Breaking Borders: The Impact of Knowledge Diffusion on the Gains from Trade

Qianxue Zhang*
University of Warwick
January 2024

Job market paper – see the latest version here

Abstract

How does international knowledge sharing affect trade patterns, economic growth, and welfare across the globe? This paper answers this question by estimating a novel dynamic trade model where heterogeneous firms innovate. I first document that knowledge diffusion and technology adoption are two key channels for acquiring foreign knowledge using comprehensive Chinese firm-level data on trade, patents, and citations. Based on these findings, I develop a dynamic general equilibrium model where firms learn from sellers when importing and choose to adopt foreign technologies when exporting. In the model, diffusion enhances productivity for all firms, and adoption further amplifies these gains by boosting the productivity of the most efficient firms. I structurally estimate the model with bilateral trade flows for the global economy. I find that knowledge diffusion substantially increases the gains from trade in all economies, ranging from 0.2% to 8.7%. However, foreign technology adoption can reduce welfare in knowledge-abundant countries as their technological advantages get eroded. Technology adoption therefore alters the conventional gains from trade, with developed countries potentially benefiting from higher trade barriers.

Keywords: Knowledge Diffusion, Technology Adoption, General Equilibrium, Trade Barriers
JEL classification: F12, F43, O33, O41

---

*Department of Economics, University of Warwick, Coventry CV4 7AL, UK. Email: qianxue.zhang@warwick.ac.uk.
I am very grateful to my supervisors, Roger Farmer and Dennis Novy, for their invaluable help and guidance. I thank Mirko Draca, James Fenske, Yuchen Lin, Marta Santamaria, David Torun and Marija Vukotic and participants of the Warwick Macro/Trade Workshop, 16th RGS Conference, 2023 CES North America Conference, 21st GEP-CEPR Postgraduate Conference, 12th UEA Conference, 5th QMUL PhD Workshop, Warwick-Oxford Macro Workshop, 11th Warwick Economics PhD Conference, 1st JIE Summer School, 3rd ASPEC Conference, 2023 NBER Innovation Research Bootcamp, 2023 EEA-ESEM-Congress and ETSG Surrey for useful suggestions.

1 Introduction

Sharing ideas, technologies, and knowledge through international trade is crucial for fostering innovation and economic growth. Previous work has explored the role of knowledge diffusion on various economic outcomes at the aggregate level (Buera and Oberfield, 2020; Cai, Li and Santacreu, 2022a), while others show that heterogeneous firms adopt foreign technologies to reduce production costs (Vishwasrao and Bosshardt, 2001). In this paper, I argue that heterogeneous firms learn foreign knowledge in two different ways — knowledge diffusion and technology adoption. The interaction between these two channels significantly amplifies the gains from trade, underscoring the importance of studying them together rather than in isolation.

This paper shows the importance of knowledge diffusion and technology adoption using a novel dynamic trade model where heterogeneous firms innovate. Leveraging comprehensive Chinese firm-level data on trade, patents, and citations, I first document that importers rely primarily on diffusion to access foreign knowledge, whereas exporters primarily on adoption. Then I develop a dynamic general equilibrium model where importers learn from sellers and exporters choose to adopt foreign technologies. I structurally estimate the model with bilateral trade flows for the global economy and I show that ignoring the interaction between the two channels leads to a substantial underestimation of the gains from trade. Furthermore, higher trade barriers attempting to maintain technological advantages in certain countries can potentially benefit them at the cost of reducing global welfare.

I build a comprehensive dataset using Chinese firm-level data from 1998 to 2007, collected from various sources, to document four stylized facts. (1) Firms that engage in international trade are more productive and file more patent applications than non-trading firms. (2) More productive and more innovative exporters sort into trade with countries that have higher patent stocks. (3) Innovative and productive exporters pay foreign technology adoption fees, which amount to approximately 60% of their total R&D expenditure. (4) Import liberalization leads firms to innovate more and cite more foreign patents. Measuring knowledge diffusion through citation, innovative importers cite more patents from the country they import more from.

Then I develop a new dynamic monopolistic trade model with knowledge diffusion and technology adoption, consistent with the stylized facts above, to quantify their impacts on the gains from trade. In my model, knowledge about how to produce goods is diffused through international trade, which leads to innovation and efficiency gains. Firms learn how to increase productivity by importing foreign goods. This is similar to Buera and Oberfield (2020), but I focus on how firms with heterogeneous productivity are differently exposed to international trade. Higher productivity, in turn, drives more firms to venture into exporting. Exporters can adopt foreign technology to further reduce their marginal production costs, leading to higher profits (similar to Bustos 2011). In addition, I assume the benefit of adoption is determined by a weighted sum of home and foreign countries' knowledge stocks, which is a parsimonious way to capture the sorting of exporters' trading partners (stylized fact 2). This assumption means that trade margins respond to trade liberalization events differentially depending on who the firms trade with. Moreover, in my dynamic framework, a one-time adjustment in trade costs can have a persistent effect on the gains from trade due to the change in the speed of knowledge accumulation and the number of new exporters.

This multi-country, multi-sector dynamic model generates predictions that are in line with the stylized facts presented above during trade liberalization. A reduction in trade costs leads to a higher speed of knowledge accumulation in the liberalizing country, which increases total sales as well as trade margins.[1] A novel prediction is that the magnitude of the increase depends on the knowledge gap between the liberalizing country and its trading partners. I provide reduced-form evidence consistent with these predictions, using China's WTO accession as an exogenous shock for Chinese manufacturing firms. Joining the WTO reduced China's import tariffs as well as trade policy uncertainty due to the permanent normalization of trade relations (PNTR) granted by the United States (Handley and Limão, 2017). Using a difference-in-differences (DiD) setting, I find that WTO accession increased Chinese firms' new patent applications and both the intensive and extensive margins are positively correlated with the number of new patent applications. Furthermore, the correlation between patent and trade margins is notably stronger when the trading partner is a patent-rich country. As the knowledge gaps between China and these countries are larger, the higher correlations are consistent with the model predictions.

Although trade liberalization can be beneficial for low-knowledge-stock coun-

[1] Here trade margins are the extensive margin (the number of exporters) and the intensive margin (sales per exporter).

tries, it may harm high-knowledge-stock countries due to the erosion of their technological advantages. I illustrate this in a symmetric two-country case, with the only difference being the initial knowledge stocks of the two countries. When both countries experience a 10% reduction in bilateral variable trade costs or fixed costs, three new results appear in general equilibrium. First, when diffusion and adoption are both at work, trade liberalization increases total sales in both countries, but the increase is higher for the country with a lower initial knowledge stock. However, separating diffusion and adoption apart, diffusion alone would lower the gains for the country with high knowledge stocks. This is because, without the chance of adoption, the high-stock country misses the potential benefits of adopting technologies when the low-stock country catches up. Second, welfare gains, measured in real income, are amplified by the knowledge diffusion and technology adoption in both countries and the long-run effects are substantial, compared to the short-run effects. Third, when the initial knowledge gap is sufficiently large, the country with a higher stock may experience a reduction in welfare even if diffusion and adoption are both at work, due to the faster erosion of its technological advantage.

To quantify the general equilibrium effects of knowledge diffusion and technology adoption across the globe, I group all countries in the world into 21 economies (accounting for more than 98% of trade) and the rest of the world (ROW)[2] and I bring the model to the data. I divide the model parameters into two groups. The first group includes parameters that appeared in previous literature and the second group includes others that are new in this paper. I calibrate the parameters in the first group following the previous literature, including sectoral linkages as in Caliendo and Parro (2015), the elasticity of substitution across varieties as in Antras et al. (2017) and the firm productivity distribution as in Shapiro and Walker (2018). The parameters for the diffusion function are new and have to be estimated sequentially. I follow the strategy of Buera and Oberfield (2020) to first estimate the knowledge stock for each country-sector pair and then estimate the parameters governing the diffusion function via the Simulated Methods of Moments (SMM).

The sectoral knowledge stocks in each economy are backed out via the structural model, using the observed bilateral trade flow data, the value-added data, the nominal and real GDP, and the calibrated sectoral linkages. After getting the

[2]Due to data availability, I only consider Australia, Austria, Belgium, Brazil, Canada, Switzerland, China, Germany, Denmark, Spain, Finland, France, the United Kingdom, Italy, Japan, Korea, the Netherlands, Norway, Sweden, Taiwan, and the United States separately.

knowledge stocks, the parameters governing the knowledge diffusion and the adoption functions can be estimated using SMM. For the diffusion function, the targeted moments are the mean and standard deviation of the knowledge stocks. For the adoption function, I use the mean of bilateral trade shares as the target, as the shares are closely related to knowledge gaps across economies. The model performs well in matching targeted and non-targeted moments. Moreover, the model successfully generates a positive correlation between knowledge stocks and empirical patent stocks, even though estimating the former doesn't require information on the latter.

As the next step, I solve the full dynamic model using dynamic hat algebra (Caliendo et al., 2019) and conduct three counterfactuals to highlight the importance of knowledge diffusion and technology adoption for the gains from trade. The first counterfactual compares the welfare gains with and without knowledge diffusion or technology adoption from 2000 to 2007. I find that global welfare would decrease by 3.6% if both channels were absent.

In a model without diffusion, all economies would have lower welfare. In general, countries that have lower initial knowledge stocks and that are more involved in global value chains benefit more from knowledge diffusion. Moreover, diffusion reduces knowledge gaps among economies, triggering a continuous reduction in the exporting productivity threshold, and allowing more firms to export.

By contrast, technology adoption does not affect all economies in the same way. Although the global welfare would be lower without adoption, certain countries would benefit. For instance, Japan, Norway, and the United States would enjoy approximately 0.9% higher welfare if exporters cannot adopt foreign technologies. These countries have higher technological advantages and thus can maintain their dominance without foreign exporters adopting their technologies.

As anticipated, the gains from adoption and diffusion do not add up linearly when both mechanisms operate simultaneously. This suggests a nuanced interaction between these two mechanisms: Diffusion increases the knowledge stocks in each economy and thus enables more firms to overcome the fixed costs to become exporters. As the number of exporters increases, more firms benefit from technology adoption and reduce their production costs. This further increases welfare. Quantitatively, adoption amplifies the welfare gain from diffusion by about 70%. This suggests that ignoring adoption can lead to an overestimation of the importance of diffusion in models attributing all gains to the latter. Thus, to provide an accurate evaluation of welfare changes and design proper trade policies, it is imperative to consider both effects, recognizing their interconnected nature in shaping the outcomes of international trade.

In the second counterfactual, I explore the repercussions of higher trade costs on welfare by assuming that China never joined the WTO. To simulate this counterfactual scenario, I assume a 5 pp increase in China's bilateral trade costs with all other economies in 2002 (Erten and Leight, 2021). I compare how welfare would change without China's WTO accession in four cases: a world with both diffusion and adoption, a world without either, a world with only diffusion and a world with only adoption. I find China's (global) welfare would be 17.15% (2.49%) lower in a world without either channel. However, the losses would be limited to 10.76% (2.18%) lower if both adoption and diffusion are at work. Considering only diffusion or adoption leads to similar results that learning helps to mitigate the losses from higher trade barriers.

The final counterfactual draws inspiration from the recent US-China trade war and examines the consequences of a unilateral 25% tariff increase imposed by the US, as measured by Egger and Zhu (2020), specifically targeting China. In this context, global welfare would decrease by 0.41% in a standard model without either learning channel. However, adding both channels leads to a 0.51% loss. These results underscore how diffusion and adoption can magnify the losses stemming from the trade dispute, with a similar impact observed whether we focus on either channel individually. The reductions in global welfare are in line with the bystander effect in Fajgelbaum et al. (2021) where the repercussions of trade disputes extend beyond the directly involved parties.

This paper contributes to five strands of literature. The first is the large literature on firm-level trade, innovation and growth, such as Lileeva and Trefler (2010), Aw et al. (2011), Bloom et al. (2016), Aghion et al. (2022). I differ by studying the general equilibrium effects of knowledge diffusion and adoption with firm heterogeneity. Additionally, I extend Bustos (2011)'s static framework into a dynamic setting. In her paper, trade margins remain constant after the shock, whereas in my model, the changes in trade margins are contingent upon the evolving knowledge stocks and can differ over time.

The second strand is the recent literature on trade models with knowledge diffusion. There has been much work on innovation, diffusion and growth but so far in a Ricardian context at the aggregate level (Buera and Oberfield, 2020; Cai, Caliendo, Parro and Xiang, 2022b; Cai, Li and Santacreu, 2022a; Lind and Ramondo, 2022, 2023). Relative to these studies, I introduce firm heterogeneity and add the technology adoption channel. I show that these new features significantly change the trade patterns and the gains from trade.

My third contribution is to the large literature on the gains from trade. Many papers quantify the gains from trade through different channels (e.g., Broda and Weinstein 2006; Edmond et al. 2015; Fajgelbaum and Khandelwal 2016). Within this domain, a branch of literature focuses on understanding dynamic gains from trade (Impullitti and Licandro, 2018; Ravikumar and Sposi, 2019). Sampson (2016) and Perla et al. (2021) both examine the impact of technology diffusion on growth and welfare by endogenous growth models but concentrate on diffusion within the home country. My work complements theirs by focusing on the dynamic gains from international knowledge diffusion. This broadens our understanding of how knowledge diffusion across borders influences the gains from trade.

This paper is related to a fourth strand of literature on the impact of trade liberalization on various economic outcomes such as Dix-Carneiro (2014) and Dix-Carneiro and Kovak (2017) on the labor market, Topalova and Khandelwal (2011) on firm productivity, Caliendo and Parro (2015) on welfare, Baldwin and Forslid (2010) on trade volume and variety, and Shu and Steinwender (2019) on innovation. More specifically, some papers study how WTO accession affects Chinese firms' productivity, markup dispersion and innovation (e.g., Yu 2015; Lu and Yu 2015; Brandt et al. 2017; Chen et al. 2017; Liu and Ma 2020; Liu et al. 2021). My work differs from these studies as it specifically examines the dynamic impact of trade liberalization on knowledge accumulation and how different learning channels amplify the gains from trade.

Related, my work also contributes to the literature studying how China's entry into the WTO has impacted the world economy. There is extensive literature on the China shock and how it has affected labor markets in other countries (e.g., David et al. 2013; Asquith et al. 2019; Caliendo et al. 2019). I complement these papers by using a general equilibrium framework to study the role of knowledge diffusion and adoption in impacting global welfare rather than focusing on individual labor markets within specific countries.

The rest of the paper is organized as follows. In Section 2, I provide four stylized facts on international trade, knowledge diffusion, and technology adoption using Chinese firm-level data from 1998 to 2007. In Section 3, I develop a dynamic monopolistic trade model with diffusion and adoption and characterize the equilibrium. In Section 4, I derive the model predictions and provide consistent evidence using China's WTO accession. I also highlight the importance of the general equilibrium effects in a two-country setting. In Section 5, I solve the model using dynamic hat algebra. I calibrate the parameters in Section 6 and evaluate the model's fit to the data. Section 7 further illustrates the importance of knowledge diffusion using three counterfactuals. Section 8 concludes.

2 Motivating facts on trade, knowledge diffusion and technology adoption

In this section, I document four stylized facts about international trade, knowledge diffusion and technology adoption that form the basis for the theoretical model in Section 3. I show that diffusion and adoption coexist in the data and their significance differs for different Chinese firms. I start by introducing the data I collected from various sources and then present the findings.

2.1 Data

I use four main firm-level datasets for the exercise, covering data from 1998 to 2007: (1) balance sheet data from the Annual Survey of Industrial Firms Database (ASIF), (2) customs data from the General Administration of Customs of the People's Republic of China (GACC), (3) patent data from He et al. (2018), where their original data is from China's State Intellectual Property Administration (SIPO) and (4) citation data obtained from the PatSnap platform. I also use two aggregate-level datasets. I use foreign countries' patent stock from the United States Patent and Trademark Office (USPTO) and foreign technology adoption fees for Chinese firms from China Science and Technology Statistics Yearbooks. Here I describe these datasets briefly, and in the empirical appendix C I provide a detailed description of all datasets, the merging processes and additional evidence.

Firm-level balance sheet data: ASIF is published annually by the National Bureau of Statistics of China and contains all industrial firms with sales above 5 million RMB before 2012. These firms comprise more than 90% of total industrial output and 97% of industrial exports in 2004 (Brandt et al., 2012). As the most comprehensive database for Chinese manufacturing firms, it includes basic firm-level information (such as a unique identifier, name, address, telephone, ownership structure, and sector), production information (such as the number of employees, total sales, value-added, intermediate inputs) and balance sheet information (such as assets, capital stock, expenditure, etc.). The database has been widely used in the literature (e.g. Hu et al. 2005; Hsieh and Klenow 2009; Song et al. 2011; Huang et al. 2013).

Customs data: The customs data is from the General Administration of Customs of the People's Republic of China, starting from 2000. The monthly transaction data contains information on firm identifiers, HS product codes, quantities, values, modes of transportation and trading partners etc. The data is aggregated into annual frequency for each firm. On average, 16% of firms engaged in international trade and the number of firms increased steadily (see Table C.5).

Firm-level patent data: The patent data is obtained from He et al. (2018), where their original data is from China's State Intellectual Property Administration. The SIPO patent database covers all published patent applications since 1985. There are three types of patents, design, invention and utility. The authors remove all patents assigned to individuals or firms outside China. The former condition is met only when the patent's inventor is also the assignee, and the assignee field does not contain any designators of the corporate form, while the latter requires the assignee to be a firm. I merge this dataset with the ASIF database and I exclude design patents for the empirical results.

Citation Data: As I measure knowledge diffusion through patent citations, I collect data on the foreign patents that were cited by Chinese firms from PatSanp. PatSnap is a global patent-searching platform covering 170 countries and regions. The data sources are generally from the National Intellectual Property Administration. PatSnap contains the universe of Chinese patents with detailed information on titles, abstracts, applicants, and International Patent Classification (IPC) and Co-operative Patent Classification (CPC) codes. I track the total number and countries of foreign patents that each Chinese firm cites for a certain patent application.

Trading partners' patent stock: I obtain sectoral patent stock data from Sampat (2011), which contain all issued utility patents from the USPTO from January 1, 1975, to December 31, 2010.[3] The database includes patent numbers, application dates, first-named assignee, the primary class and subclass of each patent,

[3] The total numbers are consistent with USPTO's statistics; see USPTO for a summary.

forward and backward citations to previous US utility patents, and the number of citations to non-patent references, etc. Due to issues with data availability, I focus on 21 economies in the database and aggregate the others into the rest of the world (ROW). The full list is in Appendix A.1. Each patent belongs to a 3-digit United States Patent Classification System (USPCS) category and a 6-digit subclass. I use the concordance table provided by Goldschlag et al. (2019) to change them into an International Standard Industrial Classification of All Economic Activities (ISIC Rev 4) category[4] and then aggregate them into 21 sectors. I calculate the annual stock of patents from 1975 by adding newly granted patents each year. Figure C.1 shows the number of patents issued in each country annually.

Expenditure on foreign technologies: I collect national expenditures of large and medium-sized Chinese manufacturing firms on foreign technologies from China Science and Technology Statistics Yearbooks from 1998 to 2007. Large and medium-sized firms are firms that have more than 300 employees or annual sales of over 30 million RMB or assets of over 40 million RMB.[5] Note this data is at the national level instead of the firm level.

2.2 Four stylized facts about trading firms

I present below four stylized facts on international trade and knowledge diffusion using Chinese manufacturing firm-level data. Those facts will be the guide for the theoretical model I develop later.

2.2.1 Trade, productivity, and innovation

Stylized fact 1: *Firms that engage in international trade have higher productivity and new patent applications.*

Using ASIF data, I follow Brandt et al. (2012) to create a 10-year firm panel. I estimate firm-level TFP following the Olley and Pakes (1996) method. Value-added, employment and intermediate goods for estimation are directly obtained from the ASIF dataset. Value-added is deflated using prices provided by Brandt et al. (2012), and provincial industrial producer prices from the National Bureau of Statistics of China when the former is missing. I estimate the real capital stocks using the perpetual inventory method. I drop all firms with fewer than eight employees, negative

[4]See the Patent crosswalk.
[5]See the definition here.

production, intermediate goods, investment or real capital stock. I also conduct a robustness check using the Generalized Method of Moments (GMM) Estimation in Wooldridge (2009). Both estimations give very similar results.[6]

Figure B.1 shows the distribution of estimated TFP in 1998 and 2007. There is a notable shift in the average productivity of firms between 1998 and 2007, meaning firms were on average more productive in 2007. The results are comparable to Elliott et al. (2016) and Lu and Lian (2012). Moreover, exporters were more productive than non-exporters in all years (see Figure B.2).

Table 1: Trade, TFP and patent

	(1) lnTFP	(2) Patent
Trade dummy	0.169***	0.215***
	(0.003)	(0.059)
Observations	2,043,423	2,043,423
R-squared	0.520	0.765
Year FE	YES	YES
Firm FE	YES	YES

Notes: Standard errors are clustered at the firm level and in parentheses. *** $p < 0.01$, **$p < 0.05$, *$p < 0.1$.

After merging the ASIF database with the GACC data and patent data, I create a trade dummy for each firm in each year, with one indicating the firm engages in international trade this year and zero otherwise. I then regress the trade dummy on the firm TFP and the patent number. Table 1 shows that trading firms in China are on average 16.9% more productive than non-trading firms and submit around 21.5% more patent applications.

2.2.2 Selection of trading partners

Stylized fact 2: *More productive and more innovative exporters sort into trade with countries that have higher patent stocks.*

As GACC provides information on the trade content and trading partner's country for each transaction, I construct a knowledge exposure index by weighting each

[6] See the empirical appendix C for the details.

trading partner's sectoral patent stock by the trade share of the Chinese firm. Specifically, for each firm, I calculate the sum of the trade-weighted patent stock of its trading partners annually as

$$W_{it} = \sum_{n=1}^{N} \sum_{j=1}^{J} trade_{int}^{j} * S_{nt}^{j},$$

where $trade_{int}^{j}$ is the share of firm i's imports from (exports to) country n, sector j at t over its total imports (exports) in sector j at t and S_{nt}^{j} indicates the patent stock in sector j, country n at t. Note $\sum_{n=1}^{N} \sum_{j=1}^{J} trade_{int}^{j} = 1$ and a firm in sector j can trade with any other sector in a foreign country. Therefore, the index also takes into account the cross-sectional correlation.

As there is generally a lag between learning and its impact, I run the following specification

$$Y_{it} = \alpha_1 \ln W_{i,t-1} + \alpha_2 X_{it} + \lambda_i + \lambda_t + \epsilon_{it}, \tag{1}$$

where Y_{it} is (log) TFP or the number of new patent applications, λ_i and λ_t are firm- and year-fixed effects. X_{it} are control variables, including firm age, capital intensity (capital per employee) and financial status (debt to asset ratio). The results are presented in Table 2. Columns (1) and (4) only include firms that both export and import. For TFP, although the coefficients are both significant, the export-weighted index has a much higher coefficient than the import-weighted one. While only the export-weighted index has a significant correlation with new patent applications. In other columns, I also include firms that just import or export, I find the positive correlations are entirely driven by exports.

2.2.3 Foreign technology adoption fees paid by Chinese firms

Stylized fact 3: *Innovative and productive exporters paid foreign technology adoption fees, which amounted to approximately 60% of their total R&D expenditure from 1998 to 2007.*

China Science and Technology Statistics Yearbooks record the total expenditure by large and medium-sized manufacturing firms on foreign technology adoption. The foreign adoption fees refer to the expenses incurred by the firm during the reporting period for the purchase of foreign (or Hong Kong/Macao/Taiwan) technologies. This includes expenses related to product design, process flow, drawings,

Table 2: TFP, Patent and trade-weighted foreign patent stock

	(1) lnTFP	(2) lnTFP	(3) lnTFP	(4) Patent	(5) Patent	(6) Patent
ln W_{it-1}^{imp}	0.005***	0.0002		0.030	0.018	
	(0.002)	(0.001)		(0.050)	(0.036)	
ln W_{it-1}^{exp}	0.014***		0.012***	0.107**		0.091***
Observations	51,572	68,801	84,405	6,543	8,199	9,553
Firm FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
Controls	YES	YES	YES	YES	YES	YES

Note: Firm-level controls include age, log of capital intensity and financial status. Standard errors are clustered at the 2-digit sector level and in parentheses. *** $p< 0.01$, **$p< 0.05$, *$p< 0.1$. Columns (1) and (4) only include firms that both import and export.

formulations, patents, and other technical documentation, as well as expenses for purchasing equipment, instruments, prototypes, and samples. This data is similar to the technology adoption fees paid by the Korean firms in (Shim, 2023). Unfortunately, the data for Chinese firms is at the national level and does not contain information on small firms. However, the 2008 Economic Census Yearbook (in its Table 1-A-1) shows that large and medium-sized firms accounted for more than 88% of total expenditure on foreign technology adoption.

Figure 1 shows that the expenditure on foreign adoption fees was increasing over time and was much higher than the expenditure on purchasing domestic technologies. Table 3 compares the annual expenditure on R&D, foreign technology adoption and domestic adoption. Although the expenditure on foreign adoption was declining over time, on average, it was around 60% of domestic R&D and eight times the expenditure on domestic adoption.

Although there is no firm-level data on technology adoption, I can classify all firms in ASIF into large and medium-sized or smaller groups based on the same criteria and compare their other characteristics. I find that the large and medium-sized firms are more innovative and more productive as well as export at an average value that is 11 times greater than that of the smaller firms. Therefore, foreign technology adoption is more important for larger, more productive and exporting firms.

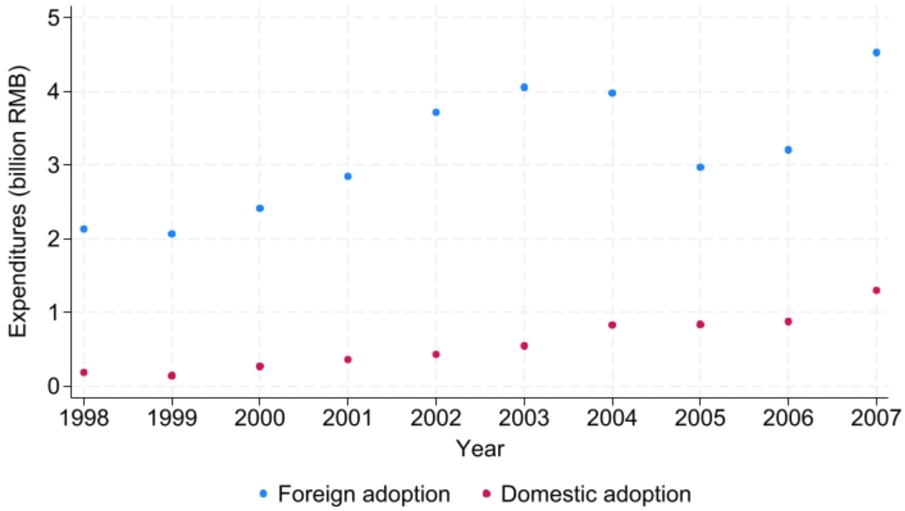

Figure 1: Foreign and domestic adoption fees

Notes: This figure shows the foreign adoption fees and domestic adoption fees paid by large and medium-sized Chinese manufacturing firms during the years 1998 to 2007. Adoption fees refer to the expenses incurred by the firm during the reporting period for the purchase of technologies.

Table 3: Foreign adoption, R&D and domestic adoption (ratios)

Year	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	Average
Foreign adoption over R&D	1.16	0.90	0.75	0.69	0.71	0.60	0.44	0.29	0.25	0.26	0.61
Foreign adoption over domestic adoption	12.60	16.35	9.98	8.42	9.28	7.97	6.04	4.39	4.60	4.31	8.39

Notes: The first row of the table shows the ratio between the expenditure of Chinese firms on foreign technology adoption and on domestic R&D from 1998 to 2007. The second row is the ratio between foreign adoption and domestic adoption in the same year.

2.2.4 Knowledge diffusion and international trade

Stylized fact 4: *Import liberalization leads firms to innovate more and cite more foreign patents. Measuring knowledge diffusion through citation, innovative importers cite more patents from the country they import more from.*

China's accession to the WTO is a well-known trade liberalization event. When China joined the WTO, it started to fulfill its tariff reduction responsibilities as a WTO member country. Figure D.1 illustrates the average import tariffs in China spanning from 1998 to 2007, notably depicting a significant decline between 2001 and 2002.[7] I explore the sector-level variation in the tariff reduction (Lu and Yu, 2015) to identify the impact of trade liberalization on firm innovation and knowledge diffusion. I use the number of new patent applications as a proxy for innovation and I measure knowledge diffusion through patent citations. PatSnap records

[7]The tariffs come from the WTO Tariff Download Facility.

all patents that a Chinese firm cites when applying for a new patent. I calculate the total foreign citations of each Chinese patent application.

To test the impact of trade liberalization on innovation and diffusion, I run the following event-study regression

$$N_{it}^{j} = \sum_{t=1998}^{t=2007} \beta_t \times D_{2001}^{j} + \lambda_i + \lambda_t + \epsilon_{it} \tag{2}$$

where N_{it}^{j} is the number of new patent applications of firm i in sector j at t or its citations of foreign patents. D_{2001}^{j} is the level of the sectoral tariff gap, i.e. tariffs in 2002 minus tariffs in 2001. λ_i and λ_t are firm and year-fixed effects. Figures 2 and 3 show that the WTO accession increased the number of new patent applications and citations of foreign patents for Chinese firms.

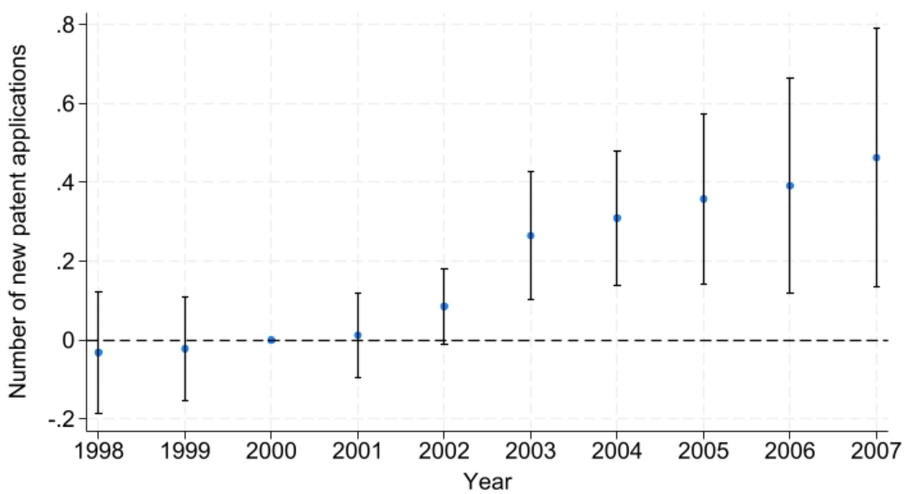

Figure 2: Tariff reduction and new patent applications

Notes: This figure shows the coefficients estimated using equation (2) with firm and year fixed effects and the 95% confidence intervals. The standard errors are clustered at the 4-digit industry level.

