INTERNATIONAL MONETARY FUND

Information, Social Media and International Trade: Theory and Evidence Using Twenty Million Online Postings

George Cui and Kailin Gao

WP/26/64

***IMF Working Papers* describe research in progress by the author(s) and are published to elicit comments and to encourage debate.**
The views expressed in IMF Working Papers are those of the authors and do not necessarily represent the views of the IMF, its Executive Board, or IMF management.

2026
APR

WORKING PAPER

 WP/26/64

IMF Working Paper
Research Department

Information, Social Media and International Trade:
Theory and Evidence Using Twenty Million Online Postings*

Prepared by George Cui and Kailin Gao

Authorized for distribution by Petia Topalova
April 2026

ABSTRACT: We employ novel data and theoretical frameworks to investigate how a social media platform facilitates information exchange among firms. Our analysis is based on an extensive dataset comprising over 20 million firm-to-firm online interactions on a prominent social platform where participants share information about international trade. We document four empirical patterns. First, we find that firms' exports grow significantly after the firm begins using the social media platform. Second, firms located geographically closer exchange more information. Third, firms in sectors that have stronger production network relationships interact more on the platform. Finally, firms in more developed regions are more likely to adopt the social media platform. Motivated by these empirical patterns, we develop a quantitative general equilibrium trade model with information frictions and endogenous learning and information sharing.

RECOMMENDED CITATION: Cui, George, Kailin Gao. 2026. "Information, Social Media and International Trade: Theory and Evidence Using Twenty Million Online Postings." IMF Working Paper 26/64.

JEL Classification Numbers:	F10, D83
Keywords:	Information; Global Value Chains; Online Platforms
Author's E-Mail Address:	gcui@imf.org, and kailingao@uibe.edu.cn

* We are grateful to Kathryn Dominguez, James Hines, Andrei Levchenko, Sebastian Sotelo, and Petia Topalova for their invaluable guidance and support. We also thank Yan Bai, Jiaqian Chen, Yan Chen, Jaedo Choi, Ting Lan, Nan Li, Tanya Rosenblat, Younghun Shim and seminar participants at the University of Michigan Economics Department and Information School, Peking University China Center of Economic Research (CCER) Summer Institute, and the Chinese Economists Society (CES) North America Annual Conference 2025 for their insightful comments. The work in this paper is partly supported by the Macroeconomic Research on Climate Change and Emerging Risks in Asia program of the Ministry of Economy and Finance of the Government of Korea. We also thank Ross China Initiative for funding support. All errors are ours. The views in this paper are those of the authors and do not necessarily reflect the views of the International Monetary Fund, its Executive Board, and IMF Management. Email: gcui@imf.org, and kailingao@uibe.edu.cn.

WORKING PAPERS

Information, Social Media and International Trade: Theory and Evidence Using Twenty Million Online Postings

Prepared by George Cui and Kailin Gao

The various ways in which the knowledge on which people base their plans is communicated to them is the crucial problem for any theory explaining the economic process, and the problem of what is the best way of utilizing knowledge initially dispersed among all the people is at least one of the main problems of economic policy—or of designing an efficient economic system.

— (*Hayek*, *1945*)

Today . . . new transportation and communications technologies allow even the smallest firms to build partnerships with foreign producers to tap overseas expertise, cost-savings, and markets . . . [T]he Chinese and Indian entrepreneurs of Silicon Valley . . . are creating social structures that enable even the smallest producers to locate and maintain mutually beneficial collaborations across long distances.

— (*Saxenian*, *2000*)

1. Introduction

There is a long-standing question on the role of information frictions in international trade, and a more recent question about the role of new technologies in alleviating these frictions. Firms often face uncertainty about foreign demand, regulatory conditions, and competitive landscapes, which can hinder their ability to enter new markets or expand trade margins. Digital platforms—particularly social media—now offer a new channel through which firms may observe and learn from the experiences of others. However, the challenge faced by researchers is that neither firms' information sets nor the flows of information between them are typically observed. As a result, most existing work must infer what firms know based on their actions.[1]

This paper introduces, for the first time, a dataset of information flows among exporters. We track firms' engagement on a major social media platform in China where they share export-related knowledge, including trade experience, policy changes, and destination-specific conditions. We link these digital interactions to firm-level customs data. This allows us to quantify how information diffusion through social media affects firms' exports. The analysis provides new evidence on how digital technologies can reduce information frictions in international trade by enabling real-time, peer-to-peer learning among firms. In particular, we use more than 20 million postings by firms on a reputable Chinese social media platform designed to facilitate information exchange about international trade. Leveraging this novel database, we (i) estimate the impact of using the platform on firms' exports, and (ii) find new evidence to support a mechanism of information flows across firms: exporters could internalize knowledge spillover through production networks. We use our findings to build a general equilibrium quantitative trade model with firm's information acquisition and knowledge sharing. The model, in which firms can chose to mitigate information frictions by using social media platforms, allows us to quantify the impact of social media on welfare and perform counterfactual simulations.

[1] Dickstein and Morales (2018).

Several features of our study make it a particularly apt setting to examine the answers to these questions. First, exporting to foreign markets is widely recognized as an information-intensive activity, requiring firms to gather and process substantial knowledge about diverse consumer preferences, competitive landscapes, distribution networks, and regulatory environments (Rhee, 1989; Rauch, 2001; Rauch and Watson, 2002; Theodosiou and Katsikea, 2013). The existence of information frictions and the externalities associated with information spillover are used as justification for export promotion policies (Hausmann and Rodrik, 2003; Volpe Martincus and Carballo, 2012; Wei, Wei and Xu, 2021). Firms need to constantly acquire information about foreign market fundamentals, policy changes and essential "know-how" to survive market competition. However, there is little empirical evidence of how firms acquire information. Our novel data fill in this gap. Different from government-led platforms studied by Carballo et al. (2022), the social media platform is completely organized by private firms. Second, we combine data on the timing of firms' adoption of the social media platform and the information flows between firms with the universe of firms' trade transactions from the Chinese Customs. The combined data allow us to measure the effect of social adoption on firms' real activity, as captured by their export volumes.

Using this comprehensive dataset, we document several new empirical patterns which suggest that firms both acquire and share information endogenously and these decisions respond to incentives. First, we find that the social media platform has a positive impact on firm-level exports. Using a matching method for panel data developed by Imai, Kim and Wang (2023), we find that adopting the platform resulted in a 3 percent increase in firm exports within the same year, followed by 8 percent and 11 percent increases in the subsequent two years, respectively. Second, geographic distance is negatively correlated with information flows between platform users. Using a gravity framework at the region-sector pair period level, we find that information flows decline with geographic distance between the information sender and receiver controlling for a rich set of fixed effects. Third, we find that information flows across sectors increase with the strength of domestic supplier-buyer relationships. This suggests that firms are more likely to share information with their domestic buyers. We attribute this pattern to a new mechanism of information transfer: suppliers have an incentive to share knowledge and guide their domestic buyers in exporting to foreign markets due to their interconnected supply-chain relationships. Fourth, we find that the rate of adoption of the social media platform at the regional level is positively correlated with the regional GDP per capita and internet accessibility. This is suggestive of adoption frictions faced by firms, which vary across Chinese regions.

Motivated by these novel empirical findings, we construct a quantitative general equilibrium trade model. The model integrates endogenous information acquisition and sharing mechanisms in a monopolistically competitive market structure with heterogeneous firms. Building on our findings regarding the significance of geographical distance and production relationships for information flows, the proposed model divides the global economy into multiple domestic regions and one foreign region, each consisting of multiple sectors interconnected through input-output linkages. Within each region-sector, the industry structure is twofold, encompassing both goods and information industries.

In the goods industry, there is a continuum of firms who use labor and inputs to produce varieties. Firms are heterogeneous in their productivity. To maximize profits, firms make production decisions on hiring labor and input purchase, foreign and domestic market entry decisions, and information acquisition decision. Firms can pay a fixed membership fee to join the social media platform and post a question about foreign market entry. Each question can be answered by a firm in any domestic-region sector.[2] The firm receives a new idea from reading the responses to her own question and enjoys spillovers from other firms in the same region-sector, implying external economies of scale (Bartelme et al., 2025).[3] The higher the quality of the idea is, the lower is the foreign market entry cost the firm faces. The expected quality of ideas depends on the online information stock faced by the firms in the region-sector. The online information stocks are exogenous to the firms but endogenously determined in general equilibrium.

In the information industry, we adopt a semi-primal approach following Bai, Jin and Lu (2023) where there is one knowledge producer (innovator) in each domestic region-sector who chooses how much information to share and maximizes local firm profits.[4] The innovator uses labor to produce information flows that help domestic firms reduce export frictions. Information flows are pooled into information stocks in each region-sector.

We calibrate our model using firm-level and regional datasets. This calibration ensures a close linkage between the empirical moments and the theoretical framework of the model. Specifically, the structural equations of the model coincide with our reduced-form regression specifications, facilitating a precise mapping from the model's key parameters to observable moments. The firm-level export response to platform adoption informs the return to information stock, the relative distance elasticity of information and trade identifies the curvature of information provision, and the regional share of platform adoption among exporters disciplines heterogeneity in adoption costs and information frictions.

Related literature. Our paper relates to three strands of literature. The first strand of literature related to our paper examines information frictions in international trade. A question in this field is why international trade is nevertheless significantly lower than domestic trade, even after accounting

[2]In this model, we do not consider information diffusion from the foreign country to China for two reasons. First, our goal is to understand knowledge flows between domestic firms, which has received much less attention in the literature. Second, we do not observe many cross-border information flows. On the platform that we study, majority of users are from China. Few users outside China adopt this Chinese platform potentially due to language barrier.

[3]Though on the platform, firms can read and benefit from postings written by others within and outside her region-sector, we make this assumption to reflect the facts that firms response more to geographically closer firms. Though we assume the firm only reads questions raised by local firms, the responses to these questions can be from anywhere in the country. Thus, the across-region and sector information diffusion is captured in the model.

[4]The term "semi-primal" refers to an approach in optimization and economic modeling that blends elements of both primal and dual formulations. In this context, it highlights a decision-making process where an innovator directly chooses the level of investment in information sharing to maximize local firm profits, while also considering the impact of her decision on other innovators and firms.

for geographic distance. One explanation is the presence of substantial bilateral trade frictions related to culture, institutions, and information (Rauch, 2001; Anderson and Van Wincoop, 2004). Reducing these frictions could significantly increase trade and welfare. Information frictions in international trade are substantial. Empirical evidence on the source of information frictions often stems from studies on incomplete information about the prices of homogenous goods, showing that improved information technology reduces price dispersion. It is rare to directly measure the information diffusion process using data. A small body of literature shows relationships between observable communication and aggregate trade flows, including travel in Cristea (2011), Bernard, Moxnes and Saito (2019), Startz (2021) and Tian and Yu (2023), telephone calls in Portes and Rey (2005), and website viewer history in Carballo et al. (2022). Using online postings, we observe communication and outcomes at the micro level, allowing us to build a fully-specified model of international trade and firm behavior with underlying heterogeneity of information frictions. Our theory builds on the rising literature about production networks (Carvalho, 2014; Bernard, Moxnes and Saito, 2019; Carvalho and Tahbaz-Salehi, 2019; Liu, 2019; Acemoglu and Azar, 2020; Arkolakis, Huneeus and Miyauchi, 2023) and on technology adoption (Choi and Shim, 2022). We show that the intricate architecture of production networks facilitates collaborative behavior among firms, thereby enhancing overall welfare.

The second literature, pioneered by Hayek (1945), is on the economics of knowledge flows.[5] Two interdependent questions in this field are: how to accurately measure knowledge flows and what are drivers of knowledge flows. Measuring knowledge flows is inherently challenging due to their intangible nature.[6] Jaffe, Trajtenberg and Henderson (1993) introduced the use of patent citations as proxies for knowledge spillovers. Subsequent applications of citation data have spanned various domains (Caballero and Jaffe, 1993; Ellison, Glaeser and Kerr, 2010; Akcigit and Kerr, 2018; Cai and Li, 2019; Liu and Ma, 2021; Fadeev, 2024; Ahn et al., 2026; Choi, 2024; Choi and Shim, 2024). A rising literature uses large-scale non-conventional data, such as news papers (Bui et al., 2022), to measure information. Our research advances this literature by pioneering the use of online postings on a social media platform to measure knowledge flows. The rise of social media platforms makes online posting a ubiquitous activity for billions of people. These online posts provide valuable data sources for economists to study previously hard-to-measure economic behaviors. While we are not the first to use online postings, to the best of our knowledge, our study is the first use this data to measure firm-to-firm knowledge flows.[7]Unlike patent citations, online postings measure knowledge that is rarely patented or traded but remains economically important, such as information about foreign market fundamentals and practical know-how. Additionally, online postings reflect deliberate information-

[5]In this paper, we sometimes interchangeably use knowledge and information.

[6]Krugman (1991) highlights this difficulty, noting: "Knowledge flows, ..., are invisible; they leave no paper trail by which they may be measured and tracked..."

[7]For example, economists use online postings to study the labor market (Antenucci et al., 2014; Atalay, Sotelo and Tannenbaum, 2022), while other social scientists explore gender equality (Wang, Li and Wu, 2023). Some researchers construct sentiment index using the content of online posts(Antweiler and Frank, 2004).

sharing decisions made by economic agents, which are often not observable through other data sources. Our data provide a direct measure of learning through social interactions complementing a rising theoretical literature on the topic (Alvarez et al., 2013; Perla and Tonetti, 2014; Lucas Jr and Moll, 2014; Sampson, 2016; Buera and Lucas Jr, 2018; Buera and Oberfield, 2020; Perla, Tonetti and Waugh, 2021).

We also shed light on the mechanism and motivation behind knowledge flows. Understanding the mechanism of knowledge flows is vital in several fields, such as endogenous growth, international trade (Jones, 2005; Aghion, Akcigit and Howitt, 2014; Buera and Oberfield, 2020; Melitz and Redding, 2021) and economic agglomeration (Marshall, 2009; Carlino and Kerr, 2015). Traditionally, theorists have assumed that knowledge spillovers occur unintentionally, creating a free-ride problem and providing a common justification for government intervention (Bloom, Van Reenen and Williams, 2019; Bartelme et al., 2025). In international trade, this assumption gives rise of the debate of "missing pioneer" issue and justifies export promotion policies (Hausmann and Rodrik, 2003; Wei, Wei and Xu, 2021). Fadeev (2024) finds that patent citations are concentrated among business partners, suggesting the measure is not about spillover but intentional cooperation. Our paper shows that intentional knowledge sharing exists between firms engaging in international trade. Our model provides a new mechanism that firms could rationally help their downstream buyers to export so as to improve their own profits.[8]

The third strand of literature that we speak to is on the economic impact of information technology. Although it is believed that the internet and related information communication technologies (ICT) significantly affect economic growth and inequality, the empirical evidence is scant. The impact of ICT on real output is hard to evaluate. Regarding equality, theoretical predictions are mixed (Rosenblat and Mobius, 2004). Individuals who choose to use ICT may differ along various dimensions which influence economic outcomes. A rising series of papers show that ICT facilitates the creation of digital public goods on platforms such as Wikipedia (Thompson and Hanley, 2018; Chen et al., 2023; Hinnosaar et al., 2023). We bridge this literature with the international trade literature to shed light on the impact of ICT on both growth and inequality. We measure the growth effect of the digital public good provided by the social media platform. We are also able to use the model to capture the heterogenous impact of ICT on firms with different productivity and facing different technology adoption costs.

This paper studies information diffusion among firms on social media platforms and its role in reducing information frictions in international trade. The analysis does not examine, promote, or lend analytical support to regulatory evasion or efforts to circumvent trade barriers or policy frameworks. The empirical and theoretical findings are descriptive and analytical in nature, and should not be interpreted as an endorsement of any particular platform governance model, trade practice, or policy

[8]In this model, for tractability, we let firms and trade intermediaries to separately take the decisions of goods production and information sharing. The trade intermediaries maximize the local firms' profits, so one can understand that the trade intermediaries decide on information sharing on behalf of the firms. One can also construct a model with finite number of firms which make the decisions together.

position.

The remainder of the paper is as follows. Section 2 introduces the data and empirical findings. Section 3 presents a quantitative trade model. Section 4 connects the model with the empirical findings. Section 5 concludes.

2. Data and Empirical Patterns

This section provides an overview of the online social media platform used in the paper and summarizes key empirical findings documented within the data.

2.1 The Social Media Platform

We obtain our dataset from one of China's largest social media platforms dedicated exclusively to international trade. Founded in 2002, this platform serves firms actively participating in international commerce, used most notably by firms' salespersons, managers and owners.[9] Registration on the platform is free but mandates users to disclose specific details, including firm affiliation, industry, and geographical location.

