Geographic Spillovers and Firm Exports: Evidence from China*

Lin Tian[†], Yue Yu[‡]

January 14, 2023

Abstract

This paper examines the importance of domestic economic geography for a country's export performance. We first develop a simple model in which the geographic integration of exporters reduces international trade frictions. We test the model's predictions by leveraging the expansion of China's high-speed rail (HSR) as a quasi-experiment to provide plausibly exogenous variation in the extent of (effective) geographic integration among Chinese exporters. The analysis reveals strong evidence that the improved integration of exporters within the same sectors positively impacts a firm's export performance, both intensively and extensively. We also provide evidence that an HSR connection strengthens knowledge spillovers among exporters to overcome frictions in accessing foreign markets.

Keywords: Agglomeration, Knowledge Diffusion, Informal Trade Barriers, China's High-speed Rail

JEL: R10, R12, F14, O18

[*]We thank ...

[†]INSEAD & CEPR. Email: lin.tian@insead.edu.

[‡]University of Toronto. Email: yueyu.yu@rotman.utoronto.ca.

1 Introduction

How do international trade and domestic economic geography interact? The literature has shown that international trade plays a significant role in reshaping the spatial distribution of economic activities domestically (see, e.g., Autor et al., 2013; Kovak, 2013; Caliendo et al., 2019). However, we know little about whether and how the domestic distribution of economic activities affects a country's trade performance.[1] In particular, despite a large literature on the role of domestic transport infrastructure in economic outcomes, there is almost no evidence on how the infrastructure alters the domestic geographic landscape by bringing firms closer together and, in turn, affects firm-level export performance. Credibly identifying the effects of domestic exporter integration on firm trade performance is important for policy making but has also been a challenging empirical problem, given the difficulty in separating the geographic integration that arises from location-specific externalities and site-specific fundamentals that are correlated with both firm location and export decisions.

This paper examines the importance of the domestic geographic integration of exporters for firms' integration with the international market. We overcome the identification challenge by exploiting a unique quasi-experiment that creates plausibly exogenous variation in the extent of geographic integration among exporters: The development of a high-speed rail (HSR) network, a passenger-only service, dramatically increases travel speed between Chinese cities and promotes geographic integration. We examine how connection to this major infrastructure project affects firm export decisions and performance.

Our assessment of the impact of geographic integration on firm export is guided by a simple model. The model proposes a particular microfoundation for how exporters may benefit from domestic integration by focusing on the diffusion of export-specific knowledge across space. We embed a knowledge diffusion model (Buera and Oberfield, 2020; Berkes et al., 2021) in a standard Melitz (2003) framework. In the model, firms acquire export-specific knowledge to overcome informal trade frictions. Whereas informal trade frictions are broadly defined, here we consider information frictions and market penetration costs. Export-specific knowledge can be acquired through exchanges with other exporters, which is subject to geographic frictions and hence is more effective between firms closer to each other. The model predicts that the concentration of export activities around a firm increases the firm's productivity in the export market—a result reminiscent of classic ag-

[1] Notable exceptions include Fajgelbaum and Redding (2022); Coşar and Demir (2016); and Cosar and Fajgelbaum (2016).

glomeration externalities. The model further predicts that in response to a reduction in geographic frictions, which promotes the effective geographic integration of exporters, firms would increase export volumes along both the extensive and intensive margins.

To examine the model's predictions, we use early-phase expansion of the HSR network in China between 2008 and 2013 as a quasi-experiment. Constructing a novel dataset that combines micro-level data on Chinese export transactions with detailed information on the HSR network expansion in China, we assess the causal impact of geographic integration on firm export performance. Our adopted measures of exporter integration are derived from the model and capture a firm's access to export activities in the same sector at all surrounding locations via the available railway transportation network. The identification strategy exploits the staggered opening of HSR stations and lines to adopt a triple-difference estimation model. We examine whether firms that are more connected with other exporters in the same sector as a result of HSR improved their export performance after the opening. Importantly, this strategy allows us to control for prefecture-specific shocks, which greatly alleviates conventional concerns about the endogenous locations and opening times of HSR stations. The validity of the triple-difference estimate depends on the parallel trends assumption with regard to relative differences in the export performance of firms in the same cities across different sectors without an HSR connection. We provide more details on the institutional background and event study results to argue that the routing and timing of HSR connections are unlikely to be correlated with the levels of and trends in export activities, as well as an extensive battery of robustness checks to examine the validity of the identifying assumptions.

We find that with the opening of an HSR station, the improved integration of exporters within the same sectors increases a firm's export revenue. This is driven by a reduction in the unit price of exported products and an increase in export volume, consistent with the hypothesis that geographic spillovers reduce variable or fixed costs in accessing foreign markets. We also find evidence of improved export performance along the extensive margin: Firms export to more destination countries, update their export product varieties more frequently, and expand their export product menu. The results are robust to considering a lagged treatment, dropping the largest sector in a city, and restricting the sample to a balanced panel of firms and firms in prefectures with an HSR station by 2016. By adopting a nonparametric permutation test for inferences—which assigns treatment randomly across all cities—we show that the p-values of the estimates are similar to the baseline estimates. We also provide evidence that the estimates are not confounded by improvement in internet connectivity.

In the second part of the paper, we explore the economic mechanisms behind the empirical results to investigate how exporters benefit from greater geographic integration. We first show that the benefits are specific to exporters using pairwise analysis that shows that connected firms converge on the set of export destinations, but not on the set of export products. Further, estimating similar triple difference specifications, we show that the baseline results are not confounded by improved integration with general production activities (domestic and export) or other types of Marshallian externalities; these include proximity to suppliers and customers and labor market pooling. Together with the results whereby the effects are stronger for firm types that tend to incur greater knowledge acquisition costs, the results support the mechanism highlighted in the model: Face-to-face interactions between firms in the export market improve export performance, because firms can more efficiently acquire export knowledge and/or because firms can more easily share knowledge with one another. Finally, we present additional evidence on the nature of the export-specific knowledge being acquired by examining changes in the export product qualities after the HSR opening. We show that firms acquire knowledge that improves the quality of products exported to existing destinations and facilitates exports to higher-quality destinations.

This paper is related to several strands of the literature. First, we enrich discussion of the interactions between globalization and economic geography. While a large literature demonstrates the heterogeneous impact of trade on local economic activities within countries (e.g., Autor et al., 2013; Kovak, 2013; Caliendo et al., 2019), we show how the domestic integration of exporting firms can drive international trade. Previous studies show how heterogeneity in shipping costs across different locations within a country can affect export (e.g., Fajgelbaum and Redding, 2022; Cosar and Fajgelbaum, 2016). Our empirical setting allows us to make a distinction between the movement of people (and their ideas) and the movement of goods. Since HSR is used exclusively for passenger service, we disentangle the effects of facilitated human interactions from the transport costs of shipping goods, in contrast to many transportation innovations that move both. This is important because some interventions (e.g., roads, air travel) necessarily move both, while others (e.g., high-speed rail, the internet) move only ideas. Distinguishing these two aspects provides guidance to policymakers when allocating scarce resources.[2]

This paper also contributes to an extensive literature on the role of geographic concentration in

[2] In this regard, our paper is also closely related to contemporaneous work by Marin et al. (2020) and Bakker (2021), both of which leverage structural models to quantitatively assess the strength of geographic spillovers in export. We complement these two papers by providing microfoundations for the reduced-form assumptions adopted in their work. Using our unique empirical setting, we provide causal empirical evidence for such export-specific agglomeration forces.

firms' activity. Many theories have suggested that learning from neighbors can determine firms' production decisions, technology adoption, and export behavior (Foster and Rosenzweig, 1995; Fernandes and Tang, 2014). However, it has always been difficult empirically to disentangle the geographic concentration that arises from positive spillover effects from other site-specific characteristics that attract firms to locate in a specific location and affect their production and export decisions (Aitken et al., 1997; Bernard and Jensen, 2004; Koenig et al., 2010; Baum-Snow et al., 2020). We address the identification problem by employing the HSR expansion in China, which promotes geographic integration and creates plausibly exogenous variations in the effective concentration of exporters.

This paper also relates to a growing literature on informal trade barriers in the export market. Recent literature suggests that one major friction for firms seeking to access foreign markets is information trade barriers, in contrast to the physical transportation cost to ship a product from one country to another (Rauch, 1999; Lovely et al., 2005; Allen, 2014; Startz, 2016; Ahlfeldt and Feddersen, 2018; Steinwender, 2018). Information trade barriers include contractual cost and information frictions regarding the price, consumer preference in destination countries, and so on. We focus on information frictions in trade and show that HSR, which promotes face-to-face interactions (Dong et al., 2020), helps exporting firms overcome information frictions.

Our theory builds on a flourishing literature on idea flows and economic growth.[3] A growing number of tractable models have been developed in this literature to microfound the learning, imitation, and knowledge diffusion between firms and regions (Lucas Jr and Moll, 2014; Perla and Tonetti, 2014; Buera and Oberfield, 2020; Huang and Zenou, 2020; Berkes et al., 2021). We adapt the insights from this literature to highlight the role of knowledge diffusion in the export market. We further develop empirical strategies, which can be adapted to other settings with changing commuting infrastructure, to investigate the existence and strength of knowledge diffusion across space.

More broadly, our work relates to the literature on spatial knowledge transmission (Marshall, 1920; Duranton and Puga, 2003; Cristea, 2011; Keller and Yeaple, 2013; Davis and Dingel, 2019; Tian, 2021). In contrast to studies that focus primarily on knowledge exchange within locations, this paper focuses on knowledge creation and transmission between exporting firms located in *different* regions. In particular, we show that the transportation infrastructure that reduces commuting time across regions, like HSR in China, can defy geographic boundaries and facilitate geographic spillovers between exporting firms. In doing so, our work expands empirical evidence on the nature

[3]See (Buera and Lucas Jr, 2018) for a review.

and extent of spatial knowledge transmission and its associated agglomeration externalities.

Finally, this paper connects to a growing literature on the impacts of transportation infrastructure investment (e.g., Giroud, 2013; Redding and Turner, 2015; Coşar and Demir, 2016; Lin, 2017; Bernard et al., 2019; Heuermann and Schmieder, 2019; Charnoz et al., 2018; Hayakawa et al., 2021). Our paper focuses on the interaction between transportation infrastructure and exporting activities. Through our analysis, we reveal the role for infrastructure in reducing travel time for individuals and thus encouraging more frequent face-to-face interactions, which spurs knowledge diffusion and idea exchanges among exporting firms.

The rest of the paper proceeds as follows. In Section 2, we develop a simple stylized model that generates testable predictions and guides the empirical analysis. Section 3 presents the identification strategy, data, and institutional background of the HSR expansion in China. Section 4 reports the empirical findings and conducts robustness checks. Section 5 investigates the economic channels that drive the empirical results. Section 6 concludes.

2 A Stylized Model

In this section, we present a model to fix ideas and motivate the empirical investigation in subsequent sections. The model embeds a knowledge diffusion model (Buera and Oberfield, 2020; Berkes et al., 2021) in a standard Melitz (2003) framework to incorporate a particular mechanism that drives how exporters benefit from domestic integration: Integration with other exporters allows firms to acquire knowledge more easily and, in turn, overcome frictions in export markets. While the empirical analysis in the next section remains agnostic on the underlying mechanism(s) that drive the improved export performance when firms are geographically integrated, we present further empirical evidence consistent with this microfoundation of geographic knowledge spillovers in Section 5.

The economy is comprised of a set of S sectors, indexed by $s \in \mathcal{S}$. We assume that the home country has N asymmetric trading partners, indexed by $n \in \mathcal{N}$.[4] Given income, a representative consumer decides how to allocate consumption over local final goods from all sectors with a Cobb-Douglas aggregator. Within each sector s, differentiated varieties are aggregated in a CES manner.

[4] We omit subscripts where appropriate for the home country to save on notation.

These preferences yield the following demand curve for variety z in country n:

$$q_n(z) = p_n(z)^{-\sigma} P_{ns}^{\sigma-1}(\alpha_s E_n),$$

where P_{ns} and $\alpha_s E_n$ are, respectively, the sectoral price index and expenditure in the export market n and sector s, and $\sigma > 1$ is the elasticity of substitution across varieties.[5]

The economy consists of a continuum of heterogeneous firms that compete monopolistically within a sector (Melitz, 2003). Each firm manufactures a unique variety with constant marginal cost and a fixed per period overhead cost, f^d.[6] The amount of labor required to produce q units for a firm with productivity level φ is $l = f^d + \frac{q}{\varphi}$. Firms enter the market by paying an entry cost, f^e, to draw their productivity from a distribution $g(\varphi)$. Conditional on its productivity draw, a firm has the option to exit the market. Finally, firms are also exogenously assigned a production location.[7]

Given our assumptions on consumer preferences and market structure, the standard Melitz results apply for firms that sell only in the domestic market, with firm revenue and profit determined by the productivity draw φ_z. Conditional on remaining in the market, a firm decides whether or not to export to a particular destination. Consistent with the standard model, exporting incurs an additional bilateral fixed cost, denoted by f_n^x, and standard bilateral iceberg transportation costs, $\tau_{sn} > 1$, for all export destinations n.

In contrast to the standard model, we explicitly incorporate in the model informal barriers to enter the export market, such as uncertainties over foreign consumer preferences (Startz, 2016); additional distribution costs (Yamawaki, 1991); and information frictions in the export market (Aitken et al., 1997; Allen, 2014). To overcome these frictions, firms need to acquire export-specific knowledge, which is endogenously determined through a two-step process following Buera and Oberfield (2020). At the beginning of each period, every firm z receives an idiosyncratic and independently distributed export-specific insight ω_z. These insights can include, but may not be restricted to, knowledge about demand in the destination market, distribution channels, production techniques, and marketing strategies. Firms invest to build their insight collection capacity, denoted as λ_z, which increases firms' insight draws on average. To improve λ_z by one unit, a firm of sector s

[5] See Appendix A.1 for model details and derivations and Appendix A.4 for proofs for all theoretical results presented in this section.

[6] We hence use z to denote both a variety and a firm.

[7] It is straightforward to extend the model to include endogenous location choices by firms, following setups in, for example, Gaubert (2018) and Tian (2021). However, we adopt the simplifying assumption of exogenously assigned location in the baseline model, since we focus on the set of firms that do not change locations over time in our empirical analysis.

incurs a unit cost c. In the second step, firms' self-collected insights are diffused as a process that involves interactions among all exporters of the same sector in the country. An exporter z can choose to adopt another exporter's insight, $\omega_{z'}$, from all $z' \in \mathcal{Z}_s \backslash z$, subject to a discount of $f(d_{zz'}) < 1$, $\forall z \neq z'$, where $f(d_{zz'})$ captures the geographic frictions that discount the quality of interactions. Specifically, $d_{zz'}$, with $f'(\cdot) < 0$, is the bilateral *effective* distance between firms z and z' and is determined by both the geographic distance and the connectivity between the two locations. This reflects the idea that better integrated regions facilitate acquisition of knowledge remotely. At the end of the period, given the vector of insights, exporters adopt one insight to maximize the profit:

$$\varphi_s^x(z) = \max\{\omega_z, \max_{z' \in \mathcal{Z}_s \backslash z} \{\omega_{zz'}\}\} = \max\{\omega_z, \max_{z' \in \mathcal{Z}_s \backslash z} \{f(d_{zz'})\omega_{z'}\}\}.^{8} \tag{1}$$

Following conventional literature (e.g., Berkes et al., 2021), we further assume that all insight draws follow a Fréchet distribution with a cumulative distribution function (CDF) given by $F(\omega) = e^{-\lambda_z \omega^{-\theta}}$, where $\theta > 1$ is the dispersion parameter and the scale parameter $\lambda_z > 0$ is the endogenously determined investment in insight collection capacity. Since multivariate Fréchet draws with the same dispersion parameter are max-stable, $\varphi_s^x(z)$ in (1) also follows a Fréchet distribution with dispersion parameter θ and scale parameter given by

$$\tilde{\lambda}_s(z) \equiv \lambda_z + \sum_{z' \in \mathcal{Z}_s \backslash z} \lambda_{z'} \left[f(d_{zz'})\right]^{\theta}. \tag{2}$$

It is worth highlighting that $\tilde{\lambda}_s(z)$ is a key variable in our model. It reflects a firm's access to other exporters (and their insights), or the extent of geographic integration. All else equal, exporters that are better geographically integrated have greater $\tilde{\lambda}_s(z)$ and tend to receive higher export-specific knowledge on average.

