Connectivity and Selective Rural Migration for Agricultural Productivity Gains*

Lin Ma
Singapore Management University

Yuan Mei
Singapore Management University

Qunfeng Wu
Peking University

Mingzhi (Jimmy) Xu
Peking University

July 30, 2025

Abstract

We study how infrastructure shapes rural development using household panel data from a nationally representative sample of Chinese villages, matched with high-resolution infrastructure maps. We find that road expansion between 2000 and 2015 prompted less productive farmers to exit agriculture, while more productive ones remained and scaled up. This selection mechanism drove substantial productivity gains, complementing within-household improvements. We quantify these mechanisms through a dynamic spatial equilibrium model with heterogeneous households and endogenous migration. Improved connectivity not only accelerates urbanization—doubling the speed of convergence—but also increases aggregate rural output by 2.2 to 5.3 percent in the long run, even as the rural workforce declines. Policy counterfactuals show that relaxing China's farmland red line and reallocating land to higher productivity areas reinforces these selection gains, underscoring the central role played by heterogeneity and sorting in shaping development outcomes.

Keywords: transportation network, migration, agricultural productivity, dynamic quantitative spatial model
JEL Code: F12, O18, Q12, R13

*We thank Davin Chor and Xiaodong Zhu for their insightful comments and seminar participants at AMES 2023 (NTU), and Jinan-SMU Conference on Urban and Regional Economics. Lin Ma, Singapore Management University, Singapore, email: linma@smu.edu.sg; Mei Yuan, Singapore Management University, Singapore, email: yuanmei@smu.edu.sg; Qunfeng Wu, SOE at Peking University, Beijing, China, email: qfwu@pku.edu.cn; Mingzhi Xu, INSE at Peking University, Beijing, China, email: mingzhixu@nsd.pku.edu.cn. Financial support from the Natural Science Foundation of China (Grant No. 72322007) is gratefully acknowledged by Mingzhi Xu.

1 Introduction

Low-income countries are often characterized by inadequate transportation infrastructure, which imposes high internal transportation costs. These frictions not only restrict access to domestic and international markets, but also slow the reallocation of labor from agriculture to more productive non-agricultural sectors, often deepening the divide between rural and urban areas (Young, 2013; Pavcnik, 2017). Improvements in transportation infrastructure are therefore widely seen as a means to reduce spatial inequality, promote market integration, and support structural transformation. Yet, the impact is nuanced: while improved connectivity can raise rural incomes (Donaldson, 2018), it may also accelerate the exit of labor from agriculture (Asher and Novosad, 2020). The trade-off between strengthening the rural economy and encouraging urban transition lies at the heart of ongoing policy debates and motivates a closer examination of how infrastructure shapes the path of development.[1]

Understanding this trade-off requires a framework that captures the multiple, interconnected channels through which transportation infrastructure influences rural economic outcomes. Improved connectivity lowers trade costs, which raises farm revenues by expanding market access for agricultural products and reducing input prices and production costs. At the same time, better transportation infrastructure facilitates migration to urban labor markets. The expansion of urban centers may subsequently increase demand for agricultural goods, generating feedback effects on rural production. Moreover, these responses depend on household-specific productivity in both agricultural and non-agricultural activities, where improved sorting across sectors can generate aggregate efficiency gains. As these micro-level adjustments interact with general equilibrium forces in land and labor markets, assessing the effects of infrastructure investment requires detailed data and a model that incorporates trade, migration, and agent heterogeneity. Such a framework is essential for understanding both the welfare implications and the broader impacts on structural transformation and urbanization.

In this article, we combine two novel datasets to examine how the expansion of China's transportation network has shaped agricultural productivity and rural economic activity. The first is the National Fixed Point Survey (NFPS), a household-level panel survey launched in 1986 by the Research Center for Rural Economy under China's Ministry of Agriculture. Covering more than 300 representative villages across 30 provinces, the NFPS provides rich

[1] In recent years, about 20 percent of World Bank lending has gone toward transportation infrastructure (Baum-Snow et al., 2017; Donaldson, 2018). According to Fan et al. (2023), "In 2016, the 47 member countries of the International Transport Forum—including OECD countries, China, and others—invested over 850 billion euros in inland transport infrastructure."

longitudinal data on agricultural production and rural-to-urban migration at both the household and individual levels. To measure evolving infrastructure access, we draw on newly constructed, high-resolution data from Ma and Tang (2024), who move beyond conventional binary road indicators by incorporating pixel-level measures of road and rail quality over time. Combined, these datasets provide an unusually detailed view of both rural economic behavior and spatial infrastructure exposure, enabling a comprehensive analysis of how improved connectivity affects production, migration, and reallocation across sectors.

We begin by examining the effects of improved village connectivity—measured by the average minimum travel time from each village to all Chinese cities—on rural economic outcomes. The causal effect of improved market access is identified by comparing households located in villages that experienced different timings and intensities of connectivity improvements, controlling for household-specific characteristics and broader regional trends. We find that improved connectivity significantly raises both total factor productivity (TFP) and labor productivity among farming households. These gains are strongly heterogeneous: more productive households benefit more from improved access, while less productive ones are more likely to exit agriculture and migrate to urban areas. This selective out-migration increases average productivity in the rural sector, as those who remain are more efficient and better able to take advantage of improved access to inputs and capital goods. Our results are robust to a range of alternative village-level connectivity measures and to an instrumental variable strategy based on the least-cost spanning tree (Faber, 2014). A decomposition exercise shows that most of the observed gains stem from within-household productivity improvements. Reallocation across households and sectoral exit further reinforce the selection mechanism: as less productive farmers leave, more productive ones remain and expand their production.

Motivated by these empirical findings, we develop a dynamic, quantitative spatial equilibrium model of trade and migration across multiple regions. We build on and extend the framework of Caliendo et al. (2019) by introducing heterogeneous rural households, each with different comparative advantages across sectors. Rural regions produce agricultural goods using land, manufactured intermediates, and capital goods, while manufacturing and capital goods are produced exclusively in urban centers. Households choose whether to remain in agriculture or migrate to the urban manufacturing sector based on their sectoral productivity and prevailing economic conditions, adjusting their decisions over time in response to evolving connectivity and wages. This intertemporal structure captures how sectoral and locational choices evolve over time as infrastructure improves. These gradual reallocations enhance efficiency in rural production through selection and scale effects, while the expan-

sion of urban markets raises demand for agricultural goods.[2] Together, these forces sustain productivity growth in agriculture even as labor shifts toward urban areas, shaping not only long-term outcomes but also the pace and trajectory of urbanization—dynamics that static models that compare only steady-state equilibria cannot capture.

We calibrate the model across 279 prefectures that account for the vast majority of China's population and economic activity, drawing on both micro-level and aggregate data. We begin by solving for the steady state consistent with the 2015 transportation network and use a shooting algorithm to compute the full transition path from the initial population distribution in 2000. This approach captures not only long-term impacts but also the timing and pace of adjustment. Key parameters governing trade and migration elasticities, productivity dispersion, and sectoral preferences are calibrated using household-level data from the NFPS, which tracks agricultural production, input use, and migration histories. Region-level statistics on sectoral output shares, migration rates, and land use provide additional inputs for the calibration. At the micro level, the calibrated model aligns closely with observed behavioral responses, including positive selection into migration and productivity gains among those who remain in agriculture, as documented in our reduced-form analysis.

The calibrated baseline economy, incorporating China's transportation network expansion between 2000 and 2015, closely replicates the country's observed macro trends. For example, China's agricultural employment declined during the same period, whereas the urbanization rate rose from 38 percent to 52 percent. In addition, total agricultural output continued to increase despite substantial rural labor outflows. Inconsistent with standard models with homogeneous agents, these patterns can be explained by the heterogeneous household productivity and dynamic sorting in our model: improved connectivity increases migration probabilities across the productivity distribution but disproportionately encourages less productive individuals to leave agriculture, while more productive farmers remain and expand their operations.[3] As a result, rural output increases by 2.2% to 5.3% over the long run (2015—2045) relative to the "no change" counterfactual in which trade and

[2]Previous work has often emphasized trade-cost reductions but has paid less attention to how migration-driven urban growth raises demand for rural output. For instance, Bryan and Morten (2019) study rural-urban migration and its productivity effects, but abstracts from its implications for demand in origin regions. As Asher and Novosad (2020) note, in the absence of such external demand, better roads alone may simply lead to rural depopulation without gains in agricultural output or income.

[3]Models with homogeneous agents, such as that of Tombe and Zhu (2019), may generate aggregate output gains without necessarily predicting corresponding growth in agricultural output. On the empirical side, Faber (2014), which abstracts from the agricultural sector and heterogeneity in migration behavior, finds that China's National Trunk Highway System reduced GDP growth in peripheral counties. By contrast, our analysis is the first to show that improved road access can support both urbanization and agricultural growth. By enabling more efficient resource reallocation, not only across sectors but also within agriculture among heterogeneous households, enhanced connectivity in our model delivers broader gains than previously recognized.

migration costs remain at 2000 levels.

At the aggregate level, the dynamic sorting of heterogeneous households substantially accelerates the pace of urbanization. In the baseline scenario, the urbanization rate converges to its steady state in roughly 115 years—more than twice as fast as in the "no change" counterfactual, where convergence takes around 270 years. A decomposition reveals that the faster urbanization is primarily driven by lower migration costs rather than by reductions in internal trade costs. Critically, however, the model also shows that migration liberalization without concurrent improvements in trade access leads to a long-term decline in agricultural output. In this case, rural producers are unable to fully capitalize on rising urban demand, and the premature loss of productive labor reduces efficiency in the agricultural sector. This outcome again hinges on heterogeneity in household comparative productivity and the resulting spatial sorting. It is not just the volume of migration that matters, but which households move and which remain. Without this heterogeneity, even a coordinated reduction in trade and migration frictions fails to raise the urbanization rate or rural output in the long run.

To further illustrate the productivity gains from selection, we conduct a counterfactual that partially relaxes China's Farmland Red Line Policy, which mandates a minimum threshold of preserved arable land at both the national and local levels. In particular, we hold the total land supply fixed but shift land toward more productive villages, equalizing the marginal return to land across villages. In theory, this reallocation can increase aggregate rural output by enabling higher productivity households to expand production. The novelty of this exercise, however, lies in its dynamic general equilibrium perspective. Unlike much of the existing literature (Adamopoulos and Restuccia, 2020; Adamopoulos et al., 2024), which abstracts from the broader economic implications of land reallocation, our analysis captures the time-path effects on factor prices, migration, urbanization, and internal trade. We find that re-optimizing land quotas across space increases total rural output by 8.2 percent. Importantly, this reallocation does not slow the urbanization process. The urbanization rate under the computed "optimal land scheme" is slightly higher than in the baseline equilibrium, reaching 47.5 percent in 2015 compared to 47.4 percent. These findings reinforce a central message of the paper: better aligning resources with high-productivity producers, whether through infrastructure investment or institutional reform, can significantly magnify the gains from structural transformation.

This paper contributes to the extensive literature on the economic impacts of investments in transportation infrastructure. The general equilibrium effects of transportation infrastructure projects have been explored in trade or spatial models (Donaldson and Hornbeck, 2016; Costinot and Donaldson, 2016; Donaldson, 2018; Asturias et al., 2019; Sotelo, 2020; Fan

et al., 2023; Bonadio, 2024; Allen and Arkolakis, 2022; Tsivanidis, 2022). Several of these studies, including those of Costinot and Donaldson (2016), Sotelo (2020), and Adamopoulos (2025), focus specifically on the implications for agricultural production. We advance this literature by developing a dynamic spatial model in which individuals' migration and occupational decisions respond not only to current infrastructure but also to its expected evolution over time. This forward-looking behavior is essential for capturing the transitional dynamics of structural transformation and the long-term reallocation of labor and resources across space.[4] Complementing these studies, related empirical works have estimated the local effects of transportation network expansion (Duranton et al., 2014; Faber, 2014; Qin, 2016; Lin, 2017; Banerjee et al., 2020; Dumas and Játiva, 2024). Most relevant to our study are Asher and Novosad (2020) and Shamdasani (2021), who document that India's rural road program enhances agricultural technology adoption and facilitates labor reallocation away from agriculture. Similarly, Adamopoulos (2025) finds that road development in Ethiopia has comparable effects. Like these studies, our paper employs a credible identification strategy to estimate the causal impacts of transportation infrastructure—specifically on rural economic activities in China.

Recent work in the spatial economics literature has explored migration dynamics in the presence of mobility costs (Monte et al., 2018; Fan, 2019; Tombe and Zhu, 2019). A complementary line of research emphasizes the role of heterogeneous labor productivity and the resulting variation in migration incentives across individuals (Lagakos and Waugh, 2013; Ngai et al., 2019; Bryan and Morten, 2019; Gai et al., 2024). Closest to our study is that of Morten and Oliveira (2024), which examines how Brazil's highway network shapes migration and trade across sub-state regions. Our paper builds on this work by combining empirical evidence and a dynamic spatial model to highlight how heterogeneity in household productivity drives selection and sorting across space. This focus on who migrates and who remains deepens our understanding of rural–urban transformation and adds a new dimension to the literature on infrastructure and migration in developing economies.

Finally, this paper complements a growing body of research on structural transformation and rural development in China using NFPS data. Previous studies have examined the role of land property rights (Chari et al., 2021; Adamopoulos et al., 2024), within-village frictions in both the land and capital markets (Adamopoulos et al., 2022), rural pension policy (Gai et al., 2024), and increasing manufacturing trade exposure (Tian et al., 2020). We extend this literature by examining how improved access to cities—driven by the expansion of China's

[4] Ma and Tang (2024) also develop a dynamic spatial general equilibrium model following Caliendo et al. (2019), focusing on the distributional effects of transportation networks across Chinese cities. However, their framework excludes agricultural production, omits rural areas entirely, and does not account for ability heterogeneity, leaving open the question of how infrastructure shapes outcomes in the rural economy.

transportation network—shapes agricultural productivity and rural economic outcomes. To our knowledge, this is the first study to calibrate a dynamic spatial model using NFPS microdata to quantify the general equilibrium effects of national infrastructure investment.

The rest of the paper is organized as follows. Section 2 outlines the institutional context of China's rural economy, migration patterns, and transportation network expansion. Section 3 describes the data and presents the empirical analysis. Section 4 introduces the dynamic spatial equilibrium model, followed by calibration in Section 5. Section 6 uses the model to quantify the welfare impacts of infrastructure expansion. Section 7 concludes.

2 Institutional Background

Small-Scale Farming and Agricultural Labor in China China has historically been an agrarian economy, dominated by a vast rural population and characterized by smallholder farmers operating fragmented plots with limited economies of scale (Adamopoulos et al., 2024).[5] While land remains collectively owned, rural households have been granted cultivation rights, allowing them to make independent production decisions. More recently, policy changes have further expanded farmers' autonomy by enabling them to transfer these cultivation rights, facilitating a more flexible allocation of land resources (Chari et al., 2021). Farmers can now sell their products at market prices through various channels, including direct sales to the government, wholesalers, retailers, and agribusiness firms.[6]

At the same time, China's small-scale farmers face significant constraints, including limited access to formal capital from urban industrial sectors and restricted opportunities outside agriculture due to trade and migration barriers. The dismantling of the collective farming system under the planned economy and its replacement with a decentralized, household-based model (McMillan et al., 1989) granted rural households greater autonomy over land

[5]Prior to the economic reforms of the late 1970s, agricultural production was organized under a collective system in which farmers had no autonomy in decision-making. Crop selection, pricing, sales, and all other aspects of production were strictly controlled by the government, leaving little room for individual incentives or market-driven adjustments. The introduction of reform and opening-up policies in the late 1970s fundamentally transformed the legal and economic structure of China's agricultural sector.

[6]As extensively discussed by Huang et al. (2008), the market has gradually replaced the state as the primary mechanism for resource allocation in China's agricultural sector since the late 1980s and early 1990s. This transition reduced distortions from domestic policies and laid the foundation for farmers' production and marketing decisions. By 1995—the beginning of our sample period—the nominal protection rate for staple grains such as rice and wheat, as well as key agricultural outputs such as soybeans, milk, sugar, and pork, had declined to nearly zero. As a result, farmers primarily generate revenue by selling their products in markets, either through private transactions or through state procurement programs. While the government continues to intervene in staple crops such as rice and wheat through the Minimum Purchase Price (MPP) program, its primary role is to ensure food security and stabilize prices, preventing extreme fluctuations that could disrupt production. Nevertheless, for the majority of rural households, market-based income remains the dominant source of earnings.

use. However, institutional barriers continue to shape entry and exit dynamics in agriculture. Exiting the sector is relatively straightforward, as farmers lose exclusive land-use rights upon permanent migration. Entry, however, remains restricted. Since rural land is collectively owned and allocated by village committees, individuals from outside the village have limited access if they lack prior land-use rights (Minale, 2018).

Expanded economic opportunities from urban sectors—driven by large-scale transportation infrastructure improvements (Ma and Tang, 2024) and the relaxation of institutional migration restrictions (Hsu and Ma, 2021)—have enabled small-scale rural households to shift labor away from farming to maximize their overall income. However, agricultural labor participation in rural China remains shaped by a combination of institutional constraints, market conditions, and non-agricultural opportunities. Unlike urban labor markets, where employment is primarily determined by firms' hiring decisions, rural households engage in farming as self-employed producers. Their entry and exit decisions are influenced by land tenure policies (Chari et al., 2021), migration opportunities (Gai et al., 2024), and the profitability of agricultural production (Imbert et al., 2022). Among these, migration serves as the most significant adjustment channel (Gai et al., 2024). China's rapid urbanization has created strong incentives for rural workers to move to cities in search of higher wages in manufacturing and services. However, the household registration (hukou) system restricts access to social benefits such as education and healthcare, creating barriers to permanent urban settlement and shaping migration decisions. As a result, many rural workers engage in seasonal or temporary migration, balancing farm work with off-farm employment to supplement household income. This flexibility enables them to reallocate labor in response to shifts in agricultural profitability and broader economic conditions.

Highway Expansion and Rural–Urban Connectivity China's modern highway development took off in the early 1990s with the launch of the National Trunk Highway System in 1992. Designed to establish a nationwide expressway network, the system aimed at not only connecting major economic hubs but also improving access to rural regions. The scale of construction was unprecedented: in 1990, China had fewer than 600 kilometers of highways, but by 2013, the total length had surpassed 100,000 kilometers, making it the largest expressway network in the world. This rapid expansion was fueled by substantial public investment, with both central and local governments funding large-scale infrastructure projects to facilitate industrialization and economic growth (Banerjee et al., 2020).

A defining aspect of China's highway expansion was its critical role in strengthening central–peripheral and rural–urban connectivity (Faber, 2014). Prior to the 1990s, many rural areas were accessible only via low-capacity roads or unpaved paths, restricting their

integration with broader markets. The development of the highway network transformed this by establishing direct, high-speed connections between rural regions and urban centers, facilitating the movement of goods, labor, and capital. Between 1995 and 2013, by our own calculation, the average travel time from villages to all cities fell by 28%, from 1,487 minutes to 1,071 minutes.[7] This improved connectivity had profound economic effects. First, it lowered transportation costs for rural producers, enabling farmers to access larger urban markets more efficiently. As highlighted in the World Bank's Transport Business Strategy 2008–2012, "One of the best ways to promote rural development is to ensure good accessibility to growing and competitive urban markets" (World Bank, 2008). Perishable goods, such as fruits and vegetables, could now be transported with less spoilage, increasing market opportunities and reducing waste. Second, improved connectivity enhanced access to agricultural inputs and capital, which are predominantly produced in urban locations.[8] Lower transportation costs made it easier for rural farmers to acquire fertilizers, machinery, and financial services, fostering productivity growth and strengthening rural–urban economic linkages. Third, better transportation infrastructure facilitated rural-to-urban migration by reducing travel costs and commute times, enabling workers to access higher wages in cities while maintaining ties to their home villages.

3 Empirical Analysis and Data Description

In this section, we empirically examine how improved access to urban markets affects agricultural production efficiency in China and explore the mechanisms behind these changes. Specifically, we focus on the impact of the country's highway network expansion on agricultural TFP and draw on household-level data to identify the underlying channels of adjustment.

3.1 Specification

We analyze the relationship between a household's proximity to input and output markets and its agricultural TFP using the following general specification:

$$\ln y_{hvt} = \beta_1 \ln Distance_{v,t-1} + \beta_2 X_{hvt} + \delta_h + \delta_{pt} + \epsilon_{hvt}. \tag{1}$$

[7] Appendix Figure B.1 presents the trend of villages' average travel time to all cities from 1995 to 2013.

[8] Agricultural production in developing countries often suffers from a lack of essential inputs (Restuccia et al., 2008) and modern capital (Caunedo and Keller, 2021), partly due to high trade barriers between urban and rural areas, including transportation costs.

Our dependent variable, y_{hvt}, measures the agricultural productivity of household h in village v at time t. We explore various alternative measures, including whether the household's primary activity is agriculture, as well as corresponding agricultural inputs and TFP. The key variable of interest, $\ln Distance_{v,t-1}$, represents the average distance from village v to all "cities" in China in the previous period, capturing proximity to input and output markets for agricultural production.[9] We include a vector of control variables, X_{hvt}, at both the household and city levels, which may be correlated with agricultural production and road network expansion. Household fixed effects (δ_h) control for time-invariant characteristics specific to households that might affect agricultural productivity, while province–year fixed effects (δ_{pt}) account for unobserved changes in regional policies or economic cycles. To handle the potential correlation of households within villages over time, we cluster standard errors at the village level. We begin by outlining the construction of each variable and its corresponding data source, with more detailed information provided in Appendix B.1. Our sample consists of 26,281 households from 377 villages and 30 provinces across mainland China.[10]

Agriculture Production Performance The analysis focuses on three key household-level outcome variables—farming decisions, agricultural output, and agricultural productivity—all derived from the NFPS. The first variable captures the extent of household participation in farming, while the latter two measure the intensity of agricultural production within each household.

The NFPS asks households to self-report their primary operating sector from nine categories: farming, forestry, animal husbandry, fishery, industry, construction, transportation, services, and others. We define the farming decision as a dummy variable that is set to 1 if the household's main sector is farming, and zero otherwise. This value can change over time within the same household. Alternatively, we identify households as being primarily engaged in agricultural production if the share of agricultural income, defined as the ratio of agricultural income to total income, is the largest compared to other sectors.

We measure agricultural output per household as the total quantity produced across thirteen major crops reported in the NFPS: wheat, rice, corn, soybeans, tubers, cotton, oil crops, sugar crops, flax, tobacco, mulberry (for silkworms), fruits, and vegetables. Instead of relying on self-reported sales—which often understate true production due to on-farm consumption—we use either physical quantities or imputed total value. This approach better reflects actual output, especially in rural China where subsistence farming is common. For

[9] As we explain later, we refer to the urban part of a prefecture-city simply as a "city".

