High-speed Rail and the Demand for Skill: Evidence from Chinese Exports*

Yumin Hu
PKU

Mingzhi (Jimmy) Xu
PKU

November 29, 2024

Abstract

We study how the extensive expansion of high-speed rail (HSR) affects the demand for skilled labor and trade patterns in China. Our empirical findings show that HSR connections significantly boost exports, elevate their skill content, and lead firms to procure inputs from more distant suppliers while increasing their reliance on domestic sources. Cities connected by HSR see an increase in the share of skilled labor and a rise in skill premiums. These trends suggest a growing relative demand for skills driven by the spatial outsourcing of tasks previously performed by unskilled workers. We develop and quantify a spatial equilibrium model with this mechanism to explore the aggregate and distributional effects of HSR. Quantitative estimates demonstrate that the cost savings from improved matching between buyers and suppliers—driven by HSR connections—are equivalent to an 11.37% reduction in iceberg trade costs for domestic intermediate inputs. This reduction leads to welfare gains equivalent to 4.27% of China's 2011 GDP over the period from 2011 to 2017, accounting for 34% of total welfare changes alongside broader economic shifts. However, the benefits of HSR disproportionately favor coastal cities with higher concentrations of skilled labor, thereby exacerbating existing spatial disparities.

Keywords: Transport infrastructure, outsourcing, skill premium, China

JEL Classification: F14, F16, R13

*We are extremely grateful to Robert C. Feenstra, Giovanni Peri, Deborah L. Swenson, and Ina Simonovska for their invaluable guidance. We thank Samuel Kortum for his assistance on this project. Appreciation also goes to Lorenzo Caliendo, Klaus Desmet, Jonathan Eaton, Benjamin Faber, Luca Macedoni, Andrés Rodríguez-Clare, Peter K. Schott, Alan Taylor, and Vladimir Tyazhelnikov, as well as seminar participants at UC Davis, Midwest, TIGN, and InsTED for their helpful comments. This paper was previously circulated under the title "Riding on the New Silk Road: Quantifying the Welfare Gains from High-Speed Railways" and has been completely rewritten. We especially thank Robert C. Feenstra for generously providing the China Customs Dataset and research support. Mingzhi Xu expresses gratitude for financial support from the China Government Scholarship (CSC) for doctoral study and from the National Science Foundation of China (Grant No. 72322007). This paper is a special gift, prepared and completed in anticipation of the baby girl of Jimmy, expected to arrive in November 2024. Please direct correspondence to Mingzhi (Jimmy) Xu at the Institute of New Structural Economics, Peking University, via email: mingzhixu@nsd.pku.edu.cn, or Yumin Hu via email: yuminhu@pku.edu.cn. All errors are ours.

1 Introduction

Improved buyer-seller linkages are redefining firm boundaries, altering the relative returns to factors such as skilled and unskilled labor and contributing to rising inequality (Bernard and Moxnes, 2018; Antràs and Chor, 2022). While the impacts of reduced trade costs and international outsourcing have been extensively studied during the globalization era (Feenstra and Hanson, 2003; Feenstra, 2007), this paper examines a new aspect: how enhanced matching efficiency between firms, driven by China's rapid expansion of its high-speed rail (HSR) passenger transportation network, affects the demand for skills and skill premiums. By enabling faster and more efficient business travel, HSR lowers coordination costs and improves access to specialized suppliers, making it easier for firms to outsource tasks and strengthen production networks. This shift lowers the demand for unskilled labor in tasks that are easily outsourced while amplifying the demand for skilled labor in tasks that remain internal.

In this paper, we make two key contributions. First, we provide the first evidence of how HSR—a modern infrastructure designed primarily for faster passenger transport—affects the relative demand for skilled labor and trade patterns in China, as reliance on outsourced intermediate inputs increases. Our findings illuminate the rising skill premiums and the skill content of exports, revealing the mechanisms that enhance production networks and outsourcing. However, these insights do not fully capture the broader economic impacts resulting from interconnected market responses. Our second contribution is the development of a general equilibrium model that incorporates spatial sourcing, frictional mobility, and trade, allowing us to quantify the impacts of these mechanisms on the overall distribution and aggregate demand for skills within the Chinese context.

By the end of 2015, China had established the world's longest HSR network, spanning 19,000 kilometers and connecting 170 cities.[1] This network has revolutionized travel for business professionals, offering faster and more efficient connectivity. In 2013, HSR ridership reached 672 million, surpassing air travel, and its reliability, frequent departures, and comfort have made it a preferred choice for medium-distance trips (Ou et al., 2014). Surveys in Bullock et al. (2014) show that business travelers increasingly favor HSR over conventional rail, with business-related trips comprising 17% more passengers on the Changchun-Jilin line and 11% more on the Beijing-Shanghai line.[2] Interviews reveal that HSR significantly reduces travel costs and time, enhancing productivity for firms involved in sales, procurement, and

[1] As of 2013, Spain had the second longest HSR network (2,515 km), followed by Japan (2,388 km) and France (2,036 km).

[2] HSR ridership has grown at an average annual rate of 39%, compared to 13% for air travel, with business travelers making up 40% of riders on the Changchun-Jilin route and 63% on the Beijing-Shanghai corridor.

partnerships. Many employees make 2 to 5 business trips monthly via HSR, underscoring its vital role in sustaining supply chains and customer networks. Notably, commuting accounts for only about 15% of total ridership, primarily for short-distance trips, indicating HSR's limited impact on migration costs (Bullock et al., 2014; Ren et al., 2023).[3]

Leveraging the staggered integration of HSR networks across Chinese cities from 2010 to 2018, we offer robust evidence that HSR connections significantly boost overall efficiency, particularly as the use of intermediate inputs increases. This enhancement disproportionately benefits skilled labor, with cities not only experiencing a rise in exports but also an increase in the skill content of those exports. To address endogenous placement, we employ an instrumental variable strategy using least-cost path spanning tree networks. Our event study further confirms the parallel trend assumption, ensuring that the HSR impact is not influenced by prior growth or anticipated future development. Our findings demonstrate that connectivity to HSR boosts city exports by an impressive 13.3% annually, while also increasing the demand for skilled labor among exporting firms by 3% as a share of total employment. This effect is primarily driven by ordinary exports that require greater domestic inputs, while exports in processing mode—which mainly involve assembling final products from imported intermediate inputs—show no significant change. Furthermore, the increase in skill content is associated with the greater distance of suppliers and a rising share of domestic outsourcing, as evidenced by the declining reliance on imported intermediate inputs alongside an overall increase in input shares. This trend highlights the enhanced domestic outsourcing activity of Chinese firms following the establishment of HSR.

To delve deeper into the potential local development effects, we analyze auxiliary data from the China Tax Survey (2010-2016), the Population Census (2010 and 2015), and the China Family Panel Studies (2010-2018), identifying additional margins of adjustment at the firm and regional levels. Our findings reveal that firms in conneocted cities experience increased wages and heightened R&D intensity in sales following HSR integration. At the local level, connected cities see an increase in both the employment share of skilled labor and a rising skill premium. We interpret these results as indicative of the rising demand for skills driven by HSR. To our knowledge, we are the first to document the causal relationship between HSR networks and skill demand in Chinese exports.

To quantify the general equilibrium effects of improved production networks from HSR expansion and assess its distributional implications, we develop a quantitative spatial trade

[3]According to Bullock et al. (2014); Ren et al. (2023), commuting trips account for about 15% of all HSR journeys on the Changchun-Jilin and Beijing-Tianjin lines. This figure is significantly lower than that of major trips, which are typically long-distance and involve travel times exceeding one hour. However, in smaller countries, such as those in the EU, HSR effectively reduces commuting costs, thereby influencing the residential and working patterns of households Heuermann and Schmieder (2019).

model grounded in Eaton et al. (2016). Our model incorporates costly trade, frictional labor mobility, and spatial outsourcing, allowing us to analyze regions with varying technology and outsourcing capabilities. We demonstrate that HSR connections increase the demand for skilled labor by enabling unskilled tasks to substitute for manufacturing through outsourcing (Feenstra and Hanson, 1996; Baumgarten et al., 2013). By linking outsourcing capabilities to the efficiency of firm-to-firm matching, we show that HSR enhances this efficiency via faster transportation. A key aspect of our model is that increased outsourcing does not necessarily reduce the relative demand for unskilled labor. In fact, it can raise both the marginal product and the demand for unskilled workers in core, non-outsourced tasks, as noted by Wright (2014); Bilal and Lhuillier (2021). The interaction between production and regional outsourcing influences the demand for various labor types and the supply of goods for both final consumption and intermediate tasks across regions. Workers choose their residences based on expected utility, shaped by local amenities, prices, wages, and personal preferences. This migration behavior ultimately determines labor supply and demand for goods, leading to equilibrium where labor distribution and factor prices ensure market clearance.

For precision in counterfactual analyses, we calibrate our model to the initial equilibrium of the Chinese economy and assess the impact of HSR with detailed disaggregated data while maintaining a tight connection with our reduced-form regression. In particular, to capture China's economic geography of goods flow, we utilize transaction records from the Taobao Platform to create a novel dataset of trade flows between cities. Migration patterns for skilled and unskilled labor are sourced from the 2010 Population Census, while regional outsourcing shares are tracked using firm-level data from the Chinese Tax Survey. The impact of HSR on matching efficiency between firms is calibrated to match the coefficients from the reduced-form estimates in the empirical section. This analysis reveals that the cost savings from improved production networks due to HSR connections are equivalent to an 11.37% reduction in iceberg trade costs for domestic intermediate inputs.

Our quantitative analysis indicates that the construction of HSR between 2011 and 2017 increased China's overall welfare by 4.27% of GDP in 2011, accounting for 34% of the total welfare change during this period, which also includes broader economic shifts such as technological advancements. However, the gains from HSR disproportionately benefit skilled labor, resulting in a welfare increase of 6.35%. In contrast, unskilled labor experiences a decline of 0.68% due to decreased demand as outsourcing rises. This widening skill gap means that regions with higher concentrations of skilled labor benefit most from HSR, exacerbating the inequality between cities. Overall inequality, measured by the Theil index, increased by 1.8%. Finally, further policy experiments suggest that this rising inequality could be mitigated through reforms aimed at reducing domestic migration costs and trade barriers.

Related Literature. This paper contributes to several key areas of the literature, with the first being the rapidly growing research on the economic impacts of HSR. Bernard et al. (2019) show that HSR enhances firm productivity by lowering search and outsourcing costs, which allows firms to engage more efficiently with higher-quality suppliers, thereby reducing marginal costs. Similarly, Charnoz et al. (2018) document that the reduced travel times from HSR restructure firms, encouraging functional specialization—where remote affiliates handle production and headquarters focus on management—resulting in productivity gains. However, these firm-level productivity gains do not consistently translate into welfare growth at the aggregate level. For instance, using reduced-form regressions, Lin (2017) finds that HSR connections have no significant effect on city-level GDP in China, while Qin (2017) reports that non-urban counties along upgraded railway lines even saw declines in GDP per capita after HSR integration. By contrast, Ahlfeldt and Feddersen (2018) finds that HSR access in Germany is linked to increases in GDP. We contribute to the literature by investigating how HSR connections shape the relative demand for skilled labor through the mechanism of outsourcing, which offers a potential explanation for these seemingly contradictory results through a unified general equilibrium theory of spatial outsourcing. Our model ties the productivity and distributional effects of HSR on skilled and unskilled workers together and examines the trade-off between these forces at the aggregate level.[4]

Second, this study contributes to the existing literature on how infrastructure investments shape the demand for skilled labor and impact inequality. Prior research demonstrates that the U.S. Interstate Highway System boosted the relative demand for skilled labor by fostering the agglomeration of trade-related activities and reducing migration costs, which disproportionately benefited skilled workers (Michaels, 2008; Cucu, 2022). Similarly, China's National Trunk Highway System raised capital intensity, which in turn increased the skill premium (Lu, 2018). Narayanan and Lebesmuehlbacher (2023) further demonstrate that the effects of capital-skill complementarity, driven by public investment, can vary depending on local labor market frictions. In Switzerland, Fretz et al. (2022) document a rise in inequality following highway construction. However, improved transportation can also expand access to education, helping reduce regional disparities by increasing the supply of skilled labor (Aggarwal, 2018; Adukia et al., 2020). Building on this, we investigate how HSR—a modern infrastructure distinct from previous developments such as highways—shapes the demand for skilled labor. We present the first reduced-form evidence that HSR increases the skill content of exports, which we explain through a general equilibrium spatial model that highlights how

[4]HSR fosters local economic development through several channels. It heightens competition, improving airline service quality (Fang et al., 2023), stimulates growth in peripheral cities, raising property values (Zheng and Kahn, 2013), and reduces travel times, facilitating knowledge exchange and academic collaboration (Dong et al., 2020)

improved transportation enhances the outsourcing of manufacturing tasks—a key mechanism driving our empirical findings.

Our analysis reveals that the distributional effects of HSR stem from an increased capacity to outsource manufacturing tasks. This connects our findings to the broader literature on outsourcing's influence on productivity and income distribution. While productivity gains from outsourcing typically reduce demand for unskilled labor (Feenstra and Hanson, 2003; Geishecker and Görg, 2008; Keuschnigg and Ribi, 2009; Hummels et al., 2014; Acemoglu et al., 2015; Carluccio et al., 2019; Bilal and Lhuillier, 2021; Bergeaud et al., 2024), this does not necessarily lead to greater inequality among different skill groups.[5] Outsourcing also generates output effects that boost demand for core manufacturing tasks performed by unskilled labor (Wright, 2014; Bilal and Lhuillier, 2021). As goods producers employ more service workers, the marginal product of unskilled labor can increase. In our model, the distributional impact of HSR depends on the interaction between outsourcing's distributional and productivity effects, which we calibrate carefully using detailed city-level economic data.

Finally, the quantitative framework of this paper builds on the recent trade literature, notably pioneered by Eaton and Kortum (2002).[6] Standard policy instruments in these models typically focus on trade costs for shipping goods or migration costs for labor movement. However, these approaches do not directly account for the improvement in firm-to-firm matching efficiency resulting from HSR connections—a critical driver of firm performance and growth, as documented by Cristea (2011), Giroud (2013), Cai and Szeidl (2017), Charnoz et al. (2018), and Bernard et al. (2019). To better capture this effect in China's economic context, we extend Eaton et al. (2016)'s framework by allowing matching efficiency to depend on both the location of the outsourcing task's origin and its destination, while introducing frictional labor mobility across regions. This adaptation quantifies the trade cost equivalent of improved matching efficiency between buyers and suppliers within a tractable model that integrates trade, migration, and outsourcing.

The paper is structured as follows: Section 2 provides an overview of China's HSR, while section 3 analyzes its causal impact on exports and skilled labor demand. Section 4 introduces a multi-regional spatial model, calibrated in section 5, with quantitative results

[5] Previous studies on productivity gains from outsourcing (Feenstra and Hanson, 1996; Amiti and Konings, 2007; Goldberg et al., 2010; Bøler et al., 2015; Halpern et al., 2015; Antras et al., 2017; Halpern et al., 2015) primarily attribute the rise in offshoring activities to reduced trade costs. In contrast, our analysis emphasizes the role of improved buyer-seller matching efficiency as a key driver of these gains.

[6] Other recent studies on this topic include Alvarez and Lucas (2007); Arkolakis et al. (2012); Hsieh et al. (2013); Di Giovanni et al. (2014); Caliendo and Parro (2015); Caliendo et al. (2017); Ramondo et al. (2016); Edmond et al. (2015), among others, with a broader literature review provided in Costinot and Rodríguez-Clare (2014). Notably, two prominent applications of these quantitative methods to China are found in the contemporaneous works of Fan (2019) and Tombe and Zhu (2019), who examine the relative importance of domestic migration and trade costs in explaining China's export performance.

discussed in section 6. We conclude in section 7.

2 Overview: China's HSR Network

The term "high-speed rail" refers to railways capable of supporting trains with top speeds exceeding 250 km/h. While early discussions of HSR development in China date back to the 1990s, it wasn't until 2002—when the newly appointed governor of the Ministry of Railways proposed the "Great Leap Forward" strategy—that concrete plans to expand railway capacity and introduce HSR gained traction. Prior to the construction boom in 2007, the only passenger-dedicated HSR line, connecting Qinhuangdao and Shenyang, was introduced in 2003, with a modest operating speed of 200 km/h.[7] In 2008, the State Council set ambitious targets through the *Mid- to Long-Term Railway Development Plan*, aiming to expand the traditional railway network to 120,000 kilometers and establish 16,000 kilometers of HSR, forming four north-south and four east-west corridors. By the end of 2015, the network had expanded to cover 170 cities—including nearly all capital cities and those with populations over 500,000—spanning a total of 19,000 kilometers.[8]

The *Mid- to Long-Term Railway Development Plan* set the primary objective of the HSR system as providing faster transportation links between capital cities. Most HSR lines are designed to accommodate bullet trains operating at average speeds of 350 km/h, substantially outpacing typical intercity lines (200–250 km/h, like the Guangzhou-Zhuhai Intercity Line) and traditional express routes (120–160 km/h).[9] HSR tickets are competitively priced, with second-class fares on 300–350 km/h lines averaging \$0.077 per kilometer. This makes them cheaper than discounted airfare and only about three times the cost of traditional express rail, reinforcing HSR's appeal as a cost-effective travel option.[10] Notably, China's HSR fares are about one-quarter the price of comparable HSR systems in other countries. This affordability is underpinned by the system's high passenger density, strong occupancy rates, and relatively low construction and operating costs.

[7] Technically, the Qinhuangdao-Shenyang line didn't qualify as HSR since its speed fell below the minimum criteria. Instead, it functioned as a pilot line to test the feasibility of HSR operations, as noted by Ou et al. (2014). China's earlier efforts focused on accelerating traditional trains through five waves of speed increases from 1997 to 2004, though these improvements only raised average speeds to 65.7 km/h by 2004.

[8] The *Mid- to Long-Term Railway Development Plan* was revised in 2015, setting new goals of 175,000 kilometers for traditional rail and 45,000 kilometers for HSR by 2030. Further background on China's HSR is provided by Ou et al. (2014), Lin (2017), and Qin (2017). Appendix Figure A.1 illustrates the network's growth from 2010 to 2018, with newly connected cities marked in red for each year.

[9] For instance, the launch of the Jinghu HSR in 2011 cut travel time between Beijing and Shanghai from 13 hours to just 5, significantly boosting connectivity between these major economic hubs.

[10] Traditional express train tickets are often difficult to obtain, and the service quality is significantly lower.

3 Empirical Analysis

To what extent do expanding supply chains and customer networks influence changes in firm production efficiency and the demand for skilled labor? We address this question by examining how HSR access impacts exports and their skill content.[11] Enhanced intercity transportation increases matching efficiency by lowering the cost of face-to-face interactions, facilitating the efficient creation of buyer-seller relationships and outsourcing partnerships. This effect is particularly significant for manufacturing tasks, which often employ unskilled labor.[12] Consequently, we expect cities connected to the HSR network to experience increased average production efficiency, reflected in the expansion of exports and higher skill content in exports. We leverage the staggered rollout of HSR across China to examine its impact on city exports and the demand for skilled labor, focusing on prefecture-level cities as our unit of analysis.[13]

3.1 Specification

We examine the relationship between a city's connection to the HSR network and its total exports and changing skill content, using the following general specifications:

$$y_{it} = \beta\ Connect_{it} + \eta X_{it} + \delta_i + \delta_t + \delta_{g(i)t} + \epsilon_{it}, \tag{1}$$

Our primary outcome variable y_{it} measures the total exports and their skill content from city i in year t. The explanatory variable of interest is a dummy variable, $Connect_{it}$, indicating whether city i was connected to the HSR network in year t. We include a vector of city control variables, X_{it}, which may affect trade and communication between producers. We also control for year (δ_t) or geographic group-year, $\delta_{g(i)t}$, fixed effects to account for national or regional macroeconomic shocks or business cycles, and city fixed effects, δ_i, to account

[11] We focus on export performance because it is less influenced by domestic factors, such as fluctuations in domestic demand, and more directly affected by supply-side changes. To the best of our knowledge, this is the first study to investigate the causal relationship between HSR connections, export performance, and the demand for skilled labor.

[12] The suitability of a supplier's product for the buyer's needs is crucial in forming buyer-supplier linkages. Without HSR, it is less convenient for buyers (e.g., firm managers on the eastern coast of China where wages are high) to frequently travel to the supplier's location (e.g., inland China where wages are low) to ensure the quality and appropriateness of outsourced inputs.

[13] A four-layer hierarchy system governs the geographical administrative units in China. The country is first categorized into provincial units, including provinces (e.g., Jiangsu Province), autonomous regions (e.g., Tibet), and municipalities directly under the central government (e.g., Beijing, Shanghai, Chongqing, and Tianjin). Prefecture-level divisions are the second level of the administrative structure. Large prefectures are subdivided into (autonomous) counties and county-level cities. Finally, townships or towns are at the fourth level of the administrative structure. In this study, the unit of analysis is the prefecture city. Our analysis includes the four municipalities (Beijing, Shanghai, Chongqing, and Tianjin) and provides robustness checks without these super-large cities. For details, see http://xzqh.mca.gov.cn/statistics/2018.html.

for any time-invariant city-specific characteristics that may influence the city exports.[14] To account for potential correlations within cities and across years, we cluster the standard errors at the city level. We begin by detailing the construction of each variable and the corresponding data sources. Comprehensive information on data sources is provided in Appendix C. Our sample includes 215 prefectures across 31 provinces in mainland China. In 2010, the median population of these prefectures was 4.2 million, with a median per capita GDP of 29 thousand Chinese Yuan.

3.1.1 City Exports and Skill Content, y_{it}

Our dependent variable captures both the total volume of exports and their skill content. We source data on Chinese exports by product and city from the China Customs Database (2010-2018), compiled by the General Administration of Customs of China. This extensive dataset covers all Chinese import and export transactions, with each entry providing detailed information on trade values in US dollars at the eight-digit Harmonized System (HS) product level, encompassing over 7,000 products. We aggregate these data to calculate total exports, further breaking them down by exporting regime—specifically distinguishing between ordinary and processing modes—at the city-year level. In our regression analysis, city exports are logged to ensure consistency and comparability.

The measurement of the skill content in exports at the city level from 2010 to 2018 is conducted in two stages. First, we adopt the methodology of Zhu and Trefler (2005) and Amiti and Freund (2008) to assess the skill intensity of a product ($SkillEmp_{p,2000}$) by calculating the ratio of skilled workers—defined as those with a college degree—to total employment in 2000. This assessment relies on data from the 2000 Population Census, which provides detailed information on individuals' employment industry and educational level. Since the 2000 Population Census uses the China Industrial Classification (CIC) system for industry classification, we match the CIC industries to the International Standard Industrial Classification (ISIC) and subsequently map them to the HS 6-digit product codes. In the second stage, we calculate the skill content of a city's exports by year using the following equation in (2):

$$SkillContent_{it} = \sum_{p} SkillEmp_{p,2000} \times \frac{X_{ipt}}{X_{it}} \tag{2}$$

where X_{ipt} represents city i's exports of HS-6 digit product p in year t, and X_{it} denotes the total exports of city i in the same year, i.e., $X_{it} = \sum_p X_{ipt}$. The variable $SkillEmp_{p,2000}$ captures the skill intensity of product p. An increase in $SkillContent_{it}$ indicates a rising

[14] Cities are grouped into three categories based on population size: those with populations under 1 million are classified as Group 1, cities with populations between 1 and 10 million are classified as Group 2, and cities with populations over 10 million are designated as Group 3.

skill content in the city's exports.

