Online Appendix for "Structural Change Within Versus Across Firms: Evidence from the United States" (Not for Publication)

Xiang Ding[*]
Georgetown University

Teresa C. Fort[†]
Tuck School of Business, CEPR and NBER

Stephen J. Redding[‡]
Princeton University, NBER and CEPR

Peter K. Schott[§]
Yale School of Management and NBER

March 2022

A Data Construction

This online appendix contains additional empirical results as well as more detailed explanations of data used in the main text.

A.1 Defining Firms

We use the new *lbdfid* variable to identify all establishments under common ownership in year t, i.e., a firm. This variable addresses the fact that certain firmids are recycled in the Business Register (BR). The Census firmid and the lbdfid contain spurious breaks by construction, whenever a single-unit firm transitions to a multi-unit, or vice-versa. This is because the Census firmid consists of "$0 \parallel EIN$" for single units, and $aplha \parallel 0000$" for multi-unit firms, where $alpha$ is the variable in the BR used to identify all establishments under common ownership of an MU firm in a given year. Our algorithm identifies these transitions and applies the MU firmid to the establishments in all years.

The longitudinal firmids created by Census for the LBD may be too restrictive for answering some research questions. As noted in the main text, they break by construction when firms grow and shrink in a particular way. That is, if a newly born single-unit (SU) firm adds another establishment in a following year, it receives a new firmid by construction because Census firmids take a different form for SU and multiple-unit (MU) firms. Likewise, if an existing MU firm sheds all but one establishment, it's firmid also will break as its firmid is changed to fit the SU pattern. A second issue with Census firmids is that they may ignore information useful for some research questions. For example, changes

[*] Georgetown University; xiang.ding@georgetown.edu.
[†] Tuck School of Business, CEPR and NBER; teresa.fort@dartmouth.edu.
[‡] Princeton University, CEPR and NBER; reddings@princeton.edu.
[§] Yale School of Management, CEPR, and NBER; peter.schott@yale.edu.

in legal ownership status or mergers and acquisitions activity can lead to breaks in Census firmids even when the firm's name and main activities are unchanged.

For this reason, studies of firm dynamics typically identify firm entry and exit using a broader conceptualization of a firm. In estimating growth rates by firm size and age categories, Haltiwanger, Jarmin, and Miranda (2013), for example, define entrants as Census firmids in year t as those with establishments that are all births in year t, and exiters as Census firmids in year t whose plants all exit in that year. Note that this approach *does not* create alternate firmids to replace the Census firmids, as they are not needed to answer the research question addressed in that paper. Instead, their approach simply identifies a firm's birth, death, or continuer status based on the status of its establishments in each year. Moreover, note that an attempt to create such firmids in the spirit of Haltiwanger, Jarmin, and Miranda (2013) likely would be unsatisfactory, particularly for research questions examining how outcomes (e.g., productivity) change within firms over time. That is because the alternate firmids to which this approach would give rise might encompass an uncomfortably large number of establishments.

A.2 Consistent Industry Codes

We assign establishments to a consistent, six-digit NAICS industry code using the following steps:

1. For analyses that span the latest years, we use the *bds_vcnaics* variable that contains the latest vintage of NAICS (NAICS 2017 in the 2019 LBD) codes for every establishment.

2. Following the directions from Fort and Klimek (2018), we merge to the $naics\_flagsYYYY$ files to pull in the number of splits for records that were randomly assigned NAICS codes. Since some establishments never contain any industry information in the data, those records are randomly assigned any industry code based solely on the distribution of codes across all industries in the aggregate economy. In other words, these codes are pure guesses. We drop these industry codes so that our analysis is contaminated by what is essentially pure noise. To do so, we sum across all the 'splits' variables in the $naics\_flagsYYYY$ files and drop industry codes for which the number of splits is greater than xxx.

3. For our regression analysis in Section 4 that spans 1997 to 2012, we use the $naicsYYYY$ files to pull the $fk\_naics2002$ variable. This provides NAICS 2002 codes for all establishments, which avoids use of any random assignments that may have been needed to map the 2002 vintage codes to NAICS 2012, which were then mapped to NAICS 2017 for the *bds_vcnaics* variable.

These steps are the most logical for most research projects using the Fort-Klimek vintage-consistent codes, and we recommend that users follow them rather than relying solely on the new LBD's *bds_vcnaics* variable.

A.3 Auxiliary Establishments

Auxiliary establishments are ones that primarily serve other establishments within the firm.

A.3.1 Identifying Auxiliaries

We use four sources of information that vary over time to identify auxiliary establishments and construct a longitudinally consistent panel.

SIC Years: During the SIC years (1977 to 2001), we use the following two sources of data to identify auxiliaries:

1. The fk_naics_aux datasets available from Census identify auxiliary establishments using the Census of Auxiliaries, the BR 'type of operation' (TOC) code, and other information. For further details, see Fort and Klimek (2018) and Chow, Fort, Goetz, Goldschlag, Lawrence, Perlman, Stinson, and White (2021).

2. The Census of Auxiliaries (AUX), that are collected in years that end in 2 and 7, through 1997. Except for in 1992 (when the AUX seems to include an implausibly large number of plants), the vast majority of establishments in the AUX are also flagged by FK as auxiliaries.

   The FK codes also flag a large number and portion of auxiliary establishments (equal to about 1/3 of the plants in the -AUX and the FK set) that are not in the AUX. While the majority is driven by LBD-only plants in the early years (i.e., plants that are not in the EC data at all and thus cannot be identified by the AUX), the FK auxes are increasingly present in the CSR data in later years. This suggests the possibility that the AUX may be missing new auxiliary establishments increasingly over time. The three or four-digit NAICS of the FK-only sectors also line up closely with those of auxiliaries identified in the AUX, suggesting they are accurately flagging aux estabs.

NAICS Years: During the NAICS years, we use the following two sources to identify auxiliaries:

1. The NAICS_AUX variable in the Business Register (BR) for years xx to xx, and then from the CBPBR starting in XX.

2. The inhouse indicator available in certain Economic Censuses (and then only for establishments that belong to multi-unit firms). The inhouse indicator in the EC data starts in 1997 in the Censuses of Services (CSR), and Transportation, Communications, and Utilities (CUT).

   - Note that there are CSR records where we calculate that MU=0 and they still have populated information for the inhouse flag, including some 1s.
   - For CRT and CWH, there are instances in which the inhouse variable is populated when MU=1. We are investigating whether these were transferred estabs from the CSR or CUT.

We investigate how these various data sources align and provide technical documentation within the FSRDC project space. Based on that work, we use all the sources of information listed above to flag auxiliary establishments. We use the plant-level information to determine which sectors contain auxiliary establishments. Specifically, we label a six-digit industry-year as one with auxiliary establishments if at least one percent of establishments and either five percent of employment or payroll is associated with auxiliary establishments within that industry and year, and if there are at least 10 auxiliary establishments. We perform the same classification at the NAICS3 level. For each aggregation level, we then count the number of auxiliary years for that industry.

We define six-digit NAICS as an auxiliary sector (i.e., a sector in which auxiliary establishments are possible) when the sector has at least two years in which it is classified as an AUX sector at the NAICS6 level, or if it is classified as an AUX sector at both the NAICS3 and the NAICS6 level in at least one year. In our final dataset, we only flag establishments as auxiliaries if they are in a six-digit NAICS industry that we classify as potentially having auxiliaries.

We also recode all 551114 establishments as auxiliaries. While the majority are coded this way in the data, there does appear to be some noise (especially with the inhouse indicator from the EC data).