[10] Appendix Figure B.2 presents a map illustrating the geographical distribution of surveyed households. The coverage spans the vast majority of mainland China.

each year, household output is calculated by summing production across all crops. While this yields a consistent measure of physical output, it may underweight high-value crops produced in smaller volumes. To account for this, we also construct a revenue-based measure of output, calculated in three steps as detailed below. First, we calculate household-specific crop prices by dividing reported sales revenue by quantities sold. Second, we construct national year-level average prices by averaging household prices for each crop, smoothing out idiosyncratic fluctuations. Third, we multiply crop quantities by these average prices and aggregate across crops to obtain household-level revenue-based output. All monetary values are adjusted to 2002 constant prices.

In addition, we estimate household-level agricultural TFP using a standard Cobb-Douglas production function, following the approach in Chari et al. (2021).[11] Details of the estimation procedure—including the construction of crop-level input and output measures using the NFPS data—are provided in Appendix B.2.

Road Connectivity of Village Our key variable of interest, the average distance of each village to all major cities in China, is constructed using transportation data, following the method developed by Ma and Tang (2024). To measure each village's connectivity to the broader country, we follow these steps: first, we pinpoint each village's exact location on the transportation map using detailed coordinates from the NFPS data; second, for each village v, we compute the shortest travel time to every city c in China for each year t using the variable $Distance_{vct}$. We use the coordinates of city centers from Ma and Tang (2024), who measure the center of the urban part of each prefecture-city. This measure builds on the methodology developed by Ma and Tang (2024), which accounts for both the quantity and quality of road infrastructure over time. Transportation quality is proxied by design speed, which varies by vintage, road class, and terrain at the pixel level.[12] Ignoring these quality differences would introduce significant bias, especially for rural villages with diverse geographical characteristics. Finally, we derive the average distance from each village v to all cities in China for each year t as follows:

$$Distance_{vt} = \sum_{c} Distance_{vct}/N_c, \tag{2}$$

[11] Chari et al. (2021) estimate TFP at the household–crop–year level; by contrast, we estimate it at the household–year level, following the procedure used in Tian et al. (2020).

[12] Following Ma and Tang (2024), we include both highways and "first-class roads" to ensure comprehensive coverage of major transportation routes, consistent with official yearbook classifications and historical records. This approach also yields a more exogenous measure of access in our empirical setting, as villages and households are unlikely to influence the construction of such higher level infrastructure.

where c denotes a city, and N_c represents the overall number of cities. By averaging the distance variable across all cities, we can assess the comprehensive impact of extensive transportation infrastructure on rural China's agricultural sectors, rather than just local effects. For added robustness, we calculate weighted average distances, applying each city's population or GDP as weights to capture their differing significance to rural villages. Appendix Figure B.3 illustrates the changes in average travel distances across regions from 1995 to 2013. Both remote and centrally located villages experienced a significant reduction in travel time to urban centers, with declines ranging from 4 to 15 hours on average, reflecting the significant impact of China's large-scale transportation infrastructure investments. Panel A of Appendix Table B.1 provides detailed summary statistics for our distance indicator.

Other Control Variables In addition to controlling for household-invariant characteristics with fixed effects, we account for a range of time-varying household characteristics that may influence agricultural performance. These are captured by dummy variables indicating whether any household member is an officer, party member, or member of the military. We also include a dummy variable for households eligible for the Chinese government's Five-Guarantee Program, which supports the most disadvantaged families.[13]

As China's growing openness to global trade during our sample period could significantly affect rural household agricultural performance and migration decisions, we account for trends related to a city's industrial structure by incorporating the initial tariffs on imported inputs and final goods imposed by China in 2000, as well as those imposed by countries importing Chinese goods. These tariffs have been shown to impact the productivity of Chinese firms (Yu, 2015). We obtain import tariffs, averaged across origins, from the World Integrated Trade Solution's Trade Analysis and Information System, using the applied tariff rates from 2000. Input tariffs are computed using the 2002 input–output table for China.[14] We map this industry data to the prefecture level using the 1997 export shares to create a weighted sum of both import and input tariffs. In our regressions, we also include interactions between these measures and time dummies to capture fluctuations in global economic conditions. Detailed descriptions of the data used in variable construction are provided in Appendix B.1.

[13]The Five-Guarantee Program in China is a rural social welfare initiative designed to support vulnerable individuals who lack family or financial means. It provides assistance to elderly, disabled, or orphaned residents in rural areas by covering five basic needs: food, clothing, housing, medical care, and funeral expenses. The program targets those who are unable to work and who have no family support, ensuring that they receive essential care. Administered at the village level, the support is provided either through direct financial and in-kind aid or through placement in welfare institutions, such as rural nursing homes, forming a critical part of China's rural social safety net.

[14]The input–output table is available for 120 industry groups ("scode" classification), of which 70 are manufacturing industries.

3.2 Empirical Results

Table 1 reports the estimated impact of transportation connectivity on household agricultural performance across several dimensions. Columns (1) and (2) show that closer proximity to cities significantly raises both total factor productivity ($\ln TFP_{ht}$) and labor productivity ($\ln LP_{ht}$), suggesting that improved access to markets, inputs, and information plays an important role in driving efficiency. Connectivity also translates into higher agricultural output, measured in terms of both total value and physical quantity, as shown in columns (5) and (6). Beyond these productivity gains, the positive and statistically significant coefficients in columns (3) and (4) suggest that improved connectivity encourages some households to shift away from farming as their primary activity, leading to a lower share of income derived from agriculture—even as overall agricultural productivity and income rise. This pattern points to uneven gains from better access to cities, with some households benefiting from increased returns to farming, while others exit agriculture altogether. These findings suggest that improved transportation access not only raises agricultural productivity but also enables rural households to earn more from farming. Based on the estimated coefficients and a 28% average reduction in travel distance between 1995 and 2013, road expansion leads to a 10.7% increase in TFP, a 23.7% rise in labor productivity, and an 11.6% increase in agricultural output (by quantity). It also results in a 4.2% increase in the value of output. At the same time, better connectivity contributes to a 6.0% decline in the likelihood of farming and a 4.6% drop in the share of household income from agriculture—evidence of structural transformation. These effects are economically significant and comparable to major policy reforms; for instance, Chari et al. (2021) report that rural land rights reform increased agricultural output and productivity by 8% and 10%, respectively.

Panel B of Table 1 explores the heterogeneity of these effects by interacting distance with the household fixed productivity ($\ln TFP_{ht_0}$).[15] The findings reveal a strong distributional effect, whereby productivity gains are concentrated among highly productive households rather than being evenly distributed across all productivity levels. The negative and significant interaction terms in columns (1) and (2) suggest a selection effect: households with higher productivity appear better positioned to benefit from enhanced access to urban centers and markets, thereby amplifying their productivity edge. The negative interaction terms in columns (3) and (4) further indicate that these high-productivity households are more inclined to specialize in crop farming and increase their income share from this sector, likely reflecting their superior capacity to scale production in response to improved mar-

[15]The household fixed productivity is calculated by capturing the (time–invariant) household fixed effect in the TFP estimation process. See Appendix B.2 for the detailed estimation procedure. This method is consistent with Chari et al. (2021), who estimate TFP at the household–crop level.

ket access. This pattern is consistent with the substantial negative coefficients observed in columns (5) and (6), underscoring that productivity levels play a critical role in maximizing the agricultural output from connectivity improvements.[16]

	Productivity		Extensive Margin		Intensive Margin	
	(1)	(2)	(3)	(4)	(5)	(6)
Dep var.	$\ln TFP_{ht}$	$\ln LP_{ht}$	$\mathbb{I}(Farm)_{ht}$	$IncSh_{ht}^{Farm}$	$\ln Output_{ht}^{V}$	$\ln Output_{ht}^{Q}$
Panel A: Average Effect						
$\ln Distance_{v,t-1}$	-0.382**	-0.845***	0.216***	0.164***	-0.150	-0.416**
	(0.149)	(0.241)	(0.063)	(0.038)	(0.251)	(0.173)
Obs	254950	303530	376057	375160	317084	317104
R^2	0.828	0.657	0.568	0.692	0.674	0.712
Panel B: Heterogeneity Pattern						
$\ln Distance_{v,t-1}$	-0.383***	-0.756***	0.031	0.137***	-0.120	-0.325*
	(0.149)	(0.232)	(0.056)	(0.036)	(0.229)	(0.184)
$\ln Distance_{v,t-1} \times \ln TFP_{ht_0}$	-0.559***	-2.336***	-0.084**	-0.147***	-3.435***	-2.877***
	(0.106)	(0.156)	(0.037)	(0.025)	(0.169)	(0.150)
Obs	254950	254949	254950	254850	254949	254950
R^2	0.829	0.664	0.503	0.689	0.679	0.722

Table 1: Impact of Transportation on Household's Agricultural Production

Note: Each observation represents a household–year, and the sample contains households that appear at least twice in the sample from 1995 to 2013. $lnTFP_{ht_0}$ refers to (time–invariant) household fixed agricultural productivity (See Appendix B.2 for the detailed estimation procedure). In column (1), the dependent variable is the household TFP in agricultural production. Column (2) examines log labor productivity, calculated as the ratio of total agricultural output value to agricultural labor-day inputs. Column (3) uses a binary indicator that is set to 1 if crop farming is the primary business, and zero otherwise, with other options including forestry, animal husbandry, fishery, industry, construction, transportation, services, and others. Column (4) measures the share of crop income in total household income, while columns (5) and (6) capture the total crop output in 2002 RMB value and quantity, respectively. All regressions control for household and province–year fixed effects. Standard errors, clustered at the village–year level, are reported in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

3.3 Robustness

Instrumental Variable Approach While we argue that our identification is less prone to reverse causality due to the aggregate nature of the shocks used in constructing village distance measures, the selection of nodal cities within the road networks may not be random.[17]

[16]These results are robust to using either quantity-based or revenue-based measures of TFP.

[17]The construction of highways and "first-class roads" is determined by higher government authorities, which minimizes the influence of individual villages and effectively mitigates reverse causality. Furthermore, when calculating each village's average distance, we consider distances to all domestic cities rather than concentrating on one or two specific cities. This method further reduces potential reverse causality arising from the influence of nearby major cities.

These selections can be influenced by a city's socio-economic characteristics, potentially affecting nearby villages and raising endogeneity concerns. To address this, we use a least-cost path (LCP) spanning tree network to construct an instrumental variable for actual highway connections. The LCP network represents a hypothetical road system that is more independent of regional economic conditions, as its placement relies primarily on exogenous factors affecting road construction feasibility. Regions located along the LCP are poised to gain disproportionately from extensive national transportation investments. Specifically, our instrumental variable in (2) builds on the approach of Baum-Snow (2007), who leverages the interaction between the 1947 U.S. national highway plan and annual variations in highway mileage to evaluate the effect of new highway construction on suburbanization:

$$IV_{vt} = D_v \times \ln TransportInvest_{t-1}, \tag{3}$$

where D_v is a dummy variable that is equal to 1 if village v lies on the LCP of China's highway network, as introduced by Faber (2014). The variable $\ln TransportInvest_{t-1}$ represents the log value of nationwide transportation investment in year $t-1$.[18] Table B.2 presents the 2SLS regression results using our instrumental variable, which align closely with the baseline findings. The results confirm that proximity to urban centers significantly enhances agricultural output in villages, with the largest gains observed among farmers who already exhibit high productivity. Notably, the instrumental variable estimates produce considerably larger coefficients than those obtained from ordinary least squares (OLS). In addition to the potential attenuation bias from measurement errors, this also suggests that the instrumental variable approach more effectively captures the full extent of spillover effects from transportation investments—effects that OLS likely underestimates.

Alternative Measures for Village Proximity to Cities Our baseline analysis uses the simple average distance of each village to all cities. As a robustness check, we also apply a population-weighted distance measure, following the market access approach developed by Donaldson and Hornbeck (2016). Table B.3 presents these findings, demonstrating that our results remain robust to this alternative measure of transportation access. In addition, we further experiment using a GDP-weighted average distance measure, which likewise supports the robustness of our original results.

Excluding Influential Cities China's large cities, especially provincial capitals, could substantially impact the economic dynamics of surrounding rural areas. To ensure that our findings are not skewed by villages influenced by these major urban centers, we conduct a

[18] This data is obtained from the National Bureau of Statistics (NBS) of China.

robustness check by excluding all provincial capitals when calculating each village's distance to nearby cities. The results, reported in Table B.4, reaffirm that proximity to urban centers enhances agricultural productivity—particularly for villages with an initial advantage in agricultural output.

3.4 Mechanism

We provide evidence on the potential mechanisms driving the rise in agricultural efficiency among Chinese villages closer to cities. First, we find that increasing connectivity can improve farmers' access to inputs produced in urban sectors, thereby enhancing productivity. Second, as access to urban job opportunities increases, less efficient farmers are more likely to exit the agriculture sector, leaving behind a workforce that is more skilled and specialized. This compositional shift raises average productivity by reallocating resources toward individuals with a comparative advantage in farming.

Agricultural Inputs and Technologies Lower transportation costs can substantially improve farmers' access to agricultural inputs and capital goods produced in urban sectors, thereby enhancing productivity. To examine this relationship, Table 2 presents regression results on how proximity to urban areas influences agricultural input use. Panel A focuses on these impacts: columns (1), (2), and (3) assess household labor inputs, capturing total days worked, the number of agricultural workers, and the share of labor devoted to farming. Columns (4) and (5) evaluate how proximity affects decisions to invest in capital goods and the use of intermediate inputs. Finally, column (6) examines changes in agricultural land use as farmers benefit from closer ties to urban centers.

The results in Panel B once again indicate that the observed improvements in agricultural productivity following road expansions are more closely linked to a selection effect than to uniform productivity gains across all households. As villages gain better access to urban centers, less productive households are, in fact, more inclined to reduce their agricultural inputs—whether in terms of labor days, agricultural workforce, capital equipment, or intermediate inputs such as fertilizers and pesticides. Notably, a reduction in distance has a more pronounced negative effect on input usage among households with lower productivity, as reflected in the significant interaction terms across the columns. This pattern suggests that improved access to urban job markets may encourage less efficient farmers to either exit agriculture altogether or shift their labor toward more productive non-agricultural sectors. Consequently, the increase in overall village productivity appears to stem not from uniform

efficiency gains, but from the reallocation of resources away from less efficient producers.[19]

	Labor Input			Capital Input	Intermediate Input	Land Input
	(1)	(2)	(3)	(4)	(5)	(6)
Dep var.	$\ln FarmDays_{ht}$	$\ln Num_{ht}^{Farm}$	$NumSh_{ht}^{Farm}$	$\ln K_{ht}$	$\ln M_{ht}$	$\ln N_{ht}$
Panel A: Average Effect						
$\ln Distance_{v,t-1}$	0.705***	0.260***	0.110**	-0.361***	-0.142	-0.005
	(0.144)	(0.063)	(0.051)	(0.119)	(0.172)	(0.104)
Obs	305868	375200	356097	305321	317487	315602
R^2	0.675	0.621	0.643	0.835	0.724	0.845
Panel B: Heterogeneity Pattern						
$\ln Distance_{v,t-1}$	0.636***	0.253***	0.084*	-0.486***	-0.267	-0.021
	(0.150)	(0.064)	(0.048)	(0.128)	(0.173)	(0.112)
$\ln Distance_{v,t-1} \times \ln TFP_{ht_0}$	-1.099***	-0.599***	-0.243***	-0.673***	-1.363***	-0.392***
	(0.090)	(0.042)	(0.033)	(0.082)	(0.104)	(0.076)
Obs	254950	254375	246647	254950	254950	254950
R^2	0.664	0.533	0.578	0.830	0.732	0.849

Table 2: Impact of Transportation on Agricultural Inputs

Note: The unit of observation is a household–year. We include households that appear at least twice in our sample period between 1995 and 2013. $lnTFP_{ht_0}$ refers to (time–invariant) household fixed agricultural productivity (see Appendix B.2 for the detailed estimation procedure). The dependent variable in column (1) is the number of labor days a household allocates to agriculture. Column (2) measures the household's agricultural labor force, while column (3) captures the share of agricultural labor relative to the total household labor force. In column (4), the dependent variable is the household's agricultural capital stock (measured in 2002 RMB, reconstructed using the perpetual inventory method). Column (5) examines household agricultural land use (measured in Chinese "mu", which equals 666.67 square meters), and column (6) focuses on the intermediate inputs in agriculture (measured in 2002 RMB). The intermediate inputs include five categories: seeds, fertilizer, diesel fuel, plastic film, and pesticides. All regressions control for household and province–year fixed effects. Standard errors, clustered at the village–year level, are reported in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

Migration Away from Agricultural Production Table 3 reports the effects of improved transportation connectivity on individual migration decisions.[20] Panel A reports the average effect of proximity to urban centers on migration. The negative and statistically significant coefficient in column (1) suggests that individuals living in villages with better access to cities are more likely to send migrants. Column (2) further shows that migration earnings are significantly higher in these villages, reinforcing the idea that proximity to urban markets enhances labor mobility and access to higher paying opportunities. Panel B

[19]Consistent with the patterns in Table 2, Appendix Figure B.5 plots the coefficients and 95% confidence intervals from a regression of household agricultural output (log values) on productivity quartiles. The results reveal a clear selection pattern: output drops sharply among households in the bottom quartile, while those in the top quartile see corresponding gains. Similar trends appear across other key outcomes. These shifts suggest that improved road connectivity facilitates migration, inducing the exit of less productive farmers and reallocating resources toward more capable producers, thereby driving aggregate productivity growth.

[20]The empirical specification follows equation (1), except that it is estimated at the individual level. The dependent variables are replaced by individual migration outcomes, and individual fixed effects are included.

examines heterogeneity by agricultural productivity. While the main effect of distance remains negative, the interaction term between distance and agricultural TFP is positive and significant for both migration incidence (column (1)) and migration income (column (2)). This suggests that migration is disproportionately driven by households in villages with lower agricultural productivity, where the opportunity cost of exiting farming is lower. These results suggest a sorting-based explanation for the observed gains in agricultural productivity, whereby improved connectivity facilitates the exit of less efficient farmers, leaving behind a more specialized and productive agricultural workforce. This pattern also aligns with the idea that better transportation reduces mobility frictions, enabling labor to reallocate more efficiently in response to economic incentives beyond the village.

Dep var.	(1) $\mathbf{1}_{it}$(Migration Out)	(2) $\ln MigEarn_{it}$
Panel A: Average Effect		
$\ln Distance_{v,t-1}$	-0.115**	-1.082*
	(0.054)	(0.593)
Obs	282899	282899
R^2	0.665	0.691
Panel B: Heterogeneity Pattern		
$\ln Distance_{v,t-1}$	-0.083	-0.527
	(0.056)	(0.631)
$\ln Distance_{v,t-1} \times \ln TFP_{ht_0}$	0.125*	1.947***
	(0.070)	(0.730)
Obs	194332	194332
R^2	0.678	0.705

Table 3: Impact of Transportation on Migration

Note: Each observation represents an individual–year. The variable $\mathbf{1}_{it}$(Migration Out) is a dummy that is equal to 1 if member i of household h located in village v migrates out (whether they work outside their hometown for more than 180 days) in year t. The variable $\ln MigEarn_{it}$ denotes the logarithm of income earned by individuals working abroad, where we apply a $\log(1+x)$ transformation to account for the large number of households with zero remittances. The sample includes non-students aged between 15 and 75 who appear at least twice in our dataset over the period 2003–2013 (individual–level data were first collected in 2003). $lnTFP_{ht_0}$ refers to (time–invariant) household fixed agricultural productivity (see Appendix B.2 for the detailed estimation procedure). All regressions control for individual and province–year fixed effects. Standard errors, clustered at the village–year level, are reported in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Decomposing Agricultural Productivity Gains To better understand the sources of agricultural productivity growth, we follow Brandt et al. (2017) and decompose the changes in average village productivity into a component within the household and three other components (between-household, entry, and exit). For each outcome variable of interest y (i.e., TFP), we normalize the current household-level value y_{ht} by y_{vt-k}, the village-wide average

in the initial period $t-k$. This decomposition enables us to distinguish between productivity changes resulting from improvements within households and those driven by shifts in the composition of agricultural producers. Let s_{ht} denote the output share of household h in year t, and y_{ht} its TFP. Village-level agricultural productivity is decomposed as follows:

$$\triangle y_{vt} = \sum_{h \in C} [\bar{s}_h \triangle y_{ht} + \triangle s_{ht}(\bar{y}_h - y_{vt-k})] + \sum_{h \in E} s_{ht}(y_{ht} - y_{vt-k}) - \sum_{h \in X} s_{ht-k}(y_{ht-k} - y_{vt-k}) \quad (4)$$

where C represents continuing households that remain in farming, contributing to productivity through efficiency gains or resource reallocation. The set E includes entrants—households that were initially outside agriculture but later joined, potentially introducing new techniques, capital, or labor. The set X comprises exiting households that were initially active but later left, often due to migration or shifts in economic activity. Equation (4) holds exactly for village-level productivity growth, $\triangle y_{vt}$, measured as the change in the share-weighted average of household-level y_{ht}.[21] The first term captures productivity gains from continuing farmers, driven by within-household improvements and resource reallocation toward more productive households. The second term reflects productivity gains from the entry of more productive farmers, while the last term accounts for potential gains from the exit of lower productivity farmers.

	(1)	(2)	(3)	(4)	(5)
		Continuing Farmers			
Dep var.	Total Effect	Within	Between	Entry Farmers	Exit Farmers
$\triangle \ln Distance_{vt}$	-1.023**	-0.735*	-0.115*	-0.062	-0.111*
	(0.442)	(0.375)	(0.065)	(0.071)	(0.064)
% Contribution	100%	72%	11%	6%	11%
Obs	780	780	780	780	780
R^2	0.166	0.171	0.110	0.145	0.162

Table 4: Decomposing Agricultural Productivity Gains

Note: Dependent variables are the terms of a linear decomposition of village-level change in productivity. The sample pools annualized growth rates over three subperiods of our studied sample (1995—2001, 2002–2008, 2009–2013) that cover continuously living villages in each subperiod (it does not cover the whole village samples since there are exits at the village level), respectively. Each regression controls for province–year fixed effects. Robust standard errors are in parentheses * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

To assess how improved road connectivity enhances agricultural productivity, we regress each decomposition term in equation (4) on distance measures. The results, presented in Table 4, are based on pooled growth rates across three subperiods of our sample: 1995–2001, 2002–2008, and 2009–2013. Column (1) shows that the total effect of improved connectivity

[21] Appendix Figure B.4 illustrates the trend in average agricultural productivity and total output during our sample period. Output increases by over 35%, from 467 million to more than 630 million tons, while average TFP more than doubles, rising from 0.88 to 2.15.

is consistent with the household-level findings in Table 1. The primary driver of productivity gains is within-household TFP growth (column (2)), but between-household reallocation as well as exit also play crucial roles (columns (3) and (5)). Improved transportation enables more productive farmers to expand their market share, potentially reinforcing their efficiency advantages. The entry margin plays a trivial role, which is likely the result of institutional restrictions on agricultural activities in rural China.[22] The exit margins have a significant impact, suggesting that within-household efficiency improvements do not solely drive the observed productivity gains. These productivity gains are also shaped by selection dynamics, where better connectivity systematically alters the composition of agricultural producers, favoring those with higher productivity.