3.1.2 City Connection to HSR, $Connect_{it}$

The initial connection time of each city to the HSR network was carefully gathered through multiple sources to ensure both accuracy and comprehensiveness. Our primary reference is the "Major Events" section of the China Railway Statistical Yearbook, where we identified the documented connection dates by searching for each city's name. To enhance this dataset, we supplemented it with information from online news sources. Using Google's advanced search features, we employed specific keywords, including city names, "Dongchezu," and "Hexiehao," to pinpoint the earliest documented connection times. This dual approach allowed us to cross-verify and compile a robust dataset of HSR connection times for all cities. In our study, we focus on HSR connections established prior to 2018.[15] Using this data, we construct the policy variable $Connect_{it}$. This dummy variable equals one for periods after city i first connects to the HSR in year t, and zero otherwise.

3.1.3 Other Control Variables, X_{it}

Other city-level socio-economic variables are primarily sourced from the China City Statistical Yearbooks for the years 2010 to 2018. These variables include city-specific data on population, average income in the manufacturing and service sectors, and the number of internet users. To control for the impact of tariffs on Chinese city exports, we utilize tariff data from the WTO's Integrated Database (IDB). Let $\tau_{n,p,t}^{CHN}$ represent the tariffs imposed by country n on imports from China at the HS 6-digit product level (p) for year t. We calculate the tariff exposure for each city c as the weighted average tariff rate, using the city's product-level exports in 2010 as the weights, as detailed in the following equation:

$$\text{Tariff}_{it} = \sum_{n,p} \frac{X_{inp,2010}}{X_{i,2010}} \tau_{n,p,t}^{CHN} \tag{3}$$

where $X_{\text{inp},2010}$ is city i's exports to country n for product p in 2010, and $X_{i,2010}$ denotes the total exports of city i in that year. Finally, our sample enables consistent tracking of 215 Chinese prefecture-level cities over time. Panel (a) of Appendix TableA.1 presents summary statistics for the variables used in the city-level regression analysis from 2010 to 2018.

3.2 Results

Columns (1) and (2) of Table 1 present the baseline results, where the dependent variable is a city's total exports to the rest of the world. The findings provide consistent evidence

[15] As shown in Appendix Figure A.1, 55 cities were connected to the HSR network by 2010, and this number grew to 215 by 2018. The year 2014 marks a particularly notable period of expansion, with 44 additional cities joining the network.

that HSR connections significantly increase a city's total exports, both with and without controlling for city-specific factors. Columns (10) and (11) focus on the skill content of city exports. As expected, the estimated coefficients show that HSR connections significantly enhance the skill content of these exports.

Table 1: Impact of HSR Connection on City Exports, 2010-2018

Dep var:	City Exports									Skill Content		
	Overall			Ordinary			Processing			in City Exports		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
$Connect_{it}$	0.048**	0.040*	0.133***	0.041**	0.049**	0.150***	0.036	0.038	0.370	0.014***	0.015***	0.030***
	(0.023)	(0.023)	(0.048)	(0.020)	(0.020)	(0.047)	(0.103)	(0.104)	(0.220)	(0.004)	(0.004)	(0.009)
Observations	1,934	1,907	1,899	1,622	1,604	1,604	1,622	1,604	1,604	1,934	1,907	1,899
R-squared	0.980	0.981	0.981	0.985	0.985	0.985	0.874	0.873	0.874	0.830	0.829	0.833
LCP IV	-	-	Y	-	-	Y	-	-	Y	-	-	Y
First-stage F-statistic	-	-	35.404	-	-	35.404	-	-	35.404	-	-	35.404
City FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Group-Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Other Controls	-	Y	Y	-	Y	Y	-	Y	Y	-	Y	Y

Notes: All regressions include fixed effects for cities, years, and group-years. Cities are classified into three groups based on population size: Group 1 comprises cities with populations below 1 million, Group 2 includes cities with populations between 1 million and 10 million, and Group 3 consists of cities with populations exceeding 10 million. The IV is the least-cost path-spanning tree instrument. Additional controls include average annual wage rates in the manufacturing and services sectors, tariff rates, population size, and internet coverage at the city level. Robust standard errors, clustered at the city level, are reported in parentheses: *** p<0.01, ** p<0.05, * p<0.1.

Instrumental Variable Approach. The selection of nodal cities in the HSR network may be not random, as it may be influenced by the socio-economic characteristics of a city, leading to potential endogeneity issues. However, official documents from the Ministry of Railways of China indicate that the HSR grid was initially designed to connect all provincial capitals and other major cities with faster transportation, with a focus on politically significant cities. This strategic design provides an opportunity to construct a least-cost path network as an instrumental variable for actual HSR placement. To address the endogeneity issue, we construct a least-cost path (LCP) spanning tree network as an instrument for the actual HSR connection. Our approach follows the methodology of Faber (2014), with a key distinction: the objective of the HSR network, as stated in official reports, is to link all provincial capitals within a single network.

The generation of the LCP involves two steps. First, we calculate the least-cost paths between all pairs of capital cities using remote sensing data on land elevation, land cover, and land use. We employ a straightforward construction cost function that increases linearly with land slope gradients and areas covered by water, wetlands, or built structures. Dijkstra's optimal route algorithm is then applied to determine the least-cost HSR paths between all bilateral destinations. In the second step, we use Kruskal's minimum spanning tree

algorithm to create a single continuous network, minimizing the overall construction cost. A more detailed description of the spanning tree instruments can be found in Appendix D.

Columns (3) and (12) of Table 1 present the results using the least-cost path instrument based on construction costs. The first-stage F-statistic of 35.4 indicates a strong correlation between the instrument and actual HSR placements. Notably, the 2SLS estimates are higher than the OLS estimates, even after controlling for city-specific factors. This discrepancy suggests that policymakers might have prioritized building HSR connections in prefectures with lower expected export performance or skill upgrading, potentially leading to an downward bias in the OLS results. Overall, the findings from the IV estimates remain robust: HSR connections significantly boost exports and enhance the skill content of those exports.

3.2.1 Robustness Checks

We perform several sensitivity checks to ensure the robustness of our results, including a placebo test using processing exports, an event study to verify parallel trends in export performance, and a falsification test to rule out the influence of unobserved stochastic trends.

Placebo Test Using Processing Exports. We differentiate a city's total exports into ordinary and processing regimes, using the processing regime as a placebo test. Processing exports are contract-based, involving Chinese firms collaborating with foreign partners to assemble final products from intermediate inputs. This process is less susceptible to the influence of accessibility to domestic input providers, which may be affected by HSR connections. We thus expect no significant impact of HSR connections on processing exports. Columns (4) to (6) of Table 1 demonstrate that HSR connections significantly enhance ordinary exports, consistent with the findings in columns (1) to (3) for total city exports. In contrast, columns (7) to (9) show that HSR connections have no impact on the export performance of processing trade firms.

Pre-trend Test. To check the parallel trends assumption, we run a variation of specification (1), incorporating the leads and lags of the initial connection dummies as controls:

$$\ln y_{it} = \sum_{m=2}^{3} \beta_m FirstConnect_{i,t-m} + \sum_{n=0}^{3} \beta_n FirstConnect_{i,t+n} + \eta X_{it} + \delta_i + \delta_t + \delta_{g(i)t} + \epsilon_{it} \quad (4)$$

where $FirstConnect_{i,t}$ is a dummy variable indicating whether city i first connected to the HSR network in year t, switching to unity only if the HSR line opens that year. The variable $FirstConnect_{i,t-m}$ represents the m-th lead, while $FirstConnect_{i,t+n}$ represents the n-th lag. Including leads controls for pre-HSR effects as a placebo, helping to distinguish anticipatory

effects from actual connection impacts. Controlling for lags allows us to trace the treatment effects following the HSR's introduction. We expect anticipatory effects to be negligible for exports, with a significant positive impact expected in the lags.

Figure 1 presents the results of an event study examining the impact of HSR connections on city exports and the skill content of these exports, based on equation (4). Here, T denotes the year when a city first connects to the HSR network, with $T-i$ and $T+i$ representing the years before and after the connection, respectively. The corresponding estimates are detailed in Appendix Table A.2. The insignificant coefficients for $T-i$ confirm the parallel trends assumption, ensuring that pre-existing trends in city exports were not influenced by the impending HSR connection. For city exports shown in Figure 1a, the estimated coefficients for $T+i$ are all positive. The impact of HSR becomes even clearer in Figures 1b and 1c, which categorize exports by regime. These figures further reveal a significant increase in ordinary exports following the HSR connection, while processing exports remain unaffected.

Randomization-based Test. To ensure the robustness of our findings and rule out the possibility that our results are driven by unobserved stochastic trends common to both HSR connections and exports, we conduct a falsification test. In this test, we randomly assign HSR connections to cities and replicate the specification from column (5) in Table 1 across 200 iterations. Appendix Figure A.2 displays the distribution of estimates from these simulations, with our baseline estimate highlighted by a vertical line for easy comparison. The simulated estimates are centered around zero, indicating no significant effect. In contrast, our benchmark estimate lies well beyond the 95th percentile of the distribution, providing strong evidence of a genuine relationship between HSR connections and exports.

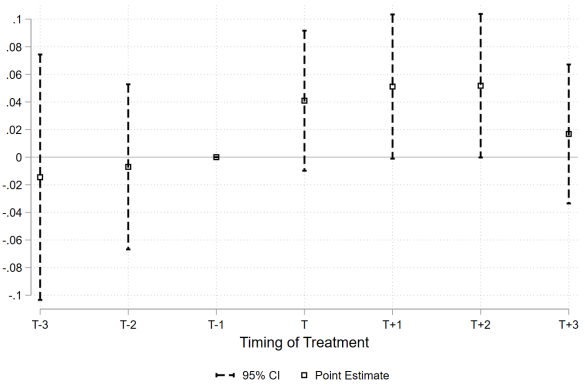
These findings confirm that HSR connections significantly enhance a city's ordinary exports and skill content, indicating a factor-biased transformative impact on the production structure. This shift occurs as firms gain access to HSR, allowing them to travel long distances more efficiently and conveniently.

3.3 Extended Analysis

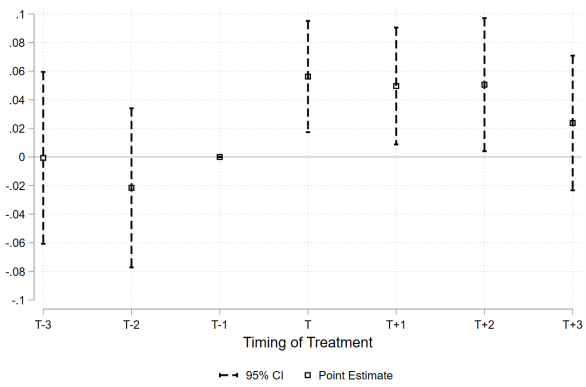
We proceed to analyze the potential mechanisms behind the surge in city exports and the heightened skill content of these exports following HSR connections, focusing on whether the time-varying changes across Chinese cities are driven by improved buyer-supplier matching efficiency, which facilitated the outsourcing of manufacturing tasks, and, in turn, boosted the demand for skilled labor.

Figure 1: Event Study of Connection to HSR

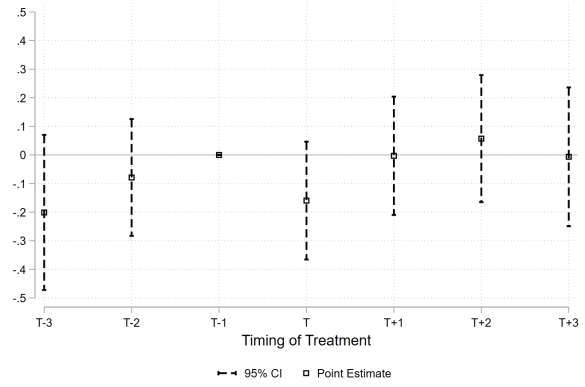
(a) City Exports



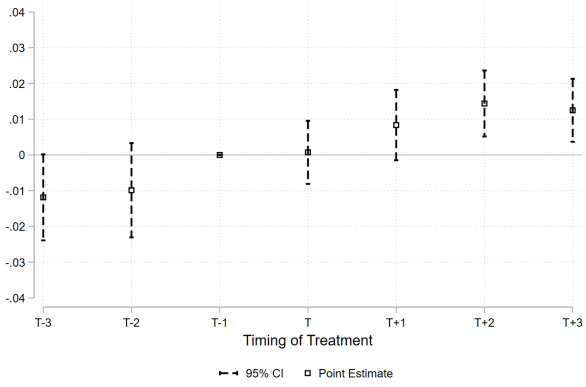
(b) Ordinary Exports



(c) Processing Exports



(d) Skill Content in Exports



Notes: The figures presents coefficients and 95% confidence intervals from the regression analysis of equation (4), with total city exports, ordinary exports, processed exports, and skilled content in exports as the respective dependent variables. We define T as the period when a city is first connected to the HSR network, with $T-i$ and $T+i$ representing the periods before and after the initial connection, respectively. Detailed estimates corresponding to this figure are reported in Appendix Table A.2.

Skill Demand. Columns (1) through (3) of Table 2 examine changes in the employment share of skilled workers—defined as those with education beyond high school—using data from the 2010 and 2015 Population Censuses. The analysis uses a long-difference specification, controlling for each city's historical employment patterns before the HSR network expansion. Column (1) presents results without city controls, showing that HSR-connected cities experienced an average 0.7 percent increase in the share of skilled employment from 2010 to 2015. However, this increase was not uniform. Columns (2) and (3), which include additional city controls, indicate that HSR connections disproportionately boosted the share of skilled employment in cities with initially higher levels of skilled labor, suggesting a pattern of HSR-induced specialization.

Table 2: Impact of HSR on Skilled Demand across Cities

Dep var:	First Difference Model ΔSkilledEmpShare, 2010 and 2015			Panel Regression Model SkillPremium, 2010 to 2018		
	(1)	(2)	(3)	(4)	(5)	(6)
$Connect_{i,t0}$	0.007*	0.004	-0.014**			
	(0.004)	(0.003)	(0.006)			
$Connect_{i,t0} \times SkilledEmpShare_{i,t0}$			0.226***			
			(0.074)			
$Connect_{it}$				0.492	0.664*	0.618*
				(0.312)	(0.361)	(0.346)
$Connect_{it} \times SkilledEmpShare_{i,t0}$						0.414
						(2.443)
Observations	206	205	205	226	224	224
R-squared	0.344	0.531	0.553	0.509	0.549	0.549
Province FE	Y	Y	Y	-	-	-
City FE	-	-	-	Y	Y	Y
Year FE	-	-	-	Y	Y	Y
Group-Year FE	-	-	-	Y	Y	Y
Other Controls	-	Y	Y	-	Y	Y

Notes: Columns (1) to (3) evaluate the impact of HSR on changes in the skilled employment share between 2010 and 2015, using a first-difference model with province-fixed effects. Columns (4) to (6) examine the effect of HSR on the skill wage premium from 2010 to 2018, using a panel regression model with fixed effects for cities, years, and group-years. Cities are grouped by population size: cities with fewer than 1 million residents are labeled as Group 1, those with 1 to 10 million residents as Group 2, and cities with more than 10 million residents as Group 3. Additional controls include average annual wage rates in the manufacturing and services sectors, tariff rates, population size, and internet coverage at the city level. Robust standard errors are clustered at the city level and are reported in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

Columns (4) to (6) of Table 2 analyze the impact of HSR connections on the skill wage premium. We estimate city-year-specific skilled premiums using data from the China Family

Panel Studies (CFPS), resulting in valid estimates for 224 city-year pairs.[16] The panel regression analysis indicates that HSR connections significantly raised the relative wage premium for skilled labor, widening the wage gap between skilled and unskilled workers. This finding suggests that cities connected by HSR have seen an increased demand for skilled labor relative to unskilled labor.

Firm Sourcing and Talent Utilization. Building on the findings of Bernard et al. (2019), who show that HSR connections reduce search and outsourcing costs, we explore how these impacts reshape firms' relative demand for skilled labor in China. We measure this demand using average firm wages, as detailed wage bills by labor type are unavailable. We also investigate how changes in outsourcing patterns affect key aspects of firm strategy and R&D intensity.

Columns (1) through (6) of Table 3 assess the impact of HSR connections on firm-level sourcing patterns. In columns (1) and (2), the outcome is the firm-year-specific sourcing distance, calculated as the weighted average distance to a firm's top five suppliers. This measure is derived from a matched sample of the China Tax Survey Data and the WIND financial dataset. Columns (3) and (4) examine the volume of intermediate input materials purchased, while columns (5) and (6) focus on the share of foreign inputs in total costs, both computed directly from the China Tax Survey Data (see Appendix C.2 for details). The results in columns (1) and (2) show that HSR connections significantly increase firms' sourcing distances, with the effect strengthening when firm-level controls are applied, suggesting an expansion of domestic sourcing networks, consistent with Bernard et al. (2019). Moreover, evidence from columns (3) through (6) indicates that firms in HSR-connected cities are purchasing more locally and relying less on foreign inputs, highlighting a shift toward domestic outsourcing driven by HSR connectivity.

Consistent with the results in Table 2, HSR connections directly boost firms' average wages in columns (7) and (8), which we use as a proxy for relative demand for skilled labor due to the absence of firm-level data distinguishing between skilled and unskilled employment. This increase aligns with the observed rise in R&D intensity in columns (9) and (10), measured by the share of R&D expenditure in total costs. These findings are in line with Charnoz et al. (2018), who demonstrates that improved travel times through HSR facilitate better management and functional specialization. Thus, HSR not only enhances

[16]The skill premium is the wage difference between skilled and unskilled workers, estimated using data from the CFPS, a nationally representative survey initiated in 2010 by the Institute of Social Science Survey at Peking University. We calculate it by regressing individual wages on a college education dummy, controlling for experience and age. The coefficient on this dummy indicates how much more skilled workers earn on average compared to their unskilled counterparts. For more details, see Appendix C.1.

productivity gained from better sourcing but also implicitly increases the demand for a more skilled workforce.

Table 3: Impact of HSR on Firm Sourcing and Talent Utilization, 2010-2018

	Firms' Sourcing Pattern									
Dep var:	Sourcing Distance		Total Materials Share		Imported Input Share		Wage		R&D Intensity	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
$Connect_{it}$	0.346**	0.856**	0.005***	0.004***	-0.008***	-0.008***	0.008***	0.014***	0.006**	0.005**
	(0.150)	(0.353)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)	(0.003)
Observations	602	185	2,832,869	2,549,346	151,339	148,977	3,204,559	2,858,058	131,702	130,380
R-squared	0.841	0.895	0.531	0.538	0.701	0.708	0.613	0.633	0.779	0.780
Firm FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Group-Year FE	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Other Firm Controls	-	Y	-	Y	-	Y	-	Y	-	Y

Notes: Each point estimate stems from a separate regression regression analysis. Columns (1)-(2) assess the impact of HSR on firms' weighted average sourcing distance from their five main suppliers, expressed in logarithmic transformation. Columns (3)-(4) examine the effect of HSR on material purchases, and columns (5) and (6) on imports, all calculated as the share of operational costs using the log(x+1) transformation. Columns (7)-(8) assess the impact on firms' average wage in logarithmic terms, and columns (9)-(10) on R&D expenses as the share of operation costs, using the log(x+1) transformation. All regressions incorporate fixed effects for firms, years, and group-years. Cities are categorized into three groups according to their population levels: those with populations below 1 million are labeled as 1, those with populations between 1 million and 10 million are labeled as 2, and cities with populations exceeding 10 million are denoted as 3. Other controls include employment and capital per employment at the firm level. Robust standard errors are clustered at the firm level and reported in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

4 Quantitative Model

Motivated by this empirical evidence, we extend the firm-to-firm trade model of Eaton et al. (2016) to quantify the aggregate and distributional effects of improved firm sourcing driven by HSR network expansion. Our extension makes task outsourcing contingent on the locations of both suppliers and buyers, capturing the heterogeneous impact of HSR connections on bilateral trade patterns. We further adapt the model to a multi-region context, incorporating individual migration decisions and frictional labor mobility.

4.1 Model Environment

The economy consists of N regions indexed by $i = 1, 2, ..., N$. Each region i is initially populated with two types of workers (skilled and unskilled, or white- and blue-collar), denoted as $\bar{L}_i^W$ and $\bar{L}_i^B$, respectively. Labor mobility across regions is allowed but subject to frictions that we specify later. Firms require services (e.g., management) and manufacturing tasks as inputs in the production process. Service tasks must be performed internally, while manufacturing tasks can be outsourced to other firms.[17] The final output can be used for consumption or as an intermediate input for manufacturing tasks in other firms, with these two uses being non-substitutable. Firms seeking inputs match randomly with sellers

[17]Alternatively, one could relax this assumption by allowing service tasks to be outsourced to a lesser extent than manufacturing tasks. This adjustment would not alter the main conclusion.

in different locations, and this matching process depends on matching efficiency at the bilateral city level. An improvement in matching efficiency between two locations (e.g., due to an HSR connection) allows firms to find more and better suppliers, decreasing marginal production costs and increasing overall production efficiency. Given substantial productivity differences across regions and labor mobility restrictions imposed by the Hukou system, we assume heterogeneous innate technology endowments across regions and frictional labor mobility to study the distributional effects of HSR among regions and different skill groups.

4.1.1 Producers

Technology. A firm j in location i can produce a quantity of output $Q_i(j)$ by combining two intermediate tasks: service tasks, denoted as S, and manufacturing tasks, denoted as M, according to the production function:

$$Q_i(j) = z_i(j) \prod_{k \in \{S,M\}} b_k^{-1} \left[\frac{\left(m_{k,i}(j)/\zeta_k\right)^{\zeta_k} \left(L_{k,i}/(1-\zeta_k)\right)^{1-\zeta_k}}{\beta_k} \right]^{\beta_k} \tag{5}$$

where $z_i(j)$ represents producer j's Hicks-neutral productivity, $m_{k,i}(j)$ denotes the input of task k, b_k is a constant to be specified later, and β_k is the Cobb-Douglas share of task k, which satisfies $\beta_k > 0$ and $\beta_S + \beta_M = 1$. To transform the input of task k into output, an assembly worker specific to that task ($L_{k,i}$) is required. We assume that labor ($L_{k,i}$) and task input ($m_{k,i}(j)$) are aggregated under Cobb-Douglas technology, with $1 - \zeta_k$ denoting the within-task assembly worker share.[18]

Intermediate tasks can be performed either by labor appropriate for that task or by an input produced by another firm. We assume skilled labor is required for the service task, and unskilled labor for the manufacturing task. If a firm uses labor for task k, it pays the wage $w_{k,i} \equiv w_i^{l(k)}$ for labor type $l(k) \in \{W, B\}$, where $k \in \{S, M\}$. The worker productivity in performing a task, $q_{k,i}(j)$, is worker-firm specific and is drawn from an extreme value distribution to be explained later. Producer j can also source intermediate inputs from suppliers to perform the tasks. Due to search frictions, buyers match with an integer number of suppliers. Variable profit is split via Nash bargaining, with buyers having all the bargaining power, as per Eaton et al. (2016). This implies that the price of intermediate input is driven down to its unit cost. We further assume that unskilled labor and intermediate inputs are perfect substitutes for performing the tasks.