A.3.2 Identifying the Industries Served by Auxiliaries

Information about the industries served by auxiliary establishments is collected at an aggregate level (e.g., not at the plant or firm level). Here, too, our procedure varies by year. During the SIC years, the data are collected at the two-digit SIC level. During the NAICS era, we observe three- and four-digit NAICS codes depending on the sector (e.g., three digits for manufacturing but four digits for wholesale).

1. $SIC\ Years$: During the SIC years, we use the fk_naics_aux files to identify these codes. Note that it is important to use only the first three digits of these codes. Note that these codes are only on a NAICS 2002 basis.

2. $SIC\ Years$: During the NAICS years, we use the NAICS_AUX variable from the BR. These are raw, native codes, so again the NAICS vintage varies by year.

B Headline numbers

This table contains the number of firms in each category.

Table B4: M and NM Employment Growth from 1977 to 2019 by Firm Type and Margin

Panel A:	**"Census Firms" (Lower Bound)**									
	Firms		Manufacturing Emp				Non-Manufacturing Emp			
	1977	2019	1977	2019	Change	Share	1977	2019	Change	Share
M Firms	285	257	17.7	12.1	-5.7	1.00	12.6	23.9	11.3	0.16
Continuers	27.5	27.5	5.6	4.5	-1.1	0.20	5.3	15.9	10.6	0.15
Net Birth/Death	257	229	12.1	7.5	-4.6	0.80	7.3	7.9	0.7	0.01
NM Firms	3211	5163					35.4	95.9	60.5	0.84
Continuers	224	224					5.6	18.2	12.6	0.18
Net Birth/Death	2987	4939					29.8	77.7	47.9	0.67
Total	3496	5420	17.7	12.1	-5.7	1.00	48.0	119.8	71.7	1.00

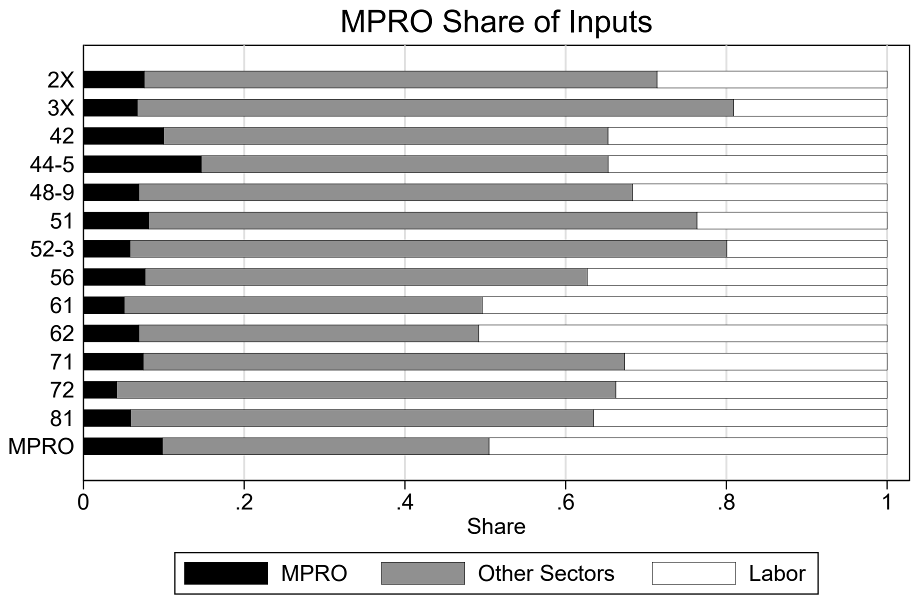
Panel B:	**"HJM Firms" (Lower Bound)**									
	Firms		Manufacturing Emp				Non-Manufacturing Emp			
	1977	2019	1977	2019	Change	Share	1977	2019	Change	Share
M Firms	314	273	17.7	12.1	-5.7	1.00	17.4	40.2	22.8	0.32
Continuers	51.5	46.0	10.8	7.2	-3.5	0.62	13.8	32.5	18.7	0.26
Net Birth/Death	262	227	7.0	4.8	-2.1	0.38	3.7	7.7	4.0	0.06
NM Firms	3183	5146					30.6	79.6	48.9	0.68
Continuers	341	332					7.1	18.5	11.4	0.16
Net Birth/Death	2842	4814					23.5	61.1	37.5	0.52
Total	3497	5419	17.7	12.1	-5.7	1.00	48.0	119.8	71.7	1.00

Source: Longitudinal Business Database (LBD) and authors' calculations. Table presents manufacturing (M) and non-manufacturing (NM) employment levels in 1977 and 2019, the change in these levels, and the share of the change accounted for by M firms, NM firms, and continuers versus net/birth day withing these firm types. M employment is the sum of employment at all US establishments in the LBD classified in manufacturing. NM employment is the sum of employment at all US establishments in the LBD classified outside manufacturing. Census M firms are those that ever have an M plant between 1977 and 2019. HJM M firms are those that ever have an establishment that was ever in a firm with an M plant in the same year. Continuing Census firms are those for which the Census *lbdfid* exists in both years. HJM continuing firms are those with an establishment in 2019 that existed in 1977. Employment is in millions.

C MPRO as a Share of Sector Inputs

Figure C1 reports the share of each two-digit NAICS sector's inputs represented by Management (NAICS 55) and Professional Services (NAICS 54) versus labor (BEA code V00100) and other sectors.

Figure C1: MPRO Share of Sector Inputs, 1997



Source: Bureau of Economic Analysis and authors' calculations. Figure displays the share of each sector's inputs accounted for by Management (NAICS 55) and Professional Services (NAICS 54). Data are from the detailed 1997 US Supply-Use Table published on the BEA website.

D Reduced-form Evidence

In this section, we describe how we construct the shocks, document variation in input versus output exposure, and discuss concerns about the standard errors.

D.1 Shock construction

D.1.1 Industry Classifications

Because of the need to concord across different trade data sources and our firm-level dataset, we work with slightly different industry classifications as that in the empirical section of the paper. Our ultimate level of analysis is at the firm level and our definition of industries only matter to the extent that they provide identifying variation. We start by working with the NAICS 1997 industry codes given the timing of the shocks we exploit is between 1997-2007.

We use one set of industry classifications for output shocks (NAICS-X) and another slightly coarsened classification for input shocks (NAICS-B). There are 440 NAICS-X and 330 NAICS-B. At the most disaggregated level, we divide the manufacturing sector into 440 industries. These are near-identical to NAICS-6 codes but with some last digits aggregated in to be concordable with HS6 codes. We call this set of codes NAICS-X. To construct input shocks, we coarsen NAICS-X slightly into NAICS-B in order to concord NAICS-X codes with BEA codes. We aggregate BEA's most disaggregated industry definitions into a set of codes we call NAICS-B—the most disaggregated level at which each NAICS-X is entirely subsumed within a NAICS-B. Our industry-level input shocks are defined at this, roughly NAICS-5 digit level. Next, many input purchases in the materials trailer are given at the 3-digit, 4-digit, or 5-digit NAICS level of clarity rather than at the 6-digit level. (That is, they have root codes that are not a part of the NAICS nomenclature). We thus prepare a set of industry shocks at each of these levels so that a firm observed to be using an input k that is a 4-digit NAICS is given the shock associated with that particular 4-digit manufacturing sub-sector.