Then I merge these results with the trade data and compare the average citations of trading firms with non-trading firms. Figure 4 displays the positive correlations between foreign citations per patent and the log of trade value for cases where the number of foreign citations is greater than zero. It indicates firms that trade a lot also tend to cite more foreign patents.

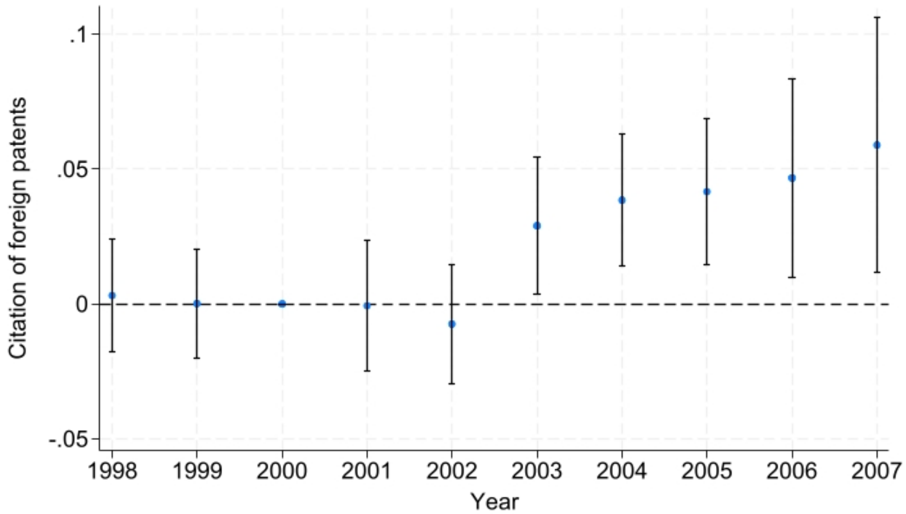

Figure 3: Tariff reduction and citation of foreign patents
Notes: This figure shows the coefficients estimated using equation (2) with firm and year fixed effects and the 95% confidence intervals. The standard errors are clustered at the 4-digit industry level.

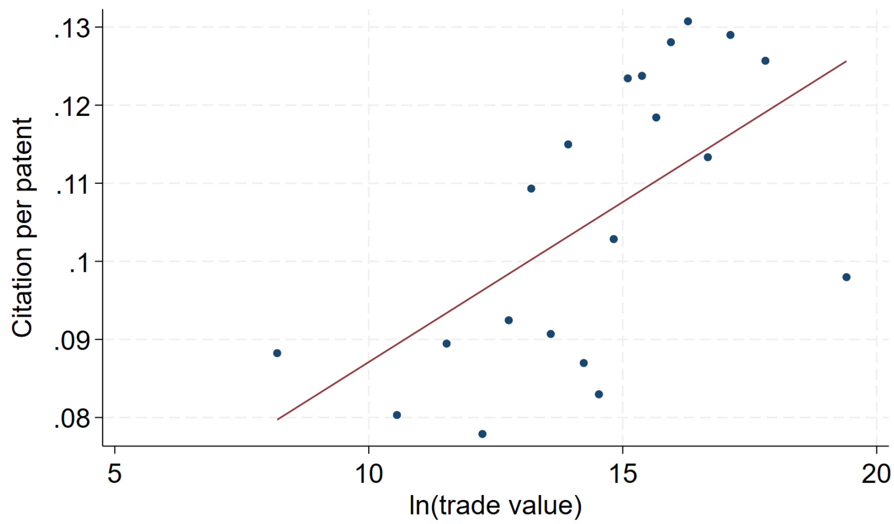

Figure 4: Foreign citations per patent and trade value
Notes: This figure shows the positive correlation between the foreign citations per number and the (log) trade value when the number of foreign citations is greater than zero. Both figures are bin plots splitting the sample into 20 bins.

Table 4 shows that importers have the highest total citations or foreign citations, followed by exporters. In both cases, trading firms cite many more patents than non-trading firms. The same pattern can be found if we only focus on firms with at least one patent application, although the gap between traders and non-traders is smaller. Moreover, I find the results are driven by firms that both export and import. For firms that are only exporters or importers, there are fewer citations and the maximum number is 15 and 17 respectively. Further separating exports and imports, I find the correlation is driven by the importers (see Table 5).

Table 4: Citation statistics

	Domestic and foreign citations			Foreign citations		
Firm type	Non-trading	Importer	Exporter	Non-trading	Importer	Exporter
Panel A: All firms						
Sample size	1903236	241527	277272	1903060	241527	277272
Mean	0.013	0.046	0.041	0.005	0.018	0.016
SD	1.529	2.511	2.351	1.364	2.116	1.975
Min	0	0	0	0	0	0
Max	1207	791	791	1168	736	736
Panel B: Firms with non-zero patent						
Sample size	47943	14248	14974	46108	13764	14455
Mean	0.534	0.776	0.752	0.197	0.320	0.305
SD	9.616	10.31	10.092	8.764	8.859	8.646
Min	0	0	0	0	0	0
Max	1207	791	791	1168	736	736

Notes: This table shows the average citation number of Chinese firms from 2000 to 2007. The first three columns are the citation of both domestic and foreign patents while the other three columns are the citations of foreign patents only. Panel A takes the average of all firms in the sample while panel B only considers firms with at least one patent. In both panels, importers and exporters are not mutually exclusive, meaning a firm can appear in both samples if it exports and imports at the same time.

Table 5: Citation per patent and trade

	Citation per patent
ln(import)	0.108***
	(0.001)
ln(export)	0.041
	(0.234)
Observations	348
R-squared	0.772
Year FE	YES
Firm FE	YES

Notes: This table shows the correlation between citation per patent application of Chinese firms and their trade values. Only the firms that both import and export are included here. Robust standard errors.

3 A model of trade, knowledge diffusion and technology adoption

I develop a dynamic general equilibrium model guided by the stylized facts above. I add two main components into an otherwise standard Melitz (2003) model. I add learning from sellers as in Buera and Oberfield (2020) to capture knowledge diffusion (stylized fact 4), although I change the setting from a Ricardian model to a monopolistic trade model to capture the firm heterogeneity (stylized fact 1). I introduce technology adoption similar to Bustos (2011) to capture stylized fact 3, where exporters can pay a fixed cost to adopt foreign technology and reduce marginal production costs.[8] I differ from her work by assuming that the reduction depends on the domestic and foreign countries' knowledge gaps. This is a parsimonious way to capture the positive correlation between productivity and the trading partner's knowledge stock in stylized fact 2. Moreover, the model features sectoral linkages as in Caliendo and Parro (2015) to account for cross-sectional knowledge exposure, which is also a key aspect of stylized fact 2. Not that in this section I always treat n as the home country and i as the exporting country.

3.1 Household problem

There are N countries and J sectors. Labor is mobile across sectors but immobile across countries. A representative household in each country maximizes utility by consuming final goods C^j_{nt} from each sector j:

$$u(C_n) = \prod_{j=1}^{J} (C^j_n)^{\alpha^j_n}, \quad \sum_{j=1}^{J} \alpha^j_n = 1. \tag{3}$$

Income consists of labor income at wage rate w_n and a share of the global fund of firm profits Π. I assume each country receives a share of profits proportional to its expenditure (Chaney, 2008).

Final consumption is a constant-elasticity-of-substitution (CES) composite of a

[8]Note exporting and importing are not mutually exclusive for firms. This assumption ensures that larger and more productive firms adopt foreign technologies. It is isomorphic to a case where firms first pay to adopt and then decide to export or not.

continuum of differentiated varieties ω

$$Q_n^j = \left[\int q_{ni}^j(\omega^j)^{1-\sigma^j} d\omega^j \right]^{\frac{1}{1-\sigma^j}}. \tag{4}$$

3.2 Intermediate goods producers

Intermediate goods $\omega^j \in [0,1]$ are produced by labor l_{nt}^j and composite intermediate goods $m_{nt}^{k,j}(\omega^j)$. Before production, intermediate goods producers draw productivity z_n^j from a Pareto distribution with CDF [9]

$$G_{z_n^j}(z) = Pr(z_n^j \leq z) = 1 - T_n^j z^{-\theta^j}.$$

Firms produce ω^j according to a constant-returns-to-scale technology

$$q_n^j(\omega^j) = z_n^j(\omega^j)[l_n^j(\omega^j)]^{\gamma_n^j} \prod_{k=1}^{J} [m_n^{k,j}(\omega^j)]^{\gamma_n^{k,j}}, \tag{5}$$

γ_n^j is the share of value-added and $1 - \gamma_n^j = \sum_{k=1}^j \gamma_n^{k,j}$. $\gamma_n^{k,j}$ is the share of the composite intermediate goods $m_n^{k,j}(\omega^j)$ from sector k for the production of ω^j.

Therefore, the unit cost of producing variety ω^j for a firm with productivity z_n^j in country n is

$$c_n^j = \eta_n^j w_n^{\gamma_n^j} \prod_{k=1}^{J} \left(P_n^k\right)^{\gamma_n^{k,j}}, \tag{6}$$

and $\eta_n^j = \prod_{k=1}^J (\gamma_n^{k,j})^{-\gamma_n^{k,j}} (\gamma_n^j)^{-\gamma_n^j}$ is a constant.

3.3 International trade

Firms can decide whether to export or not. There is a fixed cost f_{ni}^j for country i to serve the foreign market n in sector j. Moreover, shipping one unit of good from country i to country n in sector j incurs a standard 'iceberg' costs d_{ni}^j as in Samuelson (1954). Trade costs are normalized to one inside the same country and $d_{ni}^j d_{ik}^j \geq d_{nk}^j$. Firms can adopt foreign technologies when entering their markets, which reduces the marginal cost of production by ζ. Note that $\zeta \geq 1$ is assumed to be a combination of the domestic and foreign knowledge stocks (see details in

[9]Note that final goods can either be consumed or used for producing intermediate goods, $Y_{nt}^j = C_{nt}^j + \sum_{k=1}^J \int m_n^{j,k}(\omega^k) d\omega^k$.

section 3.5). Therefore, exporting q units of variety ω with productivity z_i^j from i to n costs:

$$c_{ni}^j = \frac{c_i^j d_{ni}^j}{\zeta_{ni}^j z_i^j} q_{ni}^j + f_{ni}^j. \tag{7}$$

Given the CES assumption, demand for each variety in country n and sector j is

$$x_{ni}^j(\omega^j) = \left(\frac{p_{ni}^j(\omega^j)}{P_n^j}\right)^{-\sigma^j} X_n^j \tag{8}$$

where $x_{ni}^j(\omega^j)$ is the demand for variety ω^j exported from country i to n and X_n^j is total expenditure. The profit-maximizing price p_n^j of each variety ω^j is set by firms, and P_n^j is the aggregate price

$$P_n^j = \left[\int p_{ni}^j(\omega^j)^{1-\sigma^j} d\omega^j\right]^{\frac{1}{1-\sigma^j}}. \tag{9}$$

Firms charge a constant markup over marginal costs. Thus firms in country n charge $p_{nn}^j = \frac{\sigma^j}{\sigma^j-1}\frac{c_n^j}{z_n^j}$ domestically and $p_{in}^j = \frac{\sigma^j}{\sigma^j-1}\frac{c_n^j d_{in}^j}{\zeta_{in}^j z_n^j}$ in the foreign market i.

For firms in county i exporting to country n, the zero profit condition for the marginal firm is defined by

$$s_{ni}^j(z^x) = p_{ni}^j q_{ni}^j - c_{ni}^j = 0. \tag{10}$$

Substituting $\frac{\sigma}{\sigma-1}\frac{c_i^j d_{ni}^j}{\zeta_{ni}^j z_i^j}$ for p_{ni}^j, we obtain the cut-off productivity as

$$z_{ni}^{x,j} = \frac{c_i^j d_{ni}^j}{\zeta_{ni}^j P_n^j}\left(\frac{X_n^j}{f_{ni}^j}\right)^{\frac{1}{1-\sigma^j}} \frac{1}{\sigma^j - 1}\sigma^{j\,\frac{\sigma^j}{\sigma^j-1}}. \tag{11}$$

3.4 Aggregation

Following Chaney (2008), I assume that the number of potential entrants in any country is proportional to wage income. Expenditure by country n on goods from country i equals the sum of all firm-level sales:

$$\begin{aligned} X_{ni}^j &= w_i L_i \int_{z_{ni}^{x,j}}^{\infty} p_{ni}^j q_{ni}^j dG(z) \\ &= C w_i L_i T_i^j (f_{ni}^j)^{1+\frac{\theta^j}{1-\sigma^j}} \left(\frac{c_i^j d_{ni}^j}{\zeta_{ni}^j} \right)^{-\theta^j} (P_n^j)^{\theta^j} \left(X_n^j \right)^{\frac{-\theta^j}{1-\sigma^j}} \end{aligned} \tag{12}$$

where $C = \frac{\theta^j \sigma^j}{\theta^j - \sigma^j + 1} \left(\frac{1}{\sigma^j - 1} \sigma^{j \frac{\sigma^j}{\sigma^j - 1}} \right)^{-\theta^j}$ is a constant.

Since total expenditure of country n in sector j is simply the sum of its expenditure on all trading partners, i.e. $X_n^j = \sum_{i=1}^N X_{ni}^j$, the expenditure share can be written as

$$\pi_{ni}^j = \frac{X_{ni}^j}{\sum_{i=1}^N X_{ni}^j} = \frac{w_i L_i T_i^j (f_{ni}^j)^{1-\frac{\theta^j}{\sigma^j-1}} \left(c_i^j d_{ni}^j \right)^{-\theta^j} \left(\zeta_{ni}^j \right)^{\theta^j}}{\sum_{k=1}^N w_k L_k T_k^j (f_{nk}^j)^{1-\frac{\theta^j}{\sigma^j-1}} \left(c_k^j d_{nk}^j \right)^{-\theta^j} \left(\zeta_{ni}^j \right)^{\theta^j}}. \tag{13}$$

We can calculate the aggregate price index by

$$P_n^{j\,1-\sigma^j} = \frac{\theta^j}{\theta^j - \sigma^j + 1} \sum_{i=1}^N \left(\frac{c_i^j d_{ni}^j}{\zeta_{ni}^j} \right)^{1-\sigma^j} w_i L_i T_i^j (z_{ni}^{x,j})^{\sigma^j - 1 - \theta^j} \tag{14}$$

To complete the model, we need to calculate the profits Π_n from the global fund for each country n. Defining total sales $Y_n^j = \sum_{i=1}^N X_{in}^j$, we know that gross profits earned by firms in each market are $\frac{Y_n^j}{\sigma^j}$. With producer heterogeneity, net profits are $\frac{Y_n^j(\sigma^j-1)}{\sigma^j \theta^j}$ since total entry costs are a fraction $\frac{\theta^j - \sigma^j + 1}{\theta^j}$ of gross profits. Therefore, the budget constraint becomes

$$I_n = w_n L_n + \sum_{j=1}^J \frac{\sigma^j - 1}{\sigma^j \theta^j} Y_n^j + D_n \tag{15}$$

where D_n is the trade deficit of country n.

3.5 International knowledge diffusion

I follow Buera and Oberfield (2020) to assume there is learning from foreign sellers.[10] Producers in each country of each good receive new ideas Z in a stochastic

[10] I do not consider diffusion from buyers for a couple of reasons. First, the data show that imports are the main source of diffusion. Second, using IV regressions in appendix D I show that

and exogenous manner in each period, with each idea representing a technology for producing a specific good with a productivity level of $\ln Z$. While new ideas build upon existing knowledge in the country, the adaptation of such knowledge is subject to randomness. The productivity level of a new idea upon arrival is $Z = hZ'^{\rho}, \rho \in [0, 1)$, where Z' is the knowledge drawn from another producer based on the source distribution $G_t(Z')$ and h is the original idea drawn from another exogenous distribution. It is assumed that the rate of new idea arrivals with an original component greater than h is $R_t(h)$.

Assumption 1: Between t and $t + 1$, the arrival of new ideas in each sector j follows a Poisson distribution with mean

$$R_t^j(h) = m_t^j h^{-\theta^j}.$$

Assumption 2: The initial frontier of knowledge in sector j follows an exponential distribution given by $F_0^j(Z) = e^{1-T_0^j Z^{-\theta^j}}$.

These assumptions ensure that the growth of the knowledge stock is given by [11]

$$T_{n,t+1}^j = T_{n,t}^j + m_t^j \int_0^\infty x^{\rho^j \theta^j} dG(x) = T_{n,t}^j + m_t^j \Gamma(1-\rho^j) \left[\sum_{i=1}^{N} \pi_{nit}^j \left(T_{i,t}^j \right)^{\rho^j} \right], \quad (16)$$

where $\Gamma(u) = \int_0^\infty x^{u-1} dx$. Equation (16) indicates that the growth of the knowledge stock in country n, sector j depends on the arrival rate of new ideas $m_t^j = m_0^j e^{g_m t}$ and the exposure to foreign knowledge weighted by the expenditure share π_{ni}^j. Note that the expenditure shares are endogenously determined in equilibrium. Therefore, country n learns more from country i if i has a higher knowledge stock or n spends more on i. Going back to the benefit of adopting foreign technology ζ_{ni}^j, in contrast to Bustos (2011) where it is a fixed number, I assume $\zeta_{ni}^j = \Delta^j T_{nt}^j + (1 - \Delta^j) T_{it}^j$, where $\Delta^j \in (0, 1)$. Thus the adoption of foreign technology reduces marginal costs by a combination of the two countries' knowledge stocks. This is a parsimonious way to capture the stylized fact 2 that more productive exporters sort into trade with the high-stock countries.

only imports lead to higher innovation. Third, in this framework, it is not possible to consider both cases simultaneously, and I show in section 6 that the model can generate predictions consistent with the data.

[11] See the full derivation in appendix E.7.

3.6 Equilibrium

The equilibrium consists of six equations: the production cost of each variety of intermediate goods (6), the expenditure share (13), the price index (14), and the growth of the knowledge stock (16). Moreover, the total expenditure of country n on sector j goods is given by

$$X_n^j = \sum_{k=1}^{J} \gamma_n^{j,k} \sum_{i=1}^{N} X_i^k \pi_{in}^k + \alpha_n^j (w_n L_n + D_n + \sum_{j=1}^{J} \frac{\sigma^j - 1}{\sigma^j \theta^j} Y_n^j) \tag{17}$$

with $D_n = \sum_{j=1}^{J} D_n^j$ where the deficit of sector j' is given by the difference between country n's imports from all other countries and its exports to the world, $D_n^j = \sum_{i=1}^{N} X_{ni}^j - X_{in}^j$. The national deficit is the sum of all sectoral deficits, and the sum of all countries' deficits is zero. Finally, the trade balance in country n is

$$\sum_{j=1}^{J} \sum_{i=1}^{N} X_n^j \pi_{ni}^j = \sum_{j=1}^{J} \sum_{i=1}^{N} X_i^j \pi_{in}^j + D_n. \tag{18}$$

The model is associated with a balanced growth path derived in appendix E.2 and the detrended equilibrium in appendix E.3.

4 Knowledge diffusion and dynamic trade margins

Chaney (2008) derives the intensive and extensive margins of trade and shows how changes in variable and fixed trade costs affect each margin. Here I extend this idea to a dynamic setting, where changes in trade costs impact not only current margins but also future margins through the changes in knowledge stocks. The impact on trade margins is a key dynamic adjustment mechanism leading to enhanced gains from trade in response to trade cost reductions.

4.1 How trade liberalization affects trade margins

I start by showing how the changes in bilateral variable trade cost d^j and fixed cost f^j affect the intensive (sales per exporter) and extensive (number of exporters) margins in the exporting country i. The number of exporters from i to n in sector j is calculated as $N_i = w_i L_i (1 - G(z_{ni}^{x,j})) = C_0 w_i L_i T_i^j (z_{ni}^{x,j})^{\sigma^j - 1 - \theta^j}$, where $C_0 = \frac{\theta^j}{\theta^j - \sigma^j + 1}$

is a constant. Therefore, the intensive margin equals the total exports divided by the number of exporters, which is $\sigma^j f^j_{ni}(z^{x,j}_{ni})^{1-\sigma^j}$.

Trade shocks on firm sales in the same period: Suppose there is a reduction in the variable trade cost d^j_{ni} in period t, there will be a positive effect on the total sales of exporters. Both the number of exporters and the sales per exporter increase. When there is a reduction in the fixed cost f^j_{ni}, the total sales also increase, but sales per exporter do not change.

Proposition 1: *A reduction in the bilateral variable trade cost d^j_{ni} or fixed cost f^j_{ni} increases the total sales of exporters from i to n, but only the former leads to an increase in sales per exporter.*

Proof: Same as Chaney (2008) since the knowledge level is predetermined.

However, the changes in trade costs also have intertemporal effects in this dynamic setting through changing the growth of knowledge and thus on T^j_{nt+1}. How does d^j_{ni} change the total sales from country i to n (X^j_{ni}) in the subsequent periods? The total impact can be decomposed into contemporaneous and intertemporal effects,

$$\frac{\partial X^j_{ni,t+1}}{\partial d^j_{nit}} = \underbrace{\frac{\partial X^j_{ni,t+1}}{\partial T^j_{n,t+1}}}_{\text{Contemporaneous effect}} \times \underbrace{\frac{\partial T^j_{n,t+1}}{\partial d^j_{nit}}}_{\text{Intertemporal effect}} \tag{19}$$

Starting with the second part of equation (19), T^j_{nt} is predetermined and thus does not change due to the change in d^j_{nit}. But the change in d^j_{nit} has an ambiguous effect on the knowledge stock in the next period.

Proposition 2: *A reduction in variable trade cost d^j_{ni} or fixed cost f^j_{ni} has an ambiguous impact on the knowledge stock in country n.*

Proof: *Since*

$$\frac{\partial T^j_{n,t+1}}{\partial d^j_{nit}} = \frac{\partial}{\partial d^j_{nit}} \left(T^j_{nt} + m^j_t \Gamma(1-\rho^j) \left[\sum_{i=1}^{N} \pi^j_{nit} \left(T^j_{i,t}\right)^{\rho^j} \right] \right), \tag{20}$$

a change in d^j_{nit} increases π^j_{nit} but decreases π^j_{nkt}. The overall effect depends on the weighted sum of $\left[\sum_{i=1}^{N} \frac{\partial \pi^j_{nit}}{\partial d^j_{nit}} \left(T^j_{i,t}\right)^{\rho^j}\right]$. The same results apply to the changes in fixed costs.

Corollary 1: *In a two-country example, a reduction in bilateral variable trade* *costs d^j or fixed costs f^j increases the knowledge stock in the country with a lower initial knowledge stock.*

Proof: *Since the sum of expenditure shares is one, a higher expenditure share on the foreign country means a lower share for the home country. Suppose the home country has a lower knowledge stock, then it gains from shifting its expenditure share towards the foreign country.*

Nevertheless, the impact of knowledge stock on the total sales in the same period is always positive (the first part of equation (19)),

$$\frac{\partial X^j_{ni,t+1}}{\partial T^j_{n,t+1}} = \frac{\partial(C_1 w_n L_i T^j_i f^j_{ni}(z^{x,j}_{ni})^{-\theta^j})}{\partial T^j_n} = w_i L_i f^j_{ni}(z^{x,j}_{ni})^{-\theta^j}\left(1+\frac{\theta^j\Delta^j T^j_n}{\Delta^j T^j_n+(1-\Delta^j)T^j_i}\right), \tag{21}$$

where $C_1 = \frac{\theta^j\sigma^j}{\theta^j-\sigma^j+1}$ is a constant.

Proposition 3: *An increase in the knowledge stock in country n increases the total sales of exporters in country n as well as both trade margins.*

Proof:

$$1+\frac{\theta^j\Delta^j T^j_n}{\Delta^j T^j_n+(1-\Delta^j)T^j_i} = \frac{(1+\theta^j)\Delta^j T^j_n+(1-\Delta^j)T^j_i}{\Delta^j T^j_n+(1-\Delta^j)T^j_i}$$

Since $\theta^j > 1$ and $\Delta^j \in [0,1]$, the numerator is always positive. As the extensive margin and intensive margins both decrease in $z^{x,j}_{in}$, the derivatives with respect to T^j_n are thus also positive.

4.2 Evidence consistent with the model predictions

Using Chinese firm-level data, I provide reduced-form evidence consistent with the three propositions in section 4.1. China's accession to the WTO is a well-known trade liberalization event, which reduced bilateral trade costs between China and its trading partners and reduced fixed costs resulting from policy uncertainty. When China joined the WTO, it started to fulfill its tariff reduction responsibilities as a WTO member country. Concurrently, China was granted permanent normal trade relations (PNTR) status by the United States. Before that, China only had temporary most favored nation (MFN) status, which had to be renewed every year and attracted a high level of uncertainty. Although the PNTR did not alter the actual tariff rates imposed on Chinese imports, it eliminated the uncertainty over whether

China would retain its MFN status.

In both cases, I explore the sector-level variation in the tariff reduction (Lu and Yu, 2015) and the gap between the non-NTR and NTR tariffs (Pierce and Schott, 2016) to identify the impact of trade liberalization on knowledge stocks and trade margins. The tariffs come from the WTO Tariff Download Facility and the NTR gaps are from Feenstra et al. (2002). I use the number of new patent applications as a proxy for the growth in knowledge stock. For trade margins, the intensive margin is defined as the export value of each firm and the extensive margin is defined as the probability of being an exporter.

Evidence consistent with proposition 1: To test the contemporaneous effect, I collapse the sample into two periods, one before and one after the WTO accession. I test the prediction using the following difference-in-differences (DiD) equation:

$$y_{it}^{j} = \beta Post \times G_{2001}^{j} + \lambda_i + \lambda_j + \epsilon_{it} \tag{22}$$

where y_{it}^{j} is the sales of each exporter $ln(exp_{it})$ or the number of the exporters in sector j. G_{2001}^{j} is the tariff rate of sector j in 2001 or the NTP gap of sector j in 2001. $Post$ is a dummy that equals one after the event. λ_i and λ_j are firm and sector fixed effects. The results are presented in Table 6, where columns (1)-(2) are for tariffs and (3)-(4) for the NTP gap. For the tariff reduction, the coefficients for both margins are positive and statistically significant. This suggests that both intensive and extensive margins increased more after 2002 in industries with higher tariffs in 2001 than in industries with lower tariffs. However, the coefficient for the intensive margin is not significant when the uncertainty drops. As uncertainty is more related to fixed costs, this is consistent with the theory that a change in fixed costs does not affect the sales per exporter.

Evidence consistent with proposition 2: As Figure 2 in stylized fact 4 already shows the impact of tariff reduction, I focus on the NTP gaps here. I run the same event-study regression as equation (2) but now D_{2001}^{j} is a dummy which equals one when the NTP gap is positive in 2001 and zero otherwise. Figure 2 before and Figure 5 here show that the WTO accession increased the number of new patent applications in both cases. Note this is consistent with **Corollary 1** as China had a low knowledge stock compared with other countries when China joined the WTO.

Table 6: Trade liberalization and the trade margins

	Tariff reduction		Uncertainty reduction	
	ln(exp)	number of exporters	ln(exp)	number of exporters
	(1)	(2)	(3)	(4)
$Post \times G^{j}_{2001}$	0.0417***	0.0150***	0.334	0.165***
	(0.005)	(0.001)	(0.228)	(0.007)
Observations	43,176	175,696	73,462	294,124
R-squared	0.855	0.909	0.859	0.907
Period FE	YES	YES	YES	YES
Firm FE	YES	YES	YES	YES

Notes: Standard errors are clustered at the 2-digit sector level and in parentheses. *** p< 0.01, **p< 0.05, *p< 0.1. Columns (1)-(2) are the results when using changes in tariff as the treatment and (3)-(4) are the results when using the NTP gap as the treatment.

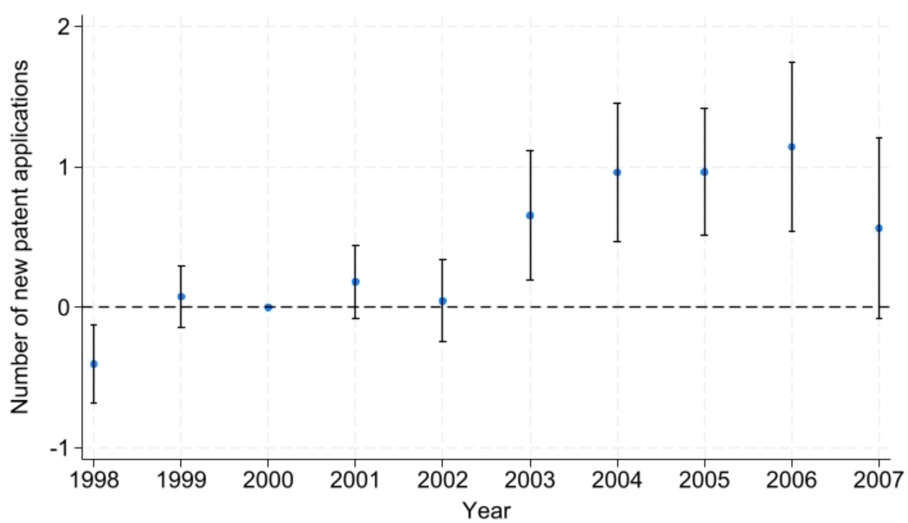

Figure 5: Uncertainty reduction and the new patent applications

Notes: This figure shows the coefficients estimated using equation (2), while changing the independent variable from sectoral tariff gaps to sectoral NTP gaps, with firm and year fixed effects and the 95% confidence intervals. The standard errors are clustered at the sector-year level.

Evidence consistent with proposition 3: Absent an exogenous shock to test the causality from knowledge stock to trad margins, I examine if the increase in the number of new patent applications correlates with the average sales per exporter and the export status of each firm. The model predicts the changes in patent stock should be positively correlated with both margins.

Specifically, the following regressions are estimated,

$$Y_{it} = \beta \ln N_{it} + \beta_1 X_{it} + \lambda_i + \lambda_t + \epsilon_{it} \tag{23}$$

where Y_{it}^{j} is the sales of each exporter (intensive margin) or is a dummy that takes the value one when the firm is exporting and zero otherwise (extensive margin). X_{it} includes firm-level controls such as capital intensity, age and financial status (debt to asset ratio). The firm- and year-fixed effects are included, and the standard errors are clustered at the firm level. The results are presented in Table 7. Both margins are significantly positively correlated with the number of new patent applications.