Our framework enriches the standard Melitz model with the addition of export-specific geographic spillovers. Firms in the export market acquire export-specific insights $\varphi_s^x(z)$, which are drawn randomly from the endogenously determined distribution described earlier. Given $\varphi_s^x(z)$,

[8] The process is similar to the "imitate or reject" knowledge diffusion framework described by Buera and Oberfield (2020). A firm adopts learned insights acquired from other firms through the interaction, but only if such insights improve the firm's own knowledge.

total revenue obtained from exporting is

$$r_s^x(z) = \left(\frac{\sigma}{\sigma - 1}\right)^{1-\sigma} \left(\sum_{n\in\mathcal{N}_z} R_{sn} P_{sn}^{\sigma-1} \tau_{sn}^{1-\sigma}\right) [\varphi_z \varphi_s^x(z)]^{\sigma-1}, \tag{3}$$

where $\mathcal{N}_z$ is the set of export markets that firm z engages in and R_{sn} and P_{sn} are, respectively, the sectoral aggregate revenue and price index in export market n. Given the Fréchet assumption of the export-specific insight draws, the expected value of export revenue is given by

$$\mathbb{E}(r_s^x(z)) = \kappa_1 \left(\sum_{n\in\mathcal{N}_z} R_{sn} P_{sn}^{\sigma-1} \tau_{sn}^{1-\sigma}\right) \left[\varphi_z \left(\tilde{\lambda}_s(z)\right)^{\frac{1}{\theta}}\right]^{\sigma-1}, \tag{4}$$

where $\kappa_1 = \left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} \left(\Gamma\left(1 - \frac{1}{\theta}\right)\right)^{\sigma-1}$ and $\Gamma(\cdot)$ is the gamma function.

Recall that potential exporters pay a unit cost of c to improve their insight collection capacity λ_z. Given the CES preferences, the total expected export profit is

$$\mathbb{E}(\pi_s^x(z)) = \kappa_2 \left(\sum_{n\in\mathcal{N}_z} R_{sn} P_{sn}^{\sigma-1} \tau_{sn}^{1-\sigma}\right) \left[\varphi_z \left(\tilde{\lambda}_s(z)\right)^{\frac{1}{\theta}}\right]^{\sigma-1} - c\lambda_z - \sum_{n\in\mathcal{N}_z} f_n^x, \tag{5}$$

where $\kappa_2 = \frac{1}{\sigma}\left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} \left(\Gamma\left(1 - \frac{1}{\theta}\right)\right)^{\sigma-1}$ and $\mathcal{N}_z$ is the set of countries z exports to. We therefore have $\lambda^*(z) \equiv \arg\max_{\lambda_z} \mathbb{E}(\pi_s^x(z))$ pinned down by the first-order condition of the expected export profit of a firm with respect to λ_z.

Finally, firms will enter the export market if the expected profit in (5) exceeds the fixed cost of exporting. Given our assumptions on consumer preferences and market structure, as well as production and export costs, we can show that firms sort not only according to their productivity, as is in Melitz (2003), but also $\tilde{\lambda}_s(z)$, the scale parameter that captures the extent of geographic integration among exporters. Formally, the productivity cutoff is given by

$$\varphi_{sn}(\tilde{\lambda}_s(z)) = \frac{\tau_{sn}}{P_{sn}\left(\tilde{\lambda}_s(z)\right)^{\frac{1}{\theta}}} \left[\frac{f_n^x}{\kappa_2 R_{sn}}\right]^{\frac{1}{\sigma-1}}, \tag{6}$$

where $\varphi_{sn}(\tilde{\lambda}_s(z))$ is the lowest productivity required in sector s to export to country n for a given value of $\tilde{\lambda}_s(z)$. Within a sector s, firms with $\varphi_{sn}(\tilde{\lambda}_s(z))$ export to n.

Relative to the results yielded by the standard Melitz model, our framework highlights the role of geographic integration in a firm's export decisions. This is reflected in (4) and (6), in which both the intensive and extensive margins of export are a function of geographic spillovers—i.e., $\tilde{\lambda}_s(z)$, a decreasing function of firm's effective distance with other exporters $d(zz')$. We formalize this result in the following proposition.

Proposition 1. *Assume that an export market's price index P_{sn} is fixed for all s and n. All else equal and conditioning on exporting, the firm's export revenue increases with its geographic integration with other exporters in the same sector. Formally, we have*

$$\frac{d\ \mathbb{E}\left(r_s^x(z)\right)}{d\,\mathbf{d}_{\mathbf{zz'}}} < \mathbf{0}; \quad \frac{d\ \mathbb{E}\left(\pi_s^x(z)\right)}{d\,\mathbf{d}_{\mathbf{zz'}}} < \mathbf{0}.$$

This result is reminiscent of the classic agglomeration externalities discussed by Marshall (1920) and is consistent with the stylized facts in Koenig et al. (2010). It is difficult empirically to identify the results using cross-sectional data due to unobserved factors that drive both the geographic location of firms and export values. In other words, cross-sectional analysis cannot separately identify the geographic concentration that arises from location-specific external economies and the geographic concentration that arises from exogenous site-specific characteristics. To aid empirical investigation and overcome this identification challenge, we consider model predictions when there is exogenous variation in the extent of geographic integration through reduction in bilateral geographic frictions between two locations—e.g., improvement in transportation infrastructure.

Proposition 2. *Assume that the export market's price index P_{sn} is fixed for all s and n. In response to an exogenous reduction in $d_{zz'}$, (i) firms already exporting increase their volume of export on average; and (ii) the share of exporting firms, denoted by Z_{sn}, increases. Formally, we have*

$$\frac{\partial\,\mathbb{E}(r_s^x(z))}{\partial d_{zz'}} < 0; \quad \frac{\partial Z_{sn}}{\partial d_{zz'}} < 0.$$

Proposition 2 provides testable predictions in the setting of transportation infrastructure improvement. First, improvement in infrastructure reduces geographic frictions and hence strengthens the integration of firms in locations that receive the new infrastructure. In particular, conditional on exporting, a firm would increase its export revenue. Firms would also increase their export activities along the extensive margin by entering into more export markets.[9]

[9]For simplicity, we consider single-product firms in the baseline model. In Appendix A.3, we extend the model, following Bernard et al. (2011), to include multi-product firms. In addition to generating the same set of theoretical predictions as in the baseline model, we further show that firms would increase the number of products exported in response to a reduction in geographic friction—another extensive margin of export.

3 Empirical Strategy

In this section, we first derive the regression specification based on the model to test predictions in Proposition 2. We then introduce the background of HSR development in China and discuss the empirical strategy in leveraging the expansion of the HSR network as a quasi-experiment to identify the effects of geographic integration on firm export performance in a triple difference model. Finally, we present the results.

3.1 Deriving Econometric Specification

Building on the theoretical insights developed in Section 2, we derive the econometric specification that allows us to identify the effect of geographic integration on firm exports. From (3), we derive the expected value of firm z's log export revenue:

$$\mathbb{E}\left(\log r_s^x(z)\right) = const + \log\left(\sum_n R_{sn} P_{sn}^{\sigma-1} \tau_{sn}^{1-\sigma}\right) + (\sigma - 1)\log\varphi_z + \beta \log \tilde{\lambda}_s(z), \tag{7}$$

where $\beta = (\sigma - 1)/\theta > 0$.[10] Assuming that (7) holds for all exporters in each period, we obtain the following estimating equation for $\mathbb{E}\left(\log r_{st}^x(z)\right)$:

$$y_{zst} = \alpha_{st} + \alpha_z + \beta x_{zst} + \iota_{zst}, \tag{8}$$

where y_{zst} is the log export revenue by firm z in sector s at time t, α_{st} is a fixed effect that is common across firms in a given sector-year, α_z is the time-invariant firm-specific productivity term, which also controls for any unobserved time-invariant location factors, $x_{zst} \equiv \log \tilde{\lambda}_{st}(z)$ is the extent of the firm's geographic integration, and $\iota_{zst} \equiv [(\sigma - 1)\ln\varphi(z;t) - \alpha_z] + [\log r_{st}^x(z) - \mathbb{E}(\log r_{st}^x(z))]$ is a combined error term that contains unobserved time-varying productivity shocks and deviations of log export revenue from its mean.

Our reduced-form coefficient of interest is β, a positive value of which lends empirical support to our model and shows the positive impact of exporter integration on firms' export performance. Empirically identifying the sign of β is challenging for a variety of reasons; paramount is that exporter locations could be endogeneous. Specifically, if certain location fundamentals (e.g., proximity to a port, availability of skilled workers) affect both the geographic concentration of firms and their export decisions, then estimating β through OLS would suffer from omitted variable bias,

[10] See Appendix A.2 for derivation details.

or $Cov(x_{zst}, \iota_{zt}) \neq 0$.[11] To address this problem, we find exogenous shocks to exporter integration that are uncorrelated with firms' export performances. We do so by exploiting a quasi-experiment: Expansion of the HSR network in China reduces geographic fictions between firms and thereby induces a plausibly exogenous change in the extent of geographic integration for exporting firms affected by the HSR network.

3.2 HSR in China

The HSR project in China is by far the world's longest high-speed railway network. It had a total length of 37,900 km at the end of 2020 and accounts for two-thirds of the world's total high-speed railway networks.[12] The development of an HSR network in China started in the early 2000s, when the State Council, in its revised Mid- to Long-Term Railway Development Plan, set the goal of a national high-speed rail grid composed of four north-south corridors and four east-west corridors, with a budget of around 4,000 billion yuan (State Council, 2008).[13] Since then, the HSR network has expanded rapidly, as shown in Figure 4.

The arrival of HSRs has reduced travel time by rail drastically across Chinese cities.[14] For example, the travel time between Beijing and Shanghai was shortened from 13 hours to 5 after the opening of the Beijing–Shanghai HSR. As a result, HSR has become a preferred commuting mode among passengers, especially business travelers. Nationally, HSR ridership increased from 300 million in 2010 to 830 million in 2014, of which the majority are business travelers (Lin, 2017; Dong et al., 2020).[15]

The criteria used by the Ministry of Railway to decide which cities are connected to the HSR network is unclear to the public. Zheng and Kahn (2013) propose two possibilities. One possible criteron is to maximize ridership by connecting megacities and cities that are expected to boom. Another possible criteron is associated with motives opposite to the first one— i.e., by connecting smaller and weaker cities to enhance growth potential and reduce spatial growth inequalities. They also test the differential economic growth between cities with and without an HSR station during 2001 and 2005. There are no significant differences in GDP growth, wage growth, or distance to

[11] We discuss other identification challenges and our empirical strategies in Section 3.4.

[12] https://en.wikipedia.org/wiki/High-speed_rail_in_China#cite_note-1

[13] Figure 5 provides more detailed discussion of the history of HSR development in China.

[14] Traditional trains run at a speed of 120 to 160 km/hr, while HSR runs at above 250 km/hr and can reach a top speed of 350 km/hr. Therefore, travel time across regions can be reduced by 40% to 60%.

[15] Competitive fare prices for HSR also contribute to the fast growth of ridership. These are on average USD 0.07/km, which is cheaper than HSR in Europe (USD 0.10-0.20/km) and Japan (>USD0.20/km) (Lin, 2017).

megacities, except for a slightly higher population growth in cities without an HSR connection.[16]

The advent of HSR has boosted local economic growth and promoted economic integration overall. For example, Lin (2017) documents that a city's employment increased by 7% after connecting to the HSR network, and Zheng and Kahn (2013) show that the opening of an HSR station is associated with rising real estate prices in nearby secondary cities. However, not all regions benefit equally from HSR expansion: Qin (2017) and Yu et al. (2019) find that connected peripheral regions experience a decrease in GDP per capita after connecting to the HSR network. The above findings are mainly based on city-level analysis, while evidence on the impacts of HSR on firms' performances is relatively limited, mainly because of the unavailability of high-quality panel data on firms over the period.

Anecdotally, however, there is evidence suggesting that HSR allows business managers to frequently visit commercial hubs hundreds of miles away to learn about market conditions and business insights. For instance, as reported by the *New York Times*, after HSR connected Changsha and Guangzhou, the sales manager of a Chinese garment export company based in Changsha increased the frequency of his business trips to Guangzhou from twice a year to once a month.[17] As he told the reporter, "More frequent access to my client base has allowed me to pick up on fashion changes in color and style more quickly... My orders have increased by 50 percent." Consequently, HSR is believed to encourage knowledge diffusion across regions, although direct empirical evidence is still limited.[18]

3.3 Measuring Geographic Spillovers

To estimate (8) empirically, we need to construct an empirical measure for the integration of exporters. Recall from (2) that we have $\tilde{\lambda}_s(z) \equiv \sum_{z' \in \mathcal{Z}_s} \lambda_{z'} \left[f(d_{zz'})\right]^{\theta}$. We adopt the following measure, which incorporates the modifications necessary to adapt to the empirical context:

$$x_{cst} \equiv \sum_{c' \neq c} \left(\sum_{z' \in \mathcal{Z}_{st_0}(c')} r^x_{st_0}(z') \right) \tilde{f}(d_{cc't}). \tag{9}$$

[16]This evidence is in line with the second criterion and is also consistent with our event study in Figure 2.

[17]*New York Times*, Sept 24 2013 (https://www.nytimes.com/2013/09/24/business/global/high-speed-train-system-is-huge-success-for-china.html).

[18]The only exception, to our knowledge, is Dong et al. (2020), who study the role of HSR in knowledge diffusion among academic researchers. They find that after HSR connects two cities, researchers from these two cities are more likely to become coauthors, and coauthor team productivity also increases.

In particular, we proxy for λ_z, the firm's own insight collection capacity, using the firm's export revenue. As shown in (4) of the model, firms that choose greater collection capacity expect higher export revenue on average.[19] Firm locations are also observed up to the prefecture level.[20] We therefore use the prefecture in which a firm is located to measure geographic frictions between firms, $\tilde{f}(d_{cc't})$. A firm's own city is excluded in the aggregation to avoid a mechanical correlation between y_{zst} and x_{cst}. Furthermore, to avoid reverse causality, we use the firm's average annual export revenue from 2000 to 2007—the last year before the rapid roll-out of the HSR network—to construct x_{cst}. We finally state the parametric specifications for the geographic friction function, $\tilde{f}(d_{cc't})$:

1. HSR connection status-weighted export revenue:

$$x_{cst} \equiv \sum_{c' \neq c} \left[\sum_{z' \in \mathcal{Z}_{st_0}(c')} r^x_{st_0}(z') \right] \times \mathbb{I}(\mathbf{HSR}_{cc't}), \tag{10}$$

where $\mathbb{I}(HSR_{cc't})$ is an indicator function that takes the value of 1 if there is a direct HSR connection between c and c' in year t, and 0 otherwise.

2. Travel time-weighted export revenue:

$$x_{cst} \equiv \sum_{c' \neq c} \left[\sum_{z' \in \mathcal{Z}_{st_0}(c')} r^x_{st_0}(z') \right] / \mathbf{T}(\mathbf{Train}_{cc't}), \tag{11}$$

where $\mathbf{T}(\mathbf{Train}_{cc't})$ is the shortest travel time by train (regular if there is no HSR).

In both (10) and (11), x_{cst} resembles the market access measure in the international trade literature—e.g., Donaldson and Hornbeck (2016)—but instead reflects a firm's access to other exporters (and, by association and through the lens of our model, their insights). Both measures incorporate the two key elements that determine exporter integration across space: All else equal, firms that are surrounded by more export activities or are better connected to other exporters are more integrated. The measures also emphasize the role the HSR connection plays in strengthening geographic integration for affected firms. The first measure in (10) adopts a more reduced-form and less restrictive form, while the second in (11) considers variations in travel time reductions

[19] More precisely, the expected export revenue depends on both the firm's exogenously given productivity φ_z and the endogenously determined insight collection capacity λ_z. However, as we show in Lemma 3 of Appendix A.3, λ_z is an increasing function of φ_z, which implies a direct positive association between a firm's expected export revenue and its insight capacity.

[20] A Chinese prefecture includes an urban core and suburban and rural areas. Mainland China is divided into 333 prefectures. We provide more details on the data in Section 3.5.

due to the HSR expansion across city pairs, but takes a stronger stance on how reduction in travel time benefits the firm's export performance. Following Lin (2017) and Kuang et al. (2021), we use commuting time via regular trains to measure commuting time between cities without HSR. We do not consider highway commuting because, as demonstrated by Lin (2017), the opening of an HSR station does not lead to any decrease in highway ridership. Next, we do not consider air travel because (a) the market share of air travel up to 2007 is fairly low (less than 1%) and (b) HSR passenger poaching effects on air travel are mild (Lin, 2017). The mismeasurement of commuting time changes will cause attenuation bias and render our estimates a lower bound of the true effect, if any.