D.1.2 Industry-Level Shocks

Output Shocks. We compute China's exports to two sets of markets, in 1997 and 2007: the EU and the US. The EU market is the basis for our "China shock" variable, and the US market is the basis for our endogenous measures of Chinese import competition. The EU market includes the following countries: Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, Netherlands, Portugal, Spain, Sweden, and United Kingdom.[¶] We use US Comtrade data at the HS 1996 classification to construct shock variables.

To isolate the portion of increased trade attributable to China's productivity and liberalization, we focus on changes in China's market *shares* in EU and US markets. We measure China's market share in the EU as China's exports to the EU divided by total imports of the EU excluding its imports from the US. We measure China's market share in the US as total US industry absorption, which we construct using the census data as total sales less exports in the CMF plus imports from the Comtrade data:

$$ChineseMktSh_{j,t}^{EU} \equiv \frac{Imports_{jt}^{EU \leftarrow Chn}}{Imports_{jt}^{EU,excl.US}},$$

$$ChineseMktSh_{j,t}^{US} \equiv \frac{Imports_{jt}^{US \leftarrow Chn}}{Sales_{jt}^{US} - Exports_{jt}^{US} + Imports_{jt}^{US}}.$$

The difference in these market shares between 1997 and 2007 measure changes in market competition from China. We define our industry (j)-level output shock variable $\Delta Output_j^{EU}$ as

$$\Delta Output_j^{EU} \equiv ChineseMktSh_{j,2007}^{EU} - ChineseMktSh_{j,1997}^{EU},$$

and, equivalently, our endogenous measure of China's competition in the US as

$$\Delta Output_j^{US} \equiv ChineseMktSh_{j,2007}^{US} - ChineseMktSh_{j,1997}^{US}.$$

Input Shocks. Increased Chinese competitiveness in industry k also affects production costs of firms in downstream industries j. We create a measure of average manufacturing input cost shocks

[¶]Belgium and Luxembourg reported as Belgium-Luxembourg in 1997.

$\Delta Input_j^{mkt}$ in each industry j (a given NAICS-B code) as

$$\Delta Input_j^{mkt} \equiv \sum_{k \in \mathcal{J}_T} \lambda_{kj} \Delta Output_k^{mkt}, \quad \forall \; mkt \in \{\text{EU, US}\},$$

where λ_{kj} are the expenditures of industry j on (manuf) input industry k as a share of total expenditures on inputs from the manufacturing sector (so the shares sum to one). Data on λ_{kj} come from the BEA's I-O tables in 1997 (at purchaser values, which includes retail wholesale margins on prices of inputs paid).

D.1.3 Firm-Level Shocks

We use cross-industry variation in the material and product trailers to create input and output shocks relevant for each firm f. A firm-level shock is a weighted average of industry-level shocks. For single-plant firms, the shock is just the shock constructed from material and product trailer information associated with that plant. For multi-plant firms, we average their plants' shocks using relative plant-level sales as weights. However, we only have shock content for the manufacturing activities of the firm. To account for the activities of firms outside of manufacturing, we scale each firms' input and output shocks using the firm's share of sales in manufacturing, defined as η_f.

We construct each type of shock at the firm-level by aggregating over plant-level information. We define our firm-level output shock measure as

$$\Delta Output_f \equiv \eta_f \sum_{p \in Mnf} s_{fp} \Delta Output_p^{EU}, \tag{D1}$$

and our firm-level input shock measure as:

$$\Delta Input_f \equiv \eta_f \sum_{p \in Mnf} s_{fp} \Delta Input_p^{EU}, \tag{D2}$$

where s_{fp} are shares that sum to one: plant p's manufacturing sales as a share of total manufacturing sales of the firm. We measure sales by industry using the variable `pv` in the product trailers of each plant, and measure total sales by aggregating total shipments (usually `tvs`) across the ECs.

In this appendix we also present additional results using endogenous measures of China's competitiveness in US output and input markets. We define these equivalently as

$$\Delta ChineseMktSh_f^{US,Output} \equiv \eta_f \sum_{p \in Mnf} s_{fp} \Delta Output_p^{US}, \tag{D3}$$

and

$$\Delta ChineseMktSh_f^{US,Input} \equiv \eta_f \sum_{p \in Mnf} s_{fp} \Delta Input_p^{US}. \tag{D4}$$

Plant-level Output Shocks. The output shock for the plant is simply a weighted average over industry output shocks in each product produced by the plant. PT data is available for all plants in the CMF so that a plant in the PT sells in at least one manuf industry.

$$\Delta Output_p^{mkt} \equiv \sum_{j \in \mathcal{J}_{mnf}} s_{pj} \Delta Output_j^{mkt}, \quad \forall\ mkt \in \{\text{EU, US}\},$$

where s_{pj} are shares of plant p's sales in industry j among all total manufacturing sales of the plant. We drop a miniscule fraction of product lines reported in the PT do not match to any manufacturing NAICS code.

Plant-level Input Shocks. Some plants have useful information from the MT (line items that match up with industry codes up to the 3-digit level). Other plants do not. We measure how much of the plant's total cost of materials and parts (`cp`) is reflected in discernible MT line items, including both manufacturing inputs and non-manufacturing inputs like agriculture. If the share of material costs that are discernible exceed 0.5, we use MT information to construct input shocks (Scenario A). Otherwise we rely on the industry codes of products *sold* by the plant (Scenario B).

(A) **MT Information.** For these plants, we construct the plant-specific input cost shock as a weighted average of input-industry output shocks over the (discernible manufacturing) materials k used by the plant:

$$\Delta Input_p^{mkt} \equiv \lambda_p^* \sum_{k \in \mathcal{J}} \lambda_{pk} \Delta Output_k^{mkt}, \quad \forall\ mkt \in \{\text{EU, US}\},$$

where λ_{pk} denote the plant's spending on material k (line-item expense variable `mc` in the MT) as a share of its total spending on manufacturing materials, and λ_p^* is the plant's expenses on materials and parts `cp` as a share of the plant's total variable costs—defined as cost of materials, resales, fuels, electricity, and production worker payroll ($\texttt{cm} + \texttt{ww}$).

(B) **PT Sales Information.** A subset of plants have discernible MT line item expenses as a share of material costs < 0.5. For these plants, we construct the input cost shock as the average input shock over the plant's industries:

$$\Delta Input_p^{mkt} \equiv \lambda_p^* \sum_{j \in \mathcal{J}} s_{pj} \Delta Input_j^{mkt}, \quad \forall\ mkt \in \{\text{EU, US}\},$$

where s_{pj} is the share of industry j in the plant's total sales, and λ_p^* is the same as defined in the scenario above ($\texttt{cp}/(\texttt{cm} + \texttt{ww})$ come from the CMF and not the material trailer, so it is available for every plant). The use of plant-level information even in this scenario makes that two plants in the same industry can have different input shocks to the extent that they have differing manufacturing input expenditure share of total variable costs, λ_p^*.

Finally, since the plant-level input shocks were created using shares that do not sum to one, we carry over the scaling factor λ_p^* to the level of the firm by passing it through the same plant sales shares:

$$\lambda_f^* \equiv \eta_f \sum_{p \in Mnf} s_{fp} \lambda_p^*. \tag{D5}$$

We include λ_f^* and s_f^* as controls in every regression specification involving input and output shocks. We refer to these as manufacturing input and output shares of the firm. The manufacturing input share controls for differences in shock measures caused by any unobserved cost shocks in the firm's non-manufacturing inputs. The manufacturing output share controls for differences in shock measures caused by any unobserved residual demand shocks in the firm's non-manufacturing output markets.

Table D5 summarizes variation in input versus output exposure measures.

