Appendix (Not for Publication)

Appendix A Additional Tables

Table AP1 Robustness checks: alternative instrument sets

	(1)	(2)	(3)	(4)
Variables	RD^a		RD^b	
IMP^{hat}	0.0187***	0.0118***	0.0130***	0.0082***
	(49.65)	(24.17)	(38.77)	(19.11)
EXP^{hat}	0.0045***	0.0040***	0.0014***	0.0014***
	(20.49)	(14.10)	(8.18)	(7.14)
firm-level controls	no	yes	no	yes
fixed effect	yes	yes	yes	yes
time effect	yes	yes	yes	yes
Observations	9,042	5,930	9,077	5,959
chi-square statistic	1177.59	1008.67	1097.41	947.69

Notes: *import-EX* and *export-EX* are excluded from the instrument set. IMP^{hat} (EXP^{hat}) represents predicted import (export) intensity from the first-stage. RD^a (RD^b) represents R&D intensity as measured by the ratio of R&D investment to the sales value (total assets). IMP^a (IMP^b), EXP^a (EXP^b) are used in the first-stage regression for RD^a (RD^b). Firm-level controls include the firm's age, size, capital intensity, leverage, collateral, and pre-sample TFP. Controls for fixed effects include a full set of 3-digit industry dummies, ownership dummies, and province dummies. The time effect is controlled for by using the year dummies. All standard errors are clustered at the city level; t-statistics are in parentheses. *, **, and *** represent the significance level of 10%, 5%, and 1%, respectively. The Chi-square statistic is the testing result for the significance of the entire model.

Table AP2 Robustness checks: clustered standard errors at province level

Variables	RD^a		RD^b	
IMP^{hat}	0.0053***	0.0027***	0.0045***	0.0007***
	(42.95)	(20.43)	(41.78)	(5.66)
EXP^{hat}	0.0000	0.0027***	0.0006***	0.0024***
	(0.05)	(23.21)	(5.43)	(22.63)
firm-level controls	no	yes	no	yes
fixed effect	yes	yes	yes	yes
time effect	yes	yes	yes	yes
Observations	9,042	5,930	9,077	5,959
chi-square statistic	1155.60	1000.62	1079.22	943.00

Notes: IMP^{hat} (EXP^{hat}) represents the predicted import (export) intensity from the first-stage. RD^a (RD^b) represents the R&D intensity measured by the ratio of R&D investment to the sales value (total asset). IMP^a (IMP^b), EXP^a (EXP^b) are used in the first-stage regression for RD^a (RD^b). Firm-level controls include the firm's age, size, capital intensity, leverage, collateral, and pre-sample TFP. Controls for the fixed effect include a full set of 3-digit industry dummies, ownership dummies, province dummies and year dummies. Time effect is controlled using the year dummies. All standard errors are clustered at the city level; t-statistics are in parentheses. *, **, and *** represent the significance level of 10%, 5%, 1%, respectively. Chi-square statistic is the testing result for the significance of the whole model.

Table AP3 Robustness checks: clustered standard errors at industry-year level

Variables	RD^a		RD^b	
IMP^{hat}	0.0053***	0.0027***	0.0045***	0.0007*
	(16.26)	(6.06)	(14.84)	(1.84)
EXP^{hat}	0.0000	0.0027***	0.0006***	0.0024***
	(0.02)	(11.28)	(4.09)	(15.61)
firm-level controls	no	yes	no	yes
fixed effect	yes	yes	yes	yes
time effect	yes	yes	yes	yes
Observations	9,042	5,930	9,077	5,959
chi-square statistic	1155.60	1000.62	1079.22	943.00

Notes: IMP^{hat} (EXP^{hat}) represents predicted import (export) intensity from the first-stage. RD^a (RD^b) represents R&D intensity measured by the ratio of R&D investment to the sales value (total assets). IMP^a (IMP^b), EXP^a (EXP^b) are used in the first-stage regression for RD^a (RD^b). Firm-level controls include the firm's age, size, capital intensity, leverage, collateral, and pre-sample TFP. Controls for fixed effects include a full set of 3-digit industry dummies, ownership dummies, and province dummies. The time effect is controlled for by using the year dummies. All standard errors are clustered at the city level; t-statistics are in parentheses. *, **, and *** represent the significance level of 10%, 5%, and 1%, respectively. The Chi-square statistic is the testing result for the significance of the entire model.