The unit cost that firm j pays to perform task k depends on the labor wage and the

[18] Specifically, the production simplifies to $Q_i(j) = z_i(j) \prod_{k \in S,M} b_k^{-1} \left(\frac{m_{k,i}(j)}{\beta_k} \right)^{\beta_k}$ if ζ_k is set to unity, mirroring the production function proposed by Eaton et al. (2016).

availability of intermediate inputs. We denote this unit cost as $c_{k,i}(j)$, which is expressed as:

$$c_{k,i}(j) = \min\left\{\frac{w_{k,i}}{q_{k,i}(j)},\ c_{k,i}^{min}(j)\right\} \tag{6}$$

where $c_{k,i}^{min}(j)$ is the lowest price of the intermediate input available to firm j, resulting in the firm's marginal cost of delivering a unit of its output to destination n being expressed as $c_{ni}(j) = d_{ni}/z_i(j) \times \prod_{k\in S,M}\left[\left(c_{k,i}(j)^{\zeta_k} w_{k,i}^{1-\zeta_k}\right)^{\beta_k}/b_k\right]$. The iceberg transportation cost of delivering one unit of final output from i to n is denoted by $d_{ni} \geq 1$, with $d_{ii} = 1$ for all i. For model tractability, we follow the assumptions made by Eaton et al. (2016) regarding the distributions of productivity, labor productivity in performing tasks, and the distribution of intermediate input prices.

In particular, we assume that the measure of potential producers in each region i with productivity greater than z ($z_i(j) \geq z$) is given by

$$\mu_i^Z(z) = T_i z^{-\theta} \tag{7}$$

where T_i represents the endowment of technology in region i, and $\theta \geq 0$ is the shape parameter governing similarities—higher values of θ imply greater similarities. Additionally, the worker productivity in performing task k in firm j, $q_{k,i}(j)$, is assumed to follow an extreme value distribution with the cumulative distribution function (c.d.f.) given by

$$F(q) = \exp(-q^{-\phi}) \tag{8}$$

where $\phi \geq 0$ reflects the similarity of workers' productivity across tasks and firms, with the restriction $\phi \leq \theta$. Given equations (7) and (8), we will later demonstrate that the measure of sellers sourcing from i with a cost below c in destination n can be derived as:

$$\mu_{ni}(c) = \Upsilon_i d_{ni}^{-\theta} c^{\theta} \tag{9}$$

where Υ_i is an endogenous variable reflecting the overall capacity of the region i, including technology level, labor market conditions, trade barriers, and the efficiency of supplier search. In contrast, the specifications of equations (7) and (8) are primitives of the model, with T_i, θ, and ϕ being exogenous parameters.

Matching with Intermediate Input Suppliers. The matching between buyers and suppliers is random, which can be interpreted as coming into contact with intermediate input suppliers and relates to the suitability of the supplier's product for the buyer's needs.

The intensity with which a seller with a unit cost of c encounters a buyer in region n searching to perform task k is governed by the following equation:

$$e_{k,ni}(c) = \lambda_{k,ni}\mu_{ni}(c)^{-\gamma} \tag{10}$$

where $\mu_{ni}(c)$ denotes the measure of suppliers with a marginal cost below c, as defined in (9). The parameter γ, restricted to $\gamma \in (0,1)$, indicates the degree of congestion caused by lower-cost suppliers. This means that as more firms offer lower costs, the likelihood of matching with an appropriate buyer decreases. The parameter $\lambda_{k,ni}$ reflects the outsourcibility of task k, which improves with an HSR connection. In our model extension, we make this outsourcibility measure dependent on the bilateral locations of both buyers and sellers, whereas in Eaton et al. (2016), it depends unilaterally on the buyer's location. This extension aims to capture the heterogeneous impact of HSR connections on bilateral trade patterns. An HSR link increases $\lambda_{k,ni}$ by facilitating buyer efforts to find and supervise sellers, ensuring the quality of intermediate inputs. For simplicity, we assume service tasks cannot be outsourced, so $\lambda_{S,ni} = 0$. The number of quotes a buyer can receive for task k at a price no higher than c follows a Poisson distribution. The expected number of quotes per buyer, represented by the Poisson parameter $\rho_{k,ni}(c)$, is calculated by aggregating over all sellers from i to n as shown in (9) and (10):

$$\rho_{k,ni}(c) = \int_0^c e_{k,ni}(x)d\mu_{ni}(x) = \frac{\lambda_{k,ni}}{1-\gamma}\Upsilon_i^{1-\gamma}d_{ni}^{-\theta(1-\gamma)}c^{\theta(1-\gamma)} \tag{11}$$

With the restriction $\gamma \in (0,1)$, the expected number of price quotes below c increases as c grows. This means that many less-productive suppliers can perform the buyer's task, leading to a higher expected number of price quotes.

Firm j can perform task k at a cost no greater than c if both the labor cost and the lowest price quote for the intermediate input do not exceed c. The probability that firm j in region n does not encounter any suppliers offering a price lower than c for task k is given by $\prod_i \exp(-\rho_{k,ni}(c))$ according to the Poisson distribution. Conversely, the labor cost is no less than c if $w_{k,n}/q_{k,n}(j) \geq c$, and the corresponding probability is $F(w_{k,n}/c)$ as defined in (8). Assuming these two events are independent, the distribution of the lowest cost of performing task k in region n is:

$$G_{k,n}(c) = 1 - \exp\left[-\left(\sum_i \frac{\lambda_{k,ni}}{1-\gamma}\Upsilon_i^{1-\gamma}d_{ni}^{-\theta(1-\gamma)}c^{\theta(1-\gamma)} + w_{k,n}^{-\phi}c^{\phi}\right)\right] \tag{12}$$

We follow Eaton et al. (2016) by setting $\gamma \equiv (\theta-\phi)/\theta$, implying that the parameters affecting

cost heterogeneity due to labor efficiency are the same as those governing heterogeneity due to intermediate inputs for a given task and buyer, thereby simplifying the cost distribution for task k as the following equation:

$$G_{k,n}(c) = 1 - exp\left(-\Xi_{k,n}c^{\phi}\right) \tag{13}$$

where

$$\Xi_{k,n} \equiv v_{k,n} + w_{k,n}^{-\phi}, \quad v_{k,n} \equiv \sum_{i} v_{k,ni}, \quad v_{k,ni} \equiv \frac{\theta}{\phi}\lambda_{k,ni}\Upsilon_{i}^{\phi/\theta}d_{ni}^{-\phi} \tag{14}$$

The endogenous parameter $\Xi_{k,n}$ captures the technological efficiency of performing task k, influenced by both labor cost ($w_{k,n}^{-\phi}$) and intermediate input efficiency ($v_{k,n}$). Lower labor costs ($w_{k,n}$) and greater overall capacity (Υ_n) in region n imply a higher level of efficiency ($\Xi_{k,n}$) for task k. Additionally, a higher matching efficiency ($\lambda_{k,ni}$) in region i corresponds to greater production efficiency for task k.

The probability that a firm outsources task k to the lowest-cost supplier is given by $\bar{\omega}_{k,n} = v_{k,n}/\Xi_{k,n}$, while the probability of performing task k in-house by hiring workers is $1-\bar{\omega}_{k,n} = w_{k,n}^{-\phi}/\Xi_{k,n}$.[19] Unlike Eaton et al. (2016), our extended model defines the efficiency of intermediate inputs for task k in location n as the sum of bilateral efficiencies ($v_{k,ni}$, $\forall i$) from all other locations. The likelihood of outsourcing task k increases with improved bilateral matching efficiency (higher $\lambda_{k,ni}$) or higher labor costs ($w_{k,n}$).[20] Since service tasks cannot be outsourced ($\lambda_{S,n} = 0$), the labor share in these tasks is entirely one. The overall labor share in production in region i is derived as follows:

$$\beta_i^L = \sum_{k\in\{S,M\}} \beta_i^{l(k)} = \beta_S + \beta_M\left(1 - \zeta_M + \zeta_M\frac{w_{M,i}^{-\phi}}{\Xi_{M,i}}\right) \tag{15}$$

The variable $\beta_i^{l(k)}$ denotes the aggregate share of labor of type l in the total production cost in region i, defined as

$$\beta_i^{l(k)} = \beta_k\zeta_k w_{k,i}^{-\phi}/\Xi_{k,i} + \beta_k(1-\zeta_k) \tag{16}$$

Here, $w_{k,i}^{-\phi}/\Xi_{k,i}$ represents the labor share in producing intermediate inputs for task k in region i. Although the production function assumes a Cobb-Douglas form, the overall labor share is no longer constant and will depend on manufacturing wages and other factors.

[19] The detailed derivation is provided in Appendix E.1.

[20] It is important to note that the comparative statics on matching efficiency assume fixed wages and do not account for general equilibrium effects, which are fully addressed in the quantitative analysis.

Cost Distribution. We now consider the cost distribution for firm j in region i, with task-specific costs as shown in (13). The marginal cost depends on task efficiency as per (6). The expected number of firms from location i that can deliver to destination n at a cost not exceeding c is expressed in the following equation, derived in Appendix E.2:

$$\mu_{ni}(c) = T_i \Xi_i \mathbb{W}_i d_{ni}^{-\theta} c^{\theta} \tag{17}$$

where Ξ_i is an aggregation of the endogenous technology parameters $\Xi_{k,i}$ defined in (14), reflecting the efficiency gains from outsourcing. The variable $\mathbb{W}_i$ is an aggregation of labor costs. These are defined by the following equations:

$$\Xi_i \equiv \prod_{k \in \{S,M\}} (\Xi_{k,i})^{\theta \zeta_k \beta_k / \phi}, \quad \mathbb{W}_i \equiv \prod_{k \in \{S,M\}} w_{k,i}^{-\theta(1-\zeta_k)\beta_k} \tag{18}$$

Aggregating across all sourcing cities i, the measure of potential sellers in region n that can deliver a unit of goods at a cost below c is derived by the following equation:

$$\mu_n(c) = \sum_i \mu_{ni}(c) = \mathbb{Y}_n c^{\theta} \tag{19}$$

where the technology parameter $\mathbb{Y}_n$ is defined by $\mathbb{Y}_n \equiv \sum_i T_i \Xi_i \mathbb{W}_i d_{ni}^{-\theta}$. The expression for $\mu_{ni}(c)$ in (17) aligns with the specification in (9). From this definition, we see that the overall capacity of the region i depends on the exogenous technology endowment T_i, trade costs d_{ni}, wages $w_{k,i}$, and matching efficiency $\lambda_{k,in}$. By combining (14) and (18), we can determine the vector Υ_i by solving the fixed point of the following system of equations, in a manner similar to that employed by Eaton et al. (2016):

$$\Upsilon_i = T_i \mathbb{W}_i \prod_{k \in \{S,M\}} \left[\sum_n \frac{\theta}{\phi} \lambda_{k,in} \Upsilon_n^{\phi/\theta} d_{in}^{-\phi} + w_{k,i}^{-\phi} \right]^{\theta \zeta_k \beta_k / \phi} \tag{20}$$

In Appendix E.3, we demonstrate that Υ_i can be computed using an iterative procedure under our assumptions. Consequently, we can calculate the probability that a potential seller delivering to region n at a unit cost below c is from region i:

$$\pi_{ni} = \frac{\Upsilon_i d_{ni}^{-\theta}}{\sum_j \Upsilon_j d_{nj}^{-\theta}} \tag{21}$$

Assuming a continuum of producers, π_{ni} also denotes the share of purchases in destination n sourced from location i. In particular, when firms are prohibited from outsourcing manu-

facturing tasks, with $\lambda_{k,i} = 0$ for all k, and there is no requirement for within-task assembly workers, with $\zeta_k = 1$ for all k, the above trade share is consistent with the results of Eaton and Kortum (2002) and Eaton et al. (2011).[21]

Impact of HSR on the Demand for Skills. The above assumptions about technology are consistent with an aggregate production function for region i, which takes the form:

$$Q_i = L_{S,i}^{\beta_S} \left\{ \left[\left(\psi \left(L_{1M,i} \right)^{\frac{\phi}{\phi+1}} + (1-\psi) \left(I_{M,i} \right)^{\frac{\phi}{\phi+1}} \right)^{\frac{\phi+1}{\phi}} \right]^{\zeta_M} \left[L_{2M,i} \right]^{1-\zeta_M} \right\}^{\beta_M} \tag{22}$$

where ψ is a constant defined as $\psi \equiv 1/\left[1 + \Gamma(1+1/\phi)^{\phi/(1+\phi)}\right]$. The other variables are defined as follows: $L_{S,i}$ denotes skilled labor for service tasks, $L_{1M,i}$ represents unskilled labor for outsourced manufacturing tasks, $L_{2M,i}$ indicates unskilled labor for non-outsourced manufacturing assembly, and $I_{M,i}$ refers to the intermediates used in manufacturing.[22] The production function indicates that intermediate inputs and unskilled labor are substitutes in manufacturing, with their elasticity of substitution governed by $1+\phi$ ($\phi > 0$). Even with enhanced outsourcing due to HSR connections, a positive amount of unskilled labor, at least $\beta_M(1-\zeta_M)$, is always required for final output. This ensures a minimum share of unskilled labor is consistently employed, regardless of the outsourcing attractiveness.

Connecting to the HSR enhances the efficiency of interactions between firms, thereby reducing the average costs associated with outsourcing manufacturing intermediate tasks ($\bar{p}_{M,i}$). This has a straightforward positive impact on the demand for skilled labor: by lowering the cost of intermediates, it boosts overall production, which in turn unambiguously increases the demand for skilled workers. However, the impact on unskilled labor is more complex. Improved outsourcing reduces demand for unskilled labor in manufacturing tasks ($L_{1M,i}$) due to substitution with cheaper intermediates. Meanwhile, reduced production costs from HSR connections boost aggregate production, increasing the demand for unskilled labor in assembly tasks ($L_{2M,i}$). The net effect on unskilled labor depends on the balance between these opposing forces.

4.1.2 Consumers

Workers are risk neutral with linear preferences over a consumption index: $U_{odi} = C_{odi}$, where C_{odi} denotes the consumption index for worker i, born in region o and working in

[21] In the scenario where $\lambda_{k,i} = 0$ for all k, and $\zeta_k = 1$ for all k, we derive $\mu_n = \sum_i T_i \left(\bar{w}_i d_{ni}\right)^{-\theta} c^{\theta}$, where $\bar{w}_i$ is the average wage in region i, defined as $\bar{w}_i \equiv \prod_{k \in S,M} w_{k,i}^{\beta_k}$. Consequently, the trade share of region i in market n is given by $\pi_{ni} = T_i \left(\bar{w}_i d_{ni}\right)^{-\theta} / \sum_j T_j \left(\bar{w}_j d_{nj}\right)^{-\theta}$.

[22] The proof of the equivalence between the production side of the model and the aggregate production function is provided in Appendix E.4.

location d. The consumption bundle depends on the consumption of the service goods ($C_{S,odi}$); consumption of the manufacturing goods ($C_{M,odi}$); and an idiosyncratic shock that is specific to the individual worker and varies with the worker's birthplace and place of work (z_{odi}). This idiosyncratic shock captures the idea that workers may have idiosyncratic reasons for migrating to different regions. In particular, we assume that the service goods and manufacturing goods are destination-specific, i.e., $C_{k,odi} = C_{k,di}\ \forall k \in \{S, M\}$, and the aggregate consumption bundle is assumed to take the Cobb-Douglas form:

$$C_{odi} = \kappa z_{odi} C_{S,di}^{\alpha_S} C_{M,di}^{\alpha_M} \tag{23}$$

where $\kappa \equiv \prod_{k\in\{S,M\}} \alpha_k^{-\alpha_k} \times a_k^{-1}$ is a constant, α_k denotes the Cobb-Douglas share of need k, with $\alpha_k > 0$ and $\alpha_M + \alpha_S = 1$, and a_k is also a constant to be explained latter in Appendix E.5. On the consumer side, we differ from Eaton et al. (2016) in three key respects: consumers can only fulfill their demands by purchasing from firms; the matching process between consumers and firms is distinct from firm-to-firm matching; and consumers can migrate across regions, albeit with frictions. These features are crucial for applying the firm's matching model to study the HSR effect in China.[23]

The meeting of consumers and firms is subject to matching frictions. We propose that the frequency with which a firm charging price p in region d meets a consumer seeking to fulfill need k is

$$\tilde{e}_{k,d}(p) = \tilde{\lambda}_{k,d} \mu_d(p)^{-\tilde{\gamma}} \tag{24}$$

where $\tilde{\lambda}_{k,d}$ governs the matching efficiency between firms and consumers for need k in location d, $\mu_d(p)$ denotes the measure of active firms in region d as defined in (19), and $\tilde{\gamma}$ captures the impact of lower-priced firms on the matching ability of others. Aggregating across firms with different prices, the number of price quotes a consumer receives for need k at prices below p follows a Poisson distribution. The Poisson parameter, representing the average number of price quotes, is given by:

$$\tilde{\rho}_{k,d}(p) = \int_0^p \tilde{e}_{k,d}(x) d\mu_d(x) = \frac{1}{1-\tilde{\gamma}} \tilde{\lambda}_{k,d} \mathbb{Y}_d^{1-\tilde{\gamma}} p^{\theta(1-\tilde{\gamma})} \tag{25}$$

We require $\tilde{\gamma} < 1$ to ensure that the expected number of price quotes increases with unit cost. The probability that a consumer does not encounter a firm offering a price lower than

[23] For instance, separating consumer-to-firm matching from firm-to-firm matching allows the HSR connection to impact only the linkages among firms, making the model more realistic.

p is $\exp\left(-\tilde{\rho}_{k,d}(p)\right)$. Therefore, the price distribution for need k is expressed as:

$$\tilde{G}_k(p) = 1 - exp\left[-\tilde{v}_{k,d}p^{\theta(1-\tilde{\gamma})}\right] \tag{26}$$

where $\tilde{v}_{k,d} \equiv \frac{1}{1-\tilde{\gamma}}\tilde{\lambda}_{k,d}\mathbb{Y}_d^{1-\tilde{\gamma}}$ measures the ease with which a consumer can purchase a product at a low price.

For consumers of type $l \in \{W, B\}$, the indirect utility function, considering prices for each need k (p_S and p_M), is defined as:

$$V_{odi}^l(p_S, p_M) = \frac{w_d^l}{a_S a_M p_S^{\alpha_S} p_M^{\alpha_M}} \tag{27}$$

where w_d^l is the wage of a consumer of type l in location d, defined as $w_d^B \equiv w_{M,d}$ for blue-collar and $w_d^W \equiv w_{S,d}$ for white-collar consumers. The parameter a_k removes task-specific utility effects. Given price randomness, after perceiving taste shocks z_{odi}, consumer i focuses only on the destination-specific component of indirect utility in d, denoted as $\nu_d^l \equiv E\left\{V_{odi}^l(p_S, p_M)\right\}$. The indirect utility is derived in the following equation:

$$\nu_d^l = B \times w_d^l \times \mathbb{Y}_d^{\frac{1}{\theta}} \tag{28}$$

The first exogenous term in (28) captures how easily a consumer finds a firm to satisfy their demand, defined as $B \equiv \left[\prod_{k\in\{S,M\}} \tilde{\lambda}_k^{\alpha_k}\right]^{1/\theta(1-\tilde{\gamma})}$. This results from the assumption that $\tilde{\lambda}_{k,d}$ is common to all cities, i.e., $\tilde{\lambda}_{k,d} = \tilde{\lambda}_k$.[24] The second term, w_d^l, represents the income effect, indicating that cities with higher incomes provide greater utility for both types of workers. Additionally, $\mathbb{Y}_d$ encompasses all other factors influencing a region's amenities. All else being equal, higher levels of technology and greater matching probabilities between buyers and sellers lead to higher values of Υ_d, and consequently, higher $\mathbb{Y}_d$ and welfare levels. However, higher wages can reduce a region's attractiveness by lowering Υ_d and $\mathbb{Y}_d$, partially offsetting the income effect and decreasing overall welfare.

Migration Across Regions. Migration is modeled as a once-in-life choice. Following Ahlfeldt et al. (2015), we account for the heterogeneity in utility workers derive from living in different regions. Upon birth, workers develop idiosyncratic tastes for various regions and decide where to work, considering the destination-specific indirect utility and the migration

[24] The derivation is provided in Appendix E.5. The full expression of the first exogenous term is $B_d \equiv \left[\prod_{k\in\{S,M\}} \tilde{\lambda}_{k,d}^{\alpha_k}\right]^{1/\theta(1-\tilde{\gamma})}$. The simplifying assumption for this welfare term does not affect our main results and allows for a more streamlined expression.

cost. The migration cost, e^l_{od}, varies by skill type and origin-destination pair. Given their idiosyncratic tastes for regions, $\{z_{odi}|d = 1, 2, ..., N\}$, workers born in region o choose to live in region d to maximize welfare:

$$\max_{d\in\{1,2,..,N\}} \left\{ \frac{z_{odi}\nu^l_d}{e^l_{od}} \right\} \tag{29}$$

where $\nu^l_d = Bw^l_d \mathbb{Y}_d^{\frac{1}{\theta}}$ denotes the amenity-adjusted expected real wage in region d for workers of type l, born in o and working in d. The migration cost, interpreted as how workers of type l discount income from the destination, is similar to the iceberg cost assumption in international trade literature. Worker i of type $l \in \{B, W\}$, born in region o, will move to region d if and only if migration provides the highest utility, i.e., $\nu^l_d z_{odi}/e^l_{od} \geq \nu^l_g z_{ogi}/e^l_{og}$, $\forall g \in \{1, 2, ..., N\}$. Following Ahlfeldt et al. (2015), we assume $\{z_{odi}|d = 1, 2, ..., N\}$ are drawn from the Fréchet distribution, allowing for correlation in individual-specific tastes across regions. The vector of idiosyncratic taste shocks for any given worker follows the Fréchet distribution (denoted as $z_{di} = z_{odi}$ for simplicity) $F(z_{1i}, z_{2i}, ..., z_{Ni}) = exp\left[-\left(\sum_d z_{di}^{-\epsilon_l}\right)^{1-\rho}\right]$, where ϵ_l measures taste dispersion for workers of type l, and ρ reflects the inter-region correlation of taste draws.[25] Under this assumption, the probability that a worker of type l from origin o moves to destination d is derived in Appendix E.6:

$$\delta^l_{od} = \frac{(\nu^l_d/e^l_{od})^{\epsilon_l}}{\sum_g (\nu^l_g/e^l_{og})^{\epsilon_l}} \tag{30}$$

Denote L^l_d as the number of workers of type l working in d, and $\bar{L}^l_o$ as the number of workers of type l born in region o. The number of workers of type l moving from o to d is $L^l_{od} \equiv \delta^l_{od}\bar{L}^l_o$. Thus, the supply of labor of type l in region d is given by:

$$L^l_d = \sum_o L^l_{od} = \sum_o \delta^l_{od}\bar{L}^l_o \tag{31}$$

Welfare. Finally, the expected utility of a type l worker born in region o, denoted as $\bar{U}^l_o$, is derived in Appendix E.7:

$$\bar{U}^l_o = \left(M^l_o\right)^{\frac{1}{\epsilon_l}} \Gamma\left[1 - \frac{1}{(1-\rho)\epsilon_l}\right] \tag{32}$$

[25] This parametric assumption is adopted by Hsieh et al. (2013), Ahlfeldt et al. (2015), Bryan and Morten (2015), and Fan (2019). A lower value of ϵ_l signifies greater dispersion. When $\rho = 0$, taste shocks are uncorrelated across regions; when $\rho = 1$, they are perfectly correlated.

where $M_o^l \equiv \sum_g \left(\nu_g^l / e_{og}^l\right)^{\epsilon_l}$ measures the welfare of being born in location o. The better connected o is to other labor markets (lower e_{og}^l) and the more attractive nearby cities are (higher ν_g^l), the higher the utility for workers born in o.[26] Migration costs prevent utility equalization across regions, allowing for regional disparity analysis and policy impact assessment. The aggregate average welfare is:

$$W = \sum_{l \in \{B,W\}} \sum_{i=1}^{N} m_i^l \times \left(M_i^l\right)^{\frac{1}{\epsilon_l}} \Gamma \left[1 - \frac{1}{(1-\rho)\epsilon_l}\right] \tag{33}$$

where $m_i^l = \bar{L}_i^l / \left(\sum_{l \in \{B,W\}} \sum_{s=1}^{N} \bar{L}_s^l\right)$ is the original share of worker of type l in region i.

