Internal Migration Restrictions, Aggregate Productivity, and Spatial Growth

Yanbin (Tracy) Xu

August, 2025

Abstract

China's hukou system ties access to public services and benefits to one's place of registration, creating large frictions to internal migration. This paper develops a dynamic spatial equilibrium model in which productivity, amenities, and migration costs evolve endogenously, allowing us to separate the dispersion force of lower migration barriers from the agglomeration forces of productivity spillovers and amenity investment. Calibrated to detailed prefecture-level data, the model shows that removing hukou barriers nationwide sets off a rapid reallocation of labor, initially toward high-amenity areas, followed by sustained gains as agglomeration effects take hold. The largest improvements in long-run welfare occur when greater mobility is paired with endogenous upgrades to both productivity and amenities, reflecting the complementarities between economic opportunity and urban livability. Partial reforms, in contrast, can yield modest or even negative welfare changes if they concentrate migration in less dynamic regions. The findings highlight that hukoureform's success hinges on the interaction between mobility and place-based growth forces, offering lessons for other economies facing institutional barriers to internal migration.

1 Introduction

Migration decisions are fundamental drivers of economic development, shaping labor supply, productivity, and the spatial distribution of economic activity. The ability of workers to move freely across regions allows for the efficient reallocation of labor to more productive areas, promoting regional convergence and long-term economic growth (Eaton & Kortum, 2002; Desmet, Nagy, & Rossi-Hansberg, 2018; Monte, Redding, & Rossi-Hansberg, 2018). Labor mobility is a well-established mechanism for enhancing aggregate productivity and supporting structural transformation (Moretti, 2012). When workers move to areas where their skills are better utilized, it leads to higher wages, increased productivity, and improved economic outcomes for individuals and regions alike (Lagakos, Marshall, Mobarak, Vernot, & Waugh, 2020; Bryan, Chowdhury, & Mobarak, 2014; Tombe & Zhu, 2019).

However, in practice, migration is often hindered by various frictions that distort labor flows and prevent workers from fully capitalizing on economic opportunities in more productive regions. Financial burdens, such as moving expenses, job search costs, and income risks, can be prohibitive, especially in developing economies where access to credit and safety nets is limited. Social factors, including cultural ties, family connections, and stigma against migrants, further discourage relocation (Bryan & Morten, 2019; Young, 2013). Compounding these challenges, many developing economies impose institutional barriers that explicitly restrict labor mobility, creating systemic obstacles to regional development and economic growth.

One of the most prominent examples of such institutional barriers is China's *hukou* system. Introduced in the 1950s, the *hukou* system, a household registration policy, remains one of the most significant institutional constraints on labor mobility. By tying access to essential public services—such as education, healthcare, and housing—to an individual's place of *hukou* registration, the system creates substantial barriers to migration. Even if individuals relocate for work, they are often unable to access critical services in their destination cities, raising the cost of mobility and exacerbating mismatches between labor supply and demand. This misallocation of labor contributes to regional inequality and dampens national productivity (Chan & Buckingham, 2008; Liu, 2005; Meng, 2012; Au & Henderson, 2006).

Existing studies, such as Chan and Buckingham (2008) and Song (2014), have highlighted how the *hukou* system restricts rural-to-urban migration by limiting the benefits associated with relocation. Using provincial-level data, Tombe and Zhu (2019) find that removing the *hukou* system could reduce inequality across provinces and boost overall welfare, while Roberts, Deichmann, Fingleton, and Shi (2012) and Bosker, Deichmann, and Roberts (2018) argue that relaxing *hukou* restrictions would reinforce urban productivity gains and amplify current core-periphery urbanization pattern. Despite these contributions, much of the literature has focused on static outcomes, such as wage differentials and migration flows, without fully capturing the dynamic interplay between labor mobility, aggregate output, and economic growth over time.

This paper bridges this gap by developing a dynamic spatial equilibrium model that explicitly incorporates the hukou system as an institutional constraint on migration, building on frameworks by Desmet et al. (2018) and Cai, Caliendo, Parro, and Xiang (2022). The model is designed to capture the nuanced interactions between regional differences, migration frictions, and endogenous growth driven by innovation and labor mobility. Migration reshapes regional market sizes, influencing firms' investment in innovation and knowledge accumulation and thus creating feedback loops between migration, productivity, and growth. The *hukou* system is modeled as friction that restricts labor mobility and results in labor misallocation across regions, highlighting how its removal could impact both productivity distribution and long-term growth trajectories.

The model captures heterogeneity across China's prefecture-level cities, where each location has distinct amenities and productivity. Amenities in the model have two main components: an exogenous component, which is fixed, and an endogenous component, which is dynamic and population-driven. The exogenous amenity reflects features that make a region inherently appealing at the initial period, such as natural landscapes, climate, and existing infrastructure. These initial qualities provide each region with an inherent advantage or disadvantage in attracting residents and businesses at the start of the model.

