Online Appendix for "Structural Change Within Versus Across Firms: Evidence from the United States" (Not for Publication)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

May 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 described in Chow et al. (2021) 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, because the Census $firmid$ consists of "$0 \,\|\, EIN$" for single units, and "$alpha \,\|\, 0000$" for multi-units, 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. A second issue with Census $firmid$s is that they may ignore information useful for some research questions. For example, changes in legal ownership status or mergers and acquisitions activity can lead to breaks in Census $firmid$s even when the firm's name and main activities are unchanged.

For these reasons, 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 et al. (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

*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.

year. Note that this approach *does not* create alternate $firmid$s to replace the Census $firmid$s, 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 et al. (2013) likely would be unsatisfactory, particularly for research questions examining how outcomes (e.g., productivity) change within firms over time, as this approach can lead to $firmid$s with 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 described in Chow et al. (2021) 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 not 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 1,000.

3. For our regression analysis in Section 3.5 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 defined as those 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 et al. (2021).

2. The Census of Auxiliaries (AUX), collected in years ending 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 Fort and Klimek (2018) as auxiliaries.

   The Fort and Klimek (2018) codes also flag a large number and portion of auxiliary establishments (equal to about 1/3 of the plants in the AUX and the Fort and Klimek (2018) 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 Fort and Klimek (2018) auxiliaries 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 sectors only in Fort and Klimek (2018) 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, which is available in the BR starting in 2002. In the RDC, the variable is available in the SSL files (which are derived from the BR for years 2002 to 2016, and then from the CBPBR starting in 2018. (These CBPBR files are also constructed from the BR and replaced the prior SSL files. RDC researchers do not have access to the underlying BR data. Updated CBPBR files may contain the NAICS_AUX variable for all available years.)

2. The in-house indicator available in certain Economic Censuses (and then only for establishments that belong to multi-unit firms). The in-house 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 in-house flag, including some 1s.
   - For CRT and CWH, there are instances in which the in-house variable is populated when MU=1. We are investigating whether these were transferred plants 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 six-digit NAICS level, or if it is classified as an AUX sector at both the three-digit NAICS and the six-digit NAICS 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 in-house 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, and 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 Support Information for Table 1

Table B7 contains the number of firms in each category reported in Table 1 in the main text.

Table B7: 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 of Change	1977	2019	Change	Share of Change
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

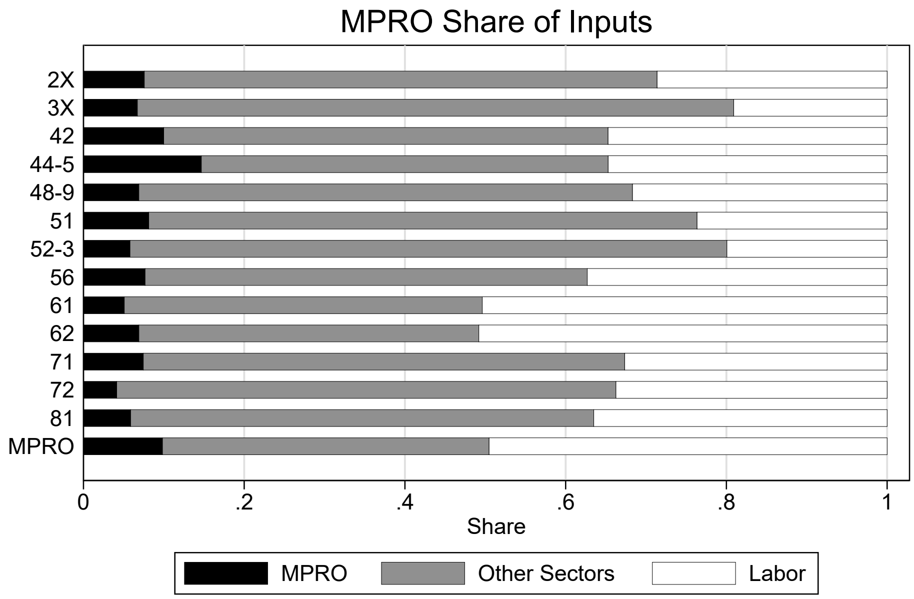
Panel B:	**"HJM Firms" (Lower Bound)**									
	Firms		Manufacturing Emp				Non-Manufacturing Emp			
	1977	2019	1977	2019	Change	Share of Change	1977	2019	Change	Share of Change
M Firms	314	273	17.7	12.1	-5.7	1.00	17.4	40.2	22.8	0.32
Continuers	51.5	46.0	10.8	7.2	-3.5	0.62	13.8	32.5	18.7	0.26
Net Birth/Death	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: 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 (top panel) are those that ever have an M plant between 1977 and 2019. HJM M firms (bottom panel) 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 C3 reports the share of each two-digit NAICS sector's inputs represented by Management (NAICS 55) and Professional, Scientific and Technical Services (NAICS 54) versus labor (BEA code V00100) and other sectors.

