High-Speed Railways and Domestic Value Added in Exports [*]

Hiau Looi Kee[†] Enze Xie[‡] Mingzhi Xu[§]

World Bank Zhejiang University, World Bank Peking University

September 25, 2024

Abstract

Since 2000, China has been upgrading its infrastructure, exemplified by the high-speed railways (HSR), while simultaneously moving up the global value chains, illustrated by the rising domestic value-added ratio (DVAR) in exports. This paper develops a novel methodology to estimate firm-level DVAR using only customs-transaction data to examine the impacts of HSR expansion on DVAR. Empirical results confirm that HSR increases DVAR through firm-level input-output linkages. Additionally, HSR is modeled to improve the connection between exporters and low-communication cost domestic suppliers, causing domestic material prices to decrease and varieties to increase, thereby substituting imported materials and raising the firm-level DVAR.

Keywords: Global value chains, domestic value-added, high-speed railways, transportation infrastructure, firm input-output linkages

JEL Classification Number: F10, H54, F14, R40, D21

[*]We thank Treb Allen, Shangjin Wei, He Wang, and seminar participants at Peking University and World Bank for valuable comments. Research for this paper has in part been supported by the World Bank's Multidonor Trust Fund for Trade and Development. Mingzhi Xu is thankful for financial support from the Natural Science Foundation of China (Grant No.72322007). The authors declare that they have no relevant or material financial interests that relate to the research described in this paper. The results and opinions presented in this paper do not represent the views of our institutions, the Executive Directors, or the countries they represent.

[†]Development Research Group, The World Bank; 1818 H ST NW, Washington, DC 20433. Tel: 1-202-473 4155; Fax: 1-202-522 1159; Email: hlkee@worldbank.org.

[‡]School of Economics, Zhejiang University, and Development Research Group, The World Bank. Email: enzexie@zju.edu.cn.

[§]Institute of New Structural Economics, Peking University. Email: mingzhixu@nsd.pku.edu.cn.

"I am taking the HSR to Shanghai for a meeting because of punctuality and it allows me to come back to Beijing on the same day. HSR improves the connection between businesses, allowing for more face to face contacts, particularly for trips up to 5 hours."

"I am traveling to Shanghai to meet my suppliers. I can introduce you to my suppliers if you need recommendations."

"I will let you know the details regarding the design after I have seen it in person, Director Pan. Don't worry, I will make sure the chips and electric board are good before agreeing to the contract."

— *three business passengers on board Beijing-Shanghai High-Speed Train G1, 06/19/2024*

1 Introduction

Since the early 2000s, China has been experiencing a dramatic upgrading of domestic infrastructure, exemplified by the construction of high-speed railways (HSR). Against this backdrop of massive infrastructure modernization, China is also deepening its engagement in global value chains (GVCs), manifested by its rising domestic value-added ratio (DVAR) in exports, driven by firms substituting domestic for imported materials (Koopman et al., 2014).[1] GVCs often require firm-specific investment and intense firm-to-firm communication (Antràs and Chor, 2022). By easing domestic travels and increasing the frequency and convenience of communication between exporting firms and domestic suppliers, could HSR provide the vehicle, pun intended, to promote the substitution of domestic for imported materials, thereby increasing firm DVAR and deepening China's engagement in GVCs?

This paper aims to explore the role of HSR in facilitating China's rising DVAR in exports through firm-level input-output linkages between domestic intermediate input suppliers and downstream final goods producers. The impacts of transportation infrastructure, such as roads and railways, on trade and development have been well studied. Faber (2014), Donaldson and Hornbeck (2016), and Donaldson (2018) investigate the economic impact of highways and railroads in different settings.[2] Some

[1] China's rising domestic contect in export has also been documented in Koopman et al. (2012) and Johnson and Noguera (2012). Kee and Tang (2016) further shows that the rising DVAR is primarily driven by the processing exporters substituting domestic for imported materials when a greater variety of domestic materials became available at lower prices, driven by input tariff reduction and foreign direct investment (FDI) liberalization during the early 2000s.

[2] See also Li and Li (2013), Baum-Snow et al. (2017), Chaudhary (2023) and Fan et al. (2023).

recent work also looks specifically at the effects of HSR on firm performance in China.[3] Most of these studies emphasize the general improvements in the exchange and acquisition of information, as well as the reduction in logistics and communication costs due to transport infrastructure upgrade, without explicitly considering input-output linkages bewteen firms. Focusing on buyer-seller relationships, Bernard et al. (2019) study the effects of HSR on Japanese firm performance through the lens of domestic outsourcing of tasks. However, the effect of modern transportation, such as HSR, on a firm's sourcing decision to choose between domestic and imported materials —one of the defining characteristics of firms engage in GVCs —has been largely overlooked. By highlighting the linkages between upstream and downstream firms along the production chain, this paper fills the gap by examining how HSR expansion may influence firms to choose domestic over imported materials, thereby driving up firm DVAR and broadening China's participation in GVCs.

This paper first proposes a new methodology to estimate firm-level DVAR from the export revenue function, drawing on insights from Olley and Pakes (1996) regarding the classic treatment of production function estimation. The key advantage of this export revenue function approach is that it only uses the customs transaction data and does not require matching with industry census data, thereby reducing data requirements compared to Kee and Tang (2016). Given that most countries have customs transaction data but lack matching industry census data, this export revenue function approach could be broadly applicable for future research and policy work on GVC measurements in both developed and developing countries. Results show that the China's DVAR has been rising over time, driven primarily by processing exporting firms. These findings echo the previous findings in Koopman et al. (2012) and Kee and Tang (2016), even with extended time coverage into the more recent years. Robustness checks confirm that the current estimated DVAR is not statistically different from the previous estimates in Kee and Tang (2016). The results remain consistent even after controlling for the presence of zeros in firm-level capital goods imports, which we use as a proxy for firm unobserved productivity in the export revenue function regressions, similar to how Olley and Pakes (1996) use firm investment as a proxy for firm productivity in the production function regressions.

Additionally, without relying on industry input-output tables, this paper proposes a novel approach to empirically capture the input-output linkages at the firm-level along the supply chains. This approach identifies the *potential domestic suppliers* for each exporting firm based on matching the product codes of the firms' imports and exports in the customs transaction data. For each exporting firm, the potential domestic suppliers are those firms that export products sharing the same HS 6-

[3]On topics such as profitability (Kuang et al., 2021), inventory spending (Cui and Li, 2019), affiliate location choice (Chen et al., 2023), investment efficiency (Wu et al., 2022), innovation (Zhang et al., 2020; Gao and Zheng, 2020), bank loan price (Geng et al., 2023) and financial asset holding (Liu et al., 2023). Beyond these firm-level analyses, Lin et al. (2023) and Wu and Yang (2023) find that connection to the HSR increases intercity investment and patent trade, respectively. Niu et al. (2020) show that the inter-regional trade increases and market segmentation decreases with HSR expansion using Chinese regional input-output tables.

digit codes with the imported materials of the exporting firm. Instead of using imported materials, exporting firm could have potentially sourced these materials domestically from firms that export the same products, thereby capturing the firm-level input-output linkages along the supply chains. Without using industry input-output tables or firm-level information about their actual suppliers, this new method of capturing firm-level input-output linkages allows for firm heterogeneity, as each exporting firm within an industry may have a different set of domestic suppliers. Moreover, it avoids selection bias since no actual transactions happen between the exporting firms and their potential suppliers.[4]

Regression results show that firm-level DVAR increases with the proportion of potential domestic suppliers connected to the exporting firms via HSR. These effects are more pronounced when HSR connects exporting firms and the domestic suppliers over medium to long distance, suggesting that HSR dominates other forms of transportation over this range for firm-to-firm interactions along the supply chain. The DVAR of foreign-invested firms, processing exporters, and firms in the electronics industry are shown to have the largest increases with the expansion of HSR. Additionally, HSR raises the DVAR of firms that import more homogeneous products, which are easier to substitute with domestic alternatives. Finally, DVAR increases among exporting firms locating in cities and regions connected by HSR.

To rationalizes the relationship between HSR and firm DVAR, this paper presents a simple two-sector model, with a monopolistically competitive final goods sector and a perfectly competitive domestic materials sector. HSR is shown to increase the probability of connecting exporting firms with domestic suppliers with low face-to-face communication costs, which leads to lower prices and more accessible varieties of domestic materials for the exporting firms, enabling them to substitute imported materials with domestic ones, thereby increasing their DVAR. These theoretical results lend supports to the previous empirical findings.

This paper speaks to several strands of literature. First, this paper is closely related to the literature on GVCs and measuring value-added trade. This literature starts with Hummels et al. (2001), who use industry input-output (IO) tables to calculate the value added-exports ratios for many countries. Recent related work includes Antràs et al. (2012), Johnson and Noguera (2012), Koopman et al. (2012), Koopman et al. (2014), Antràs (2012), and Johnson (2014). Methodologically, this paper is closer to Kee and Tang (2016), incorporating firm-level heterogeneity through the use of customs transaction data to minimize aggregation bias, thereby complementing the existing IO-based approach. Additionally, this paper lends support to the findings of Kee and Tang (2016), highlighting the importance of domestic transportation infrastructure in catalyzing the effects of trade and FDI liberalizations and pushing China up the GVC.

[4] Javorcik (2004) is the first to use input-output table to relate downstream FDI firms with upstream suppliers in order to study the vertical spillovers of FDI in Lithuania. Kee (2015) matches domestic firms with FDI firms based on the names of their common local input suppliers to study the shared-supplier spillovers of FDI in Bangladesh.

This paper further expands the growing literature on the impacts of modern transportation infrastructure projects, specifically high-speed rail, on economic development. HSR enhances the circulation of managers, which not only improves information transmission but also increases monitoring efficiency. As a result, firms are more inclined to outsource production tasks, whether to other firms (Bernard et al., 2019) or to their remote affiliates (Charnoz et al., 2018). Key studies on HSR in China include Lin (2017), Qin (2017), and Xu and Hu (2024). Each of these works highlights different aspects of HSR's influence: for instance, Lin (2017) and Qin (2017) examine how HSR connections boost income growth across Chinese cities, while Xu and Hu (2024) explores how strengthened firm-to-firm linkages, facilitated by faster transportation, lead to shifts in skill demand and inequality. Complementing these previous findings, we provide the first evidence of how the strengthened production network facilitated by HSR affects DVAR in Chinese exports.

The rest of the paper is organized as follows: Section 2 discusses the data used in this paper and the background of the HSR development in China. In Section 3, we propose a innovative method to estimate firm-level DVAR that only relies on customs transaction data. Section 4 shows the empirical specifications and the measurement of the variables of interests. Section 5 presents the results of the impacts of HSR expansion on firms' DVAR and some robustness checks. Section 6 presents the model and Section 7 concludes. The Appendix describes other datasets used in the paper and the construction of the main variables. Additional robust empirical results regarding the impact of HSR connection at the city-level on firms' DVAR are also presented there.

2 Data and Background

The data used in this project mainly consist of Chinese customs transaction data at the firm-HS 8-digit product level between 2000 and 2016, High-Speed Railway construction data from the China Railway Statistical Yearbook (2000-2016) [5], and prefecture-level socio-economic variables from the China City Statistics Yearbook (2000-2016).[6]

2.1 Chinese Customs Transaction Data

We use disaggregated product-level transaction data for each firm annually, sourced from China's General Administration of Customs. The data provide detailed information on firm exports and imports, including value, quantity, trade regime, destination country, customs name, shipment type, transportation mode, and more, for each transaction at the HS 8-digit product level.[7] The trade

[5]The geographical information of the administrative boundary and HSR network (station and line) are generously provided by Baum-Snow et al. (2017).

[6]In this paper, we use city and prefecture interchangeably.

[7]It is worth noting that, in custom data, there exists some trading companies that do not produce themselves (Ahn et al., 2011). Following Brandt et al. (2017), we delete these trading companies by identifying key words in their firm name. Moreover, we also remove imports and export transaction with China itself following Kee and Tang (2016).

regime information enables us to distinguish ordinary trade from processing trade. We identify the geographic location and the ownership type of firm through its administrative code assigned by the customs.[8]

For DVAR estimation, we estimate export revenue function by industry to obtain κ^M, the share of imports used in the production for exports. Section 3 presents the details. Industries are defined according to the classification proposed by the United Nations (UN), based on HS 2-digit code.[9] As firms may engage in trade across multiple industries, we focus on single industry firms to estimate κ^M and DVAR. A single-industry firm is defined as a firm that exports at least 90% of their products, in terms of export value, within the single industry group.

2.2 High Speed Railway Data

The prefecture-level High-Speed Railway data primarily comes from the China Railway Statistical Yearbook and are complemented and verified by manual search using HSR-relevant keywords such as 'Gaotie', 'Dongchezu', 'Hexiehao' and prefecture names. Our data cover 81 HSR lines between 2003 and 2016, connecting 185 prefectures. Figure 1 shows the number of prefecture connected per year and the cumulative number of prefecture with HSR connections.[10] The first pilot HSR line was constructed in 2003. After a four-year hiatus, HSR construction boomed, driven by the large-scale fiscal stimulus following the 2008 global financial crisis. By the end of sample period, 185 cities had been connected to the HSR network and China's HSR operational mileage exceeded 22,000 kilometers. In 2016, HSR transported 1.443 billion passengers, accounting for 52.1% of total national railway passenger traffic.

2.3 Background of the HSR Development in China

In the late 20th century, China invested heavily in transportation infrastructure, such as railways and expressways, to promote economic development. However, the previous transportation infrastructure construction mainly focused on freight transport, while passenger travel remained comparatively underdeveloped, suffering from low-quality service and constrained travel speed.

To mitigate the overloaded transportation network at the time and provide significantly improved passenger travel services, the Medium and Long Term Railway Plan (MLTRP) was approved in 2004, initially proposing the construction plan of the HSR network. HSR refers to the specially designed

[8]The first four digits of the identification code of each firm in the custom data contains information about the firms' location. We manually match the location information in the custom data with the administration code of Chinese prefecture so that we can merge custom information with the HSR network data.

[9]See https : //unstats.un.org/wiki/pages/viewpage.action?pageId = 79009148 for the UN industry classification. Appendix Table A1 shows the detail of the industry classification used in this paper.