The endogenous component of amenities evolves based on population density, reflecting how amenities improve or strain as more people move into a region. This could mean positive effects, like expanded social infrastructure (e.g., public libraries, schools, and healthcare resources), or negative effects, such as environmental pressures or overcrowding. By structuring amenities this way, the model captures both the initial appeal of a location and the way its attractiveness shifts over time in response to population changes.

The model also incorporates trade across regions, with transportation costs representing the distance and infrastructure quality between areas. These costs capture the challenges of moving goods, influenced by existing transport networks and geographical barriers.

The model is calibrated using prefecture-level data from 1990 to 2000 to estimate productivity and amenity levels and to back out mobility costs and institutional constraints, with model validation using data spanning 2010 to 2020. By conducting a counterfactual analysis, this paper evaluates the effects of *hukou* system reforms on labor mobility, spatial distribution of productivity, and aggregate productivity, output and welfare. In contrast to studies that assume free mobility (Rosen, 1979; Roback, 1982; Allen & Arkolakis, 2014), this model reflects the realities of institutional barriers, where agents make forward-looking migration decisions based on expected economic conditions, mobility costs and idiosyncratic amenity shocks. These findings offer new empirical insights into how the *hukou* system affects labor misallocation and regional inequality, particularly under dynamic economic conditions that evolve over time.

This study makes several key contributions. First, it extends the dynamic spatial equilibrium models developed by Desmet et al. (2018), Redding and Rossi-Hansberg (2017), among others, by explicitly incorporating institutional mobility frictions. Inspired by Tabuchi and Thisse (2002), Monte et al. (2018), Caliendo, Dvorkin, and Parro (2019), Ahlfeldt, Bald, Roth, and Seidel (2020) and Cai et al. (2022), this model integrates forward-looking mobile workers with heterogenous preference. Unlike earlier studies, this paper makes labor mobile but not fully mobile, providing a more realistic framework for studying migration barriers in developing economies.

Second, by incorporating forward-looking migration decisions, this paper offers new empirical insights into how the *hukou* system affects regional inequality and labor misallocation in China. While previous research has focused on static effects, this paper emphasizes the dynamic, long-term impacts on aggregate output, productivity, and welfare.

Third, this study also contributes to the broader literature on internal migration and mobility constraints in developing countries. Research on internal migration highlights how mobility shapes economic landscapes by redistributing labor across regions, often in response to economic disparities (Stark & Bloom, 1985; De Haas, 2010; Morten & Oliveira, 2018; Banerjee & Duflo, 2007). The counterfactual analysis demonstrates how relaxing *hukou* restrictions could lead to significant improvements in labor mobility, aggregate productivity, and balanced regional development. These findings offer policy-relevant insights into the economic benefits of migration policy reforms.

The rest of the paper is structured as follows. Section 2 introduces *hukou* policy and presents internal migration patterns under this constraint. Section 3 presents the dynamic spatial model. Section 4 provides empirical evidence and discusses the data used for model quantification and calibration. Section 5 presents quantitative results. Section 7 concludes, offering policy implications and directions for future research. All proofs and more detailed data descriptions are provided in the Appendix.

2 Institutional Context and Migration Constraints

2.1 *Hukou* System and Migration Barriers in China

The *hukou* system, officially established in 1958 under the People's Republic of China *Hukou* Registration Regulation, has served as a central pillar of China's demographic and labor market management for decades. The system was initially designed as a population control tool to prevent large-scale rural-to-urban migration during China's early stages of industrialization. During this period, the country faced a delicate balance between building an industrial economy and maintaining sufficient agricultural output. Fearing that unchecked migration to urban areas would lead to overcrowding, the collapse of public services, and food shortages in the countryside, the *hukou* system was introduced to strictly regulate internal migration. These restrictions are especially binding for those who migrate without obtaining a local *hukou*, as access to such services is typically limited to the area of registration. This system creates significant institutional barriers that distort labor mobility across locations.

Over two decades after the implementation of the *hukou* policy, nearly all internal movements of people in China were subject to state regulation or sponsorship (Young, 2013). The constraints on people's mobility persisted until the introduction of the Reform and Open-up Policy in 1979. This policy delegated more authority to local governments regarding setting quotas and eligibility criteria for individuals who seek to obtain *hukou*. Furthermore, workers were allowed to obtain temporary residency permits in the locations where they were employed. Concurrently, the proliferation of manufacturing factories in urban areas escalated the demand for labor in these regions.

Higher wages and better prospects in urban areas continued to attract a significant number of people, leading to a substantial influx of workers migrating from rural or inner-land areas to the more industrialized urban regions along the east coast of China. However, by creating a distinct division between rural and urban residents, *hukou* system contributes to disparities in access to education, healthcare, and social services, which, in turn, affect employment opportunities and wages.

