The Local-Area Incidence of Exporting*

Christoph E. Boehm
University of Texas at Austin

Aaron Flaaen
Federal Reserve Board of Governors

Nitya Pandalai-Nayar
University of Texas at Austin and NBER

Jan Schlupp
Amazon

September 2021

Abstract

We construct a new micro-dataset of U.S. export transactions at the plant level by mapping confidential firm-level exports to their underlying establishments. These data provide new facts about the geography of U.S. exports both across and within states, with important practical applications to empirical work linking trade to local labor market outcomes. The data reveal U.S. exports to be much more geographically concentrated than both employment and manufacturing sales. Motivated by this fact, the paper applies these data to study the effects of exporting on local labor markets during the trade collapse of the Great Recession. Counties experiencing greater declines in foreign demand performed worse in terms of employment, pay, and wages during the Great Recession. We also show that a similar analysis with imputed export data—a common practice in the literature—fails to replicate these estimates.

*Email: Boehm-chris.e.boehm@gmail.com; Flaaen- aflaaen@gmail.com; Pandalai-Nayar- npnayar@utexas.edu; Schlupp- janschlupp@gmail.com. We thank Andrei Levchenko and Justin Pierce for helpful comments. Any opinions and conclusions expressed herein are those of the authors and do not necessarily represent the views of the Board of Governors or its research staff. All results have been approved for disclosure by the Census Bureau.

1. Introduction

The remarkable expansion in the availability of micro-level data in recent decades has added much to our understanding of the heterogeneity underlying national import and export patterns. Yet, despite these advances, relatively little is known about the international trade activity of individual locations in the United States. The reason is that common U.S. datasets contain very limited information on the destination of imports and the geographic origin of U.S. exports; while information on the port of exportation or importation is available in U.S. microdata, the travels of goods within the U.S. is typically not known. Firm-level data in general does not solve this problem as large, multi-plant firms account for the vast majority of U.S. trade. As a result, studies that focus on the local effects of trade in the U.S. typically use an imputation approach that pairs the local or regional industrial structure with aggregate industry-level data on imports or exports (e.g., Hakobyan and McLaren, 2016; Feenstra, Ma and Xu, 2019).

This paper makes four contributions that together improve our understanding of local patterns of exporting in the United States. First, we construct a new dataset that maps firm-level exports to the establishment level, thereby identifying the precise location of export activity. This methodology opens up new approaches to understanding regional export patterns and provides an alternative to imputation strategies. Second, we use these new data to document several facts about exporting at the establishment level and by region. Third, we study the local labor market implications of the Great Trade Collapse of 2008-2009 and illustrate that areas in the U.S. more exposed to foreign demand shocks because of their exporting patterns were more severely impacted in terms of employment, payroll and wage declines. Fourth, we show that common approaches of imputing local area exports using product-level export data and local industry employment shares cannot replicate the local-area export incidence documented here, and should be used with caution.

At the core of our analysis is a new establishment-level dataset of export transactions with full destination and product detail. Beginning with restricted-access, firm-level export transaction data from the U.S. Census Bureau (Kamal and Ouyang 2020), we utilize several additional pieces of information to "allocate" the transaction to the appropriate establishment, and hence precise location, from which the transaction originated. This approach is an improvement on plant-level exporting data from the Census of Manufacturers which excludes certain types of trade and provides no destination or product-level detail. It is also an improvement over public-use state-level export data based on "origin-of-movement" indicators, for reasons we describe in detail below. While a number of important complications exist—the firm of export is not always the firm of production, warehousing, intra-firm export consolidation, and the like—the dataset we create provides researchers a more detailed look at the local patterns of exporting than was previously possible.[1]

Using this novel dataset, our second contribution is to document a number of new facts on the composition, heterogeneity, and concentration of export activity in the United States. First, we show that geography plays a role in export specialization patterns at the state-level as, for example, states close to the Canadian border tend to export relatively more to Canada. Second, we find that exporting is more common than previously reported, as survey estimates of export participation by manufacturers appear to under-estimate small export quantities by establishments and firms. Third and perhaps most consequentially, we find that exports are significantly more geographically concentrated than either employment or sales, and that this pattern holds both nationally and within states. Furthermore, the degree to which these variables differ in their concentration is heterogeneous across states, which further highlights problems that may arise when using employment to impute the exports to local areas.

We then use these new data to offer a fresh perspective on the connection between exporting activity and local labor markets. Aggregating the micro-data to the county-level, we study the regional heterogeneity in outcomes arising from exporting during the Great Recession and the associated trade collapse. Using an updated version of a traditional World Import Demand (WID) instrument to isolate foreign demand shocks that are plausibly orthogonal to U.S. demand declines and other confounders, we provide evidence on the employment and wage effects of export activity in the United States. Specifically, we find that counties which were more exposed to negative foreign demand shocks performed worse in terms of employment, pay, and wages during the Great Recession. In our preferred specification, we estimate that a one million dollar decline in exports led to the loss of 17 jobs in the county. Further, county-level payroll declined approximately one-for-one with a decline in exports, and a one thousand dollar drop in exports per worker led to a 1.9 percent decline in the local wage rate. We further show that these effects come not just from the direct employment effects of manufacturing exports, but also because of local, within-county spillovers to non-manufacturing industries.

Our final contribution is to assess whether the common approach of "imputing" local exports from industry-level employment data and more aggregate export data is problematic or not. To do so, we repeat our analysis of the effects of the Great Trade Collapse during the Great Recession using imputed county-level exports from publicly available data. We find

[1] It is important to note that most of these complications will also affect all publicly available export data.

the resulting estimates to be generally statistically insignificant, noisy, and suffering from weak instruments. The discrepancy with our main estimates likely reflects the fact that the imputation relies on the proportionality of employment and exports, which—as we have shown—does not hold in the data. In sum, our evidence shows that imputing local area exports can provide a distorted view of true local exporting patterns, and may imply that this practice is problematic for other analyses as well.[2]

The allocation of U.S. firm-level exports to individual establishments offers the potential for the study of other new features of trade patterns and their effects on the U.S. economy. One such example is that our new data set makes analyses at the establishment level possible. Prior to the construction of our new dataset, export and import transactions in the U.S. were only measured at the firm-level. However, many other surveys and censuses are conducted at the establishment level. To use trade data together with these datasets from the U.S. Census, researchers typically needed to aggregate establishment-level data to the firm level—which is often imprecise since most surveys only capture a subset of a firms' establishments. Our new dataset enables researchers to instead measure export transactions at the establishment level, and hence such aggregation is not necessary. The analysis can instead be performed at the establishment level where researchers can use the available sampling weights.

Further, our dataset allows us to measure exports by region more precisely than was previously possible. As our primary analysis illustrates, this allows researchers to document regional properties of export activity and to further study the effects of shocks on local labor markets.

Related Literature. This paper is related to several strands of the literature. A number of previous papers in the U.S. and other countries have attempted to construct or study sub-national and local trade patterns. For instance, Santamaría, Ventura and Yeşilbayraktar (2020) create a dataset of regional European trade using a new survey, the European Road Freight Transport survey (ERFT), from Eurostat. In the U.S., the Commodity Flow Survey (CFS) from the U.S. Census has been used to study intra-U.S. trade (see Allen and Arkolakis

[2]The practice of imputing trade using employment weights is potentially innocuous for the case of imports. The proportionality assumption underlying the imputation may better capture the effects of import competition on a given industry. The employment impacts come from the impact on overall market demand, and not from any distributional assumptions on firm or consumer location. By contrast, a foreign shock affecting demand for an industry's exports would be proportional to national employment only if export activity across regions or local areas were distributed identically to employment. An established literature dating back to Bernard and Jensen (1995) has emphasized that only a small fraction of firms export, and other work has shown exports to be more skewed than employment nationally (e.g. Mayer and Ottaviano (2008) and Bernard et al. (2007*c*)). In this paper, we show that exports are also more skewed than employment or sales across geographies, suggesting that the assumption underlying a standard imputation of exports to local geographies could generate misleading results.

(2014) and Atalay, Hortaçsu and Syverson (2014)). Our dataset supplements these studies by providing data on foreign exports by product and destination, down to the level of individual establishments and counties.

Our paper is naturally related to a long line of influential research studying local labor market effects of trade shocks. Notable papers in this literature include Autor, Dorn and Hanson (2013), Feenstra, Ma and Xu (2019) and many others (China shock); Topalova (2007) (Indian trade liberalization); Kovak (2010) and Dix-Carneiro and Kovak (2017) (Brazil); Hakobyan and McLaren (2016) (NAFTA); and Benguria and Saffie (2019) and Benguria and Saffie (2020) (U.S.-China trade war). This line of research typically constructs local exposure to a shock by using national or state-level data mapped to the local level with employment weights, and often finds large and heterogeneous effects across space. With the emphasis on estimating the employment impacts of exporting, our paper is perhaps most closely related to Feenstra, Ma and Xu (2019). Broadly speaking our application to the Great Trade Collapse extends their work with a more direct export measure and alternative empirical setup that focuses on a specific episode. Our data could also be used to distinguish between related party and non-related party trade, which is likely to be useful to study the effect of multinational activity on U.S. local labor markets (Kovak, Oldenski and Sly (2021) and Boehm, Flaaen and Pandalai-Nayar (2020)). For questions in international macro focusing on the transmission of shocks, our dataset has the advantage of containing monthly or even daily export data for local regions.[3]

Finally, our paper is also related to the large literature studying the trade collapse during the Great Recession. This literature explores reasons for the trade collapse, for instance, the role of intermediate input trade or other demand factors (Levchenko, Lewis and Tesar, 2010; Bems, Johnson and Yi, 2013; Bussière et al., 2013) or analyses the role of inventories (Alessandria, Kaboski and Midrigan, 2010*b*,*a*).[4] To the best of our knowledge, no work has directly assessed the local labor market effects of exporting during the trade collapse. In closely related work, House, Proebsting and Tesar (2019) study the geographic heterogeneity in the effects of exchange rate shocks based on heterogeneity in industry composition and state trade flows, and find empirical evidence consistent with our results for the trade collapse.

[3]See also Caliendo, Dvorkin and Parro (2019) and Caliendo et al. (2018) for quantitative models studying local labor markets and general equilibrium implications using employment share based approaches for motivating patterns.

[4]The literature on the trade collapse of the Great Recession is vast, including work by Eaton et al. (2016), Ahn, Amiti and Weinstein (2011), Bown and Crowley (2013), and Feenstra, Li and Yu (2014), among many others.

2. A New Dataset of U.S. Export Transactions at the Plant Level

This paper develops a new data resource that provides researchers a previously unavailable level of detail on U.S. export transactions. In this section we summarize the methodology behind this new establishment-level dataset of export transactions; we then subsequently highlight new facts revealed by this data along with comparisons to other data to validate its accuracy.

The new dataset we construct in this paper builds on a number of restricted-use Census Bureau datasets that have become important resources for economists studying firm-level decisions. Prior to this paper, the highest geographic granularity for export transactions would be from the survey-based indicator of total exports in the Census of Manufacturers (CMF), produced quinquennially in years ending in a 2 or 7. The CM (along with the annual survey in intervening years) provides estimates of *establishment-level* data that could ostensibly be tied to a particular labor market.[5] A long line of research dating back to Bernard and Jensen (1995) and summarized in Bernard et al. (2007*a*) have made important contributions to our understanding of establishments and firms engaged in trade using this source for exporting activity. The drawbacks of this source are that it is survey-based, infrequent, and that it contains no product, destination country, or shipment-level detail.

A second core piece of restricted-use Census data on export activity, and a primary source for the dataset used in this paper, comes from the Longitudinal Foreign Trade Transactions Dataset (LFTTD). This firm-level dataset on export and import transactions is constructed from a partnership between U.S. Customs and the Census Bureau (see Kamal and Ouyang (2020) for more details). The LFTTD contains the universe of U.S. trade transactions in goods that has been linked to firms via an Employer Identification Number (EIN) as well as other sources, and contains the date, value, quantity, and detailed product information (HS10) of a trade transaction (both import and export) along with whether the transaction was conducted between related parties or at arms-length. An advantage of these restricted access data is the ability to link to other Census Bureau datasets—most prominently the Longitudinal Business Database (LBD) containing a longitudinally consistent register of establishments/firms and their employment, payroll, etc.

An important point of emphasis is that the LFTTD export data matches an individual

[5]Another potential source of granular information on export activity comes from the Commodity Flow Survey (CFS). Unfortunately, this survey data relies on a relatively small sample, is infrequently administered, and generally lacks country of destination detail.

trade transaction to a *firm* identified as the U.S. principal party in interest (USPPI); it does not match an export transaction to an individual establishment or location within that firm.[6] Our methodology uses three primary indicators to locate the establishment within the firm identified in the LFTTD that most likely produced the goods traded in a given export transaction.[7] Each of these indicators has advantages as well as limitations, and thus by combining this information together we can arrive at a more accurate distribution of exports across the production locations of the firm.

First, we utilize a variable in the LFTTD that identifies, for each transaction, the geographic "origin of movement": where the shipment began its journey to the port of export. Of course, this origin of movement need not always be the same as the origin of production, as discussed in some detail in Cassey (2009). The three most salient reasons for a disconnect are:

- Consolidation: if a shipment combines with similar commodities from the same firm then only the location of consolidation is recorded for that shipment. This is most common with agricultural shipments.