[10]The patterns of our HSR information are comparable to those in Kuang et al. (2021) and Baum-Snow et al. (2017).

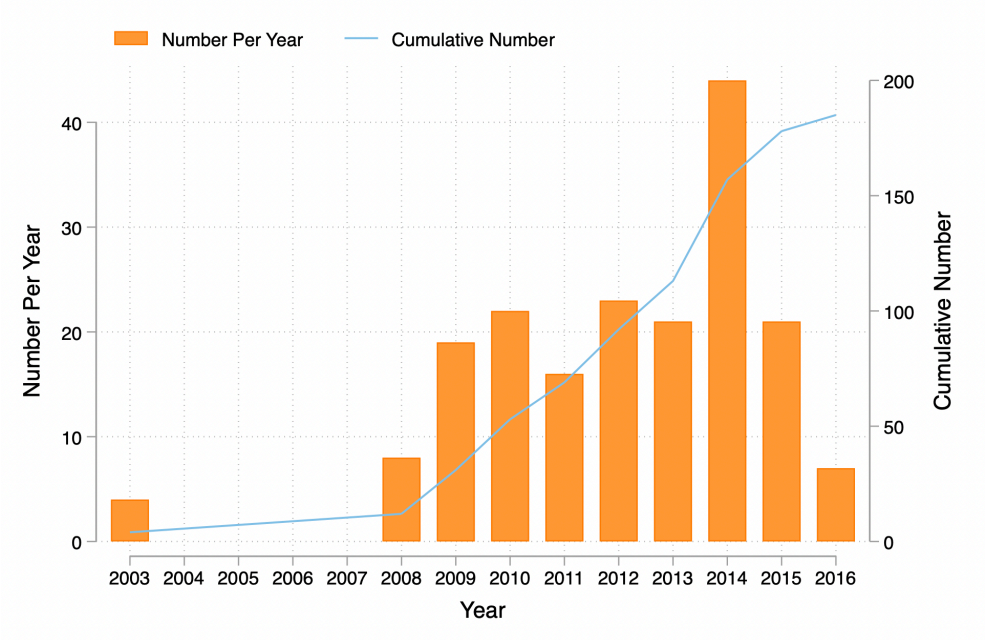


Figure 1: Number of prefecture with HSR Connection

and built railway lines on which trains can travel at speeds exceeding 250 km/h.[11] The MLTRP proposed four horizontal and four vertical form corridors, linking nearly all major cities in China. Accelerated by the fiscal stimulus aimed at helping the economy recover from the global financial crisis, the construction of the HSR speeded up since 2008. In 2016, the MLTRP was further revised to expand the network to eight horizontal and eight vertical corridors. After many years of development, now China has the world largest HSR network (Lawrence et al., 2019).

2.3.1 Comparative Analysis of HSR and Other Modes of Transportation

Compared to other modes of transportation, HSR has the following important and unique attributes: **(1) Passenger-dedicated:** HSR is exclusively used for passenger transportation instead of freight. As a result, HSR expansion reduces communication cost between regions rather than trade costs. **(2) Designed for middle-distance journeys:** As pointed out by Lawrence et al. (2019), HSR dominates trips with distance between 150 km and 800 km, and competes with air travel with distance between 800 km and 1200 km[12]. **(3) Frequent, convenient, punctual, safe and high-quality service:** Compared to airlines, HSR is less affected by weather conditions, offers simpler security checks and has shorter waiting times. HSR stations are located near the city center instead of far away from the city

[11] In comparison, the common speed limit of the national trunk highway is 120 km/h.

[12] The average travel distance in both the Beijing-Shanghai line and Beijing-Guangzhou line, the two most important HSR line in China, in 2013 is about 500km. Amos et al. (2010) suggests that HSR has comparative advantage over airlines for trips with distance less than about 800 km.

like airports. Meanwhile, passengers still have access to phone and internet on HSR, reducing the opportunity cost of travel, which is especially important for business travellers.[13] **(4) Mainly used for business purpose:** Survey evidence shows that nearly half of the HSR travel is for business, followed by leisure and commuting (Lawrence et al., 2019; Lin, 2017). **(5) Stable and relative low price:** HSR fares are relative low in China compared to other countries which increase its attractiveness to passengers across almost all income levels.

In short, the development of HSR has a tremendous impact on business travel and reshapes the interchange of information and business opportunity across regions in China, creating a better match of production and consumption, production and production.[14]

2.3.2 Network Effects of the HSR

Beyond the five attributes mentioned above, the most prominent feature of the HSR is its **network effects**. In addition to the direct HSR connection between cities, the indirect transfer connection through multiple HSR lines, or the combination of the HSR lines with other transportation modes, are even more important. For instance, the Zhengzhou-Xi'an HSR line, which was an isolated line until 2012, only transported passengers between Zhengzhou and Xi'an. However, after being connected to the Beijing-Guangzhou HSR line in 2013, both passenger volume and passenger-km skyrocketed, by 43 percent and 72 percent, respectively. By 2016, nearly 50 percent of the passengers travelling on the Zhengzhou-Xi'an line came from stations off the line itself (Lawrence et al., 2019).

Considering the network attributes, the impact of HSR opening in one city on firms' DVAR depends on how many prefectures connected directly and indirectly through the lines, as well as how many firms in those connected prefectures have potential input-output linkages with the initial firm.

Figure 2 shows the number of prefecture-pair connected through HSR, both directly and indirectly by year, during our sample period.[15] Consistent with Figure 1, HSR construction boomed after the financial crisis, and the number of prefecture-pair connected both directly and indirectly have steadily increased. However, the number of prefecture-pairs connected indirectly increased dramatically faster after 2010 due to the network effects. By the end of our sample period, the share of prefecture-pairs connected indirectly reach 36%, while the share of prefecture-pairs connected directly only account for 2.6%. In summary, Figure 2 suggests that the network effects that prefecture connected though

[13] Fang et al. (2023) point out that delay in Chinese airline industry is severe and thereby the punctuality and convenience of HSR poses a competitive shock to the airline industries.

[14] As summarized in Lawrence et al. (2019), several business interviews conducted in 2013 and 2015 in four cities newly connected to the HSR show that: HSR has limited impact in the location choice of manufacturing firms but significantly shapes the business travel. Sales, procurement and professional service staffs rely heavily on HSR for business travels.

[15] According to the Statistical Table of Administrative Divisions released by the Ministry of Civil Affairs of China in 2017, China has 334 prefecture-level administrative units. In our sample, we have 260 prefecture-level administrative units. As a result, the total number of prefecture-pairs is 33670 (or 33930 if the city itself is considered as a pair, such as Beijing-Beijing.)

line transfers are significant and tend to be larger in the late 2010s. As a result, in the subsequent analysis, if city A and City B can be connected either directly or indirectly, it is considered that an HSR connection exists between them.

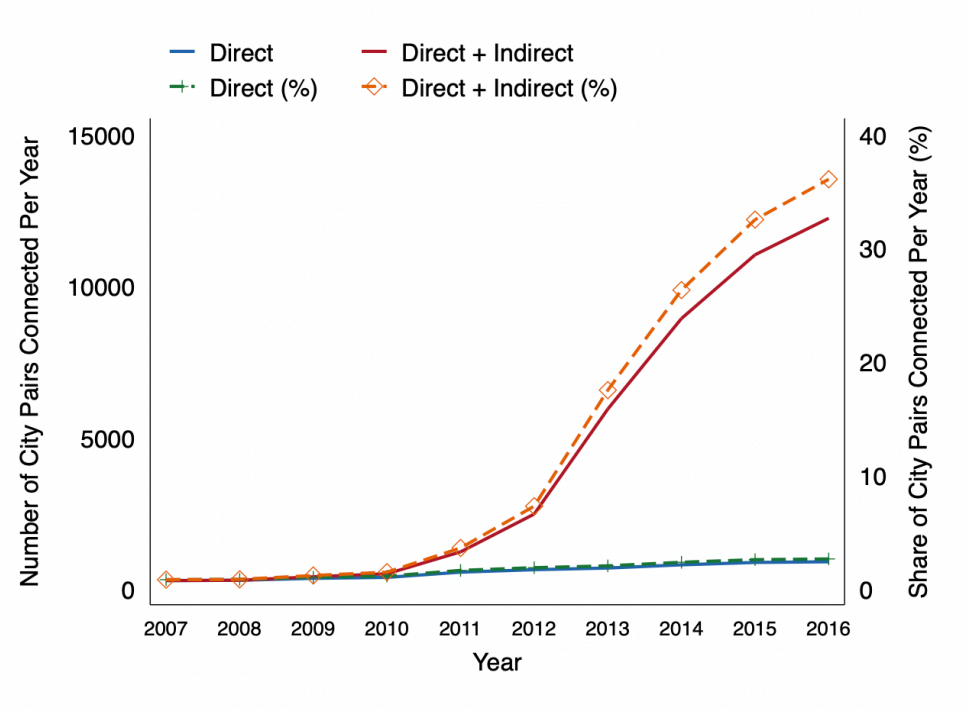


Figure 2: Number of Prefecture-Pairs with HSR Connection

3 Estimating DVAR Based on Export Revenue Function

Kee and Tang (2016) develops a methodology to construct DVAR in exports based on firm-level customs transaction data matching with the industrial census data. Domestic value added (DVA) of a firm is defined as the total value of domestic goods and services embedded in the firm's output and can be derived from the total revenue identity. For each year, t, a firm's (i) total revenue $(P_t Y_{it})$, by definition, consists of the following components: profits (π_{it}) , wages $(w_t L_{it})$, cost of capital $(r_t K_{it})$, cost of domestic materials $\left(P_t^D M_{it}^D\right)$, and cost of imported materials $\left(P_t^I M_{it}^I\right)$:

$$P_t Y_{it} \equiv \pi_{it} + w_t L_{it} + r_t K_{it} + P_t^D M_{it}^D + P_t^I M_{it}^I. \tag{1}$$

A firm's DVA thus equals the sum of its profits, wages, rental costs of capital, and both direct or indirect domestic materials purchased:

$$DVA_{it} \equiv \pi_{it} + w_t L_{it} + r_t K_{it} + P_t^D M_{it}^D. \tag{2}$$

For a processing firm that exports all its output and imports some of its intermediate inputs and capital equipment, its export (EXP_{it}) equals its revenue (P_tY_{it}) , while its import (IMP_{it}) equals the costs of imported materials, $P_t^I M_{it}^I$. Thus, equation (1) implies

$$\begin{aligned} EXP_{it} &\equiv DVA_{it} + IMP_{it} \Rightarrow \quad (3) \\ DVA_{it} &= (EXP_{it} - IMP_{it}) \text{ or} \\ DVAR_{it} &= 1 - \frac{IMP_{it}}{EXP_{it}}. \quad (4) \end{aligned}$$

Equation (3) shows that we may use $EXP_{it} - IMP_{it}$ to measure a processing firm's DVA.

For non-processing firms, Kee and Tang (2016) relies on firm level information on total materials purchase from the matching industry census to adjust for imports to obtain firm's $DVAR$. However, for most countries, while customs transaction data are available, it is often harder to obtain matching industrial census, and most of these firms do not engage in processing exports, which make the methodology of Kee and Tang (2016) not applicable.

In this section, we derive a methodology to estimate the share of imports that firms use to produce exports based on an accounting identity that relate export revenue to export expenditure and export profits.

For each firm i, total imports can be used for production for domestic sales or production for exports. Let's κ^M be the share of imports that is used as materials for the production for exports. Then from equation (3) , the total export revenue of firm i can be written as the following:

$$\begin{aligned} EXP_{it} &\equiv \kappa^M IMP_{it} + DVA_{it}, \text{ with } 0 \leq \kappa^M \leq 1 \text{ and} \\ DVA_{it} &= \pi_{it}^X + \kappa^D P_t^D M_{it}^D + \kappa^L w_t L_{it} + \kappa^K r_t K_{it} \quad (5) \end{aligned}$$

with κ^j, $j = \{D, L, K\}$ denotes the share of these respective inputs in the production of exports, and π_{it}^X is the export profits.

We do not observe $\kappa's$, and we only have data on EXP_i and IMP_i for each year. If DVA_{it} can be controlled for nonparametrically, by $D_{it}(\cdot)$, then we can turn equation (5) into a regression specification with a classical measurement error term, ε_{it}:

$$EXP_{it} = D_{it}(\cdot) + \kappa^M IMP_{it} + \varepsilon_{it}, \quad (6)$$

with the coefficient on imports, κ^M, has the interpretation of the share of imports that is used for the production of exports.

$$E\left(\widehat{\kappa^M}\right) = \kappa^M, \text{ if } E\left(IMP_{it}\varepsilon_{it}\right) = 0$$

With an unbiased estimate of κ^M, we will be able to adjust firm level imports in order to estimate

the firm level share of domestic value added in exports, $DVAR_{it}$:

$$\widehat{DVAR}_{it} = 1 - \frac{\widehat{\kappa^M} IMP_{it}}{EXP_{it}}. \tag{7}$$

Comparing equations (3) and (7), it is clear that the difference between the DVAR of processing firms and ordinary firms is due to their imports. Legally processing firms are expected to use all their imports for exports, which implies that the value of their imports should be less than the value of their exports and their κ^M should be one. Thus theoretically equation (7) is equivalent to Equation (3) for processing firms. However, empirically this may not be the case and should be verified. In our data set, about 17 percent of processing firms have imports larger than exports, spreading across all industries. For these processing firms, it seems like not all of their imports are used for their exports, which would imply they have κ^M less than one, and should be estimated based on equation (6) . The reasons behind these large imports could be due to mis-reporting, measurement errors, or leakage of imports into domestic markets, or trade between domestic processing firms.

3.1 Econometric Model

Because of the export expenditure accounting identity, the regression specification in equation (6) gives us a structural interpretation on the coefficient when we regress the value of firm's exports on the value of firm's imports. This is similar to the production function estimation, when we regression firm's output on their inputs, such as labor and materials. In this case, the estimated coefficients have the structural interpretation of input shares. Just as similar to the production function estimation, there may be factors that move both sides of the identity together which lead to simultaneity bias when we estimate κ^M based on least squares regressions. One such factor is firm productivity. An increase in firm's productivity may cause both imports and exports to increase at the same time, which causes the least squares estimate to bias upwards. On the other hand, currency depreciation may cause imports to decrease while export to increase which may lead to downward bias in the least squares estimate. The overall size and direction of bias depends on which force dominates, and can only be determined empirically.