Although reforms such as temporary residency permits have provided some flexibility, they have not fully eliminated the barriers imposed by the *hukou* system. Migrants with temporary permits continue to be excluded from public benefits like education and healthcare in their destination cities, preventing them from fully integrating into urban life. As a result, the system does not just restrict movement between rural and urban areas but also limits migration across urban centers, ultimately affecting the distribution of economic activities and growth throughout the country.

Moreover, the *hukou* system's constraints are not static; they evolve alongside broader economic and demographic changes, influencing future labor market decisions. As China's economies continue to grow, the system's rigid constraints on migration result in long-term consequences for labor allocation, productivity, and regional development. Over time, these barriers create cumulative negative effects, preventing efficient labor reallocation and constraining the country's overall economic trajectory. This dynamic interaction between migration constraints and long-term economic outcomes motivates the need for a model that captures how migration decisions evolve and impact regional productivity over time.

2.2 Migration Patterns and Economic Landscape Under the *Hukou* System

China has experienced massive internal migration over the past few decades, with millions of workers moving from less developed regions (primarily in the west and central areas) to more industrialized urban centers along the eastern coast. By 2020, nearly

380 million people, or 27% of the population, were engaged in internal migration. This large-scale movement of labor has been crucial to China's rapid economic growth and urbanization. However, the hukou system continues to constrain where migrants can settle permanently, limiting the broader benefits that fluid labor mobility would otherwise offer both individuals and the economy.