Figure C3: 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 PSTS (NAICS 54). Data are from the detailed 1997 US Supply-Use Table published on the BEA website.

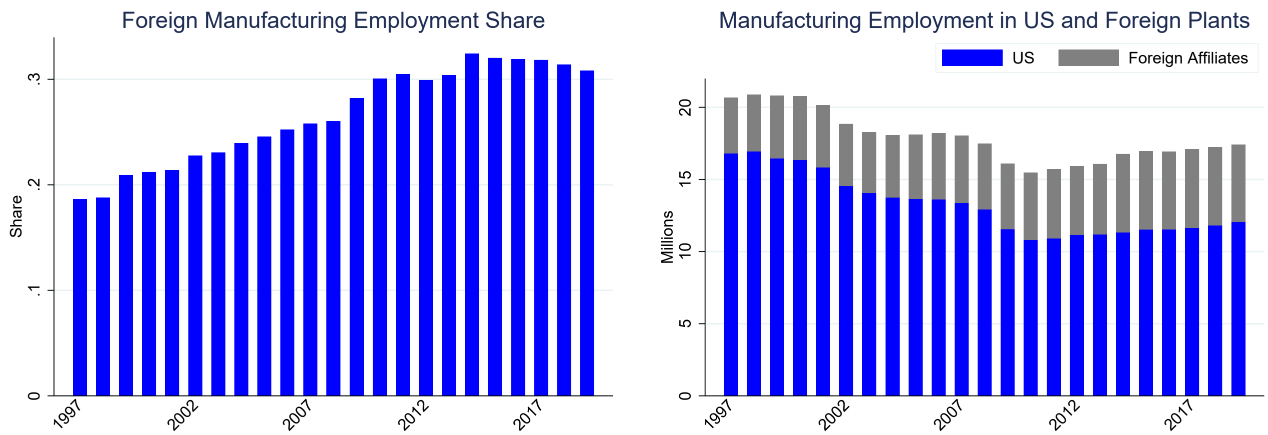
Table C8: Firms' Major NAICS Transition Matrix, 1985 to 2018

Major NAICS Sector in 1985	Major NAICS Sector in 2018: 20	30	42	44	48	51	52	53	54	56	60	70	81	Total
20	982	8	1		2		0.02	1				0.2		994
30	47	7,659	29			410	70	0.1	165	0.04	0.2	21		8,401
42		9	63	0.02	6			2			5			85
44	7	2		231			27	11				1		279
48	1	0			780			5		13				800
51		1				388			11	14	6			421
52	3	418				5	703	0.4						1,129
53								30						30
54		24						0.02	22		0.1			46
56		0							1	547				548
60										12	122			135
70						37	0.04	157				211		405
81		3						1					15	18
Total	1,040	8,124	92	231	788	840	800	207	199	587	134	233	15	13,291

Source: Computstat and authors' calculations. Sample is the 839 continuing firms for which major two-digit NAICS sector is available in COMPUSTAT in 1987 and 2018. Rows are firms' major NAICS sector in 1987. Columns are firms' major NAICS sector in 2018. Each cell reports continuing firms' global employment (in thousands) in 1987. NAICS industries are Mining (20); Manufacturing (30); Wholesale Trade (42); Retail Trade (44); Transportation and Warehousing (48); PSTS (50); Education (60); Accommodation and Food (70); Other Services (80).

D US Firms' Foreign Manufacturing Employment

Figure D4: Manufacturing Employment in the United States and in US Foreign Affiliates



Source: LBD, Bureau of Economic Analysis and authors' calculations. Figure depicts US manufacturing employment from the LBD and manufacturing employment in US firms' majority-owned foreign manufacturing affiliates from data from the Bureau of Economic Analysis.