- Wholesale/Retail: if the exporter is a wholesaler or retailer, then the location of origin of the wholesaler/retailer is recorded, and not the location of production.

- Warehousing: a shipment could be sent to a storage or warehouse facility and then subsequently exported. In the case where the export process begins at this location, then the origin of movement will be recorded as the warehouse facility, and not the location of production.

Thus, while the origin of movement variable provides useful information for the allocation of exports to plants, these features demonstrate the importance of combining it with other information where relevant. Relative to the publicly available state-level export data that is based solely on the origin of movement variable, we can identify many instances where this would provide misleading information—for instance, because the firm has no plant at the origin of movement.[8]

A second piece of information for identifying the appropriate establishment of a firm's trade transaction comes from connecting the products being exported to an establishment

[6]The USPPI is defined as the the person or legal entity in the United States that receives the primary benefit, monetary or otherwise, from an export transaction.

[7]While firms produce and export goods in most theoretical models, it is important to recognize in the data that there may exist instances in which the act of exporting is undertaken by a non-manufacturing firm (see e.g. Bernard et al. (2010), Bernard and Fort (2015)).

[8]See `https://usatrade.census.gov/` for more information on the state-level exports data.

in an industry typically associated with the production of that product. To create a mapping between products and industries, we use the "Products Trailer file" of the Census of Manufacturers, which provides product-level detail on the shipments identified by the establishment. We use this information to construct, for each exported product, a set of "Production-Associated Industries" (PAIs). This product-industry mapping can be used to narrow the potential establishments of a firm that are likely producers of a given exported product (as each establishment of a firm has a unique industry classification).

Finally, we leverage the establishment export variable as reported in the CMF and/or Annual Survey of Manufacturers (ASM) to assist in identifying likely export establishments within a firm. In this last step, where other data are silent, we utilize the shares of overall export activity across a firm's establishments from the CMF or ASM to allocate that firm's exports.

Appendix B provides the full details for how we use each of these sources of information to identify the most likely establishments associated with a firm's export transactions. As shown in Appendix Figure A1, the result is a large fraction of overall exports allocated with a high degree of confidence in the assignment to the correct establishment producing the exported good. The share of exports accounted for by unique matches to zip-code level origin of movement indicators or PAI criteria, as well as single-unit establishments (where the allocation is trivial) account for roughly three-quarters of exports in a given year.

The resulting dataset provides greater granularity and detail of U.S. export transactions than has been previously available to researchers. The wealth of details on individual export transactions—such as product, destination, value, quantity, and arms-length status—are mapped to a precise location of origination for the first time.

3. Facts on the Local Geography of Exports

We next document several facts about the incidence of exports within the U.S. that come out of this data. These facts serve three purposes. First, we aim to compare our new dataset to the extent possible to existing data. Second, since closely comparable datasets are not available, we provide several facts that serve as plausibility checks on our new data. Lastly, elements of the facts we document are novel and help further our understanding of the spatial features of exporting in the United States.

Table 1: Public Use Data Vs Allocated Data

States	Shares (in percent) Public Data	This paper	Percentage point Difference	Percent Difference
	Top 5 states: Public Data > Data in this paper			
Texas	16.3	10.5	5.7	54.6
Washington	4.5	1.8	2.6	139.7
Louisiana	1.8	1.3	0.5	42.0
Florida	4.2	3.7	0.5	13.3
Kentucky	1.9	1.6	0.3	21.1
	Bottom 5 states: Public Data < Data in this paper			
New York	5.9	6.5	-0.7	-10.3
New Jersey	2.8	3.7	-0.8	-21.7
Missouri	1.2	2.1	-0.8	-39.8
California	12.3	13.5	-1.3	-9.4
Michigan	4.2	6.5	-2.3	-35.1

Sources: Author's calculations using LFTTD, CMF as explained in the text, and Census Bureau State-Level Trade.
Notes: The figure shows the largest discrepancies of state-level exports between the public origin-of-movement data and our data for the year 2007. State-level exports are measured as a percent of total exports. A positive percent difference indicates that the public use data attributes more exports to a specific state than the data used in this paper. Hawaii is excluded due to disclosure constraints.

3.1 Comparison to State-Level Origin-of-Movement Exports

We begin by aggregating our export data to the state-level. Doing so helps compare them to the more readily-available public data that is based solely on the origin of movement variable. As our new dataset does not include agricultural and mining products we compare state-level exports to the closest analogue from of the publicly available Census data. The total value of trade is similar, as our methodology accounts for roughly 99 percent of the value identified in the Census public use data product in 2007.

The cross-sectional correlation of state-level exports between our measure and the public Census Bureau data product in 2007 is reasonably high at 0.94, which partly reflects the overlap in data sources. Appendix Figure A2 provides a scatter plot with state labels. We also correlate the time series at the state level. These correlations are around 0.9 on average, with the 10th and 90th percentiles being 0.63 and 0.99.

While we expect relatively high correlations at this level of aggregation, we next highlight notable differences. Table 1 shows for the year 2007 the top and bottom five largest discrepancies in percentage point terms between the public use data and the data constructed in this paper. Notable, this list features both small states such as Louisiana, and large states such as California and Texas.

In Appendix C we provide further documentation on the comparison to this alternative dataset, documenting a high overall correlation along with several areas of disagreement between the two sources of data.

3.2 Distance Matters Within the U.S.

The well-known gravity relationship predicts that, all else equal, trade should decrease with distance. For countries other than the United States, research has shown that distance matters for international trade even within country borders.[9] We check this feature in our new data for the U.S.—imperfectly, admittedly, as not all is held equal. Figure 1 displays the intensity of exports to select countries, where we draw out the specialization patterns by reporting the differential share of exports to a particular destination, relative to the U.S. as a whole. Specifically, for exports of state i to destination k we report:

$$\frac{x_{ik}}{\sum_j x_{ij}} - \frac{X_k^{US}}{\sum_j X_j^{US}}. \tag{3.1}$$

The figure demonstrates the well-known feature that distance matters. States close to the Canadian border have a higher-than-average share of Canadian exports (Figure 1a). In a similar fashion, exports to Mexico (Figure 1b) tend to be over-represented in states along the southern U.S. border, and exports to China (Figure 1c) are higher than average for states along the West Coast. On the other hand, Figure 1 shows that distance does not explain all important variation in state-level exporting. One important example is the clustering of North American exports in parts of the Midwest such as Michigan, Ohio, and Pennsylvania.[10] Note that these results also serve as a plausibility check on our new data.

3.3 Small Exporters Are More Common

A more novel lesson from our new data resource is that exporting among manufacturing firms is more common than previously thought, with most of these "new" exporters responsible for only small volumes of exports. This new fact stands in contrast to the prevailing stylized fact in the literature that roughly 20 percent of manufacturing firms export.

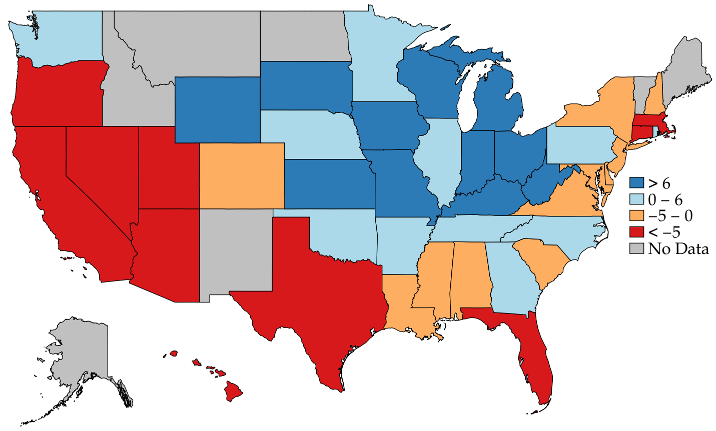
Bernard and Jensen (1995) were the first to document the fraction of exporting firms in the U.S. manufacturing sector: 14.6 percent of plants in the 1987 Census of Manufacturers.

[9]See Crozet and Koenig (2010) for an example with French data, though in their case they use firm-level data and are unable to allocate multi-unit firm exports across *régions* in France.

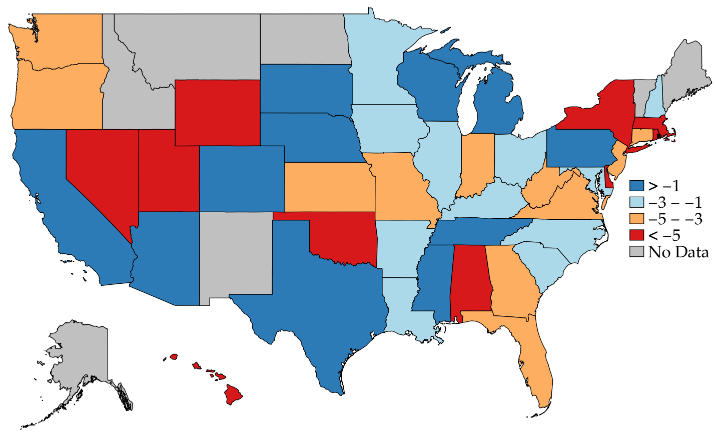
[10]See Appendix Figure A3 for additional maps corresponding to other export destinations.

Figure 1: State-Level Exports to Select Countries, Percent of Total

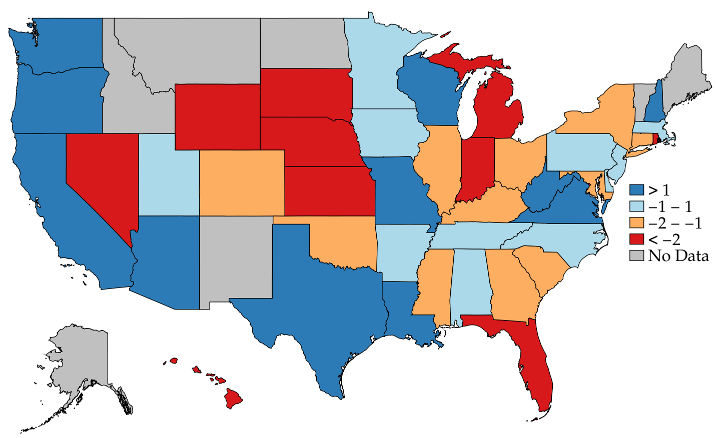
(a) State-Level Exports to Canada



(b) State-Level Exports to Mexico



(c) State-Level Exports to China



Sources: Author's calculations using LFTTD, CMF as explained in the text.
Notes: Each map reflects the differential state-share of exports to each destination, relative to the overall U.S.-level share of exports to that destination. States with grey coloring have data suppressed due to Census disclosure rules. The differential shading reflects the 25th, 50th, and 75th percentiles of each statistic, among those states with disclosed data.

Later on, Bernard et al. (2007*b*) and Bernard et al. (2007*a*) documented that 20 percent of

Table 2: Plant/Firm-Level Export Participation, by Source

	Establishment Level	Firm Level
	Percent reporting positive exports	
CMF (survey) definition	21%	19%
Allocation / LFTTD definition	41%	36%

Notes: The table reports the fraction of overall plants and firms identified as exporters based on the source of data. See Appendix C.1 for a full accounting of export participation discrepancies between these variables.

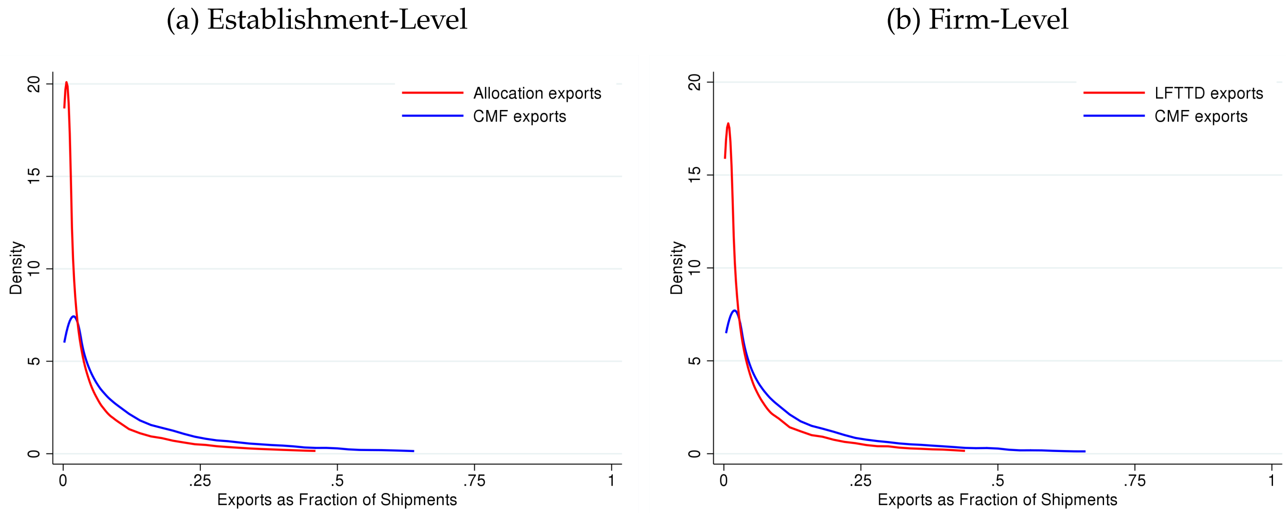
plants and 18 percent of firms export, based on the export variable in the 2002 Census of Manufacturers.[11] Using the full set of firms in the economy, the number of exporting firms is considerably lower, at only 4 percent.