Table D5: Correlation Matrix of measures of China's Import Competition in Output and Input Markets

	(1)	(2)	(3)	(4)
(1) $\Delta ChineseMktSh_f^{US,Output}$		0.523***	0.418***	0.177***
		(0.000)	(0.000)	(0.000)
(2) Output Shock	0.408***		0.306***	0.170***
	(0.000)		(0.000)	(0.000)
(3) $\Delta ChineseMktSh_f^{US,Input}$	0.151***	0.163***		0.656***
	(0.000)	(0.000)		(0.000)
(4) Input Shock	0.087***	0.050***	0.525***	
	(0.000)	(0.000)	(0.000)	

Notes: This table provides a correlation matrix between both endogenous and exogenous versions of the input and output Chinese competition measures used in the regression sample of firms. The upper-right triangle of the matrix provides the raw correlations, and the lower-right triangle of the matrix provides the correlations after residualizing regressors on the same controls used in the baseline regression specification (without interactions): $ln(emp_f^{1997})$, firm age, the log number of establishments, the share of sales in manufacturing, the share of materials in manufacturing costs, and 4-digit NAICS fixed effects.

D.2 China Shock Regressions

D.2.1 Regression Sample

Our main regression sample contains firms with manufacturing output in 1997 that continue between 1997 and 2007. We focus on firms that were manufacturers in 1997 because our input and output shocks contain variation only within the manufacturing sector. Table D6 presents summary statistics for key regressors.

Online Appendix for "Structural Change Within Versus Across Firms: Evidence from the United States" (Not for Publication)

Xiang Ding*
Georgetown University

Teresa C. Fort†
Tuck School of Business, CEPR and NBER

Stephen J. Redding‡
Princeton University, NBER and CEPR

Peter K. Schott§
Yale School of Management and NBER

March 2022

A Data Construction

This online appendix contains additional empirical results as well as more detailed explanations of data used in the main text.

A.1 Defining Firms

We use the new $lbdfid$ variable to identify all establishments under common ownership in year t, i.e., a firm. This variable addresses the fact that certain firmids are recycled in the Business Register (BR). The Census firmid and the lbdfid contain spurious breaks by construction, whenever a single-unit firm transitions to a multi-unit, or vice-versa. This is because the Census firmid consists of "$0 \parallel EIN$" for single units, and $aplha \parallel 0000$" for multi-unit firms, where $alpha$ is the variable in the BR used to identify all establishments under common ownership of an MU firm in a given year. Our algorithm identifies these transitions and applies the MU firmid to the establishments in all years.

The longitudinal firmids created by Census for the LBD may be too restrictive for answering some research questions. As noted in the main text, they break by construction when firms grow and shrink in a particular way. That is, if a newly born single-unit (SU) firm adds another establishment in a following year, it receives a new firmid by construction because Census firmids take a different form for SU and multiple-unit (MU) firms. Likewise, if an existing MU firm sheds all but one establishment, it's firmid also will break as its firmid is changed to fit the SU pattern. A second issue with Census firmids is that they may ignore information useful for some research questions. For example, changes

*Georgetown University; xiang.ding@georgetown.edu.
†Tuck School of Business, CEPR and NBER; teresa.fort@dartmouth.edu.
‡Princeton University, CEPR and NBER; reddings@princeton.edu.
§Yale School of Management, CEPR, and NBER; peter.schott@yale.edu.

in legal ownership status or mergers and acquisitions activity can lead to breaks in Census firmids even when the firm's name and main activities are unchanged.

For this reason, studies of firm dynamics typically identify firm entry and exit using a broader conceptualization of a firm. In estimating growth rates by firm size and age categories, Haltiwanger, Jarmin, and Miranda (2013), for example, define entrants as Census firmids in year t as those with establishments that are all births in year t, and exiters as Census firmids in year t whose plants all exit in that year. Note that this approach *does not* create alternate firmids to replace the Census firmids, as they are not needed to answer the research question addressed in that paper. Instead, their approach simply identifies a firm's birth, death, or continuer status based on the status of its establishments in each year. Moreover, note that an attempt to create such firmids in the spirit of Haltiwanger, Jarmin, and Miranda (2013) likely would be unsatisfactory, particularly for research questions examining how outcomes (e.g., productivity) change within firms over time. That is because the alternate firmids to which this approach would give rise might encompass an uncomfortably large number of establishments.

A.2 Consistent Industry Codes

We assign establishments to a consistent, six-digit NAICS industry code using the following steps:

1. For analyses that span the latest years, we use the *bds_vcnaics* variable that contains the latest vintage of NAICS (NAICS 2017 in the 2019 LBD) codes for every establishment.

2. Following the directions from Fort and Klimek (2018), we merge to the $naics\_flagsYYYY$ files to pull in the number of splits for records that were randomly assigned NAICS codes. Since some establishments never contain any industry information in the data, those records are randomly assigned any industry code based solely on the distribution of codes across all industries in the aggregate economy. In other words, these codes are pure guesses. We drop these industry codes so that our analysis is contaminated by what is essentially pure noise. To do so, we sum across all the 'splits' variables in the $naics\_flagsYYYY$ files and drop industry codes for which the number of splits is greater than xxx.

3. For our regression analysis in Section 4 that spans 1997 to 2012, we use the $naicsYYYY$ files to pull the $fk\_naics2002$ variable. This provides NAICS 2002 codes for all establishments, which avoids use of any random assignments that may have been needed to map the 2002 vintage codes to NAICS 2012, which were then mapped to NAICS 2017 for the *bds_vcnaics* variable.

These steps are the most logical for most research projects using the Fort-Klimek vintage-consistent codes, and we recommend that users follow them rather than relying solely on the new LBD's *bds_vcnaics* variable.

A.3 Auxiliary Establishments

Auxiliary establishments are ones that primarily serve other establishments within the firm.

A.3.1 Identifying Auxiliaries

We use four sources of information that vary over time to identify auxiliary establishments and construct a longitudinally consistent panel.

SIC Years: During the SIC years (1977 to 2001), we use the following two sources of data to identify auxiliaries:

1. The fk_naics_aux datasets available from Census identify auxiliary establishments using the Census of Auxiliaries, the BR 'type of operation' (TOC) code, and other information. For further details, see Fort and Klimek (2018) and Chow, Fort, Goetz, Goldschlag, Lawrence, Perlman, Stinson, and White (2021).

2. The Census of Auxiliaries (AUX), that are collected in years that end in 2 and 7, through 1997. Except for in 1992 (when the AUX seems to include an implausibly large number of plants), the vast majority of establishments in the AUX are also flagged by FK as auxiliaries.

   The FK codes also flag a large number and portion of auxiliary establishments (equal to about 1/3 of the plants in the -AUX and the FK set) that are not in the AUX. While the majority is driven by LBD-only plants in the early years (i.e., plants that are not in the EC data at all and thus cannot be identified by the AUX), the FK auxes are increasingly present in the CSR data in later years. This suggests the possibility that the AUX may be missing new auxiliary establishments increasingly over time. The three or four-digit NAICS of the FK-only sectors also line up closely with those of auxiliaries identified in the AUX, suggesting they are accurately flagging aux estabs.

NAICS Years: During the NAICS years, we use the following two sources to identify auxiliaries:

1. The NAICS_AUX variable in the Business Register (BR) for years xx to xx, and then from the CBPBR starting in XX.

2. The inhouse indicator available in certain Economic Censuses (and then only for establishments that belong to multi-unit firms). The inhouse indicator in the EC data starts in 1997 in the Censuses of Services (CSR), and Transportation, Communications, and Utilities (CUT).