3.4 Addressing Identification Challenges

To identify the effects of the HSR-induced strengthening of exporter integration across space, we need to address the potential endogeneity problem that arises from implementation of the infrastructure. In particular, the location and opening time of HSR stations may be correlated with long-term trends at city-year level that could potentially affect the firm's export decisions. To address this concern directly, we add to the estimating equation a prefecture-year fixed effect, α_{ct}, that controls for all location-specific shocks. Our baseline estimating equation thus becomes

$$y_{zst} = \alpha_{st} + \alpha_z + \alpha_{ct} + \beta x_{cst} + \iota_{zst}. \tag{12}$$

The regression equation in (12) adopts a triple difference model. The intuition for identification is as follows: HSR expansion is expected to bring higher export performance gains for an exporter connected with more exporters in the same sector compared with an exporter connected with fewer exporters. Therefore, the first difference is the time difference as we compare the export revenue of firms before and after the HSR connection. The second difference is the export growth in locations with and without HSR access. The last difference is the differential growth of firms located in the same city but exporting in different sectors. In adopting this setup, the treatment effect is effectively identified based on the within-city differences in the export performance of firms that export in different sectors.

This triple difference setup can address several endogeneity concerns convincingly. First, differential growth trends across sector-destination markets—e.g., strengthened export competition as a result of the HSR expansion—would be controlled by sector-year fixed effects. Second, one might be concerned that firms' location choices are endogenous to the anticipation of imminent HSR expansion. This will be addressed by firm fixed effects. Finally, a common concern associated with HSR construction is that the location of HSR stations and the timing of station openings are not random. A particular concern is that new HSR lines are introduced in areas with expected stronger or weaker export growth in the future. In our specification, this is taken care of by the prefecture-year fixed effects.

The validity of the triple difference estimate depends on the parallel trends assumption with regard to the relative differences in the export performance of firms in the same cities across different sectors in the absence of an HSR connection. The remaining threats to identification include local government lobbying for a HSR station to enhance the growth of certain sectors as well as the possibility of firms' sorting into new export hubs based on anticipating that these places will grow because of the HSR connection. When these happen, variation in the effective concentration of export activities x_{cst} would be correlated with ι_{zst}. To address the first concern, we argue that there is limited scope for local governments to influence the alignment and implementation timeline of the HSR, and show this using an event study in Section 4.1. We further examine the robustness of our results to dropping firms in the largest sector within a given prefecture, since these are firms that can potentially exert the greatest political influence to lobby for a local HSR connection. For the second concern, we control for the spatial sorting of firms by restricting the analysis to the set of firms that do not change locations for the entire study period. This is discussed further in Section 4.3.

Lastly, the success of this identification strategy also hinges on the fact that exporting firms from different sectors tend to concentrate in different regions in China. Otherwise, changes in the geographic integration of exporters would not create variations among firms from different sectors after their city connects to other cities via HSR. Reassuringly, as shown in Figure 6 in the Appendix, there is rich variation in the spatial concentration of export activities across sectors in China.

3.5 Data

We assemble a novel dataset for the empirical analysis. Our main data source is China's General Administration of Customs, which enables access to a panel dataset of all Chinese exporting firms from 2000 to 2013. Our central analysis focuses on 338,863 exporting firms that had exported at least one time before their prefecture connected to the HSR network. For each exporting firm, we observe its export revenue and export volume by product type (HS8 level) and destination country at annual frequency. We can also observe a series of firm characteristics, such as the prefecture a firm locates in, firm ownership status (private versus state owned), and the sector(s)—defined as an HS2 product category— in which it operates.

The customs data include the entire set of exporting firms regardless of firm size, which allows us to explore potential heterogeneous responses between small and large exporting firms. One limitation of the export data is that it does not include firms that never exported, and hence we cannot observe a firm's export entry decision directly. Fortunately, information about the destination countries a firm exported its products to offers an alternative way to measure its market entry decisions. As detailed in the next section, we use the numbers of export destinations and export product variety to study a firm's extensive decisions regarding entering specific export markets.

Information on the rollout of the HSR network and the opening dates of HSR stations comes from the China Railway Statistical Yearbook and ChinaMap (Li et al., 2016). We collect data on train schedules from the official website of the Chinese National Railway Administration (www.12306.cn) to calculate the amount of time it takes to commute between any two prefectures each year via regular trains and HSR. We consider both direct trains and one-stop transfers to allow HSR to reduce the commuting time for part of or an entire trip.

With the universe of export records and information on HSR rollout, we construct the two main independent variables to measure exporter integration in (10) and (11). Furthermore, we normalize each measure by the standard deviation of its change from 2007 to 2013 across firms to simplify interpretation of the economic magnitude of coefficient β. With the normalization, β can be interpreted as the change in the outcome variable if the HSR-driven increase in the geographic integration of exporters is one standard deviation higher by the end of the study period.[21]

4 Empirical Results

We present the estimation results in this section. We first show that HSR-induced improvement in the integration of exporters increases a firm's export activities both intensively and extensively. We then discuss robustness checks for a series of potential threats to identification.

[21] Figure 1 plots HSR-driven improvement in exporter integration across prefectures in China. For each prefecture, we calculate the average increase in exporter integration from 2007 (one year before the rollout of HSR) to 2013 (end of the study period) across firms in that prefecture. Under measure 1, only firms on the HSR network would experience an increase in the geographic integration of exporters. Measure 2, in contrast, accounts for the fact that a prefecture not directly connected but adjacent to the HSR network can also benefit from the HSR expansion. That said, as shown in Section 4, regressions based on the two measures produce quantitatively consistent results.

4.1 Event Study

Before estimating (12), we first conduct an event study to investigate whether there is systematic difference in terms of export growth between firms with and without an HSR opening by the end of 2013. We use the year before the HSR opening as the benchmark year and plot the coefficients in Figure 2. There are parallel trends between firms with and without an HSR station before the HSR connection. After the HSR connection, the positive response in firm export is the largest in year 0 and declines over time. Responses from the post-period are consistent with the conjecture discussed in Section 3.2, whereby HSR aims to connect cities that have relatively weaker growth potential in the future—and justifies our triple-difference research design.

4.2 Baseline Results

For the baseline analysis, we focus on investigating firms' export growth along several margins, using (12). We first estimate the impact of the strengthened geographic integration on firms' total export revenue. To examine the extensive margin of exporting decisions, we investigate, as part of the baseline analysis, changes in the number of foreign destinations to which a firm exports and the number of products a firm exports.

Estimation results are reported in Table 1. We find a statistically and economically significant increase in a firm's export revenue in response to an HSR-driven improvement in exporter integration. Table 1, Columns 1 and 2 report that a one-standard-deviation increase in integration leads to an increase in export revenue by 3% to 4%.[22] This result confirms the prediction in Proposition 2: Firms that have already exported increase their volume of exports in response to an increase in the integration of export activities. We also compare the results with and without controlling for prefecture-year fixed effects, which is equivalent to a conventional difference-in-differences method. As shown in Table 12, Panel D of the Appendix, the estimated effects would be halved without prefecture time-varying effects with larger standard errors. This finding further confirms that HSR tends to connect prefectures in which exporting firms are expected to decline, and hence it is crucial to control for prefecture-year fixed effects.

We next investigate whether the increase in export revenue is driven by price or volume effects. As shown in Table 2, the increase in export revenue is driven entirely by an increase in the export

[22] In the baseline analysis we cluster error terms at the level of prefecture, which is the level of HSR intervention. We show in Table 14, Panel B that the standard errors are virtually identical when clustering error terms by prefecture-sector-year.

volume, while the unit price goes down. A one-standard-deviation increase in exporter integration leads to an increase in export volume by 7% to 8%, while the unit price decreases by around 4%. These findings are consistent with the model assumption that the geographic integration of exporters helps firms overcome trade barriers. In doing so, affected firms would increase export quantity and reduce price (driven by the lower marginal costs).

Next, we study export firms' market entry decisions to examine the impacts on the extensive margins of export. In the baseline, we approximate a firm's entry decisions using the number of destination countries a firm exports to. As shown in Table 1, Columns 3 and 4, exporting firms sell their products to more foreign countries in response to an increase in exporter integration across space. A one-standard-deviation increase in exporter integration leads to firms exporting to 0.1 to 0.2 more foreign markets.

We study another aspect of the extensive margin: the export product variety. To conduct the analysis, we define each HS8 category as one product variety, then count the total number of product types a firm exports every year as well as the number of product categories the firm adds or drops in that year compared with the previous year's exports. As shown in Table 3, improved integration causes a firm to update its product set more frequently by adding and dropping more product categories per period (Columns 1 and 2). Columns 3 and 4 further show that firms tend to expand their exported product categories: A one-standard-deviation increase in the integration of exporters leads to an increase in export product variety by 3 to 4%. Recent research has shown that Chinese firms update their product categories frequently to overcome information frictions, especially regarding the tastes of foreign consumers (e.g., Startz, 2016). Our findings thus provide suggestive evidence that strengthened geographic integration among exporters reduces information frictions linked to accessing foreign markets. We explore the underlying economic mechanisms in more detail in Section 5.

4.3 Robustness Checks

We explore several robustness checks in this subsection. The first is that the analysis is undertaken at yearly level, but some HSR lines were only opened toward the latter part of the year—and thus there might not be enough time for firms to react to the improved exporter integration induced by HSR in the same year. To this end, we redefine the treatment of HSR using HSR status from the previous year and report the results in Table 12, Panel A of the Appendix. The results are very similar to the baseline outcome.

Next, we address remaining concerns that HSR is introduced to enhance connectivity for firms in certain sectors—e.g., if the largest sector in a prefecture expects to benefit from HSR connection, those firms may use their political connection to lobby the government to open an HSR station. This is unlikely the case, given that the planning for HSR begins years before implementation. Therefore, there is limited room for local governments to change the routes. With that in mind, we nonetheless conduct this robustness check by dropping firms in the largest sector within a given prefecture, which are likely firms with the greatest political influence. As shown in Table 12, Panel B of the Appendix, the results are in line with the baseline results.

Third, we show that the conclusions are unchanged if we restrict the sample to a balanced panel of firms to remove any potential composition effects. In the baseline, we include all firms with at least one year of export history before the HSR connection; hence firms that left the export market before the end of the study period are also included. For a robustness check, we restrict the sample to firms with a positive export volume in both 2007 and 2013. As shown in Table 12, Panel C of the Appendix, the estimates are quantitatively similar to using the baseline sample.

A related concern is that firms can decide to locate in prefectures with or without planned HSR stations, and hence firms from prefectures without any planned HSR station are not comparable to the rest. We show in Table 14, Panel A that the results barely change when we exclude firms from prefectures that did not have an HSR station by 2016.

Fifth, we conduct a permutation test to show that the p-values derived from the permutation tests are quantitatively similar to the baseline results. In particular, we randomly select a set of prefecture pairs for each year between 2008 and 2013. The number of connected prefecture pairs is set to match the actual number of connections in the corresponding year. We then use the counterfactually assigned HSR pair dummies to construct the first exporter integration measure and rerun the regression. The distribution of coefficient estimates based on 200 simulations is shown in Figure 3 (a), and none are greater than the true estimates. That corresponds to a p-value of 0. For the second measure, we reshuffle the reduction of commuting time between city pairs caused by the HSR network expansion and recompute the second measure. We rerun the regression using the counterfactual measure and plot the estimation results in Figure 3 (b). For 98% of the simulations, the true estimate is greater than the simulated result, which indicates a p-value of 0.02.

Finally, there might be concern that the growing internet penetration in China is confounded with expansion of HSR. Better internet facilitates the collaboration of teams over long distances and hence improves a firm's export performance. Table 13 in the Appendix allays this concern by showing that internet penetration is not positively correlated with HSR connection, and hence the baseline results cannot be explained by improved internet connectivity.

5 Geographic Integration and Export Performance

The empirical results from Section 4 show that the HSR infrastructure shock improves export performance, and the effects are stronger among exporters that have become more integrated with other exporters from the same sector. The results, while consistent with the model predictions, may be driven by a number of underlying mechanisms. The model in Section 2 proposes one mechanism: The improved geographic integration strengthens knowledge spillovers among exporters, which reduce informal frictions in accessing foreign markets. However, a limitation of this kind of "reduced-form" results is that, by design, it relies on exogenous variation in connectivity (or travel time) instead of access to export-specific knowledge. With this caveat in mind, we explore in more detail the economic mechanism(s) behind the empirical results in this section to better understand how exporters benefit from greater geographic integration.

5.1 Are the Spillovers Specific to Exporters?

Following an HSR connection, firms more integrated with other exporters are shown to increase their export performance along both intensive and extensive margins. The model attributes the improved performance to firms that enjoy better access to other exporters, or the presence of "export-specific spillovers". However, we need to separate export-specific spillovers from standard spillovers—from which firms benefit by being close to all firms, not just exporters. This is important because export revenue tends to be highly correlated with domestic revenue (Bernard and Jensen, 1999). Therefore, the better export performance after HSR expansion may be attributed to the possibility that firms are closer to all firms, including non-exporters, which improves their capacity to serve both domestic and foreign markets.

Our empirical strategy is to incorporate both channels directly in the regression by adding to the baseline regression a measure that accounts for a firm's integration with domestic production activities. If a firm's export decision is driven exclusively or primarily by spillovers not specific to the export market, then including the additional term would reduce the effects of x_{cst} considerably in the regression outputs.[23] The measure is constructed analogously to that for exporter integration,

[23] A more direct way to examine this is to study how improved exporter integration affects firms' performance in the domestic market. However, this is not feasible, because we do not have data on firms' domestic production.

by replacing export revenue with the same-sector value-added from all firms:

$$x_{cst}^{\text{General}} = \sum_{c' \neq c} \text{VA}_{st_0}(c') \tilde{f}(d_{cc't}),$$

where $\text{VA}_{st_0}(c')$ is the total value-added in sector s and city c' and $\tilde{f}(d_{cc't})$ is the geographic friction function defined in (10) and (11).

As shown in Table 4, the results after incorporating the additional control for integration with domestic production activities remain quantitatively similar to the baseline results, both for the intensive margin of export revenue and the extensive margin of number of export destinations. On the other hand, connecting to domestic firms does not seem to have a significant positive impact on a firm's export performance, as evidenced by the mostly insignificant estimates.

To provide additional evidence of export-specific spillovers, we next show that connected exporters have converged on the set of destinations to which they export, but not on the set of export products. Specifically, we examine, through pairwise analysis of the HSR connection between two cities, how connected firms' similarities in the set of export destinations and product varieties have evolved. We use angle distance to measure the similarity between connected firms:

$$\text{angle distance}_{zc't} = \arccos\left(\frac{\sum_d (\pi_{zdt} - \bar{\pi}_{zdt})(\pi_{c'dt} - \bar{\pi}_{c'dt})}{\sqrt{\sum_d (\pi_{zdt} - \bar{\pi}_{zdt})^2 \times (\pi_{c'dt} - \bar{\pi}_{c'dt})^2}}\right),$$

where $\pi_{zt} = (\pi_{z1t}, \ldots, \pi_{zDt})$ is a vector that summarizes the export share to set of destinations by firm z and time t, and $\bar{\pi}_{zdt}$ is the average export share to all destinations by firm z. Similarity in products is defined analogously by replacing π_{zdt} with π_{zkt}, where k denotes a product at HS8 level. We estimate the following equation:

$$\text{angle distance}_{zc't} = \alpha_{zc'} + \alpha_{zt} + \alpha_{c't} + \gamma \mathbb{I}(HSR_{zc't}) + \iota_{zc't}, \tag{13}$$

where $\alpha_{zc'}$, α_{zt}, and $\alpha_{c't}$ denote, respectively, the time-invariant firm-city pair fixed effects, firm-year fixed effects, and city-year fixed effect; $\mathbb{I}(HSR_{zc't})$ is a dummy that indicates a direct HSR connection between firm z and city c'; and γ is the coefficient of interest. The results are shown in Table 5. We find that firms export to a more similar set of destinations once they are connected by HSR; we do not find any significant impact on product varieties. This evidence shows that geographic integration affects firm export-related decisions and outcomes, which further suggests that the spillover effects identified in Section 4 are specific to exporters.

Lastly, we consider an alternative specification that controls for firm time-varying effects by examining the relative export growths within a firm across different export destinations. This test controls for any firm-level productivity improvement as a result of the exporter integration and isolate export-specific effects at the destination level. To carry out the test, we estimate the following regression:

$$y_{zdt} = \alpha_{zd} + \alpha_{zt} + \alpha_{dt} + \beta x_{cdst} + \iota_{zdt},$$

where y_{zdt} represents firm z's export to destination d in year t and x_{cdst} represents the integration with other exporting firms within the same sector s that export to the same destination d.[24] As shown in Table 6, there is evidence that a firm increases its export to destination d if it becomes more integrated with other exporters of the same sector and specialized in exporting to destination d.