Appendix B Data

This appendix presents our data-analysis procedures and discusses the representativeness of our estimation sample.

B.1 Data Source

This study uses two main sources of firm-level data. The first database is the Chinese Manufacturing Firms Data (CMFD), which is compiled by China's National Bureau of Statistics (NBS) covering the period 2000-2006. This dataset includes manufacturing firms belonging to SOEs and non-SOEs with annual sales of no less than five million *Renminbi* (equivalent to about 700,000 US dollars). These firms account for 98% of manufacturing exports. The NBS of China requires firms to report details about their operations and financial statistics, such as firm sales, exporting value, employment, and total assets, among others. In total, the dataset includes more than 100 financial variables listed on the major accounting sheets of all these firms. Most importantly, it contains information on firms' annual R&D expenditures,[1] which, when combined with annual sales, can be used to construct the R&D intensity variable to proxy the innovation intensity.

Even though the CMFD contains rich firm-level information, it does not provide detailed information about firms' participation in international trade. This leads us to employ the Chinese Customs Trade Data (CCTD), which is collected by the General Administration of Customs of China. This dataset contains detailed information about all the monthly merchandise transactions passing through Chinese customs from 2000 to 2006. The information includes firm identifiers, 8-digit HS product codes, customs regimes (ordinary trade, processing trade, or other trade), transaction quantity, value, source and destination country, and ways of transporting. We add the monthly data to determine the yearly data for every recorded firm.

B.2 Data processing

B.2.1 Data cleaning before merging

Our data-cleaning process focuses on the CMFD dataset. One purpose of cleaning the data is to exclude observations with missing values for major variables. Specifically, we eliminate the observations with missing values for sales, R&D spending, number of employees, and total production. The other purpose of this elimination is to avoid the possible bias caused by noisy information. A few observations may be misleading due to misreporting. Therefore, we adopt the following standards to rule out the possible outliers. Firstly, variables that are normally positively valued cannot be negative in the dataset. These variables include total assets, sales, employment, export value, total intermediate inputs, and R&D spending. Secondly, the number of employees at a firm must be no fewer than eight. Additionally, we follow Feenstra et al. (2014) in eliminating observations if they violate any of the following criteria: (i) a firm's identification number must exist and be unique over the sample period; (ii) total assets must be larger than total fixed assets and greater than net fixed assets; and (iii) the established time of the firm must be valid. The final CMFD used in our investigation contains 589,853 firms with 1,082,985 observations from 2000 to 2006.

B.2.2 Data merging

[1] Information on R&D expenditures is available from 2001-2006.

To perform an econometric analysis, we combine the CMFD and CCTS to identify the import status of each firm in the CMFD dataset. The CCTS is merged with the CMFD for the periods from 2000 to 2006. The validity of our estimation relies on the merging efficiency between the two datasets. On one hand, this matching process should equip the merged firm with accurate information about imports and exports. On the other hand, the matching process should maximize the number of matched observations. We keep these two rules in mind to ensure the full usage of the valid information in our data.

The procedures and outcomes of the matching process between these two datasets differ in the existing literature (Upward et al., 2013; Feng et al., 2012; Wang and Yu, 2012). We synthesize the existing procedures to improve the matching efficiency, which boils down to the following steps: (i) before the merging process, we use the approach[2] developed by Ahn et al. (2011) to exclude trading intermediaries before matching[3]; (ii) we have four rounds of matching[4]: in the first round, we merge the two datasets by firm name[5]; in the second round, we match the firms by their post-codes and the names of the declarants. A firm in the CCTS is merged with a firm in the CMFD if the firm's post-code matches the CMFD and the declarant name matches the firm's name. In the third round, we merge the datasets by post-code and telephone number; in the last round, we use the post-code and address of each company as the identifiers for matching; and finally, (iii) we perform a robustness check: we eliminate trading intermediaries using the method proposed by Upward *et al.* (2013). A summary of the number of firms included in each step is provided below in Table B1.

Table B1 here

Note that a majority of the matches originates from the first round of matching. However, the subsequent matching algorithm produces supplementary matches that may have been dropped if we had only used the firm name to match. The robustness check only finds 445 firms in the matched data; this justifies the validity of the matching procedures. We present our matching outcomes from the perspectives of imports and exports in Table B2 to Table B4.