   - Note that there are CSR records where we calculate that MU=0 and they still have populated information for the inhouse flag, including some 1s.
   - For CRT and CWH, there are instances in which the inhouse variable is populated when MU=1. We are investigating whether these were transferred estabs from the CSR or CUT.

We investigate how these various data sources align and provide technical documentation within the FSRDC project space. Based on that work, we use all the sources of information listed above to flag auxiliary establishments. We use the plant-level information to determine which sectors contain auxiliary establishments. Specifically, we label a six-digit industry-year as one with auxiliary establishments if at least one percent of establishments and either five percent of employment or payroll is associated with auxiliary establishments within that industry and year, and if there are at least 10 auxiliary establishments. We perform the same classification at the NAICS3 level. For each aggregation level, we then count the number of auxiliary years for that industry.

We define six-digit NAICS as an auxiliary sector (i.e., a sector in which auxiliary establishments are possible) when the sector has at least two years in which it is classified as an AUX sector at the NAICS6 level, or if it is classified as an AUX sector at both the NAICS3 and the NAICS6 level in at least one year. In our final dataset, we only flag establishments as auxiliaries if they are in a six-digit NAICS industry that we classify as potentially having auxiliaries.

We also recode all 551114 establishments as auxiliaries. While the majority are coded this way in the data, there does appear to be some noise (especially with the inhouse indicator from the EC data).

A.3.2 Identifying the Industries Served by Auxiliaries

Information about the industries served by auxiliary establishments is collected at an aggregate level (e.g., not at the plant or firm level). Here, too, our procedure varies by year. During the SIC years, the data are collected at the two-digit SIC level. During the NAICS era, we observe three- and four-digit NAICS codes depending on the sector (e.g., three digits for manufacturing but four digits for wholesale).

1. *SIC Years*: During the SIC years, we use the fk_naics_aux files to identify these codes. Note that it is important to use only the first three digits of these codes. Note that these codes are only on a NAICS 2002 basis.

2. *SIC Years*: During the NAICS years, we use the NAICS_AUX variable from the BR. These are raw, native codes, so again the NAICS vintage varies by year.

B Headline numbers

This table contains the number of firms in each category.

Table B4: M and NM Employment Growth from 1977 to 2019 by Firm Type and Margin

Panel A:	**"Census Firms" (Lower Bound)**									
	Firms		Manufacturing Emp				Non-Manufacturing Emp			
	1977	2019	1977	2019	Change	Share	1977	2019	Change	Share
M Firms	285	257	17.7	12.1	-5.7	1.00	12.6	23.9	11.3	0.16
Continuers	27.5	27.5	5.6	4.5	-1.1	0.20	5.3	15.9	10.6	0.15
Net Birth/Death	257	229	12.1	7.5	-4.6	0.80	7.3	7.9	0.7	0.01
NM Firms	3211	5163					35.4	95.9	60.5	0.84
Continuers	224	224					5.6	18.2	12.6	0.18
Net Birth/Death	2987	4939					29.8	77.7	47.9	0.67
Total	3496	5420	17.7	12.1	-5.7	1.00	48.0	119.8	71.7	1.00

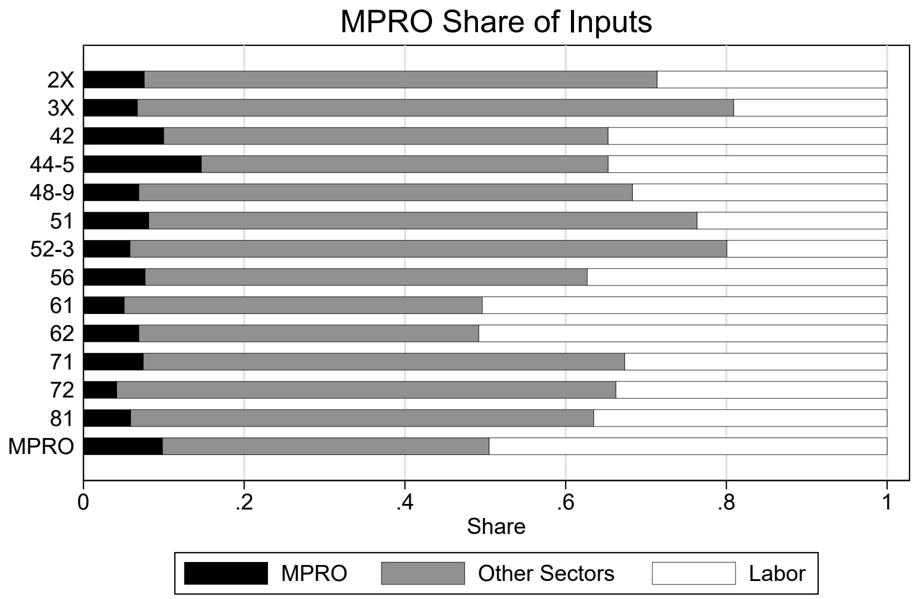
Panel B:	**"HJM Firms" (Lower Bound)**									
	Firms		Manufacturing Emp				Non-Manufacturing Emp			
	1977	2019	1977	2019	Change	Share	1977	2019	Change	Share
M Firms	314	273	17.7	12.1	-5.7	1.00	17.4	40.2	22.8	0.32
Continuers	51.5	46.0	10.8	7.2	-3.5	0.62	13.8	32.5	18.7	0.26
Net Birth/Death	262	227	7.0	4.8	-2.1	0.38	3.7	7.7	4.0	0.06
NM Firms	3183	5146					30.6	79.6	48.9	0.68
Continuers	341	332					7.1	18.5	11.4	0.16
Net Birth/Death	2842	4814					23.5	61.1	37.5	0.52
Total	3497	5419	17.7	12.1	-5.7	1.00	48.0	119.8	71.7	1.00

Source: Longitudinal Business Database (LBD) and authors' calculations. Table presents manufacturing (M) and non-manufacturing (NM) employment levels in 1977 and 2019, the change in these levels, and the share of the change accounted for by M firms, NM firms, and continuers versus net/birth day withing these firm types. M employment is the sum of employment at all US establishments in the LBD classified in manufacturing. NM employment is the sum of employment at all US establishments in the LBD classified outside manufacturing. Census M firms are those that ever have an M plant between 1977 and 2019. HJM M firms are those that ever have an establishment that was ever in a firm with an M plant in the same year. Continuing Census firms are those for which the Census *lbdfid* exists in both years. HJM continuing firms are those with an establishment in 2019 that existed in 1977. Employment is in millions.

C MPRO as a Share of Sector Inputs

Figure C1 reports the share of each two-digit NAICS sector's inputs represented by Management (NAICS 55) and Professional Services (NAICS 54) versus labor (BEA code V00100) and other sectors.

Figure C1: MPRO Share of Sector Inputs, 1997



Source: Bureau of Economic Analysis and authors' calculations. Figure displays the share of each sector's inputs accounted for by Management (NAICS 55) and Professional Services (NAICS 54). Data are from the detailed 1997 US Supply-Use Table published on the BEA website.

D Reduced-form Evidence

In this section, we describe how we construct the shocks, document variation in input versus output exposure, and discuss concerns about the standard errors.

D.1 Shock construction

D.1.1 Industry Classifications

Because of the need to concord across different trade data sources and our firm-level dataset, we work with slightly different industry classifications as that in the empirical section of the paper. Our ultimate level of analysis is at the firm level and our definition of industries only matter to the extent that they provide identifying variation. We start by working with the NAICS 1997 industry codes given the timing of the shocks we exploit is between 1997-2007.