5.2 Facilitating Export Knowledge Acquisition

The previous subsection shows that the geographic integration of exporters generates positive spillover effects that benefit exporters specifically. This subsection further explores the channel through which firms benefit from the improved exporter integration. In particular, we present evidence consistent with the underlying mechanism of the model: Better geographic integration of exporters reduces frictions in the export market by facilitating knowledge spillovers (including knowledge sharing and knowledge acquisition).

5.2.1 Heterogeneous Responses to Improved Exporter Integration

Providing direct evidence on knowledge acquisition and sharing is difficult, since it requires information about how firms interact. In general, this type of communication is either proprietary information or simply not available. Despite the data limitations, if this channel is present, we expect that firms or firm types that face greater knowledge acquisition costs will benefit more from the improved geographic integration with exporters.[25]

[24] To reduce sample size, we cluster countries into 12 country groups and consider each as one destination (Burstein et al., 2020).

[25] We extend the simple stylized model in Section 2 to incorporate heterogeneous responses to the HSR-driven geographic integration. The extended model, presented in Appendix A.3, predicts that firms that are less reliant on self-collected insight in the first place are more sensitive to the reduction in geographic frictions and improve their export performance to a larger extent. In particular, firms that face greater costs in acquiring own insight and less productive firms rely more on insights passed from other exporters, ceteris paribus.

To proxy for firms' knowledge acquisition costs in the export market, we consider a series of firm characteristics: export destinations, exporter locations, product types, exporter types, and firm sizes. We adopt the following regression specification to investigate the heterogeneous responses:

$$y_{zst} = \alpha_{st} + \alpha_z + \alpha_{ct} + \beta x_{cst} + \gamma x_{cst} \times \text{Firm Type} + \mathbb{I}(\text{Firm Type})_t + \iota_{zt}. \tag{14}$$

We pay particular attention to γ, which indicates the heterogeneous effect of better geographic integration of exporters. On top of the fixed effects included in the baseline specification, we further control for the time-varying effects of firm type, $\mathbb{I}(\text{Firm Type})_t$, to rule out concerns that the identified heterogeneous effect is driven by differential time trends between firm types.

Our first test is inspired by the results of Lovely et al. (2005), which suggest that the acquisition of knowledge about export and export destinations is easier when a firm exports to relatively open and easily accessed markets. We start by showing that firms that export to more integrated trading environments are less responsive to the integration of exporters. In particular, we compute the share of a firm's export revenue from integrated markets before the HSR opening to proxy for the firm's information acquisition cost.[26] As shown in Table 7, Columns 1 and 2, the increase in export revenue is smaller if a firm was more specialized in integrated markets before the HSR opening.

We next examine firms located in different locations. It tends to be easier for firms in centrally located regions to acquire export knowledge due to the presence of other exporters and foreign buyers (Startz, 2016).[27] As shown in Table 7, Columns 9 and 10, firms in core prefectures are less responsive to an increase in integration with other exporters, compared with firms in peripheral prefectures.[28]

Third, we explore variation in knowledge acquisition costs across product types; Rauch (1999) suggests that firms that export complex goods are subject to more information frictions. We use the Rauch index to distinguish firms that produce homogeneous products from firms that produce differentiated or complex products before the HSR connection.[29] Table 7, Columns 3 and 4 show

[26] Following Lovely et al. (2005), we classify a destination country as integrated if it has above-average trade per dollar of GDP after controlling for the size of the country and proximity to economic activity outside of the country. A firm's Destination-Integration Index is then defined as $\sum_{d \in D^0(z)} w_d^0(z) 1_{\text{Integrated},d}$, where $w_d^0(z)$ is firm z's export share to destination d before the HSR connection and $1_{\text{Integrated},d}$ turns to 1 if destination country d is an integrated economy.

[27] Firms in central regions also tend to be more productive than those in remote regions (Combes et al., 2012). We explore heterogeneity along the size dimension below.

[28] Following Yu et al. (2019), we define the following prefectures as the core prefectures and the rest as peripheral prefectures: Beijing, Shanghai, Tianjin, Chongqing, Shenzhen, Qingdao, Dalian, Suzhou, Xiamen, and Ningbo.

[29] For each firm, we calculate the percentage of exports that belong to the homogeneous products category under the Rauch classification and interact it with the integration measure.

that, as expected, firms that specialize in homogeneous products benefit less from greater integration of exporters.

We further compare export firms that specialize in processing trade with the rest—i.e., "ordinary" exporters. The former typically have prespecified production standards and fixed foreign buyers and hence face lower knowledge acquisition costs. As confirmed in Table 7, Columns 5 and 6, the higher fraction of export revenue came from processing with assembly before the HSR connection; the smaller export firm would benefit from improved geographic integration of exporters.

Finally, we turn to testing heterogeneity across firms of different sizes. Smaller firms tend to face greater knowledge acquisition costs, since they expect a smaller export market potential and find it less profitable to invest in and explore ways to penetrate the foreign market. These firms would wait for the most productive firms to explore unfamiliar contexts first and follow their successful strategies (Hausmann and Rodrik, 2003). Therefore, enhanced integration with other exporters should particularly benefit smaller firms. We define a firm as small if its export value was below the median before the HSR connection. As shown in Table 7, Columns 7 and 8, smaller firms are more responsive to better geographic integration of exporters.

Altogether, the above evidence shows that firm types that tend to face greater knowledge acquisition costs improve their export performance to a larger extent. This suggests that the HSR expansion reduces knowledge acquisition costs more for these firms, which is consistent with the presence of export-specific agglomerations that arise from knowledge spillovers.[30]

5.2.2 Controlling for Other Types of Agglomeration

Besides knowledge spillovers, Marshall (1920) emphasizes two other types of agglomeration externalities arising from geographic integration: integration with customers and suppliers, and labor market pooling. The HSR expansion can potentially facilitate a firm's connection with its upstream suppliers and improve the firm's export performance by reducing the cost for a firm to search and find a better supplier (Bernard et al., 2019). Similarly, even though HSR, as a passenger-only service, does not have any direct impact on a firm's domestic goods shipping costs, it may have indirect

[30]It is possible that by getting effectively closer to other exporters, firms also have opportunities to meet potential foreign buyers, especially in the export hubs where trade fairs often take place. This channel, for our purpose, is also part of export-specific spillovers. Yet it appears that better connection with foreign buyers may not be the main driver of the results empirically. In particular, we collect data on trade fairs by sector in China and separately estimate the effect of connecting to locations where trade fairs take place—which are also where the foreign buyers tend to show up—and the effect of connecting to the rest. As shown in Table 15 in the Appendix, both types of connections result in quantitatively similar results.

impacts on domestic trade costs.[31] This allows firms better access to international markets (and therefore their customers). For the labor market pooling channel, exporters may benefit from an HSR connection by allowing them to hire workers—particularly high-skilled workers—more easily. We next investigate the presence of these other channels.

To control for the channel of integration with suppliers and customers, we construct two measures that reflect how an HSR connection has affected exporters' effective distance to their suppliers and customers. We measure a firm's integration with its potential suppliers in a way that resembles the baseline measure for exporter integration in (9). In particular, it is defined as

$$x_{cst}^{\text{Supplier}} = \sum_{s'} \phi_{s'}^{s} \sum_{c' \neq c} R_{s't_0}(c') \tilde{f}(d_{cc't}),$$

where $\phi_{s'}^{s}$ is the input share from sector s' for a firm in sector s located in prefecture c and $R_{s't_0}(c')$ is sector s' total output (domestic and export) produced in prefecture c'. This measure is greater when HSR connects the firm to a destination prefecture with a higher presence of its supplier industry and when that supplier industry accounts for a higher input share in the firm's production. Next, we consider how HSR affects exporters' integration with their customers by examining changes in their access to ports, which are the gateway from which products are shipped to foreign countries:

$$x_{zt}^{\text{Port}} = \tilde{f}(d_{cc't}^{port(z)}),$$

where $d_{cc't}^{port(z)}$ is the effective distance between the firm and the city in which firm z's port is located.[32] The measure captures changes in a firm's access to its international markets as a result of the HSR connection. The results are shown in Table 8, which suggest that HSR-driven integration with suppliers and customers has positive effects on firms' export performance, mainly for the intensive margin. More importantly, the effects from better integration with exporters (i.e., x_{cst}) remain qualitatively and quantitatively consistent with baseline results without the additional controls.

We next investigate the labor market pooling channel. Marshall (1920) emphasizes that agglomeration can occur because workers are able to move across firms and industries. Ellison et al. (2010) highlight labor movements across firms and industries; however, can only occur if the industries use the same type of workers. We follow their approach and measure the extent to which

[31] Even though the HSR connection primarily affects passenger commuting time, it may reduce product shipping costs indirectly: Given the shift of passengers from regular trains to HSR, regular train racks have more capacity for product shipment.

[32] We use firms' choice of ports in 2007 to avoid the endogenous switching of ports due to the HSR connection.

industries across cities use similar types of labor based on the occupational employment patterns across industries and cities catalogued in the 2000 Population Census of China (IPUMS, 2020).[33] We then compute the labor correlation as a proxy that controls for access to workers in connected cities:

$$x_{zt}^{\text{Labor}} = \sum_{c' \neq c} \text{CorrL}_{zc'} \mathbb{I}(HSR_{cc't}),$$

where $\text{CorrL}_{zc'}$ is the labor correlation between firm z and city c'.[34] As shown in Table 9, while the labor market pooling effects are imprecisely estimated, the effects of exporter integration remain strong and similar to the baseline results.

Also the labor market pooling channel, we further consider the potential reallocation of high-skilled workers across prefectures as a result of HSR. Yu et al. (2019) show that HSR reinforced the concentration of skilled labor in larger markets. Consequently, in cities that experience an inflow of high-skilled workers, skill-intensive sectors are expected to have faster expansion and better export performance, and vice versa. This channel may affect baseline estimation if changes in the labor supply systematically correlate with changes in exporter integration. To investigate, we construct a control for the skill supply effect of HSR. In particular, we first define a dummy indicator for skilled-labor inflow by interacting a dummy variable that represents whether the prefecture is a core prefecture or not and a dummy variable that indicates the opening of an HSR station in that prefecture. This indicator captures the intuition whereby core cities are expected to have an inflow of high-skilled workers after an HSR opening. We then interact the indicator of skilled-labor inflow with the skill intensity of the sector of the export firm.[35] After adding this variable to the baseline regression specification, as shown in Table 10, the effects of exporter integration are still quantitatively close to the baseline results. Furthermore, the coefficients of the interaction term are also significantly positive for export revenue, which verifies the conjecture that skill-intensive firms benefit more from an inflow of high-skilled workers caused by HSR expansion.

To summarize, results from this subsection confirm the importance of geographic integration among export firms, which are partially driven by channels documented in previous literature;

[33] The Population Census provides industry-level employment in 63 occupations for all cities, based on which we define π_{ioc}^{L} as the fraction of industry i's employment in occupation o in city c.

[34] We specify the occupation vector of a firm z in city c, industry i as $[\pi_{ioc}^{L}]_{o \in O}$, which essentially summarizes the labor share by occupation for industry i in city c that firm z belongs to. We then compute destination cities' occupation vectors, $[\tilde{\pi}_{oc'}^{L}]_{o \in O}$, and finally calculate the correlation between firm i's and destination city c''s occupation share vector, $\text{CorrL}_{zc'}$.

[35] We define the skill intensity of a sector as the median percentage of workers with a college degree across all firms in that sector. We use the Chinese Annual Firm Survey (2004) to calculate this measure, given that only in this year were firms asked about the skill composition of their employees.

these include proximity to suppliers (Bernard et al., 2019) and the spatial reallocation of high-skilled workers (Yu et al., 2019). But more importantly, The results provide further evidence that improved export performance due to better exporter integration is not confounded by other types of agglomeration forces and can be attributed to an HSR connection enabling more efficient export knowledge spillovers.

5.3 Quality Improvement in Export Markets

In the final part of the analysis, we aim to shed further light on the nature of export knowledge being acquired by studying changes in product qualities across exporters following HSR-induced improvement in exporter integration.

We start by deriving the product qualities across exporters, following the approach of Khandelwal et al. (2013)—i.e., we obtain the quality of each firm-product-destination country-year observation using the residuals from the following OLS regression:

$$\ln q_{znkt} + \sigma_s \ln p_{znkt} = \alpha_{nkt} + \varepsilon_{znkt}, \tag{15}$$

where q_{znkt} and p_{znkt} are the quantities and prices of product variety k sold by firm z to destination country n and year t, respectively; α_{nkt} is fixed effects at product-destination-year level; and σ_s is the product category-specific elasticity of substitution across varieties indexed by z, the values of which are from Broda et al. (2006). The quality measure for each firm-product-country-year, denoted as $\hat{\lambda}_{znkt}$, can be inferred by $\ln \hat{\lambda}_{znkt} = \frac{\hat{\varepsilon}_{znkt}}{\sigma_s - 1}$.[36]

Next, we define quality at firm-product level as

$$\bar{\lambda}_{zkt} = \sum_n \theta_{znkt} \ln \hat{\lambda}_{znkt}, \tag{16}$$

where θ_{znkt} denotes the quantity share of product k to destination n within firm z. Finally, we compute the long difference of the quality measure for each firm-product pair:

$$\triangle \bar{\lambda}_{zkt} = \bar{\lambda}_{zkt} - \bar{\lambda}_{zkt-1}.$$

[36] Quality, denoted by λ, enters the utility function multiplicatively, with $U = \prod_{s=1}^{S} \left(\int_{\zeta \in \mathcal{Z}_{ns}} (\lambda_n(\zeta) q_n(\zeta))^{(\sigma-1)/\sigma} d\zeta \right)^{\sigma/(\sigma-1)}$, where $\sigma_s > 1$ is the product category-specific elasticity of substitution across varieties indexed by ζ. This implies that the demand is given by $q_n(\zeta) = \lambda_n(\zeta)^{\sigma-1} p_n(\zeta)^{-\sigma} P_{ns}^{\sigma-1} \alpha_s E_n$. Taking logs, the quality for each firm-product-country-year observation can be estimated as the residual from the OLS regression in Equation (15). Finally, σ_s is defined at HS3 level and the values are obtained from Broda et al. (2006).

We estimate changes in product qualities in response to improved exporter integration using a specification that follows directly from our baseline regression in (12):

$$\triangle\bar{\lambda}_{zk} = \beta\triangle x_{cs} + \alpha_c + \alpha_s + \triangle\iota_{zk}, \tag{17}$$

where $\triangle x_{cs}$ is the change in exporter integration measure for a firm in city c and sector s, and α_c and α_s are city and sector fixed effects, correspondingly. As shown in Table 11 Panel A, there is an improvement in overall firm-product qualities following HSR-induced exporter integration by 0.047 to 0.079 log points.

The results reflect the average change in product qualities for a given firm across all of its destinations. We further decompose the change in product quality into four components and quantify the relative importance of each component in overall quality changes. In particular, we can express changes in the quality of each firm-product pair as the following:

$$\begin{aligned}\triangle\bar{\lambda}_{zkt} =& \frac{1}{\bar{\lambda}_{zkt-1}}\left[\underbrace{\sum_{n\in\mathcal{N}_t\cap\mathcal{N}_{t-1}}\bar{\theta}_{znk}\left(\ln\lambda_{znkt}-\ln\lambda_{znkt-1}\right)}_{\text{Intensive Margin: Within}}+\underbrace{\sum_{n\in\mathcal{N}_t\cap\mathcal{N}_{t-1}}\left(\theta_{znkt}-\theta_{znkt-1}\right)\left(\bar{\lambda}_{znk}-\bar{\lambda}_{zk}\right)}_{\text{Intensive Margin: Across}}\right]+\\ & \underbrace{\frac{1}{\bar{\lambda}_{zkt-1}}\left[\sum_{n\in\mathcal{N}_t\setminus\mathcal{N}_{t-1}}\theta_{znkt}\left(\ln\lambda_{znkt}-\bar{\lambda}_{zk}\right)\right]}_{\text{Extensive Margin: Market Entry}}-\underbrace{\frac{1}{\bar{\lambda}_{zkt-1}}\left[\sum_{n\in\mathcal{N}_{t-1}\setminus\mathcal{N}_t}\theta_{znkt-1}\left(\ln\lambda_{znkt-1}-\bar{\lambda}_{zk}\right)\right]}_{\text{Extensive Margin: Market Exit}},\end{aligned} \tag{18}$$

where $\bar{\lambda}_{zk} = \frac{1}{2}\left(\bar{\lambda}_{zkt}+\bar{\lambda}_{zkt-1}\right)$ is the firm-product across-year mean quality; $\bar{\theta}_{znk}$ is the average export share to destination n in zk across years—i.e., $\bar{\theta}_{znk} = (\bar{\theta}_{znkt}+\bar{\theta}_{znkt-1})/2$,; and $\bar{\lambda}_{znk} = \frac{1}{2}\left(\bar{\lambda}_{znkt}+\bar{\lambda}_{znkt-1}\right)$ is the across-year average quality to destination n in firm-product zk.