Similar to production function estimation, we don't observe firm productivity, which can only be addressed with appropriate econometric techniques. Unlike firm productivity, We have empirical data on exchange rate which we can directly control for in the regression to reduce the downward bias.

One way to address the endogeneity of imports in the export revenue regression, is to use an instrumental variable (IV) that could be correlated with firms' imports but not necessary exports in order to satisfy the exclusion restriction. Suitable IV at firm level may not exist, here we instead follow the classic Olley and Pakes (1996) (hereafter referred to as OP) approach to use a proxy function to control for the unobserved firm productivity. Firms observe their productivity and make input

decisions, including imports, together with investment decision, including imported capital goods. We do not observe firm productivity, but we have data on firm imported capital goods, which we will use as a proxy for the unobserved productivity, similar to how OP uses firm investment to proxy for productivity in their production function estimation. We will also include the observed firm-level exchange rates in the control function following the method of Kee and Tang (2016).

3.1.1 Behavioral Framework

In this section we will briefly describe our estimation framework, similar to that of the OP approach. Specifically, we assume that firm's current profit depends on its own state variables, which consist of firm productivity, ω, and exchange rates, E, and current capital stock, K. Firm's current profit also depends on other state variables that are common across all firms, such as factor prices.

At the beginning of every period, an incumbent exporting firm observes the exchange rate of each source country, its productivity and capital stock, and makes two decisions simultaneously: (1) how much variable inputs, including imported materials, IMP to use in its production of exports, EXP; and (2) the optimal level of investment, including imported capital goods, I^M. Here we abstract from modeling firm's entry and exit decision of the export market, given that we cannot distinguish true exiting from missing values in the the custom transaction data. As in OP, we assume ω follows an exogenous Markov process, which leads to a Markov perfect Nash equilibrium in investment in capital good import. We then can use the observed capital good imports to control for the unobserved firm productivity, given exchange rates, capital stock, and other state variables that are common across all firms. Capital stock of each firm is constructed using perpetual inventory method with a depreciation rate of δ:

$$I_{it}^M = I(\varphi_{it}, E_{it}, K_{it}^M). \tag{8}$$

$$K_{it+1}^M = (1-\delta)K_{it}^M + I_{it}^M. \tag{9}$$

Please note that equation (9) describes the law of motion of firm level capital stock, generating through adding the current period investment of imported capital goods to the existing capital stock after depreciation. In other words, at the beginning of each period, after observing firm productivity, exchange rates and the existing capital stock, firms make investment decision to import capital goods, which will add to the existing capital stock net of depreciation, to form this new level of capital stock for the next period.

3.1.2 Estimation

The estimation of the export expenditure accounting identity based on the regression specification in equation (10) below gives us a structural interpretation on the coefficient of imports, which is the

share of imported input used in the production of exports, κ^M:

$$EXP_{it} = \alpha_i + \alpha_t + D(E_{it}, K_{it}) + \kappa^M IMP_{it} + \varphi_{it} + \epsilon_{it}, \tag{10}$$

where the dependent variable, EXP_{it} is the value of export of firm i in year t, and the variable of interest is IMP_{it}, which is the value of import. We assume that domestic value added of firm i in year t, can be decomposed into the following terms: $D(E_{it}, K_{it})$ is a function of firm exchange rate, E_{it} and capital stock, K_{it}, α_t denotes year fixed effects, which captures state variables that are common across all firms, such as wages and rental prices of capital, α_i denotes firm fixed effects, which captures state variables that are specific to each firm, such as foreign ownership and processing trade status, φ_{it} is the unobserved firm productivity and ϵ_{it} is either measurement error or unforeseeable productivity shock. While both φ_{it} and ϵ_{it} are unknown to the researchers, firm i observes φ_{it} and makes import and investment decisions based on it.

Similar to the production function estimation, estimating the export revenue function without controlling for the unobserved firm productivity, φ_{it}, will cause the least squares estimate of κ^M to have an upward bias, since a higher φ_{it} will increase both EXP_{it} and IMP_{it} simultaneously. Given that a higher φ_{it} also leads to a higher investment, provided that $I^M > 0$, we can invert equation (8) to form a control function for φ_{it} , similar to the OP approach:

$$\varphi_{it} = I^{-1}(I^M_{it}, E_{it}, K^M_{it}). \tag{11}$$

Here we are making the assumption that, given (K^M_{it}, E_{it}), the import of capital goods, which is an investment, is increasing in productivity. This assumption allows the investment function (8) to be invertible. This is similar to the original OP approach. Substituting equation (11) into equation (10) gives us the following:

$$EXP_{it} = \phi(I^M_{it}, E_{it}, K^M_{it}) + \kappa^M IMP_{it} + \epsilon_{it}, \tag{12}$$

with

$$\phi(I^M_{it}, E_{it}, K^M_{it}) = \alpha_i + \alpha_t + D(E_{it}, K^M_{it}) + I^{-1}(I^M_{it}, E_{it}, K^M_{it}), \tag{13}$$

where $\phi(I^M_{it}, E_{it}, K^M_{it})$ can be proxied semiparametrically using a higher order polynomial function in the regressions. While we will not be able to separately identify $D(E_{it}, K_{it})$ and ω_{it}, least squares estimation of equation (12) will give us a consistent estimate of κ^M.

Empirically, equation (12) is estimated for processing firms and non-processing firms, respectively, based on Chinese customs transaction data from 2000 to 2016. We expect the coefficient on imports for processing firms to be not significantly different from one, while the coefficient for non-processing firms to be significantly less than one.

3.2 Results on DVAR Estimations

Equation (12) forms the basis of our estimations at industry level, where we allow κ^M to differ across processing and non-processing firms within the same industry. We will test that κ^M of processing firms is significantly higher than that of non-processing firms and at the same time κ^M of processing firms is not different from one.

Table 1 presents the pool regression results based on equation (12). Column (1) presents the least squares estimation without any control other than firm and year fixed effects. The estimated κ^M is 0.76. Columns (2) and (3) show the regression results separately for the ordinary exporters and the processing exporters. The estimated import share for the ordinary exporters is 0.35, while for the processing exporters is 1.11, which is significantly higher. Columns (4) and (5) include the 3rd order polynomial function of investment, capital and exchange rates in the regressions. The estimated import shares are now lower, indicating that the polynomial function is effective in controlling for the unobserved firm productivity. Not only is the estimated import share of processing firms significantly higher than that of ordinary firm, it is also not different from one, which is in line with our expectations.

Table 1: Baseline Results: Share of Import Used for the Production of Exports

	(1)	(2)	(3)	(4)	(5)
Industries	All	Ordinary	Processing	Ordinary	Processing
Dependent Variable	EXP	EXP	EXP	EXP	EXP
IMP	0.756***	0.352***	1.112***	0.327***	1.014***
	(0.030)	(0.023)	(0.031)	(0.031)	(0.039)
ϕ(I, E, K)	No	No	No	Yes	Yes
Firm Fixed Effects	Yes	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes
Observations	419,030	185,493	226,042	68,250	36,691
Adjusted R^2	0.726	0.720	0.779	0.737	0.801

Notes: Robust standard errors clustered at the industry-year level are in parentheses, ***, **, * denote significance at 1%, 5%, 10% level.

Table 2 presents the regression results at the industry level with the polynomial control function. The ordinary and processing firms are together in these regressions and an interaction term between the processing firm dummy variable and imports is used to obtain the separate estimates for import shares of ordinary firms and processing firms. The result shows that processing firms in all industries have significantly higher import shares that is not different from one, while the non-processing firms import share falls in the range between 0.11 and 0.67, with the simple average only around 0.32. For all industries, we cannot reject the hypothesis that processing firms use all of their imported materials

for exports, while ordinary exporters use only part of the imported materials for their exports. Based on these results, we can set the import shares of processing firms to one when we construct their DVAR. The DVAR of ordindary firms will be constructed using the estimated import shares from this table according to the industries of the firms.

Table 2: By Industry: Share of Import Used for the Production of Exports

Industries	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Chemicals, Machinery and Transportation		Electronics, Metals and Optical Equipment		Furniture and Wood		Miscellaneous		Plastic, Rubber and Toys		Textiles and Apparels	
Dependent Variable	EXP	EXP	EXP	EXP	EXP	EXP	EXP	EXP	EXP	EXP	EXP	EXP
IMP	0.298***	0.199***	0.462***	0.373***	0.425***	0.403***	0.272***	0.114**	0.188***	0.165***	0.773***	0.673***
	(0.019)	(0.029)	(0.026)	(0.032)	(0.046)	(0.070)	(0.037)	(0.050)	(0.018)	(0.033)	(0.061)	(0.058)
Processing × IMP	0.664***	0.768***	0.578***	0.691***	0.565***	0.718***	0.929***	1.021***	0.892***	0.800***	0.542***	0.462***
	(0.027)	(0.045)	(0.038)	(0.062)	(0.071)	(0.123)	(0.054)	(0.145)	(0.041)	(0.109)	(0.070)	(0.119)
ϕ(I, E, K)	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Firm Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	91,116	37,901	124,516	48,633	27,827	3,526	49,708	4,820	43,356	6,279	77,819	3,178
Adjusted R^2	0.716	0.739	0.739	0.781	0.723	0.781	0.760	0.780	0.800	0.812	0.741	0.773

Notes: Robust standard errors clustered at the industry-year level are in parentheses, ***, **, * denote significance at 1%, 5%, 10% level.

Figure 3 presents the estimated DVAR when we aggregate up firm level estimates to the aggregate level. It is clear that from 2000 to 2016, China has been experiencing a rising DVAR in the exports, driven mainly by the processing exports. While the DVAR of ordinary exports has hovering around 0.92, the DVAR of processing exports has been rising through out the sample period, from 0.48 to 0.78.

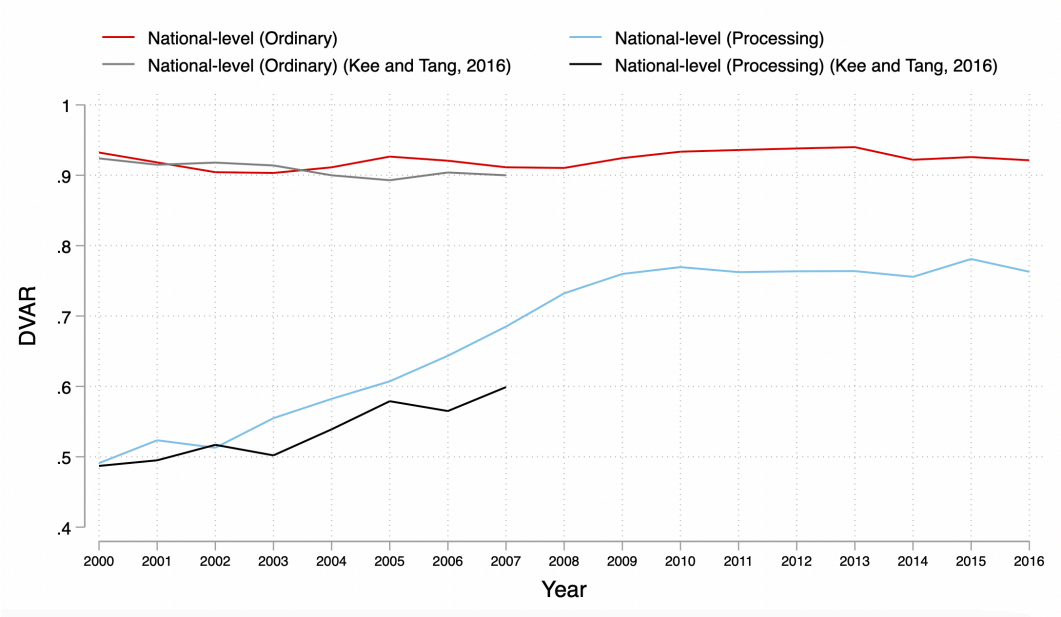


Figure 3: Domestic Value Added Ratio in Aggregate Exports of China, 2000-2016

Figure 4 present the DVARs for ordinary and processing exports at industry level. Similar to the aggregate figures, processing exports have rising DVAR across almost all industries, with the exception of miscellaneous industry.

3.2.1 Robustness Checks

As a robustness check, we compare our current DVAR estimates with the earlier estimates in Kee and Tang (2016) that based on both customs data and firm level industry census data. This is done by constructing the bootstrapped standard errors from our current estimates and perform t-tests with the null hypothesis that the difference is zero. Table 3 presents the results. For all the common years, we cannot reject the null hypothesis that our current estimates are statistically not different from the estimates from Kee and Tang (2016).