Table 7: Correlation between new patents and trade margins

	(1) ln(exp)	(2) export status
lnN	0.095***	0.009**
	(0.017)	(0.004)
Observations	15,516	32,315
R-squared	0.900	0.819
Year FE	YES	YES
Firm FE	YES	YES
Controls	YES	YES

Notes: Standard errors are clustered at the firm level and in parentheses. *** p< 0.01, **p< 0.05, *p< 0.1. $\ln N$ is the log of new patent applications.

Alternative measure: I provide alternative measures of both margins which are more closely related to theory. In Chaney (2008), the extensive margin is the change in the sales of new exporters while the intensive margin is the change in the sales of the existing exporters. Using the customs data, I separate all exporters into two groups. The incumbents are the firms that were already exporters before the WTO event, while the new entrants are the firms that started to export afterwards. The specification is the same as equation (23). As the customs database also provides me with the trading partner of each firm, I also investigate whether changes in sales differ depending on the trading partner's patent stocks. Using equation (24), the model predicts a positive coefficient of β_3 in both cases.

$$Y_{it} = \beta_1 \ln N_{it} + \beta_2 \ln S_{it} + \beta_3 \ln N_{it} * \ln S_{it} + \lambda_i + \lambda_t + \epsilon_{it} \tag{24}$$

where Y_{it} is the log of exports for incumbents or new exporters and S_{it} is the average patent stocks of all of firm i's trading partners at time t. Results are presented in Table 8. The positive and significant coefficients in columns (1) and (3) confirm the data is consistent with the model prediction that higher knowledge stock is correlated with higher exports of existing exporters (intensive margin) and with a higher number of trading partners of new entrants (extensive margin). Additionally, columns (2) and (4) are consistent with the model's predictions that higher correlations should be observed when the knowledge stocks of trading partners are higher (or when the gaps between their stocks and China's are larger).

Table 8: New patents and trade margins

	ln(exp) of existing exporters				ln(exp) of new exporters			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
lnN	0.1787***	0.1630***	0.0945**	0.1138***	0.0781***	0.0710***	0.0489**	0.0505**
	(0.0363)	(0.0346)	(0.0426)	(0.0393)	(0.0242)	(0.0229)	(0.0237)	(0.0225)
lnS			0.3064***	0.3570***			0.2272***	0.2652***
			(0.0551)	(0.0300)			(0.0036)	(0.0040)
lnN × lnS			0.0808***	0.0490***			0.0634***	0.0338***
			(0.0160)	(0.0147)			(0.0073)	(0.0069)
Observations	162,645	162,626	162,645	162,626	495,622	495,613	495,622	495,613
R-squared	0.2672	0.3770	0.2905	0.3972	0.2773	0.3744	0.2927	0.3869
Firm FE	YES	YES	YES	YES	YES	YES	YES	YES
Year FE	YES		YES		YES		YES	
Year-Sector FE		YES		YES		YES		YES

Notes: Standard errors are clustered at the firm level and in parentheses. *** p< 0.01, **p< 0.05, *p< 0.1. Columns (1)-(4) are the results for the exports of incumbents that were exporters in 2000 and (5)-(8) are the results for the firms that started to export after 2000. $\ln N$ is the log of new patent applications and $\ln S$ is the average number of patent stocks of all of a firm's trading countries.

4.3 A two-country simulation exercise

So far we have focused on a single country n which experienced a trade liberalization event. However, the impact of a higher knowledge stock in n on its trading partner country i is ambiguous. On the one hand, as country n gets richer, it buys more from country i and produces cheaper exports, indicating i should benefit from n's increase in knowledge. On the other hand, however, higher knowledge stock in n changes the expenditure share structure and thus the speed at which country i obtains new knowledge. Therefore, it could be the case that country i is now learning slower, which decreases its overall output.

Proposition 4: *An increase in the knowledge stock in country n could increase, decrease, or not change the speed of knowledge accumulation in country i.*

Proof: *In a two-country case, the knowledge diffused to country i (see equation (20)) can be written as $\pi^j_{int}\left(T^j_{n,t}\right)^{\rho^j} + \pi^j_{iit}\left(T^j_{i,t}\right)^{\rho^j}$. Since $\frac{\partial X^j_{in,t+1}}{\partial T^j_{n,t+1}} > 0$ and $\frac{\partial X^j_{ii,t+1}}{\partial T^j_{n,t+1}} = 0$, π^j_{int} increases and π^j_{it} decreases. Now contingent on the size of T^j_n and T^j_i, the overall effects could be positive, negative, or zero.*

Moreover, even though country i could have higher knowledge stocks in some cases, the welfare is not necessarily higher due to the general equilibrium effect. I present a two-country, two-sector example below for illustration. For simplicity, the two countries are symmetric, with the only difference being in their initial knowledge stock. Country One has a lower initial knowledge stocks than Country Two. Trade is balanced every period.

I perform two exercises. The first one compares the changes in trade margins and welfare when there is a reduction in bilateral trade costs, with and without diffusion and adoption. The second one compares the results when the initial knowledge gaps between the two countries vary. The parameters used for the exercises are in Tables A.3-A.4.

4.3.1 Trade liberalization with and without diffusion and adoption

Suppose there is a 10% reduction in bilateral trade costs in period five, only affecting sector one. Panels (a) and (b) in Figure 6 show the change in the number of exporters in countries one and two (extensive margin). The dashed blue curves (benchmark case) correspond to a static trade model without shock, indicating there is no change in the number of exporters. In contrast, the dash-dotted red curves show the change in the number of exporters without shock if diffusion and adoption are both at work. The numbers are always higher than the benchmark since diffusion and adoption increase the knowledge stocks in both countries. The gaps between the dash-dotted red curve and the dashed blue curve indicate the gains for both countries.

When the shock happens in period five, the solid yellow curves show the increase in the exporters in both countries in a standard static Chaney (2008) model. The dash-circled purple curves, however, present the responses with diffusion and adoption. There will be similar increases in period five, but after that, the purple curves become steeper than the red curves, indicating the amplification effect of diffusion and adoption. The amplification comes from the increase in the speed of knowledge accumulation as well as a higher number of exporters adopting more foreign technologies. However, the increases are much smaller in Country Two. The green dotted lines represent the effect with only diffusion. If diffusion is the only source of obtaining foreign knowledge, the increases in the number of exporters will be much smaller.

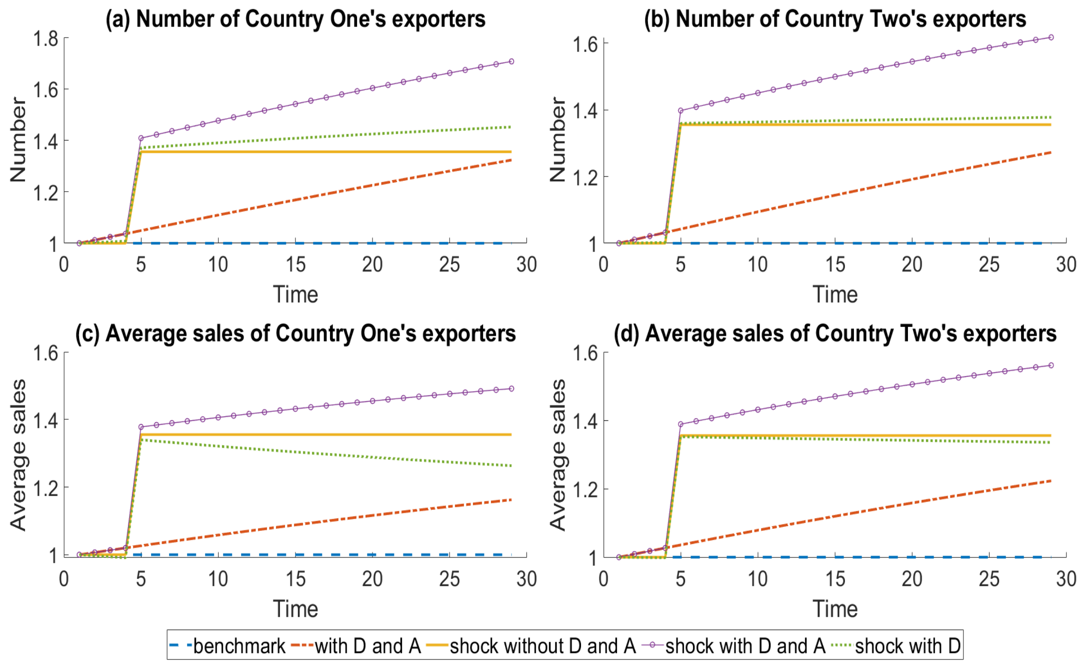

Figure 6: Changes in trade margins when tariffs drop at period five

Notes: This figure shows how a 10% bilateral tariff reduction in period five affects Country One (panels (a) and (c)) and Country Two's (panels (b) and (d)) number of exporters and sales per exporter. The dashed blue curves (benchmark) show the normalized number of exporters and sales when tariffs do not change and there is no learning. The dash-dotted red curves (with D and A) show the changes when diffusion and adoption are both at work but tariffs stay the same. The yellow solid lines (shock without D and A) show the changes when there is a 10% bilateral tariff reduction in period five without any learning. The dash-circled purple curves (shock with D and A) present the responses to the tariff reduction with diffusion and adoption. Finally, the dotted green curves (shock with D) show the responses if only diffusion is present.

Panels (c) and (d) show the average sales per exporter in each country. Different from the extensive margin, lower bilateral trade costs with both diffusion and adoption channels lead to a higher increase in Country Two's intensive margins than Country One's. If we compare the effects of diffusion with adoption separately, existing exporters will gain less with only diffusion than in a model without both channels. The relative loss is larger in Country One as diffusion squeezes the profits of existing exporters more in the low-stock country.

Figure 7 presents the welfare effects of this one-time shock, with panels (a) and (b) showing the short-run effect and panels (c) and (d) showing the whole trajectory towards the steady state. The gaps between the dash-circled purple lines and the dotted green lines are the gains from the interaction between diffusion and adoption and the gaps between the dash-circled purple lines and the dash-dotted red lines are the additional gains from trade liberalization. Panel (b) shows that

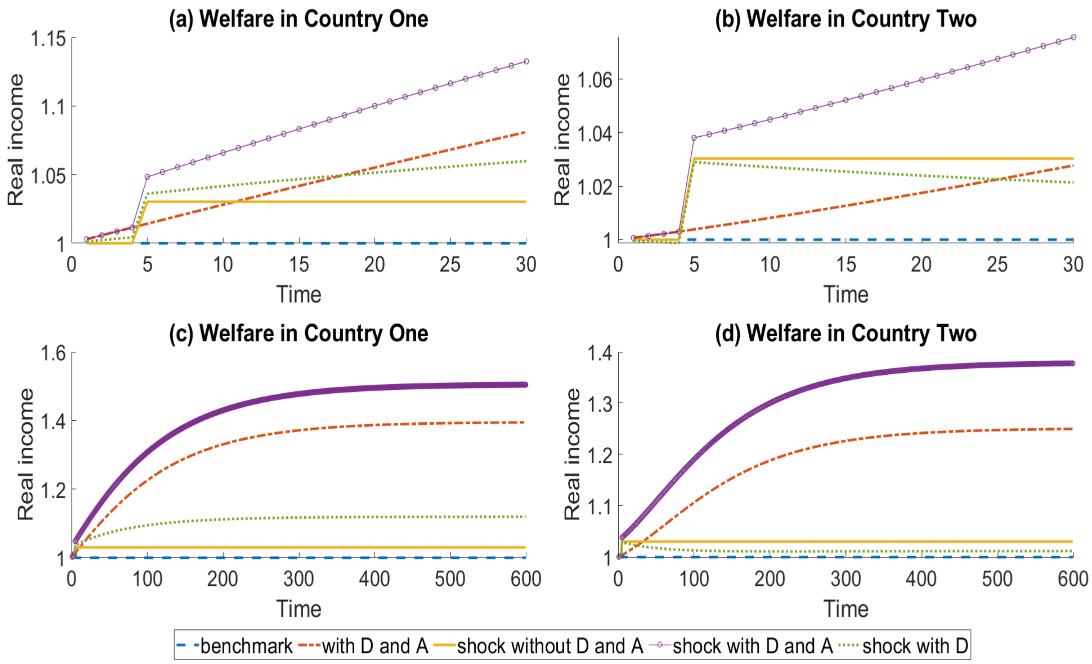

Figure 7: Changes in welfare when tariffs drop at period five

Notes: This figure shows how a 10% bilateral tariff reduction in period five affects Country One (panels (a) and (c)) and Country Two's (panels (b) and (d)) welfare in the short- and long-run. The dashed blue curves (benchmark) show the normalized number of exporters and sales when tariffs do not change and there is no learning. The dash-dotted red curves (with D and A) show the changes when diffusion and adoption are both at work but tariffs stay the same. The yellow solid lines (shock without D and A) show the changes when there is a 10% bilateral tariff reduction in period five without any learning. The dash-circled purple curves (shock with D and A) present the responses to the tariff reduction with diffusion and adoption. Finally, the dotted green curves (shock with D) show the responses if only diffusion is present.

diffusion alone reduces the welfare in Country Two. This indicates adoption and its interaction with diffusion reverses the negative impact: When Country One grows faster due to diffusion, Country Two's exporters benefit more when adopting One's technologies.

Two other points are worth mentioning. First, the long-run effects are large and the welfare gains are even greater in later periods, indicating the dynamic impacts of knowledge accumulation. Second, although Country Two starts with a higher knowledge stock, in the long run, the additional gain from trade liberalization when both learning channels are at work is still substantial. This simple exercise shows that ignoring diffusion and its interaction with adoption on trade margins leads to a significant underestimation of gains from trade.

There are similar results when I introduce a 10% reduction in fixed costs in Figures B.3-B.4, though the effects are smaller. One interesting point is shown in panels (c) and (d) of Figure B.4, where trade liberalization leads to a reduction in the intensive margin. While the theory states there should be no response from the intensive margin when there is a change in fixed cost, the reduction comes from the general equilibrium effect as both countries have lower prices.

4.3.2 Different initial knowledge gaps

The second exercise compares the changes in welfare when the initial gaps in knowledge stocks in sector one are different and there is no shock to trade costs. The initial knowledge stocks of Country Two are always normalized to 0.9 (the steady state is 1), while the stock in Country One varies from 0.2 to 0.8 in cases one to four. The results are presented in Figure 8. An interesting finding is that when the initial gap is very high, the welfare in Country Two is lower in the beginning due to faster knowledge diffusion and technology adoption, but starts to increase when Country One catches up. Country One, however, always gains from learning from Country Two. In other cases when the initial gap is small, both countries gain from diffusion and adoption. This exercise indicates the possibility of losing from trade when the knowledge stocks are too different and thus the importance of considering the interaction between countries when evaluating trade policies.

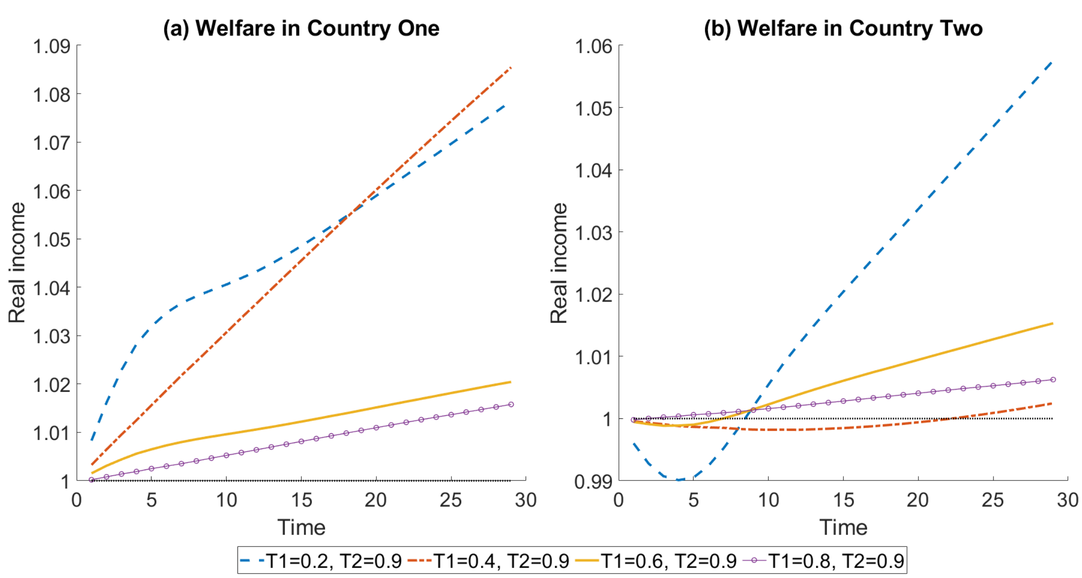

Figure 8: Changes in welfare when initial knowledge gap differs

Notes: This figure shows the impact of different initial knowledge gaps on welfare in Country One (panel (a)) and Country Two (panel (b)). The initial knowledge stocks of Country Two are always normalized to 0.9 (the steady state is 1), while the stock in Country One varies from 0.2 to 0.8 in cases one to four.

5 Solving the general equilibrium model

The exercise above highlights that the general equilibrium (GE) effect is essential and should not be ignored. To quantify the GE effect, I solve the model using the dynamic hat algebra introduced in Caliendo et al. (2019), which is developed from the technique in Dekle et al. (2007) and Dekle et al. (2008). This method does not require the levels of the fundamentals and thus maintains the advantage of using only the changes in variables and thus has low data requirements and simplifies the task. However, price is not a function of total sales in Caliendo et al. (2019). Thus, I develop a two-layer fixed-point algorithm based on their original method, where the inner loop searches for the optimal sales (X^j_{nt+1}) for a given wage w_{nt} and the outer loop finds the optimal wage that solves all equations.

Defining the detrended variables as $\hat{x}$, we can first write the changes in variable $\hat{x}$ in periods $t+1$ and t as x', i.e. $x'_{t+1} = \frac{\hat{x}_{t+1}}{\hat{x}_t}$, to represent the whole system as

$$c'^j_{n,t+1} = (w'_{n,t+1})^{\gamma^j_n} \prod_{k=1}^{J} \left(P'^k_{n,t+1}\right)^{\gamma^{k,j}_n}, \tag{25}$$

$$P'^j_{nt+1} = \left[\sum_{i=1}^{N} \pi^j_{nit} \left(\frac{c'^j_{it+1} d'^j_{nit+1}}{\zeta'^j_{nit+1}} \right)^{-\theta^j} w'_{it+1} L'_{it+1} T'^j_{it+1} \left(\frac{f'_{nit+1}}{X'^j_{nt+1}} \right)^{\frac{\sigma^j - 1 - \theta^j}{\sigma^j - 1}} \right]^{-\frac{1}{\theta^j}}, \tag{26}$$

$$\pi^j_{nit+1} = \pi^j_{nit} w'_{it+1} L'_{it+1} T'^j_{it+1} \left(\frac{f'^j_{nit+1}}{X'^j_{nt+1}} \right)^{1 - \frac{\theta^j}{\sigma^j - 1}} \left(\frac{c'^j_{it+1} d'^j_{nit+1}}{P'^j_{nt+1} \zeta'^j_{nit+1}} \right)^{-\theta^j}, \tag{27}$$

$$\hat{X}^j_{n,t+1} = \sum_{k=1}^{J} \gamma^{j,k}_n \sum_{i=1}^{N} \hat{X}^k_{i,t+1} \pi^k_{in,t+1} + \alpha^j_n \left(\hat{w}_{n,t} L_n w'_{n,t+1} + \sum_{j=1}^{J} \frac{\sigma^j - 1}{\sigma^j \theta^j} \hat{Y}^j_{nt+1} + \hat{D}_{n,t} D'_{n,t+1} \right), \tag{28}$$

$$\hat{w}_{n,t} L_n w'_{n,t+1} = \sum_{j=1}^{J} \gamma^j_n \sum_{i=1}^{N} \hat{X}^j_{i,t+1} \pi^j_{in,t+1}, \tag{29}$$

$$\hat{T}^j_{nt+1} \exp(g^j_T) - \hat{T}^j_{nt} = m^j_0 \Gamma \left[\sum_{i=1}^{N} \pi^j_{nit} \left(\hat{T}^j_{i,t} \right)^{\rho^j} \right]. \tag{30}$$

Since I have the data for annual bilateral trade flows, output, and value-added, using the calibrated parameters described in the next section, the solution for the baseline economy can be presented by the changes in wages that are consistent with the equilibrium conditions.

To quantify the impact of knowledge diffusion and technology adoption on welfare, I can simulate counterfactuals assuming no diffusion or adoption and compare the results with the benchmark. Note the benchmark solution corresponds to the equilibrium consistent with the real observations in 2000-2007. Therefore, I do not need to explicitly calculate the changes in f or d since the observed data already incorporate all information. Using a similar technique, I write the counterfactual variables as x^* and define the difference between the benchmark and counterfactual variables as $\tilde{x}$, i.e. $\tilde{x}_{t+1} = \frac{\hat{x}^*_{t+1}}{\hat{x}^*_t} / \frac{\hat{x}_{t+1}}{\hat{x}_t}$. The new system of equations is as follows

$$\tilde{c}^j_{n,t+1} = (\tilde{w}_{n,t+1})^{\gamma^j_n} \prod_{k=1}^{J} \left(\tilde{P}^k_{n,t+1}\right)^{\gamma^{k,j}_n} \tag{31}$$

$$\tilde{P}^j_{n,t+1} = \left[\sum_i^N \pi^{*j}_{ni,t} \left(\frac{\tilde{c}^j_{i,t+1} \tilde{d}^j_{ni,t+1})}{\tilde{\zeta}^j_{nit+1}} \right)^{-\theta^j} \tilde{w}_{it+1} \tilde{L}_{it+1} \tilde{T}^j_{it+1} \left(\frac{\tilde{f}_{nit+1}}{\tilde{X}^j_{nt+1}} \right)^{\frac{\sigma^j - 1 - \theta^j}{\sigma^j - 1}} \right]^{-\frac{1}{\theta^j}} \tag{32}$$

$$\pi^{*j}_{ni,t+1} = \pi'^j_{ni,t+1} \pi^{*j}_{ni,t} \tilde{w}_{it+1} \tilde{L}_{it+1} \tilde{T}^j_{it+1} \left(\frac{\tilde{f}^j_{nit+1}}{\tilde{X}^j_{nt+1}} \right)^{1 - \frac{\theta^j}{\sigma^j - 1}} \left(\frac{\tilde{c}^j_{it+1} \tilde{d}^j_{nit+1}}{\tilde{P}^j_{nt+1} \tilde{\zeta}^j_{nit+1}} \right)^{-\theta^j} \tag{33}$$

$$\hat{X}^{*j}_{n,t+1} = \sum_{k=1}^{J} \gamma^{j,k}_n \sum_{i=1}^{N} \hat{X}^{*k}_{i,t+1} \pi^{*k}_{in,t+1} + \alpha^j_n \left(\hat{w}^*_{n,t+1} L_n + \sum_{j=1}^{J} \frac{\sigma^j - 1}{\sigma^j \theta^j} \hat{Y}^{*j}_{nt+1} + \hat{D}^*_{n,t+1} \right) \tag{34}$$

$$\hat{w}^*_{n,t+1} L_n = \sum_{j=1}^{J} \gamma^j_n \sum_{i=1}^{N} \hat{X}^{*j}_{i,t+1} \pi^{*j}_{in,t+1} \tag{35}$$

$$\hat{T}^{*j}_{nt+1} \exp(g^j_T) - \hat{T}^{*j}_{nt} = \left[\sum_{i=1}^{N} \pi^{*j}_{nit} \left(\hat{T}^{*j}_{i,t} \right)^{\rho^j} \right] \tag{36}$$

Since $\hat{w}^*_{n,t+1} L_n = \tilde{w}_{n,t+1} w'_{n,t+1} \hat{w}^*_{n,t} L_n$, I can calculate it by the observed annual change in VA and the last period's counterfactual VA.

6 Parameter estimation

Now I describe how to bring the model to the data. Some parameters are standard and can be calibrated using the procedures following the literature. Others are new and must be estimated. I estimate them following a sequential structure via the simulated method of moments (SMM).

6.1 Calibration

The main data used for calibrating sectoral linkages is the World Input-Output Table (WIOT) in 2000. The WIOT contains 44 economies, and I take out 21 with high-quality patent data and then aggregate the others into the ROW (see Appendix Table A.1). These 21 economies account for 98% of global trade from 2000 to 2007. There are three agriculture, one mining, 19 manufacturing and 33 services sectors; I merge them into 19 sectors (see appendix Table A.2). To calibrate the model, γ_n^j and $\gamma_n^{j,k}$ can be obtained directly using the Table. γ_n^j is the share of value added in the production function and can be calculated as a fraction of total output, i.e. $\gamma_n^j = \frac{VA_n^j}{Y_n^j}$. $(1-\gamma_n^j)$ is the share of all intermediate goods in the production function and $\gamma_n^{j,k}$ measures the share of sector k's spending on sector j's intermediate goods in the country n. The final consumption share, α_n^j, is the final expenditure in sector j divided by the total final absorption in country n.

The elasticity of substitution across varieties is calculated similarly to Antras et al. (2017) by the ratio of sales to variable input purchases. Specifically, I use firms' sales and the sum of material inputs and wages from the ASIF database. Following the model assumptions, the ratio of sales over costs for sector j should be $\frac{\sigma^j}{\sigma^j-1}$. The calculated parameters are presented in Table 9.

For the elasticity θ^j, I explore the assumption that the distribution of firm productivities is Pareto. Since $G_n^j = 1 - T_n^j {z_n^j}^{-\theta^j}$, indicating $Pr\{z > Z\} = T_n^j {Z_n^j}^{-\theta^j}$, we can take logs on both sides to reach

$$ln(Pr\{z > Z\}) = b_0 lnT_n^j + b_1 lnZ_n^j + \epsilon_{j,n}, \tag{37}$$

where Z_n^j is the observed productivity of Chinese firms. I follow Shapiro and Walker (2018) to use the upper 90th percentile of the distribution. As they state, using only the upper tail of the firm sizes is in line with the literature, and Pareto distribution has a better fit at the right tail. One difference is that they use firm sales, while

I use TFP directly for the estimation. The estimated values are also in Table 9. Note I cannot estimate σ and θ for agriculture and services since I only have data for manufacturing firms; I take the sectoral average as Hsieh and Ossa (2016). All estimations are significant at the 1% level and consistent with the model assumption that $\theta^j > \sigma^j - 1$ (see Table 9).

6.2 Simulated Method of Moments

The remaining parameters are the diffusion rate ρ^j, initial exogenous idea arrival rate m_0^j, and the weight for the cost-reduction technique Δ^j. I follow the strategy of Buera and Oberfield (2020) to first estimate T_{nt}^j for each country-sector pair and then calculate the parameters via the Simulated Methods of Moments (SMM).

Step 1: estimate T_{nt}^j. First, note that the domestic expenditure share is

$$\pi_{nnt}^j = b^{\theta^j} w_{nt} L_{nt} {T_{nt}^j}^{1+\theta^j} \left(\frac{c_{nt}^j}{P_{nt}^j}\right)^{-\theta^j} {X_{nt}^j}^{\frac{\sigma^j - 1 - \theta^j}{1-\sigma^j}}, \tag{38}$$

and thus the level of knowledge stock at time t is

$$T_{nt}^j = \left[b^{-\theta^j} \frac{\pi_{nnt}^j}{w_{nt} L_{nt}} \left(\frac{c_{nt}^j}{P_{nt}^j}\right)^{\theta^j} {X_{nt}^j}^{\frac{\sigma^j - 1 - \theta^j}{\sigma^j - 1}} \right]^{\frac{1}{1+\theta^j}}. \tag{39}$$

Since X_{nt}^j is the total expenditure of country n in sector j, it can be calculated using total imports. Then it is straightforward to get π_{nnt}^j as the domestic share of total expenditure, where the domestic sales are equal to gross production minus total exports. Value-added can be obtained from the data. Finally, we need to calculate the cost-price ratio,

$$\frac{c_{nt}^j}{P_{nt}^j} = \frac{\eta_n^j w_n^{\gamma_n^j} \prod_{k=1}^J \left(P_n^k\right)^{\gamma_n^{k,j}}}{P_{nt}^j} = \eta_n^j \left(\frac{w_n}{P_{nt}^j}\right)^{\gamma_n^j} \prod_{k=1}^J \left(\frac{P_n^k}{P_{nt}^j}\right)^{\gamma_n^{k,j}}$$

where $\eta_n^j = \prod_{k=1}^J (\gamma_n^{k,j})^{-\gamma_n^{k,j}} (\gamma_n^j)^{-\gamma_n^j}$ is a constant. Due to the lack of detailed sectoral price data, I assume that all manufacturing sub-sectoral prices grow at the same rate, and thus I can use equation (6) for the calculation given the initial wage-price ratio for sector j. As I assume there is no change in labor, a change in wage will be the same as a change in value-added. Therefore, I approximate $\frac{w'_{nt}}{{P'}_{nt}^j}$ using the changes in real GDP in each sector. I obtain the nominal and real sectoral GDP (constant 2015 USD) from the World Bank and National Statistics Taiwan.[12] The estimated T^j_{n0} for each sector-economy pair is presented in Figure 9. A darker color indicates a higher stock in the year 2000, and each row is normalized such that the maximum is one.[13] It is clear that there is a large heterogeneity across sectors. For example, Norway has very high knowledge stocks in the mining sector, which is consistent with the fact that Norway is heavily reliant on oil and gas. The US has high stocks in a couple of sectors such as computer, transport and mineral. Certain sectors are heavily dominated by specific economies; for example, agriculture is predominantly led by the ROW, while Japan holds a significant presence in the coke sector, and Sweden plays a prominent role in the chemical sector.