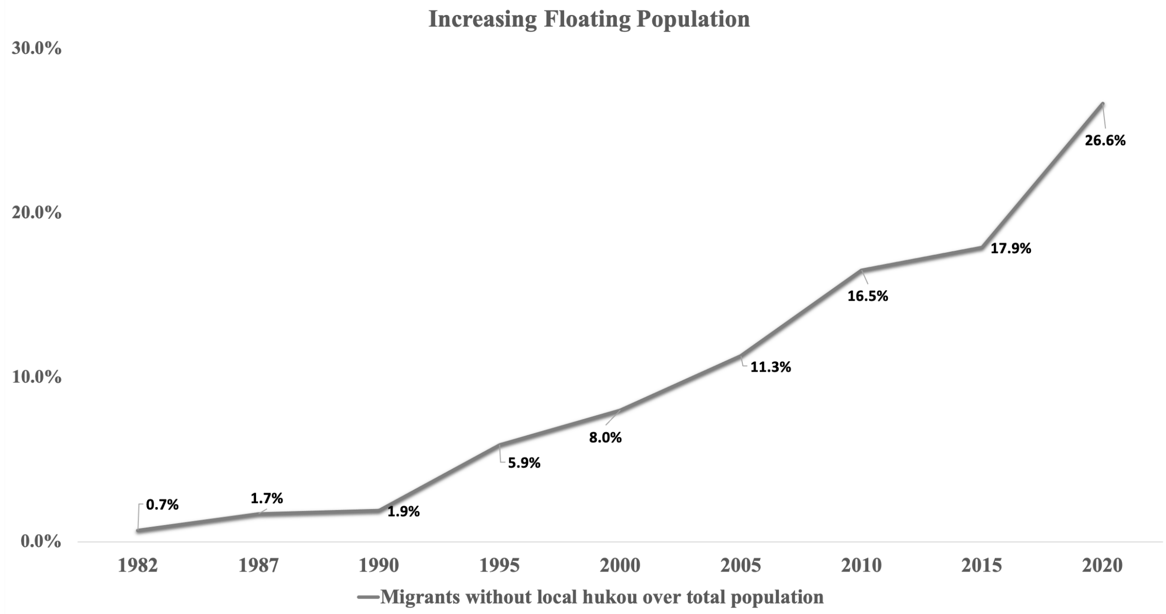


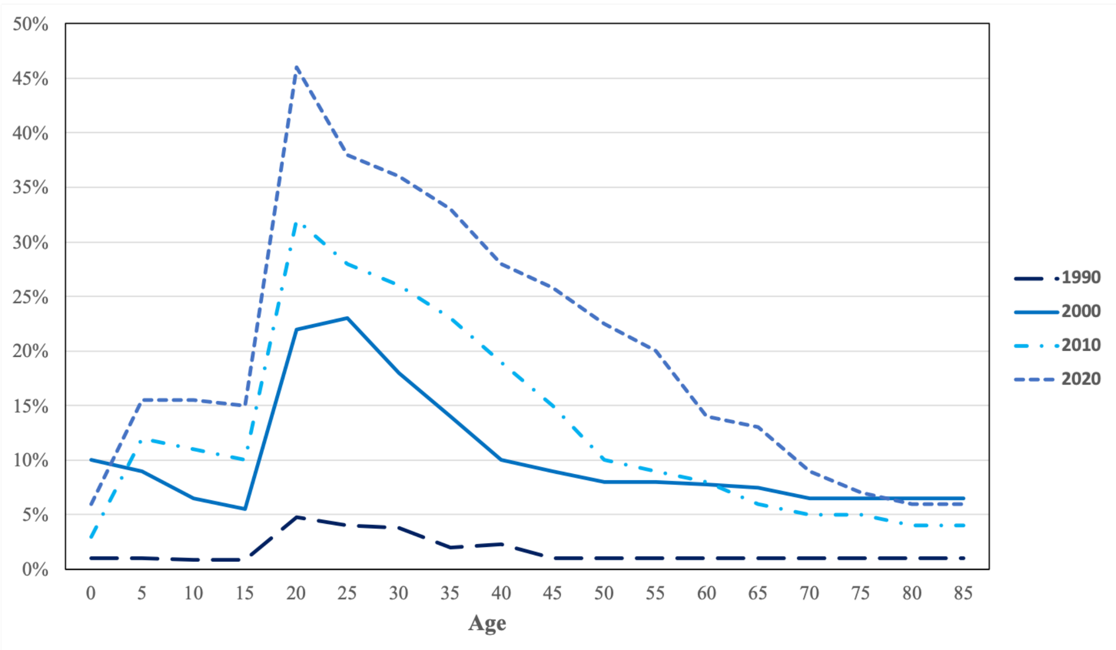
Figure 1: Floating Population

Major cities like Shanghai, Beijing, and Shenzhen have become magnets for migrants due to their economic vitality, while less-developed inland regions have seen a significant outflow of young, productive workers. This migration pattern is not just geographic but also deeply shaped by the institutional barriers of the *hukou* system. Many migrants work in informal, low-wage jobs, and because they lack local *hukou*, they cannot access the public services available to urban residents. According to the 2017 National Migrant Population Dynamic Monitoring Survey, 96% of migrant workers earned less than 900 US dollars per month, falling well below urban wage standards. These constraints reinforce income inequality and the socioeconomic stratification perpetuated by the hukou system.

2.2.1 Who Are Moving and Why?

Internal migration in China is driven by young and middle-aged workers, often referred to as the "floating population," who move in search of better economic opportunities in urban centers. Most migrants are employed in labor-intensive sectors such as construction, manufacturing, and services, but despite their critical role in the urban economy, they remain excluded from the benefits of urban life due to institutional barriers imposed by the *hukou* system.

These migrants typically have lower levels of formal education than their urban counterparts, with the majority holding only junior secondary qualifications (2017 National Migrant Population Dynamic Monitoring Survey). Although education levels among migrants have improved due to broader access to schooling, they remain concentrated in low-paying jobs with limited access to social insurance or benefits, widening the wage gap between them and local urban residents.



Source: Census 1990, 2000, 2010, 2020

Figure 2: Floating Labor

The primary motivation for migration is the search for better employment opportunities, as less-developed areas offer limited job prospects beyond agriculture. Figure 2 highlights that the vast majority of migrants move to urban centers in search of higher wages and more diverse employment prospects. However, despite these opportunities, the *hukou* system restricts their ability to settle permanently and access public services in their destination cities. As a result, many migrants are forced into temporary or circular migration arrangements, where they contribute labor to economic centers but remain socially and economically marginalized.

2.2.2 Inflow and Outflow Cities and Regional Disparities

Migration patterns in China reflect the country's broader regional development landscape. Coastal cities like Shanghai, Guangzhou, and Shenzhen are major inflow destinations due to their higher wages, economic opportunities, and advanced infrastructure. These cities attract significant numbers of migrants, but the concentration of labor also puts pressure on urban infrastructure and public services.

Meanwhile, rural provinces such as Henan, Anhui, and Sichuan have become major outflow regions, losing large portions of their young workforce to coastal areas. This exodus of labor has left these regions with economic challenges, including labor shortages and an aging population. These areas rely heavily on remittances sent by migrants to support local households and economies, underscoring the critical role of internal migration in livelihoods.

Despite these migration flows, regional disparities have widened. Urban centers benefit from labor inflows, but institutional barriers prevent migrants from fully integrating into these cities, limiting their upward mobility and exacerbating income inequality. Meanwhile, outflow regions experience long-term economic stagnation as the loss of human capital weakens their growth potential.