In sum, while the mechanism-based evidence and decomposition analysis offer useful insights into the sources of agricultural productivity gains, they do not fully disentangle the role of selection effects. At the village level, observed increases in aggregate productivity may primarily reflect a compositional shift: if improved road connectivity accelerates the exit of less productive farmers, the remaining producers will appear more efficient even if their individual productivity has not changed. Within households, selection effects may also be at play. As less productive farmers exit, more capable ones could expand production, benefiting from increasing returns to scale, which could further amplify measured productivity gains. Without explicitly modeling these selection dynamics, it remains unclear whether road improvements genuinely enhance efficiency or simply reshape the composition of agricultural producers. Therefore, a structural framework is essential to quantify these mechanisms, offering a more comprehensive assessment of how infrastructure development interacts with selection effects to drive productivity growth.

4 Model

We develop a dynamic, quantitative spatial equilibrium model of trade and migration across multiple regions, explicitly incorporating urban–rural divisions and the associated mobility and trade costs. Building on Caliendo et al. (2019), we extend the framework by introducing heterogeneous rural households who choose between agriculture and manufacturing occupations based on their comparative advantage. These endogenous occupational and location choices generate migration flows from villages to cities, enabling us to jointly analyze agricultural output, individual sorting, and urbanization within a unified framework.

[22] China's Hukou system, well known for placing strong restrictions not only on rural farmers entering urban areas, but also on outsiders entering villages' farming activities, since farm land is collectively owned and distributed by the village (Minale, 2018).

4.1 Model Environment

We consider a world with discrete time indexed as $t = 0, 1, \cdots, \infty$. The world has K locations and three broad sectors that produce agricultural, manufacturing, and capital goods. Each location is indexed by k or l so that $k, l, = 1, 2, \cdots, K$. A location can be either rural or urban, but not both. Rural locations produce agricultural products using land, manufactured intermediate inputs, and capital goods. The latter two inputs are supplied and produced by the manufacturing producers in urban locations, with labor as the only input. Agricultural products are used for consumption in all locations, and manufactured products and capital goods are either consumed or used as inputs in agricultural production. We denote the set of all rural locations as $\mathcal{R}$ and the set of all urban locations as $\mathcal{U}$, so that $\mathcal{R} \cup \mathcal{U} = \mathcal{K}$, where $\mathcal{K}$ is the set of all locations.

Households differ by type indexed as $z = 1, 2, \cdots, Z$, and differ in their time-invariant productivity in agricultural and manufacturing production, captured by the vector (a_z, m_z). As we show later, agricultural productivity, a_z, determines the TFP of the farm if the household remains in the rural region. Similarly, manufacturing productivity, m_z, is the number of efficient labor units that the household can supply to urban production if it moves to an urban location. The ratio between the agricultural and the manufacturing productivities is a measure of a household's comparative advantage in rural production. Households are forward-looking and possess perfect foresight, making decisions on where to move based on their innate abilities and the current state of the economy. We first describe the dynamic problem households face in deciding which sector to supply their labor to and where to move, conditional on a path of real wages over time, sectors, and locations. Next, we characterize the static problem in each time period, solving for prices and wages based on the labor supply in the markets.

4.1.1 Consumption

Households At time t, the utility of the household living in location k has a Cobb-Douglass preference for the final goods of sector s,

$$u_{kt} = \phi_{kt} \prod_{s \in \{a,m,c\}} (Q_k^s)^{\eta^s}, \quad \eta^a + \eta^m + \eta^c = 1, \tag{5}$$

where the first term $\phi_{kt} = \bar{\phi}_k L_{kt}^{\beta}$ captures the time-varying amenity of living in location k at time t, which depends on both a time-invariant component that captures the location-specific amenity, $\bar{\phi}_k$, and a time-varying component that allows for congestion externality with an elasticity of $\beta < 0$. The aggregators Q_k^a, Q_k^m, and Q_k^c capture the consumption of final

goods in the agricultural (i.e., $s = a$), manufactured (i.e., $s = m$), and capital goods (i.e., $s = c$) sectors, respectively. We let $\eta^s \in [0, 1]$ denote the corresponding expenditure share of each type of goods. Within each sector, there are differentiated varieties that are produced by firms from different locations, and, in particular, we make Q_k^s a constant elasticity of substitution (CES) aggregator of a continuum of goods, with the elasticity of substitution ϵ^s:

$$Q_k^s = \left[\sum_l \int_{\omega \in \Omega_{lk}^s} \left(q_{lk}(\omega) \right)^{\frac{\epsilon^s - 1}{\epsilon^s}} d\omega \right]^{\frac{\epsilon^s}{\epsilon^s - 1}}, \tag{6}$$

where each variety is denoted by ω, and $q_{lk}(\omega)$ is the consumption of good within sector s supplied by firms from region l.

Selective Migration and Labor Supply At the end of each period t, individual households in location k decide their location in the next period. The migration decision depends on bilateral migration costs from k to l, denoted by λ_{lkt}, the expected lifetime utility of living in l, and the vector of idiosyncratic preference shocks that are independently and identically distributed (i.i.d.) across time, location, and individuals. We adopt the canonical assumption frequently used in the literature, such as Caliendo et al. (2019, 2021) and Kleinman et al. (2023), that the idiosyncratic preference shocks, $\{\zeta_{lt}\}$, follow a Type-I extreme value distribution with the following cumulative distribution function:

$$F(\zeta) = \exp\left(- \exp(-\zeta - \bar{\gamma}) \right),$$

where $\bar{\gamma}$ is the Euler's constant. Moreover, individuals discount the future at the rate $\delta < 1$. Given these structures, the migration decision for a household of type-z living in location k at time t can be expressed by the following expression:

$$v_{zkt} = \log\left(\phi_{kt} \frac{y_{zkt}}{P_{kt}} \right) + \max_{l \in \mathcal{K}} \left\{ \delta \mathbb{E}[v_{zl,t+1}] - \lambda_{lkt} + \kappa \cdot \zeta_{lt} \right\}, \tag{7}$$

where v_{zkt} denotes the lifetime utility and y_{zkt} the income of an individual of type z residing in location k at time t. Households with higher productivity z tend to earn higher labor incomes. The expectation is taken over the future realizations of the idiosyncratic preference shocks. The variable P_{kt} represents the location-specific ideal price index, and κ denotes the (inverse of) the migration elasticity. Accordingly, we can express the expected lifetime utility as follows:

$$V_{zkt} = \log\left(\phi_{kt} \frac{y_{zkt}}{P_{kt}} \right) + \kappa \log\left(\sum_{l=1}^{K} \exp\left(\delta V_{zl,t+1} - \lambda_{lkt} \right)^{1/\kappa} \right). \tag{8}$$

Let $V_{zkt} \equiv \mathbb{E}[v_{zkt}]$ represent the expectation of the continuation value, where v_{zkt} is the lifetime utility of an individual of type z residing in location k at time t. The expected continuation value V_{zkt} captures the expected utility over all future periods, taking into account the stochastic nature of future idiosyncratic preference shocks and potential changes in location or income. Consistent with the standard approach in the literature, we can derive the migration probability of an individual of type z moving from location k to location l in period t as follows, which we denote as μ_{zlkt}:

$$\mu_{zlkt} = \frac{\exp\left(\delta V_{zl,t+1} - \lambda_{lkt}\right)^{1/\kappa}}{\sum_{l'=1}^{K} \exp\left(\delta V_{zl',t+1} - \lambda_{l'kt}\right)^{1/\kappa}}. \tag{9}$$

The migration probability in the expression above captures individual sorting by comparative advantage. To build intuition, consider the odds that a type-z individual currently living in rural location k chooses to migrate to urban location l rather than stay. This is given by:

$$\frac{\mu_{zlkt}}{\mu_{zkkt}} = \frac{\exp\left(\delta V_{zl,t+1} - \lambda_{lkt}\right)^{1/\kappa}}{\exp\left(\delta V_{zk,t+1}\right)^{1/\kappa}} = \exp\left(\delta(V_{zl,t+1} - V_{zk,t+1}) - \lambda_{lkt}\right)^{1/\kappa}.$$

The term $V_{zl,t+1} - V_{zk,t+1}$ captures the expected difference in lifetime utility between moving to l and remaining in k, and is directly shaped by each individual's productivity in the two locations. Since income depends on type-z productivity, this utility gap reflects where a worker's skills are relatively better rewarded. In other words, $V_{zl,t+1} - V_{zk,t+1}$ embodies the individual's comparative advantage between urban and rural sectors: it is higher for workers who are more productive—and therefore earn higher utility—in cities than in the countryside. At the same time, migration costs λ_{lkt} dampen mobility. When costs are high, even individuals with a strong comparative advantage in urban work may find it unprofitable to relocate. Thus, a reduction in λ_{lkt} increases the extent to which labor reallocates in response to productivity differences. This strengthens the sorting mechanism: less productive farmers, whose skills are better suited to urban production, are more likely to exit agriculture, while those who remain tend to be relatively more efficient. The result is a more productive allocation of workers across space.

Finally, the evolution of type-z population across markets is characterized by the following equation:

$$L_{zl,t+1} = \sum_{k=1}^{K} \mu_{zlkt} L_{zkt}. \tag{10}$$

The labor supply at each period is fully determined by forward-looking decisions made in the previous period. This provides a static structure of the economy, which enables us to solve for equilibrium wages over time, ensuring that labor markets are clear. The

total population in each location is the sum of the population across types, i.e., $L_{l,t+1} = \sum_{z=1}^{Z} L_{zl,t+1}$, where $L_{zl,t+1}$ is derived from equation (10). Next, we describe the production processes for agricultural, manufacturing, and capital goods in the economy.

4.1.2 Production

Agricultural Goods Sector A continuum of differentiated agricultural goods is produced in each rural region following an extended Krugman (1979) framework: the agricultural sector operates under a monopolistic competition market structure. Each rural household produces a unique variety of agricultural goods denoted by ω^a, using a production function that features increasing returns to scale, and competing with all other agricultural producers while taking the prices set by others as given. The income of a household with rural productivity a_z in the rural location k is equal to the profit from operating their farm, denoted by $\pi_{kt}(a_z)$.

Agricultural production requires land, along with a composite of manufactured and capital intermediate inputs. Notably, we do not model labor in agricultural production, as we assume that each farm is operated by a rural household, which serves as both the owner and the laborer. The payment to labor is, therefore, the profit earned on the farm. This abstraction of labor inputs is justified in the context of China, as hired labor was not common in the agricultural sector during our sample period (Tian et al., 2020).

The TFP in rural production comprises two components: a location-specific component ($\bar{A}_k$) that is common to all varieties in location k, and a household-specific component (a_z) that depends on the agricultural productivity of a type-z household. Given that each household produces a single variety, we can use either z or ω^a to index each variety. We also assume that the household must pay a fixed production cost, f, in units of the numeraire, each period to produce. We interpret these fixed costs as including maintenance costs for machinery, infrastructure, and land, as well as insurance and financial costs that are independent of the scale of production. The assumption of fixed costs enables us to capture increasing returns to scale in farming. The production function of a given agricultural good in location k, after paying the fixed costs, is represented by the following equation:

$$q^a_{zkt}(\omega^a) = \left(a_z \bar{A}_k\right) \cdot \left(\frac{\mathbb{n}}{\upsilon^n}\right)^{\upsilon^n} \left(\frac{\mathbb{Q}^m_{kt}}{\upsilon^m}\right)^{\upsilon^m} \left(\frac{\mathbb{Q}^c_{kt}}{1-\upsilon^n-\upsilon^m}\right)^{1-\upsilon^n-\upsilon^m}, \tag{11}$$

where υ^n, υ^m, and $(1-\upsilon^n-\upsilon^m)$ represent the cost shares of land, manufactured inputs, and capital inputs, respectively, in agricultural production. The variable $\mathbb{n}$ denotes the land usage of the firm. The composites of manufactured and capital intermediate inputs, $\mathbb{Q}^m_{kt}$ and $\mathbb{Q}^c_{kt}$, are CES aggregations of differentiated varieties of each type of good, supplied by manufactured and capital goods producers from the urban locations. These composites are

defined by the following expression:

$$\mathbb{Q}_{kt}^{m} = \left[\int_{\omega^m \in \Omega_{kt}^{m}} \left(q(\omega^m)\right)^{\frac{\epsilon^m - 1}{\epsilon^m}} d\omega^m \right]^{\frac{\epsilon^m}{\epsilon^m - 1}}, \mathbb{Q}_{kt}^{c} = \left[\int_{\omega^c \in \Omega_{kt}^{c}} \left(q(\omega^c)\right)^{\frac{\epsilon^c - 1}{\epsilon^c}} d\omega^c \right]^{\frac{\epsilon^c}{\epsilon^c - 1}},$$

where ϵ^m and ϵ^c represent the elasticity of substitution between differentiated input varieties within the composites of manufactured and capital intermediate input bundles, respectively.

We assume an endowment of N_k land in location k, owned by local, immobile landlords who consume all their income and share the same utility function as specified in equation (5).

Manufacturing and Capital Goods Sectors Each manufacturing (m) and capital (c) goods variety (ω^b, $b \in m, c$) can be produced in an urban location. The production of variety ω^b requires only labor as an input, and the production function in location k is specified below:

$$q_{kt}^{b}(\omega^b) = A_{kt}^{b} \cdot \ell \cdot \xi(\omega^b), \quad b \in \{m, c\}, \tag{12}$$

where $A_{kt}^{b} = \bar{A}_{k}^{b} \left(L_{kt}\right)^{\alpha}$ represents the location-specific productivity for output $b \in m, c$ that depends on both the location fundamental productivity $\bar{A}_{k}^{b}$ and population agglomeration with an elasticity of $\alpha > 0$. Following Eaton and Kortum (2002), we assume that the manufacturing industry is perfectly competitive. The term $\xi(\omega^b)$ denotes an idiosyncratic productivity shock that is i.i.d. across locations, time, and types. Furthermore, ξ is assumed to follow a unit Fréchet distribution with shape parameter θ. Consumers in each location purchase the variety ω^b from the supplier offering the lowest price, thus making the results from Eaton and Kortum (2002) applicable to our case. The probability that location l sources outputs from sector b from location k at time t is:

$$s_{lkt}^{b} = \frac{\left(w_k \tau_{lkt}\right)^{-\theta} \left(A_{kt}^{b}\right)^{\theta}}{\sum_{k' \in \mathcal{U}} \left(w_{k'} \tau_{lk't}\right)^{-\theta} \left(A_{k't}^{b}\right)^{\theta}}, \quad b \in \{m, c\}. \tag{13}$$

Note that, unlike the standard approach of Eaton and Kortum (2002), the denominator in our case encompasses only the set of urban regions, $\mathcal{U}$, because the rural locations do not produce in these sectors.

4.2 Market Clearing and Equilibrium

4.2.1 Rural Locations

Agricultural Goods and the Land Market Let X_{kt}^a represent the total expenditure on agricultural products in location k at time t. By aggregating the individual farm sales, as implied by the consumer utility functions in (5) and (6), along with the production function of agricultural goods in (11), we derive the aggregate agricultural sales from location k to location l as follows:

$$R_{lkt}^a = \frac{X_{lt}^a}{\left(P_{lt}^a\right)^{1-\epsilon^a}} \left(\frac{\epsilon^a}{\epsilon^a - 1} \frac{\tau_{lkt}^a \chi_{kt}}{\bar{A}_k}\right)^{1-\epsilon^a} \widetilde{L}_{kt}^r, \tag{14}$$

where $\tau_{lkt}^a \geq 1$ represents the iceberg trade cost from location k to location l in the agriculture sector. The variable $\widetilde{L}_{kt}^r$ summarizes the agricultural talent in location k at time t, defined as: $\widetilde{L}_{kt}^r = \left[\sum_{z=1}^{Z} \left(a_z\right)^{\epsilon^a - 1} L_{zkt}\right]$. The variable χ_{kt} captures the marginal cost of production in location k, which depends on land rent (r_{kt}), and the price indices of manufacturing (P_{kt}^m) and capital products (P_{kt}^c): $\chi_{kt} \equiv r_{kt}^{\nu^n} \left(P_{kt}^m\right)^{\nu^m} \left(P_{kt}^c\right)^{1-\nu^n-\nu^m}$. Finally, P_{lt}^a denotes the price index of agricultural products in location l, defined as: $P_{lt}^a \equiv \left[\int_0^1 \left(p_{lt}(\omega^a)\right)^{1-\epsilon^a} d\omega^a\right]^{1/(1-\epsilon^a)}$.[23] Total revenue of farms in location k is the sum of all sales to other locations:

$$R_{kt} = \sum_{l=1}^{K} R_{lkt}^a = \widetilde{\epsilon}^a \left(\frac{\chi_{kt}}{\bar{A}_k}\right)^{1-\epsilon^a} M_{kt}^a \times \widetilde{L}_{kt}^r, \forall k \in \mathcal{R}, \tag{15}$$

where $\widetilde{\epsilon}^a = \left(\frac{\epsilon^a}{\epsilon^a - 1}\right)^{1-\epsilon^a}$ is a constant, and M_{kt}^a is the market access measure for farms operating in location k, defined as $M_{kt}^a \equiv \sum_{l=1}^{K} X_{lt}^a / \left(P_{lt}^a\right)^{1-\epsilon^a} \left(\tau_{lkt}^a\right)^{1-\epsilon^a}$. Of the total revenue of farms, a proportion of $1/\epsilon^a$, which equals the profits, is the income of farmers. Another proportion of $\nu^n(\epsilon^a - 1)/\epsilon^a$ is the income of landowners, and the rest are payments to the intermediate inputs. Therefore, the total income of the rural location Y_{kt} is:

$$Y_{kt} = \left(\frac{1}{\epsilon^a} + \nu^n \frac{\epsilon^a - 1}{\epsilon^a}\right) R_{kt} = \frac{1 + \nu^n \left(\epsilon^a - 1\right)}{\epsilon^a} R_{kt}, \forall k \in \mathcal{R}. \tag{16}$$

The budget constraint of consumers implies that the total expenditure on agricultural products satisfies $X_{kt}^a = \eta^a Y_{kt}$.[24]

The rural land market clears when the total income of landowners is equal to the expen-

[23] The price indices of manufacturing (P_{kt}^m) and capital products (P_{kt}^c) are defined by the following equations: $P_{kt}^m \equiv \gamma^k \left[\int_0^1 \left(p_{kt}(\omega^m)\right)^{1-\epsilon^m} d\omega^m\right]^{1/(1-\epsilon^m)}$ and $P_{kt}^c \equiv \gamma^c \left[\int_0^1 \left(p_{kt}(\omega^c)\right)^{1-\epsilon^c} d\omega^c\right]^{1/(1-\epsilon^c)}$. γ^k and γ^c are constants related to the Gamma function, as defined in Appendix C.1.

[24] A more detailed derivation is provided in Appendix C.1.

diture of farmers on land, which implies that:

$$r_{kt}N_k = \nu^n \frac{\epsilon^a - 1}{\epsilon^a} R_{kt}. \tag{17}$$

Therefore, the total rural income can be expressed as a function of land rent:

$$Y_{kt} = \left[1 + \frac{1}{\nu^n (\epsilon^a - 1)}\right] r_{kt}N_k. \tag{18}$$

For a comprehensive derivation, refer to Appendix C.1.

Manufactured and Capital Goods Sector The rural expenditure on manufactured (X^m_{kt}) and capital (X^c_{kt}) goods originates from two primary sources: consumption demand and intermediate inputs:

$$X^m_{kt} = \eta^m Y_{kt} + \nu^m \frac{\epsilon^a - 1}{\epsilon^a} R_{kt}, \qquad X^c_{kt} = (1 - \eta^a - \eta^m) Y_{kt} + (1 - \nu^n - \nu^m) \frac{\epsilon^a - 1}{\epsilon^a} R_{kt}.$$

Therefore, the total expenditure on urban products is the sum of expenditure on manufactured, and capital goods in the rural area k is:

$$X^m_{kt} + X^c_{kt} = \Xi \times R_{kt}, \forall k \in \mathcal{R},$$

where Ξ is a constant, defined by the following equation: $\Xi \equiv [1 - \eta^a + (\epsilon^a - 1)(1 - \eta^a \nu^n)]/\epsilon^a$.

4.2.2 Urban Locations

Under perfect competition, the total income of an urban location equals the labor income:

$$Y_{kt} = w_k \widetilde{L}^u_{kt} \quad \forall k \in \mathcal{U}, \tag{19}$$

where $\widetilde{L}^u_{kt}$ denotes the total supply of efficiency labor at location k: $\widetilde{L}^u_{kt} = \sum_{z=1}^{Z} m_z L_{zkt}$. Note that we assume that a type-z household supplies m_z units of efficiency labor to the urban labor markets if the household resides in an urban location. The total expenditure on agricultural products is the same as in rural locations, given by $X^a_{kt} = \eta^a Y_{kt}$, and the expenditures on manufactured and capital products are $\eta^m Y_{kt}$ and $(1 - \eta^a - \eta^m) Y_{kt}$, respectively. These expressions are simpler than they are for rural locations due to the absence of intermediate inputs in urban production.

The total sales of all firms in location k are the sum of the expression above across all destinations and sectors. In addition, note that total sales must equal total income, so

$R_{kt} = Y_{kt}$:

$$R_{kt} = \sum_{l \in \mathcal{U}} \left(s^m_{lkt} X^m_{lt} + s^c_{lkt} X^c_{lt} \right) + \sum_{l \in \mathcal{R}} \left(s^m_{lkt} X^m_{lt} + s^c_{lkt} X^c_{lt} \right), \tag{20}$$

where s^b_{lkt}, with $b \in \{m, c\}$, represents location l's import of product b from location k as a share of its total imports, defined in (13).

4.2.3 Trade Balance

In equilibrium, the total expenditure on manufactured and capital products in a rural location, $X^m_{kt} + X^c_{kt} = \Xi \times R_{kt}$, represents the total imports of the rural location k from all urban locations. The trade balance condition requires this to be equal to the total sales of agricultural products to all urban locations, $\Xi \times R_{kt} = \sum_{l \in \mathcal{U}} R^a_{lkt}$, where R^a_{lkt} is defined in equation (14). By combining the trade balance condition with the labor market clearing condition, we derive a system of equations where χ_{kt} is a function of r_{kt}:

$$\Xi \times r_{kt} N_{kt} \left(\frac{\epsilon^a}{\epsilon^a - 1} \frac{1}{\nu^n} \right) = \sum_{l \in \mathcal{U}} \frac{\eta^a Y_{lt}}{\sum_{k' \in \mathcal{R}} \left(\tau^a_{lk't} \frac{\chi_{k't}}{\bar{A}_k} \right)^{1-\epsilon^a} \widetilde{L}^r_{k't}} \left(\tau^a_{lkt} \frac{\chi_{kt}}{\bar{A}_k} \right)^{1-\epsilon^a} \widetilde{L}^r_{kt}. \tag{21}$$

The equations above implicitly define a vector of r_{kt} that clears the rural markets.