The platform's core function is to facilitate the exchange of information among firms, thereby enhancing inter-firm connectivity and business opportunities. Consistent with insights from Hayek (1945) on dispersed local knowledge, critical intelligence about international market dynamics often resides with a limited group of firms. Platform interactions typically begin with a primary posting, usually an inquiry, such as "How do you locate international customers?" While initial inquiries are exploratory, subsequent responses tend to be directional and targeted, aimed explicitly at resolving the initial query.

To visualize the content of user inquiries, Figure 1 presents a word cloud constructed from posting titles, translated from Chinese to English using Google Translate. Word frequency, represented by text size, highlights significant themes and knowledge demands. Notably, frequently appearing terms include "help" and "please," signaling explicit informational needs, alongside terms such as product categories (fabrics, furniture).

Table 1 summarizes posting activities within subforums on the platform from December 2003 to July 2022. The most active subforum, labeled "Industry Communication," accounts for the largest proportion of discussion threads, underscoring the importance of industry-specific informational exchanges. Such postings collectively form a vast and publicly accessible digital repository that enhances the availability of information for all platform users. Importantly, the platform's public format allows even passive users ("bystanders") to access and benefit from information exchanges.

The design of this social media platform fosters decentralized learning through interpersonal interaction, a phenomenon referred to in existing literature as "social learning" (Mobius and Rosenblat,

[9]Other users include translators, legal advisors, certification specialists, and logistics professionals. We exclude these users in our main analysis and focus on firms directly engaging in international trade.

Figure 1: The Word Cloud of Posting Titles

Notes: This figure displays the word cloud of posting titles from the online platform. The original Chinese titles were translated into English using Google translate. The larger the key word is, the more frequent it appears in the titles.

2014). Figure 8 illustrates the typical information flows among firms via the platform. As depicted, firms may simultaneously serve as exporters and domestic market participants, establishing extensive buyer-supplier relationships. Managerial employees actively engage on the platform by posting, sharing information, and learning from peers, facilitating the identification of potential business partners and suppliers.

Crucially, the platform requires accurate disclosure of firm details, promoting accountability and enhancing the quality and reliability of shared information. Consequently, user-generated content becomes a credible mechanism through which firms establish and strengthen business relationships within and across national markets. Additionally, the platform is privately operated and derives revenue primarily from advertising rather than user fees. This revenue structure provides a strong incentive to uphold information quality. Thus, the platform employs strict monitoring, actively removing irrelevant or low-quality postings and banning users who routinely violate content guidelines. To facilitate user engagement and improve accessibility, the platform incorporates keyword-based search functionalities powered by established search engines such as Google and Baidu.

Overall, this digital infrastructure constitutes a valuable information resource for firms engaged in international trade. It simultaneously reduces informational frictions, encourages knowledge dissemination, and supports market transactions.

2.2 Data

Online Postings and User Registration Data. Our dataset includes user-generated postings and registration information collected from the platform between January 2003 to July 2022. Our main empirical analyses use postings over the 2007-2017 due to two reasons. First, we are able to obtain Chinese customs data before 2017. Second, the postings before 2007 were partially deleted by the platform. The platform database contains individual users' details, including the registration date

Table 1: Discussion Topics on the Platform

Topics (Subforum Names)	No. Threads	Share of Threads	No. Postings	Share of Postings
Industry Communication	866,243	32.69	4,256,779	16.66
Supporting Services	402,294	15.18	6,490,950	25.41
Business Communication	339,732	12.82	3,903,624	15.28
E-Commerce	221,175	8.35	2,109,699	8.26
Business Opportunities	141,634	5.34	1,020,601	4.00
Import	125,254	4.73	803,929	3.15
Regional Markets	119,229	4.50	1,482,254	5.80
Recommendations This Month	117,504	4.43	1,363,103	5.34
Wonderful Life in Foreign Trade	99,229	3.74	1,608,368	6.30
Same-City Communication	84,618	3.19	1,212,354	4.75
Related Skills	73,709	2.78	618,203	2.42
Manager Communication	39,479	1.49	403,566	1.58
Foreign Trade Related	19,884	0.75	272,327	1.07
Total	2,649,984	100	25,545,756	100

Note: This table summarizes the number of threads and postings on the subforums in our dataset. The subforums are ranked from the most to the least popular in this table.

and basic firm characteristics—specifically, industry affiliation, geographic location, and firm name. It also contains all online interactions among platform users.

For our analysis, we concentrate on 650,490 traders who provided complete city and industry information upon registration.[10] These traders represent 18.3 percent of all registered users and account for 26.2 percent of all online interactions. Among these traders, 224,718 actively participate by posting or responding in discussions, while 425,772 remain silent (passive) users. Notably, 98.77 percent of the traders in our sample originate from China.

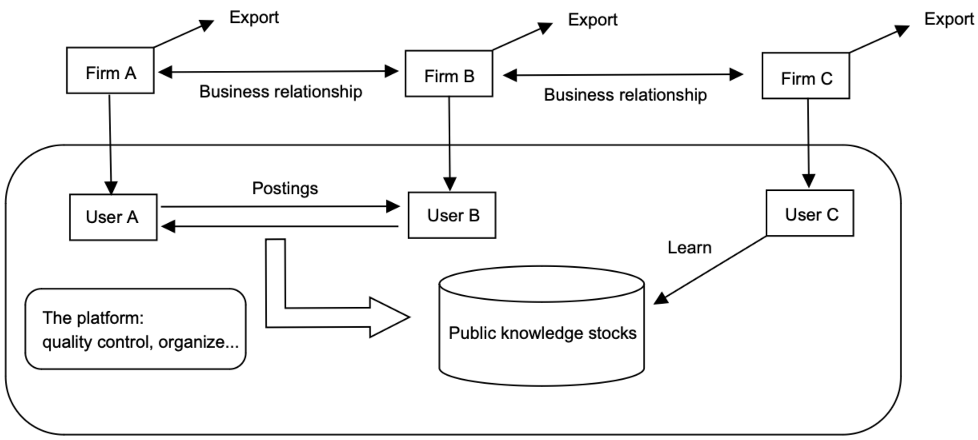
The user registration data can be seen on the platform's publicly through two sources: the platform's publicly accessible "yellow pages" directory, designed explicitly to facilitate business-to-business connections, and individual users' profile pages. From both sources, we extracted information regarding users' industrial affiliation, geographic locations, and firm names. Registered platform users can create new postings, reply to existing threads, and access historical postings, that have not yet been archived.

Chinese Customs Data. We complement the platform data with transaction-level trade records obtained from Chinese Customs, covering 2000-2017. Each customs transaction entry includes detailed information: the exporting or importing Chinese firm's name, the value of transactions in current US dollars, year of transaction,[11] product quantities, 8-digit Harmonized System (HS) product codes, and the foreign country counterpart.

[10]Typically there is only one registered user per firm. The platform requires the second user from a firm to upload supporting documents from the firm to get registered.

[11]Although the original Chinese customs records specify the exact dates (day and month) of individual transactions, our dataset lacks month and day information for recent years. To maintain consistency, we use only the year information in our analyses.

Figure 2: Across-Firm Information Flows on the Platform



Notes: This figure displays how information flows across firms on the platform. Firms export to foreign markets and have business relationships as suppliers and buyers in the domestic market. Managerial workers can register on the platform. Users post information to each other. The postings are processed by the platform by controlling the quality. The platform also organizes the postings so that they are more accessible to users. The postings become part of a public information stock which can be accessed by other users..

World Input Output Data. Lastly, we employ sector-level input-output data from the World Input-Output Database (WIOD), from 2006 to 2012, to quantify sector-to-sector economic relationships within China and internationally. These data enable us to explore detailed input-output linkages that inform our empirical analysis.

2.3 Empirical Patterns

Pattern 1: Platform Adoption and Dynamic Firm-level Export Effects

This section studies whether gaining access to the platform changes firms' export performance. We merge the platform registrations with Chinese customs transactions and define a firm's treatment date as the first year in which any of its employees registers on the platform. This definition is economically meaningful because registration unlocks the full archive of historical discussions and the ability to post questions, whereas guest users can only observe a limited subset of content.[12]

Data Matching. We match the firm names reported by platform users with exporter names in the Chinese customs data for 2000–2017. We first implement exact name matching and then apply fuzzy matching followed by manual verification. Approximately 10 percent of all Chinese exporting firms active at any point between 2000 and 2017 have at least one employee registered on the platform. We interpret this figure as a lower bound on platform adoption among exporters for several reasons. Some users do not report their firm names; some firms on the platform serve domestic exporters and

[12]Non-registered users (guest users) can read the first page of postings of recent threads.

Table 2: The Industry Distribution of Firms on the Social Media Platform and in the Customs Data

Industry (ISIC4)	Industry Description	Platform Firm No.	Platform Firm %	Customs Firm No.	Customs Firm %	Matched Firm No.	Matched Firm %
I. Production Industries							
C13-C15	Textiles, Apparel and Leather Products	106,080	17.61	179,351	18.84	14,996	16.68
C27-C28	Electrical Equipment and Machinery	94,398	15.67	147,241	15.47	14,617	16.25
C24-C25	Metal Products	64,960	10.78	79,449	8.35	9,157	10.18
C26	Computer, Electronic and Optical Products	41,832	6.94	140,760	14.79	15,152	16.85
C22-C23	Rubber, Plastic, Mineral Products	39,625	6.58	93,035	9.77	9,206	10.24
C31-C32	Furniture and Other Manufacturing	38,133	6.33	72,878	7.66	7,633	8.49
C19-C21	Petroleum, Chemicals, Pharmaceutical	36,782	6.11	68,936	7.24	7,762	8.63
C10-C12	Food Products	18,647	3.10	44,452	4.67	2,518	2.80
C16-C17	Wood and Paper Products	17,678	2.93	40,816	4.29	2,991	3.33
C29-C30	Vehicles and Transport Equipment	8,911	1.48	37,134	3.90	2,961	3.29
A01, A03	Crop, Animal, Fishing and Aquaculture	8,485	1.41	22,610	2.38	1,346	1.50
B	Mining and Quarrying	8,200	1.36	11,143	1.17	774	0.86
Others	Electricity, Support and Others	6,783	1.13	14,133	1.48	812	0.90
II. Service Industries							
H49-H53	Transportation	51,481	8.55				
G45-G47	Wholesale and Retail Trade	40,918	6.79				
M74-M75	Professional Activities and Technical Testing	8,168	1.36				
J62-J63	Programming and Information Service	7,361	1.22				
Total		602,424	100.00	951,938	100.00	89,925	100.00

Notes: This table presents the firm number in the platform, the customs data (2000-2017) and the matched sample. The industry classification for platform firms is determined by the most frequently reported industry by its worker reported on the platform (frequency method hereafter). Customs Firm industry and Forum-Customs matched firms' industry are both determined by the firm's largest import and export HS-2 digit industry from 2000-2017 (value method hereafter). The distribution of firms across industries, determined by classifying the industry of Forum-Custom matched firms using the frequency method, exhibits a high correlation of 0.98 when compared to the results obtained from the value method. We present the value method's result for the Forum-Customs matched firms since it more accurately determines a firm's industry. "Others" include ISIC Rev. 4 industries (D35-Electricity, Gas and Public Supply, J58-Publishing Activities, N-Administration and Support Service Activities, P85-Education, R_S-Other Service Activities) and HS 2-digit industries (97-98) since HS 97-98 (Art, Collection and Unclassified) do not explicitly align with any ISIC industries).

therefore do not appear in customs data; abbreviated or misspelled names can prevent successful matches; and some registered firms may be prospective exporters that have not yet exported directly.

Table 2 summarizes the distribution of platform users across industries, categorized according to the International Standard Industrial Classification of All Economic Activities (ISIC) Rev. 4. The industry distribution of platform users closely parallels the distribution of exporting firms in the customs records, suggesting that the matched sample is broadly representative rather than concentrated in a narrow set of sectors.

Empirical Strategy. Using this matched dataset, we ask: when a firm's first employee registers on the platform and gains access to information, how does the firm's export performance evolve over time? Our two main empirical challenges are familiar. First, firms may endogenously choose when to register: a firm preparing to enter new markets may be especially likely to seek information online, which would bias upward a naive estimate of the effect of platform access. Second, adoption is staggered across firms, so a simple two-way fixed effects design can be misleading in the presence of heterogeneous treatment effects (Goodman-Bacon, 2021; Callaway and Sant'Anna, 2021; Sun and Abraham, 2021; Baker, Larcker and Wang, 2022; Athey and Imbens, 2022; Liu, Wang and Xu, 2024).

To address these concerns, we estimate dynamic treatment effects using the time-series cross-sectional (TSCS) matching estimator developed by Imai, Kim and Wang (2023). The estimator combines matched comparisons with difference-in-differences logic and is well suited to staggered first adoption. We briefly introduce the method here and provide additional details in Appendix 5.1. For each firm i in year t, we observe exports Y_{it}, treatment status X_{it} indicating whether the firm has at least one employee adopting the platform in year t, and a vector of covariates Z_{it} including the number of exported product varieties, export destination count, export volume, and average unit value.[13] We set the number of leads to $F = 2$ and the number of lags to $L = 2$, given that the median exporter life span in the customs data from 2000 to 2017 is five years. The average treatment effect on the treated, evaluated F periods after adoption, is

$$\begin{aligned}\delta(F, L) =& \mathbb{E}\{Y_{i,t+F}(X_{it} = 1, X_{i,t-1} = 0, \{X_{i,t-\ell}\}_{\ell=2}^{L}) \\ & - Y_{i,t+F}(X_{it} = 0, X_{i,t-1} = 0, \{X_{i,t-\ell}\}_{\ell=2}^{L}) | X_{it} = 1, X_{i,t-1} = 0\}\end{aligned}$$

where the treated units are firms whose first observed employee registration occurs between 2000 and 2017. Intuitively, $\delta(F, L)$ compares the realized export outcome of a newly adopting firm F periods after adoption with the counterfactual outcome for an otherwise similar firm that has the same recent treatment history but does not adopt in year t.

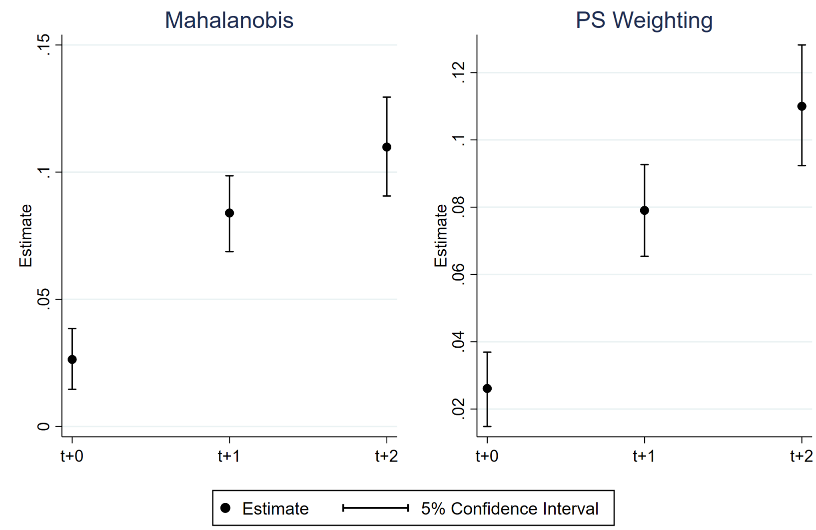
For each treated firm-year (i, t), we first build a matched set M_{it} of control observations from the same 2-digit HS sector that share the same treatment history over the previous $L = 2$ periods. This restriction absorbs cross-sector shocks such as tariff changes, demand shifts, and productivity movements that could otherwise contaminate the comparison. We then refine each matched set using either Mahalanobis matching or propensity-score weighting so that treated and control observations are similar in lagged export volume, destination count, product scope, and average unit values. The resulting estimates can be interpreted as dynamic event-study-style treatment effects at horizons $F = 0, 1, 2$.

Identification relies on a conditional parallel-trends assumption: after conditioning on recent treatment history and pre-treatment covariates, untreated firms in the matched set provide a valid counterfactual path for treated firms. Figure 9 reports standardized covariate differences in the two pre-treatment periods. Before refinement, future adopters are systematically larger and more internationally diversified than non-adopters. After either Mahalanobis refinement or propensity-score weighting, all balance statistics fall well within 0.25 standard deviations, consistent with the guideline in Stuart and Rubin (2008). These balance results do not prove the identifying assumption, but they make the comparison substantially more credible. We compute standard errors by bootstrap.

Results and Implications. Figure 3 and Table 3 report the dynamic treatment effects of platform adoption on exports over a three-year horizon. Under Mahalanobis matching, platform adoption raises export volume by 2.6 percent on impact, 7.9 percent after one year, and 11.0 percent after two

[13] We trim the upper and lower 5% of unit values following Khandelwal (2010).