Table 3: Ordinary Trade: t Statistics

year	2000	2001	2002	2003	2004	2005	2006	2007
t-Statistics	0.8	0.372	0.178	0.569	0.051	1.12	0.412	0.244

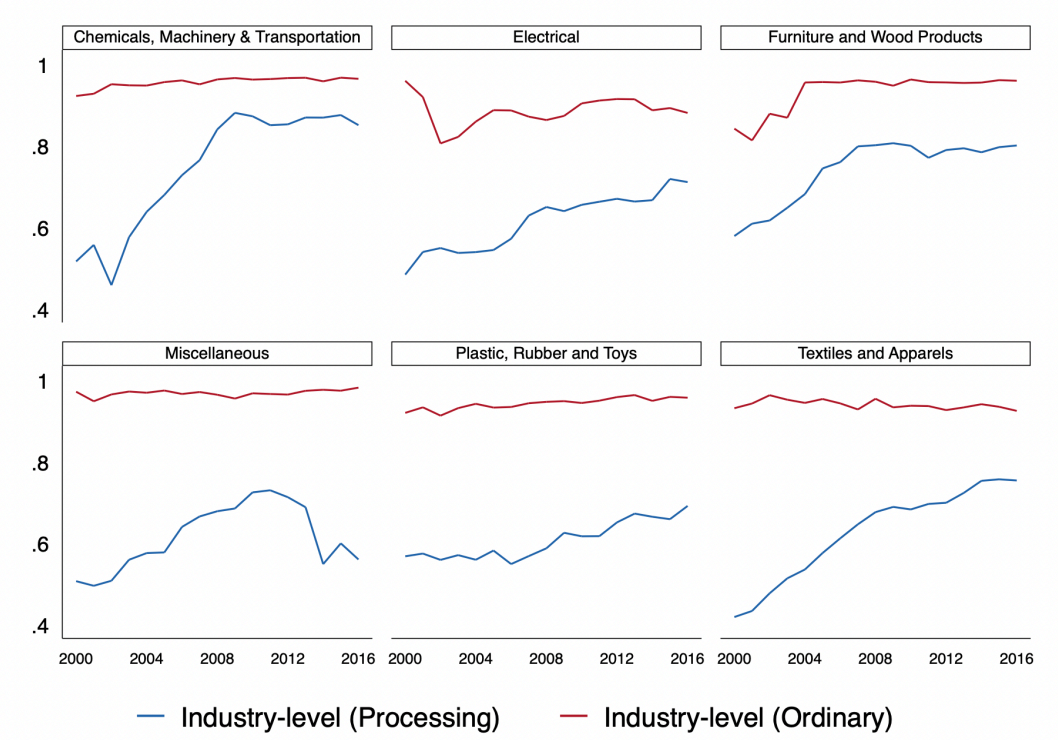


Figure 4: Domestic Value Added Ratio in Industry Exports of China, 2000-2016

There is also a concern that using capital good imports to proxy investment may run into the problems of zeros, given that not all firms import capital goods in all periods. The zeros in capital good imports may be meaningful and may be due to important omitted variables that are also correlated with imports and lead to inconsistent estimates of κ^M. To address this problem, we use a Heckman selection model to first explain the presence of positive capital good imports based on a fixed-effects conditional logit regression. We then include the inverse Mill's ratio from the logit regression in the export revenue function estimation as a control variable. Variables included in the first-stage selection model are the firm-level import weighted exchange rate, firm-level export weighted exchange rate, and firm-level tariffs, along side with the full set of year fixed-effects and firm fixed-effects.

The results show that with a Chi(2) of 3494.66, which is statistically significant, we cannot reject that the first-stage selection model has power in explaining the presence of positive capital good imports. When the inverse Mill's ratios are included in the second stage export revenue estimation, as

in Tables 1 and 2, the estimated κ's increase slightly, but are not statistically different from the original estimates, while the inverse Mill's ratio are mostly statistically significant. Thus, the results suggest that the concern of selection bias due to zero capital imports may not be relevant for this current dataset. These results are available upon requests.

4 Empirical Specifications

This section presents the empirical evidence on the impacts of HSR on the firm DVAR. Given that between 2003 and 2007, only one pilot HSR line (i.e., Qinghuangdao-Shenyang Line) was constructed.[16], the sample period of our main empirical analysis covers from 2007 to 2016.[17] The fixed effects firm-level reduced-form regressions specification for exporting firm i in year t is defined as follows:

$$\text{DVAR}_{it} = \alpha_i + \alpha_t + \beta_1 \text{CI}_{it} + \boldsymbol{X}\boldsymbol{\gamma} + \varepsilon_{it} \tag{14}$$

The dependent variable is the firm DVAR estimates from Section 3. As informed by the model in Section 6, the variable of interests is the Connectivity Index (CI), which is the share of domestic suppliers that are connected to each exporting firms through HSR, thus capturing the impact of HSR on firms' DVAR via domestic suppliers. Control variables in $\boldsymbol{X}$ include firm-level variables such as import weighted input tariff, and exchange rate appreciation (Kee and Tang, 2016), as well as city-level variables such as income level, FDI, number of fixed and mobile telephone lines, internet connection, passenger ridership, goods ridership, inner city transportation, industry structure, population density.

4.1 Potential Domestic Suppliers and Connected Index

However, to construct CI, we first need to identify the domestic suppliers of each exporting firm. Given that we don't have the data on the actual domestic suppliers of each exporting firm, and even if this information is available, the use of the actual domestic suppliers to construct CI will likely run into selection bias and endogeneity issues, and thus bias the regression analysis. To avoid these issues,

[16]As pointed out by Xu and Hu (2024), the Qinhuangdao-Shenyang line was not used as an HSR as the carrying speed was below the minimum speed criteria, and it functioned as the pilot railway to test the feasibility of HSR operations. Meanwhile, Lawrence et al. (2019) also pointed out that the Beijing-Tianjin line opened in 2008 was the first fully HSR line in China.

[17]As noted by Faber (2014), the national trunk highway in China was complete ahead of schedule by the end of 2007. By focusing between 2007 and 2016, we can exclude the confounding effect of national trunk highway construction.

CI is constructed based on the *potential domestic suppliers* of each exporting firm.[18]

Specifically, for each exporting firm, the potential domestic suppliers are firms that export products which share the same HS 6-digit codes with the imported products of the exporting firm. For example, the customs transaction data show that an apparel firm exports men's shirt and imported cotton fabric with a HS code "520819". Then all textile firms that exports cotton fabric with the same code are considered as the potential domestic suppliers of this apparel firm. The word *"potential"* indicates that there is no actual transaction observed between the textile firm and the apparel firm, but the fabric produced by the textile firm could be *potentially* used as domestic materials for the apparel firm. The assumption here is that products with the same HS 6-digit codes are substitutes.

One advantage of this strategy of matching upstream and downstream firms based on their product codes is that we no longer need to rely on matching firms by names or by industry input-output tables, which relaxes the severe data constraints when such firm-level data or input-output tables may not exist. It also allows for heterogeneity in terms of the relevant domestic suppliers for each firm, even within a narrowly defined industry. Finally, there is no endogeneity or selection bias since these are not the actual suppliers but potential suppliers.

However, several caveats need to be mentioned. This strategy could miss out domestic suppliers that do not export and are not included in the customs transaction data. This strategy also could miss out intermediate products that are not imported. Finally, this strategy assumes that products with the same HS 6-digit code are substitutes such that the import of the product can be substituted with the export of the product.

Based on the aforementioned definition of the potential domestic suppliers , the CI for a single imported product (j) of each exporting firm (i) in each year (t) is defined as the following:

$$CI_{jit} = \frac{M_{jit} \sum_{n=1}^{N_{jt}} d_{nit}}{N_{jt}}, \tag{15}$$

where N_{jt} is the total number of potential suppliers of product j in year t, which equals the total number of exporters of product j in year t. The indicator variable, M_{jit}, takes on the value 1 if firm i imports product j in year t, and 0 otherwise; Likewise, the indicator variable, d_{nit}, takes on the value 1 if the potential supplier n is connected to firm i via the HSR network in year t. So the denominator is the total number of potential suppliers of a given imported product in a given year, and the numerator

[18] Alternatively, one could rely on the industry input-output tables which relate all firms in a downstream to all firms in an upstream industry. This is the strategy used by Javorcik (2004) to study vertical spillovers of FDI on their local suppliers. However, as pointed out by Kee and Tang (2016), the use of industry input-output tables have several shortcomings: First, the aggregate input-output tables, usually at regional- or industry-level, hide the importance of firm-level heterogeneity in input-output linkages and thus generate aggregate bias. Second, the input-output tables are constructed based on large-scale firms, which might induce selection bias and underestimate the impact of the production linkages among small and medium size firms. Third, industry input-output table may have limited variation across years due to the infrequent updating, and that may cause problems in the regression analysis that based on yearly data.

is the total number of potential suppliers that are connected via HSR to the exporting firm in a given year. In short, for each imported product j, the variable, CI_{jit}, is the share of potential suppliers that are connected to firm i via HSR. The variable, CI_{jit} equals to 0 if firm i does not import product j in year t or if none of the potential suppliers are connected to the exporting firm i via HSR. Conversely, for product j, if all of the potential suppliers are connected to the exporting firm via HSR, then the variable equals 1. The higher is CI_{jit}, the higher is the likelihood that the exporting firm i can source its material j locally.

For firms that import multiple products, the variable CI_{it}, is the weighted average shares of potential suppliers that are connected to each firm via HSR, with the weights reflect the share of each product j firm i import in year t:

$$CI_{it} = \sum_{j} s_{jit} CI_{jit}, \tag{16}$$

where s_{jit} is the share of product j in the total import of firm i in year t. If a larger share of products firm i imported has potential local suppliers that are connected via HSR, the likelihood firm i may source locally will be higher and may result in a higher DVAR.

In addition, we can further refine CI depending on the distance between the potential suppliers and the exporting firms. If the distance between firm i and potential supplier n falls in the range, d_{int} takes the value of 1, otherwise 0.

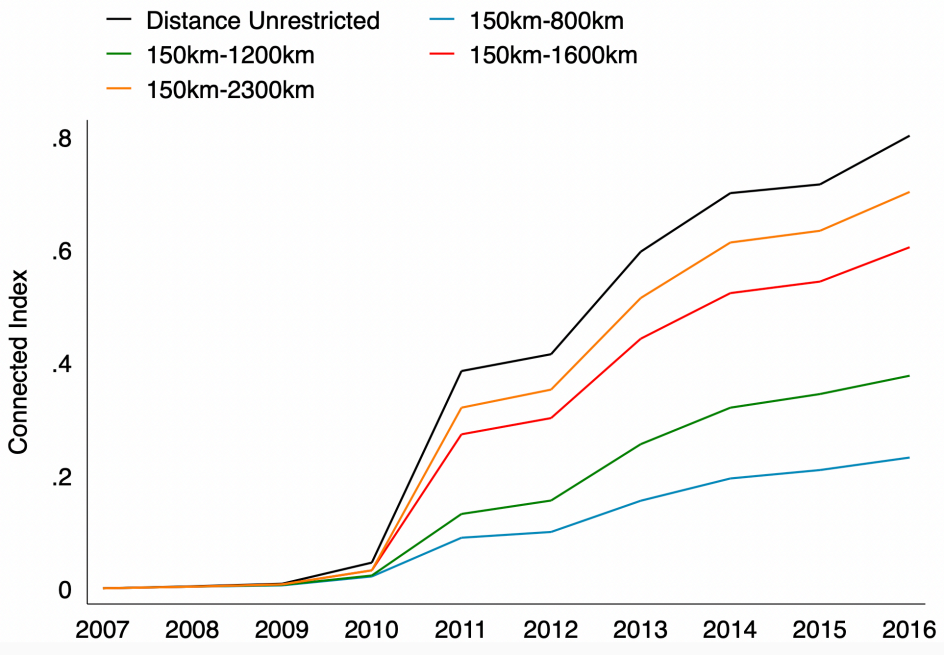


Figure 5: Connected Index Evolution from 2007 to 2016.

Figure 5 presents the evolution of CI from 2007 to 2016. The black line is the unrestricted CI

which represents the share of potential suppliers that are connected to exporting firms via HSR, both directly and indirectly, regardless of the distance. It started from nearly 0 in 2007, expanded around 2010 and reached around 0.8 in 2016. The blue line restricts the distance between the potential suppliers and the exporting firms to be between 150 km and 800 km, while the green line measures the share of the potential local suppliers that with distance between 150 km and 1200 km, follows by red line (between 150 km and 1600 km) and orange line (between 150 km and 2300 km).

5 Empirical Results

Table 4 presents the empirical results. Column (1) is the baseline regression, where we only control for firm fixed effects and year fixed effects. The coefficient of CI is positive and statistically significant, indicating that as more potential local suppliers are connected with the exporting firms via HSR, the DVAR of the exporting firm increases. Column (2) introduces the control variables. The positive coefficient on CI remains significant. Replacing year fixed effects with industry-year fixed effects in Column (3) do not change the results. Column (4) is our preferred specification where we include city-year effect to control for location specific macro variables. Given that many industries are city-specific (e.g. Shenzhen is known for its electronics, Datang is the socks town, Jinjing is the sneakers capital), industry-year effect is dropped in this specification. From 2007 to 2016, the national level aggregated DVAR increases from 0.738 to 0.846, while the average connected index of firms increases from nearly zero to 0.802. Column (4) implies that the improvement of matching with potential domestic suppliers increase firms DVAR by 0.014, which can explain 12.61% increase of DVAR.[19]

Overall, the baseline regression results are consistent with our theoretical model that HSR promotes firm DVAR through connecting exporting firms with domestic suppliers.

5.1 Robustness Check: Role of Distance

The impact of HSR on DVAR may depend on the distance between the exporting firms and the potential suppliers. Firms that are closer may not be using HSR to connect while firms that are far apart may rely on airplanes. We expect HSR to have the largest impact on the medium to long distance.

Table 5 uses CI which restricts the distance between the potential local suppliers and the exporting firms. The results suggest that HSR indeed has the largest impact on firms' DVAR when the local suppliers are within medium to long distance, that is up to 1600 km (between Beijing and Shanghai) to 2300km (between Beijing and Shenzhen) away from the exporting firms. The expansion of HSR facilitates the medium to long distance traveling, which significantly increase the accessibility of local materials for the exporting firms.

[19]The 12.61% is obtained by $0.802 \times 0.017/(0.846 - 0.738) \times 100\%$

Table 4: The Impact of HSR Connected Index on Firms' DVAR: Baseline Results

	(1)	(2)	(3)	(4)
Dependent Variable	DVAR	DVAR	DVAR	DVAR
CI	0.007***	0.007**	0.006**	0.017***
	(0.002)	(0.003)	(0.003)	(0.006)
ln(FDI)		0.176	0.217	
		(0.185)	(0.185)	
FIT		0.675***	0.673***	0.670***
		(0.020)	(0.020)	(0.020)
Δ ln(E)		-2.987***	-2.414***	-2.477***
		(0.653)	(0.653)	(0.641)
Control	No	Yes	Yes	Yes
Firm Fixed Effects	Yes	Yes	Yes	Yes
Year Fixed Effects	Yes	Yes	No	No
Industry-Year Fixed Effects	No	No	Yes	No
City-Year Fixed Effects	No	No	No	Yes
Observations	310,297	267,188	267,188	286,794
Adjusted R^2	0.572	0.577	0.579	0.578

Notes: Robust standard errors clustered at the firm level are in parentheses, ***, **, * denote significance at 1%, 5%, 10% level, respectively. The city-level control variables include GDP Per capita, total output of FDI firms, the number of fixed telephone users, the number of mobile phone users, the number of internet users, the total passenger ridership and freight ridership by transportation other than railway and the secondary industry share and service industry share, firm-level control variables include input tariffs and import-weighted exchange rate variation (appreciation). All control variables are measured in logarithm except for the secondary and third industry share, firm tariffs, and exchange rate.

