C: Additional Results

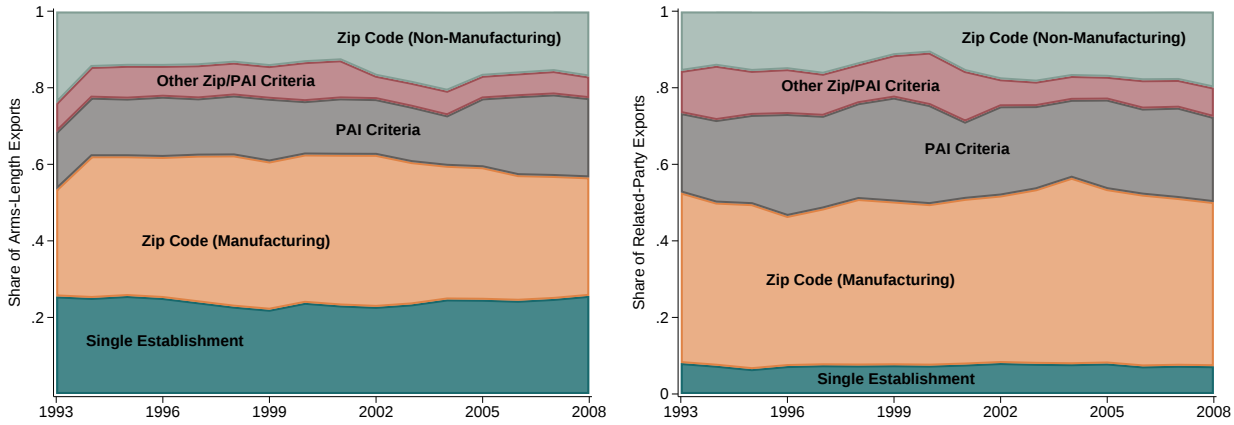
Figure C1 replicates the kernel density estimates of export intensity by source from Figure 1, but for the year 2007. Similarly, we report corresponding full detail of the matrix of alignment between the CM and LFTTD-based measures of export participation for 2007 in Table C1, analogous to Table 2 in the main text.

Changes in measured export participation by establishments and firms over two Census years–2007 and 2012—may also be informative to assess reasons for misalignment. We therefore match manufacturing establishments and firms across the 2007 CM and 2012 CM (while also noting which establishments or firms are missing in each year) and retain the various export participation measures in each year. The results are shown in Table C2. The percentages in each panel of the table sum to 100,

[15]Note that for these cases they take the top 20 establishments by employment size, as there are some firms with thousands of establishments.

Figure B1. : Share of Exports by Type of Establishment Allocation

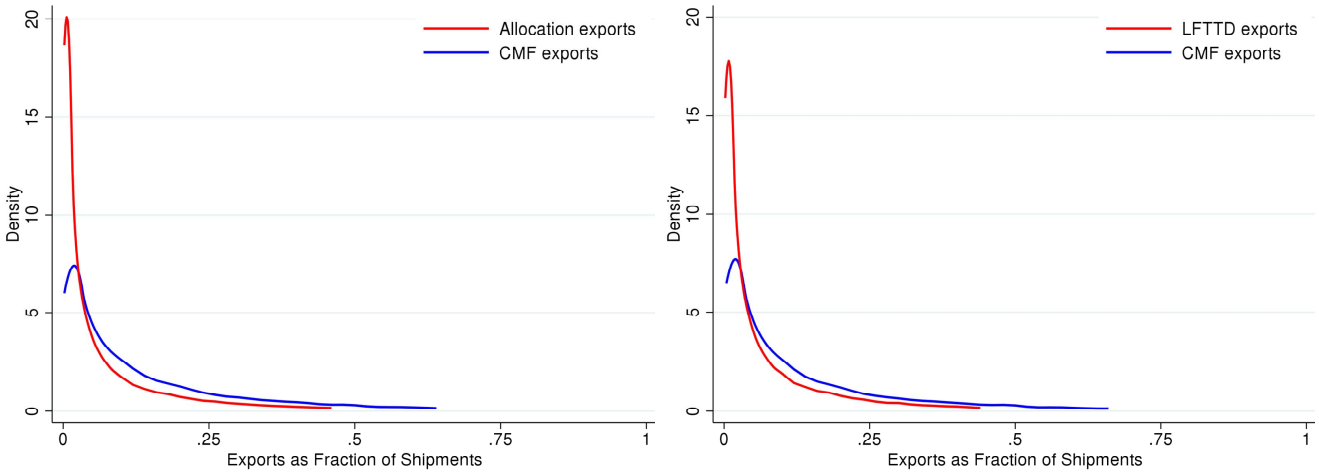
(a) Non-Related Party Exports (b) Related Party Exports



Source: Reproduced from Boehm et al. (2021)'s calculations as explained in text.