Figure 3: The Impact of Platform Adoption on Firm-level Export Volume



Notes: This figure plots the estimated ATT$(F, 2)$ for $F \in \{0, 1, 2\}$ using the TSCS matching estimator. The left panel uses Mahalanobis refinement and the right panel uses propensity-score weighting. The black dots represent point estimates, and the vertical lines indicate 95% confidence intervals. The upward-sloping profile indicates that export gains build over time after first platform access.

years. Propensity-score weighting delivers nearly identical estimates: 2.6 percent, 8.4 percent, and 11.0 percent at horizons t, $t+1$, and $t+2$, respectively. The steadily increasing profile is consistent with firms gradually absorbing and applying information obtained through the platform rather than experiencing a purely transitory shock at the moment of registration.

These magnitudes are economically meaningful. They are in the same order of magnitude as the export gains from export-promotion policies reported by Munch and Schaur (2018), smaller than the large trade-policy shocks studied by Jiao et al. (2024), and close to the 17 percent export increase found by Carballo et al. (2022) for Peruvian firms joining ConnectAmericas. We therefore interpret the estimates as sizable reduced-form effects of access to platform-mediated information, including both active participation and passive learning from the stock of archived discussions.

Heterogeneity Analysis on External Economics of Scale. Given the average positive impact of platform adoption on firm export volume, firms may experience heterogeneous effects due to differences in external economies of scale (EES). The EES mechanism, widely discussed in the rising literature of industrial policy (Bartelme et al., 2025; Lashkaripour and Lugovskyy, 2023), suggests that firms benefit from positive externalities created by other entities within the same industry or geographic area, leading to lower average costs of production. While firms do not directly induce these cost reductions, they gain from spillover effects generated by industry-wide knowledge sharing, agglomeration, and network effects.

In an online platform environment, EES primarily occurs through knowledge spillovers. Users benefit from reading and responding to others' postings, leading to the diffusion of market information

Table 3: Dynamic Effects of Platform Adoption on Firm-Level Export Volume

	Mahalanobis Matching		PS Weighting	
Year	Estimate	S.E.	Estimate	S.E.
t	0.0261	(0.0057)	0.0264	(0.0060)
$t+1$	0.0791	(0.0070)	0.0839	(0.0076)
$t+2$	0.1100	(0.0091)	0.1099	(0.0098)
No. of treated firms	23,710		23,710	
No. of untreated firms	228,073		228,073	
No. of control firms in matched set	77,760		228,073	

Notes: This table reports ATT$(F, 2)$ estimates of the dynamic effect of platform adoption on firm-level export volume using Mahalanobis matching and propensity-score weighting (PSW). Estimates represent log-point changes in exports relative to the matched control group at horizons $F = 0, 1, 2$. Standard errors are in parentheses.

and export-related strategies. Additionally, as more firms adopt the platform, offline spillovers through face-to-face interactions within the same region and sector further reinforce these effects.

To empirically test for EES, we estimate the following reduced-form specification:

$$\ln x_{zis,t} = \alpha + \beta_1 \ln H_{is,t} \times T_{zis,t} + \beta_2 T_{zis,t} + \zeta_{is,t} + \delta_z + \varepsilon_{zis,t}$$

where $H_{is,t}$ is the information stock at location-sector level measured by the number of responses to the postings generated by users in location i sector s in year t. Information stock increases with both the active user number and posting intensity. β_1 is the coefficient of interest, indicating the scale effect, or in other words, whether firms that register for the platform in location-sectors that have already built a larger stock of knowledge experience greater benefits to adoption.

The results, reported in Table 4, indicate that a 10% increase in information stock is associated with a 0.3% increase in export volume for firms that adopt the platform relative to those that don't. Column (3) further refines this analysis by weighting postings by their length, yielding similar results. The interpretation of the interaction term coefficient suggests that EES enhances the benefits of platform adoption, reinforcing the role of information diffusion in export expansion.

Pattern 2: Information Flows and Geographic Distance

Geography has been shown to be a key barrier that hinders both information and trade flows. Understanding the impact of geographic frictions on information flow is important, because it helps us understand the agglomeration of economic activities. We use distance between Chinese cities to measure geographic frictions within the country. We measure information flows at sector-region-pair-period level by aggregating the number of postings from a particular source to a particular destination. Figure 4 shows the geographic distribution of information flows and stocks. Panel A shows the information flow networks. City-pairs that have positive knowledge flows are connected by

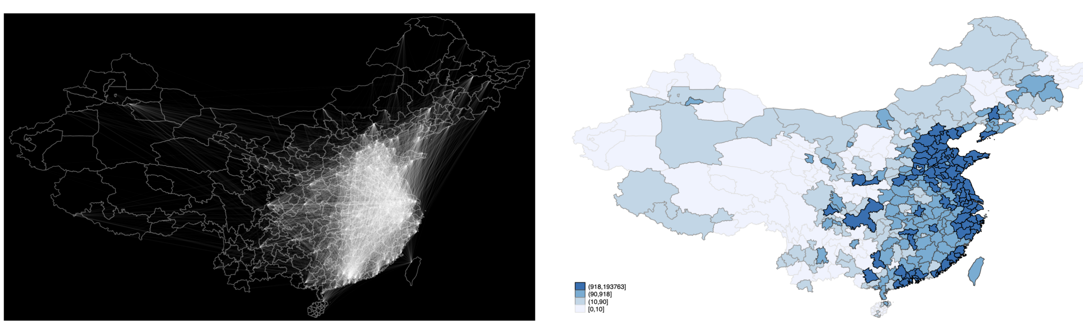
Table 4: Estimating the Impact of Information Stock on Firm-Level Export

	(1)	(2)	(3)
Adoption	0.326***	0.033	-0.063
	(0.008)	(0.049)	(0.058)
Adoption × ln(Information Stocks)		0.031***	
		(0.005)	
Adoption × ln(Weighted Information Stocks)			0.026***
			(0.004)
Observations	2,791,754	2,791,754	2,791,754
Adjusted R-squared	0.691	0.691	0.691
Province-Year FE	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes

Notes: This table reports the impact of information stock on firm-level export performance. Column (1) presents the baseline effect of platform adoption, while Column (2) introduces an interaction term between adoption and the logarithm of information stock. Column (3) further weights the information stock by posting length. Standard errors are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1..

unweighted lines. We observe that information flow across regions due to the internet. Inland regions are connected with coastal regions through the information networks. In Panel B the deeper shade indicates higher information stock. One can observe the graphic concentration of information in the coastal regions. Coastal and economically big cities are more intensively connected by information flows and receive larger information stocks. Figure 5 shows the information flow heatmap across sectors in the five largest export regions in China: Guangdong, Zhejiang, Jinagsu, Shangdong and Shanghai. The concentrated diagonal shows that within city-sector information flow is intensive. Our results show that information diffusion is local even if communication cost is homogeneous across regions on the internet.

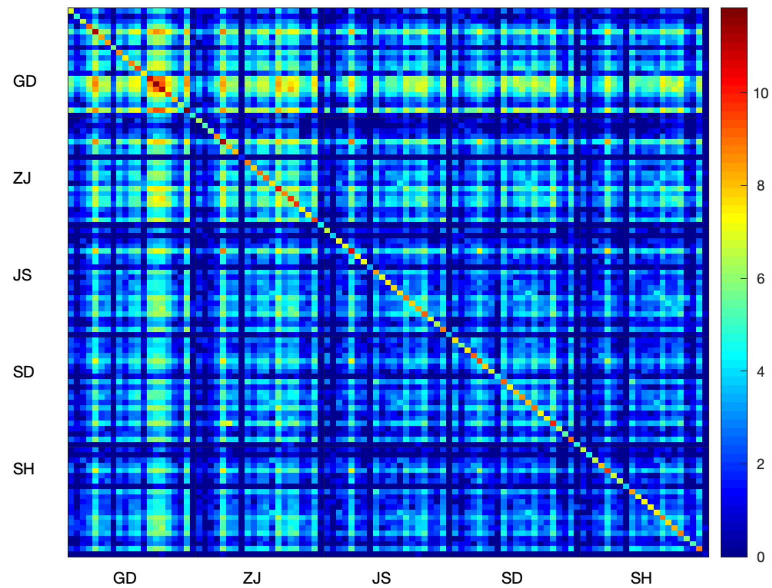
Figure 4: The Geographic Distribution of Information Flows and Stocks



A. Information Flows B. Information Stocks

Notes: This figure displays the geographic distribution of information flows measured by information posting number between cities and the information stocks measured by the total inward posting number. Panel A show the information flow networks. City-pairs that have positive information flows are connected by unweighted lines. In Panel B the deeper shade indicates higher information stock.

Figure 5: Information Flows across Regions and Sectors



Notes: This figure displays the heatmap of information flows (posting number) between sectors of the five largest Chinese exporting regions (Guangdong, Zhejiang, Jiangsu, Shandong, and Shanghai) in our dataset. The sources of information (teaching) are on the y-axis, and the destination of information (learning) are on the x-axis. The colors code the logarithm number of postings. Each region has 21 sectors. The sector labels are suppressed due to lack of space.

To formally test the local diffusion of information, we estimate the following gravity equation using Poisson pseudo-maximum likelihood (PPML):

$$\text{InfoFlow}_{aibjt} = exp\left(\beta^{Dist}\log \text{Dist}_{ij} + Z_{ij} + \delta_{abt} + \xi_{ait} + \eta_{bjt}\right) \times \varepsilon_{aibjt} \tag{2.1}$$

where InfoFlow_{aibjt} is the total number of postings responding to inquiries from location i sector a by location j sector b in period t, Dist_{ij} is the geographic distance between location i and j in kilometers[14], Z_{ij} includes two dummy variables of same city and same province accounting for the home bias in knowledge, and δ_{abt}, ξ_{ait}, and η_{bjt} indicate sector-pair-time, origin city-sector-time and destination city-sector-time fixed effects, respectively. The sector-pair-time fixed effects absorb any factors that vary across sector-pair and over time, such as input-output relationships. The origin city-sector-time and destination city-sector-time fixed effects absorb factors such as the number of users, the endowment of export knowledge, productivity, and demand for inputs in each city-sector. Furthermore, we also run the same regression by weighting the posting number with the character number of postings to proxy for the quality of information flows.

Table 5 presents the results. Column 1 shows that the elasticity of the number of postings with

[14]We impute with-city distance by 1 km following Arkolakis, Huneeus and Miyauchi (2023).

respect to distance is significant at -0.064, indicating that a 10% increase in the distance between two cities is associated with a 0.64% decrease in flows. The results in Column 2, which weigh the number of postings by the number of characters to account for posting quality, are qualitatively similar.

By comparing the estimation results with the trade literature, we find information flows decline over distance at a relatively lower rate than conventional trade flows. For example, Arkolakis, Huneeus and Miyauchi (2023) estimate a gravity equation using the municipality-to-municipality trade flow data from Chile. They report a coefficient of -1.334, meaning a 10% increase in the distance between two cities is associated with a 13.34 % decrease in aggregate trade flows from one city to another.

To the best of our knowledge, we are among the first to use online postings as a direct measure of information flows and to estimate the elasticity of information flows with respect to geographic distance. While it is widely acknowledged that distance impedes information flows, the underlying mechanisms remain unclear.[15]

A standard argument in the literature suggests that information and knowledge is more efficient to be physically transmitted through face-to-face interactions or travel. However, our findings cannot be explained by this theory, as communication costs on social media platforms are geographically invariant. Answering a question posed by someone 100 km away is no more costly than responding to a question from a nearby user.

Another possible explanation is that firms encounter more similar inquiries within the same region, leading to more frequent information exchanges. We address this concern by incorporating an exhaustive set of fixed effects and control variables in our gravity regression, yet distance continues to have a significant impact on information flows.

This finding motivates our theory that firms have stronger incentives to share information with geographically closer peers due to higher expected business returns. Consider an input supplier: she has a greater incentive to assist a nearby input buyer in expanding exports, as the buyer's success is more likely to increase future demand for her inputs. Lower geographic distance enhances incentive compatibility among firms within supply chains, fostering more localized information exchange.

In the next section, we further explore how upstream and downstream relationships shape information-sharing patterns, providing additional empirical evidence for this mechanism.

Pattern 3: Information Flows and Production Network Relationship

We examine how value chain relationships affect sector-to-sector relationships in the information networks. As is shown in Figure 4 Panel A and Figure 5, information diffuses across space. There is significant variation in cross-sector information flows.

We formally examine how the sector-to-sector value chain relationship in 2006 affects information network relationships from 2007 to 2017. Following Antràs et al. (2012) and Bui et al. (2022), we

[15]Cotterlaz and Guillouzouic (2025) find that patent citation declines with geographic distance between firms. They argue that firms' knowledge network expands through the sources of their existing knowledge sources.

Table 5: Gravity Regression of Information Flows

	(1) Unweighted	(2) Weighted
Log Distance	-0.0644***	-0.0846***
	(0.0109)	(0.0178)
Same City	2.523***	2.866***
	(0.0685)	(0.100)
Same Province	0.284***	0.322***
	(0.0398)	(0.0693)
Observations	279,172	279,172
Origin City-Sector-Time FE	Yes	Yes
Destination City-Sector-Time FE	Yes	Yes
Sector-Pair-Time FE	Yes	Yes

Notes: This table reports the results of a gravity regression estimating the relationship between information flows and geographic distance. The dependent variable in Column (1) is the total number of postings between city-sector pairs. Column (2) weights postings by character count. The key explanatory variable is the logarithm of the geographic distance (in kilometers) between cities. Control variables include indicators for same-city and same-province pairs. Fixed effects account for origin city-sector-time, destination city-sector-time, and sector-pair-time. Standard errors, clustered at the city-sector-pair level, are reported in parentheses. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

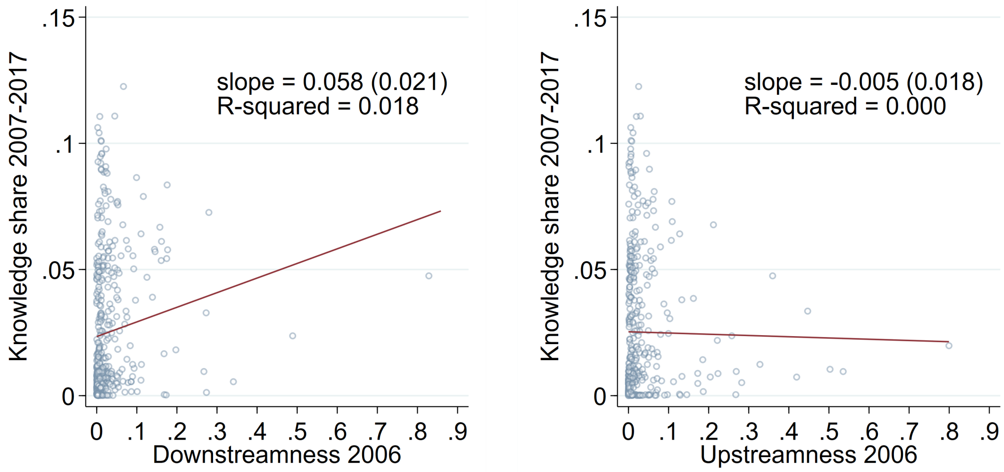
measure relationships using the posting data and global trade data from WIOD:

$$\text{InfoNetworks}_{ij,07-17} = \frac{\text{InfoFlow}_{ij,07-17}}{\text{InfoSupply}_{i,07-17}}, \quad \text{Down}_{ij,06} = \frac{\text{InputFlow}_{ij,06}}{\text{InputSupply}_{i,06}}, \quad \text{Up}_{ij,06} = \frac{\text{InputFlow}_{ji,06}}{\text{InputUsage}_{i,06}},$$

where $\text{InfoFlow}_{ij,07-17}$ is the number of replies sent from sector i to j from 2007 to 2017; $\text{InfoSupply}_{ij,07-17}$ is the number of replies sent from sector i to any domestic sector from 2007 to 2017. InfoNetworks_{ij} is the share of information owned by sector i that is shared with sector j. $\text{InputFlow}_{ij,06}$ measures the intermediate input flow from sector i to j within China in year 2006; $\text{InputSupply}_{i,06}$ is the intermediate input supplied by a Chinese sector i to any domestic and foreign sector in year 2006; $\text{InputUsage}_{i,06}$ is the intermediate input purchased by a Chinese sector i from any domestic and foreign sector in year 2006. $\text{Down}_{ij,06}$ measures the importance of sector j to sector i as a downstream input buyer, while $\text{Up}_{ij,06}$ measure the importance of sector j to sector i as a upstream input seller.