Table 5: The Impact of Distance-Adjusted HSR Connected Index on Firms' DVAR

	(1)	(2)	(3)	(4)
Distance Range (km)	150-800	150-1200	150-1600	150-2300
Dependent Variable	DVAR	DVAR	DVAR	DVAR
CI	-0.010	-0.001	0.019***	0.019***
	(0.009)	(0.007)	(0.006)	(0.006)
Control	Yes	Yes	Yes	Yes
Firm Fixed Effects	Yes	Yes	Yes	Yes
City-Year Fixed Effects	Yes	Yes	Yes	Yes
Observations	286,794	286,794	286,794	286,794
Adjusted R^2	0.578	0.578	0.578	0.578

Notes: Robust standard errors clustered at the firm level are in parentheses, ***, **, * denote significance at 1%, 5%, 10% level, respectively. The firm-level control variables include input tariffs and import-weighted exchange rate variation (appreciation).

5.2 Robustness Check: Heterogenous Impact of HSR

HSR may have differential impacts on firm DVAR, depending on the flexibility in substituting imported materials with domestic materials. Such flexibility could be affected by the ownership, industries and nature of materials used. To understand which type of firms that benefit most from the expansion of HSR, we split the sample by ownership and industries in Tables 6 and 7, based on our preferred specification. The results indicate that HSR have the largest impacts on the DVAR of FDI and processing firms, and firms in the electronic industries. Given that Chinese exports are dominated by processing exports with FDI, particularly in the electronic industry, the result suggests that the rising DVAR in Chinese exports are mainly driven by FDI or processing firms in the electronic industry using more domestic materials as more local suppliers are accessible via HSR.

Table 6: The Impact of HSR Connected Index on Firms' DVAR: Heterogenous Results

	(1)	(2)	(3)	(4)
Subsample	SOE	FIE	DPE	Processing
Dependent Variable	DVAR	DVAR	DVAR	DVAR
CI	-0.032	0.028***	0.002	0.029***
	(0.025)	(0.007)	(0.011)	(0.011)
Control	Yes	Yes	Yes	Yes
Firm Fixed Effects	Yes	Yes	Yes	Yes
City-Year Fixed Effects	Yes	Yes	Yes	Yes
Observations	8,254	196,996	80,733	106,500
Adjusted R^2	0.361	0.587	0.557	0.598

Notes: Robust standard errors clustered at the firm level are in parentheses, ***, **, * denote significance at 1%, 5%, 10% level, respectively. The firm-level control variables include input tariffs and import-weighted exchange rate variation (appreciation).

Table 7: The Impact of HSR Connected Index on Firms' DVAR: Heterogenous Results: By Industry

Industries	(1) Chemicals, Machinery and Transportation	(2) Electrical, Metals and Optical Equipment	(3) Furniture and Wood Products	(4) Miscellaneous	(5) Plastic, Rubber and Toys	(6) Textiles and Apparels
Dependent Variable	DVAR	DVAR	DVAR	DVAR	DVAR	DVAR
CI	-0.014	0.027**	-0.013	0.037**	0.001	0.030
Control	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
City-Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	74,266	88,184	18,913	32,152	29,198	37,890
Adjusted R^2	0.428	0.581	0.550	0.638	0.637	0.581

Notes: Robust standard errors clustered at the firm level are in parentheses, ***, **, * denote significance at 1%, 5%, 10% level, respectively. The firm-level control variables include input tariffs and import-weighted exchange rate variation (appreciation).

Moreover, the impact of the HSR expansion on firms input sourcing decision also depends on the substitutability between domestic input and imported input. To this end, we interact CI with a variable, $Diff_{i,t_0}$, which is the import weighted firm-level import differentiation index, which reflects the degree of product differentiation in the firm's import bundle:

$$DVAR_{it} = \alpha_i + \alpha_{ct} + \beta_1 CI_{it} + \beta_2 CI_{it} \times Diff_{i,t_0} + \boldsymbol{X}\boldsymbol{\gamma} + \epsilon it, \quad (17)$$

$$Diff_{i,t_0} = \sum_j s_{jit_0} \times Rauch_j \quad (18)$$

where s_{jit_0} denotes the share of HS 6-digits imported product, j, in the total import of firm i, in the first year that firms observed in the data, t_0, and $Rauch_j$ is the differentiated product indicator based on classification proposed by Rauch (1999), which takes the value of 1 for differentiated products, and 0 otherwise.[20] Firms that import more differentiated products will have a higher value of $Diff_{i,t_0}$.

Table 8 presents the results. Column (1) displays result based on Rauch (1999)'s conservative classification of differentiated products while Column (2) displays result based on Rauch (1999)'s liberal classification of differentiated products. The results show that, the impact of improvement in connected index on firms DVAR is larger for firms importing less differentiated products. Presumably imported materials that are less differentiated could be more easily substituted by domestic materials, which become accessible as more domestic suppliers are connected to exporting firms via HSR.

5.3 Robustness Check: Impact of HSR on Exports and Imports

Finally, to confirm that the expansion of HSR allows exporting firms to replace imported materials with domestic materials, without sacrificing their export, Table 9 changes the dependent variable to export value, export variety, import value and import variety. Results show that CI has positive impacts on exports and negative impacts on imports. These results are consistent with the findings that with HSR exporting firms gain better access to domestic materials which they use to replace imported materials.[21] Table 10 further looks at the effects of CI on the export and import prices of different types of firms, similar to Table 6. Here the results show that the export prices of FDI and processing exporters are declining, while the import prices are not affected by CI. This is consistent with hypothesis that HSR expansions allow these exporting firms to substitute the imported materials with domestic materials that are cheaper, which lead to higher DVAR and lower export prices.

[20] Rauch (1999) provides product classification at the Standard International Trade Classification (SITC) 4-digit level (version 2). Homogeneous goods include both goods traded on organized exchanges and reference-priced goods. The remaining goods are classified as differentiated products. The correspondence tables between HS 2007, HS 2012 and SITC 4 come from UN Trade Statistics (https://unstats.un.org/unsd/trade/classifications/correspondence-tables.asp).

[21] In the alternative specification with Export and Import Variety as dependent variables, the results remain similar.

Table 8: The Impact of HSR Connected Index on Firms' DVAR: The Role of Differentiation

	(1)	(2)
Rauch Classification	Conservative	Liberal
Dependent Variable	DVAR	DVAR
CI $\times$Diff$_{t_0}$	-0.007**	-0.009***
	(0.003)	(0.003)
CI	0.022***	0.022***
	(0.006)	(0.006)
Control	Yes	Yes
Firm FE	Yes	Yes
City-Year FE	Yes	Yes
Observations	245,883	245,883
Adjusted R^2	0.575	0.575

Notes: Robust standard errors clustered at the firm level are in parentheses, ***, **, * denote significance at 1%, 5%, 10% level, respectively. The firm-level control variables include input tariffs and import-weighted exchange rate variation (appreciation).

Table 9: The Impact of HSR Connected Index on Firms' Import and Export

	(1)	(2)	(3)	(4)
Dependent Variable	ln(Export Variety)	ln(Import Variety)	ln(Export)	ln(Import)
CI	0.159***	0.023	0.110**	-0.358***
	(0.021)	(0.022)	(0.043)	(0.051)
Control	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
City-Year FE	Yes	Yes	Yes	Yes
Observations	286,794	286,794	286,794	286,794
Adjusted R-squared	0.834	0.808	0.760	0.767

Notes: Robust standard errors clustered at the firm level are in parentheses, ***, **, * denote significance at 1%, 5%, 10% level, respectively. The firm-level control variables include input tariffs and import-weighted exchange rate variation (appreciation).

Table 10: The Impact of HSR Connected Index on Firms' Import Unit Value and Export Unit Value

	(1)	(2)	(3)	(4)
Panel A: Export	SOE	FIE	DPE	Processing
Dependent Variable	ln(Unit Value)	ln(Unit Value)	ln(Unit Value)	ln(Unit Value)
CI	-0.028	-0.132***	-0.031	-0.143***
	(0.143)	(0.041)	(0.049)	(0.050)
Observations	54,378	1,116,530	409,525	597,549
Adjusted R2	0.789	0.729	0.754	0.762
Panel B: Import	SOE	FIE	DPE	Processing
Dependent Variable	ln(Unit Value)	ln(Unit Value)	ln(Unit Value)	ln(Unit Value)
CI	-0.077	0.018	0.033	-0.050
	(0.162)	(0.037)	(0.066)	(0.050)
Observations	108,192	2,368,932	423,344	1,143,154
Adjusted R^2	0.740	0.744	0.760	0.797
Control	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
City-Year FE	Yes	Yes	Yes	Yes
Product FE	Yes	Yes	Yes	Yes

Notes: Robust standard errors clustered at the firm level are in parentheses, ***, **, * denote significance at 1%, 5%, 10% level, respectively. The firm-level control variables include input tariffs and import-weighted exchange rate variation (appreciation).

5.4 Robustness Check: City-level Evidence

Leveraging the large-scale expansion of HSR as a quasi-natural experiment, this robustness check studies the changes in firm DVAR as the cities they are located are served by HSR at different points in time. Without using information on domestic suppliers, this exercise intends to provide consistent supplemental evidence for the previous findings.

To this end, we adopt the difference-in-difference (DID) empirical specifications following Geng et al. (2023) and Kuang et al. (2021):

$$\text{DVAR}_{it} = \alpha_i + \alpha_{pt} + \alpha_{jt} + \beta_1 \text{HSR}_{ct} + \boldsymbol{X}\boldsymbol{\gamma} + \varepsilon_{it} \tag{19}$$

where HSR_{ct} is an indicator variable which equals 1 when prefecture c has been connected to HSR and 0 otherwise.[22] Province-year fixed effects and industry-year fixed effects are denoted by α_{pt} and α_{jt}, and $\boldsymbol{X}$ is a matrix of firm and prefecture control variables. Following Lin (2017) and Lawrence et al. (2019), the standard error ε_{it} is clustered at the prefecture level to account for the correlation between firms within each prefecture and serial correlation over time.

Table 11 presents the regression results relating our firm-level DVAR estimates to the access to the HSR network. Column (1) displays the results when we only control fixed effects. In column (2), in addition to the fixed effects, we further control city attributes such as population density, industry structure, internet connection and so on. All these results show that, the connection to the HSR network has significant positive effect on firm DVAR.[23] Overall, these city-level findings are consistent with the previous findings based on information on domestic suppliers.

6 Two-Sector Model

In this section, we present a simple two-sector model to rationalize the previous empirical results, showing how HSR may improve the DVAR of exporting firms. The main mechanism is motivated by the anecdotal evidence presented at the top of the paper before the introduction that HSR allows employees to travel from cities to cities to conduct business meetings between firms and their domestic suppliers. In abstract terms, this means that HSR expansion increases the probability of exporting firms connecting to domestic suppliers with lower communication cost, resulting in lower prices and a greater varieties of domestic materials available for the exporting firms, thus enabling the exporting firms to substitute domestic for imported materials and leading to a higher DVAR.

[22] The HSR takes the value of one only if the HSR line cross that city and that city has a HSR station along the line.

[23] More results based on this specification can be found in the Appendix D, which address the concern about potential selection bias of the HSR placement using instrumental variables approach and inconsequential unit approach following Redding and Turner (2015). Besides these results, our city-level analyses are also robust to a placebo test which we randomly assign cities' connection to the HSR network, and considering the connection month of the HSR. The empirical results are available upon requests.

Table 11: Baseline Result: HSR Connection and Domestic Value Added Ratio

VARIABLES	(1) DVAR	(2) DVAR
HSR	0.559**	0.605**
	(0.255)	(0.242)
Control	No	Yes
Firm Fixed Effects	Yes	Yes
Industry-Year Fixed Effects	Yes	Yes
Province-Year Fixed Effects	Yes	Yes
Observations	310,382	289,115
Adjusted R^2	0.575	0.575

Notes: Robust standard errors clustered at the prefecture city level are in parentheses, ***, **, * denote significance at 1%, 5%, 10% level, respectively. The control variables include GDP Per capita, total output of FDI firms, the number of fixed telephone users, the number of mobile phone users, the number of internet users, the total passenger ridership and freight ridership by transportation other than railway and the secondary industry share. All control variables are measured in logarithm except for the secondary industry and third industry share.

There are two sectors in the economy. The upstream intermediate goods sector, M, produces materials for the downstream final goods sector, Y. Sector Y is characterized as a monopolistically competitive market, with infinitely many firms, each endowed with productivity φ, and producing a unique variety ω. The intermediate goods sector, M, also has infinitely many identical firms, each producing a unique variety of material, η, in a perfectly competitive market. The economy has a large endowment of labor, L, and a numeraire sector employing L_0, which allows us to set the wage rate to one.