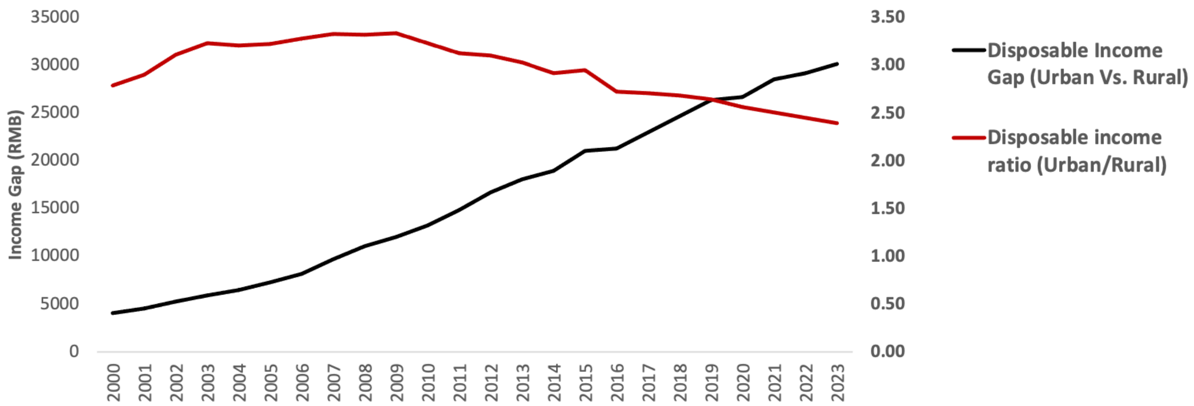


Figure 3: Urban and Rural Income Gap

The *hukou* system's restrictions on migration do not simply affect current flows; they have long-term implications for China's economic development. Coastal cities, which continue to absorb large numbers of migrants, face growing pressure on infrastructure and public services, while inland areas experience the loss of labor and human capital.

Over time, the inability to freely move and settle in high-productivity regions leads to inefficiencies that accumulate, resulting in both regional stagnation and lost economic opportunities. This highlights the dynamic nature of labor mobility and the need for models that can simulate how migration restrictions shape economic outcomes over time.

By capturing the interplay between migration flows, labor market outcomes, and regional development, this analysis provides a foundation for understanding how the *hukou* contributes to both short-term and long-term economic disparities across China. The next section will develop a dynamic spatial model to quantify these effects and explore how reforming the *hukou* system could affect labor mobility, aggregate productivity, and spatial growth.

3 The Model

This section presents a dynamic spatial equilibrium model to quantify the effects of institutional migration constraints, specifically the *hukou* system, on labor mobility, regional productivity, and aggregate economic outcomes.

We consider a closed economy with N regions indexed by $n = 1, 2, ..., N$, each endowed with a fixed supply of land $H_n > 0$, which remains constant over time. The economy is populated by a mass $\bar{L} = \sum_h L^h$ of agents, each possessing a *hukou* status h, which indicates their registered location and is time-invariant.[1] Each agent is endowed with one unit of labor, supplied inelastically at their chosen residence. Agents make location decisions based on wages, amenities, and migration costs, which include *hukou* imposed restrictions. Each agent's location choice determines the distribution of labor across regions, influencing both local and aggregate productivity. Goods are traded between locations, subject to symmetric iceberg transportation cost τ_{nj} between location n and j.

The model assumes a closed economy among Chinese prefectures, meaning that trade and migration occur only across domestic regions. This simplifying assumption allows me to focus squarely on the internal allocation of labor and goods, and to isolate the effects of institutional frictions such as *hukou* and internal trade costs on regional development.

[1]The assumption of a fixed *hukou* status, while potentially stringent, enhances the model's traceability in large-scale simulations involving multiple locations and time periods, given the minimal rate of *hukou* status changes.

In reality, each prefecture engages in trade with the rest of the world. Rather than structurally modeling international markets, I incorporate the effect of openness directly in the data by adjusting each region's expenditure to reflect net exports. Specifically, I deduct net exports (exports minus imports) from GDP when calibrating local spending. This approach preserves the tractability of a closed economy model while allowing observed expenditure to account for international trade, consistent with the practice in recent spatial equilibrium models such as Tombe and Zhu (2019)

3.1 Preferences

The economy consists of a continuum of heterogeneous, forward-looking agents. An agent's utility is derived from the consumption of differentiated goods and the enjoyment of local amenities. Agents consume their entire income each period with no savings, and thus, wealth accumulation is not considered. However, they actively observe economic conditions and make location decisions to maximize their lifetime utility, considering idiosyncratic taste shocks, mobility costs, and amenity losses associated with their *hukou* status. For an individual i with *hukou* status h, migrating from location j and residing in location n at time t, utility is given by:

$$u_t^{ih}(n) = \frac{A_t(n)C_t(n)\varepsilon_t^i(n)}{\phi(h,n)m(j,n)}, \tag{1}$$

where $A_t(n)$ is the local amenities at location n at time t; $C_t(n)$ is the consumption bundle or consumption index at n at time t; $\varepsilon_t^i(n)$ denotes idiosyncratic taste shocks, which are i.i.d. across locations, time, and individuals, following a Fréchet distribution; $\phi(h,n)$ captures the amenity loss for an individual with *hukou* status h living in n;[2] $m(j,n)$ represents the time-invariant mobility cost of moving from region j to n.