Panel A of Figure 6 illustrates that information flows are is significantly and positively correlated with the strength of the downstream relationship but not with the strength of upstream relationship. The positive and statistically significant coefficient suggests that when a sector increases its purchases of inputs from another sector by 10% (as a share of the supplying sector's total output in 2006), the supplying sector subsequently shares approximately 0.6% more knowledge with the purchasing sector over the period 2007 to 2017. This empirical pattern provides a supporting evidence of the "learning from sellers" story in the literature (Buera and Oberfield, 2020).

Figure 6: Sector-to-Sector Information Networks and Global Value Chain Relationships



A. Knowledge Share vs Downstreamness B. Knowledge Share vs Upstreamness

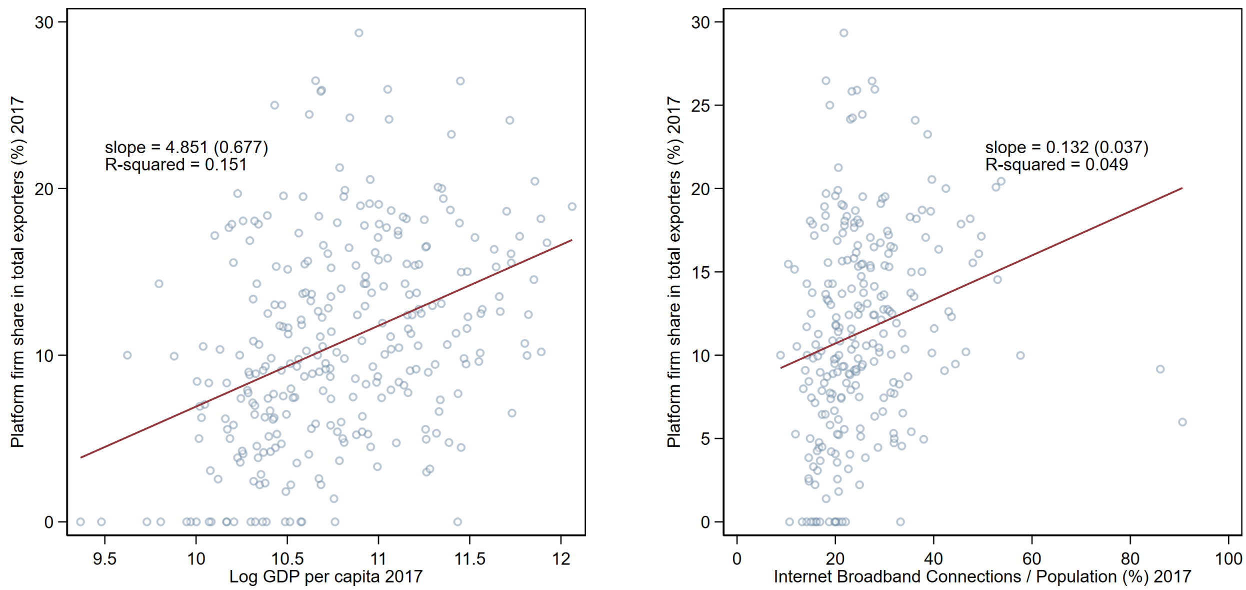
Notes: This figure displays the sactterplots of the knowledge shares between different sectors on the y-axis in panel A and B against the relative downstreamness (panel A) and the relative upstreamness (panel B). The "knowledge share" is the share of teaching sector's knowledge outflow (measured by posting number) to the learning sector divided by the teaching sector's total knowledge outflow from 2007 to 2017 on the platform. The relative downstreamness of the learning sector to the teaching sector is measured by the share of teaching sector's total input production purchased by the learning sector in 2006 WIOD. The relative upstreamness of the learning sector to the teaching sector is measured by the share of teaching sector's total input purchase supplied by the learning sector in 2006 WIOD. All panels exclude within sector observations. All plots report the bivariate regression slope coefficient, robust standard error, and the R^2. Each dot is a sector-pair observation.

Pattern 4: Platform Popularity and Regional Characteristics. Since the development of Internet, people have been debating whether it provides more equal opportunity for gaining access to information (Rosenblat and Mobius, 2004). We show some descriptive evidence on the heterogeneity in information technology adoption across regions to shed light on this question.

Figure 7 shows that the share of exporting firms in a city that have at least one employee adopting the platform significantly increases with the city-level GDP per capita and internet access, measured by the Internet broadband connection number over population in year 2017. Despite the positive impact of social media adoption on firm-level exports, not all firms choose to adopt the information technology, suggesting the existence of technology adoption costs. Though registering an user account on the platform is free, there are costs of getting access to the internet especially in low-income regions.

This pattern is somewhat surprising because one would expect an online platform would benefit low-income regions where firms have less information through off-line interactions. Two factors rationalize this pattern: First, in terms of cost, less developed regions have lower internet access and less awareness of online resources. Second, firms in less developed regions are less likely to be helped by peers because the expected economic benefit from them is low.

Figure 7: Platform Adoption and City Characteristics



Notes: This figure displays the scatterplots of platform firm share in total exporters against the city-level logarithm GDP per capita and internet access.

2.4 Empirical Summary and Connection to Theory

Our empirical analysis identifies several key patterns that shed light on the role of social media platforms in facilitating information flows and shaping firm-level export performance.

First, platform adoption has a significant and persistent impact on firm exports, with evidence suggesting that external economies of scale (EES) amplify these effects. Firms that begin using the platform experience a 3% increase in export volume in the first year, followed by 8% and 11% increases in the subsequent two years. This pattern suggests that access to trade-related information reduces informational frictions, allowing firms to expand their participation in international markets. Moreover, firms located in regions with larger information stock measured by platform interactions benefit more from information spillovers, consistent with the presence of EES. As more firms within a region or industry adopt the platform, the accumulated knowledge base grows, enhancing the quality and availability of trade-relevant information. This amplification effect suggests that individual firms not only gain from their own platform adoption but also benefit from broader industry-wide and platform-wide information diffusion.

Second, despite the digital nature of communication, information flows are constrained by geographic frictions. Our gravity regression indicate that a 10% increase in distance between two cities reduces information flows by 0.64% to 0.85%, even though the platform enables costless online interactions. The persistence of distance-related frictions suggests that the effectiveness of digital knowledge exchange remains rooted in traditional spatial economic relationships. This pattern can be explained by the complementarity between online interactions and offline business relationships.

Third, information flows align closely with supplier-buyer relationships within domestic production networks. Firms are significantly more likely to share export-relevant information with their domestic buyers rather than with unrelated firms. This pattern implies that suppliers actively assist their downstream buyers in accessing foreign markets, potentially as a strategic response to maximize their own long-term gains from supplying expanding exporters. The observed information-sharing behavior challenges the conventional assumption that knowledge spillovers occur passively. Instead, it suggests that firms internalize knowledge-sharing incentives, particularly within supply chains where downstream success directly benefits upstream firms.

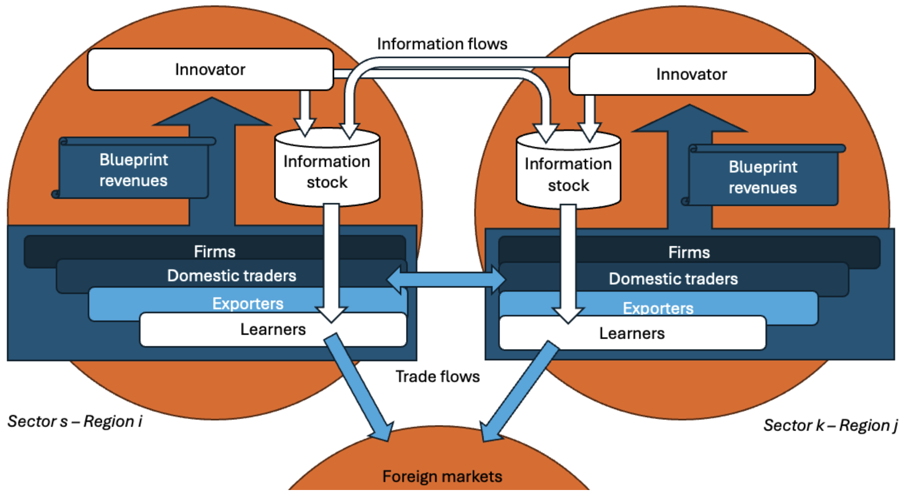
Fourth, there is substantial regional heterogeneity in platform adoption, with firms in more developed regions of China being more likely to register. The share of exporting firms using the platform is significantly higher in cities with higher GDP per capita and better internet infrastructure, while firms in less developed regions exhibit lower adoption rates. This pattern indicates that firms in low-income areas may face higher barriers to accessing online information, despite potentially greater informational needs.

The empirical findings highlight the complex nature of information flows in international trade. While digital platforms reduce direct communication costs, firms' incentives to exchange information remain shaped by geographic proximity, production linkages, and economic complementarities. The strong relationship between information flows and supply chains suggests that information acquisition is not solely an independent firm decision but rather a strategic and network-driven process.

This insight has important theoretical implications. Existing trade models often treat information frictions as exogenous constraints on firm behavior. However, our findings suggest that information diffusion is endogenous to supply-chain relationships. In particular, suppliers may strategically share information with buyers to expand their own market opportunities, reinforcing a network-based mechanism of knowledge transmission (Buera and Oberfield, 2020). Additionally, the persistence of geographic frictions in information flows suggests that even in a digital era, physical location continues to play a role in shaping trade-related knowledge diffusion.

Motivated by these findings, the next section develops a quantitative general equilibrium trade model that incorporates both endogenous information acquisition and strategic knowledge sharing. The model features geographic and sectoral frictions in information diffusion, capturing the observed decline in information flows over distance and their alignment with input-output relationships. It also formalizes firms' strategic incentives to share information within production networks, demonstrating how suppliers internalize knowledge-sharing benefits. By integrating these mechanisms, the theoretical framework provides a micro-founded explanation for how digital platforms reshape trade by mitigating—but not eliminating—information frictions.

Figure 8: The Model Environment



Notes: This figure shows the model environment. There are multiple domestic region-sectors and one foreign market. In each domestic region-sector, there is one innovator and a continuum of intermediate goods producers (firms). The innovator supplies production technologies (blueprints) to local firms and information flows (white arrows) to both the local and the other region-sector. Information flows are aggregated into information stocks in each region-sector. Firms endogenously choose to enter the other region-sector market (becoming domestic traders), enter foreign market (becoming exporters), and learn from the information stock (becoming learners). The innovator collects all local firm profits through selling blueprints, and optimally chooses information flows to maximize her profits.

3. The Theoretical Framework

In this section, we develop a general equilibrium trade model with endogenous information acquisition ("learning") and provision ("teaching"). This model is a unified framework to rationalize the patterns documented in our data. The model is built on Choi and Shim (2022) which studies technology adoption, and Costinot and Rodríguez-Clare (2014). We introduce the theoretical framework, connect the theory with empirical patterns and estimate the key structural parameters. Taking the model as a vehicle, we conduct quantitative analysis in the next section.

Set-Up. Figure 8 shows the model environment. We consider a world consisting of N domestic locations and 1 foreign location. The domestic locations are denoted by i or $j = 1, 2, ..., N$. The set of domestic locations is denoted as $\mathcal{N}$. The foreign location is denoted as x. In each location, there are S intermediate goods sectors denoted by s or $k = 1, 2, ..., S$. The exogenous measure of households in domestic and foreign locations are respectively denoted as L_i, L_j and L_x. The households inelastically supply labor and earn wage w_i. Labor is not mobile across locations, but mobile across sectors and activities within each location.

There is a continuum of firms in each location i-sector s, indexed by $\omega \in \Omega_{i,s}$. $\Omega_{i,s}$ denotes the set of firms. We denote the entire set of firms in the world economy by $\Omega \equiv \bigcup_{i \in \mathcal{N} \cup \{x\}} \bigcup_{s=1}^{S} \Omega_{i,s}$. Each firm pays a fixed entry cost $c_{i,s} f_{i,s}^e$ to purchase a "blueprint" of intermediate variety, where $c_{i,s}$ is the unit cost of production, which will be endogenously determined in equilibrium. $f_{i,s}^e$ is an entry cost shifter.

The firm produces a distinct variety that can be used as both intermediate input by other firms and final goods by households. Upon purchasing the blueprint, every firm draws a productivity z from the cumulative distribution function, $G_{i,s}(\cdot)$. Following Chaney (2008), we assume $G_{i,s}(\cdot)$ is a Pareto distribution, such that $G_{i,s}(z) = 1 - \left(z/z^{min}_{i,s}\right)^{-\theta}$, where the position parameter $z^{min}_{i,s}$ is the "natural advantage" of location i sector s and the shape parameter θ governs the dispersion of productivity. Intermediate goods and final goods are traded across locations with iceberg trade costs $\tau_{ij,s} \geq 1$ imposed on trade flow from i to j in sector s.

Each location-sector has a knowledge industry. We call the knowledge producer the "innovator". An innovator in location i-sector s supplies "blueprints" of producing varieties to local firms ($\omega \in \Omega_{i,s}$) and information flows to any domestic locations and sectors ($i \in \mathcal{N}, s \in S$). The innovator collects all profits from the local location-sector and chooses the optimal information flows to maximize the profits. To capture the domestic information diffusion, we only allow information flows between domestic locations and sectors. We take the potential information diffusion between countries as exogenously given.

Preference. The representative consumer in location j maximizes:

$$C_j = \prod_{s=1}^{S} C_{j,s}^{\beta_{j,s}}, \quad C_{j,s} = \left(\int_{\omega\in\Omega} c_{j,s}(\omega)^{(\sigma_s-1)/\sigma_s} d\omega\right)^{\sigma_s/(\sigma_s-1)} \tag{3.1}$$

where $\beta_{j,s} \geq 0$ are exogenous preference parameters satisfying $\sum_{s=1}^{S}\beta_{j,s} = 1$ and $C_{j,s}$ is total consumption of the composite good s in location j. $c_{j,s}(\omega)$ is the quantity of variety ω in sector s that is consumed by the household in location j. σ_s is the elasticity of substitution between different varieties within sector s. Each variety is sourced from only one location, so that

$$C_{j,s}^{\frac{\sigma_s-1}{\sigma_s}} = \sum_{i=1}^{N} C_{ij,s} + C_{xj,s}, \quad C_{ij,s} = \int_{\omega\in\Omega_{ij,s}} c_{j,s}(\omega)^{(\sigma_s-1)/\sigma_s} d\omega, \quad C_{xj,s} = \int_{\omega\in\Omega_{xj,s}} c_{j,s}(\omega)^{(\sigma_s-1)/\sigma_s} d\omega,$$

where $C_{ij,s}$ denotes the varieties location j sector s sourced from location i sector s and $\Omega_{ij,s}$ denotes the set of firms in $i-s$ selling varieties to j. The associated consumer price index is

$$P_j = \prod_{s=1}^{S} P_{j,s}^{\beta_{j,s}}, \quad P_{j,s}^{1-\sigma_s} = \sum_{i=1}^{N} P_{ij,s}^{1-\sigma_s} + P_{xj,s}^{1-\sigma_s}. \tag{3.2}$$

and the bilateral price index is

$$P_{ij,s} = (\int_{\omega\in\Omega_{ij,s}} p_{j,s}(\omega)^{1-\sigma_s} d\omega)^{1/(1-\sigma_s)}, \quad P_{ix,s} = (\int_{\omega\in\Omega_{ij,s}} p_{j,s}(\omega)^{1-\sigma_s} d\omega)^{1/(1-\sigma_s)}, \tag{3.3}$$

where $p_{j,s}(\omega)$ is the price of variety ω in sector s that is faced by the households in location j.

Timeline. The model timeline is as follows: First, the innovators simultaneously decide the level of information flows to provide to each domestic location-sectors and the number of blueprints to sell to local firms. Second, information flows form information stocks in each location-sector. Third, the firms purchase blueprints and make production and information learning decision taking the information stocks as given. Finally, consumers purchase final goods. We now introduce the model details and solve it backward, starting with the firms' problem.

3.1 Firms' Problem Given Information Stocks

Production Technology. The intermediate goods are produced in the same way as composite goods for final consumption:

$$I_{j,s} = \left(\int_{\omega\in\Omega} i_{j,s}(\omega)^{(\sigma_s-1)/\sigma_s} d\omega\right)^{\sigma_s/(\sigma_s-1)} \tag{3.4}$$

where $i_{j,s}(\omega)$ is the quantity of variety ω in sector s that is purchased by the firms in location j.