4.2 Aggregation and Equilibrium

We now solve for the aggregate equilibrium as described above. First, we determine the equilibrium for intermediate products, assuming labor costs are fixed. Then, we address the labor market equilibrium.

Goods Market. The aggregate production in location i includes the total revenue from selling consumption goods and the revenue from supplying intermediate inputs globally. With balanced trade, the total expenditure on final goods equals labor income, as there is no profit in this model. For region i, we have:

$$Y_i = \sum_{n=1}^{N} \pi_{ni} \left[X_n^c + \Phi_n^I Y_n\right], \quad X_n^c = \sum_{l \in \{B,W\}} w_n^l L_n^l \tag{34}$$

where $\Phi_n^I \equiv 1 - \beta_n^L$ represents the share of intermediate inputs in final production, and β_n^L denotes the overall labor share of production costs as defined in (15).

Labor Market. With balanced trade, the final expenditure X_i^C in region i is given by (34), and the equilibrium for labor of type $l \in \{B, W\}$ in region i implies:

$$w_i^l L_i^l = \beta_i^l Y_i \tag{35}$$

[26] The expected utility (32) doesn't depend on the destination d for workers from the same region, as their average welfare remains the same regardless of their location. Attractive destinations directly increase welfare by enhancing utility based on idiosyncratic taste, but they also attract workers with lower taste draws, balancing the average utility. With Fréchet-distributed taste shocks, these effects cancel out, making a worker's welfare depend solely on birthplace characteristics.

where β_i^l represents the production cost share of labor type l in region i, as given in (16), and L_i^l is the endogenous supply of workers of type l in region i, provided in (31). These equations for each labor type l in region i determine the wage w_i^l.

Finally, the competitive equilibrium, defined in Appendix E.8, is reached when workers' migration and consumption decisions, firms' production choices, and the clearing of labor and goods markets are all optimal and consistent with endogenous labor supply and expenditure patterns. To quantify the aggregate and distributional effects of HSR, we further employ the exact hat algebra method (Allen et al., 2020) to solve for the relative changes in equilibrium, as outlined in Appendix E.9.

5 Model Calibration

Before conducting counterfactual policy experiments, we first calibrate the model to the 2011 equilibrium and then analyze the impact of the HSR network expansion from 2011 to 2017.[27]The goal is to quantify the aggregate and distributional effects of HSR connections on welfare and skill disparity. Our analysis includes 101 geographic units: the top 100 export cities in China as of 2010, which together accounted for 94.4% of China's total exports, and the rest of the world. This section outlines the steps taken to calibrate the key parameters across these economic units.

5.1 Initial Equilibrium Conditions

Quantifying the model requires several key inputs, including bilateral trade shares (π_{ni}) and intermediate input shares ($v_{M,ij}/v_{M,i}$) across Chinese cities and between them and the rest of the world, inter-city bilateral migration flows by skill level (L_{od}^W, L_{od}^B), city-specific outsourcing shares ($\bar{\omega}_{M,i}$), and finally, the value-added for each task ($w_{M,n}L_{M,n}, w_{S,n}L_{S,n}$).

Trade Flows. The bilateral trade shares for the top 100 export cities in 2011 are derived from Taobao, one of China's largest e-commerce platforms. We aggregate the trade flow based on the origin and destination cities of delivery from all transaction records on the platform.[28] Since the trade flows from Taobao represent only online transactions within this platform, we scale these flows using the ratio of domestic trade to total transactions on Taobao.[29] To identify trade in intermediate inputs, we employ an estimated city-sector

[27]We chose 2011 and 2017 as the beginning and ending periods for our quantitative exercises because these are the earliest and latest years for which we have inter-city bilateral trade data.

[28]Taobao consists of a Business-to-Consumer (B2C) platform (Tmall.com) and a Consumer-to-Consumer (C2C) platform (Taobao.com). Our analysis combines all transactions from both platforms.

[29]According to our estimates, based on national input-output data from the National Bureau of Statistics of China and export data from the Chinese Statistical Yearbook, the total domestic trade volume for the

level inter-city input-output table for 2012 in China, following the methodology developed by Zheng et al. (2022), as used in Zhong et al. (2024).

Trade between Chinese cities and the rest of the world (ROW) in 2011 is calculated using China Customs Data, which provides detailed city-level export and import records by 8-digit HS code. We classify goods using the Broad Economic Categories (BEC) to differentiate between final consumption goods and intermediate inputs, aggregating them at the city level. Trade volumes within the ROW (excluding China) are obtained from the BACI trade database of 2011 (Gaulier and Zignago, 2010), which details bilateral trade flows across 200 countries at the HS 6-digit level, also aggregated into consumption and intermediate categories.

Migration Flows. Bilateral migration flows are computed using the 2010 Population Census, which provides detailed information on respondents' registered (Hukou) residence and current residence at the county (or district) level, along with their skill level. In practice, individuals with a degree above college are classified as skilled workers. We aggregate the number of migrants by origin and destination at the bilateral city level, resulting in migration flows for skilled and unskilled workers. Population data for the rest of the world are sourced from the World Bank Open Data. We assume no migration flows between Chinese cities and the rest of the world.

Outsourcing Share. The outsourcing share is calibrated based on the production costs of intermediate inputs and labor for each city. This estimation uses firm-level data on labor compensation and intermediate input purchases from the 2011 Chinese Tax Survey. Focusing on manufacturing firms, we aggregate the production costs of intermediate inputs and labor at the city level. The outsourcing share is then calculated as the ratio of intermediate input costs to the combined total of intermediate and labor costs.

Value-added by Task. The value added by manufacturing and service tasks is indirectly estimated, as the contributions of skilled and unskilled labor within a firm are not directly observable. This estimation relies on inter-city trade flow data and model parameters. The labor market clearing condition in equation (35) uniquely determines the value added for each task. Total production for region i is calculated by summing its trade flows, while the labor value-added share for each labor type l is computed using equation (16), incorporating the calibrated outsourcing share, Cobb-Douglas shares for service and manufacturing tasks, and the proportion of assembly workers in manufacturing.

top 100 export cities was 68.2 trillion CNY in 2011, while Taobao's total transaction volume for those cities was 0.23 trillion CNY. This yields a ratio of domestic trade to Taobao transactions of 294.

Other Model Parameters. The Cobb-Douglas share for the service (manufacturing) task is based on the 2011 value-added share of services in China, estimated at 0.44 (0.56) by World Bank Open Data. We calibrate the share of assembly workers using data from the 2012 CFPS, which provides detailed work profiles of surveyed individuals, including their wages, employment sectors, and occupations classified under the International Standard Classification of Occupations (ISCO-88). We classify individuals in the manufacturing sector as assembly workers, specifically those with 3-digit ISCO codes: 723 (Machinery mechanics and fitters), 724 (Electrical and electronic equipment mechanics and fitters), 828 (Assemblers), and 829 (Other machine operators not elsewhere classified). We aggregate wages to calculate labor costs for both assembly workers and the broader manufacturing workforce, determining the assembly worker share as the proportion of total labor costs attributed to this group. Our estimates indicate that assembly workers account for 25% of manufacturing labor costs ($\zeta_M = 0.75$), while we assume their share in the service sector is zero ($\zeta_S = 1$).

Last, we adopt some parameters from established literature. In particular, we set the shape parameter of the Fréchet distribution for firm productivity (θ) to 4, aligning with the preferred estimate from Simonovska and Waugh (2014). The matching parameter (γ) is set at 0.34, based on estimates from Eaton et al. (2022), who provide a strong foundation for this choice. For the correlation parameter (ρ) and the dispersion parameters (ϵ_B and ϵ_W) of the Fréchet distribution for workers' migration preferences, we rely on estimates from Fan (2019). These parameters are calibrated to $\rho = 0.36$, $\epsilon_B = 3.9$, and $\epsilon_W = 4.3$, as detailed in their study. A summary of the parameters used is provided in Appendix Table A.3.

5.2 Efficiency Change from HSR Connection

As the model outlines, HSR connections boost a location's ability to outsource manufacturing tasks to other regions while also making it a more attractive destination for outsourcing from other locations. These effects are captured by changes in the parameters $\lambda_{M,ni}$ and Ξ_i whenever either i or n is connected to the HSR network. Since these parameter changes are not directly observable, we infer the impact of HSR connections on task efficiency by combining auxiliary model equations with reduced-form estimates.

Specification of Changes in Task Efficiency. To capture the impact of HSR connections, we approximate the change in task efficiency (Ξ_{it}) in the following reduced-form relationship:

$$\ln \Xi_{it} = \ln \Xi_{it_0} + \beta \times Connect_{it} + \delta_t^{\Xi} + \epsilon_{it}^{\Xi} \tag{36}$$

where Ξ_{it_0} represents the initial level of task efficiency before the HSR connection. The term $\beta \times Connect_{it}$ reflects the change in task efficiency due to a city's HSR connection, while δ_t^{Ξ}

and ϵ_{it}^{Ξ} account for common trends and other factors affecting local task efficiency.[30] This specification suggests that, after controlling for city and time-fixed effects, the idiosyncratic shocks—including technological progress across locations—are independent of a city's HSR connection.

Specification of Technological Change. To account for the technological changes that occur during the expansion of the HSR network, we assume that a city's overall technological progress is governed by the following law of motion:

$$\ln T_{it} = \delta_i^T + \delta_t^T + \eta Z_{it} + \epsilon_{it}^T \tag{37}$$

where δ_i^T captures time-invariant factors specific to each city, and δ_t^T represents time-varying shocks that are common across all cities. The vector Z_{it} includes city-level control variables, such as population and access to ICT facilities. The term ϵ_{it}^T accounts for random shocks that are unique to each city and vary over time.

Parameterization of Foreign Trade Costs. We parameterize the trade cost $d_{ROW,it}$ as follows:

$$\ln d_{ROW,it} = \delta_i^d + \delta_t^d + \xi \times Tariff_{it} + \epsilon_{it}^d \tag{38}$$

where δ_i^d accounts for time-invariant trade barriers, such as a city's geographical distance from global markets, and δ_t^d captures time-varying factors, including global economic conditions or shifts in international trade policies that impact all cities. The term $Tariff_{it}$ represents the tariff rate specific to city i in year t, while ϵ_{it}^d captures all other idiosyncratic shocks affecting a city's trade costs when exporting to foreign markets.

Consistency with Reduced Form Regression. Returning to the model, the total exports of city i at time t are expressed as $y_{it} = \pi_{ROW,it} \times X_{ROW,t}$, where $\pi_{ROW,it}$ represents the share of city i in the total expenditure of the ROW (i.e., $X_{ROW,t}$). According to (21), the logarithm of city exports can be written as:

$$\ln y_{it} = \ln \Xi_{it} + \ln T_{it} + \ln \mathbb{W}_{it} - \theta \ln d_{ROW,it} + \delta_t^X \tag{39}$$

where δ_t^X is a time trend term defined as $\delta_t^X \equiv \ln X_{ROW,t} - \ln \sum_j \Upsilon_{j,t} d_{ROW,jt}^{-\theta}$. The labor cost parameter $\mathbb{W}_{it}$ is given by $\ln \mathbb{W}_{it} \equiv \gamma_M \ln w_{M,it} + \gamma_S \ln w_{S,it}$, where $\gamma_k = -\theta(1-\zeta_k)\beta_k$ for

[30] As indicated by equation (14), changes in technology parameters within the city or in other cities can also affect local task efficiency.

$k \in \{M, S\}$. Hence, by combining equations (36) to (38), we can rewrite (39) as follows:

$$\ln y_{it} = \beta \times Connect_{it} + X'_{it}\Gamma + \delta_i + \delta_t + \epsilon_{it} \tag{40}$$

where $\delta_i = \ln \Xi_{it_0} + \delta_i^T - \theta\delta_i^d$ and $\delta_t = \delta_t^{\Xi} + \delta_t^T + \delta_t^X - \theta\delta_t^d$ captures the city- and time-fixed effect, respectively; the variable $X_{it} = \{Z_{it}, \ln \mathbb{W}_{it}, Tariffit\}$ includes relevant covariates, and the error term $\epsilon_{it} = \epsilon_{it}^{\Xi} + \epsilon_{it}^d - \theta\epsilon_{it}^d$ is the white noise. The OLS reduced-form regression precisely reflects the model-implied estimation equation for HSR's impact. By matching the empirical specification in equation (1), we recover the change in task efficiency due to HSR connections as $\hat{\Xi}_i = \exp(\hat{\beta} \times Connect_i)$. We use the IV estimate of β, valued at 0.150, as reported in column (6) of Table 1.

5.3 Other Technology Shock

In addition to the gains in production efficiency from improved supplier access via the HSR network, other technological changes may also impact the overall production efficiency parameter. To account for these factors, we estimate the changes in the endogenous variable that reflects the location's overall capacity ($\hat{\Upsilon}_i$), capturing both the effects of HSR and any additional technological shocks.

Parameterization of Domestic Trade Costs. To do this, we first calibrate the bilateral domestic trade costs between Chinese cities. Our approach follows Fan (2019), using bilateral trade flow data aggregated from Taobao. We parameterize the trade cost from city i to city j as follows:

$$\ln d_{ij} = \sum_{k=1}^{3} \beta_k I_{ij}^k + \sum_{k=1}^{3} \beta_{k+3} \times I_{ij}^k \times \ln Dist_{ij} + \mu_{ij} \tag{41}$$

where I_{ij}^1, I_{ij}^2, and I_{ij}^3 classify whether cities i and j are in the same province, in different provinces with a shared boundary, or in different provinces without a shared boundary, respectively. These mutually exclusive dummy variables capture structural differences in trade barriers across regions. The variable $Dist_{ij}$ represents geographic trade barriers, measured as the great-circle distance between i and j.

Estimating Overall Efficiency Change with Fixed Effects. We take the trade share from (21), dividing π_{ij} by π_{ii} and taking the logarithm. Substituting the specification for bilateral domestic trade costs from (41) gives us the following equation:

$$\ln \frac{\pi_{ij}}{\pi_{ii}} = \sum_{k=1}^{3} \tilde{\beta}_k I_{ij}^k + \sum_{k=1}^{3} \tilde{\beta}_{k+3} \times I_{ij}^k \times \ln Dist_{ij} + \underbrace{\ln \Upsilon_j - \ln \Upsilon_i}_{\text{Fixed Effects}} + \epsilon_{ij} \tag{42}$$

We estimate the regression with origin ($\ln \Upsilon_j$) and destination ($\ln \Upsilon_i$) dummies to obtain annual fixed effects. We then recover $\hat{\Upsilon}_j$ as the ratio of the fixed effects for 2017 to those for 2011.[31]

6 Quantitative Results

With estimated parameters, we use the quantitative model as the laboratory to study the aggregate and distributional effects of HSR placement in a period between 2011 and 2017.[32] We first present results showing the performance of the model at predicting key data moments, then conduct various policy experiments.

6.1 Model Fit

We start by highlighting the model's accuracy in predicting key macroeconomic variables. Panel (a) of Appendix Figure A.4 compares the model's predicted city-level export growth with actual data from 2011 to 2017. Each point represents a city, with the x-axis depicting the model's predicted export change and the y-axis showing the actual change during the same period. The dashed line, with a slope of 0.112 and a t-statistic of 2.574, reveals a significantly positive correlation, indicating that the model effectively captures variations in city-level export growth. Panel (b) compares the model's predicted changes in bilateral inter-city trade flows with corresponding data from Taobao over the same period. The strong correlation further highlights the model's capability to capture shifts in inter-city trade dynamics. Panel (c) compares changes in a city's export skill content, measured as skilled workers per unit of total export sales. The model effectively predicts future skilled labor demand in exports, a moment not explicitly targeted in calibration, further affirming its robustness.

6.2 Aggregate Impact of HSR

Welfare Effect. Welfare gains from HSR are calculated by solving the system of equations in Appendix equation (A.60). To isolate the gains from HSR expansion, we compare China's equilibrium before and after HSR implementation, holding technological conditions constant between 2011 and 2017. Welfare is measured by compensating variation as a percentage of 2011 GDP, with findings shown in column (1) of Table 4. HSR connectivity yields a 4.27

[31] Estimation results are provided in Appendix Table A.4, and Appendix Figure A.3 displays the distribution of $\hat{\Upsilon}_j$ across cities. These estimates reflect shifts in broader economic conditions, including technological advancements ($\hat{T}$) and improved outsourcing capabilities in manufacturing due to HSR connections ($\hat{\lambda}$).

[32] Using the estimated parameter change, $\hat{\Upsilon}_i$, which accounts for all possible shifts in technology and outsourcing capabilities driven by HSR, we solve for changes in other endogenous variables through the system of equations detailed in Appendix E.9.

percent increase in national welfare, though these gains are unevenly distributed across cities. Skilled labor benefits significantly, with a welfare increase of 6.35 percent, while unskilled labor faces a decline of 0.68 percent. The decline in unskilled workers' welfare arises from lower intermediate task costs due to HSR connectivity. Enhanced matching efficiency boosts task outsourcing and reduces demand for unskilled labor, harming their welfare. Column (2) of Table 4 presents the welfare change accounting for technological advancements ($\hat{T}$) and enhanced outsourcing from HSR ($\hat{\lambda}$). The total welfare gain is estimated at 12.67 percent, with skilled labor benefiting more significantly at 14.12 percent compared to 9.2 percent for unskilled labor. Column (3) quantifies HSR's contribution to the overall welfare changes noted in column (2). HSR accounts for about 34% of the total welfare gain overall from 2011 to 2017, with a more significant impact on skilled workers, who see around 45% of the welfare increase attributed to HSR. Conversely, unskilled workers experience a negative welfare change relative to overall economic improvements.

Table 4: Welfare Impact of HSR Network (2011-2017)

Specification:	(1) HSR ($\hat{\lambda}$)	(2) HSR + Technological Change ($\hat{\lambda}$+$\hat{T}$)	(3) HSR Contribution: (1)/(2)
Unskilled Worker	-0.68%	9.20%	-7.36%
Skilled Worker	6.35%	14.12%	44.97%
Whole Population	4.27%	12.67%	33.73%

Notes: The table illustrates the percentage changes in welfare attributable to HSR connectivity ($\hat{\lambda}$) and the broader effects of improvements in economic conditions, including both technological advancements ($\hat{T}$) and improved outsourcing capabilities from HSR ($\hat{\lambda}$). Column (3) highlights the contribution of enhanced matching efficiency due to HSR, relative to the broader economic shifts.

Impact on Matching Efficiency. We also estimate the change in matching efficiency across cities. While the welfare effect of HSR does not depend on these estimates or their parameterization, for simplicity and without loss of generality, we assume that bilateral matching efficiency can be expressed as the product of the unilateral matching efficiencies of origin-destination city pairs ($\lambda_{M,ni} = \lambda_{M,n} \times \lambda_{M,i}$). Using this parameterization, we recover the changes in matching efficiency parameters ($\hat{\lambda}$), with detailed steps provided in Appendix E.10 and the results reported in Appendix Figure A.5. Additionally, we present histograms of the estimated changes in matching efficiency for cities with and without HSR connections as of 2017. As illustrated in Appendix Figure A.6, for cities without HSR, the average estimated $\hat{\lambda}$ is 1, indicating virtually no change in matching efficiency. In contrast, cities connected to the HSR network show a significant improvement, with an average $\hat{\lambda}$ of 1.30. The t-test for the difference in average matching efficiency change between the two city

groups yields a T-statistic of -56, confirming a significant difference.

The improved matching efficiency between buyers and suppliers facilitated by HSR connections reduces the cost of outsourcing intermediate inputs, making outsourcing a more attractive option for firms compared to hiring local unksilled workers. Part (a) of Appendix Table A.5 summarizes the reduction in intermediate input costs resulting from the introduction of HSR.[33] We categorize cities into four major regions: Northeast, East, Central, and West. The Northeast and Central regions, on average, experience more significant cost reductions in intermediate inputs due to the HSR connection, while the East benefits less. This diminished impact in the East is largely due to its stronger reliance on exports and imports, which lessens the significance of domestic inter-city trade. Overall, the HSR connection lowers the cost of intermediate inputs, equivalent to an 11.37 percent reduction in iceberg trade costs.

Impact on Technology, Intra- and International Trade. To what extent does improved matching efficiency contribute to a city's overall technological efficiency? We explore this question using a variance decomposition approach to evaluate HSR's impact on regional technology, city exports, and inter-city trade within China. We follow the method of Eaton et al. (2004), regressing each outcome variable under a counterfactual scenario with only matching efficiency changes against the corresponding outcomes under a scenario that incorporates all changes in economic fundamentals. This enables us to isolate and quantify the specific contribution of improved supplier-buyer connections relative to broader economic trends, which are reported in Appendix Table A.6. According to our estimates, HSR explains 45% of the technological changes across cities from 2011 to 2017 and 56% of the variations in city exports. However, HSR accounts for a relatively smaller proportion—25%—of the total variation in domestic trade flows among Chinese cities.[34] Nonetheless, the significant proportions explained by HSR in other areas highlight its transformative role in shaping China's economic landscape.

6.3 Distributional Impact of HSR

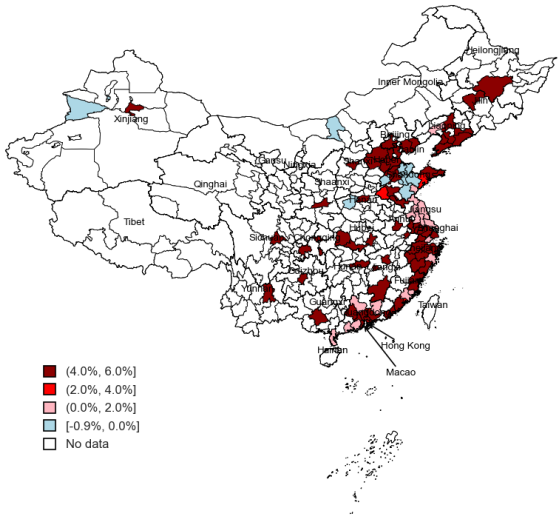
Skill-biased Impact of HSR Connections. However, the benefits of being connected to the HSR network are not evenly distributed. Figure 2a shows the geographic variation in welfare gains resulting from the HSR implementation. Areas shaded in cool tones (blue)

[33] We calculate the average unit cost of intermediate inputs sourced from region i derived in Appendix E.4, which is proportional to $v_{M,i}^{-1/\phi}$. A detailed map of these changes is reported in Appendix Figure A.7.

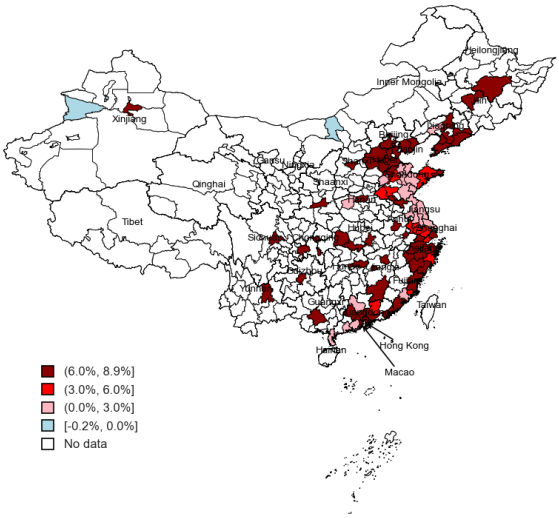
[34] The smaller impact on domestic trade could be attributed to the fact that domestic trade patterns are influenced by a broader set of factors, such as existing logistics networks, regional trade policies, and varying levels of industrial specialization across cities.