For urban locations, trade balance in equilibrium implies that total sales of manufactured and capital goods must be equal to the total imports from all rural regions, which in turn is equal to the expenditure on agricultural products. Mathematically, this is expressed as $\eta^a Y_{kt} = \eta^a R_{kt} = \sum_{l \in \mathcal{R}} \left(s^m_{lkt} X^m_{lt} + s^c_{lkt} X^c_{lt} \right)$. Combined with equation (20), the trade balance conditions for urban locations can be simplified into the following system of equations:

$$w_k \widetilde{L}^u_{kt} = \sum_{l \in \mathcal{U}} \left[\frac{\left(w_k \tau_{lkt} \right)^{-\theta} \left(A^m_{kt} \right)^{\theta}}{\sum_{k' \in \mathcal{U}} \left(w_{k'} \tau_{lk't} \right)^{-\theta} \left(A^m_{k't} \right)^{\theta}} + \frac{\left(w_k \tau_{lkt} \right)^{-\theta} \left(A^c_{kt} \right)^{\theta}}{\sum_{k' \in \mathcal{U}} \left(w_{k'} \tau_{lk't} \right)^{-\theta} \left(A^c_{k't} \right)^{\theta}} \right] w_l \widetilde{L}^u_{lt}. \tag{22}$$

A detailed derivation of the trade balance conditions for both urban and rural locations is provided in Appendix C.2.

4.2.4 Definition of Equilibrium

The endogenous state of the economy at any given moment is defined by the distribution of the sequence of endogenous variables, denoted by $\Upsilon_t = \{w_{kt}, r_{kt}, L_{zkt}, V_{zkt}, \mu_{zlkt}\}$. The economic fundamentals are a mix of deterministic factors, some of which vary over time and some of which remain constant. The time–invariant fundamentals are represented by $\bar{\Omega} = \{\bar{A}_k, \bar{A}^m_k, \bar{A}^c_k, \bar{\phi}_k, N_k\}$, while the time–variant fundamentals are denoted by $\Omega_t = \{\lambda_{lkt}, \tau_{lkt}, \tau^a_{lkt}\}$. Next, we formally define the sequential equilibrium and the steady

state of the economy based on these model parameters.

Definition 1. *Given the set of time-invariant fundamentals $\bar{\Omega}$, the time-variant fundamentals Ω_t, and the initial population $L_{zk,0}$, a sequential competitive equilibrium of the model is defined as a sequence Υ_t that satisfies the following equilibrium conditions:*

(i) Individuals maximize their lifetime utility (8) by choosing a sequence of locations so that equations (9) and (10) hold.

(ii) Farmers maximize their profits in each period so that equations (15) and (16) hold.

(iii) Rural land market clears so that equation (17) holds.

(iv) Urban firms maximize their profits in each period and trade balances, so that equation (22) holds.

A stationary equilibrium in this economy is defined by the constancy of aggregate variables over time, requiring that the fundamentals remain unchanged for all time periods. In this equilibrium, households may move between markets, but the inflows and outflows are balanced, ensuring stability. This equilibrium is characterized by the following conditions:

Definition 2. *A steady state of the economy is an equilibrium in which both the time-variant fundamentals, Ω_t, and the endogenous variables, Υ_t, remain constant over time.*

The model reveals three essential mechanisms for understanding how transportation infrastructure investment shapes agricultural transformation. First, heterogeneity in comparative productivity is crucial for capturing the selection margin: improved connectivity prompts lower productivity households to exit agriculture, while more productive ones remain and scale up. This sorting, combined with increasing returns, allows agricultural output to rise even as rural labor declines—an effect that static or representative-agent models cannot generate. Second, the model uncovers a novel source of gains: urban migration expands demand for agricultural goods, creating a market-size effect that reinforces rural productivity, beyond the standard trade-cost channel. Third, the dynamic structure is key to capturing transitional responses. Migration and production adjust gradually, shaped by expectations about future conditions and dragged by mobility frictions—patterns that static models, focused only on steady states, necessarily miss.

5 Calibration

We now bring the model to data by calibrating it to the Chinese economy. In the model, each period corresponds to five years, and each location represents either the urban

or rural part of a prefecture in China, or a region in the Rest of the World (ROW). With 279 prefectures in China, the total number of locations is $K = (279 + 1) \times 2 = 560$. This sample of 279 prefectures represents the largest common set available across the *China City Statistical Yearbook* and the NFPS. It is also highly representative, covering approximately 98 percent of China's population and 99 percent of the national GDP (Ma and Tang, 2020).

As is standard in the context of prefecture-level spatial models in China, we solve the baseline model in levels and conduct counterfactual simulations without using the dynamic hat algebra, as in Caliendo et al. (2019). The choice of the solution algorithm is guided by data availability, as prefecture-to-prefecture trade flow data do not exist in China. Our quantification and solution algorithm is similar to that in Ma and Tang (2024), and we refer readers to that paper for a detailed discussion on the (lack of) applicability of dynamic hat algebra in China.

We quantify the model in three steps. In the first step, we calibrate several parameters without solving the models. These parameters come from the literature or rely on exogenous moment conditions that can be directly mapped to the data. In the second step, we invert the model in the initial static equilibrium to recover the fundamental productivity by location and sector. In the last step, we invert the model along the transition path to steady state to infer the fundamental amenity of each location. We provide a brief overview of the quantification process in this section and refer readers to Appendix C.6 for the implementation details and data sources.

Basic Parameters We follow the literature in assigning values to several standard model parameters. The discount factor is set to $\delta = 0.97^5 \approx 0.8587$, consistent with an annual interest rate of 3 percent. Following Kleinman et al. (2023), we set the inverse of migration elasticity to $\kappa = 3\delta$. The trade elasticity is set to $\theta = 8.26$, as estimated by Eaton and Kortum (2002).[25] The agglomeration elasticity ($\alpha = 0.1$) and congestion elasticity ($\beta = -0.3$) are taken from Allen and Arkolakis (2022). We obtain the expenditure shares on agricultural and manufactured goods, $\eta^a = 0.23$ and $\eta^m = 0.52$, from the 2007 Chinese input–output table, calculated as the share of each category in total final consumption. The agricultural production parameters ($\nu^n = 0.27$ and $\nu^m = 0.68$) are estimated using a constrained linear regression, imposing that input elasticities sum to 1. We estimate fixed production costs using NFPS data, based on the city-level average depreciation expense of fixed assets in

[25] As shown in the Appendix, the functional form assumptions of the Fréchet distribution imply that $\theta > \epsilon^m - 1$ and $\theta > \epsilon^c - 1$. Given the common estimates of the elasticity of substitution, we have to use a trade elasticity that is on the higher end of the usual values. In the robustness checks, we demonstrate that our main results are robust to lower values of trade elasticity, which are closer to the theoretical lower bound.

2002. This yields an average of 119.9 RMB per household across our sample. Taking the average urban income in Beijing as the numeraire, this translates to a fixed cost of $f = 0.022$ in the model. The estimate is based on the full NFPS household sample.Table 5 summarizes the key parameters used in the quantitative analysis.

Parameter	Meaning	Source or Reason	Value
η^a	Expenditure share on agricultural products	Chinese I-O table in 2007	0.23
η^m	Expenditure share on manufactured products	Chinese I-O table in 2007	0.52
ν^n	Cost share of land	NFPS	0.27
ν^m	Cost share of manufactured inputs	NFPS	0.68
f	Fixed costs of agri. production	NFPS	0.022
δ	Discount factor	Yearly interest rate of 3 percent	0.8587
κ	Inverse of migration elasticity	Kleinman et al. (2023)	3δ
θ	Trade elasticity	Eaton and Kortum (2002)	8.26
α	Agglomeration elasticity	Allen and Arkolakis (2022)	0.1
β	Congestion elasticity	Allen and Arkolakis (2022)	-0.3

Table 5: Basic Parameter Values and Sources

Population Distribution To compute the baseline equilibrium, we begin by specifying the initial distribution of China's population across locations and productivity types. We assume that household productivity in agriculture and manufacturing is jointly distributed according to a log-normal distribution, parameterized by $(\mu_a, \mu_m, \sigma_a, \sigma_m, \rho)$, where μ_a and μ_m denote the means, σ_a and σ_m the standard deviations, and ρ the correlation between agricultural and manufacturing productivity. We estimate these parameters using a simulated method of moments (SMM) algorithm, matching model-generated moments to their empirical counterparts (see Appendix C.6 for details). Based on the estimated joint distribution, we draw a random sample of 10,000 households for each location. We then discretize the joint distribution of productivity types into a $5 \times 5 = 25$ grid for each location, enabling us to capture heterogeneity in comparative advantage across households.

Geography, Migration, and Trade Frictions The model's geography is defined by each location's fundamental amenities and productivities, as well as the migration and trade frictions across locations. We recover location fundamentals by inverting the model, as discussed later. We take the migration friction matrix λ_t from Ma and Tang (2024), who estimate bilateral frictions using China's transportation network and hukou barriers across prefectures between 1994 and 2017. We use their estimates for the years 2000, 2005, 2010, and 2015 to correspond to the first four periods in our model.[26] The trade friction matrices

[26]We do not allow rural-to-rural migration in our quantification by setting the migration costs between any two rural locations at a prohibitively high level. In the data, rural-to-rural migration is practically

for manufactured goods, τ_t, are drawn from the same source. For agricultural goods, we allow for distinct trade costs and follow Hsu and Ma (2021) in setting $\tau^A_{lkt} = 3 \times \tau_{lkt}$ for all $l \neq k$.

Static Model Inversion Conditional on the parameter specified above, the second group of parameters to quantify is the sector–location fundamental productivity $\{\bar{A}_k, \bar{A}^m_k, \bar{A}^c_k\}_{k=1}^K$. We follow the standard procedure in the literature (see e.g., Kleinman et al., 2023) to invert the model and infer these parameters. In particular, we solve the model in levels at $t = 1$, conditional on the initial population distribution, and find the vectors $\{\bar{A}_k, \bar{A}^m_k, \bar{A}^c_k\}_{k=1}^K$ to match the observed agricultural output share of each rural location, and the observed manufacturing and capital goods share of each urban location in the initial year, 2000. Despite the dynamic structure of the model, the static model inversion only requires us to solve the initial static equilibrium, but not the full dynamic transition path to a steady state.

Dynamic Model Inversion The final set of parameters to recover are the fundamental amenities $\bar{\phi}_k$ for each location. We estimate these using a dynamic model inversion procedure. Specifically, we begin by assuming that migration and trade frictions remain constant after 2015 and solve for the steady state implied by the geography observed in that year. We then compute the transition path from the initial equilibrium in 2000 to this steady state and infer the amenity vector $\bar{\phi}_k$ by matching the model's implied population distribution in 2010 (period $t = 2$) to observed census data. Importantly, this inversion relies only on the total population in each location from the 2010 Population Census, and not on the joint distribution of individuals by location and ability.

Rest of the World We introduce "Rest of the World" (ROW) as two additional locations in the model: one rural and one urban. Locations in China can trade with these locations, but workers cannot move between ROW and China. We map ROW to an aggregate economy that consists of all the trading partners of China that we have data on, using a strategy similar to that of Ma and Tang (2024). The rural and urban population and arable land endowment data come from the World Development Indicators (WDI), and the transportation costs between ROW and each of the Chinese prefectures come from Ma and Tang (2024). The rural and urban productivity of ROW is jointly calibrated with all the other productivity parameters, as discussed above. The target of this calibration exercise includes the rural and urban outputs measured in real dollar terms, which we also obtain from WDI. We assume that the populations of the two ROW locations are exogenously fixed, and, therefore, there

non-existent due to land entitlement restrictions.

is no need to calibrate the amenity parameters for ROW. The data appendix B.1 provides more details on quantifying the ROW.

6 Quantitative Results

This section presents the core quantitative findings of the paper, addressing the central question that motivates our analysis: How do transportation networks shape rural economic activity? We begin by examining how farmers adjust their production and migration decisions in response to changes in trade and migration costs. We then turn to the aggregate and distributional implications of these adjustments. Finally, we explore the welfare consequences of alternative land allocation policies, focusing on the optimal reallocation of agricultural land across regions in China.

We rely on counterfactual simulations to answer these questions. Our "baseline" equilibrium is quantified to capture the transition path from the initial conditions in 2000 to a steady state implicitly defined by the transportation networks observed in 2015. To construct the baseline economy, we first compute the steady state using the transportation networks in 2015 and then use a shooting algorithm to compute the transition path from the initial population distribution in 2000 toward the steady state. As in Caliendo et al. (2019), we do not need to assume that the model is in a steady state in the initial year. Along the baseline transition path, we incorporate the observed transportation networks for the years 2000, 2005, 2010, and 2015, and hold the network fixed beyond 2015. By construction, both trade and migration costs evolve over time along this path, allowing the model to capture the dynamic impact of transportation improvements over the 2000–2015 period. In the "no change" counterfactual, we hold trade and migration cost matrices constant at their 2000 values throughout the entire transition. That is, we simulate the evolution of the economy from the same initial population distribution in 2000, but toward a steady state defined by the transportation conditions prevailing in that year. Comparing outcomes between the "no change" and baseline scenarios isolates the general equilibrium effects of the observed transportation network expansion between 2000 and 2015. As explained in the previous section, we solve both the baseline and the counterfactual simulations in levels, not hat-changes, due to data limitations at the prefecture level.

To disentangle the channels through which transportation networks affect the economy, we further conduct two additional counterfactual simulations. In the "changing trade frictions" scenario, we allow only the trade cost matrices τ to evolve over time, while holding migration costs λ fixed at their initial 2000 levels. Conversely, in the "changing migration frictions" scenario, we allow only λ to change, keeping τ constant. Comparing each of these

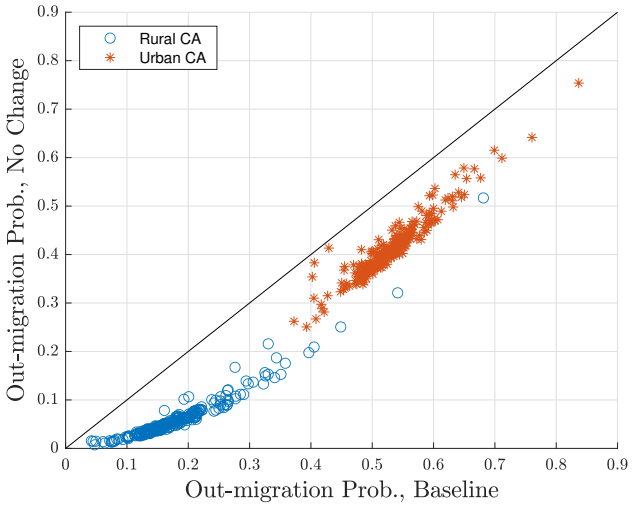


Figure 1: Transportation Improvements and Urbanization Probability

Note: This figure illustrates how transportation networks affect urbanization probabilities across individual types. Each dot represents a prefecture. The horizontal axis shows the probability of rural-to-urban migration in 2015 under the baseline simulation, while the vertical axis reports the same probability in the "no change" counterfactual, where the transportation network remains fixed at its 2000 level. Red stars denote individuals with the strongest comparative advantage in urban production ($z = 21$), and blue circles denote those with the strongest comparative advantage in rural production ($z = 5$). The 45-degree line serves as a reference.

simulations to the "no change" scenario enables us to isolate the separate contributions of declining trade and migration frictions to the overall impact of transportation improvements.

6.1 Individual Migration and Sorting

Improved transportation networks enable individuals to sort more efficiently across space according to their comparative advantage. Figure 1 illustrates this mechanism by comparing the probability that rural individuals migrate to urban areas in the baseline model versus the "no change" counterfactual.[27] Each dot in the scatter plot represents a prefecture. The red stars mark the individual type with the strongest comparative advantage in urban production, while the blue circles indicate those with the strongest comparative advantage in agriculture.[28]

Figure 1 reveals several intuitive and reassuring patterns in how transportation networks shape urbanization dynamics through individual sorting. First, the vertical distance between the baseline and counterfactual outcomes reflects the effect of improved connectivity on rural-

[27] We define the urbanization probability of a rural individual living in location i at time t as $1 - \mu_{zit}$. Since the model abstracts from rural-to-rural migration, all migration out of rural areas is directed toward urban locations.

[28] Specifically, red stars correspond to individual type $z = 21$, with the highest m_z/a_z ratio; blue circles represent type $z = 5$, with the lowest m_z/a_z ratio.

to-urban migration. The pattern is striking: individuals with a strong comparative advantage in urban production are much more likely to migrate when access improves, while those better suited to agriculture tend to remain. This sorting mechanism underscores a central feature of our model—transportation infrastructure influences not just how many people move, but who moves—thereby improving the match between individuals and locations and amplifying aggregate efficiency. Second, and as expected, urbanization increases across the board with better connectivity. This is evident from the fact that all observations lie below the 45-degree line, indicating that improved transportation raises the probability of rural-to-urban migration for all individual types, regardless of comparative advantage.

Building on the patterns shown in Figure 1, we next examine how improved connectivity affects sorting outcomes more systematically. Table 6 explores this mechanism by regressing changes in urbanization probability on local characteristics, focusing on individuals with a comparative advantage in agriculture. The results reveal a clear and intuitive pattern: highly productive rural individuals are much less likely to migrate when they are located in areas well-suited for agricultural production. Column (1) shows that an initially low urbanization rate—often a proxy for strong agricultural potential—is associated with a smaller migration response to transportation improvements. Columns (2) through (5) further demonstrate that rural locations with greater land endowment, better amenities, and stronger access to both input and output markets exhibit lower migration responses. In other words, better infrastructure facilitates sorting not only across people, but also across places: productive individuals increasingly remain in high-potential rural areas. This spatially efficient reallocation sets the stage for the aggregate gains we analyze in the next section.

The improved sorting of individuals is also reflected in higher average rural labor productivity in locations conducive to agricultural production. In particular, we find that a better transportation network leads to higher rural labor productivity in locations well-endowed with land and better market access to intermediate inputs. Panel (b) of Table 6 reports a linear probability model in which we regress an indicator variable that is equal to 1 if a rural location's average labor productivity increases with a better transportation network on the same vector of location-specific characteristics, as in panel (a) of the same table. We find that on average, doubling the land endowment of a rural location increases the probability of higher average rural productivity by 2.6 percentage points, and halving the input costs increases the probability by 34.7 percentage points.[29]

[29] As reported in column (5) of the Table 6, the coefficient on the logarithm of land endowment is 0.085, and therefore, $\log(2) * 0.085 \approx 0.026$. Similarly, the coefficient on the logarithm of initial input costs is -1.154, and therefore, $\log(0.5) * (-1.154) \approx 0.347$.

Dep var.	Changes in Out-Migration Probability					
	(1)	(2)	(3)	(4)	(5)	(6)
Panel (a): Individual Responses						
Initial Out-Mig Rate	0.437***					0.342***
	(0.027)					(0.032)
log Land Endowment		-0.015***				-0.006***
		(0.002)				(0.002)
Amenity			-0.104***			-0.064***
			(0.022)			(0.016)
log Market Access, Initial				-0.034***		-0.017
				(0.011)		(0.012)
log Input Costs, Initial					0.095***	0.041
					(0.029)	(0.030)
Obs	279	279	279	279	279	279
Adj.R^2	0.478	0.229	0.073	0.028	0.034	0.532
Dep var.	$\Pr(\Delta \bar{z} > 0)$					
	(1)	(2)	(3)	(4)	(5)	
Panel (b): Village-Level Responses						
log Land Endowment	0.075**				0.085***	
	(0.030)				(0.030)	
Amenity		-0.071			-0.038	
		(0.367)			(0.362)	
log Market Access, Initial			0.562***		0.286	
			(0.186)		(0.269)	
log Input Costs, Initial				-1.617***	-1.154*	
				(0.468)	(0.680)	
Obs	279	279	279	279	279	
Adj.R^2	0.018	-0.003	0.028	0.038	0.057	

Table 6: Effects of Transportation Improvements on Sorting of Rural-Advantaged Individuals

Note: This table reports results from regressions of the impact of transportation network improvements on the urbanization probability of rural areas, using various location-specific characteristics as explanatory variables. Each observation corresponds to a prefecture. The dependent variable is the change in the out-migration rate for type-z individuals in prefecture i in 2015, defined as $\left(1-\mu_{zii,2015}^{\text{baseline}}\right)-\left(1-\mu_{zii,2015}^{\text{no change}}\right)$. In the set of regressors, "initial" refers to values taken from the "no change" counterfactual. "Market access" is measured by M_{it}^{a}, while "input costs" correspond to χ_{it}. "Rural CA" denotes individuals with the strongest comparative advantage in agricultural production ($z=5$).

6.2 Urbanization Dynamics and Agricultural Productivity

Individual responses to improved transportation networks scale up to economy-wide shifts in both urbanization and agricultural productivity. Better connectivity accelerates the pace

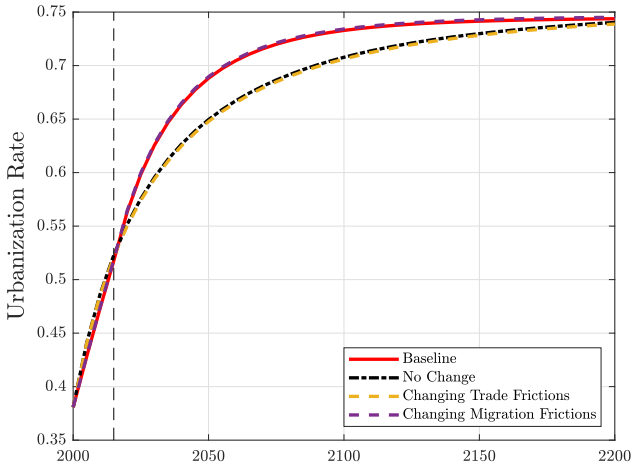


Figure 2: Transportation Networks and the Speed of Urbanization

Note: This figure plots the transition paths of the urbanization rate, measured as the share of the urban population in total population, under four simulation scenarios. The "baseline" simulation allows both trade and migration frictions to evolve over time, consistent with the data. The "no change" simulation holds both frictions fixed at their 2000 levels. The "changing trade (migration) frictions" simulations allow only one friction to evolve while keeping the other fixed at its 2000 level.

at which workers move from rural to urban areas. Figure 2 plots the transition path of the national urbanization rate under different scenarios. In the model, the productivity advantage of urban areas draws workers toward cities, while congestion and general equilibrium adjustments in factor prices moderate this pull, preventing excessive urban concentration. These opposing forces shape a gradual urbanization process: in the baseline simulation, the urbanization rate rises from 38 percent in 2000 to 52 percent by 2015, eventually converging to a steady state of 75 percent after roughly 115 years. By contrast, without transportation improvements, urbanization proceeds far more slowly, taking around 270 years to reach the same long–term level in the "no change" counterfactual.[30] A decomposition reveals that this acceleration is driven almost entirely by declining migration frictions, while reduced trade costs have little effect on the speed of transition.