Firms use both labor and intermediate inputs to produce varieties, such that the the unit production cost can be expressed as $c_{i,s} = w_i^{1-\alpha_{i,s}} \prod_{k=1}^{S} P_{i,k}^{\alpha_{i,ks}}$ where $\alpha_{i,s}$ and $\alpha_{i,ks}$ are exogenous parameters satisfying $\alpha_{i,s} + \sum_k \alpha_{i,ks} = 1$.

Learning through Platform. Firms in domestic location i-sector s face international trade cost $\tau_{ix,s}$, part of which is due to information frictions. Firms can learn about exporting so as to reduce the international iceberg trade cost.

Motivated by the platform adoption in the data, we model this learning behavior as a binary technology adoption decision (Melitz, 2003; Yeaple, 2005; Bustos, 2011; Choi and Shim, 2022): Firm can pay a fixed learning cost $c_{i,s}f_{i,s}^{L}$ to draw an idea from the information stock $H_{i,s}$. Following Eaton and Kortum (2002) and Buera and Oberfield (2020), we assume that firm in location i-sector s draws an idea z^x from a Fréchet distribution $F_{i,s}(z^x) = exp\left[-H_{i,s}(z^x)^{-\bar{\beta}}\right]$ where $H_{i,s}$ is the location i-sector s's information stocks and $\bar{\beta} > 1$ is the distribution shape parameter. The knowledge stocks in each location are taken as given by firms but are endogenously determined in the general equilibrium by innovators. We assume the idea draw is independent from the productivity draw.

Given the Fréchet distribution, the expected idea draw of a firm in location i-sector s can be expressed as:

$$\eta_{i,s} \equiv \Gamma(1-1/\bar{\beta})H_{i,s}^{1/\bar{\beta}} = \Gamma(1-1/\bar{\beta})\left(\sum_{k\in\mathcal{S}}\sum_{u\in\mathcal{N}} H_{uk,is}\right)^{1/\bar{\beta}} \tag{3.5}$$

where $H_{uk,is}$ is the information flow from location-sector $u-k$ to $i-s$. $\Gamma(1-1/\bar{\beta})$ is a constant term denoting the Gamma function of $\bar{\beta}$. The parameter $\bar{\beta}$ controls the dispersion of idea draws. Economically, it determines how effective additional information flows are in raising expected information quality. A small $\bar{\beta}$ corresponds to a fat-tailed idea distribution: information is highly variable, and expanding the information stock can occasionally generate "blockbuster" insights with large export cost reductions. In this case, the returns to information are strong but volatile. By contrast, a large $\bar{\beta}$ implies a thinner-tailed distribution: idea draws are more homogeneous, so accumulating more information stock has more predictable but incremental effects on reducing trade costs. One can also compare $1/\bar{\beta}$ with the scale elasticity in Bartelme et al. (2025), given that local knowledge stock ($H_{i,s}$) is a function of sectoral size. The realized international iceberg trade cost faced by domestic firms is $\tau_{ix}(\eta_{i,s})^{-T_{i,s}(\omega)}$ which depends on the endogenous learning decision, $T_{i,s}(\omega)$, that potentially varies across firms. Deviating from the standard trade models, our model shows that firms in the same location and sector are now potentially facing heterogeneous information spillovers and therefore different variable trade cost due to endogenous learning decision.

Firm's Problem. Firms pay a fixed entry cost $c_{i,s}f_{i,s}^{e}$ to the local innovator to purchase a blueprint. Firms draw a core productivity from a Pareto distribution: $G_{i,s}(z) = 1 - (z/z_{i,s}^{min})^{-\theta}$. In a monopolistically competitive market, the optimal prices charged by a firm ω in location i-sector s to a buyer in domestic location j and foreign location x are respectively:

$$p_{ij,s}(\omega) = \frac{\sigma_s}{\sigma_s - 1}\frac{\tau_{ij,s}c_{i,s}}{z_{i,s}(\omega)},, \quad p_{ix,s}(\omega) = \frac{\sigma_s}{\sigma_s - 1}\frac{c_{i,s}}{z_{i,s}(\omega)}\tau_{ix}(\eta_{i,s})^{-T_{i,s}(\omega)}, \tag{3.6}$$

where $\frac{\sigma_s}{\sigma_s - 1}$ is a constant mark-up, $c_{i,s}$ is the unit production cost, and $z_{i,s}(\omega)$ is the core productivity of firm ω. Because firms with the same productivity make identical decisions, we use productivity to represent firms from now on.

The export price can also be expressed as:

$$p_{ix,s}(\omega) = \frac{\sigma_s}{\sigma_s - 1}\frac{\tau_{ix}c_{i,s}}{\underbrace{z_{i,s}(\omega)}_{\text{Core productivity}}\underbrace{\left(\Gamma(1-1/\bar{\beta})H_{i,s}^{1/\bar{\beta}}\right)^{T_{i,s}(\omega)}}_{\text{Information Spillover}}}. \tag{3.7}$$

The export price is governed both by the core productivity and the spillover from the online platform. The individual firm obtains the core productivity ("blueprint") and endogenously chooses whether to get access to knowledge spillover. The level of knowledge spillover is determined by the stock of information, which is not chosen by the firm itself. The spillover term corresponds to the external economies of scale (Bartelme et al., 2025).

Similar to Melitz (2003), we assume there is a series of fixed market-specific entry costs $c_{i,s}f_{ij,s}$. A domestic firm with productivity z in location i-sector s faces the following maximization problem:

$$\max_{D_{ix,s}(z),\{D_{ij,s}(z)\}_{j=1}^{N},T_{i,s}(z)} \quad \sum_{j=1}^{N} D_{ij,s}(z)\left[\frac{\tilde{\sigma}_s}{\sigma_s}\left(\frac{c_{i,s}}{z}\tau_{ij,s}\right)^{1-\sigma_s} P_{j,s}^{\sigma_s-1}E_{j,s} - c_{i,s}f_{ij,s}\right]$$
$$+ D_{ix,s}(z)\left[\frac{\tilde{\sigma}_s}{\sigma_s}\left(\frac{c_{i,s}}{z}\frac{\tau_{ix,s}}{\eta_{i,s}^{T_{i,s}(z)}}\right)^{1-\sigma_s} P_{x,s}^{\sigma_s-1}E_{x,s} - c_{i,s}f_{ix,s}\right] - T_{i,s}(z)c_{i,s}f_{i,s}^{L}$$

where $\tilde{\sigma}_s \equiv (\frac{\sigma_s}{\sigma_s-1})^{1-\sigma}$; $D_{ix,s}(z)$ is the binary decision of foreign market entry; $\{D_{ij,s}(z)\}_{j=1}^{N}$ is a set of binary decisions of domestic market entry; and $T_{i,s}(z)$ is the binary decision of learning. Firms only enter a new market or learn information when it is profitable to do so. Thus, we can solve for a productivity cut-off which is the minimum productivity with which a firm would enter or learn. Assume the productivity cut-off for learning is larger than the productivity cut-off of exporting. Thus, productivity cut-off for entering a domestic location j and the foreign market x are respectively:

$$\bar{z}_{ij,s} = \left(\frac{\sigma_s}{\tilde{\sigma}_s}\frac{c_{i,s}f_{ij,s}}{E_{j,s}}\right)^{\frac{1}{\sigma_s-1}} \frac{c_{i,s}\tau_{ij,s}}{P_{j,s}}, \quad \bar{z}_{ix,s} = \left(\frac{\sigma_s}{\tilde{\sigma}_s}\frac{c_{i,s}f_{ix,s}}{E_{x,s}}\right)^{\frac{1}{\sigma_s-1}} \frac{c_{i,s}\tau_{ix,s}}{P_{x,s}}. \tag{3.8}$$

The marginal learners should be indifferent between the export profit with and without learning:

$$\frac{\tilde{\sigma}_s}{\sigma_s}(\frac{c_{i,s}}{z}\tau_{ix,s})^{1-\sigma_s}P_{x,s}^{\sigma_s-1}E_{x,s} = \frac{\tilde{\sigma}_s}{\sigma_s}(\frac{c_{i,s}}{z}\frac{\tau_{ix,s}}{\eta_{i,s}})^{1-\sigma_s}P_{x,s}^{\sigma_s-1}E_{x,s} - c_{i,s}f_{i,s}^{L}$$

which gives us the learner cut-off:

$$\bar{z}_{i,s}^{L} = \left(\frac{\sigma_s}{\tilde{\sigma}_s}\frac{c_{i,s}f_{i,s}^{L}}{E_{x,s}}\frac{1}{\eta_{i,s}^{\sigma_s-1}-1}\right)^{\frac{1}{\sigma_s-1}} \frac{c_{i,s}\tau_{ix,s}}{P_{x,s}}$$

The share of learners among exporters in each location-sector can be expressed as

$$\bar{S}_{i,s}^{L} = \left(\frac{\bar{z}_{i,s}^{L}}{\bar{z}_{i,s}^{x}}\right)^{-\theta} = \left(\frac{f_{i,s}^{L}}{f_{ix,s}}\frac{1}{\eta_{i,s}^{\sigma_s-1}-1}\right)^{-\frac{\theta}{\sigma_s-1}}. \tag{3.9}$$

Aggregation. When the destination is a domestic location ($j = 1, ..., N$), the origin-destination-specific price index is

$$P_{ij,s} = \left(\frac{M_{i,s}\theta\tilde{\sigma}_s(z_{i,s}^{min})^{\frac{\theta}{\sigma_s-1}}}{\theta-(\sigma_s-1)}(c_{i,s}\tau_{ij,s})^{1-\sigma_s}(\bar{z}_{ij,s})^{\sigma_s-\theta-1}\right)^{\frac{1}{1-\sigma_s}}$$

When the destination is the foreign location ($j = x$), the origin-destination-specific price index is

$$P_{ix,s} = \left(\frac{M_{is} \theta \tilde{\sigma}_s (z_{i,s}^{min})^{\frac{\theta}{\sigma_s - 1}}}{\theta - (\sigma_s - 1)} (c_{i,s} \tau_{ix,s})^{1-\sigma_s} (\tilde{z}_{ix,s})^{\sigma_s - \theta - 1} \right)^{\frac{1}{1-\sigma_s}}$$

where

$$(\tilde{z}_{ix,s})^{\sigma_s - \theta - 1} \equiv (\eta_{i,s}^{\sigma_s - 1} - 1)(\bar{z}_{i,s}^L)^{-(\theta - \sigma_s + 1)} + (\bar{z}_{ix,s})^{-(\theta - \sigma_s + 1)}$$

is a weighted average of the learner cut-off and foreign market entry cut-off. By plugging in the productivity cut-offs, we have a bilateral price index that generalizes the one in Costinot and Rodríguez-Clare (2014):

$$P_{ix,s} = \tau_{ix,s} c_{i,s} \left(\left(\frac{E_{x,s}}{c_{i,s}} \right)^{\frac{1}{1-\sigma_s}} \frac{\tau_{ix,s} c_{i,s}}{P_{x,s}} \right)^{\delta_s} \left(\frac{R_{i,s}}{c_{i,s}} \right)^{\frac{1}{1-\sigma_s}} \tilde{\xi}_{ix,s} \tag{3.10}$$

where $\delta_s \equiv \frac{\theta}{\sigma_s - 1} - 1$ and $\tilde{\xi}_{ix,s} \equiv (\frac{\tilde{\sigma}_s}{\delta_s \sigma_s})^{\frac{1}{1-\sigma_s}} \times (\frac{\sigma_s}{\tilde{\sigma}_s} \tilde{f}_{ix,s})^{\frac{\delta_s}{\sigma_s - 1}} \times (f_{i,s}^e)^{\frac{1}{\delta_s - 1}}$.

The endogenous foreign market entry cost can be expressed as

$$\tilde{f}_{ix,s} \equiv \left[\left(\eta_{i,s}^{\sigma_s - 1} - 1 \right)^{\delta_s + 1} \left(f_{i,s}^L \right)^{-\delta_s} + (f_{ix,s})^{-\delta_s} \right]^{-\frac{1}{\delta_s}} \tag{3.11}$$

which is the weighted average of the fixed learning cost $f_{i,s}^L$ and foreign market entry cost $f_{ix,s}$. In Equation 3.11, the expected idea draw $\eta_{i,s}$ is a function of the knowledge flow as in Equation 3.5 and thus is endogenously determined. Our model is a generalization of the standard monopolistic competition trade model with heterogeneous firms (Melitz, 2003). When the fixed learning cost $f_{i,s}^L$ is infinitely large, the endogenous foreign market entry cost $\tilde{f}_{ix,s}$ degenerates into the exogenous foreign market entry fixed cost $f_{ix,s}$ and the price index collapses into the standard price index as in Costinot and Rodríguez-Clare (2014). The fixed cost of export is an endogenous variable in general equilibrium because firms endogenously make the learning decision and innovators endogenously generate information stocks. Meanwhile, the bilateral price index between any origin location i and any domestic location is standard:

$$P_{ij,s} = \tau_{ij,s} c_{i,s} \left(\left(\frac{E_{j,s}}{c_{i,s}} \right)^{\frac{1}{1-\sigma_s}} \frac{\tau_{ij,s} c_{i,s}}{P_{j,s}} \right)^{\delta_s} \left(\frac{R_{i,s}}{c_{i,s}} \right)^{\frac{1}{1-\sigma_s}} \xi_{ij,s} \tag{3.12}$$

where $\xi_{ij,s} \equiv (\frac{\tilde{\sigma}_s}{\delta_s \sigma_s})^{\frac{1}{1-\sigma_s}} (\frac{\sigma_s}{\tilde{\sigma}_s} f_{ij,s})^{\frac{\delta_s}{\sigma_s - 1}} (f_{i,s}^e)^{\frac{1}{\delta_s - 1}}$.

Bilateral trade flows (domestic and international) include trade in intermediate goods and final goods, satisfying

$$X_{ij,s} = (P_{ij,s}/P_{j,s})^{1-\sigma_s} E_{j,s}. \tag{3.13}$$

3.2 Innovators' Problem

We model the supply side of knowledge as a public good provision problem: agents contribute to public information stocks which benefit them and other innovators depending on the input-output linkages. Each location sector has one innovator.[16] The innovator provides blueprints to local intermediate producers and information flows to any location sector's information stock. The information flows are homogeneous. Innovators in all locations and sectors are oligopolies in providing information flows. They play a non-cooperative Nash game by choosing the optimal level of information provision taking other innovators' decisions as given. To gain tractability, we assume the innovators take the multilateral price indexes and wages as given. The innovator problem is similar to the semi-primal approach in Bai, Jin and Lu (2023).

Because the innovator sells "blueprints" of varieties to local firms, the innovator collects all the firm profits within the location-sector. The innovator generates information flows to any domestic location-sector to maximize her profit:

$$\max_{\{H_{is,dk}\}_{d\in N, s\in S}} \Pi_{i,s} - c_{i,s} f^H_{i,s} \sum_d \sum_s \frac{(H_{is,dk})^{\gamma^H}}{\gamma^H} - c_{i,s} f^e_{i,s} M_{i,s} \tag{3.14}$$

where $\Pi_{i,s}$ is the aggregate profit of firms in location i-sector s; $f^H_{i,s}$ is the cost shifter of producing information; γ^H is the cost curvature of producing information; $f^e_{i,s}$ is the service needed to produce one "blue-print" in location i-sector s, same as the fixed market entry cost that firms pay to the innovator to purchase a "blue-print"; $c_{i,s}$ is the unit production cost; $M_{i,s}$ is the number of blueprints which is also the firm number.

The optimal information flows can be expressed as:

$$H_{is,dk} = \left(\frac{\theta \Gamma \bar{\beta}}{\sigma_s} \frac{1}{c^H_{i,s} f^H_{i,s}} \frac{X_{ij,s}}{E_j} X_{jx,k} Z_{j,k} \right)^{\frac{1}{\gamma^H - 1}}. \tag{3.15}$$

where $Z_{j,k} \equiv \left(\frac{\tilde{f}_{jx}}{f^L_j}\right)^{\delta_k} (\eta^{\sigma_k - 1}_{j,k} - 1)^{\delta_k} \eta^{\sigma_k - 2}_{j,k} H^{\bar{\beta}-1}_{j,k} P^{1-\sigma_k}_{x,k}$. The derivation is in Appendix 6.1. Information flows are positively correlated with both domestic trade flows between the teacher and learner and the learner's exports. It means that if the learner location-sector jk exports more to foreign markets, the teacher location-sector is is willing to share more information. Also, information flows

[16] One can think of the innovator as trade intermediaries or the export expert within firms. Trade intermediaries play a key role in reducing information friction. Rhee (1989) study the trade intermediaries.

increase with trade flows from the teacher location-sector is to the learner location-sector jk. The theoretical prediction coincides with the "learning from sellers" story in the literature with a new micro-foundation: instead of describing information flows as a passive spillover, we model information flow as the outcome of active and rational decisions of learner and teacher. General equilibrium is in Appendix 6.2.