Figure 2: Changes in Welfare by Skill Type and in Skill Premium across Chinese Cities

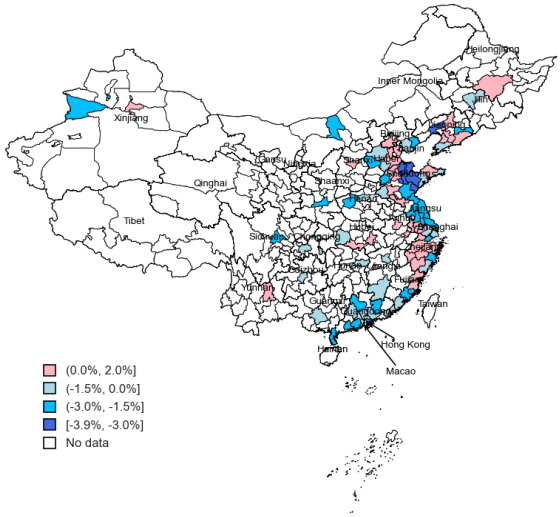
(a) Overall Welfare Change



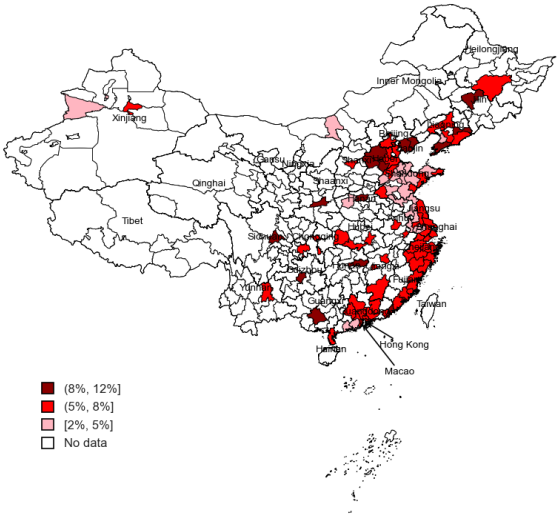
(b) Welfare Change for Skilled Workers



(c) Welfare Change for Unskilled Workers



(d) Change in Skill Premium



indicate negative welfare changes, while warm tones (red) represent positive ones. Provincial capitals stand out as the biggest beneficiaries of HSR integration, with Beijing leading as the largest "winner", followed by Wuhan, Changsha, and Shenyang. Although some cities without direct HSR links experience positive spillover effects, their welfare gains are modest, and in certain regions—such as Shandong, Hebei, Henan, and Xinjiang—some areas even suffer welfare losses.

To identify the sources of unequal gains, we calculate welfare changes by skill type across cities, as shown in Figures 2b and 2c. The disparity is largely driven by the gap between skilled and unskilled workers, with most welfare gains concentrated among the skilled. Meanwhile, unskilled workers see only modest improvements and, in many east coast regions, even experience notable welfare declines. This skill-biased impact of HSR connections is further reflected in the rising skill premium, as illustrated in Figure 2d and detailed in part (b) of Appendix Table A.5. On average, skilled workers experience substantial welfare gains, with a mean increase of 5.64%, while unskilled workers see an average welfare loss of -0.75%. This disparity is further underscored by the skill premium change, which shows an average increase of 6.17%. HSR-induced outsourcing prompts firms to shift unskilled labor-intensive tasks to lower-cost regions, lowering production costs and boosting outsourcing demand. This increased demand for outsourcing elevates the need for skilled labor, resulting in higher wages and improved welfare for skilled workers.

Overall Inequality. The skill-biased impact of HSR has significantly amplified rising regional inequality, largely driven by the uneven distribution of skilled labor across regions. To measure this, we compute the Theil index, breaking it down into within-city and between-city, as well as within-skill and between-skill, components. As shown in part (a) of Table 5, 58% of the overall increase in inequality following HSR integration is attributed to between-city disparities. Part (b) further decomposes the results by skill type. Consistent with part (a), between-city differences dominate, while HSR integration exacerbates inequality both within and between skill types. Within-skill disparities contribute 56% to the rise in inequality, with between-skill disparities within cities contributing the remaining 44%. These findings suggest that HSR infrastructure disproportionately benefits regions with higher concentrations of skilled labor, further widening the inequality gap both between cities and among different worker groups within cities.

6.4 The Role of Domestic Trade and Migration Costs

Thus far, we've shown that HSR improves welfare but widens the income gap between skilled and unskilled workers. We now investigate how domestic migration and trade costs

Table 5: Impact of HSR on Inequality

Part (a): Inequality across Cities	Within-city	Between-city	Theil Index
NO HSR Connection	0.390	1.272	1.662
HSR Connection	0.402	1.289	1.692
Percentage Change in Inequality	3.2%	1.4%	1.8%
Relative Contribution to Theil Index Change	42.1%	57.9%	100.0%
Part (b): Inequality by Skill Type	**Within-skill**	**Between-skill**	**Theil Index**
NO HSR Connection	1.278	0.383	1.662
HSR Connection	1.295	0.397	1.692
Percentage Change in Inequality	1.3%	3.5%	1.8%
Relative Contribution to Theil Index Change	55.9%	44.1%	100.0%

Notes: Part (a) presents the decomposition of the Theil index into within-city and between-city components for economies with and without HSR connections. Part (b) repeats this decomposition, but across different worker skill types.

shape HSR's impact on welfare and inequality. This analysis seeks to inform policies that sustain HSR's benefits while addressing its unequal effects.

Labour Mobility. We begin by estimating the impact of HSR under different labor mobility conditions, keeping all model parameters fixed at their calibrated values for HSR scenarios. We adjust migration costs for all worker types from 0.7 to 1.3 relative to their initial values. Panel (a) of Appendix Figure A.8 shows how HSR connections impact welfare compared to the baseline scenario with calibrated migration costs ($\hat{e}^l_{od} = 1$). As migration costs increase, wage disparities across regions widen, which intensifies firms' incentives to outsource manufacturing tasks to low-wage areas with abundant unskilled labor. This outsourcing boost raises the relative demand for unskilled labor in these regions, leading to greater welfare gains for unskilled workers as migration costs and wage disparities grow. In contrast, skilled workers benefit relatively less when their mobility is more restricted.

This pattern is evident in panel (b) of Appendix Figure A.8, which shows how inequality changes with migration costs. As outsourcing incentives and labor cost disparities intensify, HSR strengthens outsourcing and increases the demand for unskilled labor in lower-wage regions. This increased demand drives up wages for unskilled workers in those regions, reducing inequality within cities. However, this change exacerbates disparities between cities, as the wage gap between skilled and unskilled workers widens across cities with varying concentrations of skilled and unskilled labor. At the national level, welfare outcomes follow a U-shaped relationship with migration costs. Unskilled workers gain more when welfare disparities are larger, as cities with a higher concentration of skilled labor outsource more,

while skilled workers benefit when disparities are smaller, where labor cost differences are more uniform across regions and outsourcing is less attractive.[35]

Domestic Trade Costs. Finally, we examine how HSR interacts with other infrastructure that affects trade costs for shipping goods. Following a similar approach to our analysis of labor mobility, we adjust domestic trade costs, $\hat{d}_{ni}$, to values between 0.7 and 1.3. Panel (c) and (d) of Appendix Figure A.8 present the effects on welfare and inequality resulting from HSR connections, compared to the baseline scenario where trade costs remain at their initial level in the Chinese economy ($\hat{d}_{ni} = 1$). HSR connections reduce outsourcing costs, and the magnitude of this reduction depends on domestic trade costs. When these costs are high—reflecting elevated shipping costs for manufacturing tasks—the cost-saving effects of HSR are more substantial. This is because improved matching efficiency not only directly lowers outsourcing costs but also encourages more supplier entry, further enhancing the reductions, particularly when outsourcing is initially limited.[36] Stronger outsourcing increases the relative demand for skilled labor while reducing demand for unskilled labor, as evidenced by rising welfare gains for skilled workers and declining welfare for unskilled workers as trade costs rise. This shift drives an increase in overall inequality, primarily due to within-city (between-skill) disparities, as reflected in the rising Theil index.

7 Conclusion

Improved production linkages through more efficient buyer-seller matching can accelerate the outsourcing of manufacturing tasks, increasing demand for skilled labor. In this paper, we investigate how modern infrastructure investments in HSR systems, which facilitate faster passenger transportation, affect skill demand and produce significant aggregate and distributional impacts across Chinese cities. By enabling more efficient business travel, HSR enhances production networks and encourages domestic outsourcing. This shift reduces the demand for unskilled labor in easily outsourced tasks while increasing the need for skilled labor in tasks retained within firms. Our study makes several key contributions to the understanding of the impacts of large-scale infrastructure investments through the lens of firm-to-firm linkages.

Our first contribution is to show that HSR enhances production networks by enabling more frequent and efficient interactions between firms, reshaping firm boundaries, and driv-

[35] Rising heterogeneity in labor costs, driven by higher migration barriers, prompts cities with an abundance of skilled labor to increasingly outsource activities to those with a surplus of unskilled labor. This dynamic is illustrated in panel (a) of Appendix Figure A.9.

[36] As a result, reduced outsourcing costs lead to a greater share of outsourcing across all cities, as shown in panel (b) of Appendix Figure A.9, which demonstrates the rise in outsourcing share as trade costs increase.

ing production efficiency gains, a pattern consistent with Bernard and Moxnes (2018). What is novel in our findings, however, is that this improvement not only bolsters a city's export performance but also significantly transforms trade patterns, as reflected in a notable increase in the skill content of exports. These changes in trade dynamics are accompanied by substantial shifts in the relative demand for different types of labor—an effect not previously documented in the literature.

The distributional effects of HSR are enduring, with long-lasting local development impacts. Over time, firms in connected cities experience higher wages and increased R&D intensity relative to sales after HSR integration. At the local level, these cities see both a growing share of skilled labor in employment and a rising skill premium. We interpret these results as clear evidence of the increased demand for skills driven by HSR.

As the literature acknowledges, reduced-form estimates cannot fully capture the general equilibrium effects of HSR, especially with trade and migration links between cities. Another contribution of this paper is developing and calibrating a spatial trade model that incorporates frictional trade, labor mobility, and outsourcing to quantify these effects. The model reflects both sides of the outsourcing effect: it reduces unskilled labor demand by shifting tasks outside the firm while raising demand through output effects as some unskilled workers become more productive in core roles. A key strength of our quantitative analysis is leveraging transaction records from the Taobao platform to build a novel dataset on domestic trade flows between Chinese cities, offering a precise and detailed view of China's economic geography and trade linkages.

Finally, we find the impact of improved production networks is significant, equivalent to an 11.37% reduction in iceberg trade costs for domestic intermediate inputs, accounting for 34% of total welfare gains alongside broader economic shifts. The benefits of HSR, however, are concentrated in coastal cities with higher concentrations of skilled labor, intensifying existing spatial disparities. Our counterfactual analysis suggests that rising inequality could be mitigated by reforms aimed at reducing domestic migration costs and lowering trade barriers.

References

Acemoglu, D., G. Gancia, and F. Zilibotti (2015): "Offshoring and directed technical change," *American Economic Journal: Macroeconomics*, 7, 84–122.

Adukia, A., S. Asher, and P. Novosad (2020): "Educational investment responses to economic opportunity: evidence from Indian road construction," *American Economic Journal: Applied Economics*, 12, 348–376.

Aggarwal, S. (2018): "Do rural roads create pathways out of poverty? Evidence from India," *Journal of Development Economics*, 133, 375–395.

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

Ahlfeldt, G. M., S. J. Redding, D. M. Sturm, and N. Wolf (2015): "The economics of density: Evidence from the Berlin Wall," *Econometrica*, 83, 2127–2189.

Allen, T., C. Arkolakis, and Y. Takahashi (2020): "Universal gravity," *Journal of Political Economy*, 128, 393–433.

Alvarez, F. and R. E. Lucas (2007): "General equilibrium analysis of the Eaton–Kortum model of international trade," *Journal of Monetary Economics*, 54, 1726–1768.

Amiti, M. and C. L. Freund (2008): *The anatomy of China's export growth*, vol. 4628, World Bank Washington, DC.

Amiti, M. and J. Konings (2007): "Trade Liberalization, Intermediate Inputs, and Productivity: Evidence from Indonesia," *American Economic Review*, 97, 1611–1638.

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

Antras, P., T. C. Fort, and F. Tintelnot (2017): "The margins of global sourcing: Theory and evidence from us firms," *American Economic Review*, 107, 2514–64.

Arkolakis, C., A. Costinot, and A. Rodríguez-Clare (2012): "New trade models, same old gains?" *The American Economic Review*, 102, 94–130.

Baumgarten, D., I. Geishecker, and H. Görg (2013): "Offshoring, tasks, and the skill-wage pattern," *European Economic Review*, 61, 132–152.

Bergeaud, A., C. Malgouyres, C. Mazet-Sonilhac, and S. Signorelli (2024): "Technological change and domestic outsourcing," *Journal of Labor Economics*.

Bernard, A. B. and A. Moxnes (2018): "Networks and trade," *Annual Review of Economics*, 10, 65–85.

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

Bilal, A. and H. Lhuillier (2021): "Outsourcing, inequality and aggregate output," Tech. rep., National Bureau of Economic Research.

Bøler, E. A., A. Moxnes, and K. H. Ulltveit-Moe (2015): "R&D, international sourcing, and the joint impact on firm performance," *American Economic Review*, 105, 3704–39.

Bryan, G. and M. Morten (2015): "Economic development and the spatial allocation of labor: Evidence from Indonesia," *Working Paper, London School of Economics and Stanford University*, 1671–1748.

Bullock, R. G., N. Zhou, G. P. Ollivier, R. Yu, and Y. Jin (2014): "Regional Economic Impact Analysis of High Speed Rail in China: Main Report," World Bank Group, accessed: 2024-06-25.

Cai, J. and A. Szeidl (2017): "Interfirm relationships and business performance," *The Quarterly Journal of Economics*, 133, 1229–1282.

Caliendo, L. and F. Parro (2015): "Estimates of the Trade and Welfare Effects of NAFTA," *The Review of Economic Studies*, 82, 1–44.

Caliendo, L., F. Parro, E. Rossi-Hansberg, and P.-D. Sarte (2017): "The Impact of Regional and Sectoral Productivity Changes on the US Economy," *The Review of Economic Studies*.

Carluccio, J., A. Cunat, H. Fadinger, and C. Fons-Rosen (2019): "Offshoring and skill-upgrading in French manufacturing," *Journal of International Economics*, 118, 138–159.

Charnoz, P., C. Lelarge, and C. Trevien (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, 949–994.

Costinot, A. and A. Rodríguez-Clare (2014): "Trade theory with numbers: Quantifying the consequences of globalization," in *Handbook of international economics*, Elsevier, vol. 4, 197–261.

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

Cucu, F. (2022): "Roads, Internal Migration and the Spatial Sorting of US High-Skilled Workers," .

Di Giovanni, J., A. A. Levchenko, and J. Zhang (2014): "The global welfare impact of China: trade integration and technological change," *American Economic Journal: Macroeconomics*, 6, 153–183.

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

Eaton, J. and S. Kortum (2002): "Technology, geography, and trade," *Econometrica*, 70, 1741–1779.

Eaton, J., S. Kortum, and F. Kramarz (2004): "Dissecting trade: Firms, industries, and export destinations," *American Economic Review*, 94, 150–154.

——— (2011): "An anatomy of international tade: Evidence from Frech firms," *Econometrica*, 79, 1453–1498.

——— (2016): "Firm-to-Firm Trade: Imports, exports, and the labor market," Discussion papers 16048, Research Institute of Economy, Trade and Industry (RIETI).

Eaton, J., S. S. Kortum, and F. Kramarz (2022): "Firm-to-Firm Trade: Imports, exports, and the labor market," Tech. rep., National Bureau of Economic Research.

Edmond, C., V. Midrigan, and D. Y. Xu (2015): "Competition, markups, and the gains from international trade," *The American Economic Review*, 105, 3183–3221.

Faber, B. (2014): "Trade integration, market size, and industrialization: evidence from China's National Trunk Highway System," *Review of Economic Studies*, 81, 1046–1070.

Fan, J. (2019): "Internal geography, labor mobility, and the distributional impacts of trade," *American Economic Journal: Macroeconomics*, 11, 252–88.

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

Feenstra, R. C. (2007): "Globalization and its impact on labour," Tech. rep., WIIW Working Paper.

Feenstra, R. C. and G. H. Hanson (1996): "Globalization, outsourcing, and wage inequality," *The American Economic Review*, 86, 240.

——— (2003): "Global production sharing and rising inequality: A survey of trade and wages," *Handbook of international trade*, 146–185.

Fretz, S., R. Parchet, and F. Robert-Nicoud (2022): "Highways, market access and spatial sorting," *The Economic Journal*, 132, 1011–1036.

Gaulier, G. and S. Zignago (2010): "BACI: International Trade Database at the Product-Level. The 1994-2007 Version," Working Papers 2010-23, CEPII.

Geishecker, I. and H. Görg (2008): "Winners and losers: A micro-level analysis of international outsourcing and wages," *Canadian Journal of Economics/Revue canadienne d'économique*, 41, 243–270.

Giroud, X. (2013): "Proximity and investment: Evidence from plant-level data," *The Quarterly Journal of Economics*, 128, 861–915.

Goldberg, P. K., A. K. Khandelwal, N. Pavcnik, and P. Topalova (2010): "Imported intermediate inputs and domestic product growth: Evidence from India," *The Quarterly Journal of Economics*, 125, 1727–1767.

Halpern, L., M. Koren, and A. Szeidl (2015): "Imported inputs and productivity," *The American Economic Review*, 105, 3660–3703.

Heuermann, D. F. and J. F. Schmieder (2019): "The effect of infrastructure on worker mobility: evidence from high-speed rail expansion in Germany," *Journal of Economic Geography*, 19, 335–372.

Hsieh, C.-T., E. Hurst, C. I. Jones, and P. J. Klenow (2013): "The allocation of talent and us economic growth," Tech. rep., National Bureau of Economic Research.

Hummels, D., R. Jørgensen, J. Munch, and C. Xiang (2014): "The wage effects of offshoring: Evidence from Danish matched worker-firm data," *American Economic Review*, 104, 1597–1629.

Keuschnigg, C. and E. Ribi (2009): "Outsourcing, unemployment and welfare policy," *Journal of International Economics*, 78, 168–176.

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

Lu, W. J. (2018): "Transport, infrastructure and growth: Evidence from Chinese firms," *Available at SSRN 3289842*.

Michaels, G. (2008): "The effect of trade on the demand for skill: Evidence from the interstate highway system," *The Review of Economics and Statistics*, 90, 683–701.

Narayanan, A. and T. Lebesmuehlbacher (2023): "Road Infrastructure and the Skill Premium: Evidence from Indian Manufacturing," *Available at SSRN 4509296*.

Ou, J., B. Richard, Y. Jin, and N. Zhou (2014): "High-Speed Railways in China: Trafic Analysis," *The World Bank Report: China Transport Note Series*.

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

Ramondo, N., A. Rodríguez-Clare, and M. Saborío-Rodríguez (2016): "Trade, domestic frictions, and scale effects," *The American Economic Review*, 106, 3159–3184.

Ren, T., H.-J. Huang, Y. M. Nie, et al. (2023): "High-speed rail in China: Implications for intercity commuting and urban spatial structure," *Sustainable Cities and Society*, 97, 104719.

Simonovska, I. and M. E. Waugh (2014): "The elasticity of trade: Estimates and evidence," *Journal of international Economics*, 92, 34–50.

Tombe, T. and X. Zhu (2019): "Trade, migration, and productivity: A quantitative analysis of China," *American Economic Review*, 109, 1843–72.

Wright, G. C. (2014): "Revisiting the employment impact of offshoring," *European Economic Review*, 66, 63–83.

Zheng, H., J. Többen, E. Dietzenbacher, D. Moran, J. Meng, D. Wang, and D. Guan (2022): "Entropy-based Chinese city-level MRIO table framework," *Economic Systems Research*, 34, 519–544.

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

Zhong, Z., W. Zhou, J. Li, and P. Li (2024): "Regional Poverty Alleviation Partnership and E-Commerce Trade," *Marketing Science*.

Zhu, S. C. and D. Trefler (2005): "Trade and inequality in developing countries: a general equilibrium analysis," *Journal of international Economics*, 65, 21–48.

A Table Appendix

Table A.1: Descriptive Statistics

Variable	Unit	Num. Obs.	Mean	Std	Min	Max
Panel (a): City-level						
City exports (Log)	Yuan	1,934	23.10	1.89	15.33	28.38
Ordinary exports (Log)	Yuan	1,629	23.43	1.69	17.88	28.36
Processing exports (Log)	Yuan	1,629	18.82	2.89	4.51	25.57
Skill content in exports	-	1,934	0.38	0.12	0.17	1.00
GDP (Log)	Yuan	1,932	25.92	0.88	23.45	28.82
Population (Log)	-	1,933	15.23	0.65	12.18	17.34
Wage rates in manufacturing sector (Log)	Yuan	1,929	12.96	0.45	11.54	14.87
Wage rates in service sector (Log)	Yuan	1,929	12.68	0.50	10.66	14.27
Number of internet users (Log)	-	1,914	13.43	0.91	10.52	17.76
Tariff rate	-	1,934	1.06	0.04	1.00	1.51
Panel (b): Firm-level						
Sourcing distance (Log)	Kilometer					
Wage rate (Log)	Yuan	3,788,085	3.29	1.03	-11.49	13.39
R&D expenses (Log)	Yuan	195,509	7.35	2.47	-1.11	16.84
Outsourced intermediates (Log)	Yuan	3,428,840	6.59	3.68	0.00	20.69
Imports (Log)	Yuan	196,926	8.15	2.89	-1.11	19.80
Capital labor ratio (Log)	-	3,692,657	3.41	2.23	-11.62	19.38
Employment (Log)	Yuan	4,193,199	3.36	1.73	0.00	33.43

Note: All variables are presented in logarithmic form, with the exception of skill content and tariff rates.