6.1 Final Goods Sector

We begin with sector Y, the final goods sector, following the standard approach of a heterogeneous firm model with a constant elasticity of substitution (CES) preference over the final goods varieties:

$$U \equiv Y = \left[\int_{\omega} y(\omega)^{\rho} d\omega \right]^{1/\rho}, \quad \text{with} \quad 0 < \rho < 1, \tag{20}$$

with the elasticity of substitution between varieties equals $\sigma = 1/(1-\rho)$. The aggregate price index for Y, p_Y, is itself a CES function of individual price of the varieties,

$$p_Y = \left[\int_{\omega} p(\omega)^{1-\sigma} d\omega \right]^{1/(1-\sigma)}. \tag{21}$$

Given the CES preference structure, the downward sloping demand curve for each variety ω is:

$$p(\omega) = p_Y \left[\frac{y(\omega)}{Y}\right]^{-1/\sigma} \tag{22}$$

There are infinitely many firms with heterogeneous productivity, φ, in this sector. In a symmetric equilibrium, each firm chooses to produce a unique variety, ω, obeying the following constant returns to scale Cobb-Douglas (CD) production function of the composite material, M and labor, L_y:

$$\begin{aligned} y(\omega) &= \varphi M^{\theta} {L_y}^{1-\theta}, \text{ with} & (23)\\ M &= M(M^D, M^W), & (24) \end{aligned}$$

where the composite material,M, consists of the domestic material, M^D and the imported material, M^W. The elasticities of output with respect to intermediate input and labor are θ and $1-\theta$, respectively. Profit maximization implies that each firm with productivity φ sets the values of marginal revenue product of M and L equal to the corresponding marginal costs, which determines the downward sloping input demand curves:

$$\begin{aligned} M &= \theta \frac{p(\omega)y(\omega)}{p_M} & (25)\\ L_y &= (1-\theta)p(\omega)y(\omega). & (26) \end{aligned}$$

where p_M is the price of the composite material, and wage is set to be one. Profit maximization implies that the marginal cost and price for a firm with productivity φ are as follows:

$$\begin{aligned} mc(\varphi) &= \frac{{p_M}^{\theta}}{\varphi A}, \quad A \equiv (\theta)^{\theta}(1-\theta)^{1-\theta} & (27)\\ p(\omega) &= \frac{1}{\rho}\frac{{p_M}^{\theta}}{\varphi A}. & (28) \end{aligned}$$

Equation (28) is the horizontal supply curve of variety ω. Combined with equation (21), the downward-sloping demand curve of ω, gives us the equilibrium output and revenue for variety ω:

$$\begin{aligned} y(\omega) &= Y\left(\frac{{p_M}^{\theta}}{\rho\varphi A p_Y}\right)^{-\sigma} & (29)\\ r(\omega) &\equiv p(\omega)y(\omega) = p_Y Y\left(\frac{{p_M}^{\theta}}{\rho\varphi A p_Y}\right)^{1-\sigma}. & (30) \end{aligned}$$

Substituting equation (30) into equation (25) gives us the downward-sloping material input de-

mand curve:

$$
\begin{aligned}
M &= \frac{\theta p_Y Y}{{p_M}^{1-\theta(1-\sigma)}}(\rho\varphi A p_Y)^{\sigma-1} && (31)\\
\frac{\partial \ln M}{\partial \ln p_M} &= -[1+\theta(\sigma-1)] < 0 && (32)\\
p_M M &= \frac{\theta p_Y Y}{{p_M}^{\theta(\sigma-1)}}(\rho\varphi A p_Y)^{\sigma-1} && (33)
\end{aligned}
$$

Without loss of generality, similar to Kee and Tang (2016), given the prices of imported intermediate input, p_M^W, and domestic intermediate input, p_M^D, we assume that the unit material cost function of the composite intermediate input, $p_M\left(p_M^W, p_M^D\right)$, is a translog function that is symmetric and homogenous of degree one. This type of function can provide a second-order approximation to any functional form of price aggregates (Diewert and Wales, 1988):

$$
\begin{aligned}
\ln p_M\left(p_M^W, p_M^D\right) &= a_0 + a_{0W}\ln p_M^W + a_{0D}\ln p_M^D \\
&\quad + \frac{1}{2}a_{WW}\left(\ln p_M^W\right)^2 + a_{WD}\left(\ln p_M^W\right)\left(\ln p_M^D\right) + \frac{1}{2}a_{DD}\left(\ln p_M^D\right)^2. \qquad (34)
\end{aligned}
$$

The assumptions of symmetry and homogeneous of degree one imply the following restrictions on the translog parameters:

$$
\begin{aligned}
a_{WW} &< 0;\ a_{DD} < 0;\ a_{0W} + a_{0D} = 1;\ a_{WW} + a_{WD} = a_{DD} + a_{WD} = 0; \\
\text{and } a_{WW} &= a_{DD} = -a_{WD} < 0 \Rightarrow a_{WD} > 0. \qquad (35)
\end{aligned}
$$

With the translog price function, it can be shown that $\partial \ln p_M / \partial \ln p_M^W$ equals the share of imported material in total material cost, which is a log-linear function of the domestic price of materials relative to the world price of materials:

$$
\begin{aligned}
\frac{p_M^W M^W}{p_M M} &= \frac{\partial \ln p_M\left(p_M^W, p_M^D\right)}{\partial \ln p_M^W} > 0 \\
&= a_{0W} + a_{WW}\ln p_M^W + a_{WD}\ln p_M^D \qquad (36)\\
&= a_{0W} + a_{WD}\ln\frac{p_M^D}{p_M^W}.
\end{aligned}
$$

Similarly, $\partial \ln p_M / \partial \ln p_M^D$ equals the share of domestic material in total material cost, and is a

log-linear function of the domestic price of material relative to the world price of material:

$$\begin{aligned}\frac{p_M^D M^D}{p_M M} &= \frac{\partial \ln p_M\left(p_M^W, p_M^D\right)}{\partial \ln p_M^D} > 0 \\ &= a_{0D} - a_{WD} \ln \frac{p_M^D}{p_M^W}.\end{aligned} \tag{37}$$

Thus, if p_M^D drops, the unit cost of materials, p_M, will also drop in equilibrium, while the share of domestic materials will increase, given $a_{WD} > 0$. From equation (37), we can drive the demand for domestic material:

$$M^D = \frac{p_M M}{p_M^D}\left(a_{0D} - a_{WD} \ln \frac{p_M^D}{p_M^W}\right) \tag{38}$$

$$= \frac{\theta p_Y Y}{p_M^D {p_M}^{\theta(\sigma-1)}}(\rho\varphi A p_Y)^{\sigma-1}\left(a_{0D} - a_{WD} \ln \frac{p_M^D}{p_M^W}\right) \tag{39}$$

$$\ln M^D = -\ln p_M^D - \theta(\sigma-1)\ln p_M + \ln \theta p_Y Y (\rho\varphi A p_Y)^{\sigma-1} + \ln\left(a_{0D} - a_{WD} \ln \frac{p_M^D}{p_M^W}\right) \tag{40}$$

$$\frac{\partial \ln M^D}{\partial \ln p_M^D} = -1 - \theta(\sigma-1)\frac{\partial \ln p_M}{\partial \ln p_M^D} - \frac{a_{WD}}{a_{0D} - a_{WD} \ln \frac{p_M^D}{p_M^W}} \tag{41}$$

$$= -1 - \theta(\sigma-1)\left(a_{0D} - a_{WD} \ln \frac{p_M^D}{p_M^W}\right) - \frac{a_{WD}}{a_{0D} - a_{WD} \ln \frac{p_M^D}{p_M^W}} < 0. \tag{42}$$

In summary, with the constant returns to scale Cobb-Douglas production in equation (23), composite material accounts for a fixed share θ of the total revenue, as shown in equation (25). Additionally, within this fixed share, domestic and imported materials follow a translog specification, implying variable shares that depend on the relative price of domestic to imported materials, as shown in equations (36) and (37). Overall, the shares of domestic and imported materials in the total revenue are variable and also depend on the relative price of domestic and imported materials.

The analysis so far has focused only on firms in sector Y that serve only one market, which is the domestic market (superscript of D omitted for clarity). Now let's consider the export market for each variety ω of Y, assuming that the product for domestic and export is identical. The only difference between serving the domestic market and the export market is the variable trade costs, such as transportation costs and foreign tariffs that firms face due to exporting. Assuming a multiplicative iceberg ad valorem variable trade cost, $t > 1$, the marginal cost of exporting variety ω for a firm with productivity φ is

$$mc^W(\varphi) = t\, mc(\varphi), \quad t > 1 \tag{43}$$

$$= \frac{t {p_M}^{\theta}}{\varphi A}. \tag{44}$$

Thus, given the constant markup rule of profit maximization, the equilibrium price, export supply and revenue of variety ω for the export market are respectively:

$$p^W(\omega) = \frac{1}{\rho}\frac{t{p_M}^\theta}{\varphi A}, \tag{45}$$

$$y^W(\omega) = Y\left(\frac{t{p_M}^\theta}{\rho\varphi A p_Y}\right)^{-\sigma} \tag{46}$$

$$r^W(\omega) \equiv p^W(\omega)y^W(\omega) = p_Y Y\left(\frac{t{p_M}^\theta}{\rho\varphi A p_Y}\right)^{1-\sigma} \tag{47}$$

By taking ratio of domestic revenue over export revenue through dividing equation (30) by equation (47), and including the superscript D to denote domestic variables, it becomes clear that the proportion of export revenue in total revenue for variety ω is a constant, depending only on the iceberg trade cost:

$$\frac{p^D(\omega)y^D(\omega)}{p^W(\omega)y^W(\omega)} = t^{\sigma-1},\ \sigma > 1 \tag{48}$$

$$R(\omega) \equiv p^D(\omega)y^D(\omega) + p^W(\omega)y^W(\omega) \tag{49}$$

$$\frac{p^W(\omega)y^W(\omega)}{R(\omega)} = \frac{1}{1+t^{\sigma-1}}. \tag{50}$$

Equation (50) shows that as the variable trade cost increases, the share of export in the total revenue of variety ω drops.

Now consider the domestic value added (DVA) of each exporting variety ω in sector Y, which is the total revenue of the firm minus the cost of imported material:

$$DVA \equiv R(\omega) - p_M^W M^W \tag{51}$$

$$DVAR \equiv \frac{DVA}{p^W(\omega)y^W(\omega)} = (1+t^{\sigma-1})\left(1-\theta\frac{p_M^W M^W}{p_M M}\right) \tag{52}$$

$$= (1+t^{\sigma-1})\left[1-\theta\left(a_{0W} + a_{WD}\ln\frac{p_M^D}{p_M^W}\right)\right] \tag{53}$$

$$\frac{\partial DVAR}{\partial \ln p_M^D} = -(1+t^{\sigma-1})\theta a_{WD} < 0, \tag{54}$$

where equation (51) is the definition of DVA. The DVA divided by firms' total export revenue, $p^W(\omega)y^W(\omega)$, is defined as the domestic value added ratio (DVAR) in equation (52). The right-hand side of equation (52) is obtained by substituting θ for $p_M M/r(\omega)$. Substituting equation (36), the share of imported material in total material, into equation (52) to get equation (53). Finally, equation (54) shows that DVAR is negatively correlated with the price of domestic materials. Any factors that reduce the price of domestic materials increase the DVAR of final goods exporters. We will show in

the later section how the expansion of HSR may lower the price of domestic materials through raising the probability of exporting firms connecting with domestic suppliers with low communication cost, which thus will lead to an increase in DVAR of exporters. Additionally, as theoretically demonstrated by Xu and Hu (2024), while high-speed rail does not directly lower domestic transportation costs for shipping inputs, it enhances matching efficiency among firms, effectively tightening the product network. This improvement is akin to a reduction in the cost of domestic material inputs at the aggregate level.[24] Thus, the comparative statics presented further illustrate how HSR contributes to rising DVAR.

6.2 Domestic Materials Sector

For each final goods variety ω, there is a perfectly competitive domestic material sector, sector M^D, with the demand governed by the production of each ω, which is assumed to require all the available discrete domestic material varieties, indexed by $\eta = 1, ..., L(\omega)$, where $L(\omega) < \infty$ is the total number of accessible variety for firm ω:

$$M^D(\omega) = \sum_{\eta=1}^{L(\omega)} m_\eta^D. \tag{55}$$

Equation (55) shows that different varieties of domestic materials η, are considered as perfect substitutes, and all $L(\omega)$ accessible varieties are included in the composite domestic materials of firm ω. We further assume that each variety η of domestic materials is supplied by a domestic firm using one unit of labor to produce one unit of η, and the output of each firm is one unit as well. Given that labor supply is perfectly elastic with a wage rate of one, perfect competition implies that the factory-gate price for each variety of η is $p_\eta = 1$.

In addition, all domestic suppliers are assumed to be randomly scattered across cities around each final goods firm ω. Some of these cities are connected to firm ω via HSR, while others are not. Let indicator variable $d_\eta(\omega) = 1$ denote the case that domestic supplier η of the final goods firm ω is located in a city connected to firm ω via HSR, and $d_\eta(\omega) = 0$ otherwise. Assume that $d_\eta(\omega)$ follows a Bernoulli distribution, with the probability $CI(\omega) \in [0, 1]$ for $d_\eta(\omega) = 1$, indicating that η and ω are connected via HSR:

$$\begin{aligned} \text{Prob}(d_\eta(\omega)) &= \begin{cases} CI(\omega), \text{ for } d_\eta(\omega) = 1 \\ 1 - CI(\omega), \text{ for } d_\eta(\omega) = 0 \end{cases} && (56) \\ &= CI(\omega)^{d_\eta(\omega)}[1 - CI(\omega)]^{1-d_\eta(\omega)} && (57) \end{aligned}$$

This implies that for each final goods firm ω, the sum of connected suppliers follows a Binomial

[24] In the appendix model of Bernard et al. (2019), the authors represent the impact of HSR by modeling it as a reduction in domestic transportation costs for inputs, serving as a simplified shortcut.

distribution:

$$\sum_{\eta=1}^{L(\omega)} d_\eta(\omega) \sim B(L(\omega), CI(\omega)) \tag{58}$$

$$E\left[\sum_{\eta=1}^{L(\omega)} d_\eta(\omega)\right] = CI(\omega)L(\omega), \tag{59}$$

where equation (58) shows that the sum of connected suppliers has a Binomial distribution with $L(\omega)$ total accessible varieties and the probability of HSR connection between a domestic supplier and the exporting firm is $CI(\omega)$. Equation (59) shows that, for each firm ω, the share of connected domestic suppliers in the total number of accessible suppliers is $CI(\omega)$, and the expected total number of connected domestic suppliers increases with the probability of HSR connection, which is also $CI(\omega)$.