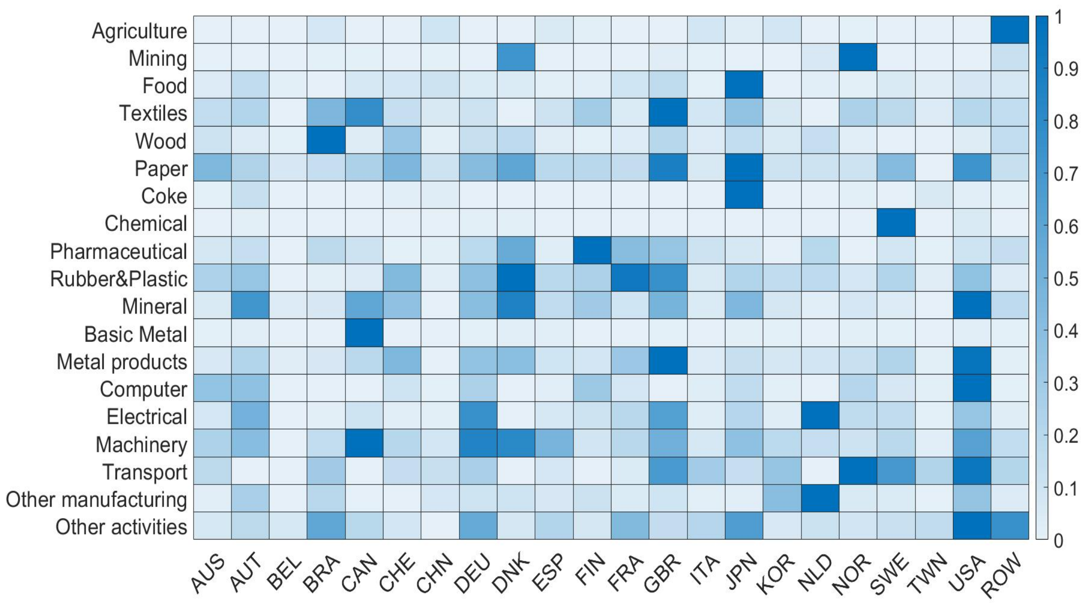

Figure 9: Estimated initial knowledge stock in the year 2000

Notes: This figure shows the estimated knowledge stock of each economy-sector pair in 2000, where a darker color indicates higher stock. Each row is normalized such that the maximum is one. Comparison across rows is infeasible here.

Step 2: estimate m^j_0 and ρ^j. After getting the initial T^j_{n0} and the changes (T'^j_{nt}), I use equation (16) to estimate m^j_0 and ρ^j. I set $g_m = 0.01$ as in Buera and Oberfield

[12] As World Bank does not provide data for Taiwan separately, I get data from National Statistics Taiwan, Table "Gross Domestic Product by Kind of Activity and Implicit Price Deflators". The price is adjusted to 2015 constant USD for consistency. The historical exchange rates are obtained from Mega Bank.

[13] Note that the colors in each column are not comparable. A darker color does not indicate a higher absolute level of knowledge.

(2020), and I choose two moment conditions for each sector to estimate m_0^j and ρ^j. Specifically, the average of the knowledge stocks T_n^j and the cross-economy variance are used for identification. m_0^j can be identified by the mean since a higher m_0^j leads to a higher average knowledge stock everywhere. The variance provides information on ρ^j as it is related to the heterogeneity in domestic expenditure shares. I use the Matlab global search function to avoid local minimums, and the preferred estimation results are presented in Table A.5. m_0^j governs the initial idea arrival rate, and a higher rate indicates a faster growth of new ideas. The ρ^j parameter determines the diffusion rate from other trading partners. Interestingly, the agriculture sector has the highest diffusion rate, followed by metal products and paper. The computer sector, however, has a very low diffusion rate, indicating that diffusion from sellers plays a limited role.

Table 9: Estimated parameter values

Sector ID	Sector Name	σ^j	θ^j	m_0^j	ρ^j	Δ^j
1	Agriculture	5.26	8.92	0.61	0.60	0.76
2	Mining	3.63	9.00	0.18	0.13	0.27
3	Food	4.48	9.06	0.08	0.12	0.28
4	Textiles	6.09	8.18	0.001	0.02	0.63
5	Wood	5.61	8.24	0.30	0.10	0.89
6	Paper	5.55	8.26	0.54	0.47	0.10
7	Coke	5.35	9.76	0.002	0.11	0.80
8	Chemical	5.19	9.21	0.01	0.29	0.37
9	Pharmaceutical	4.12	9.77	0.001	0.04	0.37
10	Rubber & Plastic	6.13	8.62	0.07	0.32	0.62
11	Mineral	4.50	8.42	0.42	0.21	0.86
12	Basic Metal	5.88	9.66	0.21	0.05	0.31
13	Metal products	5.99	8.46	0.63	0.51	0.34
14	Computer	5.61	9.26	0.35	0.03	0.59
15	Electrical	6.03	8.95	0.001	0.16	0.41
16	Machinery	5.18	8.49	0.03	0.25	0.64
17	Transport equipment	5.37	8.71	0.02	0.16	0.36
18	Other manufacturing	3.46	9.56	0.05	0.13	0.58
19	Other activities	5.26	8.92	0.50	0.42	0.20

Notes: This table presents the estimated parameter values using the calibration and SMM procedures described in Section 6. The estimations for θ^j and σ^j use Chinese firm-level data. For the other three parameters, sectoral-level bilateral trade flows for all 22 economies are used.

Step 3: estimate Δ^j. Finally, I estimate Δ^j using a strategy similar to that of Caliendo and Parro (2015). Assuming symmetric bilateral trade costs and fixed costs, take three countries i, k, n and their expenditures on the other two countries to compute the following:

$$\frac{\pi_{ni}^j}{\pi_{kn}^j}\frac{\pi_{ik}^j}{\pi_{ki}^j}\frac{\pi_{kn}^j}{\pi_{in}^j} = \left(\frac{\Delta^j T_n^j + (1-\Delta^j)T_k^j}{\Delta^j T_n^j + (1-\Delta^j)T_i^j}\frac{\Delta^j T_k^j + (1-\Delta^j)T_i^j}{\Delta^j T_i^j + (1-\Delta^j)T_k^j}\frac{\Delta^j T_i^j + (1-\Delta^j)T_n^j}{\Delta^j T_k^j + (1-\Delta^j)T_n^j}\right)^{-\theta^j}. \tag{40}$$

For 22 economies, I have $C(n,k) = \frac{22!}{3!(22-3)!} = 1540$ observations for each sector-year pair. By matching the mean of the left-hand side of empirical and simulated data, I estimate Δ^j and present the values in Table 9. The share of domestic technology is $1-\Delta^j$, which dominates the foreign share in around half of cases. Δ^j captures the effect of adopting foreign technology on domestic productivity. Interestingly, the sectors that have a higher reliance on foreign knowledge are wood, mineral, coke and agriculture. Table 10 compares the targeted moments for each sector and the results are reasonably well. The largest discrepancies are the numbers for the paper and mineral sectors in column (9). This is due to a large number of zeros appearing in the expenditure shares used for estimating T_{nt}^j and thus leads to an underestimation of the mean.

Table 10: Data and simulated moments

	Mean			S.D.			Mean		
	Data	Simulated	Ratio	Data	Simulated	Ratio	Data	Simulated	Ratio
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Agriculture	13.59	13.60	1.0	2.00	2.01	1.0	4.56	4.56	1.0
Mining	11.43	11.50	1.0	4.02	4.05	1.0	6.58	6.58	1.0
Food	7.41	7.44	1.0	1.07	1.08	1.0	1.57	1.57	1.0
Textiles	7.56	8.90	0.8	4.28	3.71	1.2	1.21	1.21	1.0
Wood	9.94	9.99	1.0	1.21	1.20	1.0	4.40	4.40	1.0
Paper	10.76	10.76	1.0	0.93	0.93	1.0	3.57	1.90	0.5
Coke	6.01	6.16	1.0	2.37	2.26	1.0	8.74	8.74	1.0
Chemical	8.51	8.77	1.0	3.85	3.96	1.0	2.12	2.12	1.0
Pharmaceutical	8.58	9.56	0.9	3.72	2.88	1.3	2.70	2.70	1.0
Rubber & Plastic	10.66	10.68	1.0	2.62	2.62	1.0	0.84	0.84	1.0
Mineral	9.86	9.88	1.0	0.96	0.98	1.0	2.27	0.71	0.3
Basic Metal	8.55	9.03	0.9	2.91	3.08	0.9	3.81	3.81	1.0
Metal products	12.14	12.15	1.0	1.42	1.44	1.0	1.13	1.13	1.0
Computer	7.54	8.00	0.9	4.06	4.32	0.9	1.22	1.22	1.0
Electrical	8.94	9.32	1.0	3.32	3.35	1.0	0.97	0.97	1.0
Machinery	9.63	9.71	1.0	2.34	2.35	1.0	1.12	1.12	1.0
Transport equipment	6.02	6.20	1.0	3.79	3.91	1.0	1.67	1.67	1.0
Other manufacturing	6.80	6.85	1.0	2.99	3.01	1.0	0.86	0.86	1.0
Other activities	19.03	19.05	1.0	1.08	1.10	1.0	3.07	3.07	1.0

Notes: This table presents the data moments and simulated moments for the SMM estimation. Columns (1) and (2) show the mean and s.d. for knowledge stocks, which are used for estimating m_0^j and ρ^j. Column (3) shows the mean of expenditure share ratios (left-hand side of equation (40)), which is used for estimating Δ^j.

Non-targeted moments: I show the non-targeted moments in Table 11 below. As estimating Δ^j only uses the mean of the expenditure ratios (left-hand side of equation (40)), I compare the average standard deviations of the data and the model simulated value. The results are very close. Moreover, the cross-sectional correlations are similar in data and simulation. To ensure the empirical patent stocks

Table 11: Non-targeted moments

Moments	Data	Simulated
S.D. of expenditure ratios	2.18	2.13
Correlation of expenditure ratios	0.96	0.84

Notes: This table presents the non-targeted data moments and simulated moments for the SMM estimation.

are a proper proxy for knowledge stocks in the model, I compare the (log of) estimated knowledge stock with the (log of) empirical patent stock for all economies between 2000 and 2007. Notably, the estimation of knowledge stocks is entirely independent of patent information. Figure 10 presents the positive correlation between these two variables. The model successfully generates a positive correlation, indicating that countries with more patents tend to have higher knowledge stocks. However, the model performs less well when the estimated knowledge stock level is low. This is again due to some very low bilateral flows used for the estimation of knowledge stocks.

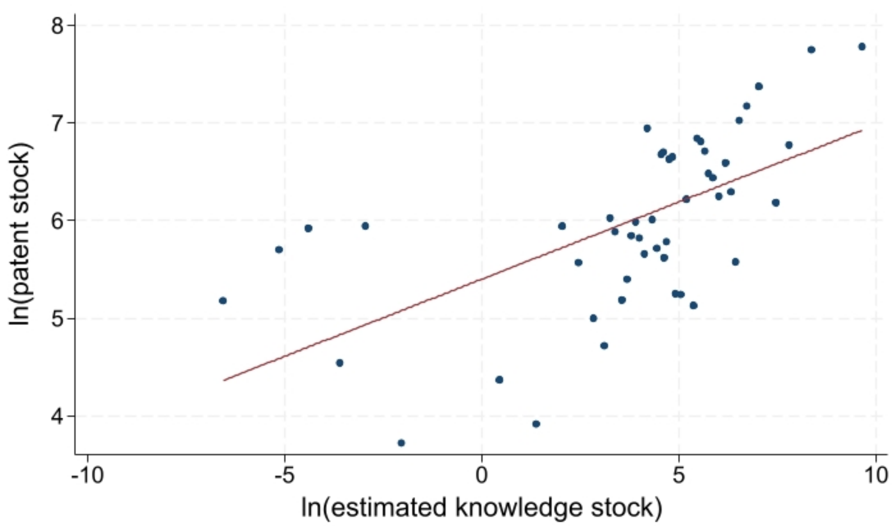

Figure 10: Correlation between estimated knowledge stock and patent stock
Notes: This figure shows the correlation between (log of) the estimated knowledge stock using (39) and (log of) the empirical patent stock using a bin plot with 50 bins. The figure uses all patents across countries from 2000 to 2007.

7 Counterfactual analysis

In this section, I evaluate how diffusion and adoption change the gains from trade and the outcomes of trade liberalization. First, I compare the results in three cases: a model with knowledge diffusion, a model with technology adoption only and a model with both channels. I then examine the impact of diffusion and adoption when there is a trade liberalization event (the WTO accession) by comparing the economic outcomes in models with and without them. The final exercise tests the impact of a unilateral increase in trade barriers, inspired by the recent US-China trade war.

7.1 Impact of knowledge diffusion and technology adoption

The first counterfactual only changes the knowledge diffusion or adoption process, and there is no change in trade costs. To get the difference in knowledge stocks when there is no diffusion, I calculate the counterfactual changes using the diffusion equation (36) by setting ρ^j to zero. Similarly, I get the results without adoption by setting ζ^j_{ni} to one.

Table 12 and Figure B.5 show the changes in welfare if there is no diffusion or no adoption. If there were no diffusion from 2000 to 2007, all economies would have lower welfare, as measured by real income. Belgium and the Netherlands, two relatively small developed countries, suffer the largest losses (-8.7% and -6.85% separately). Both countries are relatively small and rely heavily on imports. Moreover, they have relatively low knowledge stocks in high-diffusion sectors, indicating diffusion is an important source of their knowledge accumulation. Japan and Korea, on the other hand, only have 0.24% and 0.25% reduction in welfare. Surprisingly, China and Brazil, two large developing countries, do not have substantial losses. This is due to their low foreign expenditure shares compared with domestic shares.

The results for adoption are very different. Some developed countries would have higher welfare if there were no adoption. Japan, Norway and the US gain 0.9%, 0.8% and 0.9% respectively. These countries have high knowledge stocks and thus can preserve their technological advantages if foreign technology adoption is infeasible. Belgium and the Netherlands again suffer the largest losses (-10.5% and -13.5% separately). If we compare the importance of diffusion and adoption, there is a large variation across countries. For some countries such as China, Denmark, and the UK, adoption outweighs diffusion, while in Austria, Germany and Italy, diffusion contributes more to welfare than adoption. Moreover, the gains from adoption and diffusion do not add up linearly when both mechanisms operate simultaneously. This suggests a nuanced interaction between these two mechanisms: Diffusion increases the knowledge stocks in each economy and thus enables more firms to overcome the fixed costs to become exporters. As the number of exporters increases, more firms benefit from technology adoption and reduce their production costs. This interaction amplifies welfare gains in low-stock countries. However, for high-stock countries, there is a trade-off between accessing cheaper goods when low-stock countries catch up and having higher income by keeping technological advantages.

Table 12: The changes in welfare without diffusion, adoption or both

Without	Diffusion	Adoption	Both
AUS	-0.34%	0.4%	-0.34%
AUT	-1.75%	-0.1%	-1.07%
BEL	-8.70%	-10.5%	-14.72%
BRA	-0.76%	-1.9%	-2.60%
CAN	-5.04%	-6.8%	-7.96%
CHE	-5.18%	-5.9%	-7.09%
CHN	-0.38%	-3.3%	-4.10%
DEU	-1.72%	-0.1%	-1.13%
DNK	-2.91%	-6.9%	-7.71%
ESP	-0.36%	0.1%	-0.46%
FIN	-1.42%	-0.8%	-1.45%
FRA	-1.21%	0.4%	-0.70%
GBR	-2.54%	-6.0%	-6.96%
ITA	-0.72%	-0.2%	-1.32%
JPN	-0.24%	0.9%	0.57%
KOR	-0.25%	-1.6%	-1.95%
NLD	-6.85%	-13.5%	-15.42%
NOR	-0.77%	0.8%	-0.18%
SWE	-0.98%	0.2%	-0.54%
TWN	-2.31%	-3.1%	-3.50%
USA	-0.45%	0.9%	0.62%
ROW	-0.54%	-1.1%	-2.02%
Global Average	-2.1%	-2.6%	-3.6%

Notes: This table shows the relative changes in welfare (annual average) without adoption (column 1), without diffusion(column 2), or without both (column 3), compared with the benchmark case where diffusion and adoption are both at work.

To examine the impact of learning on firms' exporting decisions, I compare changes in trade margins in Figures B.6 to B.9 for diffusion and Figures B.10 to B.13 for adoption. Specifically, Figures B.6 and B.7 illustrate a comparison between the number of exporters with and without diffusion. If there were no diffusion, the number of exporters would decrease for most economies, implying that only more productive firms could engage in exporting. This observation suggests that diffusion has a similar effect to a reduction in trade costs in a standard Melitz-type model, albeit with continuous adjustments driven by the endogenous nature of the knowledge stocks. At the sectoral level, the agriculture sector experiences the most significant reduction in the number of exporters without diffusion (-5.96%), followed by the rubber and plastic sector (-2.94%). These sectors emerge as the biggest winners when diffusion is present, signifying substantial benefits for firms engaged in importing. In contrast, the chemical and other manufacturing sectors would have fewer exporters in the absence of diffusion.

In terms of the intensive margin, exporters in most economies would sell more if there were no diffusion. This suggests that diffusion plays a crucial role in suppressing average sales per existing exporter for most economies, which is consistent with the import competition channel that has been extensively studied in the literature. However, Austria, Brazil, China, Japan, Korea, and the US have lower sales without diffusion, indicating diffusion helps their exporters to enlarge the market. When examining sectoral differences, Figure B.9 demonstrates that, on average, exporters in most sectors would have higher sales without diffusion. However, there are four exceptions: the agriculture, mining, paper, and metal products sectors, which experience substantial reductions. This means that diffusion increases export sales in these sectors. In general, sectors with higher diffusion rates and more concentrated knowledge stocks tend to benefit more from diffusion. If we now compare the losses in Belgium and the Netherlands, it becomes evident that the impact in Belgium is largely driven by the extensive margin, while in the Netherlands, it is driven by the intensive margin.

In the case of adoption, all economies would have fewer exporters without technology adoptions (as shown in Figure B.10). The reductions are slightly higher compared to the scenario with no diffusion, although the variation is relatively small. Canada, in this case, experiences only a marginal loss of 0.12% of exporters, indicating that its exporters rely less on adoption to reduce production costs. Turning to sectoral changes, the computer sector stands out in Figure B.11. Without adoption, there would be an increase in the number of exporters in the computer sector. Furthermore, the existing exporters would achieve higher sales, as depicted in Figure B.13. These increases are particularly notable in countries with initially low knowledge stocks in 2000, such as Belgium, Denmark, the UK, and the Netherlands. Without adoption, there are no additional productivity gains for exporters. Therefore, the existing exporters must be the most productive firms in these low-stock countries that can still earn profits without the additional benefit of adoption. At the same time, the motivation for other firms to become exporters in these countries is much lower. Together, these two factors lead to higher intensive margins.

7.2 Impact of trade liberalization

The second counterfactual assesses the impact of trade costs on welfare and trade margins, specifically examining changes in tariffs after China's accession to the WTO at the end of 2001. As per Erten and Leight (2021), this event led to a roughly 5-percentage-point (or 27%) reduction in China's tariffs in 2002. To simulate this counterfactual scenario, I assume a 27% increase in China's trade costs for imports from other countries and a corresponding 27% increase in trade costs for all other economies when purchasing from China in 2002. Beyond 2002, there are no further changes in trade costs. I compare how welfare would change without China's WTO accession in four cases: a world with both diffusion and adoption, a world without either, a world with only diffusion and a world with only adoption. Additionally, I also investigate the impact of a 27% increase in fixed costs in these four cases.

Table A.5 presents that China's (global) welfare would be 17.15% (2.49%) lower in a world without either channel. However, the losses would be limited to 10.76% (2.18%) lower if both adoption and diffusion are at work. Considering only diffusion or adoption leads to similar results that learning helps to mitigate the losses from higher trade barriers. According to columns (2) and (3) of Table A.5, intriguingly, China's welfare would be lower in the absence of adoption but relatively higher without diffusion, compared to the scenario when both effects are present. In contrast, columns (3)-(6) show the results for higher fixed costs, where global welfare decreases by 0.43% if both channels are at work but only 0.27% if they are absent. Surprisingly, China gains when both diffusion and adoption are present or when both are absent, but losses when only either one channel is at work.

Table A.6 provides a comparison of the changes in the number of exporters and sales per exporter with and without diffusion or adoption. Specifically, columns (1)-(6) present the results for higher tariffs, while columns (7)-(12) show the results for higher fixed costs. In the benchmark case when both effects are at play and tariffs are higher, the number of exporters decreases in most economies without China's WTO accession, with China itself experiencing the largest reduction. Consequently, a reduction in bilateral trade costs with China benefits most other economies by attracting more new entrants to foreign markets. When separating the impacts of diffusion and adoption, adoption has a more significant negative effect on the extensive margin in both cases. If no technology could be adopted by exporters, all economies would have significantly fewer exporters. However, without diffusion, most countries would have more exporters in the absence of China's WTO accession, indicating that lower trade costs with China reduce their absolute advantages in terms of higher knowledge stocks and, consequently, the number of exporters.

Regarding sales per exporter, the magnitude of changes is much smaller compared to the extensive margin. Surprisingly, the changes are comparable for China in columns (4)-(5), suggesting that there isn't a substantial difference in existing exporters' sales with or without knowledge diffusion. However, without adoption, the number of exporters only reduces by 1%, showing that higher tariffs do not significantly impact the existing exporters' sales. In the case of higher fixed costs, most economies have more exporters and existing exporters also sell more, although the magnitude is much smaller than in the higher-tariff cases. One interesting finding is that the higher fixed costs are more detrimental to welfare without adoption while higher tariffs are more detrimental to welfare without diffusion.

7.3 Impact of protectionism

This final exercise assesses the impact of a unilateral tariff increase, inspired by the recent US-China trade war. On March 22, 2018, President Trump initiated a public debate by urging the United States Trade Representative to assess the possibility of imposing tariffs on Chinese goods, estimated at a value of US$ 50-60 billion. The justification for this move was grounded in Section 301 of the Trade Act of 1974, which aimed to address China's history of unfair business and trade practices, explicitly citing violations of intellectual property rights. Despite China's retaliatory actions and several waves of tariff changes, by October 2019, the additional tariffs imposed by the US on Chinese exports averaged 25% (Egger and Zhu

(2020)).

For this section's purposes, I omit China's retaliation and focus solely on the consequences of the US's heightened trade barriers. I examine how knowledge diffusion and adoption influence the gains or losses resulting from this unilateral tariff increase. Additionally, for completeness, I test the results for a 25% increase in fixed costs.

Table A.7 presents the changes in welfare. Globally, welfare would be 0.51% lower with both channels. However, in a world without either channel, the reduction in welfare would be 0.41%. Note these results are opposite to the WTO case, as learning amplifies the losses from higher trade barriers here. The rationale behind these opposite results lies in the differences between multilateral and unilateral trade barriers. Multilateral increases in trade barriers encourage countries to diversify their trading partners and learn from each other, thus alleviating some of the global losses. In contrast, unilateral increases in trade barriers do not directly affect other countries' trade with China but rather reduce China's knowledge stock by restricting its ability to learn from the United States. Although Chinese goods become relatively more expensive following the tariff hike, other countries continue to import similar shares of goods from China. As a result, even though the higher barriers are directed solely at China, many other countries also witness a decline in welfare.

If we focus on China and the US, it is evident that China would gain 0.15% in real income if there is a 25% increase in US tariffs, while the US would experience a loss of around 2.22% when both diffusion and adoption are present. Conversely, if there is no diffusion, both China and the US would see their real incomes drop by 1.92%. Without adoption, China's real income would increase by 0.06% while the US would lose 2.19%. Finally, in a world where either channel works, China would gain 3.2% but the US would lose 2.82%. These counter-intuitive results come partly from the fact that these evaluations are ex-post, which rely on the real expenditure structures and the exogenous trade balances observed in the data. The price elasticity of the US importers is low and thus they do not shift away from purchasing Chinese goods. Although the tariff targets China specifically, many other countries would also suffer from lower welfare, aligning with the bystander effects identified in Fajgelbaum et al. (2021).

In the case of higher fixed costs, the US would gain 0.16% if adoption is not present. In the other scenarios, the US's welfare still decreases, albeit to a lesser extent than with higher tariffs. However, regardless of the trade barrier imposed by the US, global welfare experiences losses as long as either learning channel is allowed.

Table A.8 (columns 1-6) illustrates the changes in trade margins resulting from higher US tariffs. An intriguing discovery is that in all economies, there would be more exporters when both diffusion and adoption are present in response to the US imposing higher tariffs on Chinese imports (column (1)). The United States experiences the most significant increase in the number of exporters. Without diffusion, most economies still witness an increase in the number of exporters, with the exceptions being China, Canada, and Switzerland. However, the losses in exporter numbers are much more pronounced when adoption is absent. Turning to the intensive margins, the US's exporters achieve higher sales in all cases, while Chinese exporters see lower sales.

However, the introduction of higher fixed costs (columns 7-12) has different effects, reducing both margins for the US and China when both channels are at work. The results diverge when we focus on cases without diffusion, where the US's margins increase in both the extensive and intensive aspects. Nonetheless, effects without adoption consistently lead to substantial reductions in all margins due to the heightened trade costs.

8 Conclusion

I investigate the impact of knowledge diffusion and foreign technology adoption on the gains from trade in a novel multi-country, multi-sector dynamic monopolistic trade model. Based on comprehensive Chinese firm-level data, I present stylized facts that heterogeneous firms learn foreign knowledge in different ways. While importers rely mainly on knowledge diffusion, exporters tend to purchase foreign technologies to reduce their production costs. Informed by these empirical findings, I embed learning from sellers and technology adoption into a general equilibrium model.

In the model, the interaction between diffusion and adoption amplifies the gains from trade. Diffusion increases the knowledge stocks in each economy and thus enables more firms to overcome the fixed costs to become exporters. As the number of exporters increases, more firms benefit from technology adoption and reduce their production costs. This further increases welfare. The model also generates dynamic changes in trade margins after trade liberalization due to the changes in the knowledge accumulation process, which is absent from a standard Melitz trade model. I provide reduced-form evidence supporting the model by leveraging China's WTO accession as an exogenous shock to firms.

After estimating the model parameters using global bilateral trade flow data, I conduct a series of counterfactuals to quantify the effects of international knowledge diffusion and technology adoption. The results show that global welfare would be 3.6% lower without both channels from 2000 to 2007. Ignoring diffusion leads to lower welfare in all economies. By contrast, technology adoption does not affect all economies in the same way: although the global welfare would be lower without adoption, some developed countries would benefit.

The interaction between diffusion and adoption is quantitatively important: on average, adoption amplifies the welfare gains from diffusion by about 70%. This suggests that ignoring adoption can lead to an overestimation of the importance of diffusion in models attributing all gains to the latter. Thus, to provide an accurate evaluation of welfare gains or losses, it is imperative to consider both effects, recognizing their interconnected nature in shaping the outcomes of international trade.

Policy counterfactuals show that multilateral and unilateral trade barriers have distinct implications when both learning channels are at work. While diffusion and adoption mitigate the losses from multilateral trade barriers, they amplify the losses from unilateral ones.

These findings carry important implications for policymakers, emphasizing the significance of facilitating knowledge diffusion and technology adoption in promoting innovation and increasing welfare. Moreover, both channels must be considered when designing trade policies.

References

Aghion, Philippe, Antonin Bergeaud, Matthieu Lequien, and Marc Melitz, "The heterogeneous impact of market size on innovation: evidence from French firm-level exports," *Review of Economics and Statistics*, 2022, pp. 1–56.

Antras, Pol, Teresa Fort, and Felix Tintelnot, "The margins of global sourcing: Theory and evidence from US firms," *American Economic Review*, 2017, *107* (9), 2514–2564.

Asquith, Brian, Sanjana Goswami, David Neumark, and Antonio Rodriguez-Lopez, "US job flows and the China shock," *Journal of International Economics*, 2019, *118*, 123–137.

Aw, Bee Yan, Mark Roberts, and Daniel Yi Xu, "R&D investment, exporting, and productivity dynamics," *American Economic Review*, 2011, *101* (4), 1312–1344.

Baldwin, Richard and Rikard Forslid, "Trade liberalization with heterogeneous firms," *Review of Development Economics*, 2010, *14* (2), 161–176.

Bloom, Nicholas, Mirko Draca, and John Van Reenen, "Trade induced technical change? The impact of Chinese imports on innovation, IT and productivity," *The Review of Economic Studies*, 2016, *83* (1), 87–117.

Brandt, Loren, Johannes Van Biesebroeck, and Yifan Zhang, "Creative accounting or creative destruction? Firm-level productivity growth in Chinese manufacturing," *Journal of Development Economics*, 2012, *97* (2), 339–351.

_ , _ , Luhang Wang, and Yifan Zhang, "WTO accession and performance of Chinese manufacturing firms," *American Economic Review*, 2017, *107* (9), 2784–2820.

Broda, Christian and David Weinstein, "Globalization and the gains from variety," *The Quarterly Journal of Economics*, 2006, *121* (2), 541–585.

Buera, Francisco and Ezra Oberfield, "The global diffusion of ideas," *Econometrica*, 2020, *88* (1), 83–114.

Bustos, Paula, "Trade liberalization, exports, and technology upgrading: Evidence on the impact of MERCOSUR on Argentinian firms," *American Economic Review*, 2011, *101* (1), 304–340.

Cai, Jie, Nan Li, and Ana Maria Santacreu, "Knowledge diffusion, trade, and innovation across countries and sectors," *American Economic Journal: Macroeconomics*, 2022, *14* (1), 104–45.

Cai, Sheng, Lorenzo Caliendo, Fernando Parro, and Wei Xiang, "Mechanics of spatial growth," *NBER Working Paper 30579*, 2022.

Caliendo, Lorenzo and Fernando Parro, "Estimates of the trade and welfare effects of NAFTA," *The Review of Economic Studies*, 2015, *82* (1), 1–44.

_ **, Maximiliano Dvorkin, and Fernando Parro**, "Trade and labor market dynamics: General equilibrium analysis of the China trade shock," *Econometrica*, 2019, *87* (3), 741–835.

Chaney, Thomas, "Distorted gravity: the intensive and extensive margins of international trade," *American Economic Review*, 2008, *98* (4), 1707–1721.

Chen, Zhiyuan, Jie Zhang, and Wenping Zheng, "Import and innovation: Evidence from Chinese firms," *European Economic Review*, 2017, *94*, 205–220.

David, David, David Dorn, and Gordon Hanson, "The China syndrome: Local labor market effects of import competition in the United States," *American Economic Review*, 2013, *103* (6), 2121–68.

Dekle, Robert, Jonathan Eaton, and Samuel Kortum, "Unbalanced trade," *American Economic Review*, 2007, *97* (2), 351–355.

_ **,** _ **, and** _ , "Global rebalancing with gravity: Measuring the burden of adjustment," *IMF Staff Papers*, 2008, *55* (3), 511–540.

Dix-Carneiro, Rafael, "Trade liberalization and labor market dynamics," *Econometrica*, 2014, *82* (3), 825–885.

_ **and Brian K Kovak**, "Trade liberalization and regional dynamics," *American Economic Review*, 2017, *107* (10), 2908–2946.

Edmond, Chris, Virgiliu Midrigan, and Daniel Xu, "Competition, markups, and the gains from international trade," *American Economic Review*, 2015, *105* (10), 3183–3221.