This appendix displays a series of tables showing the efficiency of our matching procedure from the respect of imports and exports. Table B2 shows the summary statistics of trading intermediaries in both the importer and exporter groups. In Table B3, we present the matching efficiency in terms of the percentage of matched observations and total observations in the CCTS dataset. The efficiency is evaluated according to both imports and exports. In Table B4, we present the matching efficiency from the angles of the shares of matched importing and export values in the CCTS dataset.

Table B2 here
Table B3 here

---

[2] This approach identifies the set of trading intermediaries based on Chinese characters that have the English-equivalent meaning of "importer," "exporter," and/or "trading" in their names. Specifically, firms with names including Chinese phrases, including "jin4chu1kou3," "jing1mao4," "mao4yi4," "ke1mao4," and "wai1jing1," are grouped into trading intermediaries.

[3] Although it is pointed out that the firms from the merged dataset should be firms trading directly (Upward et al., 2013), we do find merged firms that have characteristics of trading intermediaries. This procedure serves as means of ensuring the efficiency.

[4] This follows largely from Feng et al. (2012).

[5] Each firm has a unique registration code. We cannot simply merge the data according to it, however, since the firm is coded according to different coding systems depending on the dataset.

Table B4 here

B.3 Elimination of data

B.3.1 Exclusion of processing firms

In this paper, we focus on the effect of imported intermediate inputs on innovation. Much literature has emphasized the significant difference between processing-trade firms and ordinary-trade firms. The distinctions between processing-trade firms and ordinary-trade firms are well documented (Koopman et al., 2008; Manova and Yu, 2012; Yu, 2015; Fernandes and Tang, 2015). In our calculation, the export value of processing firms accounts for between one-fifth and all total exports. To accurately estimate the effect of imported intermediate inputs on R&D intensity, we exclude processing-trade firms from our estimation. To this end, we use information about the transactions of intermediates' imports. According to the concordance table of the BEC code to the HS 2002 table, we keep all the import records of intermediate inputs. For every import and/or export transaction, Chinese customs records the trade regime. The recorded trade regimes include ordinary trade and processing trade (import-and-assembly and pure-assembly), among others.

In Table B5, we list five approaches used in eliminating processing firms. In our benchmark analyses, we employ the most widely used rule labeled as Type 4; that is, firms whose largest shares of imports are processing trade are classified as processing firms. Other robust rules are 40%, 60%, 80%, and 100%, respectively.[6]

Table B5 here

B.3.2 Exclusion of imported intermediates

Our classification standard is based on the Broad Economic Classification (hereafter BEC). The specifics of classification are reported in Table B6 below. Since products are classified according to their 8-digit HS codes in the CCTD, we first transform the 8-digit HS codes into 6-digit HS codes. We then employ the HS-BEC table to select intermediate goods and capital goods, respectively, from each firm's bundle of imported products. Our sample period covers 2002-2005; we therefore employ the HS-BEC table for 2002 version to transform HS codes into BEC codes. The HS-BEC table is posted on the UN website, http://unstats.un.org/unsd/cr/registry/regcst.asp?Cl=10&Lg=1, and the correspondence table is downloadable at http://unstats.un.org/unsd/cr/registry/regot.asp?Lg=1.

Table B6 here

B.3.2 Exclusion of non-manufacturing sectors

The CMFD data set has a 2-digit industry code for each firm; the correspondence table is displayed in Table B7. We show in Column 3 of Table B5 the industry codes of high-tech industries based on China's NBS standards.

Table B7 here

B.4 Representativeness of the estimation sample

[6] For example, if the import share of processing trade is no less than 40%, we define the firm as a processing-trade firm. Otherwise, the firm is considered a non-processing firm.

In this part, we discuss the representativeness of our estimation sample. In Table B8 we display the efficiency of our data processing. Panel A reports the imports value during our sample period for samples 1 to 4; panel B reports the ratios between different samples. We find that the total imports value of our estimation sample accounts for around 30% of the matched database.

Table B8 here

Appendix C

The appendix defines and summarizes key variables used in our analysis. Definitions of all key variables are provided in Table C1 and Table C2. Table C1 summarizes the instruments for our estimation sample, and Table C2 summarizes the firm-level controls.