For each final goods firm ω to incorporate domestic material variety η in the production of $y(\omega)$, firm ω needs to conduct face-to-face meetings with supplier η. We assume that such meetings incur a positive and multiplicative ad valorem iceberg communication costs $\tau > 1$, if supplier η and firm ω are not connected by HSR. In contrast, if they are connected via HSR, then the communication cost is reduced to zero, or $\tau = 1$. For simplicity, we assume that any communication costs, when present, will be borne by firm ω. Thus, the price of domestic material η facing the final goods firm ω depends on both factory-gate price p_η and face-to-face communication costs for firm ω to travel to the domestic supplier η.

$$p_\eta^D(\omega) = \begin{cases} 1, \text{ for } d_\eta(\omega) = 1 \\ \tau, \text{ for } d_\eta(\omega) = 0 \end{cases} \tag{60}$$

Thus, the Bernoulli distribution of $d_\eta(\omega)$ further implies that $p_\eta^D(\omega)$ has the following distribution:

$$\text{Prob}(p_\eta^D(\omega)) = \begin{cases} CI(\omega), \text{ for } p_\eta^D(\omega) = 1 \\ 1 - CI(\omega), \text{ for } p_\eta^D(\omega) = \tau \end{cases} \tag{61}$$

In the perfectly competitive equilibrium, firm ω will incorporate all the accessible varieties of domestic materials into the composite domestic material, M^D, as specified by equation (55). The price of the composite domestic material for firm ω is the expected price of the different varieties of

domestic material:

$$
\begin{aligned}
M^D(\omega) &= L(\omega) && (62)\\
p_M^D(\omega) &= CI(\omega) + [1 - CI(\omega)]\tau \\
&= -(\tau - 1)CI(\omega) + \tau && (63)\\
\frac{\partial p_M^D}{\partial CI} &= -(\tau - 1) < 0, && (64)
\end{aligned}
$$

implying that for each firm ω, as the probability of connecting with domestic suppliers via HSR increases, pushing up the number of connected suppliers, the average price of domestic material drops.

For each firm ω, the equilibrium accessible variety $L(\omega)$, which is also the equilibrium composite domestic material $M^D(\omega)$, is determined by equating domestic material demand in equation (38) from the final goods sector, and the supply in equation (62), and substituting the equilibrium domestic price of materials in equation (63):

$$
L(\omega) = \frac{p_M M}{-(\tau - 1)CI(\omega) + \tau}\left[a_{0D} - a_{WD} \ln \frac{-(\tau - 1)CI(\omega) + \tau}{p_M^W}\right] \tag{65}
$$

$$
\begin{aligned}
\frac{\partial L(\omega)}{\partial CI} = &\frac{(\tau - 1)p_M M}{[-(\tau - 1)CI(\omega) + \tau]^2}\left[a_{0D} - a_{WD} \ln \frac{-(\tau - 1)CI(\omega) + \tau}{p_M^W}\right] \\
&+ \frac{p_M M}{-(\tau - 1)CI(\omega) + \tau}\left[\frac{a_{WD}(\tau - 1)}{-(\tau - 1)CI(\omega) + \tau}\right] > 0.
\end{aligned} \tag{66}
$$

Thus, any factors that increase the probability of connecting domestic materials suppliers with final goods firm ω expand the overall accessible domestic suppliers and the total varieties of domestic material in the equilibrium.

The equilibrium also implies the labor market clearing condition, that is, the sum of the overall labor demand in the final goods sector, the sum of the overall labor demand in the domestic material sector and the labor demand of the numeraire sector equals the endowment of labor L:

$$
\int_\omega L_{y(\omega)} d\omega + \int_\omega L(\omega) d\omega + L_0 = L. \tag{67}
$$

6.3 Expansion of HSR

Now consider an exogeneous expansion of HSR network. The expansion in HSR network increases the cities that are connected to HSR, thereby increasing the probability of domestic materials supplier η connecting to final goods firm ω, depicted as an increase in $CI(\omega)$. According to equations (64) and (66), the price of the composite domestic materials of firm ω decreases, and the accessible domestic materials suppliers for each firm ω increases consequently. Furthermore, equation (54) shows that, the

fall in domestic materials prices cause final goods firms' DVAR to increase. In sum, the expansion of HSR which improves the connectivity between domestic suppliers and the exporting firms will cause the DVAR of export to increase. These theoretical results are consistent with the empirical findings in Section 5.

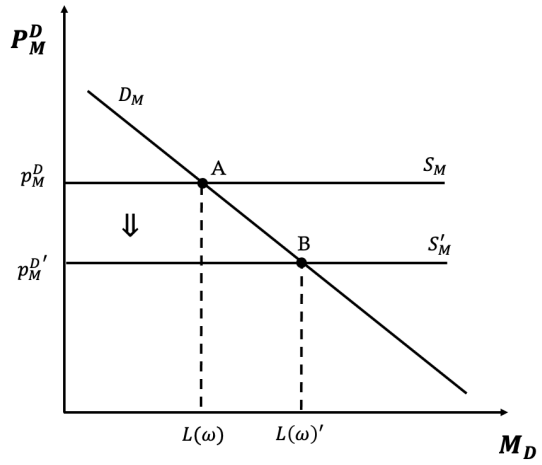


Figure 6: The Impact of the HSR Connection Improvement on Domestic Material Sector

Figure 6 illustrates the impact of HSR expansion on domestic materials sector. For firm ω, the domestic composite material market equilibrium is determined by the intersection between the downward-sloping demand curve, equation (39), and the horizontal supply curve, equation (63). The improvement of HSR connection generates an increase in CI, which reduces the price of composite domestic material and moves the horizontal supply curve downward. As a result, the equilibrium moves from point A to point B, with an increasing use of domestic composite material, from $L(\omega)$ to $L(\omega)'$, implying increases in equilibrium accessible domestic suppliers and the total varieties of domestic materials. This is because, as HSR network improved, there will be an expansion of accessible domestic suppliers in order to meet increased demand due to lower price of domestic materials.

7 Conclusions

This paper studies the effects of HSR expansion on Chinese rising DVAR in exports, focusing on the role of the HSR in connecting exporting firms to domestic suppliers. The novel estimation method of firm-level DVAR from an export revenue function only requires customs transaction data, without the need to match with industrial census data, thereby broadening its applicability.

Empirical results based on firm-level input-output linkages, which we measured directly from the custom transaction data show that the rising DVAR can be explained by the expansion of HSR which enlarges access of domestic materials by connecting exporting firms with their potential domestic

suppliers. The effects of HSR connection are more pronounced for processing firms, foreign-invested firms, and firms in the electronic industry, as well as for firms that import more homogenous products, and are connected to domestic suppliers over medium to long distance.

In addition, this paper provides a theoretical model demonstrating the impacts of HSR expansion on firm DVAR through improving the probability of the exporting firms gaining access to domestic suppliers with lower communication costs. As more domestic suppliers are connected with the exporting firms with the expansion of HSR network, more domestic materials are accessible and promote the substitution of domestic materials for imported materials, hence pushing up firm-level DVAR.

The results of this paper suggest that in addition to trade and FDI liberalizations, which have been previously identified as important factors triggering Chinese rising DVAR, upgrading domestic transportation infrastructure may also enable countries to capture more domestic value added and moving up the GVC.

References

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

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

Amos, P., Bullock, D., and Sondhi, J. (2010). *High-speed rail: The fast track to economic development?* World Bank.

Antràs, Pol, D. C. (2012). Organizing the global value chain. *Econometrica*, 81(6):2127–2204.

Antràs, P. and Chor, D. (2022). Global value chains. *Handbook of international economics*, 5:297–376.

Antràs, P., Chor, D., Fally, T., and Hillberry, R. (2012). Measuring the upstreamness of production and trade flows. *American Economic Review Papers and Proceedings*, 102(3):412–416.

Banerjee, A., Duflo, E., and Qian, N. (2020). On the road: Access to transportation infrastructure and economic growth in china. *Journal of Development Economics*, 145:102442.

Baum-Snow, N., Brandt, L., Henderson, J. V., Turner, M. A., and Zhang, Q. (2017). Roads, railroads, and decentralization of chinese cities. *Review of Economics and Statistics*, 99(3):435–448.

Bernard, A. B., Moxnes, A., and Saito, Y. U. (2019). Production networks, geography, and firm performance. *Journal of Political Economy*, 127(2):639–688.

Brandt, L., Van Biesebroeck, J., Wang, L., and Zhang, Y. (2017). Wto accession and performance of chinese manufacturing firms. *American Economic Review*, 107(9):2784–2820.

Charnoz, P., Lelarge, C., and Trevien, C. (2018). Communication costs and the internal organisation of multi-plant businesses: Evidence from the impact of the french high-speed rail. *The Economic Journal*, 128(610):949–994.

Chaudhary, S. (2023). Globalizing highways : Domestic roads and foreign inputs. *Working Paper*.

Chen, H., Cheng, K., and Zhang, M. (2023). Does geographic proximity affect firms' cross-regional development? evidence from high-speed rail construction in china. *Economic Modelling*, page 106402.

Cui, C. and Li, L. S.-Z. (2019). High-speed rail and inventory reduction: Firm-level evidence from china. *Applied Economics*, 51(25):2715–2730.

Diewert, E. and Wales, T. (1988). A normalized quadratic semiflexible function form. *Journal of Econometrics*, 37:327–342.

Donaldson, D. (2018). Railroads of the raj: Estimating the impact of transportation infrastructure. *American Economic Review*, 108(4-5):899–934.

Donaldson, D. and Hornbeck, R. (2016). Railroads and american economic growth: A "market access" approach. *The Quarterly Journal of Economics*, 131(2):799–858.

Dong, L., Du, R., Kahn, M., Ratti, C., and Zheng, S. (2021). "ghost cities" versus boom towns: Do china's high-speed rail new towns thrive? *Regional Science and Urban Economics*, 89:103682.

Dong, X., Zheng, S., and Kahn, M. E. (2020). The role of transportation speed in facilitating high skilled teamwork across cities. *Journal of Urban Economics*, 115:103212.

Duranton, G. and Turner, M. A. (2012). Urban growth and transportation. *Review of Economic Studies*, 79(4):1407–1440.

Faber, B. (2014). Trade integration, market size, and industrialization: Evidence from china's national trunk highway system. *The Review of Economic Studies*, 81(3):1046–1070.

Fan, J., Lu, Y., and Luo, W. (2023). Valuing domestic transport infrastructure: A view from the route choice of exporters. *Review of Economics and Statistics*, pages 1–18.

Fang, H., Wang, L., and Yang, Y. (2023). Competition and quality: evidence from high-speed railways and airlines. *Review of Economics and Statistics*, pages 1–47.

Gao, Y. and Zheng, J. (2020). The impact of high-speed rail on innovation: An empirical test of the companion innovation hypothesis of transportation improvement with china's manufacturing firms. *World Development*, 127:104838.

Geng, C., Li, D., Sun, J., and Yuan, C. (2023). Functional distance and bank loan pricing: Evidence from the opening of high-speed railway in china. *Journal of Banking & Finance*, 149:106810.

Hummels, D., Ishii, J., and mu Yi, K. (2001). The nature and growth of vertical specialization in world trade. *Journal of International Economics*, 54(6):75–96.

Javorcik, B. S. (2004). Does foreign direct investment increase the productivity of domestic firms? in search of spillovers through backward linkages. *American Economic Review*, 94(3):605–627.

Johnson, R. (2014). Five facts about value-added exports and implications for macroeconomics and trade research. *Journal of Economics Perspectives*, 28(2).

Johnson, R. and Noguera, G. (2012). Accounting for intermediates: Production sharing and trade in value added. *Journal of International Economics*, 86(2):224–236.

Kee, H. L. (2015). Local intermediate inputs and the shared supplier spillovers of foreign direct investment. *Journal of Development Economics*, 112:56–71.

Kee, H. L. and Tang, H. (2016). Domestic value added in exports: Theory and firm evidence from china. *American Economic Review*, 106(6):1402–36.

Koopman, R., Wang, Z., and Wei, S.-J. (2012). Estimating domestic content in exports when processing trade is pervasive. *Journal of Development Economics*, 99(1):178–189.

Koopman, R., Wang, Z., and Wei, S.-J. (2014). Tracing value-added and double counting in gross exports. *American Economic Review*, 104(2):459–494.

Kuang, C., Liu, Z., and Zhu, W. (2021). Need for speed: High-speed rail and firm performance. *Journal of Corporate Finance*, 66:101830.

Lawrence, M., Bullock, R., and Liu, Z. (2019). *China's high-speed rail development*. World Bank Publications.

Li, H. and Li, Z. (2013). Road investments and inventory reduction: Firm level evidence from china. *Journal of Urban Economics*, 76:43–52.

Li, S., Li, Z., and Yuan, J. (2020). Examining the impact of high-speed railways on land value and government revenue: Evidence from china. *China Economic Review*, 63:101502.

Lin, Y. (2017). Travel costs and urban specialization patterns: Evidence from china high speed railway system. *Journal of Urban Economics*, 98:98–123.

Lin, Y., Qin, Y., Sulaeman, J., Yan, J., and Zhang, J. (2023). Expanding footprints: The impact of passenger transportation on corporate locations. *Review of Finance*, 27(3):1119–1154.

Liu, X., Wang, Y., Li, M., and Gao, H. (2023). The importance of speed to corporate financial asset holdings: Evidence from high-speed rail. *Available at SSRN 4391283*.

Miller, D. L. (2023). An introductory guide to event study models. *Journal of Economic Perspectives*, 37(2):203–230.

Niu, D., Sun, W., and Zheng, S. (2020). Travel costs, trade, and market segmentation: Evidence from china's high-speed railway. *Papers in Regional Science*, 99(6):1799–1825.

Olley, G. S. and Pakes, A. (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica*, 64(6):1263–1297.

Qin, Y. (2017). No county left behind? the distributional impact of high-speed rail upgrades in china. *Journal of Economic Geography*, 17(3):489–520.

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

Redding, S. J. and Turner, M. A. (2015). Transportation costs and the spatial organization of economic activity. *Handbook of regional and urban economics*, 5:1339–1398.

Wu, X. and Yang, J. (2023). High-speed railway and patent trade in china. *Economic Modelling*, 123:106276.

Wu, Y., Lee, C.-C., Lee, C.-C., and Peng, D. (2022). Geographic proximity and corporate investment efficiency: evidence from high-speed rail construction in china. *Journal of Banking & Finance*, 140:106510.

Xu, M. and Hu, Y. (2024). High-speed rail and the demand for skill: Evidence from chinese exports. *INSE Working Paper*.

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

Zhang, X., Wu, W., Zhou, Z., and Yuan, L. (2020). Geographic proximity, information flows and corporate innovation: Evidence from the high-speed rail construction in china. *Pacific-Basin Finance Journal*, 61:101342.