To provide a comparison using our new data, we match establishment-level exports to the set of plants in the 2007 CMF. Using the universe of plants in that dataset, we calculate the fraction of plants identified as exporters using both the native survey-based export variable in the CMF and the LFTTD variable (derived from administrative data). It is useful to identify reasons why these export definitions may not align. For instance, exports identified by the plant could be processed via an intermediary such that the associated firm in the LFTTD would differ from the manufacturing plant reporting in the survey. This would lead to positive exports being identified in the survey but not in the LFTTD. On the other hand, a manufacturing plant may not be aware of the ultimate destination of its production, and thereby miss some value of production that is subsequently exported (perhaps by a different unit of the firm). This possibility may be more likely if the value of exports is small relative to overall production. In this case, exports would be reported by the parent firm of the plant in the LFTTD, and possibly allocated by our procedure to the plant itself, but no exports would be reported in the survey.

Table 2 reveals that our establishment-level exporting definition (derived primarily from the LFTTD) identifies a considerably higher fraction of exporting plants than that identified from the variable in the CMF survey. This is also true when identifying exporting status at the firm-level, an important point that indicates the distinction at the establishment level is

[11] It is important to note that Bernard et al. (2007*b*) show an export definition using the LFTTD transactions level trade data for firms in the 1997 Census of Manufacturers (see Table 7) which is considerably higher (27 percent) than the corresponding 2002 statistic based on the Census of Manufacturers alone. However, Bernard et al. (2007*b*) do not discuss the discrepancy based on the source of the export variable.

Figure 2: Export Share of Shipments: Distribution by Source of Exports



Sources: Author's calculations using LFTTD, CMF as explained in the text.
Notes: Per Census disclosure rules, the top and bottom 5 percent of each distribution is truncated. These distributions exclude establishments or firms that report zero exports. Since the density functions are calculated across different samples, the density estimates are relative within each measure: the figures do not indicate that the allocation-based measure includes fewer absolute establishments with a given value relative to the CMF-based measure.

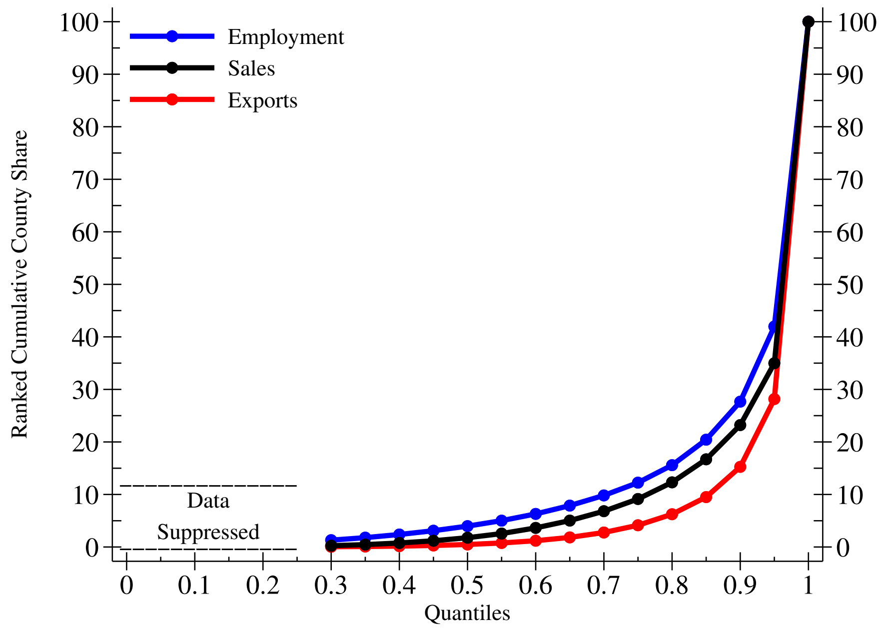
not driven by any methodological feature of our allocation procedure.

To explore this difference in export participation in greater detail, we calculate exports (from both sources) as a fraction of CMF-identified shipments. Figure 2 plots kernel density estimates of the distribution of this statistic for each export definition.[12] What is clear from the figure is that the distribution based on the LFTTD source of exports has many more instances of small export shares relative to the CMF-source, a feature consistent with the presence of under-measurement of exporting in the CMF for establishments with modest levels of export activity. Appendix C.1 provides further details on differences in export participation estimates by data source. Ultimately, this new, higher, estimate of export participation among manufacturing firms is important for researchers attempting to calibrate models featuring heterogeneous firms subject to fixed costs of export participation (see Lincoln and McCallum (2018) for a discussion of sunk costs of entry for the case of the U.S.).

Other aspects of this comparison serve to confirm the validity of the establishment-level export allocation. Figure 2 also shows that unrealistic levels of trade—exports above reported CMF shipments— are exceedingly rare (Census disclosure rules require estimates to suppress the top and bottom 5 percent of the distribution). Moreover, among establishments with positive exporting in both sources, the correlation of exports-per-shipment is high at 0.9 for firms, and 0.78 for establishments.

[12]The density function is estimated among establishments/firms with positive exports from each source, and hence do not capture those reporting zero exports.

Figure 3: County-level Concentration: Exports, Manufacturing Sales, Employment



Sources: Authors' calculations using LFTTD, CMF as explained in the text.
Notes: These cumulative shares result from ranking U.S. counties according to each variable and then calculating the county-level share relative to the total. Estimates below the third decile are suppressed due to Census disclosure rules.

3.4 Exports are Highly Concentrated Across Space

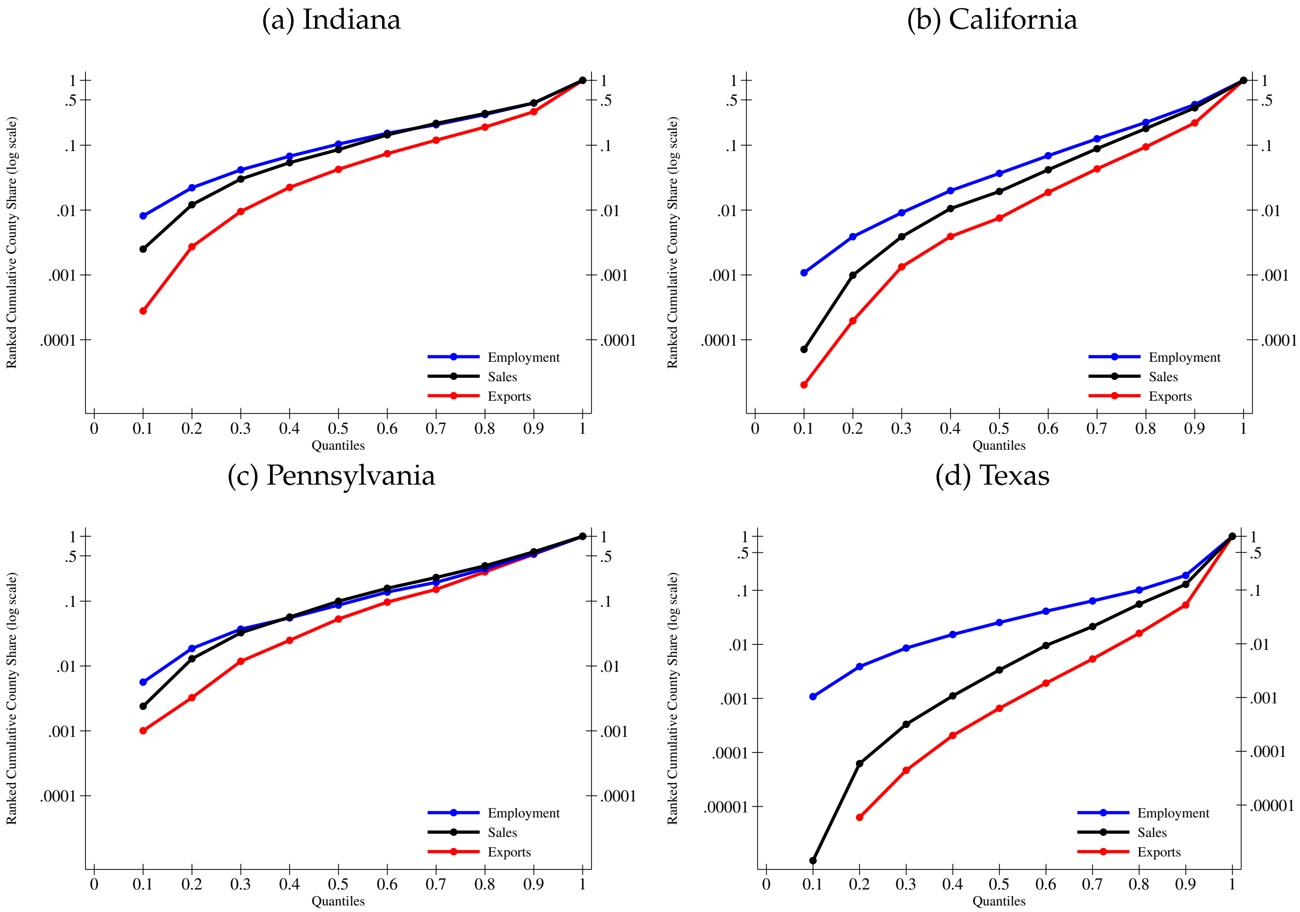
Finally, the new data described in this paper reveal aspects of exporting at a local level that were previously impossible to study with any confidence. A key new fact revealed by this perspective is the remarkable concentration of export activity in the United States. While it is well-known that economic activity is highly concentrated across space (see Axtell (2001), Gabaix (1999), Gabaix (2011)) and that export activity is concentrated in a small number of firms (see, for example, Bernard et al., 2007*c*), the geographic concentration of exporting has received less attention.

As one way to highlight this concentration, we aggregate our data to the roughly three thousand counties in the United States, rank counties by their share in aggregate exports, and then calculate the cumulative share by quantile. For comparison, we conduct a similar exercise for both employment and manufacturing sales, noting that an individual county's rank generally differs for each measure. The resulting Figure 3 is similar to a Lorenz curve, but has the cumulative *share* of exports, sales, and employment on the vertical axis.

Figure 3 not only confirms the extreme skewness of overall economic activity at this level of geography, but also highlights that export activity is even more concentrated than either employment or sales.[13] While the top 5 percent of counties by employment account for a

[13]The shares of counties below the third decile are so low that disclosure rules prevent us from quantifying

Figure 4: County-level Concentration by State: Exports, Manufacturing Sales, Employment



Sources: Author's calculations using LFTTD, CMF as explained in the text.
Notes: The first exports decile for the state of Texas is suppressed to improve legibility.

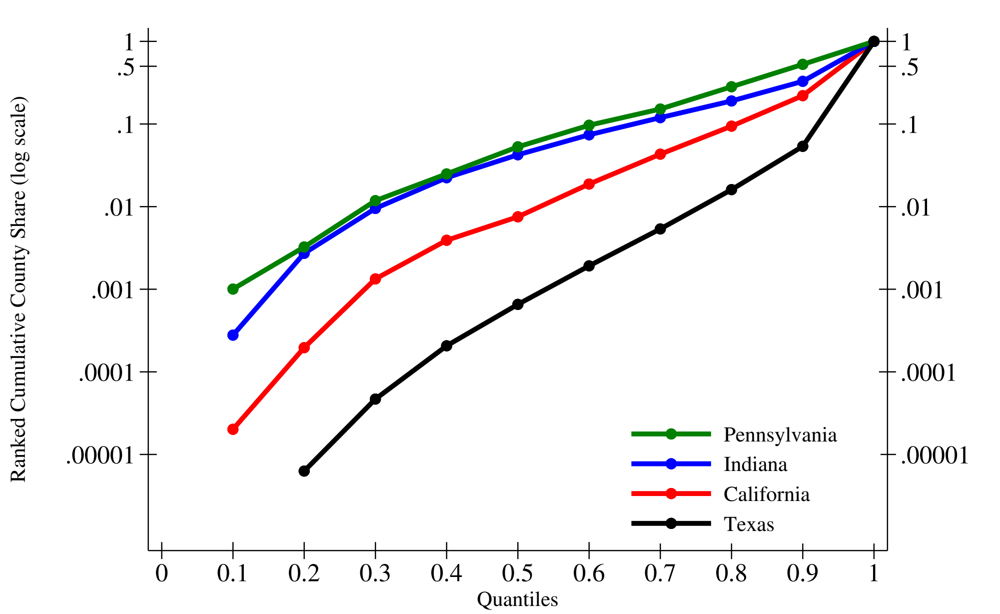
little over 55 percent of employment, the top 5 percent of counties by export volume account for over 70 percent of total exports.