E Auxiliary Employment by Sector

In this section we investigate whether the average differences in size and age for auxiliary establishments documented in Table 2 of the main text differ by industry. We present similar premia results in table E9, where we break out the auxiliary indicator by the establishments' two-digit NAICS sector. The estimates indicate that auxiliary wage differentials are particularly high in Warehousing (NAICS 49) and Professional, Scientific, and Technical Services (NAICS 54).[¶] They are absent from Transportation (NAICS 48), in which trucking is the predominant auxiliary activity.[‖]

[¶] The only six-digit NAICS industry with auxiliary establishments in Warehousing is General Warehousing and Storage (493110), for which the EC forms inquire about supply chain management, a potentially high-skill activity.

[‖] In additional results not reported here, we investigate the robustness of these results in two ways. First, we confirm that the results in Table 2 of the main text are insensitive to controlling for firm employment (rather than the analogous firm-level measure of the dependent variable), to limiting the sample to multi-unit firms, and to examining the wage premia by four-digit NAICS 51 and 54. We do not find a statistically significant estimate for wage premia in Legal Services (NAICS 5411). By contrast, Telecommunications (517) and Data Processing (518) both exhibit positive and significant wage premia. Second, we assess whether in-house manufacturing establishments display similar characteristics as these in-house "knowledge" plants. In contrast to the results we document here, manufacturing plants that provide inputs to other establishments of their firm are larger, but do not pay higher wages relative to non-in-house plants in the same detailed industries.

Table E9: Auxiliary Establishment Premia

	$ln(emp_{ijt})$	$ln(sales_{ijt})$	$ln(wage_{ijt})$
	(1)	(2)	(3)
Aux-48	-0.231**	-0.862***	-0.262
	(0.102)	(0.304)	(0.255)
Aux-49	-0.331***	-0.198***	0.059***
	(0.061)	(0.062)	(0.016)
Aux-51	0.002	-0.2	0.034
	(0.119)	(0.138)	(0.030)
Aux-54	-0.054***	-0.172***	0.075***
	(0.020)	(0.018)	(0.010)
Aux-56	-0.110***	-0.112***	0.093***
	(0.032)	(0.031)	(0.012)
Aux-81	0.097	-0.033	0.036**
	(0.068)	(0.030)	(0.015)
Adj. R-Squared	0.84	0.86	0.95
Observations	4,389	4,389	4,389

Source: EC and author's calculations. Table presents results from estimating equation (1) via OLS. Dependent variable is the log of employment, sales, or wage for establishment i, in industry j, and year t, as indicated in columns. $Auxiliary_{ijt}$ is an indicator for whether the establishment primarily serves other establishments in its firm. "$Auxiliary_{ijt}$ in:" denotes the two-digit NAICS sector of an auxiliary: 48 - Transportation, 49 - Warehousing, 51 - Information, 54 - Professional, 56 - Administrative, 81 - Repair. Sample limited to six-digit NAICS industries with auxiliary establishments. All regressions include six-digit NAICS, FIPs, and Year fixed effects. Firm Controls are firm age categories (births, 1-4, 5-9, 10-19, and 20+), the log number of establishments, and the firm-level counterpart of the dependent variable. Standard errors clustered by firm.

Table E10: Average Auxiliary Employment Among Firms with Auxiliaries, by Sector

	All	By two-digit NAICS						
	Auxes	48	49	51	54	55	56	81
Mean	0.173	0.001	0.012	0.001	0.012	0.138	0.007	0.002
Standard Deviation	0.191	0.022	0.063	0.025	0.072	0.173	0.058	0.034

Source: EC and author's calculations. Table presents the mean and standard deviation of firms' employment shares in auxiliary establishments for 14.2 million observations of continuing firms in each decade from 1977 to 2007. Statistics calculated for firm-year observations in which the firm has at least one auxiliary.

In Table E11, we interact the auxiliary dummy variable with the firms' auxiliary employment shares in each of the six sectors in which these establishments appear. These shares capture the distribution of firm employment across all of its auxiliaries and sum to one. As indicated in the table, the increased pivoting associated with auxiliary employment is driven by auxiliary employment in Professional, Scientific and Technical Services (NAICS 54), Management (55), and Warehousing (49). The industry descriptions for these sectors, and our results with respect to wages above, suggest that these are particularly high-skill or technology-intensive sectors.**