Decomposing the Sources of Agricultural Productivity Gains Despite a declining rural workforce, aggregate agricultural output continues to rise steadily along the transition path. Table 7 quantifies these gains by comparing total rural output in the baseline and the "no change" counterfactual economies for the years 2015, 2030, and 2045.[31] As shown in

[30]The steady–state level of urbanization rate depends on agricultural productivity and the expenditure share on agricultural products and only indirectly relies on transportation networks. Therefore, the steady–state levels of urbanization rate are not meaningfully different between the baseline and the "no change" counterfactual.

[31]See Figure A.1 in the appendix for the full transition path of total rural output, and Figure A.2 for rural output per capita.

panel (a), rural output is between 2.2 and 5.3 percent higher under improved connectivity, even as urbanization accelerates and the rural population shrinks (see Figure 2). These gains stem from multiple channels. Better infrastructure enhances access to input and output markets, improves household-level productivity, and triggers selection effects as less productive farmers exit the sector. To clarify the contribution of each channel, we decompose the percentage change in aggregate rural output using the model's structure—specifically equations (15) and (16). Let x' denote the level of a variable under the "no change" counterfactual, and define $\hat{x} = x/x'$ as the change due to improved connectivity. This decomposition highlights how infrastructure-induced shifts in factor allocation, prices, and productivity jointly shape the rural economy's trajectory.

$$\widehat{Y}_{kt} = \widehat{R}_{kt} = \underbrace{\left(\widehat{r}_{kt}\right)^{\nu^n(1-\epsilon^a)}}_{\text{Land Rent}} \times \underbrace{\left[\left(\widehat{P^m_{kt}}\right)^{\nu^m}\left(\widehat{P^c_{kt}}\right)^{\nu^c}\right]^{1-\epsilon^a}}_{\text{Input Costs}} \times \underbrace{\widehat{M^a_{kt}}}_{\text{Market Access}} \times \underbrace{\widehat{\tilde{L}^r_{kt}}}_{\text{Access to Efficiency Labor}} \tag{23}$$

Equation (23) breaks down the impact of improved transportation networks on rural output into four intuitive components: (i) land rents, which capture general equilibrium pressure on land prices; (ii) intermediate input costs, reflecting cheaper access to capital and materials; (iii) market access (M), which measures the ability to sell goods to a broader set of urban consumers; and (iv) efficiency-adjusted labor supply ($\tilde{L}$), which accounts for both the number and productivity of workers remaining in agriculture. Columns (2) through (5) of Table 7 report the percentage changes from each component.[32] Since $\epsilon^a > 1$, the exponents on land rent and input costs are negative, indicating that increases in production costs reduce output through general equilibrium effects. The 2015 results highlight two dominant sources of rural output gains: improved access to consumption markets (2.2 percent) and stronger selection among agricultural workers (4.0 percent). By contrast, the effect of lower intermediate input costs is modest at just 0.4 percent, while higher land rents dampen gains by 1.3 percent. Looking further ahead, the influence of urbanization becomes more visible. Although rising urbanization mechanically reduces rural labor supply, lowering output by 2.3 percent, it also boosts demand for food, intensifying the market access channel, which contributes a 4.7 percent gain. The effects of factor prices remain stable, and the net impact of transportation networks on rural output remains positive, though slightly lower at 2.2 percent.

Reducing trade costs alone can generate similar aggregate gains in the long run, though the short-run effects are considerably smaller. Panel (b) of Table 7 reports results from a

[32] Appendix C.5 provides the detailed derivation of the agricultural output decomposition. These four columns form a multiplicative decomposition of the total effect in column (1). For instance, in panel (A), $1.053 = 0.987 \times 1.004 \times 1.022 \times 1.040$.

Var.	(1) $\widehat{Y}$	(2) $\widehat{\text{Land Rent}}$	(3) $\widehat{\text{Input Costs}}$	(4) $\widehat{\text{Market Access}}$	(5) $\widehat{\text{Efficiency Labor}}$
Panel (a): Benchmark					
2015	1.053	0.987	1.004	1.022	1.040
2030	1.033	0.992	1.005	1.040	0.997
2045	1.022	0.994	1.004	1.047	0.977
Panel (b): Reducing τ					
2015	1.022	0.994	1.009	1.015	1.003
2030	1.024	0.994	1.009	1.015	1.005
2045	1.025	0.994	1.010	1.014	1.007
Panel (c): Reducing λ					
2015	1.031	0.992	0.995	1.006	1.038
2030	1.010	0.998	0.996	1.025	0.992
2045	0.998	1.000	0.995	1.033	0.971

Table 7: Changes in Total Rural Output in the Baseline Model

Note: This table reports the average impact of improved transportation networks across rural prefectures under three simulation scenarios, expressed as hat changes relative to the "no change" baseline. All values represent geometric averages across rural prefectures for multiplicative decomposability. The components are defined as follows: $\widehat{\text{Land Rent}} = (\widehat{r}_{kt})^{\nu^n(1-\epsilon^a)}$, $\widehat{\text{Input Costs}} = \left(\left[\left(\widehat{P^m_{kt}}\right)^{\nu^m}\left(\widehat{P^c_{kt}}\right)^{\nu^c}\right]\right)^{1-\epsilon^a}$, $\widehat{\text{Market Access}} = \widehat{M^a_{kt}}$, and $\widehat{\text{Efficiency Labor}} = \widehat{\tilde{L}^r_{kt}}$.

counterfactual in which trade costs decline as observed in the data, while migration costs remain fixed at their 2000 levels. By 2045, total rural output is expected to increase by 2.5 percent, comparable to the gains in the full baseline scenario. However, the mechanisms differ: with migration constrained, more workers remain in rural areas, contributing to a modest increase in efficiency-adjusted labor supply (0.7 percent). Yet, in the absence of broader labor reallocation, the productivity-enhancing effects of selection are muted, and the gains instead rely more heavily on improved access to input and output markets. However, slower urbanization dampens the expansion of urban demand, leading to a weaker market access effect of just 1.4 percent. In the short run, the differences are more pronounced. By 2015, reducing trade costs alone increases rural output by only 2.2 percent, less than half the 5.3 percent gain under the full baseline scenario. The shortfall is driven by limited labor reallocation: without falling migration costs, the efficiency gains from sorting are minimal, contributing just 0.3 percent compared to 4 percent in the baseline model.

By contrast, reducing migration costs alone, without concurrent improvements in the trade network, *lowers* total agricultural output in the long run. Panel (c) of Table 7 presents

results from a simulation in which only migration costs fall while trade costs remain fixed at their initial levels. By 2045, the efficiency-adjusted rural labor supply declines by 2.9 percent—similar to the baseline scenario. However, unlike the baseline, rural producers are unable to capitalize on rising urban demand due to stagnant market access. As a result, the contribution of market access to rural output is significantly weaker (3.3 percent versus 4.7 percent in the baseline). Ultimately, migration liberalization alone leads to a decline in total rural output, driven by the premature exit of productive labor without a corresponding demand-side expansion. These findings underscore the importance of jointly reducing trade and migration frictions to achieve both higher urbanization and sustained agricultural growth.

A central feature of our model—the heterogeneity in household productivity and the resulting spatial sorting—proves critical in shaping the impact of improved connectivity. Table A.1 in the appendix presents a version of the model with homogeneous workers, where all households share identical agricultural productivity. In this setting, even a coordinated reduction in both trade and migration costs leads to a *decline* in long–term rural output. Without selection and sorting, rising urbanization, prompted by better transportation, causes a sharp drop in the effective rural labor force, with no offsetting gains from reallocation or scale. These results underscore that heterogeneity is not merely a modeling choice, but a necessary condition for rural output to rise alongside urbanization. Improved sorting of rural households is essential for realizing the dual gains of structural transformation: expanding urban economies while maintaining agricultural productivity and output.

6.3 Land Policy and Productivity Gains Through Selection

To further illustrate the central role played by heterogeneity and selection in shaping rural productivity, we turn to an alternative policy lever: land allocation. While improved connectivity enables selection through migration, institutional reforms can foster similar gains by reallocating key production inputs. In this counterfactual, we simulate a partial relaxation of China's current farmland policy by reallocating land quotas across space, shifting arable land toward locations with stronger comparative advantage in agriculture. The aim is to assess whether better matching land with more productive producers can amplify the gains from structural transformation.

Concerns over preserving farmland have long shaped land-use policy, both in China and globally (see e.g., Brueckner and Fansler, 1983; Jiang et al., 2013; Bren d'Amour et al., 2017). In China, these concerns are formalized in the Farmland Red Line Policy, which mandates a minimum threshold of preserved arable land at both the national and local levels. While

the policy is motivated by food security, it imposes rigid constraints on rural-to-urban land conversion, limiting flexibility in land reallocation across space. By relaxing these constraints and allowing land quota to flow toward more productive rural areas, we evaluate the potential for land policy to reinforce selection and further enhance agricultural output.[33]

Dep var.	Difference in Optimal and Actual Land Allocation				
	(1)	(2)	(3)	(4)	(5)
log Prod.	0.015***				0.014***
	(0.002)				(0.002)
log Input Costs, Initial		-0.181***			-0.087***
		(0.022)			(0.033)
log Efficiency Labor, Initial			0.002		0.004**
			(0.001)		(0.001)
log Market Access, Initial				0.067***	0.026**
				(0.008)	(0.011)
Obs	279	279	279	279	279
Adj.R^2	0.229	0.183	0.001	0.162	0.364

Table 8: Gaps Between Observed and Optimal Land Allocation Across Prefectures

Note: The left-hand side of the regressions is defined as $100 \times (N_i^* - N_i)$, where N_i^* denotes the optimal land allocation that maximizes aggregate rural output in 2015, and N_i represents the observed allocation in the baseline simulation. The scaling factor of 100 is applied for ease of interpretation.

To assess how land policy can enhance productivity through selection, we use the model to simulate an optimal reallocation of farmland across prefectures. The total supply of arable land is held fixed at its baseline level, but its distribution across space is allowed to adjust. Specifically, we seek a vector of land endowments $\{N_k'\}$ that maximizes total agricultural output in 2015, $\sum_{k=1}^{K} Y_k$, subject to the constraint $\sum_{k=1}^{K} N_k' = \sum_{k=1}^{K} N_k$. The optimal allocation satisfies the first-order conditions of the problem posed by a social planner aiming to maximize agricultural output. We solve the problem numerically within the full general equilibrium structure of the model. In essence, land is reallocated toward locations where its marginal product is highest. Full implementation details are provided in Appendix C.4. Unlike previous work that abstracts from how land policy shapes broader economic forces, our framework accounts for the equilibrium effects on wages, land rent, migration, urbanization, and internal trade. This exercise enables us to evaluate land reform not in isolation, but as an integrated part of the development process.

Reallocating land quotas across prefectures yields substantial efficiency gains: our model shows that optimal redistribution could raise total rural output by 8.2 percent. Table 8

[33] Additional background on the Farmland Red Line Policy is provided in Appendix C.7.

explores the sources of these gains by regressing the gap between the optimal and observed land quotas, $N_i^* - N_i$, on location-specific characteristics. The results reveal a consistent pattern: the optimal allocation channels more quotas toward areas with higher agricultural productivity, better market access, lower input costs, and more productive workers, amplifying the effects of selection and scale. Importantly, these gains come without sacrificing urbanization. In fact, the urbanization rate under the optimal land scheme slightly exceeds that of the baseline, 47.5 versus 47.4 percent in 2015, suggesting that more efficient land allocation can reinforce rather than hinder structural transformation. Much like the impact of improved connectivity, this policy change enhances rural productivity while accelerating the reallocation of labor toward higher return opportunities, underscoring the central role of heterogeneity and spatial sorting in shaping development outcomes.

6.4 Robustness Checks

To assess the robustness of our quantitative results on the impact of transportation networks, we conduct a series of alternative simulations. These exercises evaluate how changes in model assumptions affect key outcomes such as urbanization rates and rural output, relative to the "no change" counterfactual. Table A.2 in the appendix reports the results from these exercises, alongside the baseline comparisons shown in Table 7. Specifically, we consider three modifications: (i) a setting in which agricultural production incurs no fixed cost, and (ii) two alternative values of the trade elasticity, θ, which governs the substitutability of agricultural goods across regions. For each specification, we recalibrate the model and compute both the baseline and counterfactual paths under the revised parameters. Across all scenarios, we find that the core results remain quantitatively similar to the baseline. Interestingly, we find that a higher trade elasticity leads to larger gains in agricultural output—a result that may seem at odds with standard trade theory, according to which higher elasticity typically dampens the gains from trade. The intuition in our setting, however, is different. A higher θ flattens the distribution of trade costs across sources, making trade flows more sensitive to relative productivity differences across locations. This heightened responsiveness allows more agricultural production to shift toward high-productivity regions. With land and labor still constrained by mobility frictions—such as migration barriers and land quota rules—this increased flexibility in reallocating trade flows becomes especially valuable. In short, higher trade elasticity strengthens the selection mechanism, enabling more efficient spatial sorting and, in turn, boosting rural output through better use of existing resources.

7 Conclusion

This paper examines how the expansion of China's transportation network has influenced agricultural productivity and structural transformation. Using rich household panel data linked to high–resolution infrastructure maps, we provide new empirical evidence that improved connectivity boosts both agricultural productivity and rural-to-urban migration. Importantly, these outcomes are not in conflict. Rather, they are connected through a process of selective migration: less productive farmers are more likely to exit agriculture, while more productive households remain and expand their operations. This sorting mechanism enhances efficiency within the rural sector, sustaining output growth even as the agricultural workforce contracts—an outcome that contrasts with models assuming homogeneous agents.

Motivated by our empirical findings, we develop a dynamic spatial equilibrium model with heterogeneous households, sector-specific productivity, and endogenous migration. The model matches key features of the data, including the acceleration of urbanization and sustained gains in rural output. This acceleration reflects not only higher mobility but also more efficient sorting across space and sectors. The model also reveals three distinct channels through which infrastructure improves agricultural performance: improved access to output and input markets, and—most notably—a new demand-side force. As workers migrate and urban centers expand, demand for agricultural goods increases, feeding back into rural income growth. This channel, often overlooked in previous work, emerges as a significant contributor to aggregate gains.

Finally, we use the model to explore how institutional reforms interact with heterogeneity and spatial sorting. A counterfactual reallocation of land quotas—loosening China's Farmland Red Line Policy—amplifies productivity gains by directing resources toward more productive areas without slowing urbanization. These results underscore a broader lesson: the benefits of infrastructure and policy reform hinge not only on reduced frictions but also on the underlying heterogeneity that determines how individuals and places respond. Recognizing this variation is essential for designing policies that enhance both equity and efficiency in the development process.

References

Adamopoulos, Tasso, "Spatial integration and agricultural productivity: Quantifying the impact of new roads," *American Economic Journal: Macroeconomics*, 2025, *17* (1), 343–378.

_ **and Diego Restuccia**, "Land Reform and Productivity: A Quantitative Analysis with Micro Data," *American Economic Journal: Macroeconomics*, 2020, *12* (3), 1–39.

_ **, Loren Brandt, Chaoran Chen, Diego Restuccia, and Xiaoyun Wei**, "Land Security and Mobility Frictions," *The Quarterly Journal of Economics*, 03 2024, *139* (3), 1941–1987.

_ , _ **, Jessica Leight, and Diego Restuccia**, "Misallocation, Selection, and Productivity: A Quantitative Analysis With Panel Data From China," *Econometrica*, 2022, *90* (3), 1261–1282.

Allen, Treb and Costas Arkolakis, "The Welfare Effects of Transportation Infrastructure Improvements," *The Review of Economic Studies*, 02 2022.

Asher, Sam and Paul Novosad, "Rural Roads and Local Economic Development," *American Economic Review*, 2020, *110* (3), 797–823.

Asturias, Jose, Manuel García-Santana, and Roberto Ramos, "Competition and the Welfare Gains from Transportation Infrastructure: Evidence from the Golden Quadrilateral of India," *Journal of the European Economic Association*, December 2019, *17* (6), 1881–1940.

Banerjee, Abhijit, Esther Duflo, and Nancy Qian, "On the road: Access to transportation infrastructure and economic growth in China," *Journal of Development Economics*, 2020, *145*, 102442.

Baum-Snow, Nathaniel, "Did Highways Cause Suburbanization?," *The Quarterly Journal of Economics*, 05 2007, *122* (2), 775–805.

_ **, Loren Brandt, J. Vernon Henderson, Matthew A. Turner, and Qinghua Zhang**, "Roads, Railroads, and Decentralization of Chinese Cities," *The Review of Economics and Statistics*, July 2017, *99* (3), 435–448.

Bonadio, Barthélémy, "Ports vs. Roads: Infrastructure, Market Access and Regional Outcomes," *SSRN Electronic Journal*, 2024.

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

Bren d'Amour, Christopher, Femke Reitsma, Giovanni Baiocchi, Stephan Barthel, Burak Güneralp, Karl-Heinz Erb, Helmut Haberl, Felix Creutzig, and Karen C. Seto, "Future urban land expansion and implications for global croplands," *Proceedings of the National Academy of Sciences*, 2017, *114* (34), 8939–8944.

Brueckner, Jan K. and David A. Fansler, "The economics of urban sprawl: Theory and evidence on the spatial sizes of cities," *The Review of Economics and Statistics*, 1983, pp. 479–482.

Bryan, Gharad and Melanie Morten, "The Aggregate Productivity Effects of Internal Migration: Evidence from Indonesia," *Journal of Political Economy*, 2019, *127* (5), 2229–2268.

Cai, Hongbin, J. Vernon Henderson, and Qinghua Zhang, "China's land market auctions: evidence of corruption?," *The Rand Journal of Economics*, 2013, *44* (3), 488–521.

Caliendo, Lorenzo, Luca David Opromolla, Fernando Parro, and Alessandro Sforza, "Goods and factor market integration: A quantitative assessment of the EU enlargement," *Journal of Political Economy*, 2021, *129* (12), 3491–3545.

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

Caunedo, Julieta and Elisa Keller, "Capital Obsolescence and Agricultural Productivity," *The Quarterly Journal of Economics*, 2021, *136* (1), 505–61.

Chari, Amalavoyal, Elaine M. Liu, Shing-Yi Wang, and Yongxiang Wang, "Property rights, land misallocation, and agricultural efficiency in China," *The Review of Economic Studies*, 2021, *88* (4), 1831–1862.

Chen, Ting, Laura Xiaolei Liu, Wei Xiong, and Li-An Zhou, "The Speculation Channel and Crowding out Channel: Real Estate Shocks and Corporate Investment in China," Technical Report, working paper 2016.

Costinot, Arnaud and Dave Donaldson, "How Large Are the Gains from Economic Integration? Theory and Evidence from U.S. Agriculture, 1880-1997," Technical Report w22946, National Bureau of Economic Research, Cambridge, MA December 2016.

Donaldson, Dave, "Railroads of the Raj Estimating the Impact of Transportation Infrastructure," *American Economic Review*, 2018, *108* (4-5), 899–934.

_ **and Richard Hornbeck**, " Railroads and American Economic Growth: A "Market Access" Approach," *The Quarterly Journal of Economics*, 02 2016, *131* (2), 799–858.

Dumas, Christelle and Ximena Játiva, "Better Roads, Better Off? Evidence on Upgrading Roads in Tanzania," *The World Bank Economic Review*, 04 2024, *39* (1), 104–123.

Duranton, Gilles, Peter M. Morrow, and Matthew A. Turner, "Roads and Trade: Evidence from the US," *Review of Economic Studies*, April 2014, *81* (2), 681–724.

Eaton, Jonathan and Samuel Kortum, "Technology, Geography, and Trade," *Econometrica*, Sep. 2002, *70* (5), 1741–1779.

Faber, Benjamin, "Trade Integration, Market Size, and Industrialization: Evidence from China's National Trunk Highway System," *The Review of Economic Studies*, 03 2014, *81* (3), 1046–1070.

Fan, Jingting, "Internal Geography, Labor Mobility, and the Distributional Impacts of Trade," *American Economic Journal: Macroeconomics*, 2019, *11* (3), 252–88.

_ **, Yi Lu, and Wenlan Luo**, "Valuing Domestic Transport Infrastructure: A View from the Route Choice of Exporters," *The Review of Economics and Statistics*, November 2023, *105* (6), 1562–1579.

Fei, Xuan, Yumin Hu, and Mingzhi Xu, "Misallocation in the Chinese land market," *Canadian Journal of Economics/Revue canadienne d'économique*, 2024, *57* (4), 1360–1390.

Gai, Qingen, Naijia Guo, Bingjing Li, Qinghua Shi, and Xiaodong Zhu, "Migration Costs, Sorting, and the Agricultural Productivity Gap," 2024. Working Paper.

He, Jianhua, Yaolin Liu, Yan Yu, Wenwu Tang, Weining Xiang, and Dianfeng Liu, "A counterfactual scenario simulation approach for assessing the impact of farmland preservation policies on urban sprawl and food security in a major grain-producing area of China," *Applied Geography*, 2013, *37*, 127–138.

Hsu, Wen-Tai and Lin Ma, "Urbanization policy and economic development: A quantitative analysis of China's differential hukou reforms," *Regional Science and Urban Economics*, 2021, p. 103639.

Huang, Jikun, Will Martin, Yu Liu, and Scott Rozelle, "Agricultural Trade Reform and Rural Prosperity: Lessons from China," *Food Politics & Sociology eJournal*, 2008.

Huang, Zhonghua, Xuejun Du, and Carlos Salvador Zepeda Castillo, "How does urbanization affect farmland protection? Evidence from China," *Resources, Conservation and Recycling*, 2019, *145*, 139–147.

Imbert, Clément, Marlon Seror, Yifan Zhang, and Yanos Zylberberg, "Migrants and Firms: Evidence from China," *American Economic Review*, 2022, *112* (6), 1885–1914.

Jiang, Li, Xiangzheng Deng, and Karen C. Seto, "The impact of urban expansion on agricultural land use intensity in China," *Land Use Policy*, 2013, *35*, 33–39.

Kleinman, Benny, Ernest Liu, and Stephen J. Redding, "Dynamic Spatial General Equilibrium," *Econometrica*, March 2023, *91* (2), 385–424.

Krugman, Paul, "Increasing returns, monopolistic competition, and international trade," *Journal of International Economics*, 1979, *9* (4), 469 – 479.

Lagakos, David and Michael E. Waugh, "Selection, Agriculture, and Cross-Country Productivity Differences," *American Economic Review*, April 2013, *103* (2), 948–980.