We use one set of industry classifications for output shocks (NAICS-X) and another slightly coarsened classification for input shocks (NAICS-B). There are 440 NAICS-X and 330 NAICS-B. At the most disaggregated level, we divide the manufacturing sector into 440 industries. These are near-identical to NAICS-6 codes but with some last digits aggregated in to be concordable with HS6 codes. We call this set of codes NAICS-X. To construct input shocks, we coarsen NAICS-X slightly into NAICS-B in order to concord NAICS-X codes with BEA codes. We aggregate BEA's most disaggregated industry definitions into a set of codes we call NAICS-B—the most disaggregated level at which each NAICS-X is entirely subsumed within a NAICS-B. Our industry-level input shocks are defined at this, roughly NAICS-5 digit level. Next, many input purchases in the materials trailer are given at the 3-digit, 4-digit, or 5-digit NAICS level of clarity rather than at the 6-digit level. (That is, they have root codes that are not a part of the NAICS nomenclature). We thus prepare a set of industry shocks at each of these levels so that a firm observed to be using an input k that is a 4-digit NAICS is given the shock associated with that particular 4-digit manufacturing sub-sector.

D.1.2 Industry-Level Shocks

Output Shocks. We compute China's exports to two sets of markets, in 1997 and 2007: the EU and the US. The EU market is the basis for our "China shock" variable, and the US market is the basis for our endogenous measures of Chinese import competition. The EU market includes the following countries: Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, Netherlands, Portugal, Spain, Sweden, and United Kingdom.[¶] We use US Comtrade data at the HS 1996 classification to construct shock variables.

To isolate the portion of increased trade attributable to China's productivity and liberalization, we focus on changes in China's market *shares* in EU and US markets. We measure China's market share in the EU as China's exports to the EU divided by total imports of the EU excluding its imports from the US. We measure China's market share in the US as total US industry absorption, which we construct using the census data as total sales less exports in the CMF plus imports from the Comtrade data:

$$ChineseMktSh_{j,t}^{EU} \equiv \frac{Imports_{jt}^{EU \leftarrow Chn}}{Imports_{jt}^{EU,excl.US}},$$

$$ChineseMktSh_{j,t}^{US} \equiv \frac{Imports_{jt}^{US \leftarrow Chn}}{Sales_{jt}^{US} - Exports_{jt}^{US} + Imports_{jt}^{US}}.$$

The difference in these market shares between 1997 and 2007 measure changes in market competition from China. We define our industry (j)-level output shock variable $\Delta Output_j^{EU}$ as

$$\Delta Output_j^{EU} \equiv ChineseMktSh_{j,2007}^{EU} - ChineseMktSh_{j,1997}^{EU},$$

and, equivalently, our endogenous measure of China's competition in the US as

$$\Delta Output_j^{US} \equiv ChineseMktSh_{j,2007}^{US} - ChineseMktSh_{j,1997}^{US}.$$

Input Shocks. Increased Chinese competitiveness in industry k also affects production costs of firms in downstream industries j. We create a measure of average manufacturing input cost shocks

[¶]Belgium and Luxembourg reported as Belgium-Luxembourg in 1997.

$\Delta Input_j^{mkt}$ in each industry j (a given NAICS-B code) as

$$\Delta Input_j^{mkt} \equiv \sum_{k \in \mathcal{J}_T} \lambda_{kj} \Delta Output_k^{mkt}, \quad \forall\ mkt \in \{\text{EU, US}\},$$

where λ_{kj} are the expenditures of industry j on (manuf) input industry k as a share of total expenditures on inputs from the manufacturing sector (so the shares sum to one). Data on λ_{kj} come from the BEA's I-O tables in 1997 (at purchaser values, which includes retail wholesale margins on prices of inputs paid).

D.1.3 Firm-Level Shocks

We use cross-industry variation in the material and product trailers to create input and output shocks relevant for each firm f. A firm-level shock is a weighted average of industry-level shocks. For single-plant firms, the shock is just the shock constructed from material and product trailer information associated with that plant. For multi-plant firms, we average their plants' shocks using relative plant-level sales as weights. However, we only have shock content for the manufacturing activities of the firm. To account for the activities of firms outside of manufacturing, we scale each firms' input and output shocks using the firm's share of sales in manufacturing, defined as η_f.

We construct each type of shock at the firm-level by aggregating over plant-level information. We define our firm-level output shock measure as

$$\Delta Output_f \equiv \eta_f \sum_{p \in Mnf} s_{fp} \Delta Output_p^{EU}, \tag{D1}$$

and our firm-level input shock measure as:

$$\Delta Input_f \equiv \eta_f \sum_{p \in Mnf} s_{fp} \Delta Input_p^{EU}, \tag{D2}$$

where s_{fp} are shares that sum to one: plant p's manufacturing sales as a share of total manufacturing sales of the firm. We measure sales by industry using the variable `pv` in the product trailers of each plant, and measure total sales by aggregating total shipments (usually `tvs`) across the ECs.

In this appendix we also present additional results using endogenous measures of China's competitiveness in US output and input markets. We define these equivalently as

$$\Delta ChineseMktSh_f^{US,Output} \equiv \eta_f \sum_{p \in Mnf} s_{fp} \Delta Output_p^{US}, \tag{D3}$$

and

$$\Delta ChineseMktSh_f^{US,Input} \equiv \eta_f \sum_{p \in Mnf} s_{fp} \Delta Input_p^{US}. \tag{D4}$$

Plant-level Output Shocks. The output shock for the plant is simply a weighted average over industry output shocks in each product produced by the plant. PT data is available for all plants in the CMF so that a plant in the PT sells in at least one manuf industry.

$$\Delta Output_p^{mkt} \equiv \sum_{j \in \mathcal{J}_{mnf}} s_{pj} \Delta Output_j^{mkt}, \quad \forall\ mkt \in \{\text{EU, US}\},$$

where s_{pj} are shares of plant p's sales in industry j among all total manufacturing sales of the plant. We drop a miniscule fraction of product lines reported in the PT do not match to any manufacturing NAICS code.

Plant-level Input Shocks. Some plants have useful information from the MT (line items that match up with industry codes up to the 3-digit level). Other plants do not. We measure how much of the plant's total cost of materials and parts (`cp`) is reflected in discernible MT line items, including both manufacturing inputs and non-manufacturing inputs like agriculture. If the share of material costs that are discernible exceed 0.5, we use MT information to construct input shocks (Scenario A). Otherwise we rely on the industry codes of products *sold* by the plant (Scenario B).

(A) **MT Information.** For these plants, we construct the plant-specific input cost shock as a weighted average of input-industry output shocks over the (discernible manufacturing) materials k used by the plant:

$$\Delta Input_p^{mkt} \equiv \lambda_p^* \sum_{k \in \mathcal{J}} \lambda_{pk} \Delta Output_k^{mkt}, \quad \forall\ mkt \in \{\text{EU, US}\},$$

where λ_{pk} denote the plant's spending on material k (line-item expense variable `mc` in the MT) as a share of its total spending on manufacturing materials, and λ_p^* is the plant's expenses on materials and parts `cp` as a share of the plant's total variable costs—defined as cost of materials, resales, fuels, electricity, and production worker payroll ($\texttt{cm} + \texttt{ww}$).