Table E11: Auxiliaries and Firm Outcomes

	$\Delta ln(emp_{ft})$	$\Delta ln(sales_{ft})$	$Pivot_{ft}$
	(1)	(2)	(3)
$Auxiliary_{ft}$	-0.049***	-0.096***	-0.017***
	(0.009)	(0.011)	(0.003)
$AuxEmpShare_{ft}$:			
Transportation	0.315*	0.002	-0.156***
(NAICS 48)	(0.167)	(0.251)	(0.057)
Warehousing	0.458***	0.693***	0.227***
(NAICS 49)	(0.107)	(0.131)	(0.038)
Information	0.478	1.227**	0.048
(NAICS 51)	(0.382)	(0.580)	(0.112)
Professional	0.446***	0.566***	0.156***
(NAICS 54)	(0.086)	(0.102)	(0.036)
Management	0.475***	0.926***	0.218***
(NAICS 55)	(0.059)	(0.076)	(0.019)
Administrative	0.666***	0.627***	0.047
(NAICS 56)	(0.106)	(0.110)	(0.043)
Repair	0.338***	0.503***	0.002
(NAICS 81)	(0.107)	(0.120)	(0.047)
Adj. R-Squared	0.09	0.06	0.11
Observations (M)	3.9	3.9	3.9

Source: Comtrade, Bureau of Economic Analysis, EC and author's calculations. Table presents results from estimating equation (2) via OLS. Dependent variables in columns (1) to (4) are the log change of firm employment or sales by decade. $Pivot_{ft}$ measures the share of employment in $t+1$ in the same industries as the firm's employment in t. $Auxiliary_{ft}$ is an indicator for whether the firm has an auxiliary establishment. $AuxEmpShare_{ft}$ is the firm's share of employment in auxiliary establishments. For firms with an auxiliary, the mean and standard deviation of their auxiliary employment share are 0.17 and 0.19, respectively. All regressions include fixed effects for the firm's main four-digit NAICS (by employment) and year, and control for the log of firm employment, firm age categories, and the log number of establishments, and the firm's share of non-auxiliary employment across two-digit NAICS sectors. Standard errors clustered by firm. Sample consists of 3.9 million observations of continuing firms in each decade from 1977 to 2007.

** As discussed in Footnote ¶, Warehousing includes supply-chain management.

F Shock construction

This section describes how we construct the firm-level output and input shocks, documents variation in these shock measures, and details additional specifications and checks relating to standard errors.

F.1 Industry Classifications

In this section of the paper we work with an industry classification that allows us to concord across different trade data sources and our firm-level dataset. Given our interest in firm-level outcomes, our definition of industries only matter to the extent that they provide identifying variation. We start with the NAICS six-digit (NAICS-6) industry classification in 1997, the initial year in our long difference.

We construct one industry classification for output shocks (referred to here as NAICS-X) and another slightly coarsened classification for input shocks (referred to here as NAICS-B). There are 440 NAICS-X and 330 NAICS-B. At the most disaggregated level, we divide the manufacturing sector into 440 industries. NAICS-X are near-identical to NAICS-6 codes but with some of the last digits aggregated to render them concordable with six-digit the HS codes used to track trade flows. To construct input shocks, we further coarsen NAICS-X to NAICS-B in a similar fashion to render them concordable with BEA industry codes. NAICS-B are the most disaggregated level at which each NAICS-X is entirely subsumed within a NAICS-B. Our replication provides further details.

Given that many input purchases in the materials trailer are defined at the three-, four-, or five-digit NAICS level , we 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 four-digit NAICS is given the shock associated with that particular four-digit manufacturing sub-sector.

F.2 Industry-Level Shocks

Industry 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 instrument, 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.[††]

To isolate the portion of increased trade attributable to China's productivity growth and trade liberalization with the United States between 1997 and 2007, we focus on changes in China's market share gains in the EU, which we measure as China's exports to the EU divided by total EU imports (excluding its US imports). We measure China's US import penetration as US imports from China

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

in industry j divided by total US imports plus US sales minus US exports in j:

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

Data on EU imports by origin and US imports are from Comtrade. We concord HS-level Comtrade data (1996 vintage) to NAICS as noted in the previous section. Data on US sales and US exports are from the CMF. The difference in these market shares between 1997 and 2007 measure changes in market competition from China. We define the output shock for industry j, $\Delta Output_j^{EU}$, as

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

and our endogenous measure of China's import penetration in the US as

$$\Delta ChinaImpPen_j^{US,Output} \equiv ChinaImpPen_{j,2007}^{US} - ChinaImpPen_{j,1997}^{US}.$$

Industry 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 $\Delta Input_j$ in each industry j as

$$\Delta Input_j \equiv \sum_{k \in \mathcal{J}_T} \lambda_{kj} \Delta Output_k^{EU},$$

and the endogenous measure of input cost changes using China's import penetration in the US:

$$\Delta ChinaImpPen_j^{US,Input} \equiv \sum_{k \in \mathcal{J}_T} \lambda_{kj} \Delta ChinaImpPen_j^{US,Output},$$

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} (at purchaser values, which includes retail wholesale margins on prices of inputs paid) are from the BEA's 1997 IO tables.