Figure C1. : Export Share of Shipments: Distribution by Source of Exports

(a) Establishment-Level (b) Firm-Level



Sources: Author's calculations using LFTTD and CMF as explained in the text.
Notes: Both density estimates are constructed separately and integrate to unity. They exclude establishments or firms that report zero exports. Per Census disclosure rules, the top and bottom 5 percent of each distribution is truncated. *Allocation* refers to exports measured at the establishment level using the dataset constructed in Boehm et al. (2021). Density estimates are for Census year 2007. Results for Census year 2012 are reported in Table C1.

and therefore represent the full set of establishments or firms in the two years of analysis. Beyond the issue of measurement, it is possible to construct a transition matrix for exporting, birth, and death in manufacturing over a five-year horizon from this table.

Table C1—: Plant/Firm-Level Export Participation, by Source

Panel A: Establishment-Level

LFTTD Exports	CM Exports: No	CM Exports: Yes	
No	56%	3%	59%
Yes	23%	18%	41%
	79%	21%	

Panel B: Firm-Level

LFTTD Exports	CM Exports: No	CM Exports: Yes	
No	61%	3%	64%
Yes	20%	16%	36%
	81%	19%	

Notes: The table reports the fraction of overall plants and firms identified as exporters based on the source of data. The statistics are for Census year 2007. Table 2 reports analogous numbers for Census year 2012.

Persistence of exporting

The numbers reported at the end of section III are calculated from Appendix Table C3 as follows. For establishments, conditional on having positive exports in the LFTTD and zero exports in the CM in 2007 and conditional on survival, the probability of continuing to export in 2012 according to at least one of the sources (LFTTD, CM, or both) is 25.6% + 51.4% + 1.6% = 78.6%. Similarly, conditional having positive exports in the CM and zero in the LFTTD in 2007 and conditional on survival, the probability of continuing to export in 2012 according to at least one of the sources (LFTTD, CM, or both) is 33.3% + 14.9% + 19.0% = 67.3%. The difference of 78.6% and 67.3% (= 11.4%) is reported in the text. The calculations are analogous for firms using the numbers in Panel B of Appendix Table C3.

Table C2—: Establishment/Firm-Level Export Persistence Over Time, by Source

Panel A: Shares of Establishments

	Positive LFTTD & CM	Positive LFTTD & Zero CM	Positive CM & Zero LFTTD	Zero CM & LFTTD	Don't Exist
2007			*2012*		
Positive LFTTD & CM	8.7%	2.2%	0.3%	0.6%	2.6%
Positive LFTTD & Zero CM	3.3%	6.6%	0.2%	2.7%	5.6%
Positive CM & Zero LFTTD	0.6%	0.3%	0.3%	0.6%	0.7%
Zero Exports	1.0%	3.3%	0.7%	21.5%	18.7%
Don't Exist	2.2%	3.6%	0.6%	13.3%	

Panel B: Shares of Firms

	Positive LFTTD & CM	Positive LFTTD & Zero CM	Positive CM & Zero LFTTD	Zero CM & LFTTD	Don't Exist
2007			*2012*		
Positive LFTTD & CM	7.0%	1.8%	0.2%	0.4%	3.2%
Positive LFTTD & Zero CM	2.6%	5.1%	0.2%	2.1%	5.5%
Positive CM & Zero LFTTD	0.4%	0.2%	0.3%	0.6%	0.8%
Zero Exports	0.8%	2.7%	0.7%	22.1%	20.9%
Don't Exist	2.8%	3.8%	0.7%	15.1%	

Notes: The table reports the fraction of overall plants and firms identified as exporters based on the source of data that persist as exporters in the data between 2007 and 2012. For details on the data, see this appendix. Establishment level exports are constructed by Boehm et al. (2021). The top panel reports establishment statistics and the bottom panel reports firm statistics.

Table C3—: Establishment/Firm-Level transition matrices conditional on survival

Panel A: Transition Matrix for Establishments

	Positive LFTTD & CM	Positive LFTTD & Zero CM	Positive CM & Zero LFTTD	Zero CM & LFTTD
2007		*2012*		
Positive LFTTD & CM	74.1%	18.9%	2.3%	4.8%
Positive LFTTD & Zero CM	25.6%	51.4%	1.6%	21.4%
Positive CM & Zero LFTTD	33.3%	14.9%	19.0%	32.7%
Zero Exports	3.9%	12.6%	2.5%	81.0%
Don't Exist	10.9%	18.3%	3.1%	67.7%

Panel B: Transition Matrix for Firms

	Positive LFTTD & CM	Positive LFTTD & Zero CM	Positive CM & Zero LFTTD	Zero CM & LFTTD
2007		*2012*		
Positive LFTTD & CM	74.3%	19.3%	2.2%	4.2%
Positive LFTTD & Zero CM	25.7%	51.1%	2.0%	21.1%
Positive CM & Zero LFTTD	25.8%	13.2%	21.9%	39.1%
Zero Exports	3.0%	10.2%	2.8%	84.0%
Don't Exist	12.7%	16.9%	3.1%	67.3%

Notes: The table reports the transition matrices conditional on survival between 2007 and 2012 for establishments and firms, where a state is defined as indicated by the row and column labels. The matrices are constructed from the data reported in Table C2.