(B) **PT Sales Information.** A subset of plants have discernible MT line item expenses as a share of material costs < 0.5. For these plants, we construct the input cost shock as the average input shock over the plant's industries:

$$\Delta Input_p^{mkt} \equiv \lambda_p^* \sum_{j \in \mathcal{J}} s_{pj} \Delta Input_j^{mkt}, \quad \forall\ mkt \in \{\text{EU, US}\},$$

where s_{pj} is the share of industry j in the plant's total sales, and λ_p^* is the same as defined in the scenario above ($\texttt{cp}/(\texttt{cm} + \texttt{ww})$ come from the CMF and not the material trailer, so it is available for every plant). The use of plant-level information even in this scenario makes that two plants in the same industry can have different input shocks to the extent that they have differing manufacturing input expenditure share of total variable costs, λ_p^*.

Finally, since the plant-level input shocks were created using shares that do not sum to one, we carry over the scaling factor λ_p^* to the level of the firm by passing it through the same plant sales shares:

$$\lambda_f^* \equiv \eta_f \sum_{p \in Mnf} s_{fp} \lambda_p^*. \tag{D5}$$

We include λ_f^* and s_f^* as controls in every regression specification involving input and output shocks. We refer to these as manufacturing input and output shares of the firm. The manufacturing input share controls for differences in shock measures caused by any unobserved cost shocks in the firm's non-manufacturing inputs. The manufacturing output share controls for differences in shock measures caused by any unobserved residual demand shocks in the firm's non-manufacturing output markets.

Table D5 summarizes variation in input versus output exposure measures.

Table D5: Correlation Matrix of measures of China's Import Competition in Output and Input Markets

	(1)	(2)	(3)	(4)
(1) $\Delta ChineseMktSh_f^{US,Output}$		0.523***	0.418***	0.177***
		(0.000)	(0.000)	(0.000)
(2) Output Shock	0.408***		0.306***	0.170***
	(0.000)		(0.000)	(0.000)
(3) $\Delta ChineseMktSh_f^{US,Input}$	0.151***	0.163***		0.656***
	(0.000)	(0.000)		(0.000)
(4) Input Shock	0.087***	0.050***	0.525***	
	(0.000)	(0.000)	(0.000)	

Notes: This table provides a correlation matrix between both endogenous and exogenous versions of the input and output Chinese competition measures used in the regression sample of firms. The upper-right triangle of the matrix provides the raw correlations, and the lower-right triangle of the matrix provides the correlations after residualizing regressors on the same controls used in the baseline regression specification (without interactions): $ln(emp_f^{1997})$, firm age, the log number of establishments, the share of sales in manufacturing, the share of materials in manufacturing costs, and 4-digit NAICS fixed effects.

D.2 China Shock Regressions

D.2.1 Regression Sample

Our main regression sample contains firms with manufacturing output in 1997 that continue between 1997 and 2007. We focus on firms that were manufacturers in 1997 because our input and output shocks contain variation only within the manufacturing sector. Table D6 presents summary statistics for key regressors.

Table D6: Regression Firm Sample Summary Statistics

		Mean	Standard Deviation
Output Shock		0.1438	0.0904
	x AUX	0.0037	0.0242
	x ln(emp)	0.4699	0.3476
Input Shock		0.0614	0.0425
	x AUX	0.0024	0.0152
	x ln(emp)	0.2102	0.1780
AUX Dummy		0.0487	0.2151
ln(emp)		3.4340	1.3480
Output Share		0.9623	0.1547
Input Share		0.5162	0.2138

Notes: This table presents means and standard deviations associated with key regressors in our regression sample of 73,500 firms.

D.2.2 Endogenous Specifications

Table D7 presents endogenous specifications for key firm-level outcomes. Consistent with the existing literature, we find that increases in China's competitiveness in US output markets is associated with lower firm-level sales and employment. However, we find no statistically significant impact on the input side.

D.2.3 Additional Results and Mechanisms

Table D8 presents our estimates of the impact of input and output shocks on additional outcome variables.

D.2.4 Extensive Margin

Table D9 studies the impact of output and input shocks on firm exit. We use two definitions of firm exit. Our first definition corresponds to our definition of Census firms in the rest of the paper. Additionally, we examine the robustness of our results to the HJM definition of exit: a firm in our regression sample exits under HJM's definition if it exits under Census definition and if none of its plants in 1997 survives by 2007. We modify our regression sample to include initial manufacturing firms that did not survive from 1997 to 2007.

D.2.5 Shift-Share Standard Errors

The latest research on shift-share analyses emphasizes the importance of adjusting the standard errors to address the fact that our shock is inherently an industry-level shock, whereas our observations are at the firm level. Because our analysis features multiple shift-share shocks—we assign industry-level Chinese market share gains to firms based on both their output and input shares by industry—we

Table D7: Relationship between Chinese Import Penetration in the US and Firm Outcomes

	Sales growth			Employment growth		
	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta ChineseMktSh_f^{US,Output}$	-0.360***	-0.343***	0.521***	-0.356***	-0.334***	0.548***
	(0.110)	(0.110)	(0.162)	(0.093)	(0.093)	(0.117)
$\times Aux_{ft}$		-0.921***	-0.202		-1.225***	-0.494
		(0.322)	(0.357)		(0.303)	(0.349)
$\times ln(emp_{ft})$			-0.263***			-0.269***
			(0.055)			(0.039)
$\Delta ChineseMktSh_f^{US,Input}$	-0.724	-0.650	-1.289	-0.142	-0.116	-1.019
	(0.479)	(0.478)	(0.929)	(0.460)	(0.469)	(0.822)
$\times Aux_{ft}$		-0.926	-1.522		0.182	-0.617
		(1.147)	(1.314)		(1.173)	(1.139)
$\times ln(emp_{ft})$			0.173			0.250
			(0.262)			(0.222)
R^2	0.076	0.076	0.078	0.115	0.116	0.118

Notes: Table presents results from estimating equation (32) via OLS, but using changes in China's competitiveness in output and input markets *in the US*. Aux_f^{1997} is an indicator for whether the firm has one or more auxiliary establishments in 1997. Sales and employment growth outcomes are measured as Davis-Haltiwanger-Schuh (DHS) growth rates: $DHS_f = (x_f^{2007} - x_f^{1997})/((x_f^{2007} + x_f^{1997})/2)$. All regressions include firm-level controls for Aux_f^{1997}, $ln(emp_f^{1997})$, firm age, the log number of establishments, the share of sales in manufacturing, the share of materials in manufacturing costs, and 4-digit NAICS fixed effects. Columns (2), (3), (5), and (6) also control for the input and output shares interacted with Aux_f^{1997}, and columns (3) and (6) additionally control for the input and output shares interacted with $ln(emp_f^{1997})$. Standard errors two-way clustered by firm's primary output and input NAICS. Our regression sample contains 73,500 firms.

cannot adopt the methods proposed in Adao, Morales, and Kolesar (2019) or Borusyak, Hull, and Jaravel (2021). We therefore two-way cluster our standard errors at the level of the firm's primary output and primary input industries, and perform various Monte Carlo simulations to assess the potential concerns on inference described in those papers. While our standard errors mildly overreject placebo tests at the five percent significance level, we find that they are conservative among alternative choices of standard errors.

Our firm-level input and output shocks are similar to shift-share shocks in that they are constructed by interacting firm-industry shares with industry-level shocks before aggregating to the firm-level. Both $\Delta Output_f$ and $\Delta Input_f$ in equations (D1) and (D2) can all be re-expressed in terms of standard shift-share expressions. However, their interactions with firm-specific characteristics Aux_{ft} and $\ln(emp_{ft})$ cannot. This presents a problem for off-the-shelf recommended shift-share inference as Monte Carlo exercises below reveal.