F.3 Firm-Level Shocks

We use cross-industry variation in the CMF's material trailer (MT) and product trailer (PT) files to create weighted average input and output shocks for each firm f. For single-unit firms, the shock is just the shock constructed from material and product trailer information associated with the one plant. For multi-unit firms, we compute the analogous weighted average for each plant, and then take the weighted average across all plants using plant-level sales as weights.

Plant-level Output Shocks: The output shock for a plant is simply a weighted average over industry output shocks in each *manufacturing* product produced by the plant. PT data are available for all plants in the CMF, so each manufacturing plant produces at least one manufacturing industry,

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

and similarly for the endogenous measure of US import penetration for the plant's outputs,

$$\Delta ChinaImpPen_p^{US,Output} \equiv \sum_{j \in \mathcal{J}_{mnf}} s_{pj} \Delta ChinaImpPen_j^{US,Output},$$

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

Plant-level Input Shocks: Some plants have useful information from the MT, i.e., line items that match up with industry codes up to the three-digit level. Other plants do not. We measure how much of the plant's total cost of materials and parts (variable *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:

$$\Delta Input_p \equiv \lambda_p^* \sum_{k \in \mathcal{J}} \lambda_{pk} \Delta Output_k,$$

and similarly for the endogenous measure of US import penetration among the plant's inputs

$$\Delta ChinaImpPen_p^{US,Input} \equiv \lambda_p^* \sum_{k \in \mathcal{J}} \lambda_{pk} \Delta ChinaImpPen_k^{US,Output},$$

where λ_{pk} denote the plant's spending on material k (line-item expense variable *mc*) 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 (variables *cm* + *ww*).

(B) **PT Sales Information:** For a subset of plants, total discernible MT line item expenses are less than 50 percent of total material costs. For these plants, we construct the input cost shock as the average input shock over the plant's industries:

$$\Delta Input_p \equiv \lambda_p^* \sum_{j \in \mathcal{J}} s_{pj} \Delta Input_j,$$

and similarly for the endogenous measure of US import penetration among the plant's inputs

$$\Delta ChinaImpPen_p^{US,Input} \equiv \lambda_p^* \sum_{j \in \mathcal{J}} s_{pj} \Delta ChinaImpPen_j^{US,Input},$$

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 (variables $cp/(cm{+}ww)$) come from the CMF and not the MT, so it is available for every plant).

Finally, since we define plant-level input shocks using shares that do not sum to one, we control additionally for a firm-level average of the scaling factor λ_p^* in our regressions:

$$\lambda_f^* \equiv \eta_f \sum_{p \in Mnf} s_{fp} \lambda_p^*. \tag{F1}$$

We define our firm-level output shock measure as

$$\Delta Output_f \equiv \eta_f \sum_{p \in Mnf} s_{fp} \Delta Output_p, \tag{F2}$$

and our firm-level input shock measure as:

$$\Delta Input_f \equiv \eta_f \sum_{p \in Mnf} s_{fp} \Delta Input_p, \tag{F3}$$

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 (import penetration) in US output and input markets. We define these equivalently as

$$\Delta ChinaImpPen_j^{US,Output} \equiv \eta_f \sum_{p \in Mnf} s_{fp} \Delta ChinaImpPen_p^{US,Output}, \tag{F4}$$

and

$$\Delta ChinaImpPen_f^{US,Input} \equiv \eta_f \sum_{p \in Mnf} s_{fp} \Delta ChinaImpPen_p^{US,Input}. \tag{F5}$$

Our output and input shocks are effectively constructed using "incomplete" shares that sum to λ_f^* and η_f. We include λ_f^* and η_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.

G China Shock Regressions

G.1 Regression Sample Statistics

Table G12 decomposes total non-manufacturing (NM) employment into employment in Management and Professional, Scientific, and Technical Services (NAICS 54 and 55) versus other NM.