Equation (18) naturally decomposes the overall quality change into four components. The first component measures the quality change to an existing set of destinations holding their export shares fixed. The second component accounts for changes in export shares to existing destinations, weighting those changes by the difference between a destination's average across-year quality and the overall average across-year quality within a firm-product. If qualities to existing destinations increase with further exporter integration, the within-component is positive; if qualities to existing destinations are relatively high and their export shares tend to increase, the across-component is also positive. The third component, the market entry margin, would positively respond to improving exporter integration if qualities of the new destinations entered are higher than the across-year average quality. The fourth component, the market exit margin, would negatively respond to improving exporter integration if firms exit destinations that have relatively low qualities compared with the across-year average.[37]

We regress each component against the change in exporter integration following regression specification (17). As shown in Table 11, Panel B, the intensive margin accounts for more than half (64% to 72%) of the relative improvement in export qualities. Moreover, most of the quality increase associated with the existing set of export destinations is due to the quality improvement holding export shares fixed (the "within" margin) rather than the increase in market share by relatively high-quality destinations (the "across" margin). Third, along the extensive margin, the entry of higher-quality destinations and exit of lower-quality destinations contribute more or less equally.

This evidence sheds more light on the nature of export-specific knowledge shared as firms become more geographically integrated because of HSR expansion. In particular, it suggests that firms acquire knowledge that improves the quality of products exported to existing destinations and helps them adjust the set of export destinations by entering higher-quality destinations and exiting lower-quality ones.

6 Conclusion

This paper analyzes how the geographic integration of exporters is related to firm export performance. Geographic integration plays an important role in export-specific knowledge exchange to reduce the informal trade costs in exporting. We develop a simple model, inspired by a Melitz-style framework that incorporates export-specific knowledge acquisition through stochastic interactions.Knowledge acquisition is more effective for firms located in close proximity to other firms that export in the same sector, which in turn drives down a firm's marginal production costs.

Using a comprehensive dataset of Chinese exporters, we leverage HSR planning and implementation as a unique setting to provide a rare and plausibly exogenous variation in the effective extent of geographic integration. By employing a triple difference empirical setup, we find strong and robust evidence for the existence of geographic spillovers in the export market. In particular, we find that a one-standard-deviation increase in exporter integration leads to an increase in export revenue

[37] Note that because the exit margin is subtracted from the previous two margins, negative values make a positive contribution to the overall quality change.

by 3%-4%. The effect on export revenue is driven exclusively by the increase in export volume, which suggests that export-market geographic spillovers reduce the variable cost in exporting—for example, by lowering informal trade barriers. Furthermore, we document that exporter integration also facilitates exporting on extensive margins, such as the number of destination countries and the menu of product categories.

Finally, we find compelling evidence that the improved export performance is driven by more effective export-specific knowledge spillovers induced by the HSR network. We show that connected firms converge on where they export—rather than what they export— which suggests that part of the benefits are specific to exporters. Further evidence reveals that firm types that tend to incur great knowledge acquisition costs—for example, smaller firms and firms located in peripheral cities or that export to less open destinations—have experienced larger improvement in export performance, which provides further support for the mechanism highlighted in the model. We also show that the knowledge acquired facilitates quality upgrading by both improving existing export product qualities and adapting to higher-quality export destinations. Our results provide useful guidance for policies that promote the integration of exporters and complement recent evidence on the impact of transportation infrastructure and informal trade barriers in the export market.

References

Ahlfeldt, Gabriel M and Arne Feddersen, "From periphery to core: Measuring agglomeration effects using high-speed rail," *Journal of Economic Geography*, 2018, *18* (2), 355–390.

Aitken, Brian, Gordon H Hanson, and Ann E Harrison, "Spillovers, foreign investment, and export behavior," *Journal of International Economics*, 1997, *43* (1-2), 103–132.

Allen, Treb, "Information frictions in trade," *Econometrica*, 2014, *82* (6), 2041–2083.

Autor, David H., David Dorn, and Gordon H. Hanson, "The China Syndrome: Local Labor Market Effects of Import Competition in the United States," *American Economic Review*, October 2013, *103* (6), 2121–68.

Bakker, Jan David, "Trade and Agglomeration: Theory and Evidence from France," Working Paper, UCL June 2021.

Baum-Snow, Nathaniel, Nicolas Gendron-Carrier, and Ronni Pavan, "Local Productivity Spillovers," 2020.

Berkes, Enrico, Ruben Gaetani, and Martí Mestieri, "Cities and Technological Waves," CEPR Discussion Papers 16794, C.E.P.R. Discussion Papers December 2021.

Bernard, Andrew B and J Bradford Jensen, "Exceptional exporter performance: Cause, effect, or both?," *Journal of International Economics*, 1999, *47* (1), 1–25.

_ **and** _ , "Why some firms export," *Review of Economics and Statistics*, 2004, *86* (2), 561–569.

Bernard, Andrew B., Andreas Moxnes, and Yukiko U. Saito, "Production Networks, Geography, and Firm Performance," *Journal of Political Economy*, 2019, *127* (2), 639–688.

_ **, Stephen J. Redding, and Peter K. Schott**, "MULTIPRODUCT FIRMS AND TRADE LIBERALIZATION," *The Quarterly Journal of Economics*, 2011, *126* (3), 1271–1318.

Broda, Christian, Joshua Greenfield, and David Weinstein, "From Groundnuts to Globalization: A Structural Estimate of Trade and Growth," Working Paper 12512, National Bureau of Economic Research September 2006.

Buera, Francisco J. and Ezra Oberfield, "The Global Diffusion of Ideas," *Econometrica*, January 2020, *88* (1), 83–114.

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

Burstein, Ariel, Gordon Hanson, Lin Tian, and Jonathan Vogel, "Tradability and the Labor-Market Impact of Immigration: Theory and Evidence From the United States," *Econometrica*, 2020, *88* (3), 1071–1112.

Caliendo, Lorenzo, Maximiliano Dvorkin, and Fernando Parro, "Trade and Labor Market Dynamics: General Equilibrium Analysis of the China Trade Shock," *Econometrica*, 2019, *87* (3), 741–835.

Charnoz, Pauline, Claire Lelarge, and Corentin Trevien, "Communication Costs and the Internal Organisation of Multi-plant Businesses: Evidence from the Impact of the French High-speed Rail," *Economic Journal*, 2018, *128* (610), 949–994.

Combes, Pierre-Philippe, Gilles Duranton, Laurent Gobillon, Diego Puga, and Sébastien Roux, "The Productivity Advantages of Large Cities: Distinguishing Agglomeration From Firm Selection," *Econometrica*, 2012, *80* (6), 2543–2594.

Coşar, A Kerem and Banu Demir, "Domestic road infrastructure and international trade: Evidence from Turkey," *Journal of Development Economics*, 2016, *118*, 232–244.

Cosar, A. Kerem and Pablo D. Fajgelbaum, "Internal Geography, International Trade, and Regional Specialization," *American Economic Journal: Microeconomics*, February 2016, *8* (1), 24–56.

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

Davis, Donald R and Jonathan I Dingel, "A spatial knowledge economy," *American Economic Review*, 2019, *109* (1), 153–70.

Donaldson, Dave and Richard Hornbeck, "Railroads and American Economic Growth: A Market Access Approach," *Quarterly Journal of Economics*, 2016, *131* (2), 799–858.

Dong, Xiaofang, Siqi Zheng, and Matthew E Kahn, "The role of transportation speed in facilitating high skilled teamwork across cities," *Journal of Urban Economics*, 2020, *115*, 103212.

Duranton, J and Diego Puga, "Micro-foundation of urban agglomeration economies in Henderson V. J. and Thisse JF. (eds) Handbook of Regional and Urban Economics, Vol. 4 Cities and Geography," 2003.

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

Fajgelbaum, Pablo and Stephen J. Redding, "Trade, Structural Transformation and Development: Evidence from Argentina 1869-1914," *Journal of Political Economy*, 2022, *130* (5), 1249–1318.

Fernandes, Ana P and Heiwai Tang, "Learning to export from neighbors," *Journal of International Economics*, 2014, *94* (1), 67–84.

Foster, Andrew D and Mark R Rosenzweig, "Learning by doing and learning from others: Human capital and technical change in agriculture," *Journal of Political Economy*, 1995, *103* (6), 1176–1209.

Gaubert, Cecile, "Firm Sorting and Agglomeration," *American Economic Review*, November 2018, *108* (11), 3117–53.

Giroud, Xavier, "Proximity and Investment: Evidence from Plant-Level Data," *Quarterly Journal of Economics*, 03 2013, *128* (2), 861–915.

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

Hayakawa, Kazunobu, Hans Koster, Takatoshi Tabuchi, and Jacques-François Thisse, "High-speed Rail and the Spatial Distribution of Economic Activity: Evidence from Japan's Shinkansen," Discussion papers 21003, Research Institute of Economy, Trade and Industry (RIETI) January 2021.

Heuermann, Daniel F and Johannes F Schmieder, "The Effect of Infrastructure on Worker Mobility: Evidence from High-Speed Rail Expansion in Germany," *Journal of Economic Geography*, March 2019, *19* (2), 335–372.

Huang, Jintong and Yves Zenou, "Key Sectors in Endogeneous Growth," 2020.

IPUMS, "Integrated Public Use Microdata Series, International: Version 7.3 [dataset]," Technical Report, Minnesota Population Center 2020.

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

Keller, Wolfgang and Stephen Ross Yeaple, "The gravity of knowledge," *American Economic Review*, 2013, *103* (4), 1414–44.

Khandelwal, Amit K., Peter K. Schott, and Shang-Jin Wei, "Trade Liberalization and Embedded Institutional Reform: Evidence from Chinese Exporters," *American Economic Review*, October 2013, *103* (6), 2169–95.

Koenig, Pamina, Florian Mayneris, and Sandra Poncet, "Local export spillovers in France," *European Economic Review*, 2010, *54* (4), 622–641.

Kovak, Brian K., "Regional Effects of Trade Reform: What Is the Correct Measure of Liberalization?," *American Economic Review*, August 2013, *103* (5), 1960–76.

Kuang, Chun, Zijie Liu, and Wenyu Zhu, "Need for speed: High-speed rail and firm performance," *Journal of Corporate Finance*, 2021, *66*, 101830.

Li, Yifan, Xuan Zhang, and Lex Berman, "China High Speed Railway Stations, v2," 2016.

Lin, Yatang, "Travel costs and urban specialization patterns: Evidence from China's high speed railway system," *Journal of Urban Economics*, 2017, *98*, 98–123.

Lovely, Mary E, Stuart S Rosenthal, and Shalini Sharma, "Information, agglomeration, and the headquarters of US exporters," *Regional Science and Urban Economics*, 2005, *35* (2), 167–191.

Marin, Alvaro Garcia, Andrei V. Potlogea, Nico Voigtländer, and Yang Yang, "Cities, Productivity, and Trade," NBER Working Papers 28309, National Bureau of Economic Research, Inc December 2020.

Marshall, Alfred, *Principles of economics: Unabridged eighth edition*, Cosimo, Inc., 1920.

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

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

Qin, Yu, "'No county left behind?' The distributional impact of high-speed rail upgrades in China," *Journal of Economic Geography*, 2017, *17* (3), 489–520.

Rauch, James E, "Networks versus markets in international trade," *Journal of International Economics*, 1999, *48* (1), 7–35.

Redding, Stephen J and Matthew A Turner, "Transportation costs and the spatial organization of economic activity," *Handbook of Regional and Urban Economics*, 2015, *5*, 1339–1398.

Startz, Meredith, "The value of face-to-face: Search and contracting problems in Nigerian trade," *Available at SSRN 3096685*, 2016.

Steinwender, Claudia, "Real effects of information frictions: When the states and the kingdom became united," *American Economic Review*, 2018, *108* (3), 657–96.

Tian, Lin, "Division of Labor and Productivity Advantage of Cities: Theory and Evidence from Brazil," Discussion Paper 16590, Centre for Economic Policy Research September 2021.

Yamawaki, Hideki, "Exports, and Foreign Distributional Activities: Evidence on Japanese Firms in the United States," *Review of Economics and Statistics*, 1991, *73* (2), 294–300.

Yu, Feng, Faqin Lin, Yihong Tang, and Changbiao Zhong, "High-speed railway to success? The effects of high-speed rail connection on regional economic development in China," *Journal of Regional Science*, 2019, *59* (4), 723–742.

Zheng, Siqi and Matthew E Kahn, "China's bullet trains facilitate market integration and mitigate the cost of megacity growth," *Proceedings of the National Academy of Sciences*, 2013, *110* (14), E1248–E1253.

Tables

Table 1: Impact of Exporter Integration on Firm Export Outcomes

	(1)	(2)	(3)	(4)
Dep. variables	log(export revenue)		No. exp destinations	
Ind. variables	Measure 1	Measure 2	Measure 1	Measure 2
x_{cst}	0.042***	0.030***	0.202***	0.105***
	(0.007)	(0.006)	(0.038)	(0.037)
Observations	2,032,156	2,032,156	2,032,156	2,032,156
R-squared	0.738	0.738	0.805	0.805

Note: This table shows the estimation results for (12) and shows that HSR-driven improvement in exporter integration increases a firm's export activities. Measure 1 and Measure 2 are defined in (10) and (11), respectively. Columns 1 and 2 show that conditional on exporting, a firm enjoyed higher export revenue after an improvement in exporter integration driven by expansion of the HSR network. Along the extensive margin, as shown in Columns 3 and 4, exporting firms entered more foreign markets. The outcome variable for the third and fourth columns counts the number of foreign countries a firm sells its products to. We control for firm fixed effects, prefecture time-varying effects and sector time-varying effects. Robust standard errors clustered by prefecture are in parentheses: *** $p<0.01$, ** $p<0.05$, and * $p<0.1$.

Table 2: Export Revenue Decomposition

	(1)	(2)	(3)	(4)
Dep. variables	log(price)		log(quantity)	
Ind variables	Measure 1	Measure 2	Measure 1	Measure 2
x_{cst}	-0.045***	-0.035***	0.087***	0.066***
	(0.007)	(0.009)	(0.011)	(0.010)
Observations	2,028,602	2,028,602	2,028,602	2,028,602
R-squared	0.852	0.852	0.798	0.798

Notes: We decompose export revenue into quantity and price, and use each as the outcome variable to estimate (12). Measure 1 and Measure 2 are defined in (10) and (11), respectively. Results suggest that the increase in export revenue observed in Table 1 is driven by an increase in quantity, while the unit price goes down. We control for firm fixed effects, prefecture time-varying effects and sector time-varying effects. Robust standard errors clustered by prefecture are in parentheses: *** p<0.01, ** p<0.05, and * p<0.1.

Table 3: Adjustment of Exported Products

	(1)	(2)	(3)	(4)
Dep. variables	$\frac{\text{N}_{it}^{\text{Added}}+\text{N}_{it}^{\text{Dropped}}}{\bar{\text{N}}_{i}^{\text{HS-8}}}$		$\ln \text{N}_{it}^{\text{HS-8}}$	
Ind. variables	Measure 1	Measure 2	Measure 1	Measure 2
x_{cst}	0.044***	0.026**	0.036***	0.026***
	(0.008)	(0.010)	(0.007)	(0.007)
Observations	1,638,608	1,638,608	2,032,156	2,032,156
R-squared	0.401	0.401	0.806	0.806

Notes: This table shows the estimation results for (12). Measure 1 and Measure 2 are defined in (10) and (11), respectively. This table shows that exporter integration causes a firm to update its product category (Column 1 and 2); it also raises the total number of products a firm exports (Column 3 and 4). We define each HS-8 category as one type of product. We then count the number of products a firm exports every year and the number of products a firm adds and drops compared with the previous export year. In all the regressions, we control for firm fixed effects, prefecture time-varying effects, and sector time-varying effects. Robust standard errors clustered by prefecture are in parentheses: *** p<0.01, ** p<0.05, and * p<0.1.