Egger, Peter and Jiaqing Zhu, "The US–Chinese trade war: an event study of stock-market responses," *Economic Policy*, 2020, *35* (103), 519–559.

Elliott, Robert, Liza Jabbour, and Liyun Zhang, "Firm productivity and importing: Evidence from Chinese manufacturing firms," *Canadian Journal of Economics*, 2016, *49* (3), 1086–1124.

Erten, Bilge and Jessica Leight, "Exporting out of agriculture: the impact of WTO accession on structural transformation in China," *Review of Economics and Statistics*, 2021, *103* (2), 364–380.

Fajgelbaum, Pablo and Amit Khandelwal, "Measuring the unequal gains from trade," *The Quarterly Journal of Economics*, 2016, *131* (3), 1113–1180.

—, **Pinelopi K Goldberg, Patrick J Kennedy, Amit Khandelwal, and Daria Taglioni**, "The US-China trade war and global reallocations," *NBER Working Paper 29562*, 2021.

Feenstra, Robert, John Romalis, and Peter Schott, "US imports, exports, and tariff data, 1989-2001," 2002.

Goldschlag, Nathan, Travis Lybbert, and Nikolas Zolas, "An 'algorithmic links with probabilities' crosswalk for USPC and CPC patent classifications with an application towards industrial technology composition," *Economics of Innovation and New Technology*, 2019, pp. 1–21.

Handley, Kyle and Nuno Limão, "Policy uncertainty, trade, and welfare: Theory and evidence for China and the United States," *American Economic Review*, 2017, *107* (9), 2731–2783.

He, Zi-Lin, Tony Tong, Yuchen Zhang, and Wenlong He, "A database linking Chinese patents to China's census firms," *Scientific data*, 2018, *5* (1), 1–16.

Hsieh, Chang-Tai and Peter Klenow, "Misallocation and manufacturing TFP in China and India," *The Quarterly Journal of Economics*, 2009, *124* (4), 1403–1448.

— **and Ralph Ossa**, "A global view of productivity growth in China," *Journal of International Economics*, 2016, *102*, 209–224.

Hu, Albert, Gary Jefferson, and Qian Jinchang, "R&D and technology transfer: Firm-level evidence from Chinese industry," *Review of Economics and Statistics*, 2005, *87* (4), 780–786.

Huang, Xianhai, Yi Wang, Zhujun Zhu, and Xueyin Song, "Quality of imported intermediates, innovation behaviour and markups: Firm-level evidence from China," *The World Economy*, 2022, *45* (9), 2796–2819.

Huang, Yasheng, Li Jin, and Yi Qian, "Does ethnicity pay? Evidence from overseas Chinese FDI in China," *Review of Economics and Statistics*, 2013, *95* (3), 868–883.

Hummels, David, Rasmus Jørgensen, Jakob Munch, and Chong Xiang, "The wage effects of offshoring: Evidence from Danish matched worker-firm data," *American Economic Review*, 2014, *104* (6), 1597–1629.

Impullitti, Giammario and Omar Licandro, "Trade, firm selection and innovation: the competition channel," *The Economic Journal*, 2018, *128* (608), 189–229.

Lileeva, Alla and Daniel Trefler, "Improved access to foreign markets raises plant-level productivity... for some plants," *The Quarterly Journal of Economics*, 2010, *125* (3), 1051–1099.

Lind, Nelson and Natalia Ramondo, "Global innovation and knowledge diffusion," *NBER Working Paper 29629*, 2022.

_ **and** _ , "Trade with correlation," *American Economic Review*, 2023, *113* (2), 317–353.

Liu, Qing and Hong Ma, "Trade policy uncertainty and innovation: Firm level evidence from China's WTO accession," *Journal of International Economics*, 2020, *127*, 103387.

_ **, Ruosi Lu, Yi Lu, and Tuan Anh Luong**, "Import competition and firm innovation: Evidence from China," *Journal of Development Economics*, 2021, *151*, 102650.

Lu, Xiaodong and Yujun Lian, "Estimation of total factor productivity of industrial enterprises in China: 1999–2007 (in Chinese)," *China Economic Quarterly*, January 2012, *11* (2).

Lu, Yi and Linhui Yu, "Trade liberalization and markup dispersion: evidence from China's WTO accession," *American Economic Journal: Applied Economics*, 2015, *7* (4), 221–253.

Melitz, Marc, "The impact of trade on intra-industry reallocations and aggregate industry productivity," *Econometrica*, 2003, *71* (6), 1695–1725.

Olley, Steven and Ariel Pakes, "The dynamics of productivity in the telecommunications equipment industry," *Econometrica*, 1996, *64* (6), 1263–1297.

Perla, Jesse, Christopher Tonetti, and Michael E Waugh, "Equilibrium technology diffusion, trade, and growth," *American Economic Review*, 2021, *111* (1), 73–128.

Pierce, Justin and Peter Schott, "The surprisingly swift decline of US manufacturing employment," *American Economic Review*, 2016, *106* (7), 1632–1662.

Sampat, Bhaven, "USPTO Patent and Citation Data," *Harvard Dataverse*, 2011.

Sampson, Thomas, "Dynamic selection: an idea flows theory of entry, trade, and growth," *The Quarterly Journal of Economics*, 2016, *131* (1), 315–380.

Samuelson, Paul, "The transfer problem and transport costs, II: Analysis of effects of trade impediments," *The Economic Journal*, 1954, *64* (254), 264–289.

Santacreu, Ana Maria Ravikumar and Michael Sposi, "Capital accumulation and dynamic gains from trade," *Journal of International Economics*, 2019, *119*, 93–110.

Shapiro, Joseph and Reed Walker, "Why is pollution from US manufacturing declining? The roles of environmental regulation, productivity, and trade," *American Economic Review*, 2018, *108* (12), 3814–3854.

Shim, Younghun, "From adoption to innovation: State-dependent technology policy in developing countries." PhD dissertation, The University of Chicago 2023.

Shu, Pian and Claudia Steinwender, "The impact of trade liberalization on firm productivity and innovation," *Innovation Policy and the Economy*, 2019, *19* (1), 39–68.

Song, Zheng, Kjetil Storesletten, and Fabrizio Zilibotti, "Growing like China," *American Economic Review*, 2011, *101* (1), 196–233.

Topalova, Petia and Amit Khandelwal, "Trade liberalization and firm productivity: The case of India," *Review of Economics and Statistics*, 2011, *93* (3), 995–1009.

Vishwasrao, Sharmila and William Bosshardt, "Foreign ownership and technology adoption: evidence from Indian firms," *Journal of Development Economics*, 2001, *65* (2), 367–387.

Wooldridge, Jeffrey, "On estimating firm-level production functions using proxy variables to control for unobservables," *Economics Letters*, 2009, *104* (3), 112–114.

Xu, Jin, "Profitability and capital structure: Evidence from import penetration," *Journal of Financial Economics*, 2012, *106* (2), 427–446.

Yasar, Mahmut, Rafal Raciborski, and Brian Poi, "Production function estimation in Stata using the Olley and Pakes method," *The Stata Journal*, 2008, *8* (2), 221–231.

Yu, Miaojie, "Processing trade, tariff reductions and firm productivity: Evidence from Chinese firms," *The Economic Journal*, 2015, *125* (585), 943–988.

Appendix A Additional tables

Table A.1: Economy List

ID	Abbreviation	Full name
1	AUS	Australia
2	AUT	Austria
3	BEL	Belgium
4	BRA	Brazil
5	CAN	Canada
6	CHE	Switzerland
7	CHN	China
8	DEU	Germany
9	DNK	Denmark
10	ESP	Spain
11	FIN	Finland
12	FRA	France
13	GBR	United Kingdom
14	ITA	Italy
15	JPN	Japan
16	KOR	Korea
17	NLD	Netherlands
18	NOR	Norway
19	SWE	Sweden
20	TWN	Taiwan
21	USA	United States
22	ROW	Rest of the world

Notes: This table shows the 22 economies used in this paper. The 21 economies were chosen due to the availability of their patent data. On average, they accounted for 98% of global trade from 2000 to 2010.

Table A.2: Sector List

ID	Sector	Full name	ID in WIOT
1	Agr	Crop and animal production, hunting and related service activities; Forestry and logging; Fishing and aquaculture	r1-r3
2	Mining	Mining and quarrying	r4
3	Food	Manufacture of food products, beverages and tobacco products	r5
4	Textiles	Manufacture of textiles, wearing apparel and leather products	r6
5	Wood	Manufacture of wood and of products of wood and cork, except furniture; Manufacture of articles of straw and plaiting materials	r7
6	Paper	Manufacture of paper and paper products; Printing and reproduction of recorded media	r8-r9
7	Coke	Manufacture of coke and refined petroleum products	r9
8	Chemical	Manufacture of chemicals and chemical products	r10
9	Pharmaceutical	Manufacture of basic pharmaceutical products and pharmaceutical preparations	r11
10	Rubber&Plastic	Manufacture of rubber and plastic products	r12
11	Mineral	Manufacture of other non-metallic mineral products	r13
12	Basic Metal	Manufacture of basic metals	r14
13	Metal products	Manufacture of fabricated metal products, except machinery and equipment	r15
14	Computer	Manufacture of computer, electronic and optical products	r16
15	Electrical	Manufacture of electrical equipment	r17
16	Machinery	Manufacture of machinery and equipment n.e.c.	r18
17	Transport	Manufacture of motor vehicles, trailers and semi-trailers; Manufacture of other transport equipment	r19-r20
18	Other manufacturing	Manufacture of furniture; other manufacturing	r21
19	Other activities	Other activities	r22-r56

Notes: This table shows the 19 sectors used in this paper and the sectoral ID in the World Input Output Table.

Table A.3: Parameters for simulation exercise one

	Sector 1	Sector 2
θ^j	8	8
γ_{nn}^j	0.25	0.25
γ_{ni}^j	0.25	0.25
γ_n^j	0.5	0.5
σ^j	5	5
Δ^j	0.3	0.3
ρ^j	0.5	0.5
π_{ni}^j	0.8	0.8
π_{nn}^j	0.2	0.2
w_n	40	40
L_n	1	1
α^j	0.283	0.116
X_n^j	50	30
m_0^j	0.5	0.5
$T_{1,t=0}^j$	0.6	0.6
$T_{2,t=0}^j$	0.9	0.9

Notes: This table shows parameters used in the two-country simulation exercise one in section 4.3.1.

Table A.4: Parameters for simulation exercise two

	Sector 1	Sector 2		Sector 1	Sector 2	
θ^j	8	8	$T_{1,t=0}^j$	0.2	0.2	Case1
γ_{nn}^j	0.25	0.25	$T_{2,t=0}^j$	0.9	0.9	
γ_{ni}^j	0.25	0.25	$T_{1,t=0}^j$	0.4	0.4	Case2
γ_n^j	0.5	0.5	$T_{2,t=0}^j$	0.9	0.9	
σ^j	5	5	$T_{1,t=0}^j$	0.6	0.6	Case3
Δ^j	0.5	0.5	$T_{2,t=0}^j$	0.9	0.9	
ρ^j	0.5	0.5	$T_{1,t=0}^j$	0.8	0.8	Case4
π_{ni}^j	0.8	0.8	$T_{2,t=0}^j$	0.9	0.9	
π_{nn}^j	0.2	0.2				
w_n	40	40				
L_n	1	1				
α^j	0.283	0.116				
X_n^j	50	30				
m_0^j	0.5	0.5				

Notes: This table shows parameters used in the two-country simulation exercise two in section 4.3.1.

Table A.5: Welfare changes without the WTO accession

	Higher tariffs				Higher fixed costs			
	With both	Without adoption	Without diffusion	Without both	With both	Without adoption	Without diffusion	Without both
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
AUS	-3.19%	-3.03%	-3.41%	-2.71%	-0.93%	-0.87%	-1.05%	-0.51%
AUT	-0.30%	-0.79%	-0.97%	-0.66%	-0.37%	-0.17%	0.01%	-0.13%
BEL	-1.78%	-2.03%	-2.86%	-2.36%	-0.77%	-0.38%	-0.51%	-0.68%
BRA	-2.23%	-1.53%	-0.68%	-1.41%	0.03%	-1.15%	-0.75%	-0.28%
CAN	-1.82%	-1.96%	-2.41%	-1.96%	-0.45%	-0.24%	-0.50%	-0.43%
CHE	-0.49%	-1.01%	-1.29%	-0.88%	-0.40%	0.03%	-0.04%	0.04%
CHN	-10.76%	-15.06%	-7.19%	-17.15%	0.96%	-1.61%	-0.81%	0.42%
DEU	-1.38%	-1.42%	-1.87%	-1.26%	-0.59%	-0.17%	-0.37%	-0.07%
DNK	-1.01%	-1.31%	-1.64%	-1.32%	-0.57%	-0.23%	-0.35%	-0.33%
ESP	-1.09%	-1.01%	-1.09%	-0.85%	-0.30%	-0.33%	-0.42%	-0.16%
FIN	-0.94%	-1.08%	-1.10%	-0.96%	-0.32%	-0.34%	-0.26%	-0.20%
FRA	-1.20%	-1.32%	-1.80%	-1.23%	-0.61%	-0.16%	-0.37%	-0.16%
GBR	-0.94%	-1.51%	-2.17%	-1.58%	-0.70%	0.16%	-0.38%	-0.02%
ITA	-1.41%	-1.52%	-0.89%	-1.19%	-0.28%	-0.69%	-0.58%	-0.03%
JPN	-3.22%	-3.17%	-3.98%	-3.18%	-0.95%	-0.50%	-0.72%	-0.55%
KOR	-3.43%	-3.38%	-2.31%	-3.34%	-0.34%	-1.30%	-0.84%	-0.57%
NLD	-1.51%	-1.38%	-1.87%	-1.36%	-0.58%	-0.36%	-0.64%	-0.16%
NOR	-1.26%	-1.51%	-1.83%	-1.40%	-0.47%	-0.23%	-0.36%	-0.22%
SWE	-1.11%	-1.18%	-1.80%	-1.13%	-0.47%	-0.03%	-0.31%	-0.13%
TWN	-4.91%	-5.33%	-4.16%	-5.29%	-0.63%	-1.64%	-0.87%	-1.18%
USA	-1.58%	-1.54%	-2.31%	-1.63%	-0.61%	-0.04%	-0.25%	-0.40%
ROW	-2.51%	-2.28%	-1.35%	-2.00%	-0.18%	-1.05%	-0.91%	-0.15%
Global average	-2.18%	-2.47%	-2.23%	-2.49%	-0.43%	-0.51%	-0.51%	-0.27%

Notes: This table shows the relative changes in welfare (annual average) if China did not join the WTO. Columns (1)-(4) show the results when there are higher tariffs, while (5)-(8) show the results when there are higher fixed costs. See section 7.2 for details.

Table A.6: Changes in trade margins without the WTO accession

	Higher tariffs						Higher fixed costs					
	Number of exporters			Sales per exporter			Number of exporters			Sales per exporter		
	With both	No diffusion	No adoption	With both	No diffusion	No adoption	With both	No diffusion	No adoption	With both	No diffusion	No adoption
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
AUS	-0.26%	3.48%	-0.95%	1.96%	1.77%	1.99%	0.74%	-1.16%	-6.16%	0.56%	0.68%	0.54%
AUT	-2.36%	1.18%	-2.57%	-0.12%	-0.36%	0.22%	0.23%	-1.72%	-6.46%	0.08%	0.10%	0.20%
BEL	-2.59%	2.72%	-3.27%	-0.10%	0.80%	-0.15%	0.29%	-2.67%	-7.41%	0.20%	-0.20%	-0.77%
BRA	-2.11%	1.83%	-1.84%	-0.12%	0.23%	-0.05%	0.29%	-1.19%	-6.86%	0.04%	0.19%	-0.71%
CAN	-1.47%	2.14%	-2.11%	0.87%	0.70%	4.61%	0.59%	-1.35%	-1.74%	0.16%	0.33%	3.68%
CHE	-2.92%	1.19%	-2.64%	-0.22%	-0.20%	2.27%	-0.45%	-1.96%	-5.90%	-0.08%	-0.10%	1.90%
CHN	-9.90%	-6.50%	-2.94%	-8.75%	-8.76%	-1.06%	-2.84%	-4.15%	-9.61%	-0.08%	-0.09%	-0.34%
DEU	-1.50%	2.70%	-2.20%	0.59%	0.83%	1.19%	0.33%	-1.90%	-6.35%	0.13%	0.04%	0.39%
DNK	-2.00%	1.49%	-2.59%	0.17%	-0.19%	1.86%	0.64%	-1.46%	-4.05%	0.44%	0.20%	2.14%
ESP	-2.18%	2.06%	-2.96%	0.04%	0.27%	0.13%	0.30%	-2.06%	-7.01%	0.14%	-0.03%	-0.25%
FIN	-2.65%	1.31%	-3.25%	-0.43%	-0.37%	-0.31%	-0.07%	-2.07%	-7.42%	0.04%	-0.03%	-0.44%
FRA	-1.80%	2.80%	-2.81%	0.33%	0.76%	0.42%	0.30%	-2.21%	-6.90%	0.24%	-0.06%	-0.24%
GBR	-1.47%	2.24%	-2.09%	0.70%	0.81%	3.18%	0.55%	-1.71%	-4.13%	0.28%	-0.05%	2.38%
ITA	-2.07%	2.14%	-2.89%	0.13%	0.40%	0.35%	0.30%	-2.01%	-6.85%	0.16%	-0.02%	-0.09%
JPN	-0.34%	3.41%	-1.26%	1.86%	1.71%	2.15%	0.84%	-1.34%	-5.95%	0.44%	0.54%	0.65%
KOR	-1.63%	2.17%	-1.77%	0.91%	0.90%	1.27%	1.04%	-1.28%	-5.98%	0.58%	0.28%	0.45%
NLD	-1.95%	2.85%	-3.50%	0.22%	0.93%	1.80%	0.41%	-1.85%	-3.20%	0.09%	-0.03%	2.26%
NOR	-2.26%	1.94%	-2.41%	0.17%	0.35%	0.22%	0.69%	-1.62%	-6.72%	0.29%	0.02%	-0.19%
SWE	-1.98%	2.30%	-3.09%	0.24%	0.30%	0.43%	0.38%	-2.23%	-6.75%	0.12%	0.11%	-0.04%
TWN	-1.51%	2.52%	-1.72%	1.14%	1.33%	1.39%	1.36%	-1.55%	-5.55%	0.85%	0.13%	0.50%
USA	-1.53%	2.41%	-2.19%	0.68%	0.88%	0.51%	1.11%	-1.37%	-6.70%	0.56%	0.19%	-0.23%
ROW	-0.59%	3.34%	-0.95%	1.31%	1.73%	1.34%	0.93%	-1.23%	-6.56%	0.48%	0.27%	-0.22%

Notes: This table shows the relative changes in the number of exporters and sales per exporter if China did not join the WTO. Columns (1)-(6) show the results when there are higher tariffs, while (7)-(12) show the results when there are higher fixed costs. See section 7.2 for details.

Table A.7: Changes in welfare with higher US trade barriers

	Higher tariffs				Higher fixed costs			
	With both	Without adoption	Without diffusion	Without both	With both	Without adoption	Without diffusion	Without both
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
AUS	-0.50%	-0.83%	-0.55%	-0.57%	-0.04%	-0.35%	-0.01%	-0.03%
AUT	-0.58%	-0.81%	-0.02%	-1.05%	0.02%	-0.21%	0.08%	-0.06%
BEL	-0.93%	-0.93%	-0.65%	-1.64%	0.03%	0.05%	0.28%	-0.15%
BRA	0.45%	0.34%	-0.41%	1.65%	-0.33%	-1.22%	-0.44%	0.42%
CAN	-0.74%	-0.75%	-0.78%	-1.05%	0.00%	0.18%	0.00%	-0.18%
CHE	-0.76%	-0.96%	-0.08%	-1.35%	0.05%	0.13%	0.16%	-0.12%
CHN	0.15%	0.06%	-1.92%	3.20%	-0.76%	-2.52%	-0.93%	0.64%
DEU	-0.75%	-0.87%	-0.36%	-1.03%	-0.01%	0.08%	0.12%	-0.05%
DNK	-0.64%	-0.54%	-0.30%	-0.78%	-0.01%	0.10%	0.09%	-0.10%
ESP	-0.23%	-0.42%	-0.28%	-0.21%	-0.07%	-0.24%	-0.04%	0.03%
FIN	-0.38%	-0.53%	-0.23%	-0.44%	-0.08%	-0.30%	-0.04%	0.00%
FRA	-0.77%	-0.85%	-0.36%	-1.09%	0.05%	0.12%	0.11%	-0.09%
GBR	-1.20%	-1.16%	-0.61%	-1.85%	0.27%	0.68%	0.37%	-0.29%
ITA	-0.06%	-0.30%	-0.19%	0.55%	-0.18%	-0.62%	-0.23%	0.30%
JPN	-0.50%	-0.63%	-0.13%	-0.95%	0.06%	0.33%	0.11%	-0.13%
KOR	0.21%	0.21%	-0.44%	1.29%	-0.38%	-0.89%	-0.51%	0.32%
NLD	-0.64%	-0.61%	-0.56%	-0.57%	-0.07%	-0.05%	0.04%	0.00%
NOR	-0.56%	-0.77%	-0.18%	-0.95%	0.04%	-0.10%	0.17%	-0.11%
SWE	-0.63%	-0.69%	-0.34%	-1.04%	0.04%	0.27%	0.17%	-0.17%
TWN	-0.22%	-0.28%	-0.37%	0.43%	-0.42%	-1.06%	-0.48%	0.16%
USA	-2.22%	-2.19%	-1.92%	-2.82%	-0.33%	0.16%	-0.29%	-0.61%
ROW	0.28%	0.03%	-0.47%	1.24%	-0.29%	-0.81%	-0.34%	0.38%
Global average	-0.51%	-0.61%	-0.51%	-0.41%	-0.11%	-0.29%	-0.07%	0.01%

Notes: This table shows the relative changes in welfare (annual average) if the US increases tariffs or fixed costs by 25% for China only. Columns (1)-(4) show the results when there are higher tariffs, while (5)-(8) show the results when there are higher fixed costs. See section 7.3 for details.

Table A.8: Changes in trade margins with higher US trade barriers

	Higher tariffs						Higher fixed costs					
	Number of exporters			Sales per exporter			Number of exporters			Sales per exporter		
	With both	No diffusion	No adoption	With both	No diffusion	No adoption	With both	No diffusion	No adoption	With both	No diffusion	No adoption
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
AUS	1.66%	0.13%	-2.38%	-0.62%	0.27%	0.02%	-0.72%	0.09%	-6.27%	-1.03%	0.07%	0.23%
AUT	1.56%	0.00%	-2.57%	-0.88%	0.03%	-0.22%	-0.67%	0.11%	-6.13%	-0.89%	0.18%	0.29%
BEL	1.96%	0.34%	-2.16%	-1.64%	-0.06%	0.09%	-0.90%	-0.12%	-7.08%	-1.36%	-0.26%	-0.90%
BRA	1.73%	0.13%	-2.55%	-0.10%	0.09%	-0.61%	-0.90%	-0.02%	-7.19%	-1.25%	0.00%	-0.80%
CAN	1.71%	-0.03%	2.46%	-0.40%	0.27%	3.58%	-0.82%	-0.01%	-1.92%	-1.17%	-0.07%	3.48%
CHE	1.20%	-0.09%	2.41%	-1.85%	-0.40%	2.57%	-1.39%	-0.13%	-4.93%	-1.85%	-0.09%	2.23%
CHN	1.29%	-0.35%	-2.78%	-0.75%	-0.41%	-0.62%	-1.01%	-0.03%	-6.89%	-1.35%	-0.04%	-0.21%
DEU	1.67%	0.08%	-1.75%	-1.13%	-0.05%	0.54%	-0.94%	-0.11%	-6.16%	-1.14%	-0.01%	0.33%
DNK	1.67%	0.10%	-0.64%	-0.71%	-0.02%	1.15%	-0.67%	0.12%	-4.01%	-0.80%	0.32%	2.08%
ESP	1.73%	0.13%	-2.30%	-1.10%	-0.01%	0.04%	-0.89%	-0.04%	-6.82%	-1.18%	-0.03%	-0.30%
FIN	1.58%	0.22%	-2.37%	-0.99%	0.11%	-0.09%	-0.92%	-0.01%	-7.02%	-1.24%	0.00%	-0.39%
FRA	1.76%	0.26%	-1.91%	-1.34%	-0.04%	0.32%	-0.85%	0.02%	-6.72%	-1.21%	-0.07%	-0.41%
GBR	1.73%	0.06%	0.73%	-0.97%	-0.15%	2.60%	-0.74%	0.13%	-4.22%	-1.24%	-0.07%	2.04%
ITA	1.72%	0.17%	-2.19%	-1.06%	0.03%	0.22%	-0.84%	-0.02%	-6.69%	-1.17%	-0.05%	-0.20%
JPN	1.66%	0.00%	-2.37%	-0.78%	0.17%	0.17%	-0.61%	0.13%	-6.25%	-0.87%	0.13%	0.16%
KOR	2.15%	0.31%	-2.39%	-0.43%	0.07%	-0.05%	-0.68%	0.11%	-6.52%	-0.95%	0.14%	-0.01%
NLD	1.78%	0.05%	1.60%	-0.93%	-0.02%	2.69%	-0.88%	-0.08%	-3.29%	-1.09%	-0.02%	2.06%
NOR	2.00%	0.17%	-2.51%	-0.53%	-0.03%	0.02%	-0.86%	-0.03%	-6.92%	-1.17%	-0.04%	-0.24%
SWE	1.84%	0.11%	-2.11%	-1.36%	0.09%	0.06%	-0.89%	-0.02%	-6.52%	-1.17%	-0.01%	-0.11%
TWN	2.48%	0.57%	-1.94%	-0.62%	0.01%	0.01%	-0.61%	0.13%	-6.07%	-0.75%	0.26%	-0.05%
USA	3.36%	1.46%	-1.73%	0.37%	0.91%	0.68%	-0.37%	0.32%	-6.79%	-0.72%	0.25%	-0.37%
ROW	1.85%	0.15%	-2.16%	-0.67%	-0.10%	0.01%	-0.93%	-0.10%	-6.92%	-1.28%	-0.13%	-0.56%

Notes: This table shows the relative changes in the number of exporters and sales per exporter if the US increases tariffs or fixed costs by 25% for China only. Columns (1)-(6) show the results when there are higher tariffs, while (7)-(12) show the results when there are higher fixed costs. See section 7.3 for details.

Appendix B Additional figures

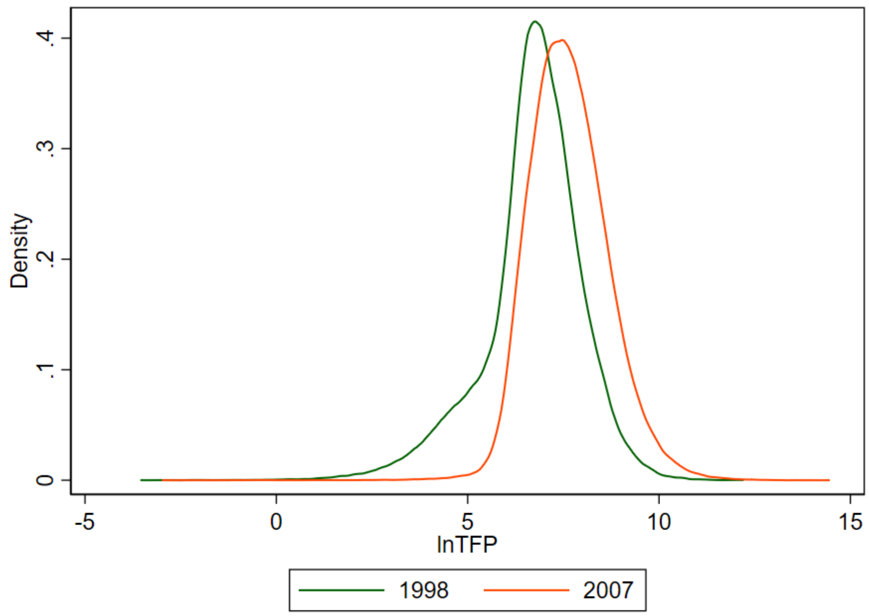

Figure B.1: Chinese manufacturing firms' TFP distribution

Notes: This figure shows the estimated TFP distribution following the method described in section 2.2.1 for Chinese manufacturing firms in 1998 and 2007.

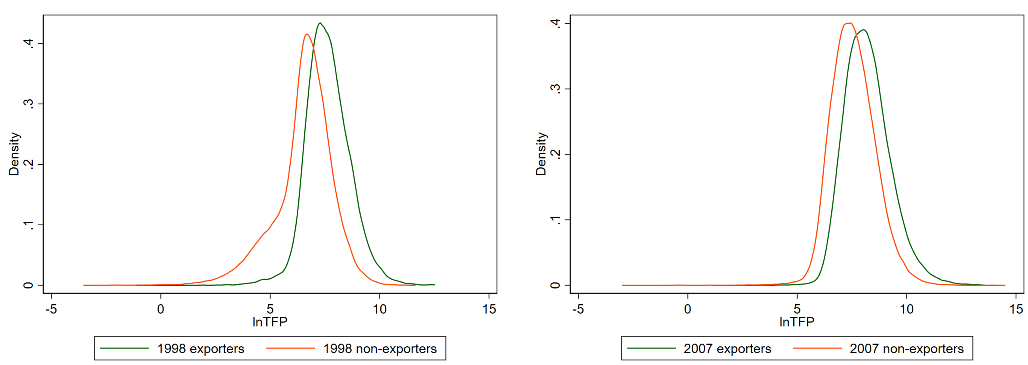

Figure B.2: Firm-level TFP of exporters and non-exporters

Notes: This figure shows the estimated TFP distribution following the method described in section 2.2.1 for Chinese manufacturing. The left panel compares the TFP distribution for exporters and non-exporters in 1998 and the right panel compares them in 2007.