A second and related fact revealed by this data is that much of variation in exporting actually occurs *within* states. In Figure 4 we conduct a similar exercise to Figure 3 but within states rather than across the entire country. Although disclosure limitations prevent a full presentation of all states, we report results for four large states covering different geographic areas of the United States. Given the smaller number of overall counties for a given state, we report these ranked shares by decile, and improve legibility by using a log scale.

Figure 4 reveals several notable features. The greater concentration of exports relative to employment or manufacturing sales holds for all states shown, though the degree of this gap differs by state, with the proportions tracking more closely in Pennsylvania than Texas. Second, the degree of export concentration also differs markedly across states. This point is

their shares.

Figure 5: County-level Export Concentration: State Comparison



Sources: Author's calculations using LFTTD, CMF as explained in the text.
Notes: The first decile for the state of Texas is suppressed to improve legibility.

more clearly visible in Figure 5, which plots just the export concentration across states. For yet another perspective, Table 3 focuses on the variation in the top decile for our selected states with as little as 42 percent of activity in the top decile (sales in Pennsylvania), to as much as 95 percent (exports in Texas). In each case, exports are more concentrated than either manufacturing sales or employment.

Table 3: Proportion in Top County Decile: Selected States

	Employment	Sales	Exports
Pennsylvania	46%	42%	47%
Indiana	55%	55%	67%
California	58%	62%	78%
Texas	81%	87%	95%

Notes: The table reports the proportion of each measure (employment, manufacturing sales, exports) occupied by the top decile of counties in each state, where the deciles are calculated separately for each measure.

The stylized facts documented in this section illustrate the high degree of variation in exports across counties across the United States, and within individual states. This variation does not always align with other indicators of economic activity, with exports typically more highly concentrated than either employment or sales. In the context of evaluating the local labor market effects of trade, this feature is particularly salient in light of the common practice of using county-level employment shares to impute the effects of export activity to local areas. As our new dataset shows, such a practice could lead to inaccurate interpretations of a local area's export exposure.

4. Local-Area Effects of Exports: Evidence from the 2008-2009 Great Trade Collapse

During the Great Recession U.S. real imports and exports both declined by around 20 percent from peak to trough—a much larger drop than overall economic activity. While a sizeable literature has analyzed the factors contributing to this "Great Trade Collapse" (e.g., Levchenko, Lewis and Tesar, 2010; Alessandria, Kaboski and Midrigan, 2010*b*), data limitations have thus far prevented researchers from studying the effects on U.S. local labor market outcomes. In this section we illustrate the advantages of our new dataset and study how the Great Trade Collapse affected county-level employment, payroll, and wages.

To isolate exogenous variation in foreign demand we employ a version of the World Import Demand (WID) instrument as used, for instance, in Hummels et al. (2014). However, doing so poses the challenge that the Great Recession also adversely affected demand in the U.S. and hence declines in foreign demand of a counties' products may be correlated with declines in domestic demand. We therefore adapt the WID instrument to only use variation in destination-specific declines in the destination's imports of a counties' product—and not variation that was common across destinations. This correction implies that our version of the WID instrument is plausibly uncorrelated with changes in domestic demand.

We use this variation to estimate to what extent counties more exposed to foreign demand declines experienced greater employment, payroll, and wage declines between 2007-2009. Our main results are as follows. A one million dollar drop in exports, induced by a drop in foreign demand, led, on average, to the loss of 17 jobs. Further, the pass-through from exports to payroll was approximately one-for-one so that a one dollar drop in exports reduced payroll in the county by also one dollar. Lastly, a one thousand dollar drop of exports per worker reduced wages, on average, by 1.9 percent.

We perform our analysis at the county level as opposed to alternative levels of aggregation (e.g., an establishment or state-level analysis). The main reason is that a key object of interest to policymakers is the net number of jobs gained or lost in a given labor market. To the extent that a county roughly aligns with the local labor market, a county-level analysis is informative about this object.[14] In contrast, an establishment-level analysis would measure a gross flow, that is, the number of jobs lost in exporting establishments without taking into account that workers might find jobs elsewhere in the local labor market. While a state-level analysis would also measure the net change in jobs, it would provide very limited variation. A second reason is that a county-level analysis aligns our application with a common level of aggregation used in previous work that studies the effects of shocks on local labor markets (e.g., Serrato and Wingender, 2016).

We first present our main estimating equation in Section 4.1. Threats to identification and the instrument are discussed in Section 4.2 and we present the results in Section 4.3. Lastly, we consider extensions in Section 4.4 and present robustness exercises in Section 4.4.2.

4.1 Regression Specification

Let c index counties and t time. We estimate specifications of the form

$$\frac{y_{c,2009} - y_{c,2007}}{\text{emp}_{c,2007}} = \alpha + \beta \frac{\text{exp}_{c,2009} - \text{exp}_{c,2007}}{\text{emp}_{c,2007}} + \text{controls} + \varepsilon_c, \tag{4.1}$$

where $y_{c,t}$ is an outcome of interest, $\text{exp}_{c,t}$ denotes the value of exports, and $\text{emp}_{c,t}$ employment. All variables are aggregated up from establishment-level data in the county. To account for the fact that counties differ in size, we divide both the change in the outcome of interest and the change in exports by the counties' employment in 2007. Notice that this division only affects the interpretation of the constant α; it does not affect the interpretation of the coefficient of interest β.

Our primary outcome variables $y_{c,t}$ are employment and payroll. Employment is measured in numbers of workers, and payroll and exports are measured in millions of U.S. dollars. Hence, when the outcome variable of interest is employment, β captures the number of jobs lost due to a drop in exports of one million U.S. dollars.[15] When the outcome variable of interest is payroll, the coefficient β measures the dollars worth of payroll lost after a drop in exports of one dollar. We also estimate equation (4.1) after replacing the left-hand side with the relative change in the average county-level wage $\frac{\text{wage}_{c,2009} - \text{wage}_{c,2007}}{\text{wage}_{c,2007}}$, expressed in

[14]Another commonly used definition of a local labor market is a Commuting Zone (CZ) (e.g. (Autor, Dorn and Hanson, 2013) among others). We do not employ CZs in our analysis in light of the extreme skewness of export activity (which does not align with employment). These patterns in the data make a comprehensive geographic unit of analysis such as counties preferable to anything else (such as CZs or CBSAs) that are not defined for some areas of the US.

[15]We also report estimates for the case in which exports are measured in constant 2007 dollars. Specifically, we deflate export transaction with the export price indexes from the BLS. Since this deflation has essentially no effect on the estimates, we report the results in the Appendix.

percent. In this case we measure exports in thousands of U.S. dollars and hence β measures the percent change in the wage rate attributable to one thousand dollar change in exports per worker.

Specification (4.1) contains a constant and is estimated at the county level. This implies that any aggregate variation over the period from 2007 to 2009 is purged and we estimate the coefficient of interest β entirely from cross-county variation. Note that this coefficient does capture local (general) equilibrium effects such as wage adjustments which may follow a change in foreign demand. As discussed earlier, county-level employment changes following the trade shock reflect *net* changes in the sense that they reflect laid off workers who do not subsequently find jobs elsewhere in the county.

When interpreting the coefficient β, it is important to note that equation (4.1) subsumes changes in domestic demand into the error term. Our objective is to estimate the labor market impact of a decline in foreign import demand on a counties' labor market outcomes relative to a county which did not experience this decline in foreign demand. Since domestic demand is not held constant in equation (4.1), the coefficient β measures the effect of a decline in foreign demand, which is not offset (or enhanced) by selling more (or less) domestically.

While we expect that the sign of β is positive so that a decline in foreign demand adversely affects employment, payroll, and wages, we do not have a strong prior regarding its magnitude. In the presence of only the direct effect—that exporting firms lay off workers following the drop in foreign demand—β captures the inverse of these workers' labor productivity. However, this effect may either be attenuated or enhanced. Some work suggests, for instance, that selling domestically and abroad are substitutes, so firms that were adversely affected by the collapse in trade may have been able to compensate for part of this decline by selling more domestically.[16] Further, some laid-off workers may subsequently find work elsewhere in the county, which could partially offset the direct effect at the plant level. Both of these mechanisms would suggest that β is relatively small. Alternatively, any initial layoffs may be aggravated due to local multiplier effects. Such a mechanism would raise the size of β. Taken together, the magnitude of β is influenced by multiple different mechanisms and their relative importance is ex-ante not clear.

We control throughout for the 2007 ratio of exports to employment $\frac{\text{exp}_{c,2007}}{\text{emp}_{c,2007}}$. This control aims to address the possibility that counties which differ in their exposure to foreign demand may differ in other dimensions that could be correlated with developments in their labor market between 2007 and 2009. To address the concern that the outcome variable of interest is trending over time, we further present specifications in which we control for the lagged

[16] See, e.g., Vannoorenberghe (2012) for a discussion of this mechanism and evidence at the firm-level.

change in the dependent variable $\frac{y_{c,2007}-y_{c,2005}}{\text{emp}_{c,2005}}$. Lastly, we also present specifications with state fixed effects. In this case the coefficient β is identified entirely from within state variation.

4.2 Identification

The objective of the empirical analysis in this section is to identify the causal effect of a decline in exports—induced by a drop in foreign demand—on counties' local labor markets. This effect is measured relative to counties which did not experience a decline in exports. Of course, the change in exports on the right-hand side of equation (4.1) and the outcome variable of interest are both endogenously determined and potentially affected by many different shocks. For instance, an adverse local supply shock may reduce both exports and employment and thus bias the estimate of β upwards.

In order for β to capture a causal effect we employ an instrumental variable strategy. In particular, we use a version of the World Import Demand (WID) instrument (Hummels et al., 2014) to isolate variation in foreign demand that is plausibly orthogonal to other shocks that affect county-level exports and employment over this period. Our version of the WID instrument is

$$\text{inst}_c^{\text{WID}} = \frac{\text{exp}_{c,2007}}{\text{emp}_{c,2007}} \cdot \sum_{d,p} s_{c,d,p,2007}^{\text{exp}} \left(\Delta \ln \text{imp}_{d,p} - \overline{\Delta \ln \text{imp}_p} \right). \tag{4.2}$$

In this expression $s_{c,d,p,2007}^{\text{exp}} = \frac{\text{exp}_{c,d,p,2007}}{\text{exp}_{c,2007}}$ is county c's export share of product p to destination d and $\Delta \ln \text{imp}_{d,p}$ denotes country d's change in total imports of product p between years 2007 and 2009. Further, $\overline{\Delta \ln \text{imp}_p} = \frac{1}{D} \sum_d \Delta \ln \text{imp}_{d,p}$ denotes the product-specific average of the change in foreign imports, where the average is taken across all possible destinations $d = 1, \dots, D$.

At its core, the WID instrument is a weighted sum of changes in foreign imports. It varies by county because different counties are differentially exposed to changes in foreign imports—as captured by the predetermined shares $s_{c,d,p,2007}^{\text{exp}}$. The fraction on the right-hand side of equation (4.2), $\frac{\text{exp}_{c,2007}}{\text{emp}_{c,2007}}$, simply scales the weighted change in foreign imports to account for the fact that some counties export more than others.

A key difference relative to alternative implementations of the WID instrument is that we subtract the product-specific average in imports ($\overline{\Delta \ln \text{imp}_p}$) from the destination and product-specific change in imports ($\Delta \ln \text{imp}_{d,p}$). This adjustment applies and extends the idea of Boehm and Pandalai-Nayar (2020) to purify the WID instrument of undesired components, and addresses the following concern. Since the Great Recession was essentially global in nature, and since downturns affect the demand for products differentially, a version of the WID instrument without this adjustment could be correlated with domestic declines in demand. It is well-known, for instance, that the demand for durable goods declines more during downturns than the demand for nondurable goods. It is also known that a disproportionate share of traded goods is durable. Hence, without purging the instrument of the average decline in demand for products, it likely that the instrument would be correlated with changes in the U.S. demand for a counties' products. Our adjustment implies that the instrument as constructed in equation (4.2) uses destination-specific and destination-product-specific variation, but not purely product-specific variation in foreign demand. Intuitively, a counties' demand from abroad was disproportionately affected by the collapse in trade if it exported to destinations whose import demand for its products declined by more than the average global demand for its products during the Great Trade Collapse.