3.1.1 Locational Amenity

Amenities in each location n consist of three components: one exogenous and two endogenous. The exogenous component, denoted by $\bar{A}(n)$, is determined by fixed geographical

[2]We should think about amenity lost due to *hukou* as a product of all amenities lost over time. However, this will require knowing the full history of an agent's migration, for which the data in the Census cannot provide sufficient information. Moreover, it will make the model very complicated. So, in this paper, I only consider the period amenities lost rather than the whole series of amenities lost.

characteristics and remains constant over time. The first endogenous component captures the effects of local population density on amenities, represented by $\bar{l}_t(n)^\lambda$, where $\bar{l}_t(n)$ is the local population density at time t and λ denotes the elasticity of amenity quality with respect to population density. This term reflects the idea that while higher density can improve the provision and diversity of amenities (e.g., cultural venues, retail, services), it may also lead to congestion or pollution, which we examine in greater detail in Section **??**.

The second endogenous component stems from fiscal investments in local public goods, such as infrastructure and environmental quality. We assume these investments are financed by land rents. Specifically, the per capita investment is proportional to the total land rent revenue $R_t(n)H(n)$ divided by the resident population $\bar{L}_t(n)$, scaled by an elasticity parameter χ. Altogether, the amenity level in location n at time t is given by: $A_t(n) = \bar{A}(n)(\frac{R_t(n)H(n)}{\bar{L}_t(n)})^\chi \bar{l}_t(n)^\lambda$.

We assume that the revenue from land rents in each location is fully reinvested into local amenities. While we do not explicitly model a government, we interpret this reinvestment as occurring indirectly via the purchase of local consumption goods that improve amenities (e.g., infrastructure, sanitation, public goods). [3]

3.1.2 *hukou* Parameter

The parameter $\phi(h, n)$ represents the utility loss experienced by an individual due to their *hukou* status h when residing outside their registered location. If h matches location n (i.e., $\phi(h, h) = 1$), the individual experiences no utility loss, enjoying the same benefits as local residents. However, if h differs from n ($\phi(h, n) \neq 1$), the individual may face restricted access to amenities, reflecting the barriers imposed by the *hukou* system.

This parameter $\phi(h, n)$ is location-specific and asymmetric but remains constant for individuals over time. Although the *hukou* status can theoretically change over time due to factors like job changes, marriage, or meeting specific local criteria, this parameter is typically assumed to be invariant for individuals, especially given the historically low rates of *hukou* changes before 2010. This assumption simplifies the model by allowing for a consistent measure of amenity loss across different time periods and locations,

[3] This assumption ensures that total expenditures match incomes, maintaining general equilibrium consistency without introducing an additional agent.

facilitating the analysis of spatial dynamics without the complexity of tracking frequent status changes.

For instance, an individual with a better educational background might have easier access to certain amenities or opportunities, such as employment in state-owned enterprises that offer more generous *hukou* quotas. However, these individual advantages do not alter the overall assumption that the utility loss parameter $\phi(n, h)$ remains constant for each person over time, as these changes are relatively rare and do not significantly impact the broader population's mobility decisions.

3.1.3 Consumption

Agents derive utility from consuming a continuum of differentiated goods, each of which is indexed by $\omega \in [0, 1]$. The quantity of good ω consumed by an agent at location n at time t is denoted by $c_t^\omega(n)$. The elasticity of substitution between these goods is represented by $\sigma > 1$, where a higher value of σ indicates that the goods are more easily substitutable for one another. The overall consumption bundle at location n at time t is then aggregated using the following CES (constant elasticity of substitution) function:

$$C_t(n) = \left[\int_0^1 c_t^\omega(n)^{\frac{\sigma-1}{\sigma}} d\omega\right]^{\frac{\sigma}{\sigma-1}}.$$

Agents supply one unit of labor inelastically at their chosen location, earning a wage $w_t(n)$ which they entirely spend on consumption at the local price of good $P_t(n)$.[4] Consequently, the utility for agent i with *hukou* h in location n is:

$$u_t^{ih}(n) = \frac{A_t(n)}{\phi(h, n)m(j, n)} \frac{w_t(n)}{P_t(n)} \varepsilon_t^i(n), \tag{2}$$

where $P_t(n)$ is the price index at location n at time t, given by:

$$P_t(n) = \left[\int_0^1 p_t^\omega(n)^{-(\sigma-1)} d\omega\right]^{-\frac{1}{\sigma-1}},$$

with $p_t^\omega(n)$ representing the price of good ω.

The idiosyncratic preference shock $\varepsilon_t^i(n)$ follows a Fréchet distribution:

$$\Pr\left[\varepsilon_t^i(n) \leq z\right] = e^{-z^{-\gamma}}, \gamma > 1, \tag{3}$$

[4]For simplicity, this model does not include income taxes, though it could be extended to account for them.

where γ is the shape parameter governing the dispersion of the amenity preference. Lower values of γ indicate greater heterogeneity in tastes.

We now describe the dynamic labor supply decisions made by migrants as they choose between different locations in the model.