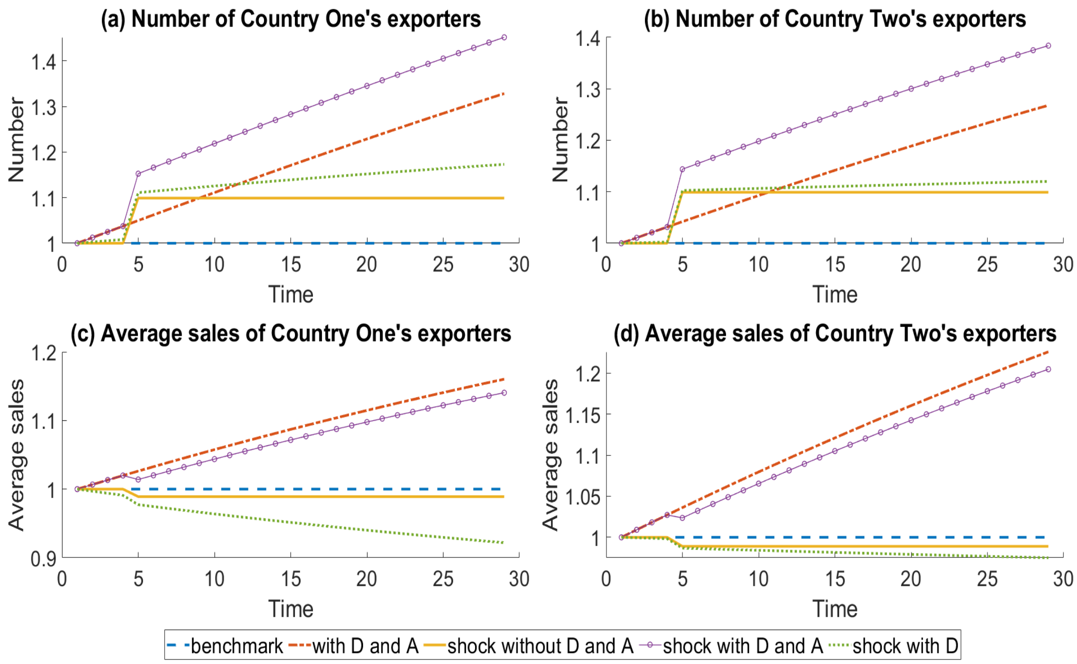

Figure B.3: Changes in trade margins when fixed costs drop at period five

Notes: This figure shows how a 10% bilateral fixed cost reduction in period five affects Country One (panels (a) and (c)) and Country Two's (panels (b) and (d)) number of exporters and sales per exporter. The dashed blue curves (benchmark) show the normalized number of exporters and sales when fixed costs do not change and there is no learning. The dash-dotted red curves (with D and A) show the changes when diffusion and adoption are both at work but fixed costs stay the same. The yellow solid lines (shock without D and A) show the changes when there is a 10% bilateral fixed cost reduction in period five without any learning. The dash-circled purple curves (shock with D and A) present the responses to the fixed cost reduction with diffusion and adoption. Finally, the dotted green curves (shock with D) show the responses if only diffusion is present. See section 4.3.1 for details.

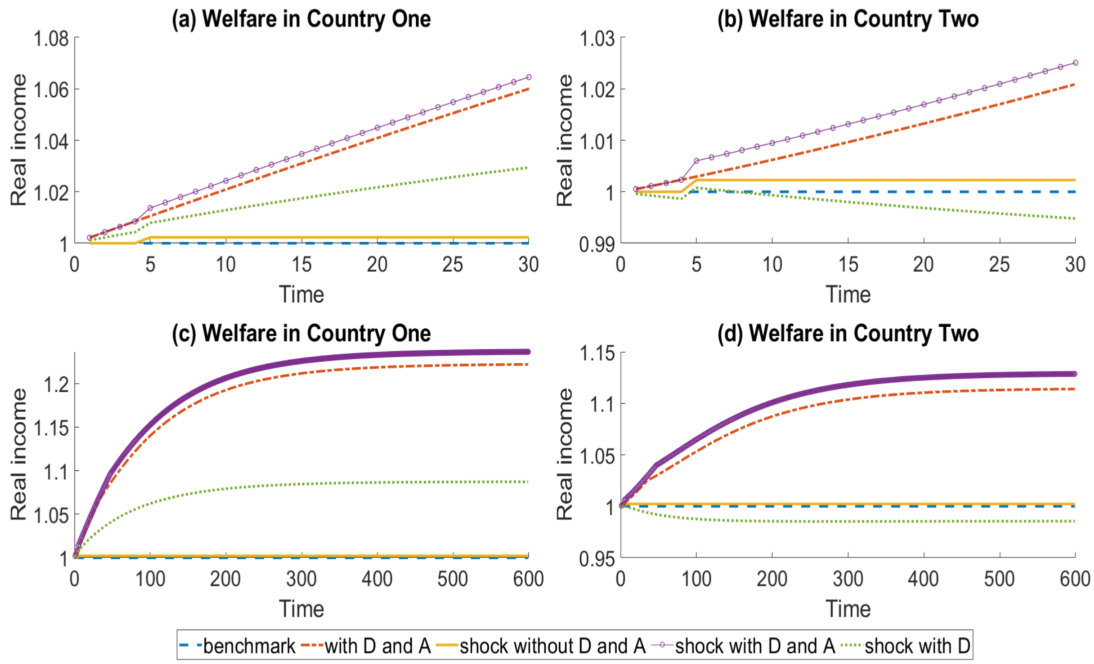

Figure B.4: Changes in welfare when fixed costs drop at period five

Notes: This figure shows how a 10% bilateral fixed cost reduction in period five affects Country One (panels (a) and (c)) and Country Two's (panels (b) and (d)) welfare in the short- and long-run. The dashed blue curves (benchmark) show the normalized number of exporters and sales when fixed costs do not change and there is no learning. The dash-dotted red curves (with D and A) show the changes when diffusion and adoption are both at work but fixed costs stay the same. The yellow solid lines (shock without D and A) show the changes when there is a 10% bilateral fixed cost reduction in period five without any learning. The dash-circled purple curves (shock with D and A) present the responses to the fixed cost reduction with diffusion and adoption. Finally, the dotted green curves (shock with D) show the responses if only diffusion is present. See section 4.3.1 for details.

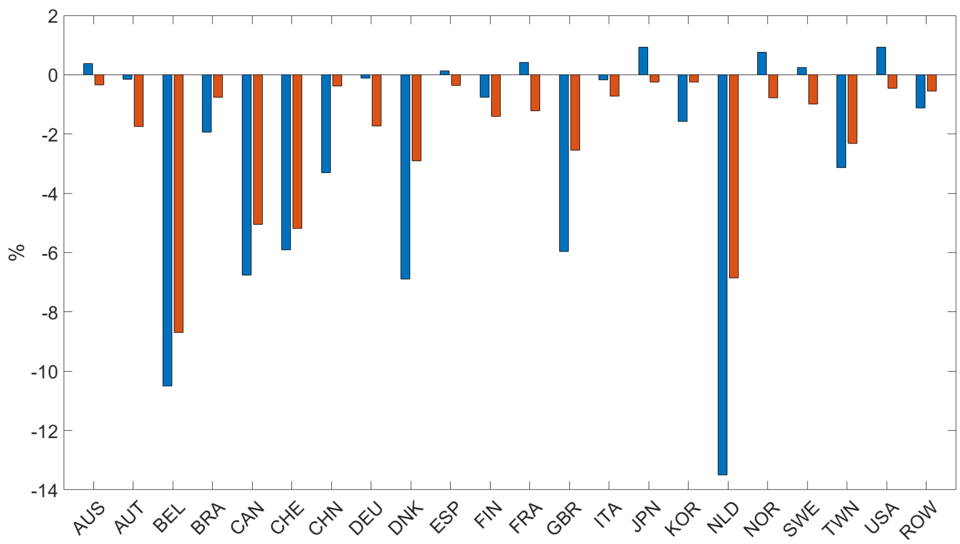

Figure B.5: Comparison of welfare without diffusion or adoption with the benchmark
Notes: This figure shows how welfare would change in a world without diffusion (orange bars) or without adoption (blue bars), compared to a world where both channels are at work (benchmark). See details in section 7.1.

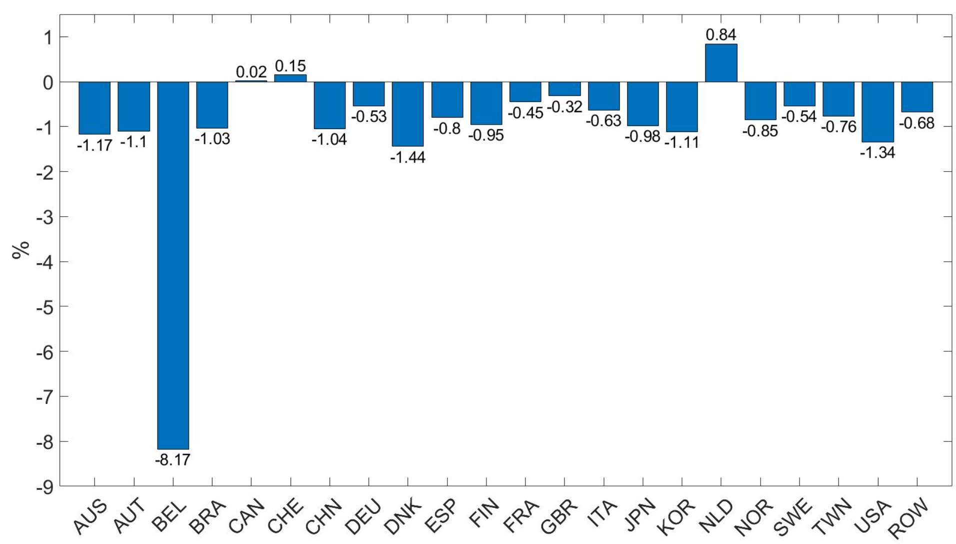

Figure B.6: Changes in national exporter number if there was no diffusion
Notes: This figure shows the changes in the number of exporters for each economy in a world without diffusion compared to a world where both channels are at work. The bars are the average value from 2000 to 2007. See details in section 7.1.

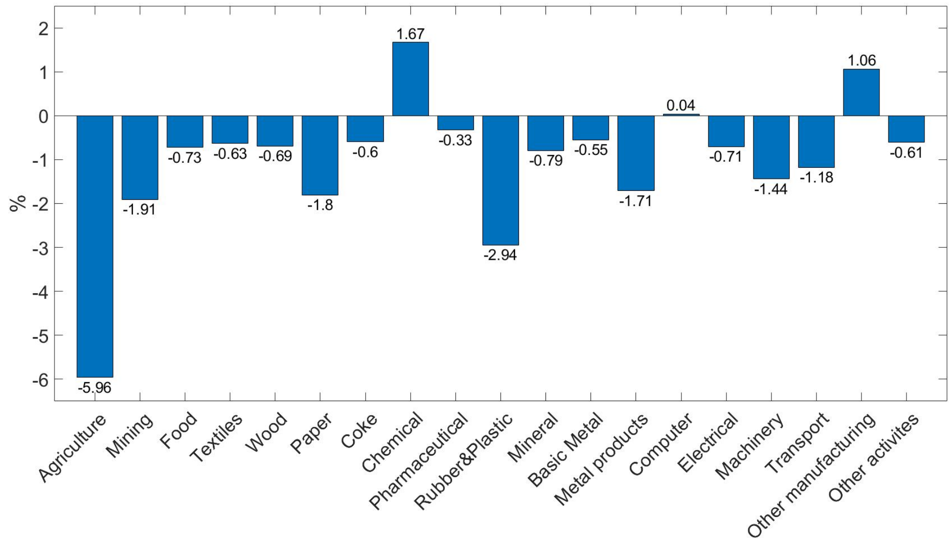

Figure B.7: Changes in sectoral exporter number if there was no diffusion

Notes: This figure shows the changes in the number of exporters for each sector in a world without diffusion compared to a world where both channels are at work. The bars are the average value from 2000 to 2007. See details in section 7.1.

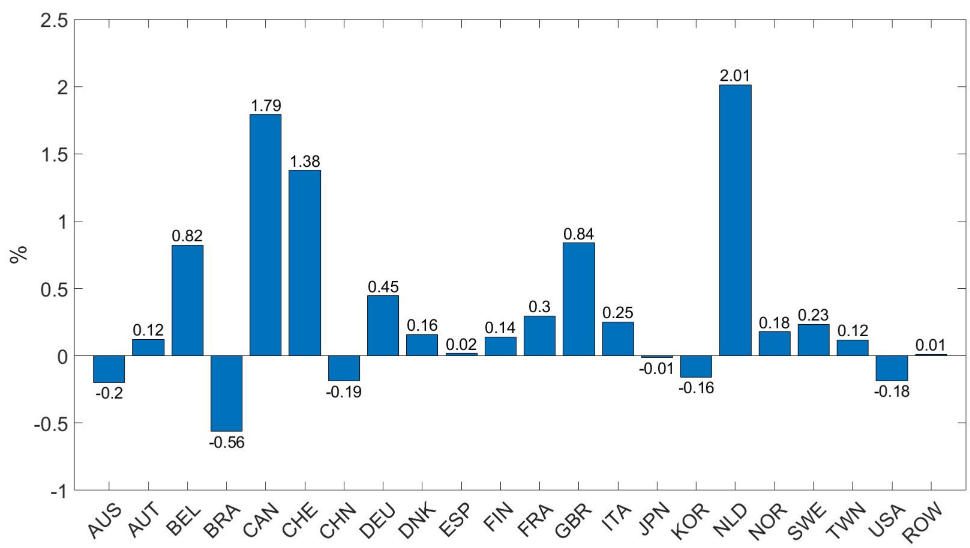

Figure B.8: Changes in the national sales per exporter if there was no diffusion

Notes: This figure shows the changes in the average sales per exporter for each economy in a world without diffusion compared to a world where both channels are at work. The bars are the average value from 2000 to 2007. See details in section 7.1.

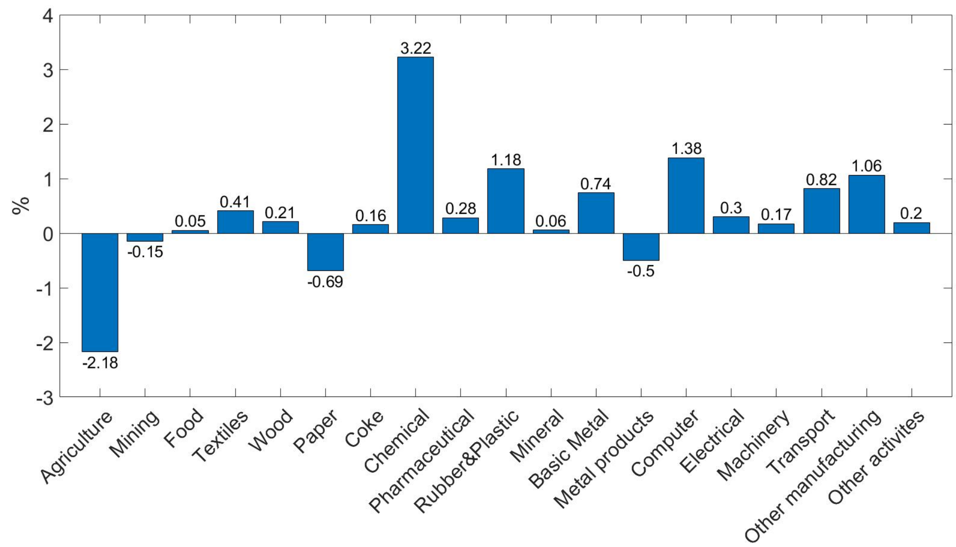

Figure B.9: Changes in the sectoral sales per exporter if there was no diffusion

Notes: This figure shows the changes in the average sales per exporter for each sector in a world without diffusion compared to a world where both channels are at work. The bars are the average value from 2000 to 2007. See details in section 7.1.

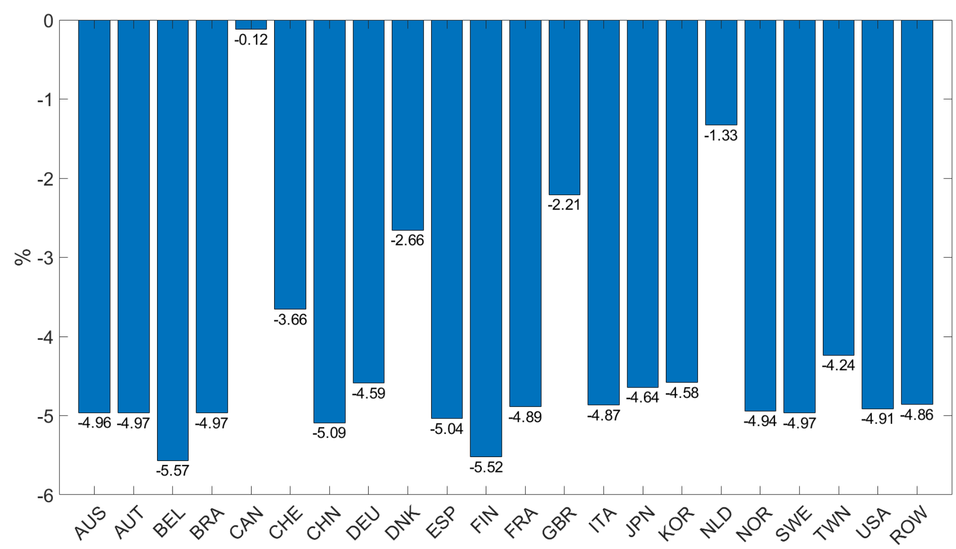

Figure B.10: Changes in national exporter number if there was no adoption

Notes: This figure shows the changes in the number of exporters for each economy in a world without adoption compared to a world where both channels are at work. The bars are the average value from 2000 to 2007. See details in section 7.1.

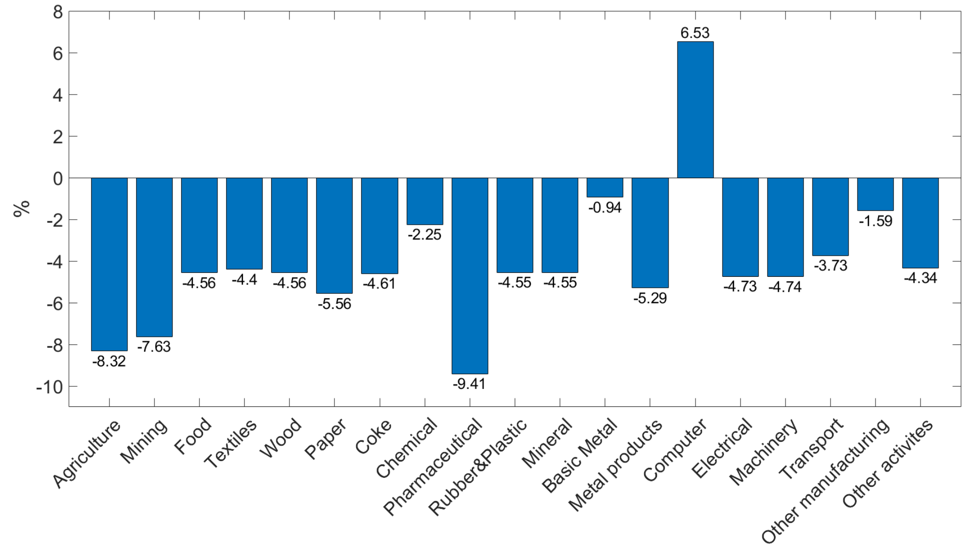

Figure B.11: Changes in sectoral exporter number if there was no adoption
Notes: This figure shows the changes in the number of exporters for each sector in a world without adoption compared to a world where both channels are at work. The bars are the average value from 2000 to 2007. See details in section 7.1.

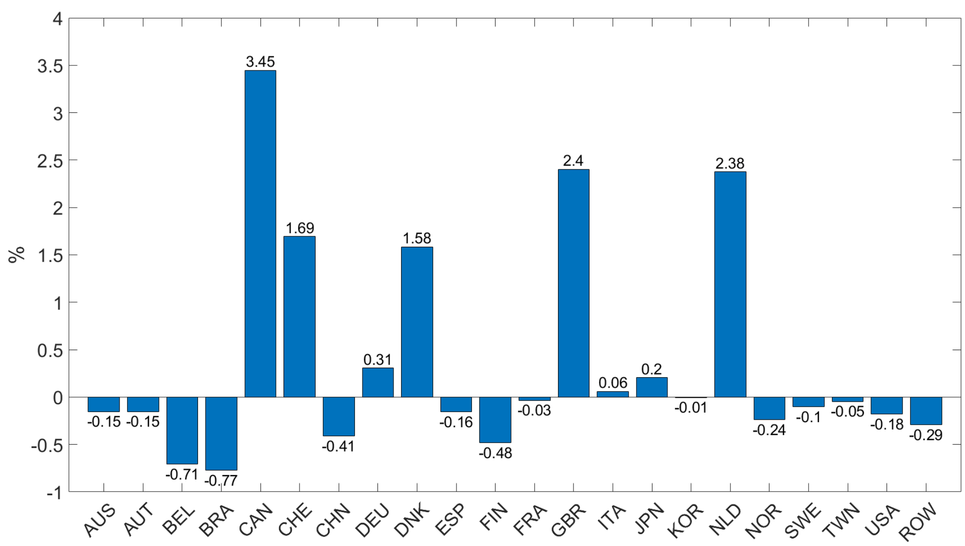

Figure B.12: Changes in national sales per exporter if there was no adoption
Notes: This figure shows the changes in the sales per exporter for each economy in a world without adoption compared to a world where both channels are at work. The bars are the average value from 2000 to 2007. See details in section 7.1.

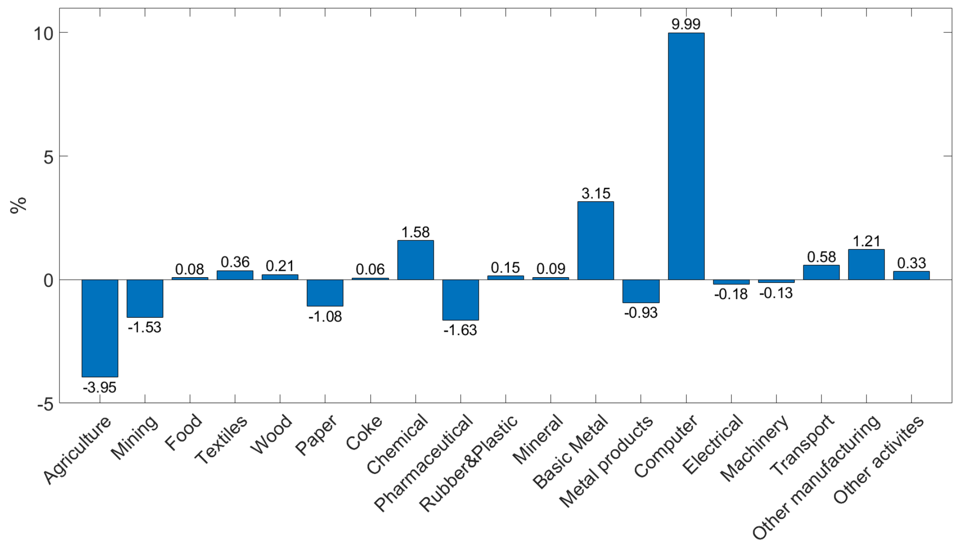

Figure B.13: Changes in sectoral sales per exporter if there was no adoption

Notes: This figure shows the changes in the sales per exporter for each sector in a world without adoption compared to a world where both channels are at work. The bars are the average value from 2000 to 2007. See details in section 7.1.

Appendix C Empirical appendix

C.1 Macro data and evidence

In this section, I present some additional evidence supporting the diffusion channel on the macro level. The patent data for each economy is from Sampat (2011), which contains all issued utility patents from the United States Patent and Trademark Office (USPTO) from January 1, 1975, to December 31, 2010.

Figure C.1 shows the number of patents issued in each economy annually. The United States and Japan had the highest number of patents annually, followed by Korea and Germany. While China started at a very low level, there was a clear increasing trend.

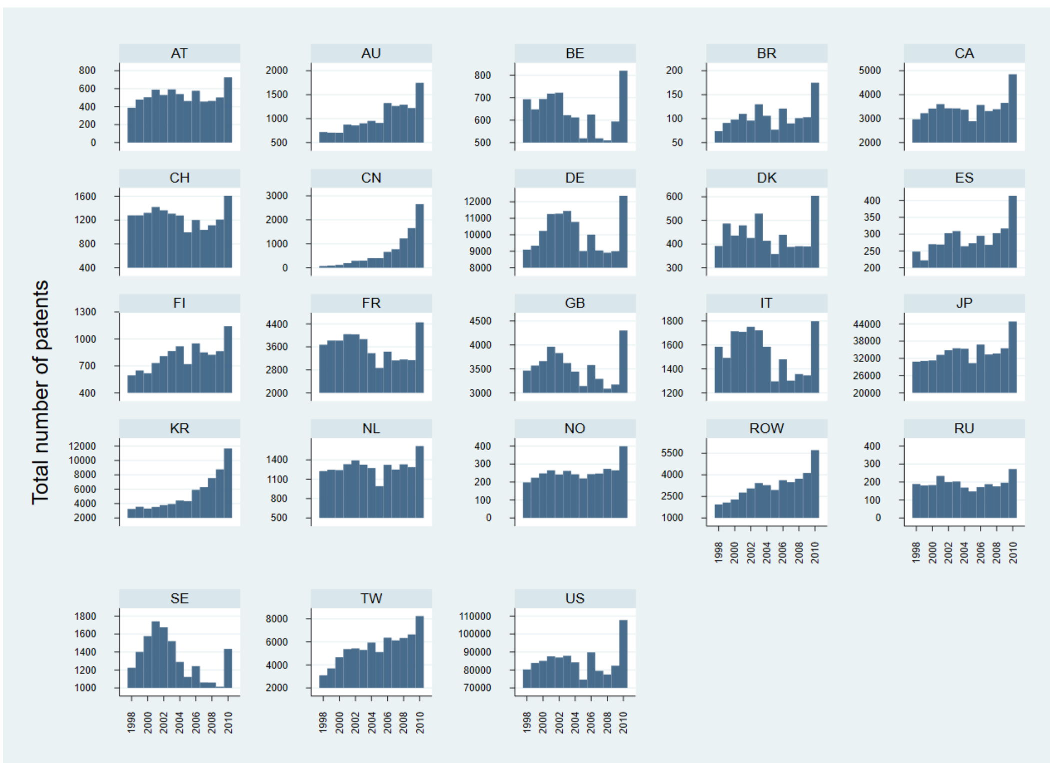

Figure C.1: Patents in each economy
Notes: This figure shows the number of patents issued in each economy annually from 1998 to 2010 using the data from Sampat (2011).

Figure C.2 presents the sectoral distribution of patents annually. Not surprisingly, the computer sector had the highest number of newly issued patents. Other sectors like electrical, machinery, and pharmaceutical all had relatively high numbers. Service, which was generally assumed to be non-tradable, also created a large number of patents.

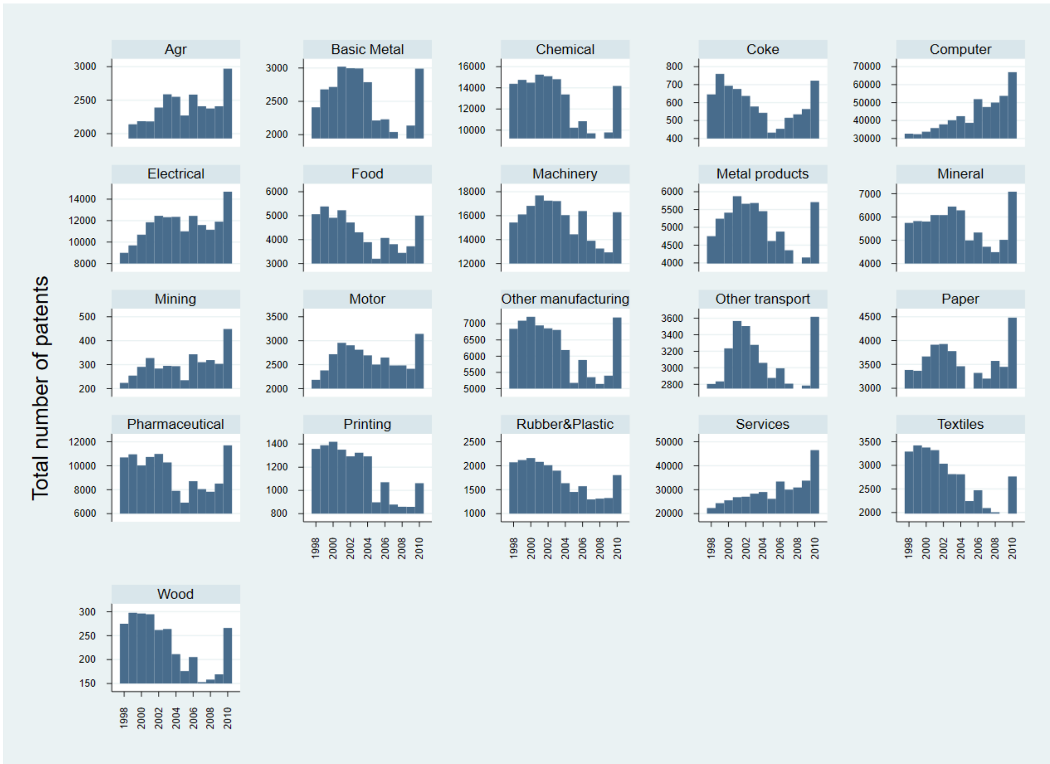

Figure C.2: Patents in each sector

Notes: This figure shows the number of patents issued in each sector annually from 1998 to 2010 using the data from Sampat (2011).

I find significant correlations between the annual growth of patents at time t and $t+1$, meaning that higher growth during this period leads to higher growth next period. To connect patents with trade, I construct the bilateral trade flows for the same span and aggregate them into the 20 sectors. The sectoral expenditure of each economy's spending on others' goods is calculated by summing all imports. Domestic sales are calculated by the total production from the WIOD minus the total exports. I plot the yearly average of sectoral imports, and the number of new patents in Figure C.4, and I find higher sectoral imports were generally correlated with more newly issued patents. To connect data to the hypothesis that knowledge is embedded in imports and importing from other countries contributes to domestic knowledge, I weigh each trading partner's patent knowledge stock by the import level and plot it against the new issue next year. There is also a significant correlation between new patents and import-weighted patent stocks (C.5). As patents could have different qualities, I adjust the number of patents by their weighted citation from the issued year until 2010,

$$Patent'^{j}_{nt} = Patent^{j}_{nt} * \frac{Citation^{j}_{nt}}{\sum_{i=1}^{N} Citation^{j}_{it}} \tag{41}$$

I find that the results are similar as before in Figure C.6.