4. Connecting Empirical Patterns and Theory

In this section, we bridge the reduced-form empirical patterns with the theory. We show the tight connections between the empirical results with their structural counterparts.

Connection 1: The Impact of Information Stock on Firm-level Export Volume

From the model, we can express the firm-level export volume as:

$$x_{ix,s}(z) = \left(\frac{\sigma_s}{\sigma_s - 1} \frac{c_{i,s}\tau_{ix,s}}{z\eta_{i,s}^{T_{i,s}(z)}} \right)^{1-\sigma_s} P_{x,s}^{\sigma_s - 1} E_{x,s} \tag{4.1}$$

Taking log on both sides , we have

$$\begin{aligned}\ln x_{ix,s}(z) = \ln \left(\frac{\sigma_s}{\sigma_s - 1} \right)^{1-\sigma_s} &+ (1-\sigma_s)\ln(c_{i,s}\tau_{ix,s}) + \ln\left(P_{x,s}^{\sigma_s-1}E_{x,s}\right) \\ &+ (\sigma_s - 1)\left(\ln\Gamma\left(1 - \frac{1}{\bar{\beta}}\right) + \frac{1}{\bar{\beta}}\ln H_{i,s}\right) T_{i,s}(z) + (\sigma_s - 1)\ln z.\end{aligned}$$

where we used $\eta_{i,s} = \Gamma\left(1 - \frac{1}{\bar{\beta}}\right) H_{i,s}^{1/\bar{\beta}}$. The reduced form specification is:

$$\ln x_{zis,t} = \alpha + \beta_1 \ln H_{is,t} \times T_{zis,t} + \beta_2 \times T_{zis,t} + \zeta_{is,t} + \delta_{zis} + \varepsilon_{zis,t}$$

where $x_{zis,t}$ is the export volume of firm z from location i-sector s in time t; α is a constant term; $H_{is,t}$ is the information stock measured by total inward posting number in location i-sector s in time t; $T_{zis,t}$ is a binary variable indicating if firm z joins the platform in t; we control for location-sector-time and firm fixed effects; $\varepsilon_{zis,t}$ is an error term. The results are shown in Table 4. Column 2 shows the main regression result. A 10% increase in information stock is associated with a 0.3% increase in the adoption firm's export volume compared with a non-adoption firm. The results are consistent with those in Column 2. The structural interpretation of the interaction term is $\beta_1 = (\sigma_s - 1)/\bar{\beta} = 0.031$. Calibrating $\sigma_s = 5$ yields $\bar{\beta} = 129$, implying a relatively thin-tailed idea distribution in which additional information generates stable but incremental improvements in export performance.

Connection 2: The Gravity of Trade and Information Flows

The model derives the gravity equations of trade:

$$X_{ij,s} = \left(\tau_{ij,s} c_{i,s} \left(\left(\frac{E_{j,s}}{c_{i,s}} \right)^{\frac{1}{1-\sigma_s}} \frac{\tau_{ij,s} c_{i,s}}{P_{j,s}} \right)^{\delta_s} \left(\frac{R_{i,s}}{c_{i,s}} \right)^{\frac{1}{1-\sigma_s}} \xi_{ij,s} \right)^{1-\sigma_s} (P_{j,s})^{\sigma_s - 1} E_{j,s}, \tag{4.2}$$

where $\xi_{ij,s} \equiv (\frac{\tilde{\sigma}_s}{\delta_s \sigma_s})^{\frac{1}{1-\sigma_s}} \times (\frac{\sigma_s}{\tilde{\sigma}_s} f_{ij,s})^{\frac{\delta_s}{\sigma_s - 1}} \times (f^e_{i,s})^{\frac{1}{\delta_s - 1}}$, and knowledge flows:

$$H_{is,dk} = \left(\frac{\theta \Gamma \bar{\beta}}{\sigma_s} \frac{1}{c^H_{i,s} f^H_{i,s}} \frac{X_{ij,s}}{E_j} X_{jx,k} Z_{j,k} \right)^{\frac{1}{\gamma^H - 1}}. \tag{4.3}$$

where $Z_{j,k} \equiv \left(\frac{\tilde{f}_{jx}}{f^L_j} \right)^{\delta_k} (\eta_{j,k}^{\sigma_k - 1} - 1)^{\delta_k} \eta_{j,k}^{\sigma_k - 2} H_{j,k}^{\bar{\beta}-1} P_{x,k}^{1-\sigma_k}$.

Notice that the only bilateral term in the knowledge gravity equation is the bilateral trade flow $(X_{ij,s})^{\gamma^H - 1}$. This explains our empirical pattern that knowledge flow decreases with distance. In the empirical section, we estimated the knowledge gravity equation using city-sector-pair level data. However, trade data at this level is limited to us. Thus, we use province-sector-pair level data for both knowledge and trade to estimate the gravity equations. Using the estimation results in Table 6, the ratio of the distance elasticity of knowledge and trade identifies $\frac{1}{\gamma^H - 1} = 0.338/0.672 = 0.5$ and $\gamma^H = 3$.

Table 6: Gravity Estimation of Information and Trade Flows at Province-sector-pair Level

	(1) Information	(2) Trade
Log Distance	-0.338***	-0.672***
	(0.00785)	(0.00164)
Observations	73,902	3,335,724
Source-Time FE	Yes	Yes
Destination-Time FE	Yes	Yes
Sector-Pair-Time FE	Yes	Yes
Clustered SE (Pair)	Yes	Yes

Notes: This table presents the results of a gravity regression estimating the relationship between information flows and trade flows at the province-pair-sector level. Column (1) reports the results for information flows, while Column (2) reports the results for trade flows. The key explanatory variable is the logarithm of geographic distance between provinces. All regressions include source-time, destination-time, and sector-pair-time fixed effects. Standard errors are clustered at the province-pair level and reported in parentheses. ***$p < 0.01$, **$p < 0.05$, *$p < 0.1$.

Connection 3: The Platform Adoption Share and Region Characteristics

The share of learners among exporters in each location-sector can be expressed as

$$\bar{S}^L_{i,s} = \left(\frac{\bar{z}^L_{i,s}}{\bar{z}^x_{i,s}}\right)^{-\theta} = \left(\frac{f^L_{i,s}}{f_{ix,s}} \frac{1}{\eta_{i,s}^{\sigma_s-1} - 1}\right)^{-\frac{\theta}{\sigma_s-1}}. \tag{4.4}$$

The learner share corresponds to the fourth empirical pattern: there is a large variation of social media adoption across regions. The learner share is higher when the fixed learning cost $f^L_{i,s}$ is lower and when the expected idea draw $\eta_{i,s}$ is high. Importantly, the learner share is governed by the relative size of learning cost compared with foreign market entry cost $f_{ix,s}$. This explains why firms in richer cities are more likely to adopt social media and learn: even if their foreign market entry cost $f_{ix,s}$ is low, they face even lower learning cost. Also, they could have higher expected idea draw from the information stock. Thus, as will be seen soon, we can infer the relative size $\frac{f^L_{i,s}}{f_{ix,s}}$ using the observed learner share in the data $\bar{S}^L_{i,s}$ and the estimated average idea draw $\eta_{i,s}$.

5. Conclusions

This paper studies information diffusion among firms on social media platforms and its role in reducing information frictions in international trade. The analysis does not examine, promote, or lend analytical support to regulatory evasion or efforts to circumvent trade barriers or policy frameworks. The empirical and theoretical findings are descriptive and analytical in nature, and should not be interpreted as an endorsement of any particular platform governance model, trade practice, or policy position.

The invention of social media has profoundly transformed the ways in which people share information and acquire knowledge. Unlike traditional information sources, such as television and newspapers, social media fosters an environment that facilitates decentralized information production and consumption. In this paper, we explore how a professional social media platform aids exporting firms in learning and sharing information related to international trade. We have gathered unique data, documented novel empirical patterns, and expanded the conventional trade model.

References

Acemoglu, Daron, and Pablo D Azar. 2020. "Endogenous production networks." Econometrica, 88(1): 33–82.

Aghion, Philippe, Ufuk Akcigit, and Peter Howitt. 2014. "What do we learn from Schumpeterian growth theory?" In Handbook of economic growth. Vol. 2, 515–563. Elsevier.

Ahn, JaeBin, Chan Kim, Nan Li, and Andrea Manera. 2026. "Knowledge Diffusion Through FDI: Worldwide Firm-Level Evidence." Journal of International Economics, 160: 104205.

Akcigit, Ufuk, and William R Kerr. 2018. "Growth through heterogeneous innovations." Journal of Political Economy, 126(4): 1374–1443.

Alvarez, Fernando E, Francisco J Buera, Robert E Lucas, et al. 2013. "Idea flows, economic growth, and trade." National Bureau of Economic Research.

Anderson, James E, and Eric Van Wincoop. 2004. "Trade costs." Journal of Economic literature, 42(3): 691–751.

Antenucci, Dolan, Michael Cafarella, Margaret Levenstein, Christopher Ré, and Matthew D Shapiro. 2014. "Using social media to measure labor market flows." National Bureau of Economic Research.

Antràs, Pol, Davin Chor, Thibault Fally, and Russell Hillberry. 2012. "Measuring the upstreamness of production and trade flows." American Economic Review, 102(3): 412–416.

Antweiler, Werner, and Murray Z Frank. 2004. "Is all that talk just noise? The information content of internet stock message boards." The Journal of finance, 59(3): 1259–1294.

Arkolakis, Costas, Federico Huneeus, and Yuhei Miyauchi. 2023. "Spatial production networks." National Bureau of Economic Research.

Atalay, Enghin, Sebastian Sotelo, and Daniel I Tannenbaum. 2022. "The geography of job tasks." National Bureau of Economic Research.

Athey, Susan, and Guido W Imbens. 2022. "Design-based analysis in difference-in-differences settings with staggered adoption." Journal of Econometrics, 226(1): 62–79.

Bai, Yan, Keyu Jin, and Dan Lu. 2023. "Technological rivalry and optimal dynamic policy in an open economy." National Bureau of Economic Research.

Baker, Andrew C, David F Larcker, and Charles CY Wang. 2022. "How much should we trust staggered difference-in-differences estimates?" Journal of Financial Economics, 144(2): 370–395.

Bartelme, Dominick, Arnaud Costinot, Dave J Donaldson, and Andrés Rodríguez-Clare. 2025. "The textbook case for industrial policy: Theory meets data." Journal of Political Economy, 133(5): 1527–1573.

Bernard, Andrew B, Andreas Moxnes, and Yukiko U Saito. 2019. "Production networks, geography, and firm performance." Journal of Political Economy, 127(2): 639–688.

Bloom, Nicholas, John Van Reenen, and Heidi Williams. 2019. "A toolkit of policies to promote innovation." Journal of economic perspectives, 33(3): 163–184.

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

Buera, Francisco J, and Robert E Lucas Jr. 2018. "Idea flows and economic growth." Annual Review of Economics, 10: 315–345.

Bui, Ha, Zhen Huo, Andrei A Levchenko, and Nitya Pandalai-Nayar. 2022. "Noisy global value chains." National Bureau of Economic Research.

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

Caballero, Ricardo J, and Adam B Jaffe. 1993. "How high are the giants' shoulders: An empirical assessment of knowledge spillovers and creative destruction in a model of economic growth." NBER macroeconomics annual, 8: 15–74.

Cai, Jie, and Nan Li. 2019. "Growth through inter-sectoral knowledge linkages." The Review of Economic Studies, 86(5): 1827–1866.

Callaway, Brantly, and Pedro HC Sant'Anna. 2021. "Difference-in-differences with multiple time periods." Journal of econometrics, 225(2): 200–230.

Carballo, Jerónimo, Marisol Rodriguez Chatruc, Catalina Salas Santa, and Christian Volpe Martincus. 2022. "Online business platforms and international trade." Journal of International Economics, 137: 103599.

Carlino, Gerald, and William R Kerr. 2015. "Agglomeration and innovation." In Handbook of Regional and Urban Economics. Vol. 5, 349–404. Elsevier.

Carvalho, Vasco M. 2014. "From micro to macro via production networks." Journal of Economic Perspectives, 28(4): 23–48.

Carvalho, Vasco M, and Alireza Tahbaz-Salehi. 2019. "Production networks: A primer." Annual Review of Economics, 11(1): 635–663.

Chaney, Thomas. 2008. "Distorted gravity: the intensive and extensive margins of international trade." American Economic Review, 98(4): 1707–21.

Chen, Yan, Rosta Farzan, Robert Kraut, Iman Yeckehzaare, and Ark Fangzhou Zhang. 2023. "Motivating experts to contribute to digital public goods: A personalized field experiment on Wikipedia." Management Science.

Choi, Jaedo. 2024. "From adoption to innovation: State-dependent technology policy in developing countries." IMF Working Papers, 2024(154): 1.

Choi, Jaedo, and Younghun Shim. 2022. "Technology adoption and late industrialization." Working Paper.

Choi, Jaedo, and Younghun Shim. 2024. "Industrialization and the Big Push: Theory and Evidence from South Korea." International Monetary Fund.

Costinot, Arnaud, and Andrés Rodríguez-Clare. 2014. "Trade theory with numbers: Quantifying the consequences of globalization." In Handbook of international economics. Vol. 4, 197–261. Elsevier.

Cotterlaz, Pierre, and Arthur Guillouzouic. 2025. "The percolation of knowledge across space." Journal of International Economics, 153: 104026.

Cristea, Anca D. 2011. "Buyer-seller relationships in international trade: Evidence from US States' exports and business-class travel." Journal of international economics, 84(2): 207–220.

Dickstein, Michael J, and Eduardo Morales. 2018. "What do Exporters Know?*." The Quarterly Journal of Economics, 133(4): 1753–1801.

Eaton, Jonathan, and Samuel Kortum. 2002. "Technology, geography, and trade." Econometrica, 70(5): 1741–1779.

Ellison, Glenn, Edward L Glaeser, and William R Kerr. 2010. "What causes industry agglomeration? Evidence from coagglomeration patterns." American Economic Review, 100(3): 1195–1213.

Fadeev, Evgenii. 2024. "Creative Construction: Knowledge Sharing and Cooperation Between Firms." Duke University. Working paper, June 2024.

Goodman-Bacon, Andrew. 2021. "Difference-in-differences with variation in treatment timing." Journal of econometrics, 225(2): 254–277.

Hausmann, Ricardo, and Dani Rodrik. 2003. "Economic development as self-discovery." Journal of development Economics, 72(2): 603–633.

Hayek, FA. 1945. "The Use of Knowledge in Society." The American Economic Review, 35(4): 519–530.

Hinnosaar, Marit, Toomas Hinnosaar, Michael Kummer, and Olga Slivko. 2023. "Wikipedia matters." Journal of Economics & Management Strategy, 32(3): 657–669.

Imai, Kosuke, In Song Kim, and Erik H Wang. 2023. "Matching methods for causal inference with time-series cross-sectional data." American Journal of Political Science, 67(3): 587–605.

Jaffe, Adam B, Manuel Trajtenberg, and Rebecca Henderson. 1993. "Geographic localization of knowledge spillovers as evidenced by patent citations." the Quarterly journal of Economics, 108(3): 577–598.

Jiao, Yang, Zhikuo Liu, Zhiwei Tian, and Xiaxin Wang. 2024. "The impacts of the US trade war on Chinese exporters." Review of Economics and Statistics, 106(6): 1576–1587.

Jones, Charles I. 2005. "Growth and ideas." In Handbook of economic growth. Vol. 1, 1063–1111. Elsevier.

Krugman, Paul. 1991. Geography and Trade. Cambridge, MA:MIT Press.

Lashkaripour, Ahmad, and Volodymyr Lugovskyy. 2023. "Profits, scale economies, and the gains from trade and industrial policy." American Economic Review, 113(10): 2759–2808.

Liu, Ernest. 2019. "Industrial policies in production networks." The Quarterly Journal of Economics, 134(4): 1883–1948.

Liu, Ernest, and Song Ma. 2021. "Innovation networks and r&d allocation." National Bureau of Economic Research.

Liu, Licheng, Ye Wang, and Yiqing Xu. 2024. "A practical guide to counterfactual estimators for causal inference with time-series cross-sectional data." American Journal of Political Science, 68(1): 160–176.

Lucas Jr, Robert E, and Benjamin Moll. 2014. "Knowledge growth and the allocation of time." Journal of Political Economy, 122(1): 1–51.

Marshall, Alfred. 2009. Principles of economics: unabridged eighth edition. Cosimo, Inc.

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

Melitz, Marc J, and Stephen J Redding. 2021. "Trade and innovation." National Bureau of Economic Research 28945.

Mobius, Markus, and Tanya Rosenblat. 2014. "Social learning in economics." Annu. Rev. Econ., 6(1): 827–847.