Table A.2: Event Study of Connection to HSR

Dependent Variable:	City Exports			Skill Content
	Overall	Ordinary	Processing	in City Exports
	(1)	(2)	(3)	(4)
$Connect_{it}$ (T-3)	-0.015	-0.001	-0.201	-0.012
	(0.045)	(0.031)	(0.138)	(0.006)
$Connect_{it}$ (T-2)	-0.007	-0.022	-0.079	-0.010
	(0.031)	(0.028)	(0.104)	(0.007)
$Connect_{it}$ (T)	0.041	0.056***	-0.160	0.001
	(0.026)	(0.020)	(0.105)	(0.005)
$Connect_{it}$ (T+1)	0.051*	0.050**	-0.003	0.008*
	(0.027)	(0.021)	(0.106)	(0.005)
$Connect_{it}$ (T+2)	0.052*	0.051**	0.057	0.014***
	(0.027)	(0.024)	(0.113)	(0.005)
$Connect_{it}$ (T+3)	0.017	0.024	-0.006	0.012***
	(0.026)	(0.024)	(0.124)	(0.004)
Observations	1,907	1,604	1,604	1,907
R-squared	0.981	0.985	0.874	0.831
City Fixed Effects	Y	Y	Y	Y
Year Fixed Effects	Y	Y	Y	Y
Group-Year Fixed Effects	Y	Y	Y	Y
Other Controls	Y	Y	Y	Y

Notes: Each point estimate stems from a separate regression regression analysis. All regressions incorporate fixed effects for cities, years, and group-years. Cities are categorized into three groups according to their population levels: those with populations below 1 million are labeled as 1, those with populations between 1 million and 10 million are labeled as 2, and cities with populations exceeding 10 million are denoted as 3. Other controls include average annual wage rates in the manufacturing and services sectors, tariff rates, population size, and internet coverage at the city level. We define T as the period when a city is initially linked to the High-Speed Rail (HSR) network, with $T - i$ and $T + i$ representing the periods before and after the initial connection, respectively. Robust standard errors are clustered at the city level and reported in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

Table A.3: Summary of Key Parameter Calibration in the Model

Parameter	Description	Source / Target	Value
π_{ni}	Trade in final goods	Taobao, Chinese Customs Data, BACI Trade	
$v_{M,ij}/v_{M,i}$	Trade in intermediate inputs	Inter-City Input-Output (IO) Table	
L_{od}^W, L_{od}^B	Migration flows	Chinese Population Census	
$\bar{\omega}_{M,i}$	Outsourcing share	China Tax Survey	
$w_{M,i}L_{M,i}$, $w_{S,i}L_{S,i}$	Value-added	Market Clearing Conditions	
β_M, β_S	Cobb-Douglas cost share	World Bank Open Data	0.56, 0.44
$1-\zeta_M$, $1-\zeta_S$	Assembly worker share	China Family Panel Studies (CFPS)	0.25, 0
θ	Elasticity of trade	Simonovska and Waugh (2014)	4
γ	Matching parameter	Eaton et al. (2022)	0.34
ρ	Correlation parameter	Fan (2019)	0.36
ϵ_B, ϵ_W	Taste dispersion	Fan (2019)	3.9, 4.3

Table A.4: Estimation Results of Trade Costs in 2011 and 2017

Dependent Variable: $\ln \pi_{ij} - \ln \pi_{ii}$	(1) Year 2011	(2) Year 2017
I_{ij}^1 (Same Province)	4.734***	2.959***
	(0.455)	(0.321)
I_{ij}^2 (Different Provinces, Common Border)	12.156**	4.955**
	(4.736)	(2.182)
I_{ij}^3 (Different Provinces, No Common Border)	1.936***	3.306***
	(0.171)	(0.122)
I_{ij}^1 (Same Province) $\times \ln Dist_{ij}$	-0.579***	-0.357***
	(0.038)	(0.026)
I_{ij}^2 (Different Provinces, Common Border) $\times \ln Dist_{ij}$	-1.242***	-0.538***
	(0.406)	(0.185)
I_{ij}^3 (Different Provinces, No Common Border) $\times \ln Dist_{ij}$	-0.377***	-0.387***
	(0.012)	(0.009)
Observations	9,897	9,900
R-Squared	0.969	0.973

Notes: This table reports the estimates of trade costs. Robust standard errors are reported in parentheses; *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A.5: Impacts of HSR on Input Costs, Welfare, and Skill Premium

	(1) Mean	(2) SD	(3) Min	(4) Median	(5) Max
Part (a): Impacts of HSR on Intermediate Input Costs					
Northeast Region	-14.48%	5.18%	-18.36%	-15.78%	-0.93%
Eastern Region	-10.28%	7.38%	-18.54%	-14.89%	1.07%
Central Region	-14.61%	4.94%	-17.35%	-15.87%	0.98%
Western Region	-12.41%	8.19%	-18.01%	-16.25%	1.97%
Overall	-11.37%	7.17%	-18.54%	-15.31%	1.97%
Part (b): Impacts of HSR on welfare and skill premium					
Unskilled Worker	-0.75%	1.53%	-3.86%	-0.48%	1.98%
Skilled Worker	5.64%	2.55%	-0.23%	6.72%	8.94%
Skill Premium	6.17%	2.29%	1.29%	6.02%	11.97%

Notes: Part (a) of the table presents summary statistics for city-level percentage changes in intermediate input prices resulting from the HSR connection across various regions. The Northeast region includes cities in Liaoning, Jilin, and Heilongjiang provinces. The Eastern region encompasses cities in Beijing, Tianjin, Hebei, Shanghai, Jiangsu, Zhejiang, Fujian, Shandong, Guangdong, and Hainan. The Central region comprises cities in Shanxi, Anhui, Jiangxi, Henan, Hubei, and Hunan. Lastly, the Western region includes cities in Inner Mongolia, Guangxi, Chongqing, Sichuan, Guizhou, Yunnan, Tibet, Shaanxi, Gansu, Qinghai, as well as the Ningxia Hui and Xinjiang Uyghur autonomous regions. Part (b) of the table presents the summary statistics for city-level percentage changes in welfare due to HSR connections for both unskilled and skilled workers, along with the percentage changes in the skill premium, measured as the wage difference between skilled and unskilled workers.

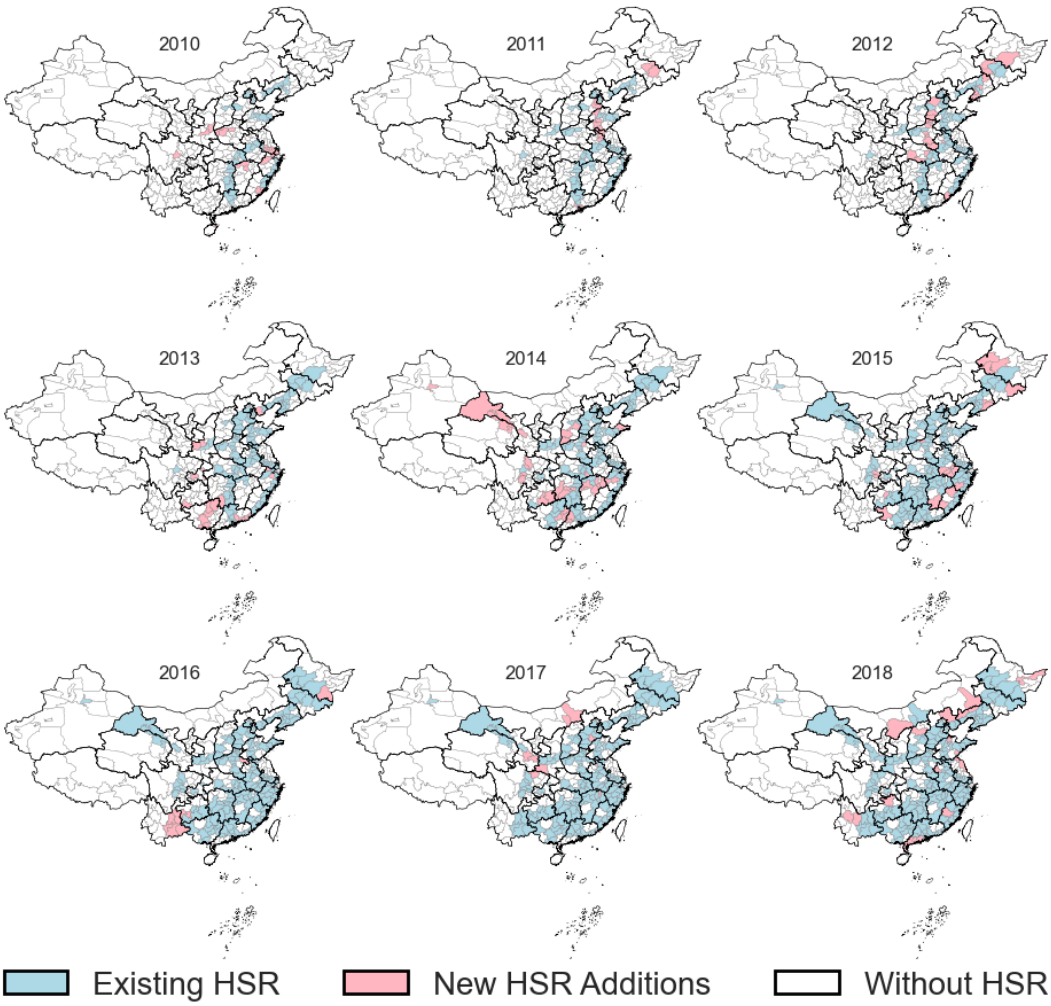
Table A.6: Impact of HSR on Technology, Domestic, and International Trade

	Variations Explained by HSR
Change in Overall Technology	45%
Change in City Exports	56%
Change in Inter-City Trade	25%

Notes: This table presents the percentage of total variance in technology change ($\hat{\Upsilon}_i$), city exports, and inter-city trade explained by HSR connections, as predicted by the model from 2011 to 2017.

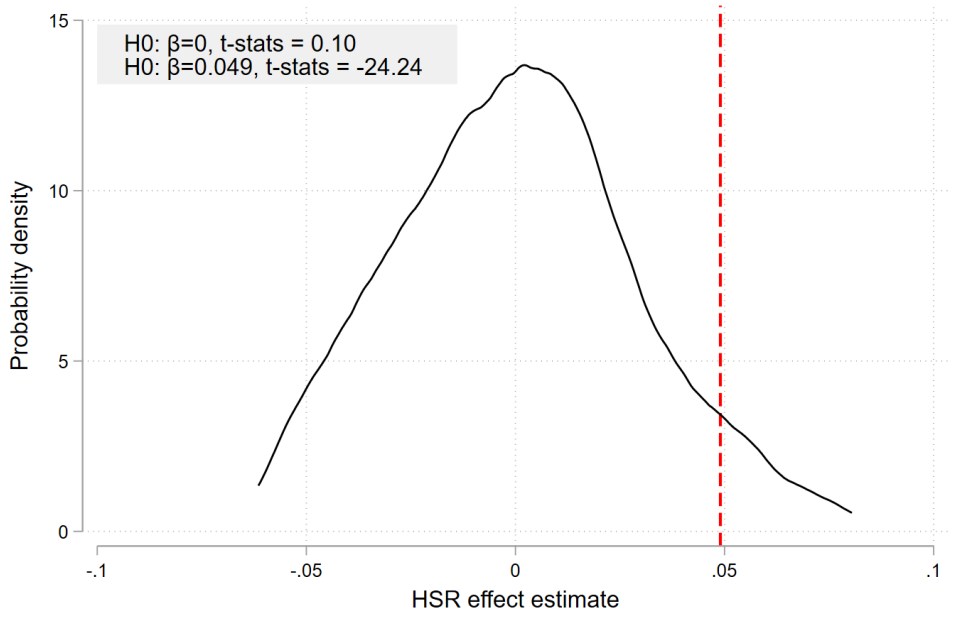
B Figure Appendix

Figure A.1: Evolution of HSR Expansion from 2010 to 2018



Notes: The figure illustrates the evolution of prefectures connected to the HSR network from 2010 to 2018. Areas in blue represent prefectures that were connected prior to the current year, while areas in red denote prefectures connected during the current year.

Figure A.2: Distribution of Estimated Coefficients in the Falsification Test



Notes: The figure shows the probability density of the estimated coefficients from 200 simulations randomly assigning HSR connection to cities. The vertical line presenting the estimate of column (5) in Table 1.

Figure A.3: Estimation Result of Technology Shock across Cities

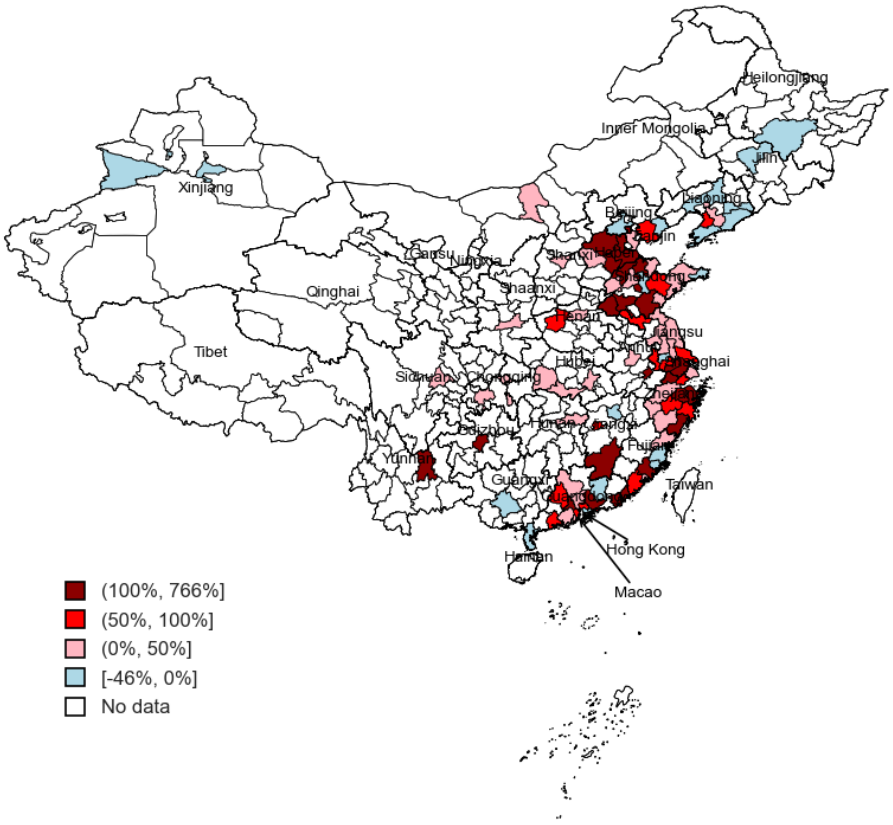
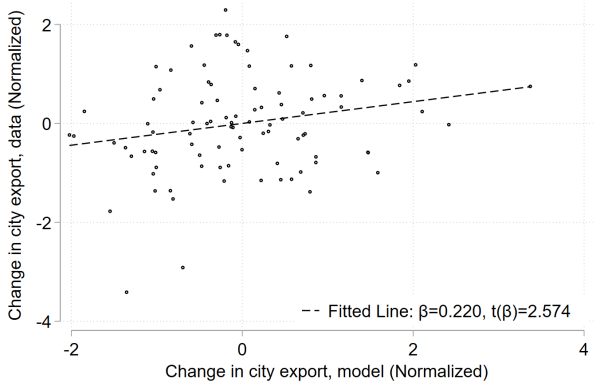


Figure A.4: Model Fit of Predicted Trade and Skill Demand Against Data (2011-2017)

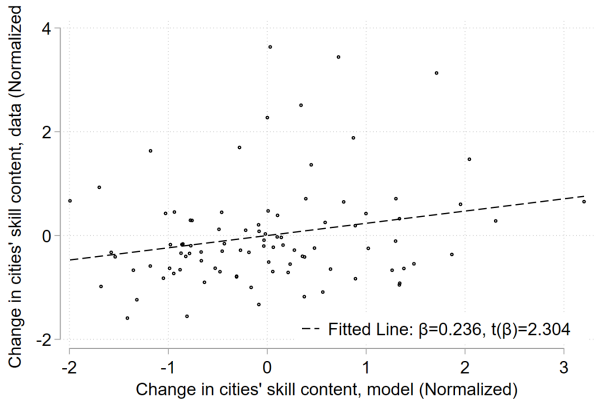
(a) City exports



(b) Bilateral city trade



(c) City's skill content in export



Notes: Panel (a) presents a scatter plot comparing the model's predicted changes in exports with actual data across cities from 2011 to 2017. Panel (b) extends this analysis to bilateral city trade data, showing consistent results that further validate the model's reliability in capturing trade dynamics. Panel (c) compares the changes in skill content within exports, measured by the number of skilled workers per unit of total export sales. In each case, a fitted line is included, displaying the beta coefficient and corresponding t-statistic, underscoring the model's predictive accuracy. We normalize the x-axis and y-axis measures in the scatter plot by subtracting the mean and dividing by the standard deviation, ensuring comparability across variables.

Figure A.5: Changes in Matching Efficiency Across Cities

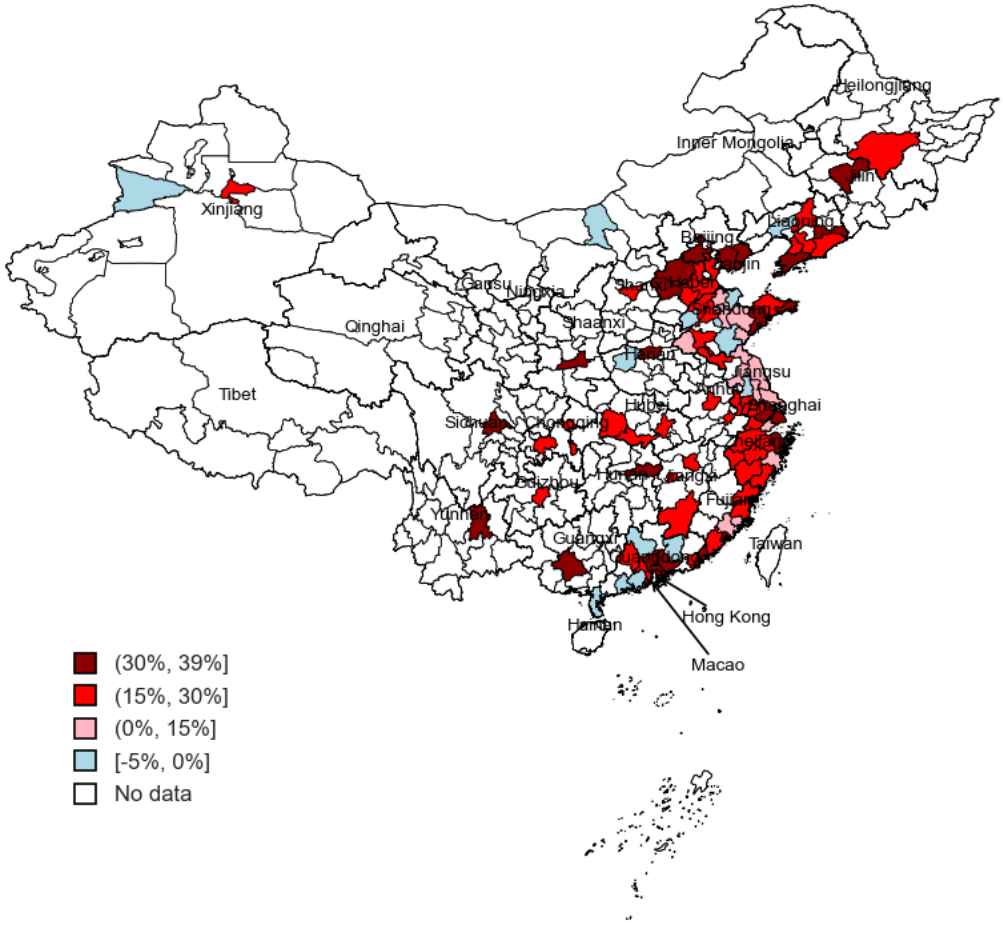


Figure A.6: Distribution of Matching Efficiency Change across Cities

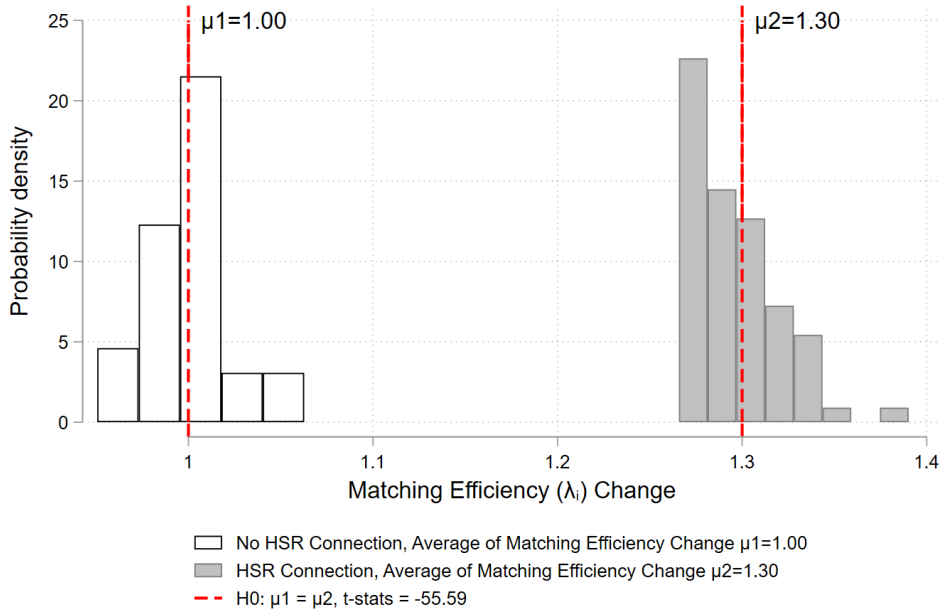


Figure A.7: Changes of Intermediate Inputs Price across Cities

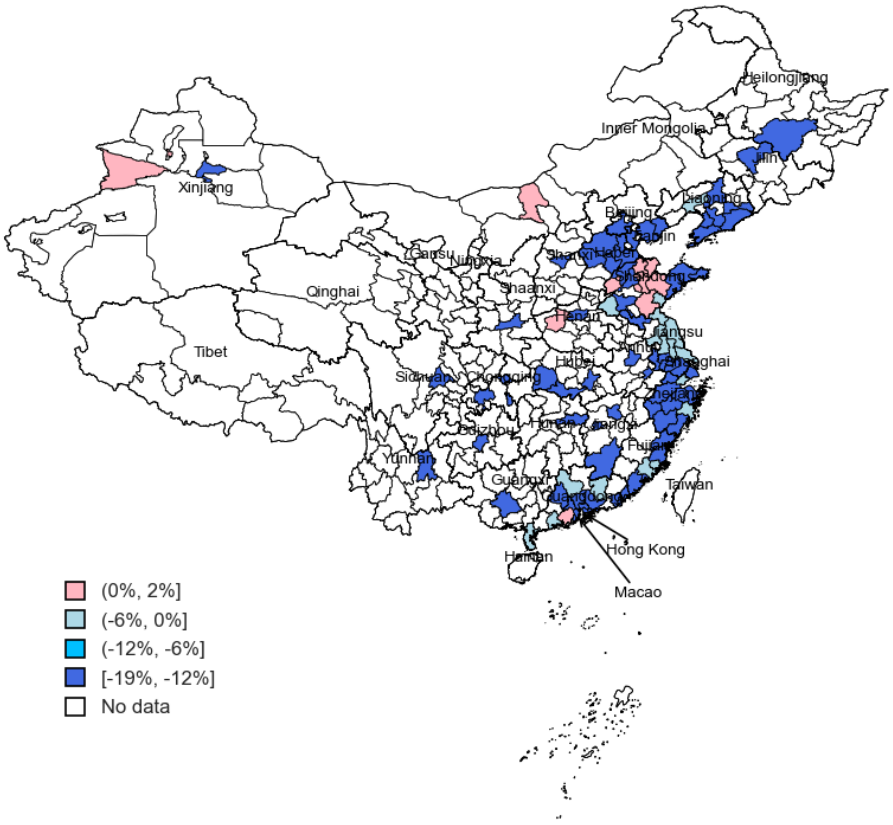
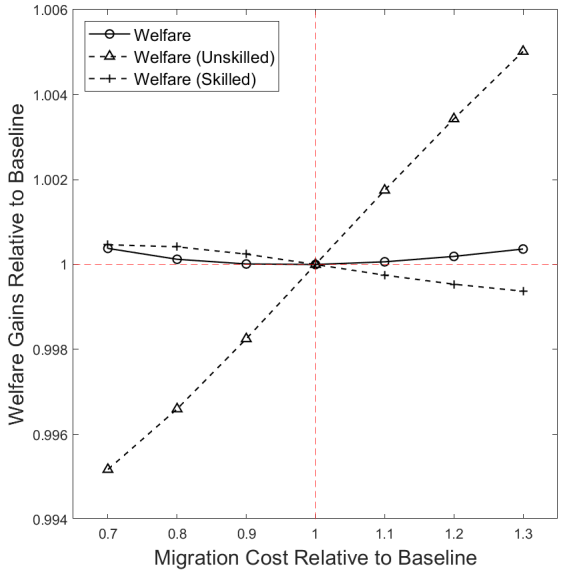
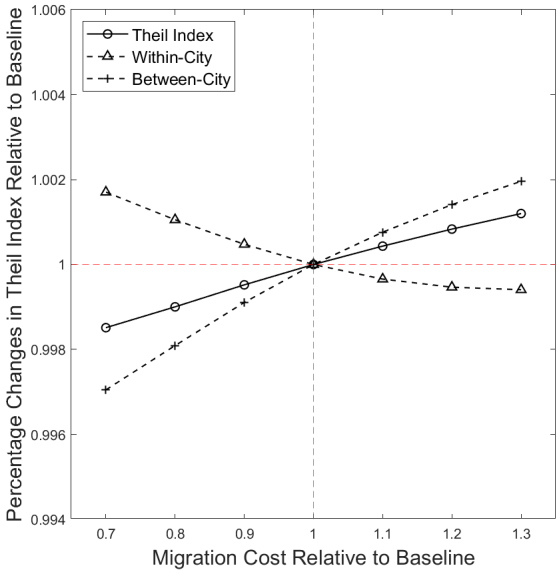