3.2 Labor Mobility

At the end of each period, workers observe the economic conditions and realize idiosyncratic amenity shocks. Based on this information, at the beginning of the next period, they decide where to relocate to optimize their present discounted value of utility, subject to institutional constraints and mobility costs. Each worker supplies one unit of labor inelastically at their chosen location and earns a wage $w_t(\cdot)$. They spend their entire income on goods at the local price $P_t(\cdot)$.

The value function of a worker i with *hukou* status h residing in region n at time t, who considers moving to region s at time $t+1$, is expressed as:

$$\begin{aligned} V_t^{ih}(n) &= \max_s \left[u_t^{ih}(n) + \beta \mathbb{E}\left(V_{t+1}^{ih}(s)\right)\right] \\ &= \max_s \frac{A_t(n)}{\phi(h,n)m(j,n)} \frac{w_t(n)}{P_t(n)} \varepsilon_t^i(n) + \beta \mathbb{E}[V_{t+1}^{ih}(s)], \end{aligned} \tag{4}$$

where β is the exogenous discount factor.

The probability that an agent with *hukou* status h will relocate from location n (at time t) to location s in period $t+1$ is given by:

$$\mathbb{P}\left[V_{t+1}^{ih}(s) \geq \max_{s \neq n} V_{t+1}^{ih}(n) | V_t^{ih}(n) \geq \max_{j \neq n} V_t^{ih}(j)\right].$$

Given that the idiosyncratic amenity shocks $\varepsilon_t^i(n)$ are i.i.d. and follow a Fréchet distribution, we can substitute (4) and integrate over $\varepsilon_t^i(n)$ and obtain the probability of individuals with *hukou* h and relocate from location n to location s at time t

$$\Omega_t(s,n)^h = \frac{\frac{A_t(s)w_t(s)}{\phi(h,s)P_t(s)}^{\gamma} m(n,s)^{-\gamma}}{\sum_{k=1}^{N} \frac{A_t(k)w_t(k)}{\phi(h,k)P_t(k)}^{\gamma} m(n,k)^{-\gamma}}. \tag{5}$$

Since forward-looking individuals decide where to live and supply labor in the future by evaluating the relative net future value of each location, the probability of the individual moving to a location depends on the net value she obtained from one location relative to all locations. The presence of migration costs and idiosyncratic shocks leads to a gradual adjustment of labor supply in response to changes in the economic environment. If the proportion of workers living in location n at time $t-1$ with *hukou* status h is $\mu_{t-1}(n)^h$, then the ratio of these individuals moving to location s at time t relative to the total labor population is $\mu_{t-1}(n)^h\Omega_t(n,s)^h$. Total labor population with *hukou* h at time t in location s is thus the sum of net labor inflows to s with *hukou* h and labor with*hukou* h staying at s:

$$\begin{aligned} H(s)\bar{l}_t(s)\mu_t(s)^h &= \sum_n^N \mu_{t-1}(n)^h\Omega_t(n,s)^h H(n)\bar{l}_{t-1}(n) \\ &= \Omega_t(n,s)^h \sum_n^N \mu_{t-1}(n)^h H(n)\bar{l}_{t-1}(n) \\ &= \Omega_t(n,s)^h \bar{L}^h_{t-1}, \end{aligned} \tag{6}$$

where $\bar{L}^h_{t-1}$ represents the total population with *hukou* h in the country at time $t-1$.

3.3 Absent Landlord

In this model, land is owned by immobile landlords who do not participate in the labor market. These landlords earn income solely from renting land to local firms. The rental income per unit of land at time t in location n denoted as $R_t(n)$, is collected and used exclusively for local public investment.

Unlike previous formulations where landlords consume goods locally, in this framework the total rental income $R_t(n)H(n)$ is not directed toward private consumption. Instead, it is fully allocated to the provision of local amenities through fiscal investment by local governments. These investments enhance local infrastructure, public services, environmental quality, and other dimensions of location-specific amenity value.

This assumption ensures that land rents play a direct role in shaping spatial equilibrium by improving local amenities, which in turn influence migration decisions and labor allocation. By excluding landlord consumption, the model eliminates the need to track landlord preferences or spending patterns, thereby simplifying the economic environment while preserving the fiscal significance of land rents.

3.4 Production, Innovation, and Growth

This section outlines the structure of production at both the firm and regional levels, incorporating firm-level heterogeneity, regional agglomeration dynamics, and endogenous productivity evolution. We adopt a tractable framework where regional productivity evolves over time due to path dependence and local labor concentration, while firm-level heterogeneity arises from idiosyncratic productivity draws. This dual structure allows us to connect micro-level firm decisions to macroeconomic aggregates in a spatial and dynamic context.