Table 4: Controlling for General Production

	(1)	(2)	(3)	(4)
Dep. variables	log(export revenue)		No. exp destinations	
Ind. variables	Measure 1	Measure 2	Measure 1	Measure 2
x_{cst}	0.036***	0.029***	0.171***	0.094**
	(0.009)	(0.010)	(0.049)	(0.043)
x_{cst}^{General}	-0.022*	-0.011	0.009	0.003
	(0.011)	(0.009)	(0.044)	(0.046)
Observations	2,018,616	2,018,616	2,018,616	2,018,616
R-squared	0.805	0.805	0.805	0.805

Notes: This table shows the estimation results for (12), with an additional control x_{cst}^{General} for integration with general production activities. Measure 1 and Measure 2 are defined in (10) and (11), respectively. This table presents evidence that the effects of exporter integration remain quantitatively similar to the baseline results after controlling for better integration with general production activities. In all regressions, we control for firm fixed effects, prefecture time-varying effects, and sector time-varying effects. Robust standard errors clustered by prefecture are in parentheses: *** p<0.01, ** p<0.05, and * p<0.1.

Table 5: Pairwise Analysis: Changes in Export Destinations and Product Varieties

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. variables	Dest. similarity			Product similarity		
$\mathbf{I}(HSR_{zc't})$	0.002*	0.003***	0.004***	-0.000	0.000	-0.000
	(0.001)	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)
Firm-Dest-City FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm-Year FE	No	No	Yes	No	No	Yes
Dest-City-Year FE	No	Yes	Yes	No	Yes	Yes
Year FE	Yes	No	No	Yes	No	No
R-squared	0.738	0.741	0.912	0.762	0.770	0.811

Notes: This table shows the estimation results for (13). Connected firms export to a more similar set of destinations but there is no impact on product similarity. The evidence is consistent with geographic spillovers that reduce frictions in market penetration. Robust standard errors clustered by prefecture are in parentheses: *** p<0.01, ** p<0.05, and * p<0.1.

Table 6: Firm-Destination Regression

	(1)	(2)	(3)	(4)
Dependent Variable	log(export revenue)		Export dummy	
Independent Variable	Measure 1	Measure 2	Measure 1	Measure 2
x_{csdt}	0.012**	0.003	0.001**	0.001
	(0.005)	(0.005)	(0.001)	(0.000)
Observations	6,397,648	6,397,648	22,849,152	22,849,152
R-squared	0.838	0.838	0.743	0.743

Notes: This table presents evidence about exporters' adjustments at the firm-destination level. We control for firm-destination fixed effects, firm time-varying effects, and destination time-varying effects. To reduce sample size, we cluster countries into 12 country groups and consider each as one destination. Robust standard errors clustered by prefecture are in parentheses: *** p<0.01, ** p<0.05, and * p<0.1.

Table 7: Heterogeneous Responses to Exporter Integration

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Dep. variables	log(export revenue)									
Ind. variables	Measure 1	Measure 2	Measure 1	Measure 2	Measure 1	Measure 2	Measure 1	Measure 2	Measure 1	Measure 2
x_{cst}	0.016*	0.011*	0.043***	0.032***	0.082***	0.064***	0.047***	0.033***	0.053***	0.041***
	(0.009)	(0.050)	(0.008)	(0.006)	(0.012)	(0.009)	(0.007)	(0.006)	(0.010)	(0.008)
$x_{cst} \times$Small	0.060**	0.090***								
	(0.024)	(0.027)								
$x_{cst} \times$Simple			-0.022	-0.029*						
			(0.020)	(0.015)						
$x_{cst} \times$Destination Integration					-0.068***	-0.055***				
					(0.013)	(0.012)				
$x_{cst} \times$Process							-0.399***	-0.288***		
							(0.116)	(0.100)		
$x_{cst} \times$Core Prefecture									-0.026*	-0.025***
									(0.014)	(0.010)
Observations	2,032,156	2,032,156	2,023,497	2,023,497	1,974,606	1,974,606	2,032,156	2,032,156	2,032,156	2,032,156
R-squared	0.740	0.740	0.737	0.737	0.735	0.735	0.739	0.739	0.738	0.738

Notes: This table shows estimation results for (14). Measure 1 and Measure 2 are defined in (10) and (11), respectively. In Column 1 and 2, we define a firm as small if its export value before the HSR connection was below the median. In Column 3 and 4, *Simple* is defined as the fraction of export revenue that comes from the homogeneous product category under Rauch classification before the HSR connection. Next, to define the degree of market integration of a firm's export market, we classify a destination country as integrated if it has above average trade per dollar of GDP after controlling for the size of the country and proximity to economic activity outside of the country (Lovely et al., 2005). A firm's Destination-Integration Index used in Column 5 and 6 is then defined as $\sum_{d \in D^0(z)} w_d^0(z) 1_{\text{Integrated},d}$, where $w_d^0(z)$ is firm z's export share to destination d before the HSR connection and $1_{\text{Integrated},d}$ turns to 1 if destination country d is an integrated economy. In Column 7 and 8, we define variable *Process* as the fraction of export revenue from processing with assembly before the HSR expansion. In Column 9 and 10, the following prefectures are assigned to the core city group while the rest are in the peripheral city group: Beijing, Shanghai, Tianjin, Chongqing, Shenzhen, Qingdao, Dalian, Suzhou, Xiamen, and Ningbo (Yu et al., 2019). Robust standard errors clustered by prefecture are in parentheses: *** p<0.01, ** p<0.05, and * p<0.1.

Table 8: Controlling for Integration with Suppliers and Customers

	(1)	(2)	(3)	(4)
Dep. variables	log(export revenue)		No. exp destinations	
Ind. variables	Measure 1	Measure 2	Measure 1	Measure 2
x_{cst}	0.036***	0.029***	0.171***	0.094**
	(0.009)	(0.010)	(0.049)	(0.043)
$x_{cst}^{\text{Supplier}}$	0.100***	0.059**	0.201*	0.116
	(0.023)	(0.029)	(0.116)	(0.109)
x_{zt}^{Port}	0.094***	0.029	0.315*	0.116
	(0.029)	(0.018)	(0.164)	(0.079)
Observations	2,004,824	2,004,824	2,004,824	2,004,824
R-squared	0.737	0.737	0.804	0.804

Notes: This table shows the estimation results for (12), with additional controls for integration with suppliers and access to port. Measure 1 and Measure 2 are defined in (10) and (11), respectively. This table presents evidence that the effects of exporter integration remain quantitatively similar to the baseline results after controlling for better integration with suppliers and customers. We control for firm fixed effects, sector time-varying effects, and prefecture time-varying effects. Robust standard errors clustered by prefecture are in parentheses: *** p<0.01, ** p<0.05, and * p<0.1.

Table 9: Controlling for Labor Market Pooling

	(1)	(2)	(3)	(4)
Dep. variables	log(export revenue)		No. exp destinations	
Ind. variables	Measure 1	Measure 2	Measure 1	Measure 2
x_{cst}	0.041***	0.029***	0.200***	0.103***
	(0.007)	(0.006)	(0.038)	(0.037)
x_{cst}^{Labor}	0.006	0.025*	-0.058	0.038
	(0.006)	(0.013)	(0.036)	(0.067)
Observations	2,016,197	2,009,924	2,016,197	2,009,924
R-squared	0.737	0.737	0.805	0.804

Notes: This table shows the estimation results for (12), with additional control for access to the potential labor pool. Measure 1 and Measure 2 are defined in (10) and (11), respectively. This table presents evidence that the effects of exporter integration remain quantitatively similar to the baseline results after controlling for the channel of labor market pooling. We control for firm fixed effects, sector time-varying effects, and prefecture time-varying effects. Robust standard errors clustered by prefecture are in parentheses: *** p<0.01, ** p<0.05, and * p<0.1.

Table 10: Controlling for Flow of Skilled Labor

	(1)	(2)	(3)	(4)
Dep. variables	log(export revenue)		No. exp destinations	
Ind. variables	Measure 1	Measure 2	Measure 1	Measure 2
x_{cst}	0.036***	0.024***	0.184***	0.086***
	(0.008)	(0.007)	(0.039)	(0.033)
Flow of skilled labor	0.038***	0.037***	0.119	0.124
	(0.014)	(0.014)	(0.090)	(0.089)
Observations	2,020,651	2,020,651	2,020,651	2,020,651
R-squared	0.737	0.737	0.805	0.805

Notes: This table shows the estimation results for (12), with an additional control for the potential flow of skilled labor. Measure 1 and Measure 2 are defined in (10) and (11), respectively. This table presents evidence that the effects of exporter integration remain quantitatively similar to the baseline results after controlling for the effects of HSR-induced reallocation of high-skilled workers. We control for firm fixed effects, sector time-varying effects, and prefecture time-varying effects. Robust standard errors clustered by prefecture are in parentheses: *** p<0.01, ** p<0.05, and * p<0.1.

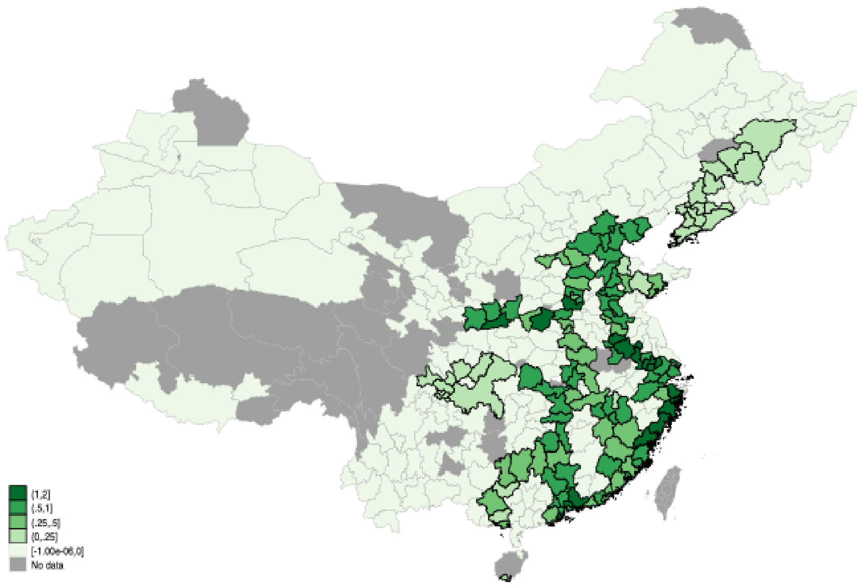
Table 11: Change in Quality and Decomposition of Margins

	(1)	(2)	(3)	(4)
Dep. variables	Product quality			
Ind. variables	Measure 1		Measure 2	
	Panel A. Overall Effect			
Overall	0.079***	0.058***	0.075***	0.047***
	Panel B. Effect Decomposition			
Intensive	0.051***	0.042***	0.048***	0.032***
Within	0.044***	0.037***	0.040***	0.026***
Across	0.007**	0.005	0.008**	0.007*
Extensive	0.027***	0.016*	0.027***	0.015***
Entry	0.013***	0.009*	0.014***	0.011***
Exit	-0.015***	-0.007*	-0.013***	-0.005
Sector FE	No	Yes	No	Yes
Observations	137,856	137,856	137,856	137,856

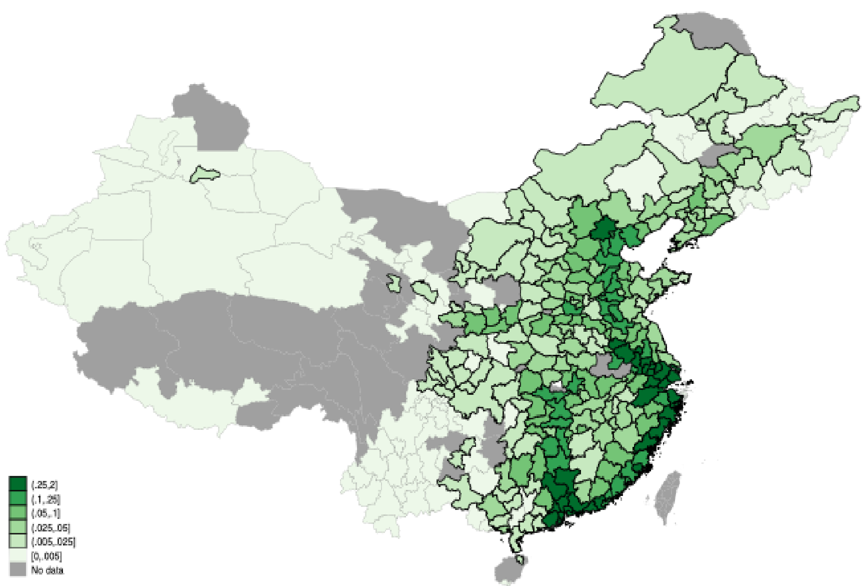
Notes: This table shows the estimation results for (17). Measure 1 and Measure 2 are defined in (10) and (11), respectively. The table suggests that firms acquire knowledge that improves qualities of products exported to existing destinations and adjust the set of export destinations by entering into higher-quality destinations and exiting lower-quality ones. Robust standard errors are clustered by prefectures: *** p<0.01, ** p<0.05, and * p<0.1.

Figures

Figure 1: Spatial Distribution of Changes in Exporter Integration



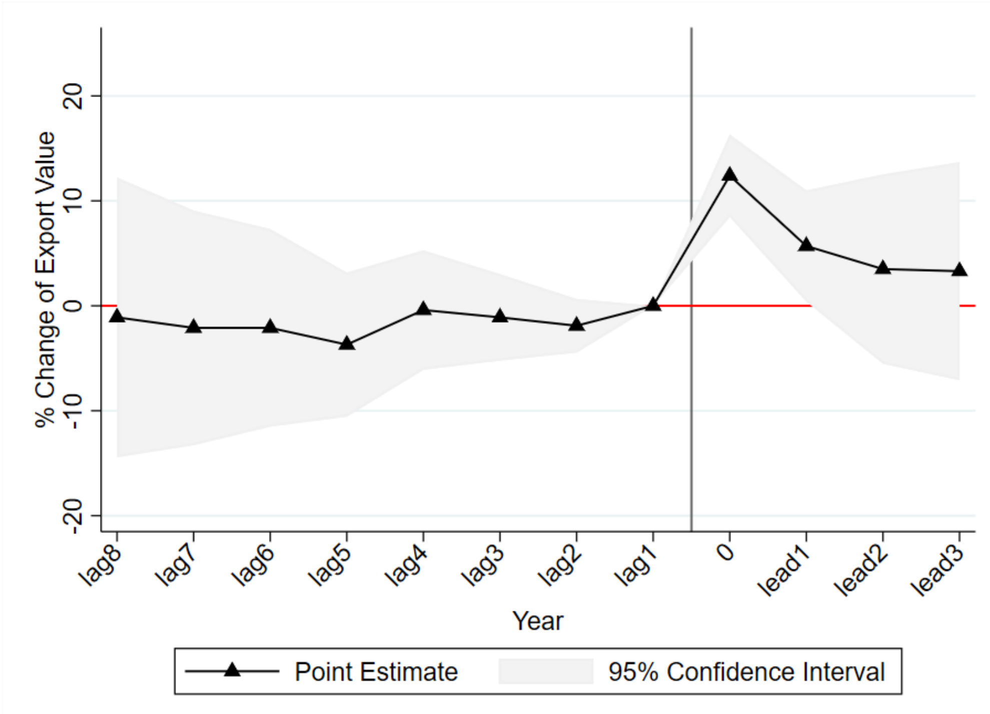
(a) Measure 1



(b) Measure 2

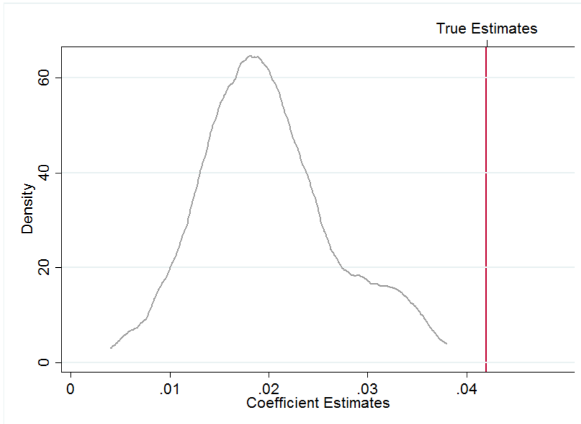
Notes: This figure plots HSR-driven improvement in exporter integration across prefectures in China. Measure 1 and Measure 2 are defined in (10) and (11), respectively. We calculate the increase in the measure of exporter integration from 2007 to 2013 across firms, take the average at prefecture level, and show the values on the map.