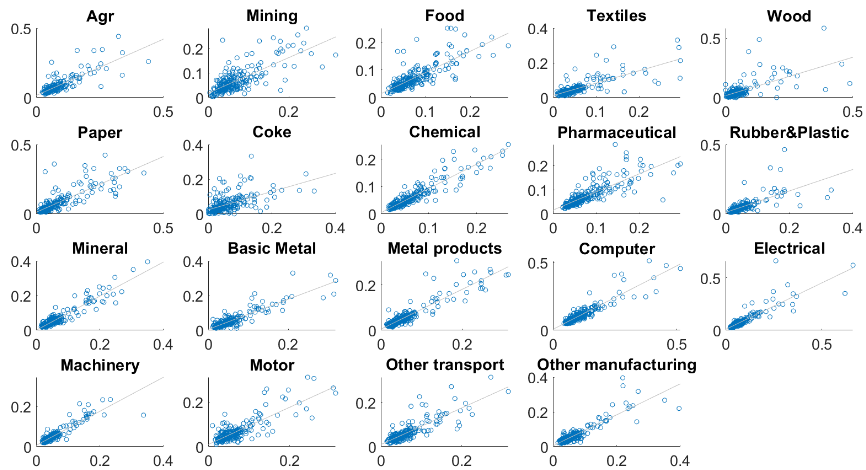

Figure C.3: Correlation between annual growth rates

Notes: Each subplot represents the correlation between growth at t and $t+1$ in one sector over the year 2000-2010. Each point indicates one specific observation in one economy in a certain year.

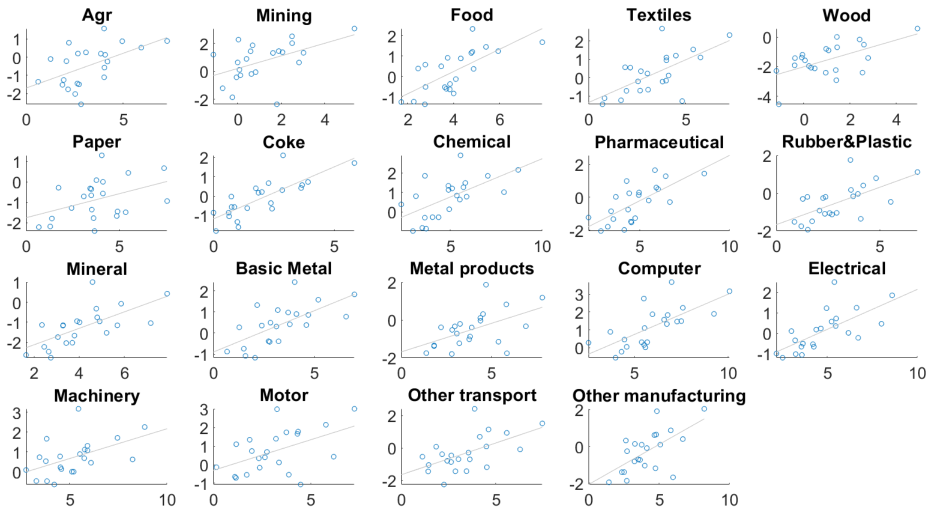

Figure C.4: Correlation between imports and newly granted patents

Notes: Each subplot represents the correlation between averaged imports and new patents in one sector from 2000 to 2010. The (log of) number of patents is on the x-axis, while the (log of) imports is on the y-axis.

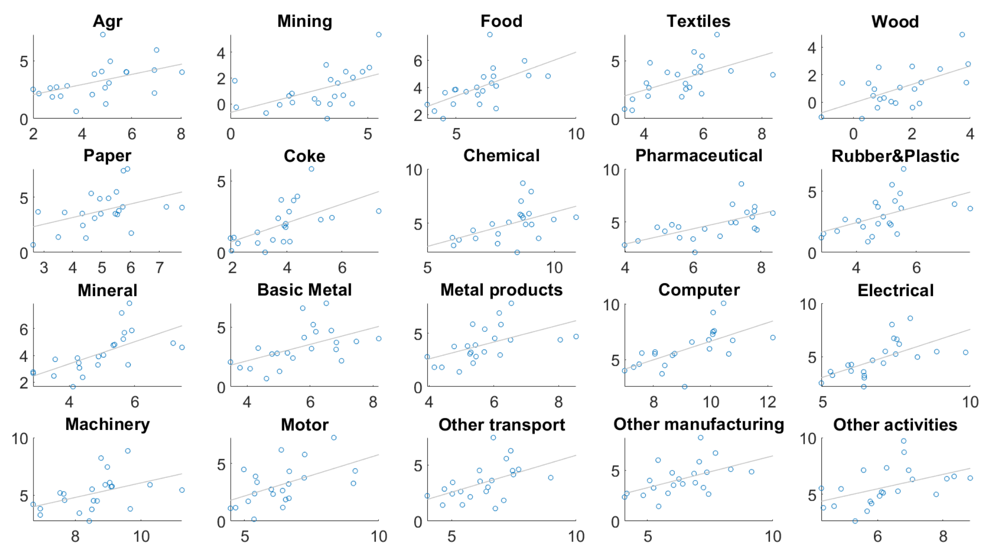

Figure C.5: Correlation between import-weighted stocks and newly granted patents

Notes: Each subplot represents the correlation between import-weighted patent stocks of all trading partners and new patents in one sector from 2000 to 2010. The (log of) number of new patents is on the y-axis, while the (log of) import-weighted stocks is on the x-axis.

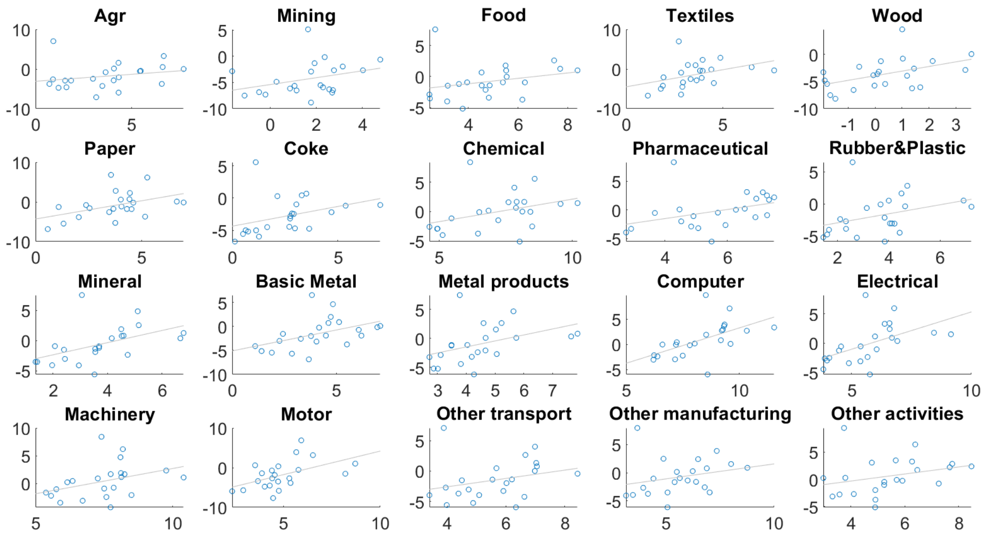

Figure C.6: Correlation between import-weighted patent stocks and citation-adjusted new patents
Notes: Each subplot represents the correlation between import-weighted patent stocks of all trading partners and citation-weighted patent stocks in one sector from 2000 to 2010. The (log of) number of citation-weighted new patent stocks is on the y-axis, while the import-weighted stocks is on the y-axis.

C.2 Micro data and evidence

In this appendix, I present how I constructed the firm-level panel used for the stylized facts and how I estimate the firm-level TFP. This section introduces four datasets: (1) The Annual Survey of Industrial Firms Database (ASIF), (2) the patent data from China's State Intellectual Property Administration (SIPO), (3) customs data from the General Administration of the Customs People's Republic of China (GACC) and the (4) expenditure on technical activities in from Statistical Yearbook on Science and Technology. The data-cleaning process for the main empirical results is described.

The Annual Survey of Industrial Firms Database (ASIF for short, 工业企业数据库 in Chinese) is published by the National Bureau of Statistics of China annually and contains all industrial firms with sales above 5 million RMB before 2012. These firms comprise more than 95% of total industrial output and 98% of industrial exports. As the most comprehensive database, it includes basic firm-level information (such as unique identifier, name, address, telephone, ownership structure, sector), production information (such as the number of employees, total sales, value-added, intermediate inputs) and balance sheet information (such as assets, capital stock, expenditure etc.) The database has been widely used in the literature (e.g. Hu et al. (2005), Hsieh and Klenow (2009), Song et al. (2011),Huang et al. (2013)).

C.3 Creating a panel for firms

I follow Brandt et al. (2012) to create a 10-year firm panel. The whole process includes four steps. First, I trim the data by removing punctuation marks and symbols (as "?", " $*$ ", and "[" in names or telephone numbers) and deleting duplicate rows. After pre-processing, I merge two consecutive years by matching the unique firm id. The legal person's name is used in the next round if the firms cannot be matched by id. Next, I create a new variable containing each firm's telephone number, provincial code, and sectoral code and compare this variable between two years to match them. Finally, I match the firm's opening year and the main product to confirm any remaining matches. Following Brandt et al. (2012), firms that appear in the first year may be missing in the second year, only to reappear in the third year with different information, such as a changed ID or legal person. To capture the maximum amount of information available, I construct a three-year panel to identify such firms, whereby if firm A appears in year one and is unmatched in year two but reappears as firm B in year three, it is considered a match. After constructing the three-year panel, I merge all the files to create a ten-year panel. The matching results are shown in Table C.1. The "after cleaning" column drops the firms with negative or zero production or value-added, and the "continuing" column indicates firms that appear in two consecutive years in the final panel.

C.4 Estimating firm-level TFP

After creating the panel, firm-level TFP is estimated using the panel data created in the above section by Olley and Pakes (1996) method (OP hereafter). This seminal work deals with the simultaneity and selection biases when estimating productivity using a Cobb–Douglas (CD) production function. The simultaneity issue arises since the firm makes the production decision observing the productivity shock, while econometricians do not. The selection bias results from firms with more capital stock being less likely to exit the market than firms with lower stock, given a certain productivity shock.

Table C.1: Panel for firms

Year	Raw number	After cleaning	Continuing
1998	165,116	151,648	142,546
1999	162,033	150,017	140,754
2000	162,885	152,089	136,616
2001	171,256	162,757	151,547
2002	181,557	172,242	159,848
2003	196,222	189,517	163,352
2004	279,092	268,627	237,521
2005	270,043	263,769	250,717
2006	301,961	294,695	279,652
2007	336,769	330,379	

Notes: This table shows the firm panel created following Brandt et al. (2012) using annual ASIF data. “Continuing" counts the firms that appear in two consecutive years in the panel.

Following their approach, I assume firm i's production technology is captured by $Y_{it} = F(A_{it}, K_{it}, L_{it}, M_{it}, a_{it})$, where A_{it} is the productivity shock, L_{it} is the labour input, M_{it} is the intermediate input and a_{it} is the age for the firm. For estimation, write the following equation assuming the production technique is CD and lower-case letters mean log of the original variable except for age,

$$y_{it} = \beta_0 + \beta_k k_{it} + \beta_l l_{it} + \beta_m m_{it} + \beta_a a_{it} + A_{it} + \epsilon_{it}. \tag{42}$$

OP assumes a firm's investment decision is a function of A_{it}, K_{it}, a_{it} and thus A_{it} will be a function of investment I_{it}, capitalK_{it} and age a_{it}. In the first step, we replace A_{it} by the function $g(i_{it}, k_{it}, a_{it})$,

$$y_{it} = f(i_{it}, k_{it}, a_{it}) + \beta_l l_{it} + \beta_m m_{it} + \epsilon_{it}. \tag{43}$$

where $f(i_{it}, k_{it}, a_{it}) = \beta_0 + \beta_k k_{it} + \beta_a a_{it} + g(i_{it}, k_{it}, a_{it})$. Since the simultenity issue is solved, β_l and β_m are consistent.

In the second step, the survival probability P_{it} is estimated via a probit model. A firm will leave the market if the negative shock is huge, making it no longer profitable. This probability is a function of its capital stock and age.

In the final step, the remaining parameters are estimated via nonlinear least squares,

$$y_{it} - \hat{\beta}_l l_{it} - \hat{\beta}_m m_{it} = \beta_k k_{it} + \beta_a a_{it} + H(\hat{g}_{t-1} - \beta_k k_{i,t-1} - \beta_a a_{i,t-1}, \hat{P}_{it}) + \eta_{it} + \epsilon_{it}, \quad (44)$$

where $H(.)$ is approximated by ploynominals of all arguments and $\hat{}$ denotes estimation from previous steps. The data for estimation is obtained directly from the dataset and is summarised in Table C.2.

Table C.2: Summary statistics (mean)

Year	Observations	Employment	Capital	Production	Intermediate	Age	Investment
1998	146,935	389.8375	39,022.56	43,383.40	32,897.05	11.657	39,022.56
1999	125,135	377.2193	46,865.61	48,838.52	37,044.30	11.78	13,531.45
2000	123,526	347.2181	47,953.70	56,293.08	43,095.61	11.591	13,788.50
2001	134,169	328.3776	50,526.32	59,809.26	45,625.28	10.796	13,287.58
2002	142,648	312.5096	48,154.29	63,611.69	48,508.33	10.631	10,582.46
2003	159,706	300.3619	50,183.62	74,074.19	56,600.56	10.018	12,819.54
2004	231,214	239.7522	41,651.35	71,599.27	55,253.41	8.309	16,149.90
2005	224,037	257.6549	47,332.40	91,736.33	70,469.76	8.65	14,498.66
2006	253,660	245.6682	48,290.11	98,814.65	75,574.33	8.728	14,391.03
2007	284,320	230.6658	48,638.14	109,095.80	83,549.58	8.582	14,786.19

Notes: This table shows the summary statistics for Chinese manufacturing firms after creating the firm panel. Capital, production, intermediate, and investment are in thousand RMB.

The outputs are then deflated using provincial industrial producer prices from NBS. Real capital stocks are estimated using the perpetual inventory method stated in Brandt et al. (2012). There were changes in sectoral codes in 2003, so I adjust the code to ensure consistency. I drop all firms with less than eight employees, negative production, intermediate goods, investment or real capital stock. Then I create an exiting flag if a firm is no longer in the dataset after a certain year. I use the program by Yasar et al. (2008) in STATA for estimation. The results are presented in tables C.3 and C.4.

C.5 Linking balance sheet data to customs data

The customs data are from the General Administration of Customs People's Republic of China (GACC), starting from 2000. The monthly transaction data contains information on firm identifiers, HS product codes, quantities, values, modes of transportation and trading partners etc. The data is then aggregated into yearly data for each firm. A similar data cleaning procedure as in section C.3 was con-

Table C.3: Estimated TFP

	lny
age	-0.004***
	(0.0001)
lnK	0.017***
	(0.001)
lnL	0.120***
	(0.001)
lnM	0.776***
	(0.001)
year	0.030***
	(0.0002)
province	-0.002***
	(7.79e-05)
sector	0.0003***
	(4.12e-05)
type	0.05***
	(0.001)
Observations	1,844,416

Notes: This table shows the estimated coefficients for the production function. 40 sectors, 31 provinces, and five types of firms are included for estimation. Standard errors are in parentheses. *** p< 0.01, **p< 0.05, *p< 0.1.

ducted before merging the customs data with the balance sheet data. I use firm names, telephone and zip codes for matching. On average, 16% of firms engaged in trade and the number of firms increased steadily.

I test the correlation between trade and TFP by running the following OLS regression:

$$ln(TFP_{it}) = \alpha_0 + \alpha_1 Trade_{it} + \alpha_2 X_{it} + \beta_{province} + \gamma_{firmtype} + \eta_{sector} + \delta_{year} + \epsilon_{it}. \quad (45)$$

$Trade_{it}$ is a dummy, where one indicates firm i engaging in trade in year t. Next, I separate trade into exporting and importing behaviours and run a similar regression. The results are in Table C.6. Not surprisingly, firms that participate in trade have higher TFP levels. Splitting between import and export shows the former is associated with higher TFP. The results are again very close to Elliott et al. (2016), though they use the level instead of log TFP.

Table C.4: Estimated lnTFP

Year	Obs.	Mean	SD	Min	Max
1998	146935	5.535	1.697	1.922	18.965
1999	125135	6.206	1.809	1.919	18.946
2000	123526	6.044	1.835	1.919	18.967
2001	134169	6.118	1.843	1.92	18.964
2002	142648	6.706	2.048	1.919	18.968
2003	159706	6.735	2.108	1.919	18.967
2004	231214	6.357	2.070	1.919	18.967
2005	224037	6.668	2.182	1.92	18.968
2006	253660	7.191	2.351	1.92	18.968
2007	284320	7.612	2.467	1.919	18.965

Notes: This table shows the summary statistics of the estimated log of TFP for Chinese manufacturing firms from 1998 to 2007 after dropping all firms with less than eight employees, negative production, intermediate goods, investment or real capital stock.

Table C.6: correlation between Trade and TFP

	(1)	(2)
Variables	lnTFP	lnTFP
Trade	0.0430***	
	(0.001)	
Import		0.0549***
		(0.001)
Export		0.0427***
		(0.001)
Observations	1,527,473	1,974,962
R-squared	0.277	0.291
Firm-level controls	YES	YES
Province FE	YES	YES
Sector FE	YES	YES
Year FE	YES	YES
Firm_type FE	YES	YES

Notes: Standard errors are clustered at the firm level and in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Column (1) has fewer observations due to grouping importing and exporting firms together.

Table C.5: Trading firms statistics

Year	Trading firms	Total firms	Ratio
2000	21199	162,885	13.01%
2001	24957	171,256	14.57%
2002	27282	181,557	15.03%
2003	32090	196,222	16.35%
2004	50206	279,092	17.99%
2005	49900	270,043	18.48%
2006	60515	301,961	20.04%
2007	62161	336,769	18.46%

Notes: This table shows the matched data using the ASIF data and customs data. Trading firms include both exporters and importers. A firm that both exports and imports is counted as one.

C.6 Linking firm data to patent data

The patent data is obtained from He et al. (2018), where their original data is from China's Stata Intellectual Property Administration (SIPO 知识产权局). The SIPO patent database covers all published patent applications since 1985. There are three types of patents, design, invention and utility. The authors remove all patents assigned to individuals or firms outside China (including Hong Kong, Macao and Taiwan). The former condition is met only when the patent's inventor is also the assignee, and the assignee field does not contain any designators of the corporate form, while the latter requires the assignee to be firm. After some pre-processing, the authors then matched the SIPO dataset with the ASIF database (called ASIE in their paper) I described in 2.1. The final results can be freely downloaded from the Havard Dataverse repository.

As I need to compare the firms with or without a patent, a necessary task is to merge their dataset with the firm-level panel dataset I created in section 2.1. Several points need to be mentioned. First, names and IDs are used for matching. However, a lot of unmatched firms remained. As He et al. (2018) remove various designators of corporate forms (e.g. 有限公司, 股份有限公司,总公司,分公司, 工厂etc.) to obtain stem names for further matching, the same procedure is conducted to eliminate the discrepancy. Second, in He et al. (2018), the match is done with a unique firm identifier-name combination. This means there could be two firms that have the same name but different identifiers or the same identifier is linked to two firms. In their paper, they assign the patent to both firms. As I have created a panel for firms, it is easier for me to link the patent to the firm as long as the identifier or name is captured in any year. I remove the duplicates by matching the firm name and id in two steps. Third, the authors conducted a manual check for false matches by the automatic process, and they left both the true and false matches. I only keep the true matches in the dataset. I end up with 456,920 records, covering the years 1998 to 2007.

To test the correlation between TFP and patent, I run the following specification:

$$Patent_{it} = \alpha_0 + \alpha_1 lnTFP_{it} + \alpha_2 X_{it} + \beta_{province} + \gamma_{firmtype} + \eta_{sector} + \delta_{year} + \epsilon_{it}. \quad (46)$$

Table C.7 shows a clear positive correlation between the patent number and productivity.

Table C.7: Correlation between TFP and number of patents

VARIABLES number	(1) OLS all firms	(2) OLS all firms	(3) Poisson all firms	(4) Poisson all firms	(5) OLS with patent	(6) Poisson with patent
lnTFP	0.323***	0.291***	2.433***	2.168***	3.890***	0.848***
	(0.0123)	(0.0122)	(0.0631)	(0.0681)	(0.336)	(0.0660)
size		0.00121		0.0154***	-0.0181	-0.00379
		(0.000913)		(0.00571)	(0.0207)	(0.00505)
age		0.00277***		0.0181***	0.000192	-0.000483
		(0.000353)		(0.00243)	(0.00999)	(0.00212)
finance		-2.16e-05		0.000156	0.139	0.0302
		(3.27e-05)		(0.000333)	(0.227)	(0.0477)
capital		0.0147***		0.113***	-0.0818	-0.0173
intensity		(0.00181)		(0.0155)	(0.0739)	(0.0153)
Observations	1,751,878	1,733,348	1,751,878	1,733,348	47,500	47,500
R-squared	0.006	0.006			0.031	
Region FE	YES	YES	YES	YES	YES	YES
Sector FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
Firm_type FE	YES	YES	YES	YES	YES	YES

Notes: Columns (1) and (2), (3) and (4) only differ in firm-level controls. Column (4) uses the same controls as column (2). Columns (5) and (6) only keep the firms with at least one patent. Standard errors are clustered at the firm level and in parentheses. *** p< 0.01, **p< 0.05, *p< 0.1.

Appendix D Suggestive causal evidence

In this section, I present some suggestive causal evidence from trade to productivity and innovation. To test the causality, I use changes in import tariff and world export supply as instruments for imports and world import demand as an instrument for exports as Chen et al. (2017) and Hummels et al. (2014). World export supply (WES_{fkt}) is country f's total supply of product k to the world minus supply to China at time t. World import demand (WID_{fkt}) is country f's total imports of product k from the world minus imports from China at time t. Import tariffs have been widely used in similar studies (e.g. Huang et al. (2022), Chen et al. (2017), Xu (2012)). There were significant reductions in tariffs after China entered WTO at the end of 2001 (see Figure D.1). Although there have been some discussions about the event, there were a lot of uncertainties, and firms could not know when the tariff reductions would be enforced. As stated and tested in Lu and Yu (2015), the lengthy and uncertain process of China's accession to WTO and the strict compliance with the WTO agreement make the tariff reductions plausibly exogenous.

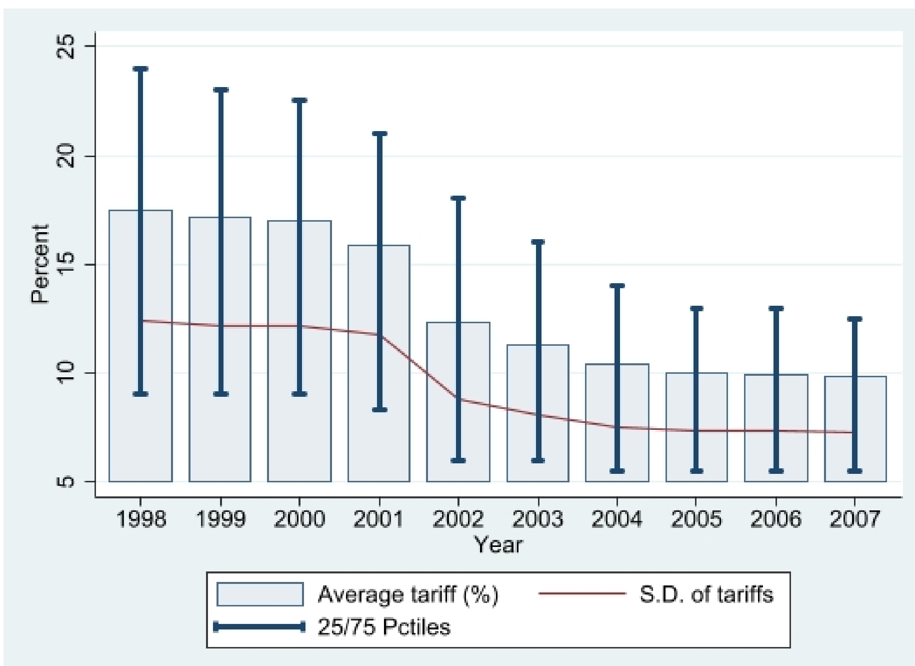

Figure D.1: Average tariff over time

Notes: This figure shows the averaged sectoral import tariffs in China from 1998 to 2007 and the reduction after the WTO accession at the end of 2001.

I construct firm-specific instruments for imports and exports. To be more specific, the instrument for firm i is a shift-share instrument $I_{it} = \sum_f \sum_p s_{2000}^{f,p} \times instrument_{f,p,t}$. Instruments include $\{tariff_{f,p,t}, WES_{f,p,t}, WID_{f,p,t}\}$, where f and p indicates product p from economy f and $s_{2000}^{f,p}$ is the share of product that was imported from or exported to region f in the year 2000, which is the earliest available data. I collect other economies' product-level trade data from the CEPII database and the HS 6-digit level tariff data from the WTO Tariff Download Facility. I match them with each firm's import transaction data. As one firm can import or export different goods, I take the mean of all transactions. Then the firm-specific average WES, WID and tariff are used in the first stage estimation as IVs for the imports or exports per firm.

Using the instruments, I first estimate the following:

$$trade_{it} = \alpha_1 I_{it} + \alpha_2 X_{it} + \lambda_t + \lambda_i + \epsilon_{it}. \tag{47}$$

The predicted value of imports or exports from the first stage is then used in the second-stage estimation,

$$Y_{it} = \alpha_1 \widehat{trade}_{it} + \alpha_2 X_{it} + \lambda_t + \lambda_i + \epsilon_{it}. \tag{48}$$

where Y means firm i's TFP or patent number. Note $\widehat{trade}_{it}$ is the predicted value using tariff from the first stage. To further examine the mechanism, I also include the interaction between import and the patent stock of the foreign country. The results are displayed in Table D.1.

Table D.1: Trade and TFP

Panel A: Exports						
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	OLS	IV	IV	IV	IV
log export	0.008***	0.008***	0.009**	0.009**	0.008**	0.007
	(0.001)	(0.001)	(0.004)	(0.004)	(0.004)	(0.005)
log export $\times$ foreign stock					0.001	
					(0.002)	
log export $\times$ high						0.003
						(0.005)
Observations	51,211	51,211	51,211	51,211	51,211	51,211
R-squared	0.752	0.754	0.751	0.753	0.751	0.753
First stage F-Stat		387.43	387.43	387.43	387.43	387.43
IV: World import demand						
Panel B: Imports						
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	OLS	IV	IV	IV	IV
log import	0.005*	0.005*	0.005*	0.005*	0.004*	0.004
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
log import $\times$ foreign stock					0.003**	
					(0.001)	
log import $\times$ high						0.003
						(0.004)
Observations	42,074	42,074	42,074	42,074	41,994	42,074
R-squared	0.769	0.771	0.769	0.771	0.770	0.771
First stage F-Stat		583.77	583.777	583.77	583.77	583.77
IV: World export supply						
Year FE	YES	YES	YES	YES	YES	YES
Firm FE	YES	YES	YES	YES	YES	YES
Controls	NO	YES	NO	YES	YES	YES

Notes: *** $p< 0.01$, **$p< 0.05$, *$p< 0.1$. Standard errors are clustered at the firm level and in parentheses. Both panels use the same IV and the same independent variables. Firm-level controls include age, log of capital intensity (capital per employee) and financial status (debt to total asset ratio).

Table D.2: Trade and innovation

Panel A: Exports						
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	IV	IV	IV	IV	IV
log export	0.003**	-0.015**	-0.015**		-0.017**	-0.024***
	(0.002)	(0.007)	(0.007)		(0.007)	(0.008)
log export (lag 1),				-0.003		
				(0.020)		
log export (lag 2),				0.004		
				(0.019)		
log export (lag 3),				-0.040*		
				(0.023)		
log export × foreign stock					0.003	
					(0.003)	
log export × high						0.013*
						(0.008)
Observations	51,211	51,211	51,211	10,496	51,211	51,211
R-squared	0.593	0.593	0.593	0.737	0.593	0.593
First stage F-Stat		387.7	387.7	387.7	387.7	387.7
IV: World import demand						
Panel B: Imports						
	(1)	(2)	(3)	(4)	(5)	(6)
	OLS	IV	IV	IV	IV	IV
log import (IV)	0.009*	0.009*	0.009*		0.008	0.004
	(0.002)	(0.005)	(0.005)		(0.005)	(0.006)
log import (lag 1),				0.004		
				(0.016)		
log import (lag 2),				0.039*		
				(0.023)		
log import (lag 3),				0.006		
				(0.015)		
log import × foreign stock					0.005**	
					(0.002)	
log import × high						0.010
						(0.008)
Observations	42,074	42,074	42,074	8,429	42,003	42,074
R-squared	0.608	0.608	0.608	0.730	0.609	0.608
First stage F-Stat		583.77	583.777	583.77	583.77	583.77
IV: World export supply						
Year FE	YES	YES	YES	YES	YES	YES
Firm FE	YES	YES	YES	YES	YES	YES
Controls	NO	NO	YES	YES	YES	YES

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors are clustered at the firm level and in parentheses. Both panels use the same IV and the same independent variables. Firm-level controls include age, log of capital intensity (capital per employee) and financial status (debt to total asset ratio).

Table D.3: First-stage results

VARIABLES	(1) lnEXP	(2) lnIMP
lnWIDs	0.304***	0.066***
	(0.020)	(0.019)
lnWESs	0.114***	0.327***
	(0.012)	(0.015)
Wtariff	-0.002	-0.021***
	(0.002)	(0.003)
Observations	27,402	27,402
R-squared	0.803	0.822
Year FE	YES	YES
Firm FE	YES	YES
First stage F-Stat	111.569	176.237

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors are clustered at the firm level and in parentheses. lnWIDs and lnWESs are logs of world import demand and world export supply. Wtariff is the weighted import tariff.

Table D.4: Firms both import and export

	(1) lnTFP	(2) lnTFP	(3) lnN	(4) lnN
log export	0.014**	0.014**	-0.020*	-0.020*
	(0.005)	(0.005)	(0.011)	(0.011)
log import	-0.001	-0.001	0.028***	0.027***
	(0.004)	(0.004)	(0.008)	(0.008)
Observations	27,402	27,402	27,402	27,402
R-squared	0.765	0.766	0.615	0.616
Year FE	YES	YES	YES	YES
Firm FE	YES	YES	YES	YES
Controls	NO	YES	NO	YES

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Standard errors are clustered at the firm level and in parentheses. lnN is the log of one plus new patent applications. Firm-level controls include age, log of capital intensity (capital per employee) and financial status (debt to total asset ratio).

The results indicate the exogenous positive shocks of imports or exports lead to a higher productivity level for trading firms. Moreover, importing from countries with a higher stock of patents contributes more to the domestic TFP, indicating an impact on knowledge diffusion through imports. However, opposite results are found in terms of innovation. Imports lead to higher innovation at home, while exports reduce innovation. Running a separate regression for each sector shows while there is clearly heterogeneity, any specific sector does not drive the results.