Table G12: NM Employment in Regression Sample Relative to Economy Totals

	PSTS Employment				Other NM Employment			
	1997		Δ 1997-2007		1997		Δ 1997-2007	
	Level	Share	Level	Share	Level	Share	Level	Share
Firms in Regression Sample	1.78	0.20	0.40	0.23	6.97	0.09	3.49	0.20
Firms without Auxiliaries	0.02	0.00	0.12	0.07	0.27	0.00	0.29	0.02
Firms with Auxiliaries	1.75	0.20	0.28	0.16	6.70	0.09	3.21	0.18
Firms Outside Regression Sample	7.07	0.80	1.33	0.77	70.97	0.91	14.27	0.80
Economy Total	8.85	1.00	1.73	1.00	77.94	1.00	17.76	1.00

Source: Economic Census and author's calculations. Table presents Management and Professional, Scientific, and Technical Services (PSTS) and non-manufacturing (NM) employment levels and shares in 1997 and changes from 1997 to 2007 for firms in the regression sample, by their auxiliary status. Regression sample contains 73,500 continuing firms with M employment in 1997, of which 3,600 have an auxiliary establishment. Administrative Records from the Census of Manufactures are excluded from the regression sample since all their sales and input purchases are imputed. Employment is in millions.

G.2 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 G13 presents summary statistics for key regressors.

G.3 Endogenous Specifications

Table G14 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.

G.4 First-Stage Regressions

Table G15 shows that Chinese market share gains in the EU in a firm's inputs and outputs predict increased import penetration in the United States in those measures. See Online Appendix F for how we construct each firm's output and input-based Chinese market share changes in the US.

Table G13: Regression Firm Sample Summary Statistics

	Mean	Standard Deviation
Output Shock	0.1438	0.0904
Output Shock $\times Aux_f^{1997}$	0.0037	0.0242
Output Shock $\times ln(Emp_f^{1997})$	0.4699	0.3476
Input Shock	0.0614	0.0425
Input Shock $\times Aux_f^{1997}$	0.0024	0.0152
Input Shock $\times ln(Emp_f^{1997})$	0.2102	0.1780
Aux_f^{1997}	0.0487	0.2151
$ln(Emp_f^{1997})$	3.4340	1.3480
Output Share, η_f	0.9623	0.1547
Input Share, λ_f^*	0.5162	0.2138

Source: Comtrade, Bureau of Economic Analysis, Economic Census and author's calculations. This table presents means and standard deviations associated with key regressors in our regression sample of 73,500 firms. We also present statistics on η_f, the firm's share of sales in manufacturing, and λ_f^*, the firm's share of total variable costs on materials and parts.

Table G15: First Stage Relationship between Chinese Market Shares in EU and US Import Penetration

Dependent variable is change is Chinese Import Penetration in the US, with relevant interactions

	$\Delta ChinaImpPen_f^{US,Output}$				$\Delta ChinaImpPen_f^{US,Input}$			
Interactions:	None	None	$\times Aux_f^{1997}$	$\times ln(Emp_f)$	None	None	$\times Aux_f^{1997}$	$\times ln(Emp_f^{1997})$
Output Shock	0.346***	0.246***	-0.011***	-0.818***	0.028***	0.038***	0.001	-0.029
	(0.056)	(0.061)	(0.003)	(0.214)	(0.010)	(0.011)	(0.001)	(0.028)
$\times Aux_f^{1997}$		-0.059**	0.511***	-0.075		0.022***	0.079***	0.132***
		(0.029)	(0.057)	(0.166)		(0.006)	(0.023)	(0.049)
$\times ln(Emp_f^{1997})$		0.031***	0.002***	0.593***		-0.003*	-0.001***	0.035**
		(0.009)	(0.001)	(0.073)		(0.002)	(0.000)	(0.015)
Input Shock	0.136***	0.263***	0.002	0.394	0.259***	0.194***	-0.015**	-0.209**
	(0.047)	(0.086)	(0.007)	(0.441)	(0.064)	(0.061)	(0.006)	(0.095)
$\times Aux_f^{1997}$		0.061*	0.048	0.058		-0.054**	0.179***	-0.248*
		(0.034)	(0.050)	(0.216)		(0.022)	(0.063)	(0.134)
$\times ln(Emp_f^{1997})$		-0.037**	0.000	0.000		0.018***	0.005***	0.325***
		(0.016)	(0.002)	(0.125)		(0.005)	(0.002)	(0.065)
R^2	0.644	0.645	0.639	0.654	0.665	0.667	0.757	0.700
F-stat	18.54	40.54	115.6	38.06	25.60	168.9	121.8	74.76

Source: Comtrade, Bureau of Economic Analysis, EC and author's calculations. 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. The left panel presents results for changes in Chinese import penetration in the US in a firm's outputs. The right panel presents results for the equivalent change in Chinese import penetration in the US in a firm's inputs. 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 four-digit NAICS fixed effects. Standard errors two-way clustered by firm's primary output and input NAICS. Regression sample contains 73,500 firms.