Our research design connects to a recent literature formalizing the properties of shift-share instruments and establishing sufficient conditions for consistent estimation. Our identification argument particularly relates to Borusyak, Hull and Jaravel (Forthcoming) and—adopting their terminology—relies on the quasi-random assignment of "shocks", which in our setting are $\Delta \ln \text{imp}_{d,p} - \overline{\Delta \ln \text{imp}}_p$. An important feature of our analysis is that these shocks vary by product *and* destination. A sufficient condition for identification is that these shocks are uncorrelated with other shocks that drive appropriately constructed averages of counties' employment or payroll changes that vary at the destination and product level. Subtracting the product-specific average of the growth rate of imports from the shock is analogous to a "shock-level control" and ensures that shock-level confounders that vary purely at the product level—such as a decline of domestic demand for a product—do not pose a problem for identification. Further, our setup features "incomplete shares" since our exposure weights for a given county c, $\frac{\text{exp}_{c,2007}}{\text{emp}_{c,2007}} s^{\text{exp}}_{c,d,p,2007} = \frac{\text{exp}_{c,d,p,2007}}{\text{emp}_{c,2007}}$, do not sum to unity, but to exports per worker $\frac{\text{exp}_{c,2007}}{\text{emp}_{c,2007}}$. Since it is undesirable to use this variation for identification, we include exports per worker as a control in our preferred specifications—as discussed above. In additional checks suggested by Borusyak, Hull and Jaravel (Forthcoming), we establish that the effective sample size of shocks, computed as the inverse Herfindahl index of normalized exposure shares, is large (greater than 180) and that the largest exposure share to shocks is smaller than 0.06.[17] Since the structure of the shift-share instrument implies that conventional formulas for computing standard errors do often not apply, we follow the recommendation of Borusyak, Hull and Jaravel (Forthcoming) and report standard errors

[17] Both of these conditions support the asymptotic thought experiment required for consistent estimation—an appropriate law of large numbers at the shock level.

which are clustered by product and destination in all results below. We also report the shock-level F-statistics for the first stages for each regression. These are typically lower than conventional first stages, and so provide a conservative assessment of the strength of our instrument.

We use annual imports from the United Nations Comtrade database to construct the WID instrument. A product in equation (4.2) is measured at the Harmonized System (HS) 3-digit level. The change in imports, $\Delta \ln \text{imp}_{d,p}$, is constructed as the relative change in imports from all partner countries except for the U.S.[18]

4.3 Baseline estimates

Panel A of Table 4 presents the effects on employment. We begin with specification (1), which reports estimates with only the exports-per-worker control. The estimated coefficient on the change in exports is around 17, implying that a one million dollar decline in exports to foreign countries led to the loss of 17 jobs. This estimate is significant at the one percent level. Little changes when we include the lagged change in employment to account for pre-trends (specification (2)). The coefficient estimate on the change in exports remains stable near 17 and highly significant.

In specification (3) we additionally add state fixed effects and therefore identify the coefficient of interest only from within-state variation across counties. The estimate remains near 17. We note that the first-stage F statistics exceed the conventionally applied threshold of 10 by a large margin for all specifications reported in Panel A of Table 4. This suggests that the instrument is uniformly strong.

We next turn to the effects on payroll, which are reported in Panel B of Table 4. Specification (1) reports the estimate with only the exports-per-worker control. This estimate is 1.04. In specification (2) we add the lagged change in payroll. The coefficient of interest remains stable. A coefficient of 1.04 implies that a one dollar decline in exports reduces payroll by 1.04 dollars or, put differently, that the decline in exports is passed through to payroll approximately one-for-one. Again the estimate remains stable when we include state fixed effects, as reported in specification (3).

Table 5 reports the effects on average pay per worker—to which we henceforth refer to as wages. More specifically, the dependent variable is the relative change in wages from 2007 to 2009, expressed in percent. Further, for ease of interpretation, exports are now measured in thousands of U.S. dollars. The estimate of interest in specification (1) with only the exports-

[18]We have alternatively constructed the instrument at the HS 4-digit level in undisclosed results.

Table 4: Baseline estimates: Employment and Payroll

Panel A: Employment

Dependent variable: $\frac{\text{emp}_{c,2009}-\text{emp}_{c,2007}}{\text{emp}_{c,2007}}$

	(1)	(2)	(3)
$\frac{\text{exp}_{c,2009}-\text{exp}_{c,2007}}{\text{emp}_{c,2007}}$	16.9***	16.4***	17.7***
	(4.78)	(4.65)	(6.51)
	[4.89]	[4.77]	[6.47]
$\text{exp}_{c,2007}/\text{emp}_{c,2007}$	0.00***	0.00***	0.00***
	(0.00)	(0.00)	(0.00)
$\frac{\text{emp}_{c,2007}-\text{emp}_{c,2005}}{\text{emp}_{c,2005}}$		0.01	-0.01
		(0.01)	(0.01)
State fixed effects	no	no	yes
First stage F	37.57	37.42	31.27
Shock-level first stage F	25.29	25.18	21.84

Panel B: Payroll

Dependent variable: $\frac{\text{pay}_{c,2009}-\text{pay}_{c,2007}}{\text{emp}_{c,2007}}$

	(1)	(2)	(3)
$\frac{\text{exp}_{c,2009}-\text{exp}_{c,2007}}{\text{emp}_{c,2007}}$	1.04***	0.96***	0.96***
	(0.35)	(0.34)	(0.44)
	[0.35]	[0.34]	[0.45]
$\text{exp}_{c,2007}/\text{emp}_{c,2007}$	0.06	0.05	0.07
	(0.05)	(0.04)	(0.06)
$\frac{\text{pay}_{c,2007}-\text{pay}_{c,2005}}{\text{emp}_{c,2005}}$		0.04***	0.02
		(0.02)	(0.02)
State fixed effects	no	no	yes
First stage F	37.57	36.89	31.28
Shock-level first stage F	25.29	24.62	21.81

Notes: The table displays two-stage least squares coefficient estimates based on equation (4.1), where we use the WID instrument (4.2) to address the endogeneity of exports. Exports and payroll are measured in millions of dollars and employment in numbers of workers. The number of observations—rounded to the closest 100 to comply with Census disclosure requirements—is 3000 in all specifications. We report two types of standard errors. First, we report standard errors clustered by state in parentheses. Second, we report in square brackets standard errors from the "shock-level" specification and clustered by product and destination using the approach by Borusyak, Hull and Jaravel (Forthcoming). ***, **, and * indicate significance at the 1, 5, and 10 percent level. Significance is based on the "shock-level" standard errors where available.

Table 5: Baseline estimates: Wage

Dependent variable: $\frac{\text{wage}_{c,2009}-\text{wage}_{c,2007}}{\text{wage}_{c,2007}}$ (percent)			
	(1)	(2)	(3)
$\frac{\text{exp}_{c,2009}-\text{exp}_{c,2007}}{\text{emp}_{c,2007}}$	1.85***	1.92***	1.56**
	(0.87)	(0.77)	(0.75)
	[0.86]	[0.77]	[0.75]
$\text{exp}_{c,2007}/\text{emp}_{c,2007}$	0.02	0.02	0.00
	(0.12)	(0.10)	(0.10)
$\frac{\text{wage}_{c,2007}-\text{wage}_{c,2005}}{\text{wage}_{c,2005}}$		-0.21***	-0.22***
		(0.02)	(0.02)
State fixed effects	no	no	yes
First stage F	37.57	37.37	31.29
Shock-level first stage F	25.29	25.06	21.82

Notes: The table displays two-stage least squares coefficient estimates based on equation (4.1), where we use the WID instrument (4.2) to address the endogeneity of exports. The dependent variable is the relative change in the wage from 2007 to 2009, expressed in percent. Exports are measured in thousands of dollars and employment in numbers of workers. The number of observations—rounded to the closest 100 to comply with Census disclosure requirements—is 3000 in all specifications. We report two types of standard errors. First, we report standard errors clustered by state in parentheses. Second, we report in square brackets standard errors from the "shock-level" specification and clustered by product and destination using the approach by Borusyak, Hull and Jaravel (Forthcoming). ***, **, and * indicate significance at the 1, 5, and 10 percent level. Significance is based on the "shock-level" standard errors where available.

per-worker control is 1.85, implying that a decline in exports per worker by 1000 U.S. dollars reduces the county-level wage by 1.85 percent. This estimate is roughly stable when we add the lagged change in wages as a control. When we add state fixed effects in specification (3) the estimate falls slightly.

OLS estimates. For comparison Table 6 shows OLS estimates. For all three dependent variables, these estimates are positive. While they typically remain statistically significant, the estimates tend to be smaller than our 2SLS estimates. The reduced magnitude could potentially reflect classical measurement error in our independent variable. It is also consistent with any omitted variable which affected exports and the dependent variable with opposite signs during the Great Recession.

4.4 Extensions and robustness

We next consider an extensions in which we explore how the foreign demand shock during the Great Recession affected the manufacturing and the non-manufacturing sector separately.

Table 6: OLS estimates: Employment, Payroll and Wages

Dependent variable:	$\frac{\text{emp}_{c,2009}-\text{emp}_{c,2007}}{\text{emp}_{c,2007}}$		$\frac{\text{pay}_{c,2009}-\text{pay}_{c,2007}}{\text{emp}_{c,2007}}$		$\frac{\text{wage}_{c,2009}-\text{wage}_{c,2007}}{\text{wage}_{c,2007}}$	
	(1)	(2)	(3)	(4)	(5)	(6)
$\frac{\text{exp}_{c,2009}-\text{exp}_{c,2007}}{\text{emp}_{c,2007}}$	4.38***	3.80***	0.18***	0.15***	0.15**	0.08
	(0.68)	(0.71)	(0.03)	(0.03)	(0.09)	(0.07)
$\text{exp}_{c,2007}/\text{emp}_{c,2007}$	0.00	0.00***	-0.04***	-0.03*	-0.20***	-0.18***
	(0.00)	(0.00)	(0.02)	(0.02)	(0.04)	(0.04)
Lagged dep. variable	0.01	-0.01	0.05***	0.02	-0.20***	-0.21***
	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
State fixed effects	no	yes	no	yes	no	yes
R^2	0.01	0.13	0.05	0.19	0.09	0.17

Notes: The table displays ordinary least squares coefficient estimates based on equation (4.1). In specifications (1)-(4) exports and payroll are measured in millions of dollars and employment in numbers of workers. In specifications (5) and (6) exports are expressed in thousands of dollars. The change in wages between 2007 and 2009 is expressed in percent. The number of observations—rounded to the closest 100 to comply with Census disclosure requirements—is 3000 in all specifications. Standard errors are clustered by state. ***, **, and * indicate significance at the 1, 5, and 10 percent level.

We subsequently discuss additional robustness checks.

4.4.1 Manufacturing versus non-manufacturing

We begin with studying the degree to which the foreign demand shock affected manufacturing sectors and non-manufacturing sectors differentially. Since the manufacturing sector produces the large majority of traded goods, 78 percent in 2007 (see Census, 2008, p.1 of FT-900 Supplement), one would expect that a decline in foreign demand affects this sector disproportionately. On the other hand, lost income of manufacturing workers may trigger declines in the demand for locally produced goods and services, thereby leading to contractions in the non-manufacturing sectors through local equilibrium effects. This section aims to estimate the magnitudes of these effects.

To do so we consider the same outcome variables as before—i.e. employment, payroll, and wages—but construct them separately for manufacturing and non-manufacturing industries within each county. We identify manufacturing plants based on their 2-digit NAICS classifications (31, 32, and 33) and refer to aggregates over the manufacturing industries as the manufacturing *sector* and to aggregates of non-manufacturing industries as the non-manufacturing sector.

Table 7 presents the results. We report coefficient estimates of our preferred specifications, which include the the 2007 ratio of exports per worker as well as a lagged dependent variable (from 2005 to 2007) as controls, but do not include state fixed effects.[19] Specifications (1) and (2) show that employment was impacted by similar magnitudes in the manufacturing and non-manufacturing sectors. More precisely, the estimates suggest that a decline of exports of one million U.S. dollars reduced manufacturing employment by 9.72 jobs while it reduced non-manufacturing employment by 7.95 jobs. Both estimates are significantly different from zero at conventional levels, but the effect for the non-manufacturing sector is estimated with somewhat lower precision. While the large magnitude of the effect on employment for non-manufacturing may seem at first surprising, recall that the level of manufacturing employment is considerably lower than non-manufacturing employment in most counties.

Specifications (3) and (4) show the effects on payroll. The estimates suggest that for every dollar decline in exports payroll fell by 52 cents in manufacturing sectors and by 87 cents in non-manufacturing sectors. While the latter effect is very large, it is imprecisely estimated. Specifications (5) and (6) show—as expected—that the impact on wages is greater in the manufacturing sector.

4.4.2 Robustness

Our findings in the previous section illustrated that counties differ substantially in their export behavior and hence in their exposure to declines in foreign demand. In particular, we showed that over 70 percent of exports are accounted for by just five percent of counties. While this differential exposure implies that the trade collapse during the Great Recession provides a useful source of variation for our exploration, it also raises the concern that our results could be driven by a relatively small number of counties. To ensure that our estimates are not driven by extreme observations we winsorize all variables used in our empirical analysis.

Employment weights. We have re-estimated the effect of a drop in foreign demand on employment, payroll, and wages, weighing counties with their employment share. In this case, the coefficient estimates, in particular for employment, shrink and become insignificant (estimates not reported). This exercise may suggest that our main estimates are driven by a set of relatively small counties with large changes in exports per worker. We note, however, that it is unclear whether to use employment weights in the estimation. As discussed in Solon, Haider and Wooldridge (2015), in the presence of heterogeneous effects, neither weighted nor unweighted estimation generally ensure consistent estimation of the average partial effect

[19]Results are robust to including state fixed effects and available on request.