There are two potential sources of variation in the input and output shocks. The first source comes from the shares—the firm's distribution of sales over plants, plants' distribution of sales over industries, and plants' distribution of input expenses over manufacturing inputs. These are potentially endogenous to the error term in the regression: firms that are increasing their growth of AUX employ-

Table D8: Additional Outcomes

	Pivot		Δ Manuf Share of:		Import Growth	AUX Growth:	
	Sales	Emp	Sales	Emp		Sales	Emp
Output Shock	0.152	0.138	0.097	0.083	-0.108	0.116**	0.231***
	(0.220)	(0.219)	(0.078)	(0.080)	(0.327)	(0.056)	(0.079)
Output Shock $\times Aux_f^{1997}$	-0.209**	-0.212**	-0.273**	-0.214**	-1.521***	-0.200	-0.738**
	(0.098)	(0.103)	(0.132)	(0.106)	(0.289)	(0.272)	(0.287)
Output Shock $\times ln(emp_f^{1997})$	-0.055	-0.051	-0.036*	-0.029	0.076	-0.037**	-0.071***
	(0.036)	(0.037)	(0.020)	(0.021)	(0.077)	(0.016)	(0.025)
Input Shock	-0.409	-0.380	-0.094	-0.035	1.420	-0.699***	-0.229
	(0.451)	(0.459)	(0.160)	(0.159)	(1.443)	(0.162)	(0.271)
Input Shock $\times Aux_f^{1997}$	-0.098	-0.181	-0.175	-0.184*	1.473***	0.849***	0.894
	(0.175)	(0.178)	(0.152)	(0.108)	(0.515)	(0.316)	(0.607)
Input Shock $\times ln(emp_f^{1997})$	0.101	0.092	0.016	-0.003	-0.399	0.196***	0.061
	(0.103)	(0.106)	(0.038)	(0.038)	(0.302)	(0.049)	(0.078)
R^2	0.087	0.081	0.080	0.061	0.107	0.105	0.183

Notes: Table presents results from estimating equation (32) via OLS. Aux_f^{1997} is an indicator for whether the firm has one or more auxiliary establishments in 1997. Import growth and AUX growth are measured as Davis-Haltiwanger-Schuh (DHS) growth rates: $DHS_f = (x_f^{2007} - x_f^{1997})/((x_f^{2007} + x_f^{1997})/2)$. All regressions include firm-level controls for Aux_f^{1997}, $ln(emp_f^{1997})$, firm age, the log number of establishments, the share of sales in manufacturing, the share of materials in manufacturing costs, these shares interacted with Aux_f^{1997} and $ln(emp_f^{1997})$, and 4-digit NAICS fixed effects. Standard errors two-way clustered by firm's primary output and input NAICS. Our regression sample contains 73,500 firms.

ment may have decided to specialize in certain industries or manufacturing with certain techniques that are reflected in their input shares. The second source comes from the shocks themselves—certain industries in which China gained competitiveness on the world market.

We argue that the shocks at the industry level X_j^{EU} are plausibly exogenous to the firm-level error-terms. Borusyak, Hull, and Jaravel (2019) and Adao, Kolesar, and Morales (2020) show how to do statistical inference in these settings. However, their approaches do not apply in our setting where we have four different shocks, two of which created by interacting industry-level shocks with different firm-level shares.

Monte Carlo Simulations We assess how our standard errors (two-way clustering at the main output and input industries of the firm) holds up to Monte Carlo simulations. We draw 1000 random samples of industry shocks (over 440 NAICS-X industries) with mean and variance equal to the empirical distribution of industry shocks (changes in EU Import penetration from China). We then construct firm-level shocks using the same steps described in the paper. We repeat our regression specifications using the same outcome variables and controls while iterating over the 1000 samples of firm-level shocks.

The true impact of such output and input shocks should be null, so we compute the fraction of instances the regression specification rejects the null at a given significance level. Overrejection occurs when coefficients are statistically significant in a greater share of the sample than the allowed significance level.

Table D9: Probability of Firm Exit in Response to Chinese Import Competition

	Census Definition of Firm Exit			HJM Definition of Firm Exit		
	(1)	(2)	(3)	(4)	(5)	(6)
Output Shock	0.096	0.091	0.038	0.053	0.049	-0.074
	(0.088)	(0.088)	(0.126)	(0.068)	(0.069)	(0.099)
Output Shock $\times Aux_{ft}$		0.244**	0.131		0.183*	0.036
		(0.113)	(0.128)		(0.102)	(0.094)
Output Shock $\times ln(emp_{ft})$			0.019			0.043**
			(0.027)			(0.020)
Input Shock	0.063	0.057	-0.764**	-0.046	-0.047	-0.503*
	(0.124)	(0.131)	(0.381)	(0.109)	(0.115)	(0.288)
Input Shock $\times Aux_{ft}$		0.038	-0.499**		-0.048	-0.369*
		(0.195)	(0.218)		(0.174)	(0.193)
Input Shock $\times ln(emp_{ft})$			0.240**			0.134**
			(0.093)			(0.066)
R-squared	0.117	0.117	0.118	0.105	0.105	0.106

Notes: Table presents results from estimating equation (32) via OLS. The outcome variable is a dummy for firm exit between 1997 and 2007. Aux_f^{1997} is an indicator for whether the firm has one or more auxiliary establishments in 1997. All regressions include firm-level controls for Aux_f^{1997}, $ln(emp_f^{1997})$, firm age, the log number of establishments, the share of sales in manufacturing, the share of materials in manufacturing costs, and 4-digit NAICS fixed effects. Columns (2), (3), (5), and (6) also control for the input and output shares interacted with Aux_f^{1997}, and columns (3) and (6) additionally control for the input and output shares interacted with $ln(emp_f^{1997})$. Standard errors two-way clustered by firm's primary output and input NAICS. Our regression sample of continuers and exiters across contains 168,000 firms.

We find overrejection rates of 2-3x on the input shock (both pure and interacted) and overrejection rates of only about 1.5x for the output shock. An overrejection rate of 2x would mean that 5% significance level tests would reject the (true) null hypothesis that $\beta = 0$ for approximately 10 percent of the samples simulated. We find that these overrejection rates are much lower than the up to 5-6x overrejection rates we find when we (incorrectly) use the AKM0 and AKM1 and BHJ standard errors.

Therefore, while our standard error formula biases us towards being more likely to find significant results, it does not do so at an alarmingly high rate. In particular, our simulations find that 1% significant level tests would reject the (true) null that $\beta = 0$ for approximately 5 percent of the samples simulated. Therefore, given our results all come in with p-values < 0.01, a conservative adjustment would put their simulated p-value at 0.05.

One potential reason why we are more likely to overreject (find false positives) for the input shocks than for the output shocks is that the input shocks are constructed using coarser (more aggregated) shares, so there is more correlation structure across firms based on groupings of industries. Another reason is that intrinsically manufacturing production requires only a few key types of inputs, so there is more dispersion in the output shock than in the input shock.

D.2.6 Pre-trends / Falsification Test

We also do a falsification test where we regress changes in firm outcomes between 1987 and 1997 on the same firm's shocks we use in our main regressions—output and input shocks from the change in China's import penetration into the EU from 1997 to 2007 weighted by the firm's *1997* relative output and input shares. All other control variables (age, firm size, industry fixed-effect etc) are created using the firm's 1987 characteristic. Checking the key specifications in the paper, we do not find that the China shock significantly affects the firm's MPRO employment growth in that pre-period.