Table G14: Relationship between Chinese Import Penetration in the US and Firm Outcomes

	Sales growth			Employment growth		
	(1)	(2)	(3)	(4)	(5)	(6)
$\Delta ChinaImpPen_f^{US,Output}$	-0.360***	-0.343***	0.521***	-0.356***	-0.334***	0.548***
	(0.110)	(0.110)	(0.162)	(0.093)	(0.093)	(0.117)
$\times Aux_f^{1997}$		-0.921***	-0.202		-1.225***	-0.494
		(0.322)	(0.357)		(0.303)	(0.349)
$\times ln(Emp_f^{1997})$			-0.263***			-0.269***
			(0.055)			(0.039)
$\Delta ChinaImpPen_f^{US,Input}$	-0.724	-0.650	-1.289	-0.142	-0.116	-1.019
	(0.479)	(0.478)	(0.929)	(0.460)	(0.469)	(0.822)
$\times Aux_f^{1997}$		-0.926	-1.522		0.182	-0.617
		(1.147)	(1.314)		(1.173)	(1.139)
$\times ln(Emp_f^{1997})$			0.173			0.250
			(0.262)			(0.222)
R^2	0.076	0.076	0.078	0.115	0.116	0.118

Source: Comtrade, Bureau of Economic Analysis, EC and author's calculations. Table presents results from estimating equation (32) via OLS, but using changes in China's import penetration 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 four-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.

G.5 Shift-Share Standard Errors

The latest research on shift-share analyses emphasizes the importance of adjusting 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 cannot adopt the methods proposed in Adão et al. (2019) or Borusyak et al. (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 over-reject placebo tests at the five percent significance level, we find that they are conservative compared to 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 (F2) and (F3) can all be re-expressed in terms of standard shift-share expressions.

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 auxiliary employment 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 our industry-level shocks $\Delta Output_j^{EU}$ are plausibly exogenous to the firm-level error terms. Borusyak et al. (2021) and Adão et al. (2019) show how to do statistical inference in these settings. However, their approaches cannot be applied in our setting where we have up to six different shocks regressors, four of which are created by interacting industry-level shocks with different firm-level characteristics (Aux_f^{1997}, or $ln(Emp_f^{1997})$).

Monte Carlo Simulations We conduct Monte Carlo simulations to assess the validity of our standard errors (two-way clustering at the main output and input industries of the firm). We draw 1000 random placebo samples of industry shocks (changes in EU Import penetration from China in each of the 440 NAICS-X industries) with mean and variance equal to the empirical distribution of industry shocks. For each sample we construct firm-level shocks using the same steps described in the main text. We repeat our regression specifications using the same outcome variables and controls in each of the 1000 samples.

The true impact of such output and input shocks should be null. We compute the fraction of simulated samples where our two-way clustered standard errors reject the null at a given significance level. Over-rejection occurs when coefficients are statistically significant in a greater share of the sample than the allowed significance level.

We find over-rejection 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 implies that 5% significance level tests reject the (true) null hypothesis that $\beta = 0$ for approximately 10 percent of the samples simulated. We find that these over-rejection rates are much lower than the up to 5-6x over-rejection rates we find when we (incorrectly) use the AKM0 and AKM1 and BHJ standard errors.

While our standard error formula still biases us towards statistical significance, we find that it is the most conservative (i.e., least-biased) among alternatives. Our simulations suggest that 1% significance level tests would reject the (true) null that $\beta = 0$ for approximately 5 percent of the samples simulated. Therefore one rule-of-thumb adjustment that can be applied on our p-values is to multiply them by 5. For example, any coefficient that we estimate with a p-value of < 0.01 will still retain a p-value of < 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 manufacturing production requires only a few key types of inputs, so there is more cross-industry dispersion in the output shock than in the input shock.