Munch, Jakob, and Georg Schaur. 2018. "The effect of export promotion on firm-level performance." American Economic Journal: Economic Policy, 10(1): 357–387.

Perla, Jesse, and Christopher Tonetti. 2014. "Equilibrium imitation and growth." Journal of Political Economy, 122(1): 52–76.

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

Portes, Richard, and Helene Rey. 2005. "The determinants of cross-border equity flows." Journal of international Economics, 65(2): 269–296.

Rauch, James E. 2001. "Business and social networks in international trade." Journal of economic literature, 39(4): 1177–1203.

Rauch, James E, and Joel Watson. 2002. "Entrepreneurship in international trade." National Bureau of Economic Research 8708.

Rhee, Yung W. 1989. Small trading companies and a successful export response: lessons from Hong Kong. World Bank Industry and Energy Department, Policy, Planning and Research.

Rosenblat, Tanya S, and Markus M Mobius. 2004. "Getting closer or drifting apart?" The Quarterly Journal of Economics, 119(3): 971–1009.

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

Saxenian, AnnaLee. 2000. "Silicon Valley's new immigrant entrepreneurs."

Startz, Meredith. 2021. "The value of face-to-face: Search and contracting problems in Nigerian trade." Working Paper.

Stuart, Elizabeth A, and Donald B Rubin. 2008. "Best Practices in Quasi-Experimental Designs: Matching Methods for Causal Inference." In Best Practices in Quantitative Methods. , ed. Jason W Osborne, 155–176. Sage Publications.

Sun, Liyang, and Sarah Abraham. 2021. "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects." Journal of econometrics, 225(2): 175–199.

Theodosiou, Marios, and Evangelia Katsikea. 2013. "The export information system: An empirical investigation of its antecedents and performance outcomes." Journal of International Marketing, 21(3): 72–94.

Thompson, Neil, and Douglas Hanley. 2018. "Science is shaped by Wikipedia: evidence from a randomized control trial."

Tian, Lin, and Yue Yu. 2023. "Geographic Integration and Firm Exports Evidence from China." Working Paper.

Volpe Martincus, Christian, and Jerónimo Carballo. 2012. "Export promotion activities in developing countries: What kind of trade do they promote?" The Journal of International Trade & Economic Development, 21(4): 539–578.

Wang, Zhihan Helen, Jun Li, and Di Andrew Wu. 2023. "Mind the Gap: Gender Disparity in Online Learning Platform Interactions." Working Paper.

Wei, Shang-Jin, Ziru Wei, and Jianhuan Xu. 2021. "On the market failure of "missing pioneers"." Journal of Development Economics, 152: 102705.

Yeaple, Stephen Ross. 2005. "A simple model of firm heterogeneity, international trade, and wages." Journal of international Economics, 65(1): 1–20.

Appendix

5.1 TSCS Matching

The details of the matching procedures are as follows. For each treated observation, we first find a set of control observations that have the identical treatment history up to the pre-specified number of periods. We call this group of matched control observations a matched set. Once a matched set is selected for each treated observation, we further refine it by adjusting for observed confounding via standard matching and weighting techniques, such as Mahalanobis and Propensity Score (PS) Weighting, so that the treated and matched control observations have similar covariate values. Finally, we apply the difference-in-differences estimator in order to account for an underlying time trend.

Here we only briefly introduce the panel matching method. Suppose there are N individuals and T periods in the balanced panel (this method is also applicable to the unbalanced panel). Let i and t denote each individual and each period, where $i = 1, 2, 3, ...N$, $t = 1, 2, 3..., T$.

Let Y_{it} denote the dependent variable, and TR_{it} denote the dummy variable for whether individual i receives the shock at period t. Z_{it} represents K other matching variables. Let F denote F periods after the shock and L denote the L period before the shock. To find the control group, they use a matching method to adjust for covariates up to L periods before the shock.

Then, the Average Treated Effect (ATT) is expressed as:

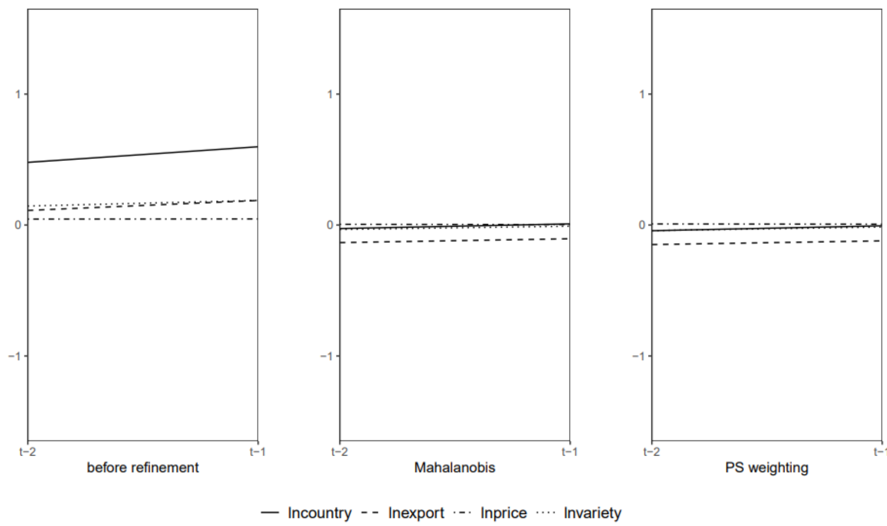
$$\delta(F, L) = \mathbb{E}\{Y_{i,t+F}(TR_{it} = 1, TR_{i,t-1} = 0, \{TR_{i,t-\ell}\}_{\ell=2}^{L}) -$$
$$Y_{i,t+F}(TR_{it} = 0, TR_{i,t-1} = 0, \{TR_{i,t-\ell}\}_{\ell=2}^{L}) | TR_{it} = 1, TR_{i,t-1} = 0\}$$

The key assumption of the DiD approach is the parallel trend assumption, that is, given the shock history, the history of the dependent variable (excluding the past period t-1) and the history of the control variable, the trend of the dependent variable after the shock and its hypothetical unaffected trend are the same, that is:

$$\mathbb{E}[Y_{i,t+F}(TR_{it} = 1, TR_{i,t-1} = 0, \{TR_{i,t-\ell}\}_{\rho=0}^{L}) - Y_{i,t-1}|$$
$$TR_{it} = 1, TR_{i,t-1} = 0, \{TR_{i,t-\ell}, Y_{i,t-\ell}\}_{\ell=2}^{L}\{\mathbf{Z}_{i,t-\ell}\}_{\ell=0}^{L}] =$$
$$\mathbb{E}[Y_{i,t+F}(TR_{it} = 0, TR_{i,t-1} = 0, \{TR_{i,t-\ell}\}_{\ell=2}^{L}) - Y_{i,t-1}|$$
$$TR_{it} = 0, TR_{i,t-1} = 0, \{TR_{i,t-\ell}, Y_{i,t-\ell}\}_{\ell=2}^{L}, \{\mathbf{Z}_{i,t-\ell}\}_{\ell=0}^{L}]$$

$V_{i,t}$ indicates the matching variable for unit i at period t. $D_{it} = X_{it}(1 - X_{i,t-1}) \cdot \mathbf{1}\{|\mathcal{M}_{it}| > 0\}$, and w_{iit} represents the non-negative normalized weight. Note that $D_{it} = 1$ only if observation (i, t) changes the treatment status from the control condition at time t- 1 to the treatment condition at time t and has at least one matched control unit.

Figure 9: Covariate Balance Test over Pre-Treatment Periods



Notes: This figure presents covariate balance over the two pre-treatment periods, $t-2$ and $t-1$. The vertical axis reports the standardized balance score, so values closer to zero indicate better balance between treated and control observations. The left panel shows substantial imbalance before refinement, especially for export volume, destination count, product scope, and average unit value. The middle and right panels show the balance after Mahalanobis matching and propensity-score weighting, respectively. Both procedures sharply reduce imbalance and support the credibility of the identifying comparison used in the main text.

Define the covariate balance for variable j at the pre-treatment time period $t-l$ as,

$$B_{it}(j,\ell) = \frac{V_{i,t-\ell,j} - \sum_{i' \in \mathcal{M}_{it}} w_{it}^{i'} V_{i',t-\ell,j}}{\sqrt{\frac{1}{N_1 - 1} \sum_{i'=1}^{N} \sum_{t'=L+1}^{T-F} D_{it'} \left(V_{i',t'-\ell,j} - \bar{V}_{t'-\ell,j}\right)^2}}$$

6. Expected Draw from the Fréchet Idea Distribution

This appendix derives the expected draw from the Fréchet distribution used in the main text. Let z^x denote the random quality of an idea drawn from the information stock. Suppose the cumulative distribution function (CDF) is

$$F(z) = \exp\{-Hz^{-\bar{\beta}}\}, \qquad z > 0,\ H > 0,\ \bar{\beta} > 1. \tag{6.1}$$

This is the standard Fréchet form, with H indexing the size of the information stock and $\bar{\beta}$ the shape parameter governing tail thickness.

Lemma 1. *For any $r < \bar{\beta}$, the r-th moment of z^x exists and is*

$$\mathbb{E}\left[(z^x)^r\right] = \Gamma\left(1 - \frac{r}{\bar{\beta}}\right) H^{r/\bar{\beta}}. \tag{6.2}$$

Proof. Differentiate $F(z)$ to obtain the density

$$f(z) = H\bar{\beta}\, z^{-(\bar{\beta}+1)} e^{-Hz^{-\bar{\beta}}}, \qquad z > 0. \tag{6.3}$$

The r-th moment is

$$\mathbb{E}[z^r] = \int_0^\infty z^r f(z)\, dz = H\bar{\beta} \int_0^\infty z^{r-\bar{\beta}-1} e^{-Hz^{-\bar{\beta}}}\, dz. \tag{6.4}$$

Make the change of variables $u = Hz^{-\bar{\beta}}$ so that $z = (H/u)^{1/\bar{\beta}}$ and $dz = -\frac{1}{\bar{\beta}} H^{1/\bar{\beta}} u^{-1/\bar{\beta}-1}\, du$. Substituting gives

$$\mathbb{E}[z^r] = H^{r/\bar{\beta}} \int_0^\infty u^{-r/\bar{\beta}} e^{-u}\, du = H^{r/\bar{\beta}}\, \Gamma\!\left(1 - \frac{r}{\bar{\beta}}\right), \tag{6.5}$$

where the last step uses the definition of the Gamma function. The condition $r < \bar{\beta}$ ensures the integral converges. □

Corollary.. Setting $r = 1$ gives the expected draw:

$$\mathbb{E}[z^x] = \Gamma\!\left(1 - \frac{1}{\bar{\beta}}\right) H^{1/\bar{\beta}}. \tag{6.6}$$

Interpretation.. The shape parameter $\bar{\beta}$ determines the dispersion of the idea distribution. Smaller $\bar{\beta}$ implies fatter tails: the distribution of ideas is more dispersed, and firms occasionally obtain "blockbuster" ideas that yield large reductions in export costs. Larger $\bar{\beta}$ implies thinner tails: idea draws are more homogeneous and the returns to expanding the information stock H are more incremental. In the reduced-form specification, the coefficient on $\ln H \times T$ maps to

$$\beta_1 = \frac{\sigma_s - 1}{\bar{\beta}}, \tag{6.7}$$

showing that the effect of information stock on exports is stronger when $\bar{\beta}$ is small (high dispersion) and weaker when $\bar{\beta}$ is large (low dispersion).

6.1 Innovator's Problem

The profit is

$$
\begin{aligned}
\Pi_{i,s} &= \frac{1}{\sigma}\left(\sum_{j=1}^{N} X_{ij,s} + X_{ix,s}\right) \\
&= \frac{1}{\sigma}\left(\sum_{j=1}^{N}(P_{ij,s}/P_{j,s})^{1-\sigma_s} e_{j,s}E_j + (P_{ix,s}/P_{x,s})^{1-\sigma_s} e_{x,s}E_x\right) \\
&= \frac{1}{\sigma}\left(\sum_{j=1}^{N}(P_{ij,s}/P_{j,s})^{1-\sigma_s} e_{j,s}R_j + (P_{ix,s}/P_{x,s})^{1-\sigma_s} e_{x,s}R_x\right) \\
&= \frac{1}{\sigma}\left(\sum_{j=1}^{N}(P_{ij,s}/P_{j,s})^{1-\sigma_s} e_{j,s} \sum_{k}\left(\sum_{j'=1}^{N} X_{jj',k} + X_{jx,k}\right) + (P_{ix,s}/P_{x,s})^{1-\sigma_s} e_{x,s}R_x\right) \\
&= \frac{1}{\sigma}\left(\sum_{j=1}^{N}(P_{ij,s}/P_{j,s})^{1-\sigma_s} e_{j,s} \sum_{k}\left(\sum_{j'=1}^{N} X_{jj',k} + (P_{jx,k}/P_{x,k})^{1-\sigma_k} e_{x,s}E_x\right) + (P_{ix,s}/P_{x,s})^{1-\sigma_s} e_{x,s}R_x\right)
\end{aligned}
$$

where we used that total expenditure equals total producer revenue $E_i = R_i$.

The partial impact of knowledge stock on export is

$$
\frac{\partial \Pi_{i,s}}{\partial \eta_{jk}} = \frac{1}{\sigma_s}\left(\frac{P_{ij,s}}{P_{j,s}}\right)^{1-\sigma_s} e_{j,s} \times (1-\sigma_k) P_{jx,k}^{-\sigma_k} e_{x,s} E_x \frac{\partial P_{jx,k}}{\partial \eta_{jk}}
$$

$$
P_{ix,s} = \tau_{ix,s} c_{i,s}\left[\left(\frac{e_{x,s}}{v_x}\frac{w_x}{c^x_{ix,s}}\right)^{\frac{\delta_s}{1-\sigma_s}} \frac{\tau_{ix,s} c_{i,s}}{P_{x,s}}\right]^{\delta_s} \left(\frac{r_{i,s}}{v_i}\frac{w_i}{c_{i,s}}\right)^{\frac{\delta_s}{1-\sigma_s}} \tilde{\xi}_{ix,s} \tag{6.8}
$$

where $\delta_s \equiv \frac{\theta}{\sigma_s - 1} - 1$ and $\tilde{\xi}_{ix,s} \equiv (\frac{\tilde{\sigma}_s}{\delta_s \sigma_s})^{\frac{1}{1-\sigma_s}} \times (\frac{\sigma_s}{\tilde{\sigma}_s}\tilde{f}_{ix})^{\frac{\delta_s}{\sigma_s - 1}} \times (f^e_i)^{\frac{1}{\delta_s - 1}}$

and

$$
\tilde{f}_{ix} \equiv \left[(\eta_{i,s}^{\sigma_s - 1} - 1)^{\frac{\theta}{\sigma_s - 1}} (f^L_i)^{-\delta_s} + (f_{ix})^{-\delta_s}\right]^{-\frac{1}{\delta_s}}
$$

$$
\frac{\partial P_{ix,s}}{\partial \tilde{f}_{ix}} = \frac{P_{ix,s}}{\tilde{\xi}_{ix,s}} \times \frac{\partial \tilde{\xi}_{ix,s}}{\partial \tilde{f}_{ix}} = (\frac{\delta_s}{\sigma_s - 1})\frac{P_{ix,s}}{\tilde{f}_{ix}}
$$

$$\begin{aligned}\frac{\partial \tilde{f}_{ix,s}}{\partial \eta_{i,s}} &= -\frac{1}{\delta_s}\left[(\eta_{i,s}^{\sigma_s-1}-1)^{\frac{\theta}{\sigma_s-1}}(f_i^L)^{-\delta_s}+(f_{ix})^{-\delta_s}\right]^{-\frac{1}{\delta_s}-1}(f_i^L)^{-\delta_s}\frac{\theta}{\sigma_s-1}(\eta_{i,s}^{\sigma_s-1}-1)^{\frac{\theta}{\sigma_s-1}-1}\times(\sigma_s-1)\eta_{i,s}^{\sigma_s-2}\\ &= -\frac{\theta}{\delta_s}\tilde{f}_{ix}^{1+\delta_s}(f_i^L)^{-\delta_s}(\eta_{i,s}^{\sigma_s-1}-1)^{\frac{\theta}{\sigma_s-1}-1}\times\eta_{i,s}^{\sigma_s-2}\\ &= -\frac{\theta}{\delta_s}\tilde{f}_{ix}^{1+\delta_s}(f_i^L)^{-\delta_s}(\eta_{i,s}^{\bar{\sigma}_{s,1}}-\eta_{i,s}^{\bar{\sigma}_{s,2}})^{\frac{\theta}{\sigma_s-1}-1}\end{aligned}$$

where $\bar{\sigma}_{s,1} \equiv (\sigma-1)+\frac{(\sigma-2)(\theta-\sigma+1)}{\sigma-1}$ and $\bar{\sigma}_{s,2} \equiv \frac{(\sigma-2)(\theta-\sigma+1)}{\sigma-1}$.