Table 7: Effects on Manufacturing and non-Manufacturing Sectors

Dependent variable	Employment		Payroll		Wages	
	Mfg.	Non-Mfg.	Mfg.	Non-Mfg.	Mfg.	Non-Mfg.
	(1)	(2)	(3)	(4)	(5)	(6)
$\frac{\text{exp}_{c,2009}-\text{exp}_{c,2007}}{\text{emp}_{c,2007}}$	9.72***	7.95***	0.52***	0.87	2.08***	0.90
	(2.76)	(4.49)	(0.12)	(0.60)	(0.79)	(0.88)
	[3.90]	[3.70]	[0.19]	[0.62]	[0.86]	[0.92]
$\text{exp}_{c,2007}/\text{emp}_{c,2007}$	0.00	0.00	0.02	0.09	-0.04	-0.00
	(0.00)	(0.00)	(0.02)	(0.08)	(0.09)	(0.11)
Lagged dep. variable	-0.01	-0.01	-0.09***	0.01	-0.19***	-0.21***
	(0.01)	(0.01)	(0.02)	(0.01)	(0.02)	(0.04)
State fixed effects	no	no	no	no	no	no
Shock-level first stage F	25.02	25.26	24.65	25.12	25.27	24.89

Notes: The table displays two-stage least squares coefficient estimates based on equation (4.1), where we use the WID instrument (4.2) to address the endogeneity of exports. Abbreviations: Mfg.—Manufacturing; Non-Mfg.—Non-manufacturing. In specifications (1) and (2) the dependent variable is $(\text{emp}^s_{c,2009} - \text{emp}^s_{c,2007})/\text{emp}_{c,2007}$, where $s \in \{\text{Mfg.}, \text{Non-Mfg.}\}$ indexes the sector. In specifications (3) and (4) the dependent variable is $(\text{pay}^s_{c,2009} - \text{pay}^s_{c,2007})/\text{emp}_{c,2007}$. In specifications (1) through (4) exports and payroll are measured in millions of dollars and employment in numbers of workers. In specifications (5) and (6) the dependent variable is $(\text{wage}^s_{c,2009} - \text{wage}^s_{c,2007})/\text{wage}^s_{c,2007}$, expressed in percent, and exports are measured in thousands of dollars. *Lagged dep. variable* refers to the dependent variable from 2005 to 2007. The number of observations—rounded to the closest 100 to comply with Census disclosure requirements—is 3000 in all specifications. We report two types of standard errors. First, we report standard errors clustered by state in parentheses. Second, we report in square brackets standard errors from the "shock-level" specification and clustered by product and destination using the approach by Borusyak, Hull and Jaravel (Forthcoming). ***, **, and * indicate significance at the 1, 5, and 10 percent level. Significance is based on the "shock-level" standard errors where available.

in the population. Instead, they recommend directly exploring the source of heterogeneity. This is what we aim to do in future iterations of this paper.

CBP employment. As additional robustness, to ensure our employment results are not driven by changes in the vintage of the LBD or other Census products, we also estimate the effect of a foreign demand shock on county-level employment using the publicly available County Business Patterns (CBP) dataset to measure employment. The estimated declines in employment per million dollars of foreign export declines are greater and range from 24 workers in the specification with only the exports-per-worker control to 26.9 workers in the specification with state fixed effects. This helps assuage any concerns about the cleaning process we implement on the establishment-level employment data while aggregating to county-level within the Census.

5. Estimates using Publicly Available Data

In this section we illustrate how an alternative approach for identifying the local employment effect of exporting during the Great Trade Collapse is likely problematic. We do so by replicating the analysis in the previous section, but instead of using our new data on county-level exports, we impute these series from state-level exports and county-level employment shares. This imputation follows a common approach in the literature and pairs more aggregate trade data with county-level employment shares (e.g., Hakobyan and McLaren, 2016; Feenstra, Ma and Xu, 2019). All else, most notably the sample, is kept the same.

5.1 Imputation of County-Level Exports

To use the maximum amount of information available and therefore give this approach the best possible chance to provide accurate estimates, we use state-level export data from the U.S. Census Bureau for the imputation. This data is the most geographically disaggregated data with comprehensive coverage that is publicly available and provides exports from state s of product p to destination d at time t.[20] The imputation further uses the share of county c employment in industry p of overall state-level employment in industry p—again measured at time t.[21] We use County Business Patterns data (also from the U.S. Census Bureau) with adjustments for missing values as detailed in Eckert et al. (2020). Formally, the imputed value of county-level exports is

$$\widehat{\exp}_{c,d,p,t} = \frac{\text{emp}^{\text{CBP}}_{c,p,t}}{\sum_{c' \in s} \text{emp}^{\text{CBP}}_{c',p,t}} \cdot \exp_{s(c),d,p,t}, \tag{5.1}$$

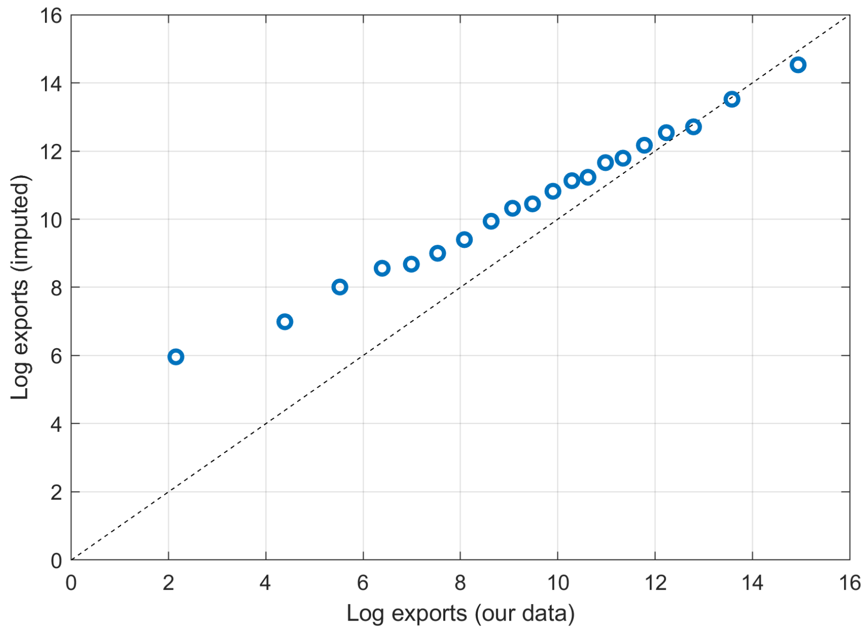
where $s(c)$ denotes the state s of county c.

Figure 6 shows a a binned scatterplot of the quantiles of county-level export, aggregated over products and destinations, between the imputed data and our new dataset on county-level exports. While the figure demonstrates a strong positive correlation, it also highlights the troubling feature that imputed exports systematically exceed measured exports for all but the largest counties. Note that this imputation error is consistent with the high concentration of exports documented above.

[20] We access this dataset via the USATrade Online database. As described in detail above, the state-level attribution relies on the "origin of movement" identified in Customs records and therefore may have some error relative to true source of export activity.

[21] Note that we are defining product and industry to be identical in this application, which follows from the classification of Census products according to a NAICS classification. At the level of available detail—3-digit NAICS—the industry and product-level classifications coincide.

Figure 6: Binned Scatterplot of Imputed vs allocated County-Level (log) Exports



Sources: Author's calculations using LFTTD, CMF as explained in the text. Counties are aggregated into 20 quantiles. Exports are measured in thousands of U.S. dollars.

5.2 Estimates using Publicly Available Data

We next repeat the analysis from the previous section, replacing measured exports with imputed exports, but keeping all else equal. Imputed exports enter the estimation of equation (4.1) in three distinct places. First, they enter the change in exports on the right-hand side of the estimating equation. Second, they enter into the exports per worker control. Third, they enter into the construction of the instrument: the exposure weights as well as the scaling of exports per worker are now constructed from the imputed data.

Table 8 collects the results of this exercise. In all columns the change in exports on the right hand side and the exports-per-worker control are constructed using imputed exports. Further, columns (1)-(3) use our version of the WID instrument, constructed as described in section 4.2, but where county-level exposure shares are constructed using imputed county-level exports. Hence, these columns reflect how a researcher might proceed without access to our new dataset (hence the label "public"). The estimates tend to be large and imprecise. Most importantly, the first stage is so weak that this research design does not appear viable and the estimates are not informative.

In columns (4)-(6), we continue to use imputed exports to construct all variables—except the instrument. While this research design is not feasible without access to our data, this

Table 8: Estimates using Public Data

	Dependent variable: $\frac{\text{emp}_{c,2009}-\text{emp}_{c,2007}}{\text{emp}_{c,2007}}$					
	(1)	(2)	(3)	(4)	(5)	(6)
$\frac{\widehat{\text{exp}}_{c,2009}-\widehat{\text{exp}}_{c,2007}}{\text{emp}_{c,2007}}$	21.9	22.4	29.9	39.9***	40.9***	76.4**
	(27.7)	(28.6)	(28.8)	(12.7)	(13.0)	(43.1)
	[21.4]	[22.1]	[27.7]	[9.63]	[10.2]	[36.9]
$\widehat{\text{exp}}_{c,2007}/\text{emp}_{c,2007}$	0.00	0.00	0.00	0.00***	0.00***	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$\frac{\text{emp}_{c,2007}-\text{emp}_{c,2005}}{\text{emp}_{c,2005}}$		-0.03	-0.03		-0.04	-0.02
		(0.03)	(0.02)		(0.03)	(0.03)
State fixed effects	no	no	yes	no	no	yes
Instrument	Public	Public	Public	Baseline	Baseline	Baseline
Shock-Level First stage F	0.11	0.11	0.02	8.33	7.68	3.23

Notes: The table displays two-stage least squares estimates based on equation (4.1), where we use publicly available trade data to impute export growth at the county-level as described in the text. In columns (1)-(3) the WID instrument (4.2) is constructed using exposure shares based on the imputed data to address the endogeneity of exports. In columns (4)-(6) the instrument is simply our baseline instrument used in Section 4.3. The dependent variable is the relative change in CBP employment from 2007 to 2009, expressed in percent. Exports are measured in millions of dollars and employment in numbers of workers. The number of observations—rounded to the closest 100 to comply with Census disclosure requirements—is 3000 in all specifications. We report two types of standard errors. First, we report standard errors clustered by state in parentheses. Second, we report in square brackets standard errors from the "shock-level" specification and clustered by product and destination using the approach by Borusyak, Hull and Jaravel (Forthcoming). ***, **, and * indicate significance at the 1, 5, and 10 percent level. Significance is based on the "shock-level" standard errors where available.

exercise provides a tentative answer as to whether the discrepancy between our main estimates and the estimates in columns (1)-(3) result from measurement error in the change in exports or whether the imputation introduces too much noise for the instrument to have a strong first stage (or both). As columns (4)-(6) show, the instrument strength improves and the estimates rise, suggesting the imputation is certainly a problem for the instrument's first stage.

6. Concluding Remarks

This paper presents a new perspective on the patterns and effects of exporting at the county-level in the U.S. We first improve on existing data on the geography of U.S. exports by developing a novel method for allocating export transactions to a firm's establishments. Applying this methodology to microdata from the U.S. Census Bureau results in a new dataset that offers researchers greater detail about the origin of shipments to complement the high-level of detail that transaction-level export data allows. We then exploit the increased geographic breakdown to document a number of new stylized facts. We show that export activity is more geographically concentrated than either employment or sales. We further document that the differential degree of concentration amongst these variables varies substantially across U.S. states. These features of export concentration imply that conventional imputation measures that assign export exposure to local areas based on industry-level employment shares might be problematic and might—as in our application—prevent identification of the effect of interest.

We expect that this new dataset will have a diverse array of research applications. When applied to the collapse in trade during the Great Recession in the U.S., we showed that the exposure to declines in foreign demand exacerbated employment, payroll, and wage declines in local labor markets. Because these results are purged of aggregate effects, we leave for further research a full accounting of the relationship between changes in foreign demand and job losses during the Great Recession. We hope the tighter geographic link between trade activity and other economic variables will be useful for future research, and in turn, to policymakers eager to understand the impact of trade on the broader economy.

References

Ahn, JaeBin, Mary Amiti, and David E. Weinstein. 2011. "Trade Finance and the Great Trade Collapse." *American Economic Review*, 101(3): 298–302.

Alessandria, George, Joseph P. Kaboski, and Virgiliu Midrigan. 2010*a*. "Inventories, Lumpy Trade, and Large Devaluations." *American Economic Review*, 100(5): 2304–39.

Alessandria, George, Joseph P Kaboski, and Virgiliu Midrigan. 2010*b*. "The Great Trade Collapse of 2008–09: An Inventory Adjustment?." *IMF Economic Review*, 58(2): 254–294.

Allen, Treb, and Costas Arkolakis. 2014. "Trade and the Topography of the Spatial Economy." *Quarterly Journal of Economics*, 129(3): 1085–1140–48.

Atalay, Enghin, Ali Hortaçsu, and Chad Syverson. 2014. "Vertical Integration and Input Flows." *American Economic Review*, 104(4): 1120–48.

Autor, David H., David Dorn, and Gordon H. Hanson. 2013. "The China Syndrome: Local Labor Market Effects of Import Competition in the United States." *American Economic Review*, 103(6): 2121–68.

Axtell, R.L. 2001. "Zipf Distribution of U.S. Firm Sizes." *Science*, 293: 1818–1820.