A Tables

Table A1: Industry Classification Based on HS 2 Code

Industry	Sub Industry	HS 2 Code
Chemicals, Machinery and Transportation	Chemicals and Allied	28-38
	Machinery	84
	Transportation	86-89
Textiles and Apparels	Textiles	50-60, 63
	Apparel	61-62
Electrical, Metals and Optical Equipment	Electrical	85
	Metals	72-83
	Optical Equipment	90-92
Plastic, Rubber and Toys	Plastic and Rubber	39-40
	Toys	95
Furniture and Wood Products	Wood Products	44-49
	Furniture	94
Miscellaneous	Foodstuffs and Beverage	16-24
	Raw hides and Leather	41-43
	Footwear and Headgear	64-67
	Stone and Glass	68-71

B Data Appendix

B.1 China City Statistical Yearbooks

We collect prefecture-level socioeconomic characteristics from China City Statistical Yearbooks (2000-2016). The variables include population density, GDP per capita, the industry structure measured by the share of the secondary industry and tertiary industry, the total output value of the above-scale foreign invest enterprises (FIE), the ridership of passengers by road, waterway and airline and the ridership of freight by road, waterway and airline, the number of fixed-line telephone users, the number of mobile phone users, the number of internet users and within city public transport density.

Table B1 reports the mean value of the prefecture-level control variables between 2007 and 2016 for different groups based on the whether and when the city connected to the HSR network. 'Treated'

group refers to prefectures that connected to the HSR network between 2007 and 2016. 'Not yet Treated' groups refers to prefectures that connected to the HSR network between 2017 and 2021, so they are actually never treated group in our sample period, that is between 2007 and 2016. 'Never Treated' groups refers to prefectures that are not connected to the HSR network until 2021.

Compared with the other two groups, 'Treated' group has higher population density, higher average income, more foreign invested enterprises' output and more developed transportation in both passengers and freight, denser telecommunication network and inner city public transport. Meanwhile, 'Not yet Treated' group is more comparable with the 'Treated' group than the 'Never Treated' group. As pointed out by Miller (2023), never treated units could be problematic comparison for the treatment group. As a result, 'Never Treated' groups are omitted from the sample.[25]

Table B1: Prefecture-level Characteristics across Different Cities based on HSR Connection Status

Variable	Unit	Treated	Not yet Treated	Never Treated
Population Density	People per KM2	1,097.63	796.13	551.89
GDP Per Capita	RMB	56,182.31	42,295.85	44,307.57
Second Industry Share of GDP	%	50.32	49.27	49.87
Tertiary Industry Share of GDP	%	43.91	42.99	41.06
Above Scale FIE Output	10k RMB	5,359,041.90	1,638,813.70	715,734.59
Other Ridership of Freight	10k Ton	13,599.44	9,791.81	7,150.59
Other Ridership of Passengers	10k People	11,687.15	7,058.58	4,249.74
Fixed-Line Telephone	10k Households	118.88	66.46	44.15
Mobile Phone	10k Households	459.75	280.13	185.71
Internet	10k Households	88.33	41.51	25.17
Public Transport Per Capita	unit per 10k People	8.82	6.04	5.70
Number of Observations	Unit	1,741	456	695

Notes: "Not yet Treated" refers to prefectures that connected to the HSR network between 2017 and 2021, so they are not yet treated during our sample period, that is from 2000 to 2016.

B.2 Exchange Rate Data

Information on exchange rate is derived from the Penn Table, which provides rich information covering 183 countries between 1950 and 2019, and is complemented by the exchange rate data from the World Bank and International Monetary Fund. The correspondence tables of country code between Chinese customs data and exchange rate data are constructed manually.

[25] Dong et al. (2020) study the impact of the HSR connection on facilitating high skilled teamwork across cities. They keep those cities that are already connected or will be definitely connected by the HSR according to the construction plan issued by the Chinese government.As a result, the cities that never connect to the HSR network are excluded from the sample to have a higher level of homogeneity across cities. The logic is same here.

B.3 Import Tariff Data

The data about tariff of China imposed on import from other countries comes from World Integrated Trade Solution (WITS) (https://wits.worldbank.org), which provides highly disaggregated tariff data at HS6 product-origin country level. We use effectively applied tariff in our analysis. The correspondence tables of country code between Chinese customs data and WITS tariff data are constructed manually.

B.4 City-Pair Distance Data

The distance between two cities used for constructing distance-adjusted connected index comes from Amap (Chinese version Google Map), which is the minimum driving distance between cities.

C Variable Construction

C.1 Computing Firm-specific Exchange Rate Indices

We use the Tornqvist method to construct an firm-specific time-varying exchange rate. For each firm i, let I_{it} be the set of countries firms in firm i import from in two consecutive years, t and $t-1$. Denote country c's currency price of a RMB in year t and $t-1$ by E_{ct} and E_{ct-1} and denote country c's share in firm $i's$ total imports in year t and $t-1$ by s_{cit} and $s_{ci,t-1}$. The firm-specific rate of RMB appreciation with respect to the countries from which firm i imports in year t is defined as

$$\triangle \ln E_{it} = \sum_{c \in I_{it}} \frac{1}{2} \left(s_{cit} + s_{ci,t-1}\right) \left(\ln E_{ct} - \ln E_{c,t-1}\right) \tag{D.1}$$

Using this weighted average of appreciation rates, the firm-specific exchange rate for imports is defined as

$$E_{it} = E_{i,t-1} \exp\left(\triangle \ln E_{it}\right), \tag{D.2}$$

with E_{it} normalized to 1 in the base year or any starting year for each firm.

C.2 Computing Firm-specific Input Tariffs

Following Yu (2015), the firm-level input tariffs are constructed according to the following equations:

$$\text{FIT}_{it} = \sum_{c,k \in O} \frac{m_{i,t,c,k}}{\sum_{c,k \in M} m_{i,t,c,k}} \tau_{t,c,k} \tag{D.3}$$

where FIT_{it} denotes firm-level input tariff of firm i in year t. $m_{i,t,c,k}$ is firm i's import of HS 6 digit product k from country c in year t. τ is the import tariff. Note that $O \cup P = M$, where O, P, M

denotes the product sets of ordinary imports, processing imports, and total imports respectively. The set of processing imports does not appear in (D.3) because processing imports are duty-free.

C.3 Computing City-pair HSR Connection Indicator

We construct city-pair matrix to identify cities that connected directly and indirectly through the HSR network to each city annually. The steps are as follows: First, for each prefecture, the HSR indicator takes the value of one only if the HSR line passes the prefecture, there is a HSR station on the line, and the station is within the administrative boundary of the prefecture. Second, by assigning opening year to the city-line information, we create a matrix of prefecture-pair with cells taking the value of one if the prefecture-pair is connected directly through the HSR line for each year, respectively. Third, through matrix calculation, we create a matrix of prefecture-pair with cells takes the value of one if the prefecture-pair is connected indirectly through the HSR line for each year, respectively.

D City-level Additional Analyses

D.1 Identification of Causal Effects

As emphasized by Redding and Turner (2015), the biggest challenge in identifying the causal effects of the transportation improvement is the endogeneity concern stemming from the non-random selection of the treatment group. As stated in the official documents of the Ministry of Railway of China, the objective of this HSR grid is to connect all of the provincial capitals and other major cities with faster transportation. The placement of HSR lines should target politically on those important and prosperous cities and give them priority in the HSR network. It is likely that policy makers choose to build HSR near the cities that they expect to have higher economic growth, which is unobserved by the researcher. As a result, one potential caveat of our empirical setting is that the HSR network is not randomly distributed, that is, the location of the HSR station and line may suffer selection bias. Though we have already control for multi-dimensional fixed effects and use control variables to mitigate the endogenity concern, it might still be not sufficient.

As emphasized in section 2.3, the construction of the HSR network is designed for enhance interactions between large cities but it is also affected by the engineering cost of the construction which is determined by the natural conditions (Lin, 2017). Meanwhile, the HSR lines between large cites are inevitably cross and connect cities along the route, which can be regarded as more exogeneous treatment group. Based on these backgrounds, in order to further verify that our results are not affected by the endogeneity concern, we adopt the three methods suggested by Redding and Turner (2015): (1) Historical route instrument variable approach; (2) Least cost path spanning tree networks;

and (3) Inconsequential units approach.[26]

D.1.1 Historical Route Instrumental Variables Approach

Duranton and Turner (2012) is the pioneering paper using historical route and map information as a instrument for the modern interstate highways in the US. Baum-Snow et al. (2017) use historical railway and road information in China in 1962 as instrument variables for urban railroad and highway. Similar approaches are adopted in research investigating the impact of the HSR, where Kuang et al. (2021), Chen et al. (2023), Wu and Yang (2023), Wu et al. (2022) and Dong et al. (2020) use 1962 historical railways stations as instruments for the HSR placement, Geng et al. (2023) and Wu et al. (2022) use the number of Ming Dynasty post stations in each city as instruments for the HSR placement. The identification assumption is that, the cities connected by the historical railway or road have a larger possibility of being connected by the HSR currently. However, the historical transportation infrastructure has limited impact on firms' DVAR except through the channel of HSR connection.

Following the same logic, we use 1962 railway information and the 1700 road information as instruments for the current HSR connection. Since the instrument variable is at time-invariant at city-level, we use the interaction between these variables and year dummies as the instruments. Table D1 column (1) and column (2) shows the second-stage results. The coefficient of HSR is significantly positive, suggesting that the baseline results is robut to IV approach.

Table D1: Robustness Check: Instrumental Variable Method

VARIABLES	(1) DVAR	(2) DVAR	(3) DVAR	(4) DVAR
HSR	1.427**	2.924**	2.357***	2.705***
	(0.660)	(1.240)	(0.784)	(0.907)
Firm Fixed Effects	Yes	Yes	Yes	Yes
Industry-Year Fixed Effects	Yes	Yes	Yes	Yes
Province-Year Fixed Effects	Yes	Yes	Yes	Yes
IV	1962	1700	Euclid	LCP
First Stage F-Stat	1864.882	3309.884	4001.33	3760.124
Observations	289,115	289,115	289,115	289,115

Notes: Robust standard errors clustered at the prefecture city level are in parentheses, ***, **, * denote significance at 1%, 5%, 10% level, respectively. The control variables include GDP Per capita, total output of FDI firms, the number of fixed telephone users, the number of mobile phone users, the number of internet users, the total passenger ridership and freight ridership by transportation other than railway and the secondary industry share. All control variables are measured in logarithm except for the secondary industry and third industry share.

[26]There are also some papers using alternative instrument variables to tackle the possible endogeneity issue, such as geographical slope terrain of each city (Geng et al., 2023) and construction engineering difficulties (Lin, 2017; Zhang et al., 2020).

D.1.2 Least Cost Path Spanning Tree Networks

To address this endogenous placement problem, we construct a least-cost path (LCP) spanning tree network as an instrument for the actual HSR connection, in a way similar to Faber (2014), Dong et al. (2021), Banerjee et al. (2020) and Gao and Zheng (2020), to evaluate China's HSR expansion. For robustness, we also construct the Euclidean spanning tree network as the alternative instrument. All these IVs correspond to the question of which routes the central government would have been likely to construct if it had wanted only to link all of the provincial capitals in a single network with the lowest construction cost.

The construction method follows Faber (2014). The first step is to compute the least cost paths between all capital city pairs on the basis of remote sensing data on land elevation, land cover and land use. To do so, we adopt a simple construction cost function that increases linearly with land slope gradients and lands that are covered by water, wetland, or built structures. We then use Dijkstra's optimal route algorithm to construct the least cost HSR paths between all bilateral destinations. In the second step, we use Kruskal's minimum spanning tree algorithm to generate a single continuous network subject to minimization of the global construction cost.[27]

Column (3) to Column (4) in Table D1 report the results for the least cost path instruments according to the Euclidean distance and construction cost, respectively. The F-statistics of the first stage indicates that all these instruments are relevant to the actual HSR placements. The positive coefficients all confirm that the HSR connection contributes to the firm's DVAR increase.

D.1.3 Inconsequential Unit Approach

To further address the concern regarding endogeneity of the HSR location selection, we adopt the inconsequential unit approach following Chen et al. (2023), Lin et al. (2023), Ahlfeldt and Feddersen (2018) and Li et al. (2020). The intuition behind inconsequential unit approach is that the economically insignificant units lying between the large cities are only traversed when they are on the convenient route between the large cities by chance. As a result, the unobserved characteristics of these units are inconsequential to the choice of route and the connection to the transportation system can be regarded as random.

According to the Medium- and Long-Term Railway Network Plan issued by State Council of China, "four horizontal" and "four vertical" passenger transportation line are designed to connect 22 node cities.[28] For these node cities, the HSR construction is highly dependent on the population,

[27] As our instrument variable, i.e., the least-cost-path HSR network, is time-invariant while we control for city fixed effects in the baseline regression, we instrument HSR_{ct} using the interaction of least-cost-path HSR network and year dummies.

[28] These cities are Beijing, Shanghai, Wuhan, Guangzhou, Shenzhen, Shenyang, Harerbin, Dalian, Hangzhou, Ningbo, Fuzhou, Xuzhou, Zhengzhou, Lanzhou, Nanchang, Changsha, Qingdao, Shijiazhuang, Taiyuan, Nanjing, Chongqing, Chengdu.

administration level and economic development status. Beyond these cities, the other cities between these node cities are less artificially selected and have less endogeneity concerns. Alternatively ,the capital city of each province and four municipality (i.e., Beijing, Tianjin, Shanghai, Chongqing) are also vital to economic growth and serve as potential or real node cities and their HSR connection may also suffers more endogeneity. As a result, we can use subsample with observations from these node cities dropped to re-conduct the empirical analysis. The results are presented in columns (1) and column (2) of Table D2, respectively. The significantly positive coefficients of the HSR confirm the improving effects of the HSR connection on firms' domestic value added ratio.

Table D2: Robustness Check: Inconsequential Unit Approach

VARIABLES	(1) DVAR	(2) DVAR
HSR	0.736**	0.699**
	(0.329)	(0.342)
Control	Yes	Yes
Firm Fixed Effects	Yes	Yes
Industry-Year Fixed Effects	Yes	Yes
Province-Year Fixed Effects	Yes	Yes
Observations	170,138	204,850
Adjusted R2	0.586	0.600

Notes: Robust standard errors clustered at the prefecture city level are in parentheses, ***, **, * denote significance at 1%, 5%, 10% level, respectively. The control variables include GDP Per capita, total output of FDI firms, the number of fixed telephone users, the number of mobile phone users, the number of internet users, the total passenger ridership and freight ridership by transportation other than railway and the secondary industry share. All control variables are measured in logarithm except for the secondary industry and third industry share.