Figure A.8: The Role of Labor Mobility and Domestic Trade Costs in HSR's Impact

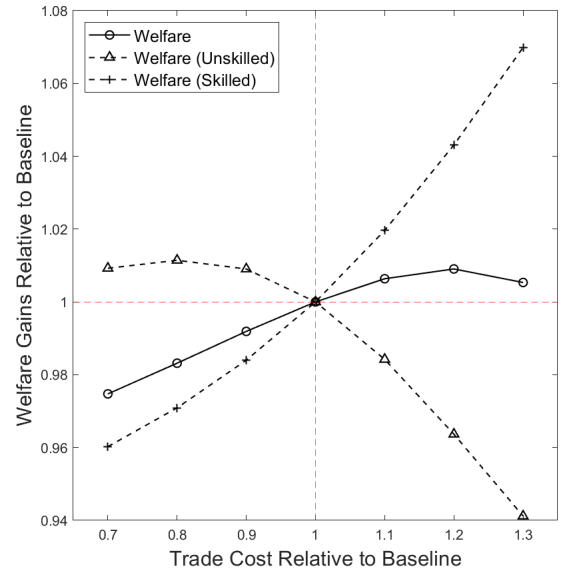
(a) Welfare Effect by Labor Mobility



(b) Inequality Effect by Labor Mobility



(c) Welfare Effect by Domestic Trade Costs



(d) Inequality Effect by Domestic Trade Costs

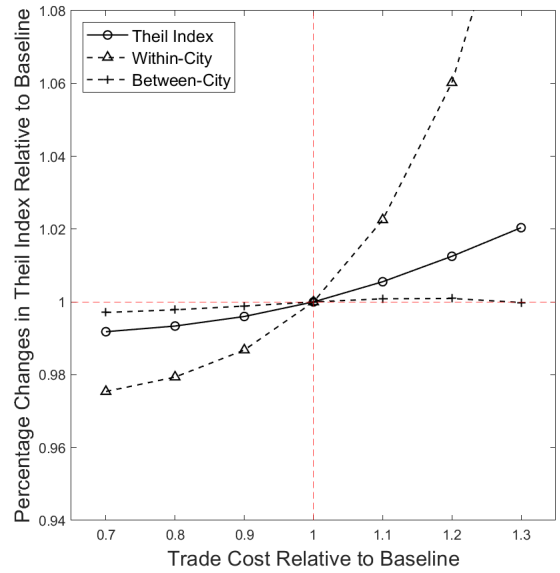
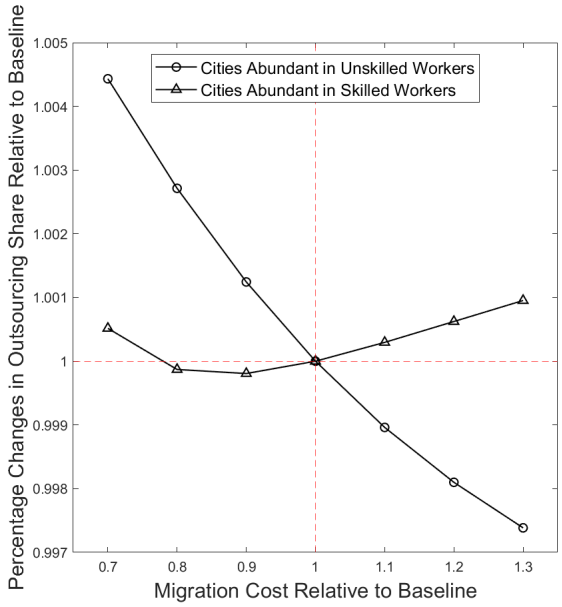
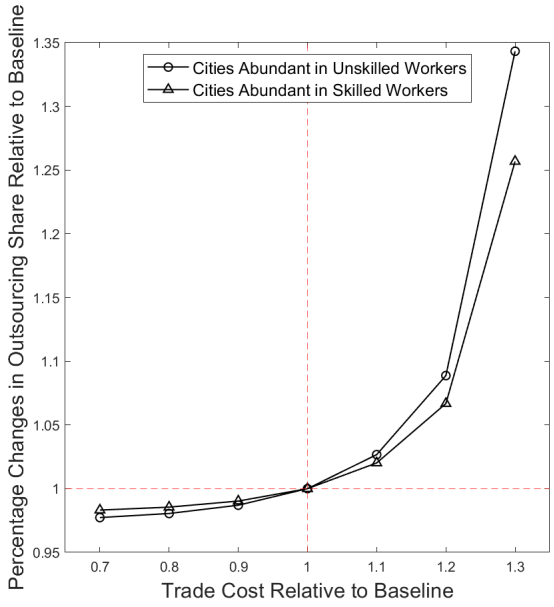


Figure A.9: The Role of Labor Mobility and Domestic Trade Costs in Outsourcing

(a) Impact of Labor Mobility



(b) Impact of Domestic Trade Costs



C Data Appendix

C.1 Skill Premium

The skill premium refers to the wage differential between skilled and unskilled workers. To estimate the skill premium, we employ the data from the CFPS, a nationally representative, biennial longitudinal survey of Chinese communities, families, and individuals, initiated in 2010 by the Institute of Social Science Survey at Peking University. The specification adopted is provided as:

$$\ln w_i = \beta_0 + \beta_1 College_i + \beta_2 Age_i + \beta_3 Age_i^2 + \gamma X_i + u_i \tag{A.1}$$

where w_i represents the wage of individual i, $College_i$ is a binary variable indicating whether individual i has completed tertiary education beyond high school, and Age_i is used as a proxy for work experience. The vector X_i includes control variables for marital status, gender, ethnicity, and sector of employment. This model is estimated separately for each city and year. The coefficient β_1 on the education dummy variable quantifies the wage premium for skilled workers, reflecting the additional earnings they receive compared to their unskilled counterparts within the same city and year.

C.2 Firm Data

Chinese tax survey: The firm-level data in this study is primarily sourced from the enterprise income tax records maintained by the Chinese State Administration of Tax (SAT), which functions similarly to the Internal Revenue Service (IRS) in the United States. The SAT collects and audits tax payments and maintains comprehensive firm-level records. These records include various firm characteristics and financial statement details, such as tax identification numbers, locations, industry classifications, employment numbers, cash flow statements, balance sheets, and income statements. For this analysis, we utilize data spanning from 2010 to 2016, covering variables such as registration location, capital, employment, labor compensation, Research and Development (R&D) expenses, materials purchased, exports and imports, and operational sales and costs.

WIND financial data: The second dataset employed in our firm analysis is from WIND, a leading provider of financial data on Chinese firms. This dataset includes detailed purchasing information from the top five suppliers for each listed firm. It is important to note that these suppliers are not necessarily listed firms themselves. To create a comprehensive dataset, we combined the firm characteristics from the Chinese Tax Survey with the purchasing data

from WIND. This integration has yielded a detailed firm-level dataset that encompasses both buyer and seller information spanning from 2010 to 2018, although production information from the tax survey is only available up to 2016.

Measure for sourcing distance: To calculate the sourcing distance for each listed firm based on the purchasing data, we first determine the geographical distances ($Dist_{ni}$) between each listed firm (n) and its suppliers (i), using the city locations of each. The sourcing distance for a firm is then calculated as the weighted average of the distances to its top five suppliers, with the weights being the purchase amounts (Q_{ni}) from each supplier. This is expressed as $SourcingDistance_n = \sum_{i=1}^{5} Dist_{ni} \times \frac{Q_{ni}}{\sum_{k=1}^{5} Q_{nk}}$.

Measurement of additional firm variables: Additionally, we assess other relevant firm variables such as wage, R&D expenses, materials cost, and firm imports. Wages are estimated by dividing total labor costs by the number of employees. The shares of R&D expenses, materials, and imports are then calculated as proportions of total operational costs, using data from the tax survey dataset.

D Least-Cost HSR Network

The section describes the method of constructing the least-cost path networks of HSR that is used as the instrument variables for being connected to HSR in reality. The construction method follows Faber (2014) and we differ in that the objectives of HSR is to link all the provincial capitals on a single network, as stated in the official report. The involved data, such as land usage, land cover, and elevation, are obtained from Institute of Geographic Sciences and Natural Resources Research Chinese Academy of Sciences.

Firstly, we adopt the construction cost function as used in Faber (2014),[37]

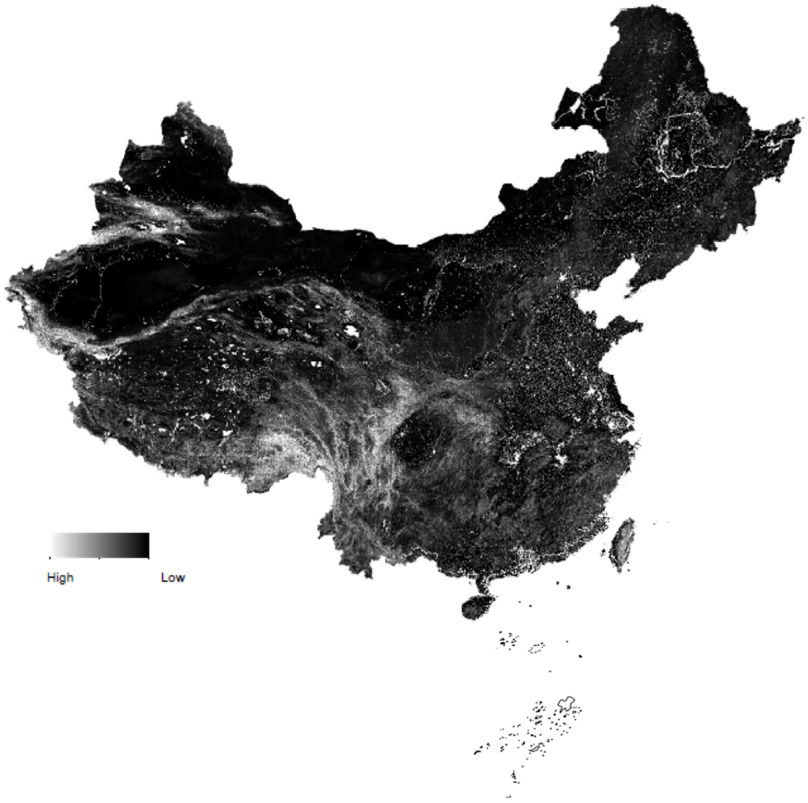
$$c_i = 1 + Slope_i + 25 \times Built_i + 25 \times Water_i + 25 \times Wetland_i,$$

where c_i is the incurred cost if the HSR crosses a pixel of land i, $Slope_i$ stands for the average slope gradient of land i, $Built_i$ is a dummy variable that equals one if land i is covered by any industrial, mining and residential building, $Water_i$ and $Wetland_i$ indicates whether the pixel is covered by water or wetland. This simple cost function implies that the high costs are associated crossing built structures, water area as well as long and steep routes. We use the satellite sensing data on elevation, land cover and land use to compute the cost (c_i) for

[37]Despite the fact that the cost model in Faber (2014) is for high way system, we believe that the construction cost of HSR is qualitatively similar.

each pixel. In the calculation, we reclassify the elevation, land cover and land use grids to $1km \times 1km$ grid cells. Figure A.10 graphically demonstrate the construction cost surface. The color ranges from white (the very high cost of crossing) to the black (the very low cost of crossing).

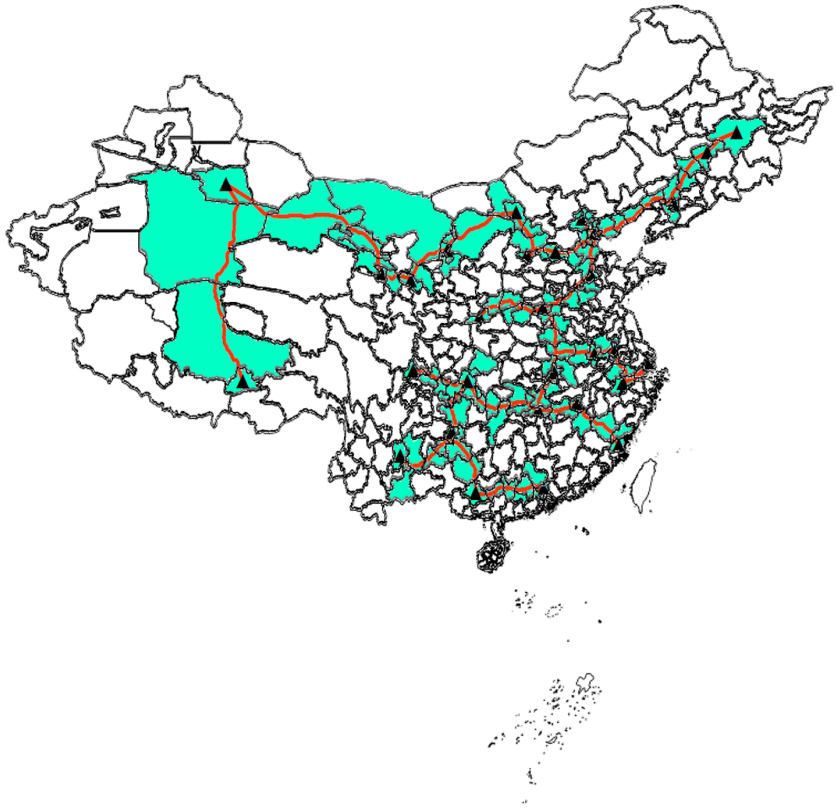
Figure A.10: The Estimated Construction Cost of HSR in China



After obtaining the constructing cost raster as the input, we then proceed to construct the least cost HSR paths between the 30 capital cities.[38] This generates 435 ($\frac{30*29}{2}$) possible bilateral paths between all pairs of the targeted city nodes. We then extract the aggregate construction cost of each bilateral path and use it as the input to compute the Prim's minimum spanning tree algorithm. Figure A.11 depicts the LCP network of HSR.

[38]The 30 *capital cities* cover 21 provinces excluding Hainan and Taiwan, 5 autonomous regions as well as 4 municipalities.

Figure A.11: The Least Cost Path Spanning Tree Network



E Theoretical Appendix

E.1 Deriving the Probability of Outsourcing

The probability that firm j in region i is not able to encounter any suppliers with price lower than c for task k is $\prod_n exp(-\rho_{k,in}(c))$ by Poisson distribution. The cost distribution of outsourcing the task, given parameter restriction $\gamma = (\theta - \phi)/\theta$, is expressed as

$$G^o_{k,i}(c) = 1 - exp(-v_{k,i}c^{\phi}) \tag{A.2}$$

where $v_{k,i} \equiv \sum_n \frac{\theta}{\phi}\lambda_{k,in}\Upsilon_n^{\phi/\theta}d_{in}^{-\phi}$. The probability that labor cost $(w_{k,i}/q_{k,i}(j))$ is no less than c is $F(w_{k,i}/c)$, and the cost distribution of hiring labor is thus

$$G^l_{k,i}(c) = 1 - exp(-w_{k,i}^{-\phi}c^{\phi}) \tag{A.3}$$

If the cost of using labor to fulfill task k is p, then the probability that hiring labor costs less than outsourcing is equal to $1 - G^o_{k,i}(p)$. Then we integrate this for all possible p under density $dG^l_{k,i}(p)$ to obtain

$$\int_0^{+\infty}(1 - G^o_{k,i}(p))dG^l_{k,i}(p) = \int_0^{+\infty} exp(-v_{k,i}p^{\phi})dG^l_{k,i}(p) \tag{A.4}$$

$$= -w_{k,i}^{-\phi}\int_0^{+\infty}\phi p^{\phi-1}exp\left[-(v_{k,i} + w_{k,i}^{-\phi})p^{\phi}\right]dp \tag{A.5}$$

$$= \frac{w_{k,i}^{-\phi}}{\Xi_{k,i}} \tag{A.6}$$

where $\Xi_{k,i} = v_{k,i} + w_{k,i}^{-\phi}$. The probability of fulfilling task k by outsourcing is then $v_{k,i}/\Xi_{k,i}$.

E.2 Derivation of Cost Distribution $\mu_{ni}(c)$

$$\mu_{ni}(c) = E\left\{\mu_i^Z\left(\frac{d_{ni}}{c}\prod_{k\in\{S,M\}}\frac{(c_{k,i}(j)^{\zeta_k}w_{k,i}^{1-\zeta_k})^{\beta_k}}{b_k}\right\}\right. \tag{A.7}$$

$$= T_i d_{ni}^{-\theta} c^{\theta}\prod_{k\in\{S,M\}}\int_0^{+\infty} b_k^{\theta}c_k^{-\theta\zeta_k\beta_k}w_{k,i}^{-\theta(1-\zeta_k)\beta_k}dG_{k,i}(c_k) \tag{A.8}$$

$$= T_i d_{ni}^{-\theta} c^{\theta}\prod_{k\in\{S,M\}} b_k^{\theta}w_{k,i}^{-\theta(1-\zeta_k)\beta_k}\Xi_{k,i}^{\theta\zeta_k\beta_k/\phi}\int_0^{+\infty} x^{-\theta\zeta_k\beta_k/\phi}exp(-x)dx \tag{A.9}$$

$$= T_i d_{ni}^{-\theta} c^{\theta}\prod_{k\in\{S,M\}} b_k^{\theta}w_{k,i}^{-\theta(1-\zeta_k)\beta_k}\Xi_{k,i}^{\theta\zeta_k\beta_k/\phi}\Gamma(1-\frac{\theta\zeta_k\beta_k}{\phi}) \tag{A.10}$$

$$= T_i\Xi_i\mathbb{W}_i d_{ni}^{-\theta}c^{\theta} \tag{A.11}$$

where we define $x_k \equiv \Xi_{k,i}c_k^{\phi}$ in the third equality. The last equality is derived by assuming $b_k = \Gamma(1-\frac{\theta\zeta_k\beta_k}{\phi})^{-\frac{1}{\theta}}$, which require $\theta\zeta_k\beta_k/\phi < 1$ in order to make $\mu_{ni}(c)$ meaningful. Furthermore, we define $\Xi_i \equiv \prod_{k\in\{S,M\}}(\Xi_{k,i})^{\theta\zeta_k\beta_k/\phi}$ and define $\mathbb{W}_i \equiv \prod_{k\in\{S,M\}} w_{k,i}^{-\theta(1-\zeta_k)\beta_k}$ to obtain the last equality.

E.3 Algorithm of Computing Υ

Υ is denoted as the vector of $N\times 1$, whose i-th element is defined as (20). We denote the function form of (20) as $\Upsilon = f(\Upsilon)$. To show there is a fixed point, we transform the fixed point problem in form of $\ln\Upsilon = F(\ln\Upsilon)$,where $F(x) \equiv \ln f(exp(x))$. There exists a unique fixed point of F if it is a contraction. $F_i(x)$, the i-th element of vector $F(x)$ is derived as:

$$F_i(x) = \ln\left\{T_i\mathbb{W}_i\prod_{k\in\{S,M\}}\left[\sum_n\frac{\theta}{\phi}\lambda_{k,in}exp(\frac{\phi}{\theta}x_n)d_{in}^{-\phi}+w_{k,i}^{-\phi}\right]^{\theta\zeta_k\beta_k/\phi}\right\} \tag{A.12}$$

$$= \ln\left\{T_i\mathbb{W}_i w_{S,i}^{-\theta\zeta_S\beta_S}\left[\sum_n\frac{\theta}{\phi}\lambda_{M,in}exp(\frac{\phi}{\theta}x_n)d_{in}^{-\phi}+w_{M,i}^{-\phi}\right]^{\theta\zeta_M\beta_M/\phi}\right\} \tag{A.13}$$

where the second equality comes from the assumption that the service task cannot be outsourced, $\lambda_{S,i} = 0$.

To apply Blackwell's contraction mapping theory, we need to verify both monotonicity and discounting properties for $F(x)$ to confirm the sufficient conditions. For monotonicity, let x and y satisfy that $x \le y$, as all the exogenous parameters are positive then

$$F_i(x) = \ln\left\{ T_i \mathbb{W}_i w_{S,i}^{-\theta\zeta_S\beta_S} \left[\sum_n \frac{\theta}{\phi} \lambda_{M,in} exp(\frac{\phi}{\theta} x_n) d_{in}^{-\phi} + w_{M,i}^{-\phi} \right]^{\theta\zeta_M\beta_M/\phi} \right\} \tag{A.14}$$

$$\leq \ln\left\{ T_i \mathbb{W}_i w_{S,i}^{-\theta\zeta_S\beta_S} \left[\sum_n \frac{\theta}{\phi} \lambda_{M,in} exp(\frac{\phi}{\theta} y_n) d_{in}^{-\phi} + w_{M,i}^{-\phi} \right]^{\theta\zeta_M\beta_M/\phi} \right\} \tag{A.15}$$

$$= F_i(y) \tag{A.16}$$

For discounting, for $\forall a > 0$

$$F_i(x+a) = \ln\left\{ T_i \mathbb{W}_i w_{S,i}^{-\theta\zeta_S\beta_S} \left[\sum_n \frac{\theta}{\phi} \lambda_{M,in} exp(\frac{\phi}{\theta} x_n + \frac{\phi}{\theta} a) d_{in}^{-\phi} + w_{M,i}^{-\phi} \right]^{\theta\zeta_M\beta_M/\phi} \right\} \tag{A.17}$$

$$\leq \ln\left\{ T_i \mathbb{W}_i w_{S,i}^{-\theta\zeta_S\beta_S} \left[\sum_n \frac{\theta}{\phi} \lambda_{M,in} exp(\frac{\phi}{\theta} x_n + \frac{\phi}{\theta} a) d_{in}^{-\phi} + w_{M,i}^{-\phi} exp(\frac{\phi}{\theta} a) \right]^{\theta\zeta_M\beta_M/\phi} \right\} \tag{A.18}$$

$$= \ln\left\{ T_i \mathbb{W}_i w_{S,i}^{-\theta\zeta_S\beta_S} \left[\sum_n \frac{\theta}{\phi} \lambda_{M,in} exp(\frac{\phi}{\theta} x_n) d_{in}^{-\phi} + w_{M,i}^{-\phi} \right]^{\theta\zeta_M\beta_M/\phi} exp(\zeta_M\beta_M a) \right\} \tag{A.19}$$

$$= F_i(x) + \zeta_M\beta_M a \tag{A.20}$$

where the inequality comes from that $exp(\frac{\phi}{\theta}a) > 1$. As $\beta_M < 1$ and $\zeta_M \leq 1$, discounting is verified.

Therefore, we are able to compute the fixed point ($\ln \Gamma$) by iterating on:

$$\ln \Upsilon_{t+1} = F(\ln \Upsilon_t)$$

with the initial value of $\ln \Upsilon_0 = 0$. After we obtain $\ln \Upsilon$, we are able to calculate Υ. The result directly implies that each element of $\ln \Upsilon$ increases with technology parameter T_i anywhere, decreases with any task-specific labor cost $w_{k,i}$ in any region, and increases with the task-specific arrival rate of price quotes $\lambda_{k,ni}$ in any region, though the comparative statics only capture the partial effect and do not predict the general equilibrium outcomes.