Bems, Rudolfs, Robert C. Johnson, and Kei-Mu Yi. 2013. "The Great Trade Collapse." *Annual Review of Economics*, 5(1): 375–400.

Benguria, Felipe, and Felipe Saffie. 2019. "Dissecting the Impact of the 2018-2019 Trade War on U.S. Exports." *SSRN Electronic Journal*.

Benguria, Felipe, and Felipe Saffie. 2020. "The Impact of the 2018-2019 Trade War on U.S. Local Labor Markets." *SSRN Electronic Journal*.

Bernard, Andrew B., and J. Bradford Jensen. 1995. "Exporters, Jobs, and Wages in U.S. Manufacturing: 1976 – 87." *Brookings Papers on Economic Activity*, 1995: 67–112.

Bernard, Andrew B., and Teresa C. Fort. 2015. "Factoryless Goods Producing Firms." *American Economic Review Papers and Proceedings*, 105(5): 518–523.

Bernard, Andrew, Bradford Jensen, and Peter Schott. 2009. "Importers, Exporters, and Multinationals: A Portrait of Firms in the U.S. that Trade Goods." In *Producer Dynamics:*

New Evidence from Micro Data. , ed. Timothy Dunne, J. Bradford Jensen and Mark J. Roberts, 513–552. Chicago:University of Chicago Press.

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

Bernard, Andrew, J. Bradford Jensen, Stephen J. Redding, and Peter K. Schott. 2007*b*. "Firms in International Trade." NBER Working Paper 13054.

Bernard, Andrew, J. Bradford Jensen, Stephen J. Redding, and Peter K. Schott. 2007*c*. "Global Firms." *Journal of Economic Literature*, 56(2): 565–619.

Bernard, Andrew, J. Bradford Jensen, Stephen J. Redding, and Peter K. Schott. 2010. "Intrafirm Trade and Product Contractibility." *American Economic Review*, 100(2): 444–48.

Boehm, Christoph, Aaron Flaaen, and Nitya Pandalai-Nayar. 2020. "Multinationals, Offshoring, and the Decline in U.S. Manufacturing." *Journal of International Economics*, 127.

Boehm, Christoph, and Nitya Pandalai-Nayar. 2020. "Convex Supply Curves." National Bureau of Economic Research, Inc NBER Working Papers 26829.

Borusyak, Kirill, Peter Hull, and Xavier Jaravel. Forthcoming. "Quasi-Experimental Shift-Share Research Designs." *Review of Economic Studies*.

Bown, Chad P., and Meredith A. Crowley. 2013. "Import protection, business cycles, and exchange rates: Evidence from the Great Recession." *Journal of International Economics*, 90(1): 50–64.

Bussière, Matthieu, Giovanni Callegari, Fabio Ghironi, Giulia Sestieri, and Norihiko Yamano. 2013. "Estimating Trade Elasticities: Demand Composition and the Trade Collapse of 2008-2009." *American Economic Journal: Macroeconomics*, 5(3): 118–151.

Caliendo, Lorenzo, Fernando Parro, Esteban Rossi-Hansberg, and Pierre-Daniel Sarte. 2018. "The Impact of Regional and Sectoral Productivity Changes on the U.S. Economy." *The Review of Economic Studies*, 85(4): 2042–2096.

Caliendo, Lorenzo, Maximiliano Dvorkin, and Fernando Parro. 2019. "Trade and Labor Market Dynamics: General Equilibrium Analysis of the China Trade Shock." *Econometrica*, 87(3): 741–835.

Cassey, Andrew. 2009. "State Export Data: Origin of Movement vs. Origin of Production." *Journal of Economic and Social Measurement*, 34: 241–268.

Census, US. 2008. "FT-900 SUPPLEMENT December 2007." Exhibit 1.

Crozet, Matthieu, and Pamina Koenig. 2010. "Structural Gravity Equations with Intensive and Extensive Margins." *Canadian Journal of Economics*, 43(1): 41–62.

Dix-Carneiro, Rafael, and Brian K. Kovak. 2017. "Trade Liberalization and Regional Dynamics." *American Economic Review*, 107(10): 2908–46.

Eaton, Jonathan, Samuel Kortum, Brent Neiman, and John Romalis. 2016. "Trade and the Global Recession." *American Economic Review*, 106(11): 3401–38.

Eckert, Fabian, Teresa C. Fort, Peter K. Schott, and Natalie J. Yang. 2020. "Imputing Missing Values in the US Census Bureau's County Business Patterns." National Bureau of Economic Research, Inc NBER Working Papers 26632.

Feenstra, Robert C., Hong Ma, and Yuan Xu. 2019. "U.S. Exports and Employment." *Journal of International Economics*, 120: 46–58.

Feenstra, Robert C., Zhiyuan Li, and Miaojie Yu. 2014. "Exports and Credit Constraints under Incomplete Information: Theory and Evidence from China." *The Review of Economics and Statistics*, 96(4): 729–744.

Gabaix, Xavier. 1999. "Zipf's Law for Cities: An Explanation." *Quarterly Journal of Economics*, 114(3): 739–767.

Gabaix, Xavier. 2011. "The Granular Origins of Aggregate Fluctuations." *Econometrica*, 79: 733–772.

Hakobyan, Shushanik, and John McLaren. 2016. "Looking for Local Labor Market Effects of NAFTA." *The Review of Economics and Statistics*, 98(4): 728–741.

House, Christopher L, Christian Proebsting, and Linda L Tesar. 2019. "Regional Effects of Exchange Rate Fluctuations." National Bureau of Economic Research Working Paper 26071.

Hummels, David, Rasmus Jørgensen, Jakob Munch, and Chong Xiang. 2014. "The Wage Effects of Offshoring: Evidence from Danish Matched Worker-Firm Data." *American Economic Review*, 104(6): 1597–1629.

Kamal, Fariha, and Wei Ouyang. 2020. "Identifying U.S. Merchandise Traders: Integrating Customs Transactions with Business Administrative Data." Center for Economic Studies, U.S. Census Bureau Working Papers 20-28.

Kovak, Brian K. 2010. "Regional Labor Market Effects of Trade Policy: Evidence from Brazilian Liberalization." Research Seminar in International Economics, University of Michigan Working Papers 605.

Kovak, Brian, Lindsey Oldenski, and Nicholas Sly. 2021. "The Labor Market Effects of Offshoring by U.S. Multinational Firms." *Review of Economics and Statistics*, 103(2): 381–396.

Levchenko, Andrei A, Logan T Lewis, and Linda L Tesar. 2010. "The Collapse of International Trade during the 2008–09 Crisis: In Search of the Smoking Gun." *IMF Economic Review*, 58(2): 214–253.

Lincoln, William F., and Andrew H. McCallum. 2018. "The Rise of Exporting by U.S. Firms." *European Economic Review*, 102: 280–297.

Mayer, Thierry, and G Ottaviano. 2008. "The Happy Few: New Facts on the Internationalisation of European Firms." *Intereconomics - Review of European Economic Policy*, 43(3): 135–148.

Pierce, Justin, and Peter Schott. 2012. "A Concordance Between U.S. Harmonized System Codes and SIC/NAICS Product Classes and Industries." *Journal of Economic and Social Measurement*, 37(1-2): 61–96.

Santamaría, Marta, Jaume Ventura, and Uğur Yeşilbayraktar. 2020. "Borders within Europe." National Bureau of Economic Research w28301, Cambridge, MA.

Serrato, Juan Carlos Suárez, and Philippe Wingender. 2016. "Estimating Local Fiscal Multipliers." National Bureau of Economic Research, Inc NBER Working Papers 22425.

Solon, Gary, Steven J. Haider, and Jeffrey M. Wooldridge. 2015. "What Are We Weighting For?" *Journal of Human Resources*, 50(2): 301–316.

Topalova, Petia. 2007. "Trade Liberalization, Poverty and Inequality: Evidence from Indian Districts." In *Globalization and Poverty. NBER Chapters*, 291–336. National Bureau of Economic Research, Inc.

Vannoorenberghe, G. 2012. "Firm-level volatility and exports." *Journal of International Economics*, 86(1): 57–67.

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). 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 to these trade transactions was linked to other Census datasets, thus improving the dimensions of information on firm-level trade for study by economists. The principal method of linkage was the employer identification number (EIN), as it was recorded by Customs Bureau documents accompanying individual shipments and also existed as part of the establishment/firm register in the Census Bureau.[22] 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

[22]On the export side, the EIN is listed on the "Shippers Export Declaration. The exception 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.

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.

B. Details on Establishment-Level Allocation Methodology

B.1 Industry-Based Matching

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.[23] 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, we attempt to match at a more aggregated level. Through an iterated process we 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

[23]The 1992 version of the CM used SIC-based product codes.

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. We 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 we 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, we designate an industry as a PAI of product p provided that its share S_{pj} passes a certain threshold – which we set at 5 percent [24]. We define the set of industries for product p for which $S_{pj} > 0.05$ as J_p.

We match individual years of the LFTTD data to the closest available Census year. To summarize, this procedure allows one to attach a set of industries to each exported product from the LFTTD; the industry detail of establishments in the LBD can then be used to match products to establishments.

B.2 Location-Based Matching

Along with industry-based information on production, we also use geographic information to narrow the set of potential establishments involved with a particular trade transaction. Part of the shippers export declaration form (now electronically administered via the Automated Exporter System) asks for address information on the USPPI where "goods begin their

[24]We have varied this threshold without affecting our results

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, we 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 the limitations outlined in the main text. We will discuss below the issue of whether these establishments are the location of production or the location of export processing.

B.3 Survey-Based Information

A final set of information used to allocate firm-level exports to individual establishments is the export variable included in the CMF and ASM. Although the CMF 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 we use the export indicator from the CMF/ASM as a way of distinguishing between a firm's plants when the industry/location information specified above yields multiple plants associated with a given transaction.

B.4 Allocation Procedure

The paragraphs below outline how we combine these sources of information to allocate all firm-level LFTTD exports to the most likely establishment associated with that export. This allocation 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 impossible 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 allocation 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, we 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 we take the cartesian product of these firm-level transactions to the full set of firm-establishments from the LBD in each associated year – essentially making copies of each export observation to attach 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, we 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, we 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 allocation is a trivial exercise for those firms having only one establishment. We 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. We separately flag whether the establishment records positive export shipments in the ASM/CM.

- **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 our PAI criteria, then we 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, we use employment weights to allocate the exports across establishments.

- At this stage for any unallocated export transactions, we 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, simply allocate all trade to the unique PAI establishment.

- **Case 5: No zip code match but non-unique PAI establishments.** For multiple establishments matching a PAI, we must 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.** We 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 is rare), we continue to use employment shares as weights.

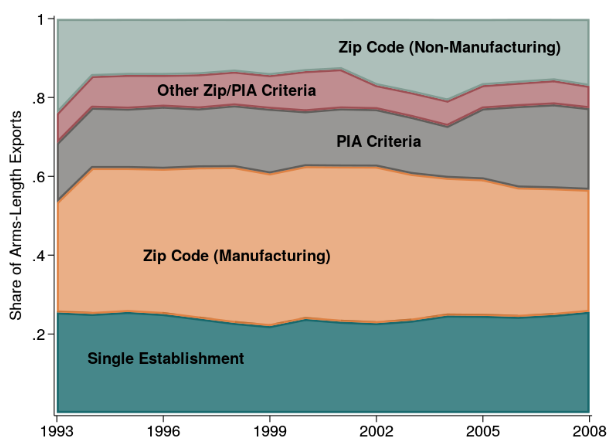
- **Case 7: No zip code match and no PAI establishments.** For this final case, we first allocate the export transaction to all manufacturing establishments of the associated firm (using either ASM/CM export share weights or employment weights). We also prioritize establishments in the wholesale (NAICS 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 allocates the export transaction to all other establishments based on employment shares.[25]

B.5 Characteristics of the Allocation

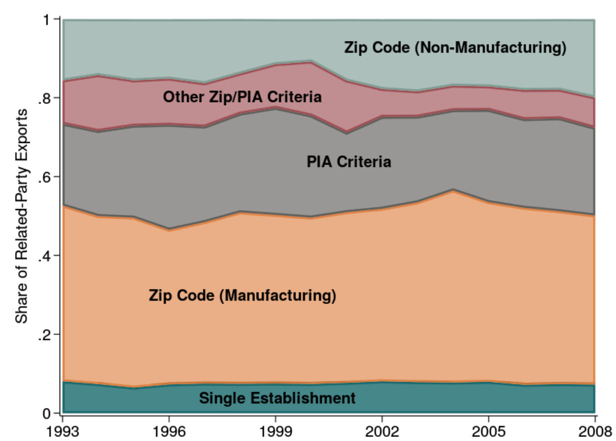
Figure A1 documents the share of overall U.S. exports that is allocated according to the hierarchy of methods as described above.

Figure A1: Share of Exports by Type of Establishment Allocation

(a) Non-Related Party Exports



(b) Related Party Exports



Source: Author's calculations as explained in text.

[25]In reality, for these cases we take the top 20 establishments by employment size, as there are some firms with thousands of establishments.