Figure 2: Event Study

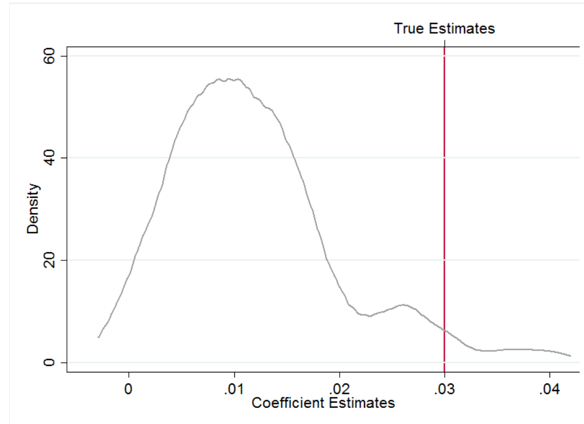


Notes: This figure shows that parallel trends hold for exporting firms with and without HSR stations before the HSR connection. After the HSR connection, the positive response in firm export is the largest in year 0 and declines over time. In particular, we estimate the following regression: $\ln r_{zct} = \sum_{k=\ldots,-3,-2,0,1,2,\ldots} \theta^k \mathbb{I}(HSR)^k_{ct} + \alpha_z + \gamma_t + \epsilon_{zct}$, where $\mathbb{I}(HSR)^k_{ct} = 1$ if prefecture c is connected to the HSR network in year $t - k$.

Figure 3: Permutation Test



(a) Measure1



(b) Measure2

Notes: This figure shows that the p-values derived from the permutation tests are quantitatively similar to the baseline results. Measure 1 and Measure 2 are defined in (10) and (11), respectively. For the first measure, we randomly select a set of prefecture pairs for each year during 2008 and 2013. The number of connected prefecture pairs is specified to be the same as in reality in the corresponding year. We then use the falsely assigned HSR pair dummies to construct the first exporter integration measure and rerun the regression. For the second measure, we reshuffle the reduction of commuting time between city pairs caused by the HSR network expansion and recompute the second measure. The distribution of coefficient estimates based on 200 simulations is shown in (a) and (b) correspondingly. The p-values of the estimates are 0 for Measure 1 and 0.02 for Measure 2, both of which are close to the baseline estimates.

Appendix

A Theoretical Appendix

In Appendix A, we provide details of the model and derivations of theoretical results presented in Section 2.

A.1 Model Details

A.1.1 Preferences

Recall that we consider an open economy with N countries, indexed by $n \in \mathcal{N}$. We omit the country index where possible to save on notation. In the home country, consumers have identical Cobb-Douglas preferences—that is, given income, a representative consumer decides how to allocate consumption over local final goods from all sectors, indexed by $s \in \{1, \dots, S\}$, with a Cobb–Douglas aggregator:

$$U = \prod_{s=1}^{S} Q_s^{\alpha_s}, \tag{19}$$

where $\alpha_s > 0$ is the expenditure share for each sector s, with $\sum_{s=1}^{S} \alpha_s = 1$. Within each sector, differentiated varieties are aggregated in a CES manner,

$$Q_s = \left(\sum_{z \in \mathcal{Z}_s} q(z)^{\frac{\sigma}{\sigma-1}} \right)^{\frac{\sigma-1}{\sigma}}, \tag{20}$$

where $\mathcal{Z}_s$ is the set of available varieties within sector s in the country, and $\sigma > 1$ is the elasticity of substitution across varieties. The corresponding sectoral price index in each country is given by

$$P_s = \left(\sum_{z \in \mathcal{Z}_s} p(z)^{1-\sigma} \right)^{\frac{1}{1-\sigma}}.$$

A.1.2 Firm Revenues

Given our assumptions on consumer preferences and market structure, the price is a constant markup over marginal cost. A firm with productivity φ in the domestic market charges

$$p_s^d(z) = \frac{\sigma}{\sigma - 1} \frac{1}{\varphi_z},$$

which only depends on the productivity draw.

To access the foreign markets, firms are subject to the standard fixed cost of export f_n^x and a bilateral iceberg trade cost τ_{sn}, as well as informal trade barriers, which is an inverse function of export-specific knowledge $\varphi_s^x(z)$. Therefore, the export price for firm z is given by

$$p_{sn}(z) = \frac{\sigma}{\sigma - 1} \frac{\tau_{sn}}{\varphi_z \varphi_s^x(z)}. \tag{21}$$

From here, we can derive the firm revenue in the domestic market:

$$r_s^d(z) = R_s^d \left(\frac{p_s^d(z)}{P_s^d} \right)^{1-\sigma} = \left(\frac{\sigma}{\sigma - 1} \right)^{1-\sigma} R_s^d \left(P_s^d \right)^{\sigma - 1} \left(\frac{1}{\varphi_z} \right)^{1-\sigma}, \tag{22}$$

where R_s^d and P_s^d are the expenditure and price index in sector s within the domestic market. Similarly, we obtain the export revenue in foreign market n:

$$\begin{aligned} r_{sn}^x(z) &= R_{sn} \left(\frac{p_{sn}(z)}{P_{sn}} \right)^{1-\sigma} \\ &= \left(\frac{\sigma}{\sigma - 1} \right)^{1-\sigma} R_{sn} P_{sn}^{\sigma-1} \left(\frac{\tau_{sn}}{\varphi_z \varphi_s^x(z)} \right)^{1-\sigma}, \end{aligned}$$

where R_s and P_s are the expenditure and price index in sector s in n. Summing up across all export destinations, we obtain (3).

A.1.3 Productivity Cutoffs

The entry costs—f^e to draw from the productivity distribution and f_n^x to export, which becomes sunk once the firms receives export-specific insight draws—and the revenue conditions in (3) and (22) give us the cutoff conditions for participating in the domestic and the export markets, as in Melitz (2003). In particular, the first cutoff defines the least productive firm that is active, which is the firm selling in the domestic market only and exactly offsetting the fixed costs of production with its operating profits: $\pi^d = \frac{r_s^d(\underline{\varphi}_s^d)}{\sigma} - f^d = 0$, where $\underline{\varphi}_s^d$ is the first cutoff. Firms with productivity lower than $\underline{\varphi}_s^d$ exit the market.

The second cutoff also depends on the firm's export-specific insights $\varphi_s^x(z)$. Expanding (5), we get

$$\mathbb{E}(\pi_s^x(z)) = \kappa_2 \left(\sum_{n \in \mathcal{N}_z} R_{sn} P_{sn}^{\sigma-1} \tau_{sn}^{1-\sigma} \right) \left[\varphi_z \left(\lambda_z^* + \sum_{z' \in \mathcal{Z}_s \setminus z} \lambda_{z'}^* \left[f(d_{zz'}) \right]^\theta \right)^{\frac{1}{\theta}} \right]^{\sigma-1} - c\lambda_z^* - \sum_{n \in \mathcal{N}_z} f_n^x, \quad (23)$$

where λ_z^* is a function of φ_z for all $z \in \mathcal{Z}_s$. In particular, $\lambda_s^*(z) \equiv \arg\max_{\lambda_z} \mathbb{E}(\pi_s^x(z))$, which is pinned down by the following first-order condition of the expected export profit of a firm with respect to λ_z:

$$\kappa_2 \frac{\sigma - 1}{\theta} \left(\sum_{n \in \mathcal{N}_z} R_{sn} P_{sn}^{\sigma-1} \tau_{sn}^{1-\sigma} \right) \varphi_z^{\sigma-1} \left(\lambda_z^* + \sum_{z' \in \mathcal{Z}_s \setminus z} \lambda_{z'} \left[f(d_{zz'}) \right]^\theta \right)^{\frac{\sigma-1}{\theta}-1} - c = 0. \quad (24)$$

Note that there is a unique solution if $\sigma < 1 + \theta$, which requires the expected export profit to be a concave function of λ_z—a condition we assume to hold throughout. The marginal exporting firm is just indifferent between exporting to a foreign market and not. Finally, the lowest productivity required in sector s to export to destination n for a given value of λ_z^* is pinned down by the following equation:

$$\kappa_2 R_{sn} P_{sn}^{\sigma-1} \tau_{sn}^{1-\sigma} \left[\lambda_z^* + \sum_{z' \in \mathcal{Z}_s \setminus z} \lambda_{z'} \left[f(d_{zz'}) \right]^\theta \right]^{\sigma-1} - f_n^x = 0, \quad (25)$$

a transformation of which leads to (6).

A.1.4 Aggregation

From (6), the share of firms that export to n as a function of λ_z is

$$Z_{sn}(k; \lambda_s(z)) = \sum_{y=k}^{\infty} p(y),$$

where recall that $p(\cdot)$ is the pmf of firm productivity draws and k is the minimal value of all productivity draws. Further, the total share of firms that export to n is

$$Z_{sn}(a_{sn}) = \sum_{j=a_{sn}}^{\infty} f_{sn}^\lambda(j) \sum_{y=\varphi(a_{sn})}^{\infty} f(y), \quad (26)$$

where $f_{sn}^\lambda(\cdot)$ is the empirical pmf of λ_z and a_{sn} is the productivity cutoff given by (6).

A.2 Expected Log Export Revenue

Taking logs of (3), we get

$$\log r_s^x(z) = const + \log\left(\sum_{n\in\mathcal{N}_z} R_{sn}P_{sn}^{\sigma-1}\tau_{sn}^{1-\sigma}\right) + (\sigma-1)\log\varphi_z + (\sigma-1)\log\varphi_s^x(z). \tag{27}$$

In (27), we define $K \equiv \log\varphi_s^x(z)$. Since $\varphi_s^x(z)$ follows a Fréchet distribution, K follows a log-Fréchet distribution with CDF

$$F(k) = e^{-\lambda_s(z)e^{-\theta k}}.$$

The moment-generating function for K is given by

$$M_K(\varphi) = e^{\frac{\varphi}{\theta}\log(\lambda_s(z))} + \Gamma(1-\frac{\varphi}{\theta}).$$

Hence, the expected value is given by

$$\begin{aligned}\mathbb{E}(K) &= M_K'(0)\\ &= \frac{\log\lambda_s(z)}{\theta} + \gamma,\end{aligned}$$

where γ is the Euler-Mascheroni constant. Finally, taking the expectation of (27), we obtain (7):

$$\begin{aligned}\mathbb{E}\left(\log r_s^x(z)\right) &= const + \log\left(\sum_{n\in\mathcal{N}_z} R_{sn}P_{sn}^{\sigma-1}\tau_{sn}^{1-\sigma}\right) + (\sigma-1)\log\varphi_z + (\sigma-1)\,\mathbb{E}(\log\varphi_s^x(z))\\ &= const + \log\left(\sum_{n\in\mathcal{N}_z} R_{sn}P_{sn}^{\sigma-1}\tau_{sn}^{1-\sigma}\right) + (\sigma-1)\log\varphi_z + \beta\log\tilde{\lambda}_s(z),\end{aligned}$$

where $\beta = (\sigma-1)/\theta$.

A.3 Model extensions

In this subsection, we present extensions to the baseline model in Section 2.

A.3.1 Multi-product firms

We now augment the baseline model to allow firms to supply multiple products, following Bernard et al. (2011). In addition to productivity φ and export-specific insight φ^x, there is an additional component of product characteristics, which is captured by χ. Accordingly, once the sunk entry cost has been incurred, firm z also observes its product attribute for product k, χ_k, which is drawn from a continuous distribution with CDF $G(\chi)$. We assume further that the productivity, export insight, and product attribute distributions are independent across firms and of one another, and that product attribute distributions are independent across products.

Upon observing φ_z and χ_k, the firm decides whether to enter and what products and countries to supply. In addition to the market-specific fixed cost of exporting, firms also face fixed costs of supplying each product to a foreign country—i.e. f_n^x. As more products are supplied to a market, total fixed costs rise but average fixed costs fall.

Since the demand structure stays the same, we can easily derive the equilibrium revenue received by a firm in its home country exporting a product to country n:

$$r_{sn}^x(z,k) = \left(\frac{\sigma}{\sigma-1}\right)^{1-\sigma} R_{sn} P_{sn}^{\sigma-1} \left(\frac{\tau_{sn}^k}{\varphi_z \varphi_s^x(z)\chi_k}\right)^{1-\sigma}, \tag{28}$$

which closely resembles the export revenue equation in the baseline model in (3), except for the new term χ_k. It is also obvious from the equation above that stronger geographic spillovers would increase the firm's export revenue to destination n even in the multi-product setting.

We now turn to the extensive margin. Since firms need to pay for a fixed cost for each product they export, there is a productivity cutoff for each product-destination given by

$$\varphi_{snk}(\tilde{\lambda}_s(z), \chi_k) = \frac{\tau_{sn}}{P_{sn}\left(\tilde{\lambda}_s(z)\right)^{\frac{1}{\theta}} \chi_k} \left[\frac{f_{nk}^x + c_s \lambda_z^*}{\kappa_2 R_{sn}}\right], \tag{29}$$

where $\varphi_s(\tilde{\lambda}_s(z), \chi_k)$ is the lowest productivity required in sector s to export for given values of $\tilde{\lambda}_s(z)$ and χ_k. Within a sector s, firms with $\varphi > \varphi_{sn}(\tilde{\lambda}_s(z), \chi_k)$ export product k to destination n. From here, it is obvious that

$$\frac{\partial \varphi_{snk}(\tilde{\lambda}_s(z), \chi_k)}{\partial \tilde{\lambda}_s(z)} < 0,$$

which implies, consistent with the baseline model, that a reduction in geographic frictions reduces productivity cutoffs and firms tend to export a greater number of products.

A.3.2 Incorporating heterogeneous firm types

In this subsection, we present an extended version of the model that incorporates different firm types to generate heterogeneous effects in response to an improvement in geographic integration.

This is achieved by assuming heterogeneity in the costs to build firm insight collection capacity. That is, to improve λ_z by one unit, a firm of type s incurs a unit cost c_s.[38]

With this extension, the total expected export profit is

$$\mathbb{E}(\pi_s^x(z)) = \kappa_2 \left(\sum_{n \in \mathcal{N}_z} R_{sn} P_{sn}^{\sigma-1} \tau_{sn}^{1-\sigma} \right) \left[\varphi_z \left(\tilde{\lambda}_s(z) \right)^{\frac{1}{\theta}} \right]^{\sigma-1} - c_s \lambda_z, \tag{30}$$

where the costs of building the insight collection capacity, c_s, differ across firm types. The optimal investment in λ_z, defined as $\lambda_s^*(z) \equiv \arg\max_{\lambda_z} \mathbb{E}(\pi_s^x(z))$, is again pinned down by the following first-order condition of the expected export profit of a firm with respect to λ_z:

$$\kappa_2 \frac{\sigma-1}{\theta} \left(\sum_{n \in \mathcal{N}_z} R_{sn} P_{sn}^{\sigma-1} \tau_{sn}^{1-\sigma} \right) \varphi_z^{\sigma-1} \left(\lambda_z^* + \sum_{z' \in \mathcal{Z}_s \backslash z} \lambda_{z'} \left[f(d_{zz'}) \right]^{\theta} \right)^{\frac{\sigma-1}{\theta}-1} - c_s = 0. \tag{31}$$

Note that from the optimality condition, it is also apparent that geographic spillovers reflect a notion of externalities: Firms choose their optimal level of capacity investment by internalizing their own benefits and costs and ignore the additional spillover effects of investment on other firms. From (31), we obtain the following result:

Lemma 3. *λ_z^* is an increasing function of φ_z and a decreasing function of c_s—i.e.,*

$$\frac{\partial \lambda_z^*}{\partial \varphi_z} > 0; \quad \frac{\partial \lambda_z^*}{\partial c_s} < 0.$$

Recall that λ_z^* is a firm's investment in its own export insight collection capacity. The first result in Lemma 3 is driven by the complementarity between a firm's own productivity and the export-specific insight in the firm's export profit function, as reflected in (5). Therefore, the marginal benefits of investing to build up insight collection capacity, λ_z, are greater for more productive firms. These firms will thus invest relatively more in their own capacity for acquiring export-specific insight. On the other hand, certain firm types would incur greater insight collection costs and would therefore invest less due to the higher marginal costs. In equilibrium, firms with greater φ and lower c_s would choose a greater λ_z^*, and hence tend to receive a great insight draw ω_z on average—and thereby rely more on self-collected insights and less on insight exchanges with other exporters.

[38] We can more generally think of this as also reflecting heterogeneities in both the costs and benefits of insight collection capacities.