Table C1 here
Table C2 here

References

Ahn, J., A. K. Khandelwal, and S. J. Wei (2011). The role of intermediaries in facilitating trade. *Journal of International Economics 84* (1), 73-85.

Feenstra, R. C., Z. Li, and M. Yu (2014). Exports and credit constraints under incomplete information: Theory and evidence from China. *Review of Economics and Statistics 96* (4), 729-744.

Feng, L., Z. Li, and D. L. Swenson (2012). The connection between imported intermediate inputs and exports: Evidence from Chinese firms. Working Paper 18260, National Bureau of Economic Research.

Fernandes, A. P., and H. Tang (2015). Scale, scope, and trade dynamics of export processing plants. *Economics Letters 133*, 68-72.

Koopman, R., Z. Wang, and S. J. Wei (2008). How much of Chinese exports is really made in China? Assessing domestic value-added when processing trade is pervasive. Working Paper 14109, National Bureau of Economic Research.

Manova, K., and Z. Yu (2012). Firms and credit constraints along the global value chain: processing trade in China. Working Paper 18561, National Bureau of Economic Research.

Upward, R., Z. Wang, and J. Zheng (2013). Weighing China's export basket: The domestic content and technology intensity of Chinese exports. *Journal of Comparative Economics 41* (2), 527-543.

Wang, Z., and Z. Yu (2012). Trading Partners, Traded Products and Firm Performances of China's Exporter-Importers: Does Processing Trade Make a Difference? *The World Economy 35* (12), 1795-1824.

Yu, M. (2015). Processing trade, tariff reductions and firm productivity: evidence from Chinese firms. *The Economic Journal 125* (585), 943-988.

Table B1 Data matching process and efficiency

	Matching process	Matched observations
Deletion of trade intermediaries	total pre-matching	314825
	intermediaries pre-matching	58186
Matching procedure	By name	92849
	By post code and declarant	12222
	By post code and telephone number	3597
	By post code and address	925
Robustness check	Deletion of trading intermediaries	(-)445
Final matches		109148

Table B2 Total observations of matched importers and exporters

	Importers	Exporters
Total obs.	633145	741383
Obs. of trading intermediaries	101228	128950
Obs. of non-trading intermediaries	531917	612433
Matched obs.	310760	379891

Table B3 Matched observations for importers and exporters

	Importers			Exporters		
year	Obs.	Matched Obs.	percentage	Obs.	Matched Obs.	Percentage
2000	62,789	31,952	50.89%	62,771	33,898	54.00%
2001	67,588	35,986	53.24%	68,487	38,877	56.77%
2002	77,303	41,165	53.25%	78,612	45,893	58.38%
2003	87,934	46,178	52.51%	95,688	54,747	57.21%
2004	102,242	50,791	49.68%	120,590	64,140	53.19%
2005	113,454	52,376	46.16%	144,030	69,763	48.44%
2006	121,835	52,312	42.94%	171,205	72,573	42.39%
Total	633,145	310,760	49.08%	741,383	379,891	51.24%

Table B4 Trading value of matched importers and exporters by year

year	Importing value	Import values of matches	percentage	Exporting value	Export values of matches	percentage
2000	2251	1080	47.98%	2492	1212	48.64%
2001	2661	1295	48.67%	2906	1506	51.82%
2002	2952	1533	51.93%	3256	1799	55.25%
2003	4131	2207	53.43%	4385	2572	58.65%
2004	5608	2997	53.44%	5936	3605	60.73%
2005	6571	3446	52.44%	7567	4605	60.86%
2006	7883	3982	50.51%	9685	5751	59.38%

Note: the unit value is 100 million US dollars.

Table B5 Summary of importers of intermediate inputs

year	CMFD-CCTD merged sample	Estimation sample: without importers of processing trade				
		type4 (biggest share)	40%	60%	80%	100%
2001	19,858	8,009	6,186	6,563	7,092	11,154
2002	22,469	9,786	7,554	8,013	8,664	13,451
2003	25,783	11,939	9,406	9,996	10,826	16,392
2004	35,830	16,827	13,255	14,106	15,274	22,681
2005	36,487	17,270	13,970	14,822	16,035	23,544
2006	38,155	18,811	15,187	16,130	17,420	25,008
Total	178,582	82,642	65,558	69,630	75,311	112,230