Lichtenberg, Erik and Chengri Ding, "Assessing farmland protection policy in China," *Land Use Policy*, 2008, *25* (1), 59–68.

Lin, Yatang, "Travel Costs and Urban Specialization Patterns: Evidence from China's High Speed Railway System," *Journal of Urban Economics*, 2017, *98*, 98–123.

Ma, Lin and Yang Tang, "Geography, trade, and internal migration in China," *Journal of Urban Economics*, 2020, *115*, 103181.

_ **and** _ , "The distributional impacts of transportation networks in China," *Journal of International Economics*, 2024, p. 103873.

McMillan, John, John Whalley, and Lijing Zhu, "The impact of China's economic reforms on agricultural productivity growth," *Journal of Political Economy*, 1989, *97* (4), 781–807.

Minale, Luigi, "Agricultural Productivity Shocks, Labour Reallocation and Rural–Urban Migration in China," *Journal of Economic Geography*, 2018, *18* (4), 795–821.

Monte, Ferdinando, Stephen J. Redding, and Esteban Rossi-Hansberg, "Commuting, Migration, and Local Employment Elasticities," *American Economic Review*, December 2018, *108* (12), 3855–90.

Morten, Melanie and Jaqueline Oliveira, "The Effects of Roads on Trade and Migration: Evidence from a Planned Capital City," *American Economic Journal: Applied Economics*, April 2024, *16* (2), 389–421.

Ngai, L. Rachel, Christopher A. Pissarides, and Jin Wang, "Chinaś Mobility Barriers and Employment Allocations," *Journal of the European Economic Association*, October 2019, *17* (5), 1617–1653.

Pavcnik, Nina, "The impact of trade on inequality in developing countries," Technical Report, National Bureau of Economic Research 2017.

Qin, Yu, "'No county left behind?' The distributional impact of high-speed rail upgrades in China," *Journal of Economic Geography*, 06 2016, *17* (3), 489–520.

Restuccia, Diego, Dennis Tao Yang, and Xiaodong Zhu, "Agriculture and Aggregate Productivity: A Quantitative Cross-Country Analysis," *Journal of Monetary Economics*, 2008, *55* (2), 234–250.

Shamdasani, Yogita, "Rural road infrastructure & agricultural production: Evidence from India," *Journal of Development Economics*, 2021, *152*, 102686.

Sotelo, Sebastian, "Domestic Trade Frictions and Agriculture," *Journal of Political Economy*, July 2020, *128* (7), 2690–2738. Publisher: The University of Chicago Press.

Tian, Yuan, Junjie Xia, and Rudai Yang, "Trade-induced urbanization and the making of modern agriculture," *Working Paper*, 2020.

Tombe, Trevor and Xiaodong Zhu, "Trade, Migration, and Productivity: A Quantitative Analysis of China," *American Economic Review*, May 2019, *109* (5), 1843–72.

Tsivanidis, Nick, "Evaluating the Impact of Urban Transit Infrastructure: Evidence from Bogotá's TransMilenio," 2022. Working Paper.

World Bank, "Safe, Clean, and Affordable: Transport for Development," Technical Report, World Bank 2008.

Young, Alwyn, "Inequality, the urban-rural gap, and migration," *The Quarterly Journal of Economics*, 2013, *128* (4), 1727–1785.

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

Yu, Yue, "Land-use regulation and economic development: Evidence from the farmland red line policy in China," 2019. Working Paper.

Online Appendix

A Additional Figures and Tables

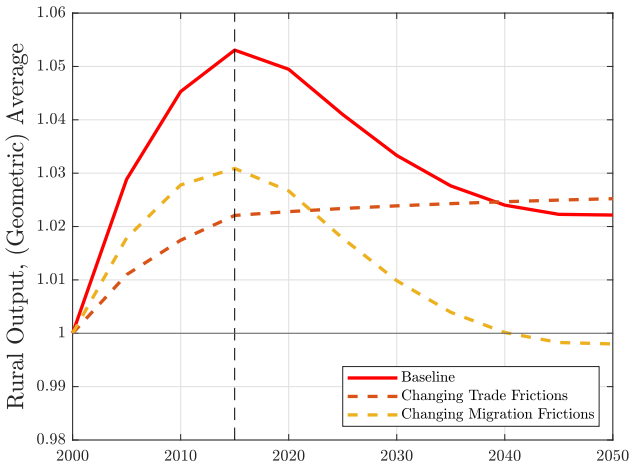


Figure A.1: Changes in Total Rural Output

Note: This figure plots the transition paths of the aggregate rural output in three simulations relative to the "no change" scenario in which the infrastructure network is fixed at the initial level in 2000. The "baseline" simulation allows both trade and migration frictions to change. The "reducing trade frictions" simulation allows only the trade frictions to change, and the "reducing migration frictions" simulation allows only the migration frictions to change. The vertical dotted line indicates 2015, the last year with transportation network data.

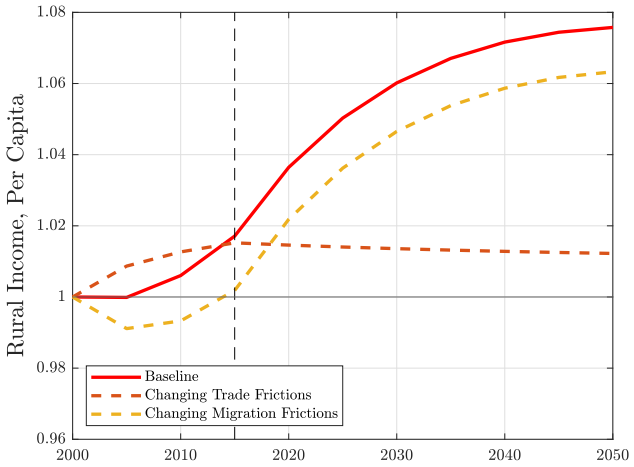


Figure A.2: Changes in Per Capita Rural Output

Note: This figure plots the transition paths of the aggregate rural output in three simulations relative to the "no change" scenario in which the infrastructure network is fixed at the initial level in 2000. The "baseline" simulation allows both trade and migration frictions to change. The "reducing trade frictions" simulation allows only the trade frictions to change, and the "reducing migration frictions" simulation allows only the migration frictions to change. The vertical dotted line indicates 2015, the last year with transportation network data.

Var.	(1) $\widehat{Y}$	(2) Land Rent	(3) Input Costs	(4) Market Access	(5) Efficiency Labor
Panel (a): Benchmark					
2015	1.053	0.987	1.006	1.021	1.038
2030	1.012	0.997	1.007	1.044	0.966
2045	0.981	1.005	1.007	1.056	0.918
Panel (b): Reducing τ					
2015	1.023	0.994	1.009	1.015	1.004
2030	1.026	0.994	1.010	1.015	1.007
2045	1.027	0.993	1.010	1.015	1.010
Panel (c): Reducing λ					
2015	1.030	0.993	0.996	1.006	1.035
2030	0.987	1.003	0.998	1.029	0.959
2045	0.955	1.012	0.997	1.041	0.909

Table A.1: The Changes in Total Rural Output, with Homogeneous Agents

Note: This table summarizes the impacts of transportation networks in each prefecture under the three simulations relative to the "no change" scenario in hat changes. The numbers in the table are the **simple average** of hat changes across all Chinese prefectures. Figure 1 is the total output, which is equivalent to the weighted average.

Var.	Δ UR	$\widehat{Y}$	$\widehat{\text{Market Access}}$	$\widehat{\text{Efficiency Labor}}$
	(1)	(2)	(3)	(4)
Baseline	0.038	1.022	1.047	0.976
No Fixed Costs	0.038	1.022	1.047	0.976
Low θ	0.038	1.016	1.032	0.977
High θ	0.038	1.024	1.053	0.976

Table A.2: Robustness of Long-Term Outcomes Under Alternative Model Assumptions

Note: This table reports robustness checks across three alternative model specifications. The first row reproduces the baseline results for comparison. The second row eliminates fixed costs in agricultural production. "Low θ" and "High θ" correspond to elasticity values of $\theta = 7$ and $\theta = 9$, respectively. For each specification, we recalibrate the model parameters and compute both the baseline and counterfactual simulations. The reported figures reflect the difference in outcomes between the recalibrated baseline and the "no change" scenario in the year 2045.

B Empirical Appendix

B.1 Data Appendix

National Fixed Point Survey: Our main analysis combines microdata from the National Fixed Point Survey (NFPS) with digitized national transportation maps of China. The NFPS is an origin-based longitudinal survey administered by the Research Center for Rural Economy under China's Ministry of Agriculture and Rural Affairs, in collaboration with the Central Policy Research Office. It covers over 300 villages across 30 provinces and features a relatively low attrition rate, making it one of the most comprehensive and consistent data sources on rural China. The survey includes village- and household-level data from 1986 onward, and individual-level data starting in 2003. The household module provides detailed information on agricultural production, employment, and income, while the individual module reports demographic characteristics, educational attainment, and labor migration patterns, including days and earnings from working outside the village.

To align with the transportation data and capture the period of rapid infrastructure expansion, we focus on household-level data from 1995 to 2013 and individual-level data from 2003 to 2013. Our empirical analysis is restricted to households located in villages that appear consistently throughout the study period. Because crop-level input data are unavailable before 2003, we aggregate inputs at the household level to ensure consistency over time. For individual-level analysis, we use self-reported migration information to study how improved access to transportation shapes migration decisions. This period coincides with China's era of double-digit growth, accelerated rural-to-urban migration, and major investments in transportation infrastructure—providing an ideal setting in which to examine how improved connectivity affects household-level agricultural outcomes and migration behavior.

Transportation Network Datasets: We borrow heavily from Ma and Tang (2024) to construct our measurements of rural villages' access to cities. On the one hand, the authors construct a detailed layer of transportation networks in China, tracking the evolution of every individual railroad and highway from its inception in the country. They digitized national maps to obtain binary connectivity of pixels with a length of around 500 meters, and cross-referenced it with the annual Collection of Provincial Transportation Maps. This provides a comprehensive foundational dataset for geo-coding our village-level analysis.

On the other hand, they construct a novel measure of transportation networks by incorporating variation in both the quality and quantity of roads over time and space. This is

achieved by using information on their "design speed" as well as transportation expansion resulting from government design changes and transportation infrastructure investments. As Ma and Tang (2024) point out in their paper, the Chinese government has not only invested heavily in transportation but also placed a strategic emphasis on elevating the construction standards of highways and railways over the years. A "highway" constructed in 1988 might only be appropriately classified as a "first-rate road" by the 2014 standards, which provides a source of quality variation across both the temporal and spatial dimensions that has been overlooked in other literature.

Based on this foundational dataset and variable construction method, we pinpoint the exact location of each village on the transportation map using detailed latitude and longitude coordinates provided by the NFPS data. We then calculate the average distance of each village to all major cities in China, using this as our main independent variable of interest, according to Equation 2.

Other Datasets: To conduct our empirical and quantitative analysis, we also use alternative datasets from various sources. City-level population and GDP are obtained from the City Statistical Yearbook for various years. The initial average tariff is from the World Integrated Trade Solution–Trade Analysis and Information System, using the applied tariff rates from 2000. The 2002 input–output table provided by the NBS of China is used to construct input tariffs. The expenditure share on different kinds of goods is constructed from the 2007 Chinese input–output table provided by the NBS of China. Population census datasets for 2000 and 2015, and the industrial census data for 2004 are also obtained from the NBS of China.

We obtain data related to the ROW from the World Development Indicators (WDI), hosted by the World Bank. We define ROW as all the countries included in WDI, excluding China. Recall that we need the initial population and land endowments for the year 2000. In that year, the total population of China was 1.26 billion, and the ROW was $6.14-1.26 = 4.88$ billion. Out of these 4.88 billion people, 49.84 percent are urbanized and the rest are rural. These countries together have an arable land endowment of 16 million square kilometers, which is $16/1.2 \approx 13.3$ times that of China.

Summary Statistics

Variable	Description	Mean	SD	Min	Max	Obs
Panel A: Village-specific Variable						
$Distance_{vt}$	Avg. Distance to Cities (min)	1225.8	383.6	642.1	3768.4	5994
Panel B: Household-specific Variable						
$Output^{Q}_{ht}$	Output in Quantity (kg)	4050.4	5286.8	0	62810	376057
$Output^{V}_{ht}$	Output in Value (RMB)	22261.0	262966.8	0	10640000	376057
$\mathbb{I}(\text{Farm})_{ht}$	Farming Indicator (1=Farm)	0.719	0.45	0	1	376057
$IncSh^{Farm}_{ht}$	Farming Income Share	0.333	0.287	0	1	375160
TFP_{ht}	Total Factor Productivity	1.151	0.645	0.001	26.541	254950
LP_{ht}	Labor Productivity (RMB per worker)	167.8	7189.0	0	3532144.3	306333
$FarmDays_{ht}$	Agricultural Labor Days	166.7	162.8	0	844	375802
Num^{Farm}_{ht}	Agricultural Labor Force	1.494	1.161	0	5	375200
$NumSh^{Farm}_{ht}$	Agricultural Labor Share	0.594	0.384	0	1	356168
K_{ht}	Capital (RMB)	1996.7	4656.1	0	45625.9	366757
N_{ht}	Land (mu)	7.2	11.3	0	90	376057
M_{ht}	Intermediate Input (RMB)	1006.5	1314.8	0	17964.2	376057
Panel C: Individual-specific Variable						
$\mathbf{1}_{it}$(Migration Out)	Migration Indicator ($\geq$ 180 days)	0.32	0.467	0	1	282899
$MigEarn_{it}$	Migration Income (RMB)	3717.58	6622.42	0	43217.51	282899

Table B.1: Summary Statistics

Note: This table provides summary statistics of main dependent variables and independent variables in our analyses. Distance to Cities is a variable measured at the village–year level, which is calculated as the unweighted average of each village to all Chinese cities in each year, following the method developed by Ma and Tang (2024). Output in Quantity is the aggregate output quantity in kilograms over thirteen crops at the household–year level, while Output in Value is the aggregate market value of all crops', of which each crop's market value is calculated by multiplying each crop's national-year average price by household–crop–year outputs. Farming Indicator is a dummy variable equals to 1 if the family self-reported farming as its main family business. Farming Income Share is the household's share of agricultural income over its total income. Total Factor Productivity is calculated from the estimation of a simple Cobb–Douglas production function using land, capital, intermediate input, and agricultural labor input, as Tian et al. (2020) and Chari et al. (2021) have also adopted. Labor Productivity is the average agricultural output of each agricultural worker. Agricultural Labor Days, Agricultural Labor Force, and Agricultural Labor Share are household–year level labor input in days, in heads, and the share of agricultural labor over total household labor force, respectively. Agricultural Capital is calculated using Perpetual Inventory System. Land scale is directly obtained from the data, while Intermediate Input is the aggregate of seeds and seedlings, fertilizers, agricultural films, and pesticides, expressed in monetary terms. The Migration Indicator is a dummy variable that is equal to 1 if the individual reports working out of town for at least 180 days in the year, and zero otherwise. Migration Income is the income from migration, adjusted for inflation. All monetary terms are inflation-adjusted to 2002.

B.2 Agricultural TFP Estimation

We estimate household agricultural total factor productivity (TFP) following Chari et al. (2021) and Tian et al. (2020) by using a standard Cobb-Douglas production function, which is specified below:

$$\ln y_{hvt} = \alpha \ln L_{ht} + \beta \ln N_{ht} + \gamma \ln K_{ht} + \delta \ln M_{ht} + \phi_{hvt}, \tag{B.1}$$

where $\ln y_{hvt}$ represents the output value of crop farming for household h in village v in year t; $\ln L_{ht}$, $\ln N_{ht}$, $\ln K_{ht}$, and $\ln M_{ht}$ denote labor days in agricultural production, and the inputs of land, capital, and intermediates, respectively. In the NFPS data, we observe the annual output of each crop type in kilograms, alongside four key categories of agricultural inputs: land (measured in hectares), labor (number of workdays dedicated to agricultural activities), capital (initial book value of all agricultural machinery), and intermediate inputs. The intermediate inputs include expenses related to seeds and seedlings, fertilizers, agricultural films, and pesticides, expressed in monetary terms. The recorded capital value of machinery is based on its initial book value, but it is adjusted over time for inflation, depreciation, and missing values using the perpetual inventory method to estimate the household-level capital stock more accurately. An important consideration is that, while output data for each crop is consistently recorded throughout the study, some input categories—such as labor time and machinery usage—are reported as aggregate values for all or a subset of crops rather than individually for each crop in certain years. This limitation is understandable, as farmers may find it difficult to accurately track labor hours and machinery use for each crop when managing multiple crops simultaneously. As a result, our baseline specification calculates household-level TFP using total household output and input data, providing a more reliable measure given these reporting constraints.

After estimating (B.1), we obtain a household–year specific measure of TFP, defined as $\hat{\phi}_{hvt} = \ln y_{hvt} - \hat{\alpha} \ln L_{ht} - \hat{\beta} \ln N_{ht} - \hat{\gamma} \ln K_{ht} - \hat{\delta} \ln M_{ht}$, where $\hat{\alpha}$, $\hat{\beta}$, $\hat{\gamma}$, and $\hat{\delta}$ represent the estimated elasticities for labor, land, capital, and intermediate inputs, respectively. This approach provides a clear measure of TFP at the household-year level, reflecting the efficiency with which these inputs are used in agricultural production. We further decompose the productivity term into two components: (i) a fixed household component, which captures the household's innate ability to farm a given crop, and (ii) a village-year component, which reflects time-varying productivity shocks common to all households within the village. This follows the fixed effects decomposition, as shown below:

$$\hat{\phi}_{hvt} = \hat{\phi}_{vt} + \hat{\phi}_{h} + \hat{e}_{hvt}, \tag{B.2}$$

In the empirical analysis, we focus on two key measures of TFP: the household-year specific measure of TFP ($\hat{\phi}_{hvt}$), and a time-invariant household measure ($\hat{\phi}_h$), derived by estimating the fixed effect from the above regression. We use the first measure to examine how China's transportation network expansion affects household agricultural TFP. The second measure is employed to construct interactions with our policy variable, which enables us to analyze heterogeneous responses to improved access to input and output markets. In addition to the revenue-based TFP measure, we also estimate a quantity-based TFP by replacing the output variable with the household's total physical agricultural output and reapplying the same estimation process.

B.3 Table Appendix

	Productivity		Extensive Margin		Intensive Margin	
	(1)	(2)	(3)	(4)	(5)	(6)
Dep var.	$\ln TFP_{ht}$	$\ln LP_{ht}$	$\mathbb{I}(Farm)_{ht}$	$IncSh_{ht}^{Farm}$	$\ln Output_{ht}^{V}$	$\ln Output_{ht}^{Q}$
Panel A: Average Effect						
$\ln Distance_{v,t-1}$	-3.203	-10.913**	1.928	0.663	-14.992**	-14.987***
	(2.175)	(4.593)	(1.221)	(0.731)	(6.473)	(5.639)
Obs	234816	277753	342396	341588	289775	289795
Weak instr. (F-statistic)	27.986	26.346	27.100	27.100	26.216	26.216
Panel B: Heterogeneity Pattern						
$\ln Distance_{v,t-1}$	-3.197	-6.102*	1.876**	0.978**	-3.429	-4.597**
	(2.176)	(3.480)	(0.817)	(0.480)	(3.543)	(2.332)
$\ln Distance_{v,t-1} \times \ln TFP_{ht_0}$	-0.289	-3.348***	-0.207**	-0.088	-4.516***	-3.125***
	(0.248)	(0.489)	(0.095)	(0.062)	(0.485)	(0.390)
Obs	234816	234815	234816	234734	234815	234816
Weak instr. (F-statistic)	14.000	14.049	14.000	14.000	14.000	14.000

Table B.2: Robustness: Instrumental Variable Approach

Note: The dependent variables are the same as in Table 1. The instrument variable is the interaction of a time-invariant dummy variable equal to 1 if village v is located in the least-cost-path (LCP) of China's highway network, as introduced by Faber (2014), and the nationwide transportation investment in the previous year. All regressions control for household and province–year fixed effects. Standard errors, clustered at the village–year level, are reported in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

	Productivity		Extensive Margin		Intensive Margin	
	(1)	(2)	(3)	(4)	(5)	(6)
Dep var.	$\ln TFP_{ht}$	$\ln LP_{ht}$	$\mathbb{I}(Farm)_{ht}$	$IncSh_{ht}^{Farm}$	$\ln Output_{ht}^{V}$	$\ln Output_{ht}^{Q}$
Panel A: Average Effect						
$\ln Distance_{v,t-1}$	-0.356***	-0.766***	0.204***	0.148***	-0.146	-0.400**
	(0.138)	(0.222)	(0.059)	(0.035)	(0.233)	(0.161)
Obs	254950	303530	376057	375160	317084	317104
R^2	0.828	0.657	0.568	0.692	0.674	0.712
Panel B: Heterogeneity Pattern						
$\ln Distance_{v,t-1}$	-0.357***	-0.687***	0.028	0.122***	-0.115	-0.321*
	(0.138)	(0.214)	(0.052)	(0.034)	(0.212)	(0.171)
$\ln Distance_{v,t-1} \times \ln TFP_{ht_0}$	-0.553***	-2.298***	-0.086**	-0.145***	-3.395***	-2.848***
	(0.104)	(0.155)	(0.037)	(0.025)	(0.169)	(0.149)
Obs	254950	254949	254950	254850	254949	254950
R^2	0.829	0.664	0.503	0.689	0.679	0.722

Table B.3: Robustness: Alternative Distance Measurement

Note: The dependent variables are the same as in Table 1. The independent variable is the weighted average of each village's distance to all Chinese cities, using cities' population as weights. All regressions control for household and province–year fixed effects. Standard errors, clustered at the village–year level, are reported in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

	Productivity		Extensive Margin		Intensive Margin	
	(1)	(2)	(3)	(4)	(5)	(6)
Dep var.	$\ln TFP_{ht}$	$\ln LP_{ht}$	$\mathbb{I}(Farm)_{ht}$	$IncSh_{ht}^{Farm}$	$\ln Output_{ht}^{V}$	$\ln Output_{ht}^{Q}$
Panel A: Average Effect						
$\ln Distance_{v,t-1}$	-0.400***	-0.921***	0.217***	0.179***	-0.172	-0.529***
	(0.153)	(0.246)	(0.064)	(0.040)	(0.256)	(0.178)
Obs	248087	291240	357113	356241	303999	304019
R^2	0.828	0.657	0.564	0.688	0.672	0.711
Panel B: Heterogeneity Pattern						
$\ln Distance_{v,t-1}$	-0.401***	-0.795***	0.015	0.139***	-0.147	-0.385**
	(0.153)	(0.238)	(0.057)	(0.037)	(0.233)	(0.189)
$\ln Distance_{v,t-1} \times \ln TFP_{ht_0}$	-0.552***	-2.373***	-0.077**	-0.147***	-3.449***	-2.903***
	(0.107)	(0.158)	(0.037)	(0.026)	(0.171)	(0.151)
Obs	248087	248086	248087	247995	248086	248087
R^2	0.828	0.663	0.501	0.689	0.679	0.722

Table B.4: Robustness: Excluding Influential Cities

Note: The dependent variables are the same as in Table 1. This robustness check excludes all household–level samples residing in influential cities, i.e., each province's capital city, to avoid potential concerns driven by those special observations. All regressions control for household and province–year fixed effects. Standard errors, clustered at the village–year level, are reported in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

B.4 Figure Appendix

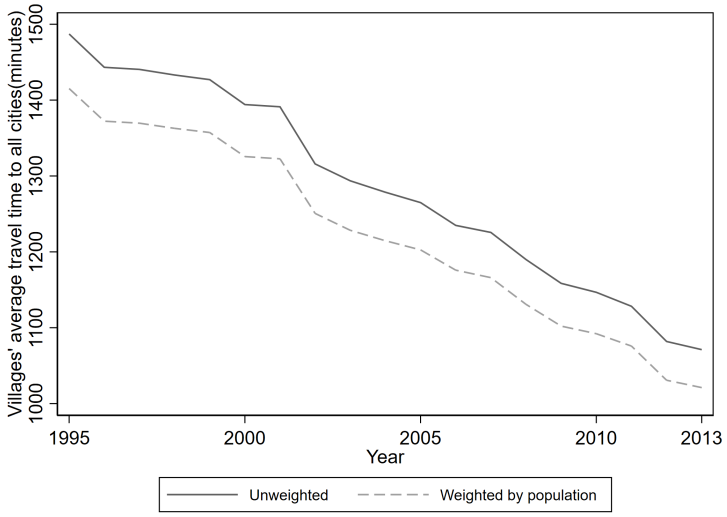


Figure B.1: Trend in Average Travel Time to All Cities from 1995 to 2013

Note: This figure plots the average travel time from villages to all cities between 1995 and 2013 (in minutes). The solid line shows the unweighted average across villages, while the dashed line presents the population-weighted average using population as weights— which is more closely aligned with the concept of "market access" developed by Donaldson and Hornbeck (2016).