$$\eta_{i,s} \equiv \Gamma \times H_{i,s}^{\bar{\beta}} = \Gamma \times (\sum_{k\in\mathcal{S}}\sum_{u\in\mathcal{N}} H_{uk,is})^{\bar{\beta}}$$

$$\frac{\partial \eta_{i,s}}{\partial H_{uk,is}} = \Gamma\bar{\beta}H_{i,s}^{\bar{\beta}-1}$$

In the end

$$\begin{aligned}\frac{\partial \Pi_{i,s}}{\partial H_{is,jk}} =& \frac{1}{\sigma_s}\left(\frac{P_{ij,s}}{P_{j,s}}\right)^{1-\sigma_s} e_{j,s}\times(1-\sigma_k)P_{jx,k}^{-\sigma_k}e_{x,s}E_x\times(\frac{\delta_k}{\sigma_k-1})\frac{P_{jx,k}}{\tilde{f}_{jx}}\\ &\times(-\frac{\theta}{\delta_k})\tilde{f}_{jx}^{1+\delta_k}(f_j^L)^{-\delta_k}(\eta_{j,k}^{\sigma_k-1}-1)^{\frac{\theta}{\sigma_k-1}-1}\times\eta_{j,k}^{\sigma_k-2}\times\Gamma\bar{\beta}H_{j,s}^{\bar{\beta}-1}\end{aligned}$$

$$\frac{\partial \Pi_{i,s}}{\partial H_{is,jk}} = \frac{\theta\Gamma\bar{\beta}}{\sigma_s}X_{ij,s}Z_{j,k}$$

where we used $X_{ij,s} = (P_{ij,s}/P_{j,s})^{1-\sigma_s}e_{j,s}E_j$ and we define the destination-specific term as

$$Z_{j,k} \equiv \frac{E_x}{E_j}P_{jx,k}^{1-\sigma_k}e_{x,s}\left(\frac{\tilde{f}_{jx}}{f_j^L}\right)^{\delta_k}(\eta_{j,k}^{\sigma_k-1}-1)^{\delta_k}\eta_{j,k}^{\sigma_k-2}H_{j,k}^{\bar{\beta}-1}.$$

The first order condition of knowledge posting flow is

$$\frac{\partial \Pi_{i,s}}{\partial H_{is,jk}} = c_{i,s}^H f_{i,s}^H (H_{is,dk})^{\gamma^H-1}.$$

Plug into the left-hand side, we have

$$\frac{\theta\Gamma\bar{\beta}}{\sigma_s}X_{ij,s}Z_{j,k} = c_{i,s}^H f_{i,s}^H (H_{is,dk})^{\gamma^H-1}.$$

The optimal knowledge posting flow is

$$H_{is,dk} = \left(\frac{\theta \Gamma \bar{\beta}}{\sigma_s} \frac{X_{ij,s} Z_{j,k}}{c^H_{i,s} f^H_{i,s}} \right)^{\frac{1}{\gamma^H - 1}}.$$

6.2 Equilibrium

We denote the net trade deficit of location j as D_j. Assume the trade deficit is a constant share of either local GDP ($\mu = 1$) or global GDP ($\mu = 0$):

$$D_j = \Xi_j Y_j^{\mu} \tag{6.9}$$

The multilateral price index $P_{j,s}^{1-\sigma_s} = \sum_i P_{ij,s}^{1-\sigma_s}$ can be expressed as

$$P_{j,s} = \left(\sum_i ((1 + t_{ij,s}) \tau_{ij,s} c_{i,s})^{(1-\sigma_s)(1+\delta_s)} \left(\frac{E_{j,s}}{c_{i,s}} \right)^{\delta_s} \left(\frac{R_{i,s}}{c_{i,s}} (\tilde{\xi}_{ij,s})^{1-\sigma_s} \right) \right)^{\frac{1}{(1-\sigma_s)(1+\delta_s)}}. \tag{6.10}$$

The expenditure share $\lambda_{ij,s} = P_{ij,s}^{1-\sigma_s} / P_{j,s}^{1-\sigma_s}$ can be expressed as

$$\lambda_{ij,s} = \frac{(\tau_{ij,s}(1 + t_{ij,s}) c_{i,s})^{-\varepsilon_s} c_{i,s}^{-\delta_s} \frac{R_{i,s}}{c_{i,s}} \chi_{ij,s}}{\sum_l (\tau_{lj,s}(1 + t_{ij,s}) c_{l,s})^{-\varepsilon_s} c_{l,s}^{-\delta_s} \frac{R_{l,s}}{c_{l,s}} \chi_{lj,s}} \tag{6.11}$$

where the trade elasticity is $\varepsilon_s \equiv (\sigma_s - 1)(1 + \eta_s)$ and $\chi_{ij,s} \equiv \xi_{ij,s}^{1-\sigma_s}$.

The total expenditure of location j on sector s can be expressed as

$$E_{j,s} = \beta_{j,s}(Y_j + D_j + T_j) + \sum_{k=1}^{S} \alpha_{j,sk} R_{j,k} \tag{6.12}$$

where $T_j = \sum_{i=1}^N \sum_{s=1}^S \frac{t_{ij,s}}{1+t_{ij,s}} X_{ij,s}$ is tariff revenue. Given that $X_{ij,s} = \lambda_{ij,s} E_{j,s}$, we have

$$T_j = \sum_{i=1}^{N} \sum_{s=1}^{S} (\frac{\sigma_s - 1}{\sigma_s}) \frac{t_{ij,s}}{1 + t_{ij,s}} \lambda_{ij,s} \left(\beta_{j,s}(Y_j + D_j + T_j) + \sum_{k=1}^{S} \alpha_{j,sk} R_{j,k} \right).$$

By rearranging the equation, we have

$$T_j = \frac{\sum_{i=1}^N \sum_{s=1}^S (\frac{\sigma_s - 1}{\sigma_s}) \frac{t_{ij,s}}{1+t_{ij,s}} \lambda_{ij,s} \left(\beta_{j,s}(Y_j + D_j) + \sum_{k=1}^S \alpha_{j,sk} R_{j,k} \right)}{1 - \sum_{i=1}^N \sum_{s=1}^S (\frac{\sigma_s - 1}{\sigma_s}) \frac{t_{ij,s}}{1+t_{ij,s}} \lambda_{ij,s} \beta_{j,s}}.$$

Substituting tariff revenues in the equation for expenditure, we have

$$E_{j,l} = \frac{\beta_{j,l}\left(Y_j + D_j + \sum_{i=1}^{N}\sum_{s=1}^{S}(\frac{\sigma_s-1}{\sigma_s})\frac{t_{ij,s}}{1+t_{ij,s}}\lambda_{ij,s}\left(\sum_{k=1}^{S}\alpha_{j,sk}R_{j,k}\right)\right)}{1-\sum_{i=1}^{N}\sum_{s=1}^{S}(\frac{\sigma_s-1}{\sigma_s})\frac{t_{ij,s}}{1+t_{ij,s}}\lambda_{ij,s}\beta_{j,s}} + \sum_{k=1}^{S}\alpha_{j,sk}R_{j,k}. \tag{6.13}$$

The location-sector-level revenues are expressed as

$$R_{i,s} = \sum_{i=1}^{N}\left(1+\frac{t_{ij,s}}{\sigma_s}\right)\frac{\lambda_{ij,s}}{1+t_{ij,s}}E_{j,s}. \tag{6.14}$$

The location-level income is expressed as

$$Y_i = \sum_{s=1}^{S}(1-\alpha_{i,s})R_{i,s}. \tag{6.15}$$

General Equilibrium. For given trade imbalance $\{D_i\}$ and tariffs $\{t_{ij,s}\}$, a general equilibrium is summarized by bilateral expenditure shares at the location-pair-sector level $\{\lambda_{ij,s}\}$, location-sector level expenditures $\{E_{i,s}\}$, location-sector level revenues, $\{R_{i,s}\}$, and location level wages, $\{w_i\}$, satisfying the following system of equations:

(Unit production cost)

$$c_{i,s} = w_i^{1-\alpha_{i,s}}\prod_{k=1}^{S}P_{i,k}^{\alpha_{i,ks}}, \quad i=1,\dots,N,x, \quad \forall s \in \mathcal{S} \tag{6.16}$$

(Multilateral price index)

$$P_{j,s} = \left(\sum_i((1+t_{ij,s})\tau_{ij,s}c_{i,s})^{(1-\sigma_s)(1+\delta_s)}\left(\frac{E_{j,s}}{c_{i,s}}\right)^{\delta_s}\left(\frac{R_{i,s}}{c_{i,s}}(\tilde{\xi}_{ij,s})^{1-\sigma_s}\right)\right)^{\frac{1}{(1-\sigma_s)(1+\delta_s)}}, \quad j=1,\dots,N,x, \quad \forall s \in \mathcal{S} \tag{6.17}$$

(Endogenous market entry cost term)

$$\tilde{\xi}_{ix,s} = (\frac{\tilde{\sigma}_s}{\delta_s\sigma_s})^{\frac{1}{1-\sigma_s}}\left(\frac{\sigma_s}{\tilde{\sigma}_s}\left[(\eta_{i,s}^{\sigma_s-1}-1)^{\frac{\theta}{\sigma_s-1}}(f_{i,s}^{L})^{-\delta_s}+(f_{ix,s})^{-\delta_s}\right]^{-\frac{1}{\delta_s}}\right)^{\frac{\delta_s}{\sigma_s-1}}(f_{i,s}^{e})^{\frac{1}{\delta_s-1}}, \quad i=1,\dots,N, \quad \forall s \in \mathcal{S} \tag{6.18}$$

$$\tilde{\xi}_{ij,s} = (\frac{\tilde{\sigma}_s}{\delta_s\sigma_s})^{\frac{1}{1-\sigma_s}}\left(\frac{\sigma_s}{\tilde{\sigma}_s}\left[f_{ij,s}\right]\right)^{\frac{\delta_s}{\sigma_s-1}}(f_{i,s}^{e})^{\frac{1}{\delta_s-1}}, \quad i=1,\dots,N,x, \quad j=1,\dots,N, \quad \forall s \in \mathcal{S} \tag{6.19}$$

$$\tilde{\xi}_{xx,s} = (\frac{\tilde{\sigma}_s}{\delta_s\sigma_s})^{\frac{1}{1-\sigma_s}}\left(\frac{\sigma_s}{\tilde{\sigma}_s}\left[f_{xx,s}\right]\right)^{\frac{\delta_s}{\sigma_s-1}}(f_{x,s}^{e})^{\frac{1}{\delta_s-1}}, \quad \forall s \in \mathcal{S} \tag{6.20}$$

(Knowledge stocks and flows)

$$\eta_{i,s} = \Gamma(1-1/\bar{\beta})(\sum_{k\in\mathcal{S}}\sum_{u\in\mathcal{N}} H_{uk,is})^{\bar{\beta}}, \quad i=1,...,N, \quad \forall s\in\mathcal{S} \tag{6.21}$$

$$H_{is,jk} = \left(\frac{\theta\Gamma\bar{\beta}}{\sigma_s}\frac{1}{c_{i,s}f^H_{i,s}}\lambda_{ij,s}\frac{E_{j,s}}{\sum_l E_{j,l}}\lambda_{jx,k}E_{x,k}\left[\frac{(\eta_{j,k}^{\sigma_k-1}-1)^{\delta_k}\eta_{j,k}^{\sigma_k-2}(f^L_{j,k})^{-\delta_k}}{(\eta_{j,s}^{\sigma_k-1}-1)^{\delta_k+1}(f^L_{j,k})^{-\delta_k}+(f_{jx,k})^{-\delta_k}}\left(\frac{\eta_{j,k}}{\Gamma}\right)^{\frac{\bar{\beta}-1}{\bar{\beta}}}P_{x,k}^{1-\sigma_k}\right]\right)^{\gamma^H-1}, \tag{6.22}$$

$$i,j=1,...,N, \quad \forall s,k\in\mathcal{S} \tag{6.23}$$

(Expenditure share)

$$\lambda_{ij,s} = \frac{(\tau_{ij,s}(1+t_{ij,s})c_{i,s})^{-\varepsilon_s}c_{i,s}^{-\delta_s}\frac{R_{i,s}}{c_{i,s}}(\tilde{\xi}_{ij,s})^{1-\sigma_s}}{\sum_l(\tau_{lj,s}(1+t_{ij,s})c_{l,s})^{-\varepsilon_s}c_{l,s}^{-\delta_s}\frac{R_{l,s}}{c_{l,s}}(\tilde{\xi}_{lj,s})^{1-\sigma_s}}, \quad i=1,...,N,x, \quad j=1,...,N,x, \quad \forall s\in\mathcal{S} \tag{6.24}$$

(Expenditure)

$$E_{j,l} = \frac{\beta_{j,l}\left(Y_j+D_j+\sum_{i=1}^N\sum_{s=1}^S(\frac{\sigma_s-1}{\sigma_s})\frac{t_{ij,s}}{1+t_{ij,s}}\lambda_{ij,s}\left(\sum_{k=1}^S\alpha_{j,sk}R_{j,k}\right)\right)}{1-\sum_{i=1}^N\sum_{s=1}^S(\frac{\sigma_s-1}{\sigma_s})\frac{t_{ij,s}}{1+t_{ij,s}}\lambda_{ij,s}\beta_{j,s}} + \sum_{k=1}^S\alpha_{j,sk}R_{j,k}, \quad j=1,...,N,x, \quad \forall l\in\mathcal{S} \tag{6.25}$$

(Revenues)

$$R_{i,s} = \sum_{i=1}^N\left(1+\frac{t_{ij,s}}{\sigma_s}\right)\frac{\lambda_{ij,s}}{1+t_{ij,s}}E_{j,s}, \quad i=1,...,N,x, \quad \forall s\in\mathcal{S} \tag{6.26}$$

(Income)

$$Y_i = w_iL_i = \sum_{s=1}^S(1-\alpha_{i,s})R_{i,s}, \quad i=1,...,N,x. \tag{6.27}$$

Welfare. The expenditure share can be expressed as

$$\lambda_{ij,s} = \frac{((1+t_{ij,s})\tau_{ij,s}c_{i,s})^{-\varepsilon_s}(\frac{E_{j,s}}{c_{i,s}})^{\delta_s}(\frac{R_{i,s}}{c_{i,s}})^{\delta_s}(\tilde{\xi}_{ij,s})^{1-\sigma_s}}{P_{j,s}^{-\varepsilon_s}}$$

where $\varepsilon_s = (\sigma_s-1)(1+\delta_s)$. Set $i=j$ and using $(1+t_{jj,s})\tau_{jj,s}=1$, we have

$$P_{j,s} = \lambda_{jj,s}^{1/\varepsilon_s}(\frac{R_{j,s}}{c_{j,s}})^{-1/\varepsilon_s}(\frac{E_{j,s}}{c_{j,s}})^{-\delta_s/\varepsilon_s}c_{j,s}(\tilde{\xi}_{jj,s})^{1-\sigma_s}$$

Similar to Costinot and Rodríguez-Clare (2014), we can express the real income $C_j = (Y_j + D_j + T_j)/P_j$ as

$$C_j = \left(Y_j + D_j + \frac{\sum_{i=1}^{N+1} \sum_{s=1}^{S} (\frac{\sigma_s - 1}{\sigma_s}) \frac{t_{ij,s}}{1+t_{ij,s}} \lambda_{ij,s} (\beta_{j,s}(Y_j + D_j) + \sum_{k=1}^{S} \alpha_{j,sk} R_{j,k})}{1 - \sum_{i=1}^{N+1} \sum_{s=1}^{S} (\frac{\sigma_s - 1}{\sigma_s}) \frac{t_{ij,s}}{1+t_{ij,s}} \lambda_{ij,s} \beta_{j,s}} \right) \tag{6.28}$$

$$\times \frac{1}{Y_j} \prod_{s=1}^{S} \prod_{k=1}^{S} \left(\lambda_{jj,k}^{-1} (\frac{R_{j,k}}{Y_j}) (\frac{E_{j,k}}{Y_j})^{\delta_k} (\tilde{\xi}_{jj,s})^{1-\sigma_s} \right)^{\frac{\beta_{j,s} \tilde{a}_{j,sk}}{\varepsilon_k}}. \tag{6.29}$$

PUBLICATIONS

Information, Social Media and International Trade: Theory and Evidence Using Twenty Million Online Postings

Working Paper No. WP/2026/064