C. Comparisons to Alternative Sources of Data

C.1 Census of Manufacturers Export Variable

When focusing on exporting by manufacturing plants, the only other available indicator of export status comes from the Census of Manufacturers (and the corresponding Annual Survey of Manufacturers in non-Census years). Unlike the data described in this paper, the export variable in the CMF lacks any product or destination country detail, and is simply an annual total in dollar terms. Additionally, researchers may worry that such a survey-based measure may lack the quality of reporting that typically comes from administrative data.

To evaluate how our data compare to this other source for measure of exports, we merge our establishment export identifiers to the full CMF in 2007 and calculate the alignment between exporting status according to each measure. The result is shown in Panel A of Table A1. 85 percent (18/21 percent) of those establishments recording positive exports in the CMF also record positive exports in our allocated export measure. On the other hand, only 44 percent of establishments recording positive exports in our allocation also record positive exports in the CMF. The most likely explanation for the considerably higher export participation based on our data is under-reporting of export behavior in the CMF due to low value export shipments that are missed or unknown to survey participants.

Table A1: Plant/Firm-Level Export Participation, by Source

Panel A: Establishment-Level

		CMF Exports		
		No	**Yes**	
Alloc. Exports	**No**	56%	3%	59%
	Yes	23%	18%	41%
		79%	21%	

Panel B: Firm-Level

		CMF Exports		
		No	**Yes**	
LFTTD Exports	**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.

This explanation is confirmed when we replicate this exercise at the firm-level in Panel B of Table A1. We do this by first summing export variables by firm before merging the two datasets together. The results are largely the same as in establishment participation. @@4 percent (16/@@ percent) of firms reporting positive exports in the CMF also record positive exports in the LFTTD measure of exports. Note that one reason why there would be this type of discrepancy would be if the firm used a third-party (importer/exporter or wholesale firm) to conduct their export transactions, and hence the associated identifier from forms submitted to U.S. Customs would not align with the manufacturer. Conversely, only 4@@ percent of firms recording positive exports in the LFTTD also record positive exports in the CMF.

C.2 State-Level Origin-of-Movement Exports

Broadly speaking, there is a high correlation between the allocated exports described in this paper and the state-level export data based solely on the origin of movement variable; there are nevertheless important differences that generally support the value of our methodology. First, Figure A2 demonstrates the high overall correlation visually in a scatterplot, showing the (logged) share of state-level exports in the U.S. total between our measure and the publicly available data. Though the logged shares provide a broad view of all 50 states, it masks the few areas where the differences between the publicly available data and our data are significant: Texas (6 percentage points) and Florida (1 percentage point), and Michigan (-3 percentage points). The possibility of over-counting exports at states with ports and under-counting industrial areas is consistent with the drawbacks of the origin of movement-based approach highlighted above, and demonstrates the value-added of combining these other sources of information to provide a more accurate match location.

D. Additional Results

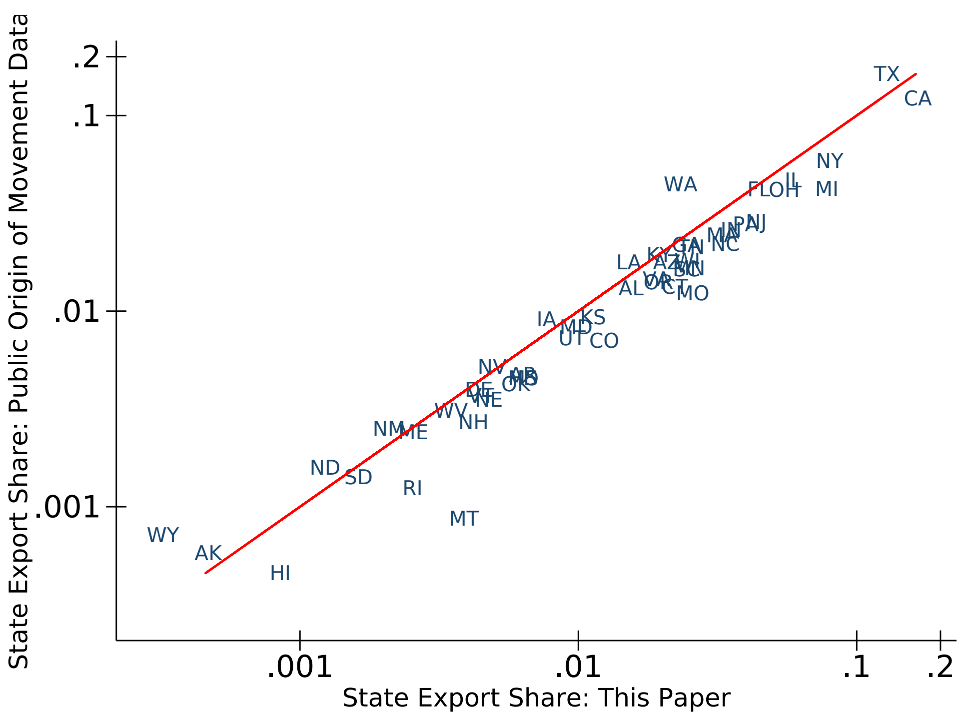
D.1 Public vs Allocated

Table A2: Economic Activity HHI Index

employment	pay	Total Trade	TVS	year
0.004	0.007	0.020	0.057	2007
0.004	0.006	0.020	0.004	2009

D.2 Concentration Facts

Figure A2: State-Level Alignment: Our Measure and Census State-Level Exports



Sources: Author's calculations using LFTTD, CMF as explained in the text. Census Bureau State-Level Trade
Notes: This figure provides the share of state-level exports in total exports (in log scale) between our measure (in the x-axis) and the state-level trade database maintained by the U.S. Census Bureau.

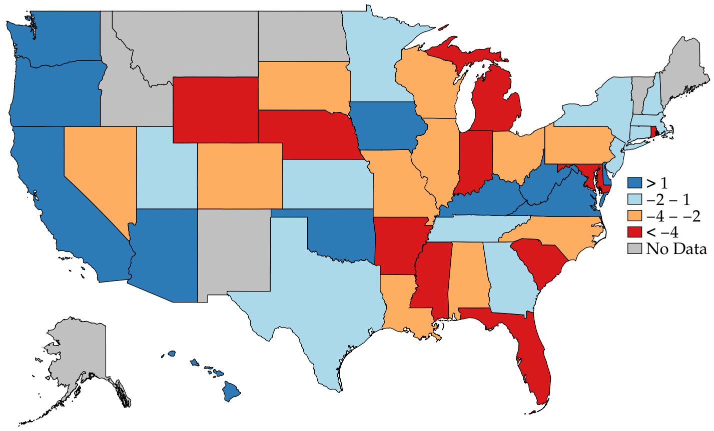
Table A3: Public Use Data Vs Allocated Data

State	Share in public data (in percent)	Share in allocation (in percent)	Percentage point difference
Texas	16.3	10.3	6.0
Florida	4.2	3.3	0.9
Louisiana	1.8	1.3	0.5
Kentucky	1.9	1.7	0.3
Washington	4.5	4.2	0.3
Georgia	2.2	1.9	0.3
Vermont	0.4	0.1	0.3
Maryland	0.8	0.6	0.2
Iowa	0.9	0.7	0.2
Utah	0.7	0.5	0.2
Nevada	0.5	0.4	0.2
Idaho	0.5	0.3	0.1
Arizona	1.8	1.7	0.1
NewMexico	0.3	0.2	0.1
Virginia	1.5	1.4	0.1
SouthDakota	0.1	0.1	0.1
WestVirginia	0.3	0.3	0.1
NorthDakota	0.2	0.1	0.1
Wyoming	0.1	0.0	0.0
Tennessee	2.1	2.1	0.0
Alabama	1.3	1.3	0.0
Maine	0.2	0.2	0.0
Alaska	0.1	0.0	0.0
Kansas	0.9	0.9	0.0
Arkansas	0.5	0.5	0.0
Mississippi	0.5	0.5	0.0
Delaware	0.4	0.4	0.0
Oklahoma	0.4	0.5	-0.1
Nebraska	0.4	0.4	-0.1
NewHampshire	0.3	0.3	-0.1
RhodeIsland	0.1	0.2	-0.1
Massachusetts	2.4	2.6	-0.2
Oregon	1.4	1.6	-0.2
SouthCarolina	1.6	1.9	-0.2
Montana	0.1	0.3	-0.2
Ohio	4.2	4.4	-0.3
Wisconsin	1.8	2.1	-0.3
NewJersey	2.9	3.1	-0.3
Colorado	0.7	1.0	-0.3
Pennsylvania	2.8	3.1	-0.3
NorthCarolina	2.2	2.6	-0.4
NewYork	5.9	6.3	-0.4
Indiana	2.6	3.0	-0.4
Connecticut	1.3	1.9	-0.5
Illinois	4.7	5.2	-0.5
Minnesota	1.7	2.2	-0.6
Missouri	1.2	1.9	-0.7
California	12.3	13.2	-0.9
Michigan	4.2	7.0	-2.8

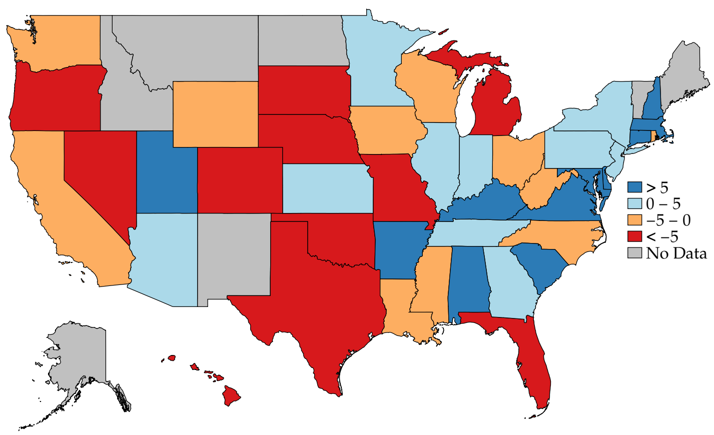
Sources: Author's calculations using LFTTD, CMF as explained in the text. Census Bureau State-Level Trade
Notes: This chart shows the share of trade each state accounts for in its respective data-source in 2007, as well as the percentage difference between the two. A positive percent difference indicates that the public use data attributes more exports to a specific state than the establishment allocation
Hawaii is excluded for disclosure purposes

Figure A3: State-Level Exports to Select Regions, Percent of Total

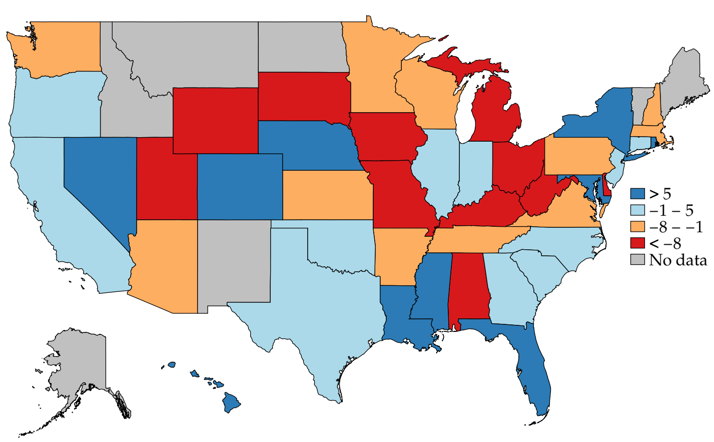
(a) State-Level Exports to Asia (excl. China)



(b) State-Level Exports to Europe



(c) State-Level Exports to Rest of World



Sources: Author's calculations using LFTTD, CMF as explained in the text.
Notes: Europe is defined as... Rest of World excludes Europe, Canada, Mexico, and Asia (including China).

Table A4: Lorenz Curves

half-decile	employment share	pay share	county tvs share	total trade share
6	0.01303	0.00768	0.00248	0.00048
7	0.01786	0.01074	0.00458	0.00093
8	0.02379	0.01466	0.00761	0.00168
9	0.03103	0.01947	0.01188	0.00290
10	0.03974	0.02532	0.01770	0.00476
11	0.05016	0.03248	0.02556	0.00767
12	0.06296	0.04133	0.03636	0.01196
13	0.07883	0.05249	0.05027	0.01828
14	0.09820	0.06626	0.06817	0.02760
15	0.12250	0.08438	0.09137	0.04139
16	0.15580	0.10990	0.12310	0.06237
17	0.20430	0.14710	0.16690	0.09504
18	0.27670	0.20530	0.23210	0.15260
19	0.41980	0.33100	0.34980	0.28170
20	1.00000	1.00000	1.00000	1.00000

Sources: U.S. Census; authors' calculations.

Notes: Table displays the cumulative sums of the column variable at a national level by counties sorted into half-deciles (20 bins) for each variable.
Bottom 5 half-deciles suppressed due to disclosure protocol

Table A5: Economic Activity HHI Index

employment	pay	total trade	tvs	year
0.004	0.007	0.020	0.057	2007
0.004	0.006	0.020	0.004	2009

Sources: U.S. Census; authors' calculations.

Notes: Table displays the Herfindahl–Hirschman Index for each variable in 2007 and 2009
Bottom 5 half-deciles suppressed due to disclosure protocols