Table B6 BEC classifications of intermediates

		Sum of Categories
Intermediate goods	111	Food and beverages, primary, mainly for industry
	121	Food and beverages, processed, mainly for industry
	21	Industrial supplies not elsewhere specified, primary
	22	Industrial supplies not elsewhere specified, processed
	31	Fuels and lubricants, primary
	322	Fuels and lubricants, processed (other than motor spirit)
	42	Parts and accessories of capital goods (except transport equipment)
	53	Parts and accessories of transport equipment

Table B8 Sample representatives

Panel A						
	2001	2002	2003	2004	2005	2006
1.intermediate imports in CCTD	192.1	212.5	296.2	405	488.4	585.7
2.intermediate Imports in CCTD-CIAS merged	77.82	88.5	128	186.8	214.3	252.4
3.intermediate Imports in estimation sample	72.52	84.36	122.9	175.9	204.3	240.5
4. intermediate imports in estimation sample without processing trade firms	24.48	23.19	38.55	53.81	59.89	72.6
Panel B						
ratio1=2/1	40.51%	41.65%	43.21%	46.12%	43.88%	43.09%
ratio2=3/1	37.75%	39.70%	41.49%	43.43%	41.83%	41.06%
ratio3=4/1	12.74%	10.91%	13.01%	13.29%	12.26%	12.40%
ratio4=3/2	93.19%	95.32%	96.02%	94.16%	95.33%	95.29%
ratio5=4/2	31.46%	26.20%	30.12%	28.81%	27.95%	28.76%

Note: The Unit in panel A is billion US dollars.

Table B7 Chinese 2-digit industry code and high-tech industries

2-digit code	industry	high-tech
13	Processing of Food from Agricultural Products	
14	Manufacture of Foods	
15	Manufacture of Liquor, Beverages and Refined Tea	
16	Manufacture of Tobacco	
17	Manufacture of Textile	
18	Manufacture of Textile, Wearing Apparel and Accessories	
19	Manufacture of Leather, Fur, Feather and Related Products and Footwear	
20	Processing of Timber, Manufacture of Wood, Bamboo, Rattan, Palm and Straw Products	
21	Manufacture of Furniture	
22	Manufacture of Paper and Paper Products	
23	Printing and Reproduction of Recording Media	
24	Manufacture of Articles for Culture, Education, Arts and Crafts, Sport and Entertainment Activities	
25	Processing of Petroleum, Coking and Processing of Nuclear Fuel	253
26	Manufacture of Raw Chemical Materials and Chemical Products	2665
27	Manufacture of Medicines	27
28	Manufacture of Chemical Fibers	
29	Manufacture of Rubber and Plastics Products	
30	Manufacture of Non-metallic Mineral Products	
31	Smelting and Pressing of Ferrous Metals	
32	Smelting and Pressing of Non-ferrous Metals	
33	Manufacture of Metal Products	
34	Manufacture of General Purpose Machinery	
35	Manufacture of Special Purpose Machinery	
36	Manufacture of Automobiles	368
37	Manufacture of Railway, Ship, Aerospace and Other Transport Equipments	376
38	Manufacture of Electrical Machinery and Apparatus	
39	Manufacture of Computers, Communication and Other Electronic Equipment	
40	Manufacture of Measuring Instruments and Machinery	40
41	Manufacture of Measuring Instruments and Machinery	411,412,4141, 4154,4155,4190
42	Utilization of Waste Resources	
43	Repair Service of Metal Products, Machinery and Equipment	

Note: The classification of high-tech industries is based on China Statistics Yearbook on High Technology Industry (2003-2006) by China's National Bureau of Statistics.

Table C1 Summary statistics of instruments: estimation sample

Variable	Mean	Std. Dev	Min	Max
log(tariff)	.0356	.0449	0	.7608
log(import-EX)	.568	.3258	0	2.0480
log(WES)	11.282	2.386	0	18.042
log(export-EX)	.623	.2915	0	2.646
log(WID)	10.370	2.245	0	17.256

Table C2 Summary statistics of firm-level controls: estimation sample

Variable	Mean	Std. Dev	Min	Max
age	15.425	8.602	6	58
Size [log(employee)]	5.165	1.107	2.565	7.859
Log(k/l)	4.141	1.238	-.755	6.623
leverage	.5465	.234	.017	1.383
collateral	.288	.1770	.0060	.863
tfp00	4.409	1.040	-.926	8.932