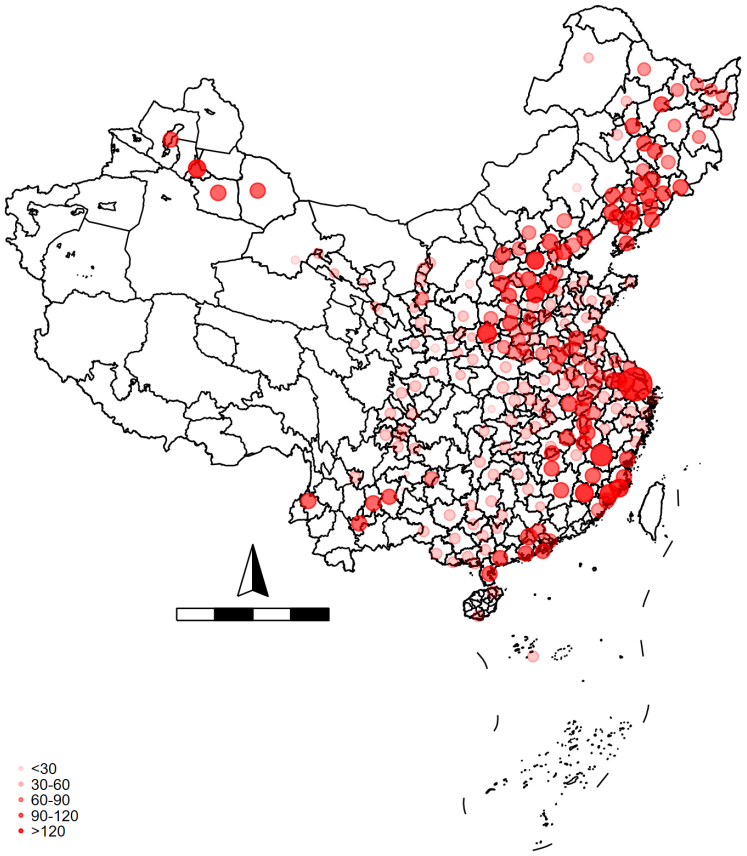


Figure B.2: Geographical Distribution of NFPS Sampled Households

Note: This figure displays the size and geographical distribution of sampled households in the NFPS in 1995. The radius of each circle reflects the number of surveyed households in each county. Additional details on the data are provided in Section B.1.

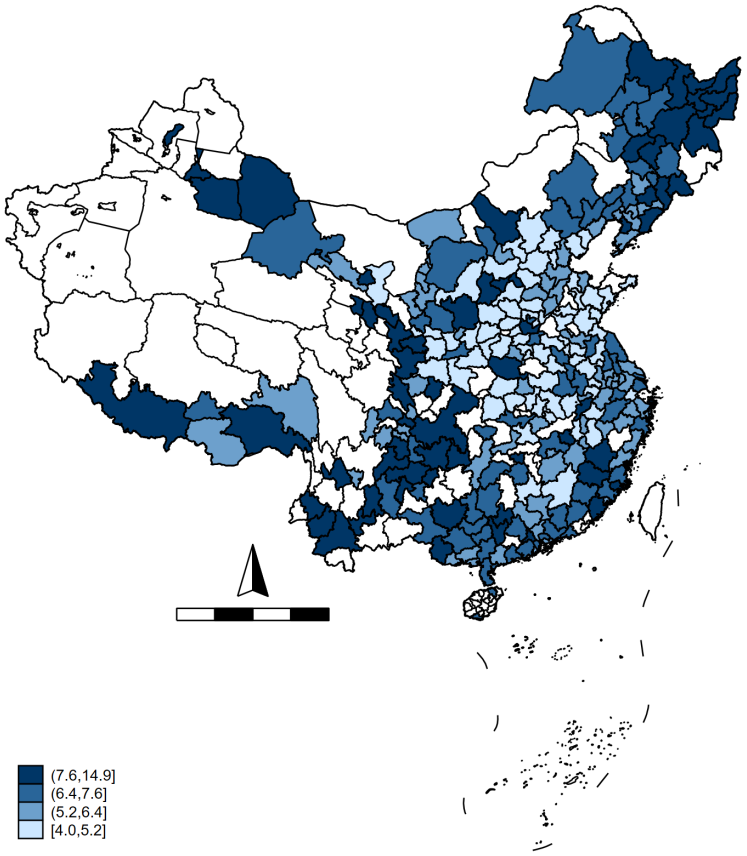


Figure B.3: Change in Village Travel Time to Cities, 1995–2013 (in Hours)

Note: This figure presents the geographical distribution of our unweighted distance measure across regions and over time. The underlying data are sourced from Ma and Tang (2024). We calculate each village's average travel distance, measured in hours to all major Chinese cities, following the data and methodology outlined in Equation 2 in the main text. This map illustrates the changes in this distance measure across cities from 1995 to 2013.

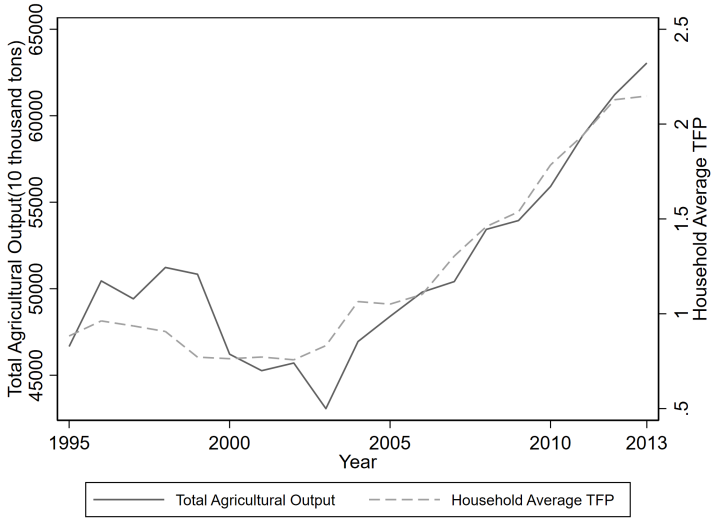


Figure B.4: Trends in Average Household TFP and Total Agricultural Output

Note: This figure depicts the annual trend of average TFP across all households (dashed line, right y-axis) alongside total agricultural output in quantity (solid line, left y-axis).

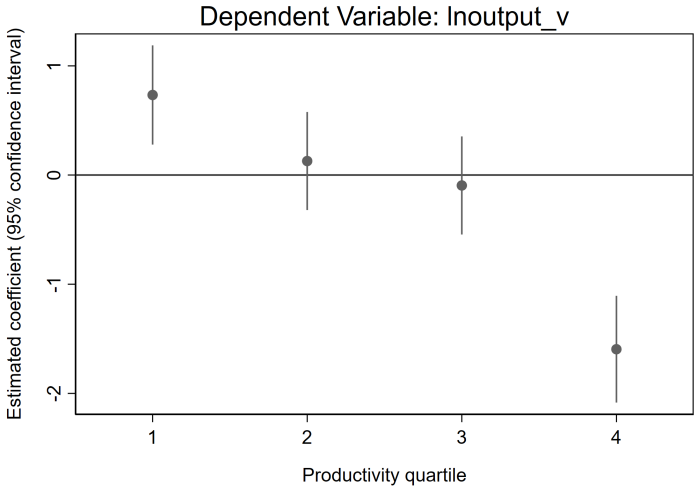


Figure B.5: Impact of Transportation on Output Across Productivity Quartiles

Note: This figure presents the estimated coefficients and 95% confidence intervals from a regression of the logarithm of household agricultural output (measured in value terms) on productivity quartiles. These quartiles are determined based on household-level TFP, which is time-invariant and calculated according to Equation B.2.

C Theory Appendix

C.1 Model Aggregation

Let X_{kt}^a represent the total expenditure on agricultural products in location k at time t. Under monopolistic competition, the revenue earned in location l by a type-z farmer from rural region k is:

$$R_{zlkt}^a = \frac{1}{(P_{lt}^a)^{1-\epsilon^a}} \left(\frac{\epsilon^a}{\epsilon^a - 1} \frac{\tau_{lkt}^a \ \chi_{kt}}{A_k \ a_z} \right)^{1-\epsilon^a} X_{lt}^a, \tag{C.1}$$

where $\tau_{lkt}^a \geq 1$ is the iceberg trade cost from k to l in the agriculture sector. The variables P_{lt}^a and χ_{kt} denote the price index of agricultural products in location l and the marginal cost of production in location k, respectively, as defined in the following equations:

$$P_{lt}^a \equiv \left[\int_0^1 \left(p_{lt}(\omega^a) \right)^{1-\epsilon^a} d\omega^a \right]^{\frac{1}{1-\epsilon^a}}, \quad \chi_{kt} \equiv r_{kt}^{\nu^n} \left(P_{kt}^m \right)^{\nu^m} \left(P_{kt}^c \right)^{1-\nu^n-\nu^m} \tag{C.2}$$

In the above equations, the variable r_{kt} is the land rent, and P_{kt}^m and P_{kt}^c are the price indices of manufacturing and capital products in location m, respectively:

$$P_{kt}^m = \left[\int_0^1 \left(p_{kt}(\omega^m) \right)^{1-\epsilon^m} d\omega^m \right]^{\frac{1}{1-\epsilon^m}}, \quad P_{kt}^c = \left[\int_0^1 \left(p_{kt}(\omega^c) \right)^{1-\epsilon^c} d\omega^c \right]^{\frac{1}{1-\epsilon^c}}. \tag{C.3}$$

Aggregating the individual farm sales from equation (C.1), the aggregate agricultural sales from k to l is:

$$\begin{aligned} R_{lkt}^a &= \sum_{z=1}^{Z} R_{zlkt}^a L_{zkt} = \sum_{z=1}^{Z} \frac{X_{lt}^a}{(P_{lt}^a)^{1-\epsilon^a}} \left(\frac{\epsilon^a}{\epsilon^a - 1} \frac{\tau_{lkt}^a \chi_{kt}}{A_k a_z} \right)^{1-\epsilon^a} L_{zkt} \\ &= \frac{X_{lt}^a}{(P_{lt}^a)^{1-\epsilon^a}} \left(\frac{\epsilon^a}{\epsilon^a - 1} \frac{\tau_{lkt}^a \chi_{kt}}{A_k} \right)^{1-\epsilon^a} \widetilde{L}_{kt}^r \end{aligned}$$

where $\widetilde{L}_{kt}^r$ summarizes the agricultural talent in location k at time t, which is defined by the following equation:

$$\widetilde{L}_{kt}^r = \left[\sum_{z=1}^{Z} \left(a_z \right)^{\epsilon^a - 1} L_{zkt} \right], \tag{C.4}$$

The total revenue of farms in location k is the sum of all sales to other locations:

$$\begin{aligned} R_{kt} &= \sum_{l=1}^{K} R_{lkt}^{a} \\ &= \sum_{l=1}^{K} \frac{X_{lt}^{a}}{\left(P_{lt}^{a}\right)^{1-\epsilon^{a}}} \left(\frac{\epsilon^{a}}{\epsilon^{a}-1} \frac{\tau_{lkt}^{a} \chi_{kt}}{A_{k}}\right)^{1-\epsilon^{a}} \widetilde{L}_{kt}^{r} \\ &= \widetilde{\epsilon}^{a} \left(\frac{\chi_{kt}}{A_{k}}\right)^{1-\epsilon^{a}} M_{kt}^{a} \times \widetilde{L}_{kt}^{r}, \forall k \in \mathcal{R}. \end{aligned} \tag{C.5}$$

where $\widetilde{\epsilon}^{a} = \left(\frac{\epsilon^{a}}{\epsilon^{a}-1}\right)^{1-\epsilon^{a}}$ is a constant, and M_{kt}^{a} is the market access measure for farms operating in location k, defined as $M_{kt}^{a} \equiv \sum_{l=1}^{K} X_{lt}^{a} / \left(P_{lt}^{a}\right)^{1-\epsilon^{a}} \left(\tau_{lkt}^{a}\right)^{1-\epsilon^{a}}$.

The aggregate income of farmers in location k is $\frac{R_{kt}}{\epsilon^{a}}$, while the income of landowners is $\frac{R_{kt}\nu^{n}(\epsilon^{a}-1)}{\epsilon^{a}}$. Consequently, the total income of the rural location, Y_{kt}, is:

$$Y_{kt} = \left(\frac{1}{\epsilon^{a}} + \nu^{n} \frac{\epsilon^{a}-1}{\epsilon^{a}}\right) R_{kt} = \frac{1+\nu^{n}\left(\epsilon^{a}-1\right)}{\epsilon^{a}} R_{kt}, \forall k \in \mathcal{R}. \tag{C.6}$$

The total expenditure on agricultural products is given by $X_{kt}^{a} = \eta^{a} Y_{kt}$. Expenditure on manufactured and capital goods comes from two sources: consumption demand and intermediate inputs:

$$\begin{aligned} X_{kt}^{m} &= \eta^{m} Y_{kt} + \nu^{m} \frac{\epsilon^{a}-1}{\epsilon^{a}} R_{kt} \\ &= \left[\eta^{m} \frac{1+\nu^{n}\left(\epsilon^{a}-1\right)}{\epsilon^{a}} + \nu^{m} \frac{\epsilon^{a}-1}{\epsilon^{a}}\right] R_{kt}, \\ X_{kt}^{c} &= \left(1-\eta^{a}-\eta^{m}\right) Y_{kt} + \left(1-\nu^{n}-\nu^{m}\right) \frac{\epsilon^{a}-1}{\epsilon^{a}} R_{kt} \\ &= \left[\left(1-\eta^{a}-\eta^{m}\right) \frac{1+\nu^{n}\left(\epsilon^{a}-1\right)}{\epsilon^{a}} + \left(1-\nu^{n}-\nu^{m}\right) \frac{\epsilon^{a}-1}{\epsilon^{a}}\right] R_{kt}. \end{aligned}$$

Combining the terms, the total expenditure on urban products is:

$$\begin{aligned} X_{kt}^{m} + X_{kt}^{c} &= \Xi \times R_{kt} \\ \Xi &= \left(1-\eta^{a}\right) \frac{1+\nu^{n}\left(\epsilon^{a}-1\right)}{\epsilon^{a}} + \left(1-\nu^{n}\right) \frac{\epsilon^{a}-1}{\epsilon^{a}} \\ &= \frac{1-\eta^{a}+\nu^{n}(\epsilon^{a}-1)-\eta^{a}\nu^{n}(\epsilon^{a}-1)+(\epsilon^{a}-1)-\nu^{n}(\epsilon^{a}-1)}{\epsilon^{a}} \\ &= \frac{1-\eta^{a}-\eta^{a}\nu^{n}(\epsilon^{a}-1)+(\epsilon^{a}-1)}{\epsilon^{a}} \\ &= \frac{1-\eta^{a}+\left(\epsilon^{a}-1\right)\left(1-\eta^{a}\nu^{n}\right)}{\epsilon^{a}}, \end{aligned}$$

The rural land market clears when the total income of landowners is equal to the expen-

diture of farmers on land:

$$r_{kt}N_k = \nu^n \frac{\epsilon^a - 1}{\epsilon^a} R_{kt}. \tag{C.7}$$

By combining equations (C.6) and (C.7), we can also express rural income as a function of land rent:

$$\begin{aligned} Y_{kt} &= \frac{1 + \nu^n (\epsilon^a - 1)}{\epsilon^a} \left(\frac{r_{kt}N_k}{\nu^n} \frac{\epsilon^a}{\epsilon^a - 1} \right) \\ &= \left[1 + \frac{1}{\nu^n (\epsilon^a - 1)} \right] r_{kt}N_k \end{aligned} \tag{C.8}$$

C.2 Trade Balance

C.2.1 Rural Locations

In equilibrium, the total expenditure on manufactured and capital products of a rural location $X^m_{kt} + X^c_{kt} = \Xi R_{kt}$ is the total imports of the rural location k from all the urban locations. Trade balance implies that this should be equal to the total sales of agricultural products to all the urban locations:

$$\begin{aligned} \Xi R_{kt} = X^m_{kt} + X^c_{kt} &= \sum_{l \in \mathcal{U}} R^a_{lkt} \\ &= \sum_{l \in \mathcal{U}} \frac{X^a_{lt}}{(P^a_{lt})^{1-\epsilon^a}} \left(\frac{\epsilon^a}{\epsilon^a - 1} \tau^a_{lkt} \frac{\chi_{kt}}{A_k} \right)^{1-\epsilon^a} \widetilde{L}^r_{kt} \\ &= \sum_{l \in \mathcal{U}} \frac{\eta^a Y_{lt}}{\sum_{k' \in \mathcal{R}} \widetilde{\epsilon}^a \left(\tau^a_{lk't} \frac{\chi_{k't}}{A_{k'}} \right)^{1-\epsilon^a} \widetilde{L}^r_{k't}} \widetilde{\epsilon}^a \left(\tau^a_{lkt} \frac{\chi_{kt}}{A_k} \right)^{1-\epsilon^a} \widetilde{L}^r_{kt} \\ &= \sum_{l \in \mathcal{U}} \frac{\eta^a Y_{lt}}{\sum_{k' \in \mathcal{R}} \left(\tau^a_{lk't} \frac{\chi_{k't}}{A_{k'}} \right)^{1-\epsilon^a} \widetilde{L}^r_{k't}} \left(\tau^a_{lkt} \frac{\chi_{kt}}{A_k} \right)^{1-\epsilon^a} \widetilde{L}^r_{kt} \end{aligned}$$

In the second to last line, we use the solution of the agricultural price index derived later in equation (C.15). Substitute in the land market clearing condition:

$$r_{kt}N_{kt} \left(\Xi \frac{\epsilon^a}{\epsilon^a - 1} \frac{1}{\nu^n} \right) = \sum_{l \in \mathcal{U}} \frac{\eta^a Y_{lt}}{\sum_{k' \in \mathcal{R}} \left(\tau^a_{lk't} \frac{\chi_{k't}}{A_k} \right)^{1-\epsilon^a} \widetilde{L}^r_{k't}} \left(\tau^a_{lkt} \frac{\chi_{kt}}{A_k} \right)^{1-\epsilon^a} \widetilde{L}^r_{kt}. \tag{C.9}$$

Note that the above equations implicitly define a vector of $\{r_{kt}\}$ that clears the rural markets, as χ_{kt} are functions of r_{kt} as well. We solve the vector r_{kt} similarly to w_{kt}, as explained below.

Alternative Trade Balance Instead of thinking about the trade between urban and rural locations, we can also write down the trade balance conditions from the income–expenditure equations. The total revenue of all the farmers in location k, R_{kt}, must be equal to the sales to all the rural and urban locations:

$$\begin{aligned} R_{kt} &= \sum_{l=1}^{K} \frac{\eta^a Y_{lt}}{\sum_{k' \in \mathcal{R}} \left(\tau^a_{lk't} \frac{\chi_{k't}}{A_k}\right)^{1-\epsilon^a} \widetilde{L}^r_{k't}} \left(\tau^a_{lkt} \frac{\chi_{kt}}{A_k}\right)^{1-\epsilon^a} \widetilde{L}^r_{kt} \\ r_{kt} N_{kt} \frac{\epsilon^a}{\epsilon^a - 1} \frac{1}{\nu^n} &= \sum_{l=1}^{K} \frac{\eta^a Y_{lt}}{\sum_{k' \in \mathcal{R}} \left(\tau^a_{lk't} \frac{\chi_{k't}}{A_k}\right)^{1-\epsilon^a} \widetilde{L}^r_{k't}} \left(\tau^a_{lkt} \frac{\chi_{kt}}{A_k}\right)^{1-\epsilon^a} \widetilde{L}^r_{kt}, \end{aligned} \tag{C.10}$$

where Y_{lt} takes different forms depending on whether l is rural or urban, using equations (C.8) and (C.11), respectively.

C.2.2 Urban Locations

In an urban location, due to the assumption of perfect competition, the total income is simply:

$$\begin{aligned} Y_{kt} &= w_k \sum_{z=1}^{Z} m_z L_{zkt}, \forall k \in \mathcal{U} \\ &= w_k \widetilde{L}^u_{kt}. \end{aligned} \tag{C.11}$$

in which the second term is the total supply of efficiency labor at location k:

$$\widetilde{L}^u_{kt} = \sum_{z=1}^{Z} m_z L_{zkt}. \tag{C.12}$$

The total expenditure on agricultural products is the same as in the rural locations at $X^a_{kt} = \eta^a Y_{kt}$. The expenditure on the manufactured and capital products is $\eta^m Y_{kt}$ and $(1 - \eta^a - \eta^m) Y_{kt}$. The expressions are simpler as compared to the rural locations due to the absence of intermediate inputs in urban production.