Appendix E Mathematical appendix

E.1 Equilibrium

Equilibrium consists of six equations (with one redundant). The production cost of each variety of intermediate goods is,

$$c_n^j = \eta_n^j w_n^{\gamma_n^j} \prod_{k=1}^{J} \left(P_n^k\right)^{\gamma_n^{k,j}}, \tag{49}$$

and $\eta_n^j = \prod_{k=1}^{J} (\gamma_n^{k,j})^{-\gamma_n^{k,j}} (\gamma_n^j)^{-\gamma_n^j}$ is a constant.

The total expenditure of region n on sector j goods is thus given by

$$X_n^j = \sum_{k=1}^{J} \gamma_n^{j,k} \sum_{i=1}^{N} X_i^k \pi_{in}^k + \alpha_n^j (w_n L_n + \sum_{j=1}^{J} \frac{\sigma^j - 1}{\sigma^j \theta^j} Y_n^j). \tag{50}$$

The trade balance in country n is

$$\sum_{j=1}^{J} \sum_{i=1}^{N} X_n^j \pi_{ni}^j = \sum_{j=1}^{J} \sum_{i=1}^{N} X_i^j \pi_{in}^j + D_n. \tag{51}$$

The price index at each country n is,

$$P_n^j = \left[\frac{\theta^j}{\theta^j - \sigma^j + 1}\right]^{-1/\theta^j} \frac{\sigma^j}{\sigma^j - 1} \sigma^{j\,\frac{\sigma^j - 1 - \theta^j}{-\theta^j(\sigma - 1)}} \left[\sum_{i=1}^{N} w_i L_i \left(\frac{c_i^j d_{ni}^j}{\zeta_{ni}^j}\right)^{-\theta^j} T_i^j \left(\frac{f_{ni}}{X_n^j}\right)^{\frac{\sigma^j - 1 - \theta^j}{\sigma^j - 1}}\right]^{-\frac{1}{\theta^j}} \tag{52}$$

This also implies the labour market clearing condition is,

$$w_n L_n = \sum_{j=1}^{J} \gamma_n^j \sum_{i=1}^{N} X_i^j \pi_{in}^j \tag{53}$$

The expenditure share is,

$$\pi_{ni}^j = \frac{X_{ni}^j}{\sum_{i=1}^{N} X_{ni}^j} = \frac{w_i L_i T_i^j (f_{ni}^j)^{1 - \frac{\theta^j}{\sigma^j - 1}} \left(c_i^j \tilde{d}_{ni}^j\right)^{-\theta^j}}{\sum_{k=1}^{N} w_k L_k T_k^j (f_{nk}^j)^{1 - \frac{\theta^j}{\sigma^j - 1}} \left(c_k^j \tilde{d}_{nk}^j\right)^{-\theta^j}}, \tag{54}$$

where $\tilde{d} = \frac{d^j_{ni}}{\zeta^j_{ni}}$.

Finally, the growth of knowledge stock follows,

$$T^j_{n,t+1} - T^j_{n,t} = m^j_t \Gamma(1-\rho^j) \left[\sum_{i=1}^{N} \pi^j_{nit} \left(T^j_{i,t}\right)^{\rho^j} \right], \tag{55}$$

where $\Gamma(u) = \int_0^\infty x^{u-1} dx$.

E.2 Derivation of the balanced growth path

The model is associated with a balanced growth path where all variables are growing at a constant rate. First, note that the growth rate of T^j_n will be the same across sector j in all countries because of the technology diffusion. Then, restricting expenditure shares from exploding, we need them to be growing at the same rate for all sectors. Balanced trade indicates the total expenditure in each country grows at the same rate (51). (53) shows wage growth at the same rate as the total expenditure. Since total entry costs are a fraction of gross profits, which is a share of total expenditure, f^j_{ni} has to be growing at the same rate since the potential entrants are proportional to total labour income. Together with (52), we know that cost c^j_{it} also grows at the same rate across countries. Going back to the function of production cost (49), we find prices will grow at the same rate in the same sector.

Formally, from (52), the growth rate of price in region n and sector j is,

$$g_{P^j_n} = -\frac{1}{\theta^j}(g_T) + g_{c^j} - \frac{1}{\sigma^j - 1} g_w, \tag{56}$$

(49) indicates

$$g_{c^j} = \gamma^j_n g_{w_n} + \sum_{k=1}^{J} \gamma^{kj}_n g_{P^k_n} \tag{57}$$

We care about the welfare of the economy, which is measured by the real income $W_n = I_n/P_n$. The growth rate of nominal income I_n equals g_w since from (53) we know that wage and expenditure share grow at the same rate. The relative proportions of $w_n L_n$ and $\frac{\sigma^j-1}{\sigma^j\theta^j} X^j_n, \forall j$ in the sum remain constant over time, so the total income has to grow at the same rate too. Therefore, the growth rate of welfare is

$$g_{W_n} = g_w - g_{p_n} = g_w - \sum_{j=1}^{J} \alpha^j g_{p^j_n}. \tag{58}$$

Given (56) and (57), we can solve for g_{c^j} and $g_{P_n^j}$ as a function of g_{w_n} and g_A.

$$g_{P_n^j} = -\frac{1}{\theta^j} g_T + \left(\gamma_n^j - \frac{1}{\sigma^j - 1}\right) g_w + \sum_{k=1}^{J} \gamma_n^{kj} g_{P_n^k}, \tag{59}$$

Consider the matrix representation,

$$\begin{bmatrix} g_{P_n^1} \\ g_{P_n^2} \\ \dots \\ g_{P_n^3} \end{bmatrix} = \begin{bmatrix} \gamma_n^{11} & \gamma_n^{21} & \dots & \gamma_n^{J1} \\ \gamma_n^{12} & \gamma_n^{22} & \dots & \gamma_n^{J2} \\ \dots \\ \gamma_n^{1J} & \gamma_n^{2J} & \dots & \gamma_n^{JJ} \end{bmatrix} \begin{bmatrix} g_{P_n^1} \\ g_{P_n^2} \\ \dots \\ g_{P_n^3} \end{bmatrix} + \begin{bmatrix} -\frac{1}{\theta^1} \\ -\frac{1}{\theta^2} \\ \dots \\ -\frac{1}{\theta^J} \end{bmatrix} g_T + \begin{bmatrix} \gamma_n^1 - \frac{1}{\sigma^1 - 1} \\ \gamma_n^2 - \frac{1}{\sigma^2 - 1} \\ \dots \\ \gamma_n^J - \frac{1}{\sigma^3 - 1} \end{bmatrix} g_w$$

Define $\Omega \equiv \begin{bmatrix} \gamma_n^{11} & \gamma_n^{21} & \dots & \gamma_n^{J1} \\ \gamma_n^{12} & \gamma_n^{22} & \dots & \gamma_n^{J2} \\ \dots \\ \gamma_n^{1J} & \gamma_n^{2J} & \dots & \gamma_n^{JJ} \end{bmatrix}$, we can solve for the price indices in each sector,

$$\begin{bmatrix} g_{P_n^1} \\ g_{P_n^2} \\ \dots \\ g_{P_n^3} \end{bmatrix} = [I - \Omega]^{-1} \left(\begin{bmatrix} -\frac{1}{\theta^1} \\ -\frac{1}{\theta^2} \\ \dots \\ -\frac{1}{\theta^J} \end{bmatrix} g_T + \begin{bmatrix} \gamma_n^1 - \frac{1}{\sigma^1 - 1} \\ \gamma_n^2 - \frac{1}{\sigma^2 - 1} \\ \dots \\ \gamma_n^J - \frac{1}{\sigma^3 - 1} \end{bmatrix} g_w \right)$$

Now we can get the growth of the real income,

$$\sum_{k=1}^{J} \sum_{j=1}^{J} \alpha_n^j \frac{\Phi_n^{jk}}{\theta^k} g_T + \sum_{k=1}^{J} \sum_{j=1}^{J} \frac{\alpha_n^j}{\sigma^j - 1} \Phi_n^{jk} g_w \tag{60}$$

where $\Phi = [I - \Omega]^{-1}$ and Φ^{jk} means the element on row j and column k of the matrix Φ.

To get the intuition behind this, let us assume there is only one sector and no intermediate inputs, then (60) becomes

$$g_{W_n} = \frac{g_T}{(1 - \gamma^{11})\theta^1} = \frac{g_T}{\gamma^1 \theta^1} + \frac{g_w}{\gamma^1 (\sigma^1 - 1)} = \frac{g_T}{\theta^1} + \frac{g_w}{\sigma^1 - 1} \tag{61}$$

The real income grows at the rate that nominal income grows minus the speed of price increase. The differences between the two growth rates are the real growth of the productivity that can be attributed to the workers' productivity, weighted by the consumption share ($\alpha^1 = 1$ here). Moreover, in this particular setting, profits are closely related to labour income since it determines the new entrants and the cutting productivity level. As the second effect is stronger, wage growth leads to welfare gains.

Let us now go back to (55) to derive the growth rate of technology on the balanced growth path. Denote $\hat{X}$ as the detrended variable.

$$\hat{T}^j_{nt+1}\exp(g^j_T(t+1))-\hat{T}^j_{nt}\exp(g^j_T t) = m^j_0\exp(g_{m^j}t)\Gamma\left[\sum_{i=1}^{N}\pi^j_{nit}\left(\hat{T}^j_{i,t}\exp(g_T t)\right)^{\rho^j}\right] \tag{62}$$

After simplification, we get

$$\hat{T}^j_{nt+1}\exp(g^j_T) - \hat{T}^j_{nt} = m^j_0\exp(g_{m^j}t + g^j_T(\rho^j-1)t)\Gamma\left[\sum_{i=1}^{N}\pi^j_{nit}\left(\hat{T}^j_{i,t}\right)^{\rho^j}\right] \tag{63}$$

In order for the right-hand-side of (63) to be constant, the growth rate of T has to satisfy

$$g^j_T = \frac{g_{m^j}}{1-\rho^j}. \tag{64}$$

E.3 Detrended equilibrium

The detrended equilibrium is written in relative changes. Let us first denote the detrended model as the following. Normalise wage to a benchmark and call $\bar{w}$ the growth rate of all wages, then $\hat{w}_{nt} = \frac{w_{nt}}{\bar{w}}$ and $\hat{X}^j_n = \frac{X^j_n}{\bar{w}}$ are growing at a constant rate along the BGP. We have solved the growth rates of price and cost as functions of g_w and g_T. Thus we can normalize $\hat{P}^j_{nt} = \frac{P^j_{nt}}{\bar{A}^{\Psi^j_1}\bar{w}^{\Psi^j_2}}$, where Ψ^j_1 and Ψ^j_2 means the j^{th} element of

$$\Psi_1 = [I-\Omega]^{-1}\begin{bmatrix}-\frac{1}{\theta^1}\\ -\frac{1}{\theta^2}\\ \cdots\\ -\frac{1}{\theta^J}\end{bmatrix} \text{ and } \Psi_2 = [I-\Omega]^{-1}\begin{bmatrix}\gamma^1_n-\frac{1}{\sigma^1-1}\\ \gamma^2_n-\frac{1}{\sigma^2-1}\\ \cdots\\ \gamma^J_n-\frac{1}{\sigma^3-1}\end{bmatrix}$$

Now we can normalize the costs via equation (56) such that $\hat{c}^j_{nt} = \frac{\hat{c}^j_{nt}}{\bar{T}^{\Psi^j_1+1/\theta^j}\bar{w}^{\Psi^j_2}}$.

The normalized equilibrium equations are:

The cost of production,

$$\hat{c}_{nt}^{j} = \eta_n^j \hat{w}_{nt}^{\gamma_n^j} \prod_{k=1}^{J} \left(\hat{P}_{nt}^{k}\right)^{\gamma_n^{k,j}}, \tag{65}$$

The expenditure share in each sector,

$$\pi_{nit}^{j} = \frac{\hat{w}_{it} L_{it} \hat{T}_{it}^{j} (f_{nit}^{j})^{1-\frac{\theta^j}{\sigma^j-1}} \left(\hat{c}_{it}^{j} \tilde{d}_{nit}^{j}\right)^{-\theta^j}}{\sum_{k=1}^{N} \hat{w}_{kt} L_{kt} \hat{T}_{kt}^{j} (f_{nkt}^{j})^{1-\frac{\theta^j}{\sigma^j-1}} \left(\hat{c}_{kt}^{j} \tilde{d}_{nkt}^{j}\right)^{-\theta^j}}, \tag{66}$$

Price of the intermediate goods in each sector,

$$\hat{P}_{nt}^{j} = b \left[\sum_{n}^{N} \left(\frac{\hat{c}_{it}^{j} d_{nit}^{j}}{\zeta_{ni}^{j}} \right)^{-\theta^j} \hat{w}_{it} L_i T_{it}^{j} \left(\frac{f_{nit}}{\hat{X}_{nt}^{j}} \right)^{\frac{\sigma^j - 1 - \theta^j}{\sigma^j - 1}} \right]^{-\frac{1}{\theta^j}} \tag{67}$$

where $b = \left[\frac{\theta^j}{\theta^j - \sigma^j + 1}\right]^{-1/\theta^j} \frac{\sigma^j}{\sigma^j - 1} \sigma^{j\,\frac{\sigma^j - 1 - \theta^j}{-\theta^j(\sigma^j - 1)}}$ is a constant.

The total expenditure on sector j's goods,

$$\hat{X}_{nt}^{j} = \sum_{k=1}^{J} \gamma_n^{j,k} \sum_{i=1}^{N} \hat{X}_{it}^{k} \pi_{int}^{k} + \alpha_n^j \left(\hat{w}_{nt} L_{nt} + \sum_{j=1}^{J} \frac{\sigma^j - 1}{\sigma^j \theta^j} \hat{Y}_n^j \right). \tag{68}$$

The labour market clearing condition,

$$\hat{w}_{nt} L_{nt} = \sum_{j=1}^{J} \gamma_n^j \sum_{i=1}^{N} \hat{X}_{it}^{j} \pi_{int}^{j} \tag{69}$$

The knowledge diffusion process,

$$\hat{T}_{nt+1}^{j} \exp(g_T^j) - \hat{T}_{nt}^{j} = m_0^j \Gamma \left[\sum_{i=1}^{N} \pi_{nit}^{j} \left(\hat{T}_{i,t}^{j} \right)^{\rho^j} \right] \tag{70}$$

E.4 Derivation of equations (12) and (13)

Expenditure from country n on i equals the sum of all firm-level sales:

$$
\begin{aligned}
X_{ni}^j = w_i L_i \int_{z_{ni}^{x,j}}^{\infty} p_{ni}^j q_{ni}^j dG(z) \\
&= w_i L_i \int_{z_{ni}^{x,j}}^{\infty} \left(\frac{p_{ni}^j}{P_n^j}\right)^{1-\sigma^j} X_n^j dG(z) \\
&= \theta^j w_i L_i \left(\frac{p_{ni}^j(z_{ni}^{x,j})}{P_n^j}\right)^{1-\sigma^j} X_n^j \int_{z_{ni}^{x,j}}^{\infty} \left(\frac{p_{ni}^j(z)}{p_{ni}^j(z^x)}\right)^{1-\sigma^j} T_i^j (z_{ni}^j)^{-\theta^j-1} dz \\
&= \theta^j w_i L_i \sigma^j f_{ni}^j \int_{z_{ni}^{x,j}}^{\infty} \left(\frac{p_{ni}^j(z)}{p_{ni}^j(z^x)}\right)^{1-\sigma^j} T_i^j (z_{ni}^j)^{-\theta^j-1} dz \\
&= \theta^j w_i L_i T_i^j \sigma^j f_{ni}^j \int_{z_{ni}^{x,j}}^{\infty} \left(\frac{z^x}{z}\right)^{1-\sigma^j} (z)^{-\theta^j-1} dz \\
&= \frac{\theta^j \sigma^j}{\theta^j - \sigma^j + 1} w_i L_i T_i^j f_{ni}^j (z_{ni}^{x,j})^{-\theta^j} \\
&= \frac{\theta^j \sigma^j}{\theta^j - \sigma^j + 1} w_i L_i T_i^j f_{ni}^j \left(\frac{c_i^j d_{ni}^j}{\zeta_{ni}^j P_n^j} \left(\frac{X_n^j}{f_{ni}^j}\right)^{\frac{1}{1-\sigma^j}} \frac{1}{\sigma^j - 1} \sigma^{j\,\frac{\sigma^j}{\sigma^j-1}}\right)^{-\theta^j} \\
&= \frac{\theta^j \sigma^j}{\theta^j - \sigma^j + 1} \left(\frac{1}{\sigma^j - 1}\sigma^{j\,\frac{\sigma^j}{\sigma^j-1}}\right)^{-\theta^j} w_i L_i T_i^j (f_{ni}^j)^{1+\frac{\theta^j}{1-\sigma^j}} \left(\frac{c_i^j d_{ni}^j}{\zeta_{ni}^j}\right)^{-\theta^j} (P_n^j)^{\theta^j} \left(X_n^j\right)^{\frac{-\theta^j}{1-\sigma^j}}
\end{aligned} \tag{71}
$$

The aggregate price index is

$$
\begin{aligned}
{P_n^j}^{1-\sigma^j} &= \sum_{i=1}^{N} w_i L_i \left(\int_{z_{ni}^{x,j}}^{\infty} (p_{ni}^j)^{1-\sigma^j} dG(z)\right) \\
&= \sum_{i=1}^{N} w_i L_i \left(\int_{z_{ni}^{x,j}}^{\infty} \left(\frac{c_i^j d_{ni}^j}{\zeta_{ni}^j z_{ni}^j}\right)^{1-\sigma^j} dG(z)\right) \\
&= \theta^j \sum_{i=1}^{N} w_i L_i T_i^j \left(\frac{c_i^j d_{ni}^j}{\zeta_{ni}^j}\right)^{1-\sigma^j} \left(\int_{z_{ni}^{x,j}}^{\infty} z^{\sigma^j-1-\theta^j-1} dz\right) \\
&= \frac{\theta^j}{\theta^j - \sigma^j + 1} \sum_{i=1}^{N} \left(\frac{c_i^j d_{ni}^j}{\zeta_{ni}^j}\right)^{1-\sigma^j} w_i L_i T_i^j (z_{ni}^{x,j})^{\sigma^j-1-\theta^j}
\end{aligned} \tag{72}
$$

E.5 Derivation of equation (60)

Consider the matrix representation of sector price in region n,

$$\begin{bmatrix} g_{P_n^1} \\ g_{P_n^2} \\ \ldots \\ g_{P_n^3} \end{bmatrix} = [I-\Omega]^{-1} \left(\begin{bmatrix} -\frac{1}{\theta^1} \\ -\frac{1}{\theta^2} \\ \ldots \\ -\frac{1}{\theta^J} \end{bmatrix} g_T + \begin{bmatrix} \gamma_n^1 - \frac{1}{\sigma-1} \\ \gamma_n^2 - \frac{1}{\sigma-1} \\ \ldots \\ \gamma_n^J - \frac{1}{\sigma-1} \end{bmatrix} g_w \right)$$

where $[I-\Omega] = \begin{bmatrix} 1-\gamma_n^{11} & \gamma_n^{21} & \ldots & \gamma_n^{J1} \\ \gamma_n^{12} & 1-\gamma_n^{22} & \ldots & \gamma_n^{J2} \\ \ldots & & & \\ \gamma_n^{1J} & \gamma_n^{2J} & \ldots & 1-\gamma_n^{JJ} \end{bmatrix}$.

For illustration, suppose there are two sectors only. We can write the conditions as the following:

$$\begin{bmatrix} g_{P_n^1} \\ g_{P_n^2} \end{bmatrix} = \Theta \begin{bmatrix} 1-\gamma_n^{22} & -\gamma_n^{21} \\ -\gamma_n^{12} & 1-\gamma_n^{11} \end{bmatrix} \begin{bmatrix} -\frac{1}{\theta^1} g_T + (\gamma_n^1 - \frac{1}{\sigma-1}) g_w \\ -\frac{1}{\theta^2} g_T + (\gamma_n^2 - \frac{1}{\sigma-1}) g_w \end{bmatrix}$$

where $\Theta \equiv (1-\gamma^{11}-\gamma^{22}+\gamma^{11}\gamma^{22}-\gamma^{12}\gamma^{21})^{-1}$. We can write the price indices as

$$g_{P_n^1} = \Theta \left[(\gamma_n^1 - \gamma_n^1\gamma^{22} - \gamma_n^2\gamma^{21}) g_w + \left(-\frac{1-\gamma_n^{22}}{\theta^1} + \frac{\gamma^{21}}{\theta^2} \right) g_T + \frac{\gamma_n^1}{1-\sigma} g_w \right] \tag{73}$$

$$g_{P_n^2} = \Theta \left[(\gamma_n^2 - \gamma_n^2\gamma^{11} - \gamma_n^1\gamma^{12}) g_w + \left(-\frac{1-\gamma_n^{11}}{\theta^2} + \frac{\gamma^{12}}{\theta^1} \right) g_T + \frac{\gamma_n^2}{1-\sigma} g_w \right] \tag{74}$$

Now we can write $\alpha^1 g_{P_n^1} + \alpha^2 g_{P_n^2}$ as the sum of two terms,

$$\begin{aligned} &\Theta g_w[\alpha^1(\gamma_n^1 - \gamma_n^1\gamma^{22} - \gamma_n^2\gamma^{21}) + \alpha^2(\gamma_n^2 - \gamma_n^2\gamma^{11} - \gamma_n^1\gamma^{12})] + \\ &\Theta g_A \left[\alpha^1 \left(-\frac{1-\gamma_n^{22}}{\theta^1} + \frac{\gamma^{21}}{\theta^2} \right) + \alpha^2 \left(-\frac{1-\gamma_n^{11}}{\theta^2} + \frac{\gamma^{12}}{\theta^1} \right) \right] + \\ &\Theta g_w[\alpha^1 \frac{\gamma_n^1}{1-\sigma} + \alpha^2 \frac{\gamma_n^2}{1-\sigma}] \end{aligned} \tag{75}$$

Think about the first term in (75),

$$
\begin{aligned}
&\Theta g_w[\alpha^1(\gamma_n^1 - \gamma_n^1\gamma^{22} - \gamma_n^2\gamma^{21}) + \alpha^2(\gamma_n^2 - \gamma_n^2\gamma^{11} - \gamma_n^1\gamma^{12})] \\
&= \Theta g_w[\alpha^1((1-\gamma^{11}-\gamma^{21})(1-\gamma^{22}) - (1-\gamma^{12}-\gamma^{22})\gamma^{21}) + \\
&\alpha^2((1-\gamma^{12}-\gamma^{22})(1-\gamma^{11}) - (1-\gamma^{11}-\gamma^{21})\gamma^{12})] \\
&= \Theta g_w[\alpha^1[(1-\gamma^{22})(1-\gamma^{11}) - \gamma^{21}(1-\gamma^{22}) - \gamma^{21}(1-\gamma^{12}-\gamma^{22})] + \\
&\alpha^2[(1-\gamma^{22})(1-\gamma^{11}) - \gamma^{12}(1-\gamma^{11}) - \gamma^{12}(1-\gamma^{11}-\gamma^{21})]] \\
&= g_w(\alpha^1 + \alpha^2)\frac{1}{\Theta}\Theta = g_w
\end{aligned}
\tag{76}
$$

Think about the second term in (75),

$$
\Theta g_T \left[\alpha^1\left(-\frac{1-\gamma_n^{22}}{\theta^1} + \frac{\gamma^{21}}{\theta^2}\right) + \alpha^2\left(-\frac{1-\gamma_n^{11}}{\theta^2} + \frac{\gamma^{12}}{\theta^1}\right)\right] \tag{77}
$$

It is the sum of the weighted growth by the matrix $[I-\Omega]^{-1}$ and the trade elasticity, which can thus be represented as

$$
-\sum_{k=1}^{J}\sum_{j=1}^{J}\alpha_n^j\frac{\Phi_n^{jk}}{\theta^k}g_T \tag{78}
$$

where $\Phi_n = [I-\Omega_n]^{-1}$ and Φ_n^{jk} means the element on row j and column k of the matrix Φ_n.

Finally, the third term can be summarised as

$$
-\frac{1}{\sigma-1}\sum_{j=1}^{J}\alpha_n^j\sum_{k=1}^{J}\Phi_n^{jk}g_w \tag{79}
$$

Now we can get the growth of the real income,

$$
g_{W_n} = g_{w_n} - \sum_{j=1}^{J}\alpha_n^j g_{p_n^j} = \sum_{k=1}^{J}\sum_{j=1}^{J}\alpha^j\frac{\Phi^{jk}}{\theta^k}g_A + \frac{1}{\sigma-1}\sum_{j=1}^{J}\alpha_n^j\sum_{k=1}^{J}\Phi_n^{jk}g_w \tag{80}
$$

E.6 Derivation of equation (66)

$$\hat{P}^j_{nt+1} = \left[\sum_n^N \left(\frac{\hat{c}^j_{it+1} d^j_{nit+1}}{\zeta^j_{nit+1}}\right)^{-\theta^j} \hat{w}_{it+1} L_{it+1} T^j_{it+1} \left(\frac{f_{nit+1}}{\hat{X}^j_{nt+1}}\right)^{\frac{\sigma-1-\theta^j}{\sigma-1}}\right]^{-\frac{1}{\theta^j}} \tag{81}$$

$$\begin{aligned}\hat{P}^j_{nt} {P'}^j_{nt+1} = \Bigg[\sum_i^N &\left(\frac{\hat{c}^j_{it} d^j_{nit}}{\zeta^j_{ni}}\right)^{-\theta^j} \hat{w}_{it} L_i T^j_{it} \left(\frac{f_{nit}}{\hat{X}^j_{nt}}\right)^{\frac{\sigma-1-\theta^j}{\sigma-1}} \\ &\left(\frac{\hat{c'}^j_{it+1} {d'}^j_{nit+1}}{{\zeta'}^j_{nit+1}}\right)^{-\theta^j} \hat{w'}_{it+1} {L'}_{it+1} {T'}^j_{it+1} \left(\frac{{f'}_{nit+1}}{\hat{X'}^j_{nt+1}}\right)^{\frac{\sigma-1-\theta^j}{\sigma-1}}\Bigg]^{-\frac{1}{\theta^j}}\end{aligned} \tag{82}$$

$$\pi^j_{nit} \sum_{k=1}^N \hat{w}_{kt} L_{kt} \hat{T}^j_{kt} (f^j_{nkt})^{1-\frac{\theta^j}{\sigma-1}} \left(\hat{c}^j_{kt} \tilde{d}^j_{nkt}\right)^{-\theta^j} = \hat{w}_{it} L_{it} \hat{T}^j_{it} (f^j_{nit})^{1-\frac{\theta^j}{\sigma-1}} \left(\hat{c}^j_{it} \tilde{d}^j_{nit}\right)^{-\theta^j}, \tag{83}$$

$$\begin{aligned}\hat{P}^j_{nt} {P'}^j_{nt+1} = \Bigg[\sum_i^N \pi^j_{nit} \sum_{k=1}^N \hat{w}_{kt} L_{kt} \hat{T}^j_{kt} &\left(\frac{f_{nit}}{\hat{X}^j_{nt}}\right)^{\frac{\sigma-1-\theta^j}{\sigma-1}} \left(\hat{c}^j_{kt} \tilde{d}^j_{nkt}\right)^{-\theta^j} \left(\frac{\hat{c'}^j_{it+1} {d'}^j_{nit+1}}{{\zeta'}^j_{nit+1}}\right)^{-\theta^j} \\ \hat{w'}_{it+1} {L'}_{it+1} {T'}^j_{it+1} &\left(\frac{{f'}_{nit+1}}{\hat{X'}^j_{nt+1}}\right)^{\frac{\sigma-1-\theta^j}{\sigma-1}}\Bigg]^{-\frac{1}{\theta^j}}\end{aligned} \tag{84}$$

$${P'}^j_{nt+1} = \left[\sum_i^N \pi^j_{nit} \left(\frac{\hat{c'}^j_{it+1} {d'}^j_{nit+1}}{{\zeta'}^j_{nit+1}}\right)^{-\theta^j} \hat{w'}_{it+1} {L'}_{it+1} {T'}^j_{it+1} \left(\frac{{f'}_{nit+1}}{\hat{X'}^j_{nt+1}}\right)^{\frac{\sigma-1-\theta^j}{\sigma-1}}\right]^{-\frac{1}{\theta^j}} \tag{85}$$

E.7 Derivation of the knowledge diffusion function (16)

The derivation of the knowledge diffusion function follows Cai et al. (2022b)'s Appendix A closely with some deviations. Here I only present some key parts. The frontier of knowledge $F_t(z)$ (omit sector for simplicity) changes from t to $t+1$ if we have some better ideas associated with higher productivity. At $t+1$, we then have $F_{t+1}(z) = \Pr[$the best productivity is no greater than z at $t+1] = F_q(z) \cdot F^{\text{best new}}(q)$ $= F_0(z) \cdot \prod_{\tau=0}^{t} F_\tau^{\text{best new}}(z)$

I have assumed the initial distribution at time 0 follows an exponential distribution, $F_0(z) = \exp(1 - T_0 z^{-\theta})$. Therefore, the frontier of knowledge will be exponential at any t since: $F_t(z) = \exp\left(1 - \left(T_0 + \sum_{\tau=0}^{t-1} m_\tau \int_0^\infty x^{\rho\theta} dG_\tau(x)\right) z^{-\theta}\right) =$ $\exp(1 - T_t z^{-\theta})$. Given the exponential distribution, the transformed distribution of $Z = lnz$ is Pareto.

From here we can derive the low of motion of the knowledge stock at each time,

$$T_{t+1} = T_t + m_t \int_0^\infty x^{\rho\theta} dG_t(x).$$

Assume that at time t in country n, when a new idea arrives, the insight from any goods that are selling in n contributes to the creation of the new idea.

Then $G_{n,t}(z') = \Pr[$the insight is no greater than $z']$

$= \sum_{i=1}^{N} \Pr[$the goods with the insight are from i, sold in n at t$]$
$\cdot \Pr[$the insight is no greater than $z'|$the goods with the insight are from i and sold in n at t
$= \sum_{i=1}^{N} \pi_{nit} F_{i,t}(z')$ where N is the total number of economies, and $\pi_{nit} F_{i,t}(z')$ represents the probability that a good is from i but sold in n at time t has an insight component no greater than z'.

Therefore, we can substitute it into the following integral

$$\int_0^\infty x^{\rho\theta} dG_t(x) = \Gamma(1-\rho) \sum_{i=1}^{N} \left[\pi_{nit} (T_{i,t})^\rho\right].$$

Finally, the law of motion of the stock of knowledge is given by:

$$T_{n,t+1} - T_{n,t} = m_t \Gamma(1-\rho) \sum_{i=1}^{N} \pi_{nit} (T_{i,t})^\rho.$$