E.4 Aggregate Production Function

The proof follows Eaton et al. (2016). We only focus on the manufacturing tasks, since service tasks uses the skilled workers only. The cost distribution of outsourcing the task, given parameter restriction $\gamma = (\theta - \phi)/\theta$, is expressed as

$$G^o_{M,i}(c) = 1 - exp(-v_{M,i}c^{\phi}) \tag{A.21}$$

where $v_{M,i} \equiv \sum_n \frac{\theta}{\phi}\lambda_{M,in}\Upsilon_n^{\phi/\theta}d_{in}^{-\phi}$. The average unit cost of using intermediate input sourced from region i is

$$\begin{aligned}
\bar{p}_{M,i} &= \int_0^{+\infty} p dG^o_{M,i}(p) && \text{(A.22)}\\
&= \int_0^{+\infty} \phi \times v_{M,i}p^{\phi} \times exp\left(-v_{M,i}p^{\phi}\right) dp && \text{(A.23)}\\
&= \int_0^{+\infty} \left(\frac{x}{v_{M,i}}\right)^{1/\phi} \times exp(-x)dx && \text{(A.24)}\\
&= v_{M,i}^{-1/\phi}\Gamma\left(1 + \frac{1}{\phi}\right) && \text{(A.25)}
\end{aligned}$$

The aggregate labor share in the production of task M can be rewritten as (by written $v_{M,i}$ as function of average price $\bar{p}_{M,i}$):

$$\begin{aligned}
\beta_i^{l(M)} &= \beta_M \times \frac{w_{M,i}^{-\phi}}{v_{M,i} + w_{M,i}^{-\phi}} && \text{(A.26)}\\
&= \beta_M \times \frac{w_{M,i}^{-\phi}}{\left[\bar{p}_{M,i}/\Gamma\left(1 + \frac{1}{\phi}\right)\right]^{-\phi} + w_{M,i}^{-\phi}} && \text{(A.27)}
\end{aligned}$$

We now turn to the production function shown in equation (22). For each manufacturing task in region i, the representative firm could hire unskilled labor $L_{M,i}$ at blue-collar wage rate w_i^B, and purchase the composite intermediate task $I_{M,i}$ at the average price $\bar{p}_{M,i}$. The first order condition of cost minimization of manufacturing task yields:

$$\frac{w_{M,i}L_{M,i}}{\bar{p}_{M,i}I_{M,i}} = \Gamma\left(1+\frac{1}{\phi}\right)^{-\phi} \times \left(\frac{w_{M,i}}{\bar{p}_{M,i}}\right)^{-\phi} \tag{A.28}$$

$$= \frac{w_{M,i}^{-\phi}}{\left[\bar{p}_{M,i}/\Gamma\left(1+\frac{1}{\phi}\right)\right]^{-\phi}} \tag{A.29}$$

Then it is straightforward that the labor share generated by the aggregate production function is same with the one implied by the assumptions on technology in model.

E.5 Derivation of Expected Indirect Utility

Given taste shocks z_{odi}, the expected indirect utility is expressed as

$$\nu_d^l = E\left\{V_{odi}^l(p_S, p_M)\right\} \tag{A.30}$$

$$= \int\int \frac{w_d^l}{a_S a_M p_S^{\alpha_S} p_M^{\alpha_M}} d\tilde{G}(p_S) d\tilde{G}(p_M) \tag{A.31}$$

$$= w_d^l \prod_{k\in\{S,M\}} a_k^{-1} \int_0^{+\infty} p_k^{-\alpha_k} d\tilde{G}(p_k) \tag{A.32}$$

$$= w_d^l \prod_{k\in\{S,M\}} a_k^{-1} \int_0^{+\infty} p_k^{-\alpha_k} \tilde{v}_{k,d}\theta(1-\tilde{\gamma}) p_k^{\theta(1-\tilde{\gamma})-1} exp\left[-\tilde{v}_{k,d} p_k^{\theta(1-\tilde{\gamma})}\right] dp_k \tag{A.33}$$

$$= w_d^l \prod_{k\in\{S,M\}} a_k^{-1} \tilde{v}_{k,d}^{\frac{\alpha_k}{\theta(1-\tilde{\gamma})}} \Gamma\left(1-\frac{\alpha_k}{\theta(1-\tilde{\gamma})}\right) \tag{A.34}$$

$$= B_d \times w_d^l \times \mathbb{Y}_d^{\frac{1}{\theta}} \tag{A.35}$$

where the third equality comes from the assumption that the prices p_S and p_M are independently drawn. Here B_d is defined as

$$B_d \equiv \left[\prod_{k\in\{S,M\}} \tilde{\lambda}_{k,d}^{\alpha_k}\right]^{\frac{1}{\theta(1-\tilde{\gamma})}} \tag{A.36}$$

and a_k is cautiously chosen as

$$a_k = (1-\tilde{\gamma})^{-\frac{\alpha_k}{\theta(1-\tilde{\gamma})}} \times \Gamma\left(1-\frac{\alpha_k}{\theta(1-\tilde{\gamma})}\right) \tag{A.37}$$

to eliminate the effect of need k on utility. We also require $\alpha_k < \theta(1-\tilde{\gamma})$ to make the Gamma function well-defined.

E.6 Derivation of Migration Share

The migration share from o to d of workers of type l is expressed as:

$$\delta_{od}^{l} = \mathbf{Pr}(\frac{\nu_{d}^{l} z_{di}}{e_{od}^{l}} \geq \frac{\nu_{g}^{l} z_{gi}}{e_{og}^{l}},\ \forall g) \tag{A.38}$$

$$= \mathbf{Pr}(z_{gi} \leq \frac{\nu_{d}^{l}/e_{od}^{l}}{\nu_{g}^{l}/e_{og}^{l}} z_{di},\ \forall g) \tag{A.39}$$

$$= \int_{0}^{+\infty} \left[\int ... \int_{z_{gi} \leq \frac{\nu_{d}^{l}/e_{od}^{l}}{\nu_{g}^{l}/e_{og}^{l}} z_{di}, \forall g \neq d} \frac{\partial^{n} F(\mathbf{z}_{-d}, z_{d})}{\partial z_{1} \partial z_{2} ... \partial z_{N}} d\mathbf{z}_{-d} \right] dz_{d} \tag{A.40}$$

$$= \int_{0}^{+\infty} \left\{ \int ... \int_{z_{gi} \leq \frac{\nu_{d}^{l}/e_{od}^{l}}{\nu_{g}^{l}/e_{og}^{l}} z_{di}, \forall g \neq d,j} \left[\int_{0}^{\frac{\nu_{d}^{l}/e_{od}^{l}}{\nu_{j}^{l}/e_{oj}^{l}}} \left(d \underbrace{\frac{\partial^{n-1} F(\mathbf{z}_{-d-j}, z_{j}, z_{d})}{\partial z_{1} \partial z_{2} ... \partial z_{g} ... \partial z_{N}}}_{g \neq d,j} / dz_{j} \right) dz_{j} \right] d\mathbf{z}_{-d-j} \right\} dz_{d} \tag{A.41}$$

$$= \int_{0}^{+\infty} \left\{ \int ... \int_{z_{gi} \leq \frac{\nu_{d}^{l}/e_{od}^{l}}{\nu_{g}^{l}/e_{og}^{l}} z_{di}, \forall g \neq d,j} \left[\frac{\partial^{n-1} F(\mathbf{z}_{-d-j}, \frac{\nu_{d}^{l}/e_{od}^{l}}{\nu_{j}^{l}/e_{oj}^{l}} z_{di}, z_{d})}{\underbrace{\partial z_{1} \partial z_{2} ... \partial z_{g} ... \partial z_{N}}_{g \neq d,j}} \right] d\mathbf{z}_{-d-j} \right\} dz_{d} \tag{A.42}$$

$$= .....$$

$$= \int_{0}^{+\infty} \frac{\partial F(\frac{\nu_{d}^{l}/e_{od}^{l}}{\nu_{1}^{l}/e_{o1}^{l}} z_{di}, ..., \frac{\nu_{d}^{l}/e_{od}^{l}}{\nu_{N}^{l}/e_{oN}^{l}} z_{di})}{\partial z_{d}} dz_{d} \tag{A.43}$$

where $\mathbf{z}_{-d}$ denotes the vector of $\{z_{gi}\}$ which does not include z_{di}, and the same notation meaning for $\mathbf{z}_{-d-j}$. Given the formula of F, it follows that

$$\delta_{od}^{l} = \int_{0}^{+\infty} \left[\sum_{g} \left(\frac{\nu_{d}^{l}/e_{od}^{l}}{\nu_{g}^{l}/e_{og}^{l}} \right)^{-\epsilon_{l}} z_{di}^{-\epsilon_{l}} \right]^{-\rho} (1-\rho) \epsilon_{l} z_{di}^{-\epsilon_{l}-1} exp \left\{ -z_{di}^{-\epsilon_{l}(1-\rho)} \times \left[\sum_{g} \left(\frac{\nu_{d}^{l}/e_{od}^{l}}{\nu_{g}^{l}/e_{og}^{l}} \right)^{-\epsilon_{l}} \right]^{1-\rho} \right\} dz_{di} \tag{A.44}$$

$$= \int_{0}^{+\infty} \frac{1}{\sum_{g} \left(\frac{\nu_{d}^{l}/e_{od}^{l}}{\nu_{g}^{l}/e_{og}^{l}} \right)^{-\epsilon_{l}}} \times dexp \left\{ -z_{di}^{-\epsilon_{l}(1-\rho)} \times \left[\sum_{g} \left(\frac{\nu_{d}^{l}/e_{od}^{l}}{\nu_{g}^{l}/e_{og}^{l}} \right)^{-\epsilon_{l}} \right]^{1-\rho} \right\} \tag{A.45}$$

$$= \frac{\left(\nu_{d}^{l}/e_{od}^{l}\right)^{\epsilon_{l}}}{\sum_{g} \left(\nu_{g}^{l}/e_{og}^{l}\right)^{\epsilon_{l}}} \tag{A.46}$$

E.7 Derivation of Expected Utility

The joint probability of utility for worker of type l from city o is denoted by vector $\mathbf{u}_o^l = \{u_{od}^l | d = 1, 2, ...N\}$ where $u_{od}^l \equiv \frac{z_{di}\nu_d^l}{e_{od}^l}$. Then the joint distribution of utility is derived as

$$U_o^l(\mathbf{u}_o^l) = \mathbf{Pr}\left(u_{o1}^l \leq u_1, ..., u_{oN}^l \leq u_N\right) \tag{A.47}$$

$$= \mathbf{Pr}\left(..., z_{di} \leq \frac{e_{od}^l u_d}{\nu_d^l}, ...\right) \tag{A.48}$$

$$= exp\left\{-\left[\sum_d \left(\frac{e_{od}^l u_d}{\nu_d^l}\right)^{-\epsilon_l}\right]^{1-\rho}\right\} \tag{A.49}$$

Therefore, the distribution of $u_o^l \equiv \max_{d\in\{1,2,..,N\}}\left\{\frac{z_{di}\nu_d^l}{e_{od}^l}\right\}$, which is denoted as $\mathbb{U}_o^l(u)$:

$$\mathbb{U}_o^l(u) = \mathbf{Pr}\left(u_o^l \leq u\right) \tag{A.50}$$

$$= \mathbf{Pr}\left(u_{od}^l \leq u,\ \forall d\right) \tag{A.51}$$

$$= U_o^l(u, u, ..., u) \tag{A.52}$$

$$= exp\left\{-\left[\sum_d \left(\frac{e_{od}^l u}{\nu_d^l}\right)^{-\epsilon_l}\right]^{1-\rho}\right\} \tag{A.53}$$

$$= exp\left\{-\left[\sum_d \left(\frac{e_{od}^l}{\nu_d^l}\right)^{-\epsilon_l}\right]^{1-\rho} u^{-(1-\rho)\epsilon_l}\right\} \tag{A.54}$$

$$= exp\left[-\left(M_o^l\right)^{1-\rho} u^{-(1-\rho)\epsilon_l}\right] \tag{A.55}$$

where $M_o^l \equiv \sum_d \left(\frac{e_{od}^l}{\nu_d^l}\right)^{-\epsilon_l}$. Then the expected utility born in city o is then

$$\bar{U}_o^l = \int_0^{+\infty} u d\mathbb{U}_o^l(u) \tag{A.56}$$

$$= \int_0^{+\infty} \left(M_o^l\right)^{1-\rho}(1-\rho)\epsilon_l u^{-(1-\rho)\epsilon_l} exp\left[-\left(M_o^l\right)^{1-\rho} u^{-(1-\rho)\epsilon_l}\right] du \tag{A.57}$$

$$= \left(M_o^l\right)^{\frac{1}{\epsilon_l}} \int_0^{+\infty} x^{-\frac{1}{(1-\rho)\epsilon_l}} exp\left(-x\right) dx \tag{A.58}$$

$$= \left(M_o^l\right)^{\frac{1}{\epsilon_l}} \Gamma\left[1 - \frac{1}{(1-\rho)\epsilon_l}\right] \tag{A.59}$$

where the derivation of the third equality uses the transformation $x = \left(M_o^l\right)^{1-\rho} u^{-(1-\rho)\epsilon_l}$ and the last equality comes from the definition of Gamma function.

E.8 Definition of Equilibrium

Definition. *A competitive equilibrium of the economy is defined as a set of prices and allocations such that the following conditions are satisfied:*

1. *The migration decisions for workers of both types are optimal, i.e.,* (29) *and* (30) *are satisfied.*
2. *The distribution of labor L_i^l, and the final consumption expenditure X_i^C are consistent with the endogenous supply of workers, i.e.,* (31) *and* (34).
3. *The decisions about firm's production are optimal.*
4. *The decisions about worker's consumption are optimal.*
5. *Labor markets and goods markets clear, i.e.,* (34) *and* (35).

E.9 Relative Change in Equilibrium

To quantify the aggregate and distributional effects of HSR, we further apply the exact hat algebra method (Allen et al., 2020). This approach models the impact of changes in exogenous variables, such as bilateral matching efficiency (from $\lambda_{M,ni}$ to $\lambda'_{M,ni}$). A variable with a hat, $\hat{x}$, represents its relative change, defined as $\hat{x} = x'/x$.

Assuming no change in $\hat{d}_{ni}$, $\hat{e}_{od}$, and $\hat{T}_i$, i.e., $\hat{d}_{ni} = \hat{e}_{od} = \hat{T}_i = 1$, we can solve 15 simultaneous equations given the matching efficiency shock ($\hat{\lambda}_{M,ni}$) through equations (A.61) to (A.75) and determine the impact of HSR connections on various variables including $\{\hat{w}_{M,i}$, $\hat{w}_{S,i}$, $\hat{\Xi}_i$, $\hat{\Xi}_{M,i}$, $\hat{\Upsilon}_i$, $\hat{\mathbb{Y}}_i$, $\hat{v}_{M,i}$, $\hat{\pi}_{ni}$, $\hat{\delta}^B_{od}$, $\hat{\delta}^W_{od}$, $\hat{L}_{M,i}$, $\hat{L}_{S,i}$, Y'_i, $\Phi^{I\prime}_n$, $X_i^{c\prime}\}$. Therefore, the relative change in aggregate average welfare is derived by the following expression:

$$\hat{W} = \sum_{l\in\{B,W\}} \sum_{i=1}^{N} \tilde{\omega}_i^l \times \hat{\tilde{U}}_o^l \tag{A.60}$$

where the change in the expected utility of a type l worker born in o, denoted as $\hat{\tilde{U}}_o^l$, and the original welfare share of a worker of type l in region i, denoted as $\tilde{\omega}_i^l$, are defined by:

$$\hat{\tilde{U}}_o^l = \left(\sum_g \delta_{og}^l (\hat{w}_g^l \hat{\mathbb{Y}}_g^{1/\theta} / \hat{e}_{og}^l)^{\epsilon_l}\right)^{\frac{1}{\epsilon_l}}, \quad \tilde{\omega}_i^l = \frac{m_i^l \times \left(M_i^l\right)^{\frac{1}{\epsilon_l}} \times \Gamma\left[1 - \frac{1}{(1-\rho)\epsilon_l}\right]}{\sum_{l'\in\{B,W\}} \sum_{i'=1}^{N} m_{i'}^{l'} \times \left(M_{i'}^{l'}\right)^{\frac{1}{\epsilon_{l'}}} \times \Gamma\left[1 - \frac{1}{(1-\rho)\epsilon_{l'}}\right]}$$

Since changes in matching efficiency are not directly observable, we estimate changes in task efficiency ($\hat{\Xi}_i$) and use these to compute the relative changes in equilibrium due to HSR connections and recover changes in matching efficiency at the same time. Specifically, given $\hat{\Xi}_i$ and that $\hat{d}_{ni} = \hat{e}_{od} = \hat{T}_i = 1$, we solve equations (A.61) to (A.75) to obtain $\{\hat{w}_{M,i}, \hat{w}_{S,i}, \hat{\lambda}_{ni}, \hat{\Xi}_{M,i}, \hat{\Upsilon}_i, \hat{\mathbb{Y}}_i, \hat{v}_{M,i}, \hat{\pi}_{ni}, \hat{\delta}^B_{od}, \hat{\delta}^W_{od}, \hat{L}_{M,i}, \hat{L}_{S,i}, Y'_i, \Phi^{I\prime}_n, X^{c\prime}_i\}$.

Furthermore, we assess changes in broader technology ($\hat{\Upsilon}_i$), which allows us to compute the relative shifts in equilibrium due to both HSR connections and other technological advancements. This exercise assumes that $\hat{d}_{ni} = \hat{e}_{od} = 1$ and solve for $\{\hat{w}_{M,i}, \hat{w}_{S,i}, \hat{\mathbb{Y}}_i, \hat{\pi}_{ni}, \hat{\delta}^B_{od}, \hat{\delta}^W_{od}, \hat{L}_{M,i}, \hat{L}_{S,i}, Y'_i, \Phi^{I\prime}_n, X^{c\prime}_i\}$ through equations (A.65) to (A.75).

$$\hat{v}_{M,i} = \sum_j \frac{v_{M,ij}}{v_{M,i}} \hat{\lambda}_{M,ij} \hat{\Upsilon}_j^{\phi/\theta} \hat{d}_{ij}^{-\phi} \tag{A.61}$$

$$\hat{\Xi}_{M,i} = \bar{\omega}_{M,i} \hat{v}_{M,i} + (1 - \bar{\omega}_{M,i}) \hat{w}_{M,i}^{-\phi} \tag{A.62}$$

$$\hat{\Xi}_i = \hat{\Xi}_{M,i}^{\theta\zeta_M\beta_M/\phi} \hat{w}_{S,i}^{-\theta\zeta_S\beta_S} \tag{A.63}$$

$$\hat{\Upsilon}_i = \hat{T}_i \hat{w}_{M,i}^{-\theta(1-\zeta_M)\beta_M} \hat{\Xi}_{M,i}^{\theta\zeta_M\beta_M/\phi} \hat{w}_{S,i}^{-\theta\beta_S} \tag{A.64}$$

$$\hat{\mathbb{Y}}_n = \sum_j \pi_{nj} \hat{\Upsilon}_j \hat{d}_{nj}^{-\theta} \tag{A.65}$$

$$\hat{\pi}_{ni} = (\hat{\Upsilon}_i \hat{d}_{ni}^{-\theta}) / \hat{\mathbb{Y}}_n \tag{A.66}$$

$$\hat{\delta}_{od}^B = \frac{(\hat{w}_{M,d} \hat{\mathbb{Y}}_d^{1/\theta} / \hat{e}_{od}^B)^{\epsilon_B}}{\sum_g \delta_{og}^B (\hat{w}_{M,g} \hat{\mathbb{Y}}_g^{1/\theta} / \hat{e}_{og}^B)^{\epsilon_B}} \tag{A.67}$$

$$\hat{\delta}_{od}^W = \frac{(\hat{w}_{S,d} \hat{\mathbb{Y}}_d^{1/\theta} / \hat{e}_{od}^W)^{\epsilon_W}}{\sum_g \delta_{og}^W (\hat{w}_{S,g} \hat{\mathbb{Y}}_g^{1/\theta} / \hat{e}_{og}^W)^{\epsilon_W}} \tag{A.68}$$

$$\hat{L}_{M,d} = \sum_o \hat{\delta}_{od}^B \frac{\delta_{od}^B \bar{L}_o^B}{L_{M,d}} \tag{A.69}$$

$$\hat{L}_{S,d} = \sum_o \hat{\delta}_{od}^W \frac{\delta_{od}^W \bar{L}_o^W}{L_{S,d}} \tag{A.70}$$

$$Y_i' = \sum_{n=1}^N \pi_{ni} \hat{\pi}_{ni} \left[X_n^{c\prime} + \Phi_n^{I\prime} Y_n' \right] \tag{A.71}$$

$$\Phi_n^{I\prime} = 1 - \beta_S - \beta_M(1 - \zeta_M) - \beta_M \zeta_M (1 - \bar{\omega}_{M,n}) \frac{\hat{w}_{M,n}^{-\phi}}{\hat{\Xi}_{M,n}} \tag{A.72}$$

$$X_n^{c\prime} = w_{M,n} L_{M,n} \hat{w}_{M,n} \hat{L}_{M,n} + w_{S,n} L_{S,n} \hat{w}_{S,n} \hat{L}_{S,n} \tag{A.73}$$

$$w_{M,i} L_{M,i} \hat{w}_{M,i} \hat{L}_{M,i} = \left[\beta_M \zeta_M (1 - \bar{\omega}_{M,i}) \frac{\hat{w}_{M,i}^{-\phi}}{\hat{\Xi}_{M,i}} + \beta_M (1 - \zeta_M) \right] Y_i' \tag{A.74}$$

$$w_{S,i} L_{S,i} \hat{w}_{S,i} \hat{L}_{S,i} = \beta_S Y_i' \tag{A.75}$$

E.10 Estimating the Change in Matching Efficiency

We decompose the bilateral matching efficiency as the interaction of unilateral matching efficiencies of origin-destination city pairs ($\lambda_{M,ni} = \lambda_{M,n} \times \lambda_{M,i}$). Then, equation (A.61) can be rewritten as

$$\hat{v}_{M,i} = \hat{\lambda}_{M,i} \sum_j \frac{v_{M,ij}}{v_{M,i}} \hat{\lambda}_{M,j} \hat{\Upsilon}_j^{\phi/\theta} \hat{d}_{ij}^{-\phi} \tag{A.76}$$

To solve for $\hat{\lambda}_{M,i}$, we define function $F(x)$ from $\mathbb{R}_+^N$ to $\mathbb{R}^N$ as:

$$F(x_i) = x_i \sum_j \frac{v_{M,ij}}{v_{M,i}} x_j \hat{\Upsilon}_j^{\phi/\theta} \hat{d}_{ij}^{-\phi} - \hat{v}_{M,i} \tag{A.77}$$

Let's denote the derivative of $F(x)$ with respect to x_i as $F'(x_i)$, which we calculate as the following equation:

$$F'(x_i) = \sum_j \frac{v_{M,ij}}{v_{M,i}} x_j \hat{\Upsilon}_j^{\phi/\theta} \hat{d}_{ij}^{-\phi} + x_i \frac{v_{M,ii}}{v_{M,i}} \hat{\Upsilon}_i^{\phi/\theta} \hat{d}_{ii}^{-\phi} \tag{A.78}$$

Since $F'(x_i) > 0$, $F(x)$ is a monotonically increasing function. Given that $F(x = 0) < 0$ and $\liminf F(x) > 0$, there exists a unique solution for $\hat{\lambda}_{M,i}$ such that $F(\hat{\lambda}_{M,i}) = 0$ for all i.