Recall that the total expenditure on the manufactured and capital products at location k is $X^b_{kt}, b \in \{m, c\}$. The sales from location k to l at time t is then

$$R^b_{lkt} = \frac{(w_k \tau_{lkt})^{-\theta} \left(A^b_k\right)^{\theta}}{\sum_{k' \in \mathcal{U}} (w_{k'} \tau_{lk't})^{-\theta} \left(A^b_{k'}\right)^{\theta}} X^b_{lt}, b \in \{m, c\}. \tag{C.13}$$

The total sales of all the firms in location k is the summation of the expression above across destinations and sectors. Also note that the total sales must equal total income, so that

$R_{kt} = Y_{kt}$:

$$\begin{aligned}
R_{kt} &= \sum_{l=1}^{K} \frac{(w_k\tau_{lkt})^{-\theta}(A_k^m)^{\theta}}{\sum_{k'\in\mathcal{U}}(w_{k'}\tau_{lk't})^{-\theta}(A_{k'}^m)^{\theta}} X_{lt}^m + \sum_{l=1}^{K} \frac{(w_k\tau_{lkt})^{-\theta}(A_k^c)^{\theta}}{\sum_{k'\in\mathcal{U}}(w_{k'}\tau_{lk't})^{-\theta}(A_{k'}^c)^{\theta}} X_{lt}^c \\
&= \sum_{l\in\mathcal{U}} \left[\frac{(w_k\tau_{lkt})^{-\theta}(A_k^m)^{\theta}}{\sum_{k'\in\mathcal{U}}(w_{k'}\tau_{lk't})^{-\theta}(A_{k'}^m)^{\theta}} X_{lt}^m + \frac{(w_k\tau_{lkt})^{-\theta}(A_k^c)^{\theta}}{\sum_{k'\in\mathcal{U}}(w_{k'}\tau_{lk't})^{-\theta}(A_{k'}^c)^{\theta}} X_{lt}^c \right] \\
&\quad + \sum_{l\in\mathcal{R}} \left[\frac{(w_k\tau_{lkt})^{-\theta}(A_k^m)^{\theta}}{\sum_{k'\in\mathcal{U}}(w_{k'}\tau_{lk't})^{-\theta}(A_{k'}^m)^{\theta}} X_{lt}^m + \frac{(w_k\tau_{lkt})^{-\theta}(A_k^c)^{\theta}}{\sum_{k'\in\mathcal{U}}(w_{k'}\tau_{lk't})^{-\theta}(A_{k'}^c)^{\theta}} X_{lt}^c \right].
\end{aligned}$$

Note that the terms in the second line are the total exports to all the rural regions. Trade balance in equilibrium implies that it must be equal to the total imports from all the rural regions, which is in turn equal to the expenditure on agricultural products:

$$\sum_{l\in\mathcal{R}} \left[\frac{(w_k\tau_{lkt})^{-\theta}(A_k^m)^{\theta}}{\sum_{k'\in\mathcal{U}}(w_{k'}\tau_{lk't})^{-\theta}(A_{k'}^m)^{\theta}} X_{lt}^m + \frac{(w_k\tau_{lkt})^{-\theta}(A_k^c)^{\theta}}{\sum_{k'\in\mathcal{U}}(w_{k'}\tau_{lk't})^{-\theta}(A_{k'}^c)^{\theta}} X_{lt}^c \right] = X_{kt}^a = \eta^a Y_{kt} = \eta^a R_{kt}.$$

Substitute this expression back into the trade balance condition and simplify:

$$\begin{aligned}
R_{kt} &= \sum_{l\in\mathcal{U}} \left[\frac{(w_k\tau_{lkt})^{-\theta}(A_k^m)^{\theta}}{\sum_{k'\in\mathcal{U}}(w_{k'}\tau_{lk't})^{-\theta}\left(A_{k'}^m\right)^{\theta}} X_{lt}^m + \frac{(w_k\tau_{lkt})^{-\theta}(A_k^c)^{\theta}}{\sum_{k'\in\mathcal{U}}(w_{k'}\tau_{lk't})^{-\theta}\left(A_{k'}^c\right)^{\theta}} X_{lt}^c \right] \\
&\quad + \eta^a R_{kt} \\
(1-\eta^a) w_k \sum_{z=1}^{Z} m_z L_{zkt} &= \sum_{l\in\mathcal{U}} \frac{(w_k\tau_{lkt})^{-\theta}(A_k^m)^{\theta}}{\sum_{k'\in\mathcal{U}}(w_{k'}\tau_{lk't})^{-\theta}\left(A_{k'}^m\right)^{\theta}} \eta^m w_l \left(\sum_{z=1}^{Z} m_z L_{zlt} \right) \\
&\quad + \sum_{l\in\mathcal{U}} \frac{(w_k\tau_{lkt})^{-\theta}(A_k^c)^{\theta}}{\sum_{k'\in\mathcal{U}}(w_{k'}\tau_{lk't})^{-\theta}\left(A_{k'}^c\right)^{\theta}} (1-\eta^a-\eta^m) w_l \left(\sum_{z=1}^{Z} m_z L_{zlt} \right) \\
&= \sum_{l\in\mathcal{U}} \left[\frac{(w_k\tau_{lkt})^{-\theta}(A_k^m)^{\theta}}{\sum_{k'\in\mathcal{U}}(w_{k'}\tau_{lk't})^{-\theta}\left(A_{k'}^m\right)^{\theta}} + \frac{(w_k\tau_{lkt})^{-\theta}(A_k^c)^{\theta}}{\sum_{k'\in\mathcal{U}}(w_{k'}\tau_{lk't})^{-\theta}\left(A_{k'}^c\right)^{\theta}} \right] (1-\eta^a) w_l \left(\sum_{z=1}^{Z} m_z L_{zlt} \right),
\end{aligned}$$

which simplifies to the familiar expression of trade balance:

$$w_k \widetilde{L}_{kt}^u = \sum_{l\in\mathcal{U}} \left[\frac{(w_k\tau_{lkt})^{-\theta}(A_k^m)^{\theta}}{\sum_{k'\in\mathcal{U}}(w_{k'}\tau_{lk't})^{-\theta}\left(A_{k'}^m\right)^{\theta}} + \frac{(w_k\tau_{lkt})^{-\theta}(A_k^c)^{\theta}}{\sum_{k'\in\mathcal{U}}(w_{k'}\tau_{lk't})^{-\theta}\left(A_{k'}^c\right)^{\theta}} \right] w_l \widetilde{L}_{lt}^u, \tag{C.14}$$

C.3 Price Indices

In equilibrium, the agricultural price index of a location l can be written as:

$$\begin{aligned} P_{lt}^a &= \left(\int_0^1 \left(p_{lt}(\omega^a) \right)^{1-\epsilon^a} d\omega^a \right)^{\frac{1}{1-\epsilon^a}} \\ &= \left[\sum_{k \in \mathcal{R}} \left(\frac{\epsilon^a}{\epsilon^a - 1} \tau_{lkt}^a \frac{\chi_{kt}}{A_k} \right)^{1-\epsilon^a} \left(\sum_{z=1}^{Z} a_z^{\epsilon^a - 1} L_{zkt} \right) \right]^{\frac{1}{1-\epsilon^a}} \\ &= \left[\sum_{k \in \mathcal{R}} \widetilde{\epsilon}^a \left(\tau_{lkt}^a \frac{\chi_{kt}}{A_k} \right)^{1-\epsilon^a} \left(\sum_{z=1}^{Z} a_z^{\epsilon^a - 1} L_{zkt} \right) \right]^{\frac{1}{1-\epsilon^a}} \\ &= \left[\sum_{k \in \mathcal{R}} \widetilde{\epsilon}^a \left(\tau_{lkt}^a \frac{\chi_{kt}}{A_k} \right)^{1-\epsilon^a} \widetilde{L}_{kt}^a \right]^{\frac{1}{1-\epsilon^a}} . \end{aligned} \tag{C.15}$$

In the expression above, $\widetilde{\epsilon}^a$ is defined in the main text. As it is standard in the Eaton-Kortum models, the price indices for manufactured and capital goods are, respectively:

$$P_{lt}^m = \gamma^m \left[\sum_{k' \in \mathcal{U}} \left(w_{k'} \tau_{lk't} \right)^{-\theta} \left(A_{k'}^m \right)^{\theta} \right]^{-1/\theta} \tag{C.16}$$

$$P_{lt}^c = \gamma^c \left[\sum_{k' \in \mathcal{U}} \left(w_{k'} \tau_{lk't} \right)^{-\theta} \left(A_{k'}^c \right)^{\theta} \right]^{-1/\theta} , \tag{C.17}$$

where $\gamma^b = \left[\Gamma \left(\frac{\theta + 1 - \epsilon^b}{\theta} \right) \right]^{1/(1-\epsilon^b)}, b \in \{m, c\}$, and $\Gamma(\cdot)$ is the Gamma function. Note that the Gamma function imposes the following restrictions on parameters: $\epsilon^b < \theta + 1$.

Lastly, the price index in location l is derived as the following equation:

$$P_{lt} = \left(P_{lt}^a \right)^{\eta^a} \left(P_{lt}^m \right)^{\eta^m} \left(P_{lt}^c \right)^{1-\eta^a-\eta^m} \tag{C.18}$$

C.4 Side Note: Optimal Land Allocation

Consider a land allocation $\{N_k\}_{k=1}^K$ that maximizes aggregate rural income defined as $Y_t^r = \sum_{k=1}^K Y_{kt} = \left[1 + \frac{1}{\nu^n (\epsilon^a - 1)} \right] \sum_{k=1}^K r_{kt} N_k$. The necessary condition for optimization is $r_{kt} = r_t, \forall k$. To illustrate this, suppose an allocation results in $r_{k't} > r_{kt}$ for some k' and k. In this case, moving a marginal amount of land from k to k' would increase $r_{kt} N_k + r_{k't} N_{k'}$, and therefore lead to a higher total income.

C.5 Decomposing Agricultural Output

A central question in our analysis is how much of the observed change in agricultural outcomes can be attributed to average productivity gains versus selection effects driven by reallocation. To unpack this distinction, we begin with the definition of effective agricultural talent from equation (C.4). Let us denote the average talent of $(a_z)^{\epsilon^a-1}$ as $\widetilde{a}$:

$$\widetilde{a} = \frac{\sum_{z=1}^{Z} (a_z)^{\epsilon^a-1}}{Z}. \tag{C.19}$$

The endowment can be decomposed into two terms:

$$\begin{aligned}
\widetilde{L}^r_{kt} &= \sum_{z=1}^{Z} \left[(a_z)^{\epsilon^a-1} - \widetilde{a} + \widetilde{a} \right] L_{zkt} \\
&= \widetilde{a} \sum_{z=1}^{Z} (L_{zkt}) + \sum_{z=1}^{Z} \left[(a_z)^{\epsilon^a-1} - \widetilde{a} \right] (L_{zkt}) \\
&= \widetilde{a} L_{kt} + \sum_{z=1}^{Z} \left[(a_z)^{\epsilon^a-1} - \widetilde{a} \right] (L_{zkt}) \\
&= \widetilde{a} L_{kt} + S_{kt}
\end{aligned} \tag{C.20}$$

The expression above decomposes the total effective agricultural endowment into two distinct components: (i) changes in average talent, captured by $\widetilde{a}$, and (ii) a selection term, reflecting how the distribution of rural labor shifts toward households with higher productivity—driven by dispersion in a_z. This second component captures the reallocation of labor across talent types and can be negative ($S_{kt} < 0$) in locations dominated by below-average farmers. Together, this decomposition enables us to disentangle productivity gains arising from improved average talent from those due to selection, offering a clearer view of how enhanced connectivity and migration influence agricultural performance.

To formalize this decomposition, let the counterfactual level of effective agricultural talent be denoted by $\left(\widetilde{L}^r_{kt}\right)' = \left[\sum_{z=1}^{Z} (a_z)^{\epsilon^a-1} L'_{zkt}\right]$, where L'_{zkt} is the counterfactual level of population distribution. Then, the level changes in effective agricultural talent can be

decomposed as:

$$\begin{aligned}
\Delta\widetilde{L}^r_{kt} = \left(\widetilde{L}^r_{kt}\right)' - \widetilde{L}^r_{kt} &= \sum_{z=1}^{Z} (a_z)^{\epsilon^a-1} \left(L'_{zkt} - L_{zkt}\right) = \sum_{z=1}^{Z} \left[(a_z)^{\epsilon^a-1} - \widetilde{a} + \widetilde{a}\right] \left(L'_{zkt} - L_{zkt}\right) \\
&= \widetilde{a} \sum_{z=1}^{Z} \left(L'_{zkt} - L_{zkt}\right) + \sum_{z=1}^{Z} \left[(a_z)^{\epsilon^a-1} - \widetilde{a}\right] \left(L'_{zkt} - L_{zkt}\right) \\
&= \underbrace{\widetilde{a}\left(L'_{kt} - L_{kt}\right)}_{\text{Population Effect}} + \underbrace{\sum_{z=1}^{Z} \left[(a_z)^{\epsilon^a-1} - \widetilde{a}\right] \left(L'_{zkt} - L_{zkt}\right)}_{\text{Selection Effect}} \\
&= \tilde{a}\Delta L_{kt} + \Delta S_{kt}
\end{aligned}$$

In this expression, the first term reflects changes in effective agricultural talent due purely to shifts in population size, while the second term isolates the contribution from selection. A positive selection effect, $\Delta S_{kt} > 0$, indicates that higher productivity farmers are disproportionately represented in location k under the counterfactual, amplifying gains in effective talent beyond what would be predicted by average productivity alone. We can decompose the familiar hat changes as:

$$\begin{aligned}
\widehat{\widetilde{L}^r_{kt}} = \frac{\left(\widetilde{L}^r_{kt}\right)'}{\widetilde{L}^r_{kt}} &= \frac{\widetilde{a}\left(L'_{kt}\right) + \sum_{z=1}^{Z} \left[(a_z)^{\epsilon^a-1} - \widetilde{a}\right] \left(L'_{zkt}\right)}{\widetilde{L}^r_{kt}} \\
&= \frac{\widetilde{a}\left(L'_{kt}\right)}{\widetilde{a}L_{kt}} \frac{\widetilde{a}L_{kt}}{\widetilde{L}^r_{kt}} + \frac{\sum_{z=1}^{Z} \left[(a_z)^{\epsilon^a-1} - \widetilde{a}\right] \left(L'_{zkt}\right)}{\sum_{z=1}^{Z} \left[(a_z)^{\epsilon^a-1} - \widetilde{a}\right] \left(L_{zkt}\right)} \frac{\sum_{z=1}^{Z} \left[(a_z)^{\epsilon^a-1} - \widetilde{a}\right] \left(L_{zkt}\right)}{\widetilde{L}^r_{kt}} \\
&= \widehat{L}_{kt} \left(1 - \frac{S_{kt}}{\widetilde{L}^r_{kt}}\right) + \widehat{S}_{kt} \frac{S_{kt}}{\widetilde{L}^r_{kt}}.
\end{aligned}$$

This equation shows that the hat change in effective agricultural talent can be interpreted as a weighted average of the hat change in population ($\widehat{L}_{kt}$) and the selection term ($\widehat{S}_{kt}$), where the weight $\frac{S_{kt}}{\widetilde{L}^r_{kt}}$ captures the relative contribution of selection in the baseline equilibrium. Notably, this weight can be negative if location k attracts farmers with below-average productivity ($S_{kt} < 0$). Building on this, the hat change in aggregate rural income at location k, as given by equation (C.6), can be decomposed into four distinct channels:

$$\widehat{Y}_{kt} = \widehat{R}_{kt} = \left(\widehat{\chi}_{kt}\right)^{1-\epsilon^a} \times \widehat{M}^a_{kt} \times \left[\widehat{L}_{kt} \left(1 - \frac{S_{kt}}{\widetilde{L}^r_{kt}}\right) + \widehat{S}_{kt} \frac{S_{kt}}{\widetilde{L}^r_{kt}}\right]. \tag{C.21}$$

The first two hat terms capture the effects of changing production costs and market access, respectively. The terms in brackets then isolate the contribution of population changes and

the selection effect, highlighting how shifts in both scale and composition of rural labor influence income.

C.6 Quantification Appendix

To implement the quantitative analysis, we use the 2007 Chinese input–output table and calculate input cost shares in agriculture, manufacturing, and farm income based on the NFPS data.

Population Distribution To compute the baseline equilibrium, we need China's initial population in 2000, organized by location and ability. We assume that the two productivities are drawn from a joint log-normal distribution with parameters $(\mu_a, \mu_m, \sigma_a, \sigma_m, \rho)$, where μ_a (μ_m) stands for the mean of agricultural (manufacturing) productivity, and σ_a (σ_m) stands for the standard deviation of agricultural (manufacturing) productivity, and ρ the correlation between the agricultural and manufacturing productivities.

The moments used in calibration compare the agricultural (or manufacturing) income of a household with the average level of agricultural (or manufacturing) income in the same village. Consider a household z with productivity drawn (a_z, m_z) in a village k. Let π_{zt} and w_{zt} denote the revenues generated by agricultural and manufacturing activities in household z in year t. According to the model, the log ratio between household z's agricultural income in year t relative to the average agricultural income of village k in year t is provided as:

$$\ln \tilde{\pi}_z = (\epsilon^a - 1)(\ln a_z - \ln \bar{a}_z). \tag{C.22}$$

We could derive household z's manufacturing income relative to his village mean as:

$$\ln \tilde{w}_z = \ln m_z - \ln \bar{m}_z \tag{C.23}$$

where $\bar{a}_z$ and $\bar{m}_z$ denote the arithmetic average of agricultural and manufacturing productivity of village k.

We pool separately the transformed agricultural ($\ln \tilde{\pi}_z$) and manufacturing ($\ln \tilde{w}_z$) income ratio across households and village. For each type, we compute the mean, median, and the standard deviation of each income ratio to obtain our first two sets of moment conditions:

$$M1 = \Big\{ Mean(\ln \tilde{\pi}_z),\ Median(\ln \tilde{\pi}_z),\ Std(\ln \tilde{\pi}_z) \Big\} \tag{C.24}$$

and

$$M2 = \Big\{ Mean(\ln \tilde{w}_z),\ Median(\ln \tilde{w}_z),\ Std(\ln \tilde{w}_z) \Big\} \tag{C.25}$$

To capture potential correlation between agricultural and manufacturing productivities, we compute the correlation between $\ln \tilde{\pi}_z$ and $\ln \tilde{w}_z$ as our third moment:

$$M3 = Cor(\ln \tilde{\pi}_z, \ln \tilde{w}_z) \tag{C.26}$$

Taken together, the above procedure generates 7 moment conditions, which we use in the SMM algorithm, which we denote by $\mathbf{M}^{data}$. Specifically, we independently draw J^{sim} set of (a_z, m_z) from the joint log-normal distribution with parameters $(\mu_a, \mu_m, \sigma_a, \sigma_m, \rho)$, one for each location k. Within location, we compute the relative agricultural and manufacturing productivities relative to the village average following equations C.22 and C.23. In this way, we can use our simulated data to construct the moments summarized from equations C.24 to C.26, and we denote it by $\mathbf{M}^{sim}$.

To precisely estimate the key parameters Θ, we first take the differences between observed and simulated moments $\Delta\mathbf{M}(\Theta) = \mathbf{M}^{data} - \mathbf{M}^{sim}$. There are 11 moment conditions as implied above. The true parameter Θ_0 will satisfy $\mathbb{E}[\Delta\mathbf{M}(\Theta_0)] = 0$. So we search for the optimal $\widehat{\Theta}$ to minimize the weighted sum of squares, i.e., $\Delta\mathbf{M}(\Theta)'\mathbf{W}\Delta\mathbf{M}(\Theta)$. Ideally, $\mathbf{W}$ is the positive semi-definite weighting matrix, which is chosen at $\mathbf{W} = \mathbf{V}^{-1}$, where $\mathbf{V}$ is the variance-covariance matrix of the moments $\Delta\mathbf{M}(\Theta)$. However, the true variance-covariance matrix is unobserved, and we use its empirical estimates:

$$\widehat{\mathbf{V}} = \frac{1}{N^{Bstrp}} \sum_{i=1}^{N^{Bstrp}} (\mathbf{M}^{data} - \mathbf{M}_i^{Bstrp}) (\mathbf{M}^{data} - \mathbf{M}_i^{Bstrp})' \tag{C.27}$$

where $\mathbf{M}^{data}$ is moment conditions from the data, and $\mathbf{M}_i^{Bstrp}$ denotes the moments generated from a random sample (sample i) that is drawn with replacement from the original households by location in the dataset. N^{Bstrp} is the number of bootstrap samples. We invert $\widehat{\mathbf{V}}$ using the Moore-Penrose pseudo-inverse. After obtaining the key parameters of the joint log-normal distribution through the above process, we generate a random sample of 10,000 for each location. Then, we discretize the ability distribution of a_z and m_z into $5 \times 5 = 25$ grid points to characterize the ability distributions for each location.

C.7 Background on the Farmland Red Line Policy

Institutional Context of Land Ownership in China Since the founding of the People's Republic of China in 1949, land ownership has been either collectively owned or state-owned. This system was further codified in the 2007 Property Law, which reinforced the separation between land ownership and land–use rights. Foreign entities are not permitted to own

land in China. Instead, land users obtain transferrable land–use rights for a fixed period, depending on the intended purpose: 70 years for residential use, 50 years for industrial use, and 40 years for commercial, tourism, or recreational use.[34] Land–use rights are granted through contracts signed with local land administration departments at the municipal or county level.

Overview of the Farmland Red Line Policy In addition to regulating land transactions and conversion, the Chinese government imposes institutional constraints on land use that directly shape spatial development. Among the most influential of these is the Farmland Red Line Policy, which originated in the 1999 Land Administration Law of the People's Republic of China and was formally codified under the Eleventh Five-Year Plan in 2006 as part of a national strategy to safeguard food security. The policy mandates the preservation of at least 1.8 billion mu (approximately 120 million hectares) of arable land nationwide.

The Red Line is enforced through a top–down quota system. Provincial and local governments are assigned minimum targets for protected farmland and are held accountable for compliance. Converting designated farmland to non-agricultural use requires approval from higher level authorities and is subject to strict oversight and regulation. Violations carry penalties, and enforcement is supported by satellite monitoring and cadastral systems. At the planning level, the Red Line is embedded into official land–use master plans at both the provincial and county levels. It constrains not only the spatial extent of urban expansion but also the location and timing of new development. Since its adoption, the policy has played a central role in land governance and continues to shape patterns of land allocation, investment, and urbanization.

Over time, the Farmland Red Line Policy has emerged as one of the key institutional tools balancing development and conservation, especially in peri-urban and rapidly urbanizing areas. While its primary objective is to protect arable land and ensure national food security, the policy also acts as a binding constraint on local land-use flexibility. An expanding body of literature is exploring the broader economic implications of this policy, including effects on urban form, land markets, and regional development outcomes (Lichtenberg and Ding, 2008; He et al., 2013; Huang et al., 2019; Yu, 2019; Fei et al., 2024).

[34] See Cai et al. (2013) and Chen et al. (2016) for a comprehensive review of China's land market institutions.