With the type-specific heterogeneities, exporters may not benefit equally from the reduction in geographic frictions. Through the lens of the model, this is generated by differences in the optimal investment in insight collection capacity, λ_z^*, across firms. Intuitively, firms with stronger capacity tend to rely more on self-collected insight than geographic spillovers from other firms. Therefore, they would derive lower marginal benefits from the strengthened geographic spillovers that arise from the reduction in $d_{zz'}$. Recall from Lemma 3 that more productive firms (higher φ) and firms in sectors in which it is less costly to collect insights directly (lower c_s) choose higher λ_z^*. These are the firms that tend to benefit less from a reduction in $d_{zz'}$. Combining, we have the following results:

Proposition 4. *Assume that the export market's price index P_{sn} is fixed for all s and n. In response to an exogenous reduction in $d_{zz'}$, (i) within a sector, less productive exporters will increase their export revenue more; (ii) across sectors, exporters in sectors with greater c_s will increase their export revenue more. Formally, we have*

$$\frac{\partial^2\,\mathbb{E}(r_s^x(z))}{\partial d_{zz'}\partial\varphi_z} > 0; \quad \frac{\partial^2\,\mathbb{E}(r_s^x(z))}{\partial d_{zz'}\partial c_s} < 0.$$

Proposition 4 provides a further set of testable predictions in the setting of transportation infrastructure improvement, since firms may be differentially sensitive to reduction in $d_{zz'}$. The heterogeneity is determined by firm's own productivity. We would expect that within a sector, smaller firms would increase their export revenue relatively more. Exporters in different sectors will also be affected by reductions in $d_{zz'}$ differently. This is determined by c_s, the cost of acquiring export-specific insight by the firms themselves. For example, one may expect firms that produce more complex products to face greater information barriers in the trade market, thus making it more costly for firms themselves to collect relevant information in the foreign market. As a result, they may respond more to an infrastructure-induced increase in geographic spillovers than firms producing simpler products (Rauch, 1999). In Section 5, we test the predictions here using a series of firm characteristics associated with the costs of acquiring export-specific knowledge.

A.4 Proofs

In this section, we provide proofs of all theoretical results in Section 2 and Appendix A.

Proof for Proposition 1

Proof. Differentiating expected revenue in (4) with respect of $\mathbf{d_z} = (\{d_{zz'}\}_{z' \in \mathcal{Z}_s})'$, we get

$$\frac{\partial\, \mathbb{E}(r_s^x(z))}{\partial \mathbf{d_{zz'}}} = \frac{\partial\, \mathbb{E}(r_s^x(z))}{\partial \mathbf{f}(\mathbf{d_{zz'}})} \frac{\partial \mathbf{f}(\mathbf{d_{zz'}})}{\partial \mathbf{d_{zz'}}} < \mathbf{0}.$$

Following similar steps, we obtain

$$\frac{\partial\, \mathbb{E}(\pi_s^x(z))}{\partial \mathbf{d_{zz'}}} < \mathbf{0}.$$

□

Proof for Proposition 2

Proof. Differentiating (4) with respect to $d_{zz'}$, we obtain

$$\frac{\partial\, \mathbb{E}(r_{sc}^x(z))}{\partial d_{zz'}} = \frac{\partial\, \mathbb{E}(r_{sc}^x(z))}{\partial f(d_{zz'})} \frac{\partial f(d_{zz'})}{\partial d_{zz'}} < 0.$$

From (26), it is straightforward to see that $Z_s'(a) < 0$. Therefore, we have

$$\frac{\partial Z_s}{\partial d_{zz'}} < 0.$$

□

Proof for Lemma 3

Proof. From (24), we have

$$h(\lambda_z^*) = \kappa_2 \frac{\sigma - 1}{\theta} \left(\sum_{n \in \mathcal{N}_z} R_{sn} P_{sn}^{\sigma-1} \tau_{sn}^{1-\sigma} \right) \varphi_z^{\sigma-1} \left(\lambda_z^* + \sum_{z' \in \mathcal{Z}_s \setminus z} \lambda_{z'} \left[f(d_{zz'}) \right]^{\theta} \right)^{\frac{\sigma-1}{\theta} - 1} - c_s = 0.$$

Applying implicit function theorem, we obtain

$$\frac{\partial \lambda_z^*}{\partial \varphi_z} = -\frac{\partial h(\lambda_z^*)}{\partial \varphi_z} / \frac{\partial h(\lambda_z^*)}{\partial \lambda_z^*} > 0;$$

and

$$\frac{\partial \lambda_z^*}{\partial c_s} = -\frac{\partial h(\lambda_z^*)}{\partial c_s} / \frac{\partial h(\lambda_z^*)}{\partial \lambda_z^*} < 0.$$

□

Proof for Proposition 4

Proof. Using results from Proposition 2, we have

$$\frac{\partial^2 \mathbb{E}(r_s^x(z))}{\partial d_{zz'} \partial \lambda_z^*} > 0;$$

combining results from Lemma 3, we have

$$\frac{\partial^2 \mathbb{E}(r_s^x(z))}{\partial d_{zz'} \partial \varphi_z} = \frac{\partial^2 \mathbb{E}(r_s^x(z))}{\partial d_{zz'} \partial \lambda_z^*} \frac{\partial \lambda_z^*}{\partial \varphi_z} > 0;$$

and

$$\frac{\partial^2 \mathbb{E}(r_s^x(z))}{\partial d_{zz'} \partial c_s} = \frac{\partial^2 \mathbb{E}(r_s^x(z))}{\partial d_{zz'} \partial \lambda_z^*} \frac{\partial \lambda_z^*}{\partial c_s} < 0.$$

□

B HSR Background

The idea of building HSR can be traced back to 1978, when Deng Xiaoping visited Japan and was deeply impressed by Japan's high-speed rail, Shinkansen. The planning proposal was completed in 1990 and a feasibility study was conducted in 1994. The attitude toward building an HSR network in China was ambivalent at the time, as there were doubts about whether this very costly infrastructure would generate enough benefits to justify the upfront investment. The economic boom during the 2000s rekindled interest in building HSR in China, and the first official HSR line was opened in 2008, which runs between Beijing and Tianjin. Expansion of the HSR network during our study period is illustrated in Figure 4. The HSR network reached an important milestone in 2016 with completion of the so-called "four plus four network," which consists of four horizontal lines and four vertical lines. The Chinese government plans to extend the HSR network to eight horizontal lines and eight vertical lines and complete the construction by 2030. Figure 5 below further illustrates the history of HSR development in China.

Figure 4: Expansion of the HSR Network

(a) 2008

(b) 2009

(c) 2010

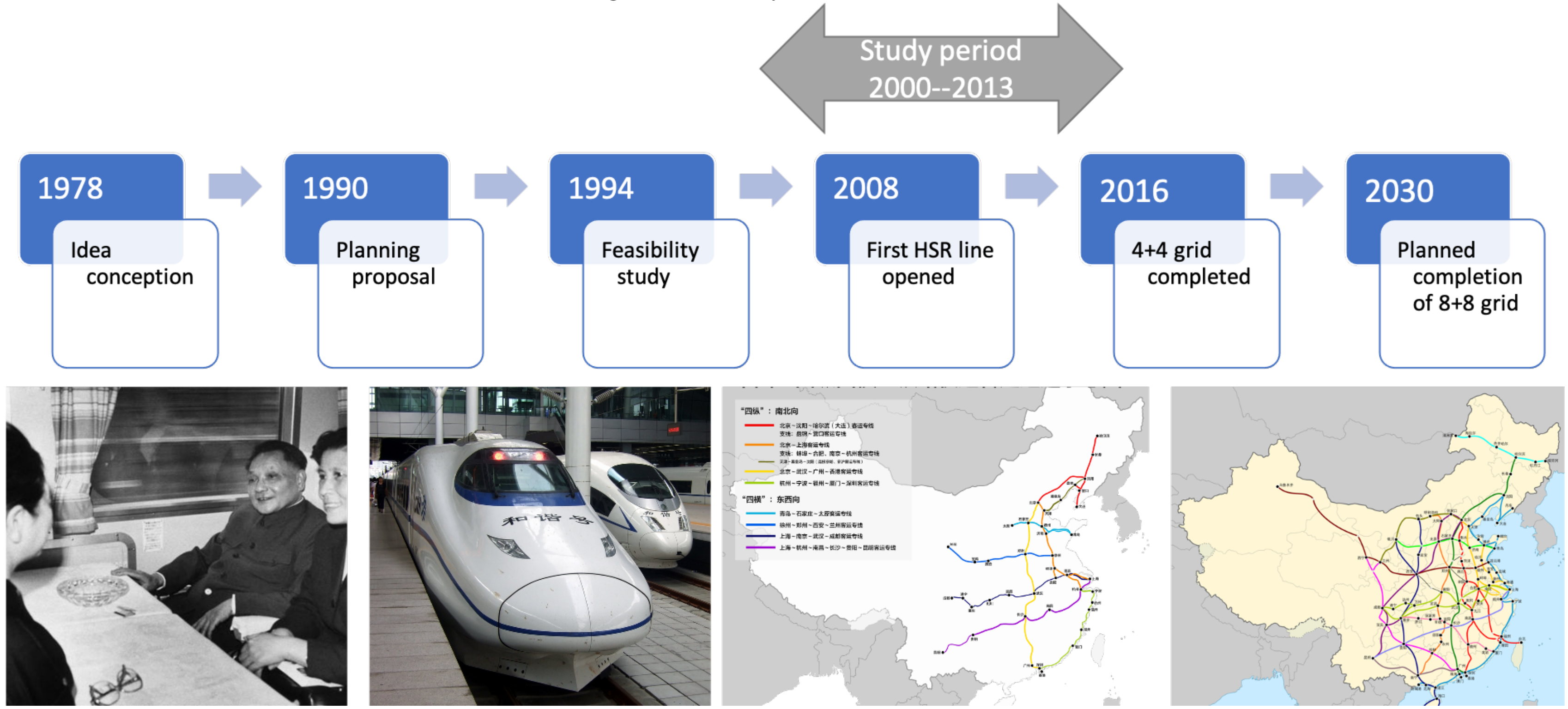
(d) 2011

(e) 2012

(f) 2013

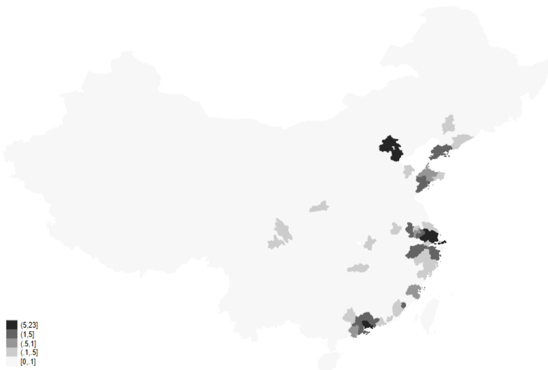
Notes: This figure shows the rapid expansion of HSR lines during 2008 and 2013. As a result, more and more city pairs were connected by the HSRs, which significantly reduced the commuting time between them.

Figure 5: History of HSR in China



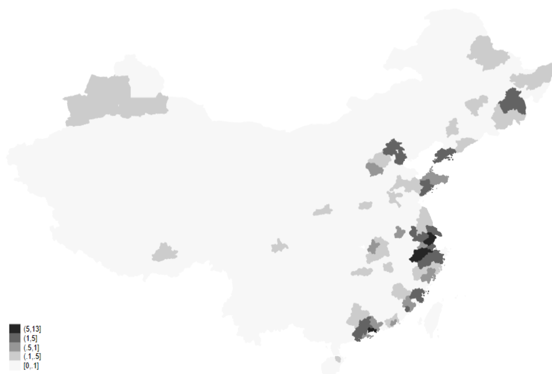
C Additional Figure and Tables

Figure 6: Spatial Distribution of Export Revenue Share



(a) Electrical equipment, 22.5%

(b) Machinery, 18.1%



(c) Apparel (not knitted), 4.9%

(d) Apparel (knitted), 4.7%

(e) Iron steel articles, 2.6%

(f) Vehicles, 2.2%

Notes: This figure shows that there were rich variations in the spatial concentration of export activities across the major export sectors in China before rollout of the HSR network. We aggregate export revenue from 2000 to 2007 by HS2 sector and prefecture. We display the spatial distribution of export revenue share for six major export sectors in China, which account for 55% of total export revenue during this period.

Table 12: Robustness Checks

	(1)	(2)	(3)	(4)
Dep. variables	log(export revenue)		No. exp destinations	
Ind. variables	Measure 1	Measure 2	Measure 1	Measure 2
	Panel A. Lagged treatment			
x_{cst}	0.044***	0.027***	0.288***	0.180**
	(0.013)	(0.008)	(0.101)	(0.074)
Observations	2,032,156	2,032,156	2,032,156	2,032,156
	Panel B. Excluding the largest sector			
x_{cst}	0.048***	0.030***	0.182***	0.145***
	(0.008)	(0.007)	(0.040)	(0.036)
Observations	1,740,255	1,740,255	1,740,255	1,740,255
	Panel C. Balanced panel			
x_{cst}	0.047***	0.040***	0.229***	0.170***
	(0.009)	(0.008)	(0.047)	(0.057)
Observations	975,743	975,743	975,743	975,743
	Panel D. No control for prefecture time-varying effects			
x_{cst}	0.026*	0.033**	0.052	0.102
	(0.013)	(0.014)	(0.063)	(0.065)
Observations	2,032,190	2,032,190	2,032,190	2,032,190

Notes: This table shows that the baseline results are robust to a series of alternative regression specifications. Measure 1 and Measure 2 are defined in (10) and (11), respectively. Panel A shows that results are quantitatively similar if we define treatment by HSR using HSR status from the previous year. Panel B shows that results barely change when dropping firms in the largest sector within a given prefecture. Panel C shows that the estimates become even greater if we restrict the sample to firms that exported in both 2007 and 2013. Panel D shows that the baseline effects shrink by one-third to one-half without controlling for prefecture time-varying effects. We always control for firm fixed effects, prefecture time-varying effects, and sector time-varying effects, except in Panel D. Robust standard errors clustered by prefecture are in parentheses: *** p<0.01, ** p<0.05, and * p<0.1.

Table 13: HSR and Internet Connectivity

	(1)	(2)
Dep. variables	Internet usage	log(Internet usage)
HSR connection	1.176	-0.041
	(4.849)	(0.036)
Observations	2,271	2,271
R-squared	0.437	0.993

Notes: This table shows that internet penetration is not positively correlated with HSR connection. Therefore, the results cannot be explained by improved internet service. We conduct the regression at prefecture-year level and use data from 2000 to 2013. The outcome variable is defined as the number of users who subscribed to internet service, and comes from the China City Statistics Yearbook. We control for prefecture fixed effects and year fixed effects, and robust standard errors clustered by prefecture are in parentheses: *** p<0.01, ** p<0.05, and * p<0.1.

Table 14: Additional Robustness Checks

	(1)	(2)	(3)	(4)
Dep. variables	log(export revenue)		No. exp destinations	
Ind. variables	Measure 1	Measure 2	Measure 1	Measure 2
Panel A. Dropping Prefectures with no HSR Station by 2016				
x_{cst}	0.046***	0.032***	0.210***	0.112***
	(0.008)	(0.006)	(0.040)	(0.039)
Observations	1,834,222	1,834,222	1,834,222	1,834,222
Panel B. Clustering Error Terms at Prefecture-sector-year Level				
x_{cst}	0.042***	0.030***	0.202***	0.105**
	(0.007)	(0.007)	(0.034)	(0.044)
Observations	2,032,156	2,032,156	2,032,156	2,032,156

Notes: This table shows that the results are robust to restricting to firms located in prefectures that had an open HSR station by 2016 (Panel A). Panel B shows further that the central results are robust to clustering the error term at the prefecture-sector-year level. Measure 1 and Measure 2 are defined in (10) and (11), respectively.

Table 15: Change in Quality and Decomposition of Margins

	(1)	(2)	(3)	(4)
Dep. variables	log(export revenue)			
Ind. variables	Measure 1	Measure 2	Measure 1	Measure 2
$x_{cst}^{\text{No Trade Fair}}$	0.052***	0.033***	0.065***	0.036***
	(0.020)	(0.010)	(0.012)	(0.011)
$x_{cst}^{\text{Trade Fair}}$	0.028	0.028**		
	(0.028)	(0.013)		
Observations	2,032,156	2,032,156	2,032,156	2,032,156
R-squared	0.738	0.738	0.738	0.738

Notes: Robust standard errors are clustered by prefecture: *** $p<0.01$, ** $p<0.05$, and * $p<0.1$.