3.4.1 Firm-Level Production

At each location n, a continuum of immobile firms indexed by $\omega \in [0,1]$ produce differentiated varieties of goods using labor and land as inputs. The production function for firm ω in region n at time t is specified as:

$$y_t^\omega(n) = Z_t^\omega(n) H_t^\omega(n)^{(1-\iota)} L_t^\omega(n)^\iota, \tag{7}$$

where:

- $H_t^\omega(n)$ is land input,
- $L_t^\omega(n)$ is labor input,
- $\iota \in (0,1)$ is the elasticity of output with respect to labor or labor share in production,
- $Z_t^\omega(n)$ denotes firm-level total factor productivity (TFP).

To simplify, we normalize by land input and express production per unit of land:

$$q_t^\omega(n) = z_t^\omega(n) l_t^\omega(n)^\iota \tag{8}$$

where $z_t^\omega(n) \equiv Z_t^\omega(n)$, and $l_t^\omega(n) = L_t^\omega(n)/_t^\omega(n)$ denotes labor per unit of land. This formulation captures decreasing returns to scale in labor, given fixed land supply.

3.4.2 Firm-Level Productivity and Heterogeneity

We follow the approach of Eaton and Kortum (2002) and Tombe and Zhu (2019), modeling firm-level productivity as:

$$Z_t^{\omega}(n) = Z_t(n)\epsilon_t^{\omega}(n),$$

Here $Z_t(n)$ denotes the region-specific average productivity level in region n, and $\epsilon_t^{\omega}(n)$ is an idiosyncratic firm-specific shock, drawn independently across firms and time from a Fréchet distribution:

$$F(z) = exp(-T(n)\epsilon^{-\delta}),$$

where $-T(n)$ is a location-specific scale parameter (reflecting the local technological potential) and the shape parameter $\delta > 1$ governs the dispersion of firm productivity draws.

This setup enables the evolution of a firm's productivity driven by local knowledge stock $Z_t(n)$ with uncertainty. The Frchet specification implies that a few firms in each location will have extremely high productivity draws, allowing them to dominate trade flows. It also ensures that key aggregate outcomes, such as trade shares and price indices, can be expressed in closed form, which is essential for tractability in dynamic spatial models.

3.4.3 Regional Productivity Dynamics

While firm-level productivity is heterogeneous and static within a period, regional productivity evolves endogenously over time. The evolution $Z_t(n)$, the mean productivity of firms in location n, is given by:

$$Z_{t+1}(n) = Z_t(n)^{\alpha}\bar{l}_t(n)^{\lambda_1}\epsilon_{t+1}(n), \tag{9}$$

where:

- $\alpha \in (0, 1)$ captures the persistence of past productivity or path dependence,
- $\bar{l}_t(n)$ is labor density in location n,
- $\lambda_1 > 0$ captures the strength of the agglomeration effects, reflecting the productivity gains from the concentration of labor in a particular region,
- $\epsilon_{t+1}(n)$ is a location-specific stochastic shock, i.i.d. over time and space.

This expression formalizes the notion that regions with denser economic activity (higher labor per land) experience faster technological progress due to learning, knowledge spillovers, or scale economies. The multiplicative shock $\epsilon_t(n)$ introduces randomness to reflect unforeseen local innovations or disruptions.

More importantly, we do not assume endogenous innovation at the firm level in this model. Instead, firms are passive recipients of location-specific productivity growth, which reflects aggregate local forces rather than micro-level investment decisions. This choice allows us to simplify the model while retaining endogenous spatial dynamics.

3.4.4 Firm Behavior and Profit Maximization

Each firm ω in region n maximizes profits by choosing labor inputs, taking the locational wage $w_t(n)$, rent $R_t(n)$ and its own productivity draw $\epsilon_t^\omega(n)$ as given. Firms do not invest in innovation. Instead, their productivity is $Z_t^\omega(n) = Z_t(n)\epsilon_t^\omega(n)$ as discussed in section 3.4.2.

These firms operate under Bertrand competition and face constant elasticity of substitution (CES) demand from consumers. After drawing an idiosyncratic productivity shock $\epsilon_t^\omega(n)$, each firm takes local factor prices–wages $w_t(n)$ and land rents $R_t(n)$–as given and chooses its labor input to maximize profit per unit of land:

$$\max_{l_t^\omega(n)} p_t^\omega(n,n) z_t^\omega(n) l_t^\omega(n)^\iota - w_t^\omega(n) l_t^\omega(n) - R_t^\omega(n),$$

where:

- $p_t^\omega(n,n)$ is the price charged in the local market,
- $w_t^\omega(n)$ is the wage per worker,
- $R_t^\omega(n)$ is the rental price per unit of land.

Note that land is in fixed supply and normalized at the firm level. All costs and revenues are expressed per unit of land.

To solve the firm's problem, we take the first-order condition with respect to labor $l_t^\omega(n)$:

$$\iota p_t^\omega(n,n) z_t^\omega(n) l_t^\omega(n)^{\iota-1} = w_t^\omega(n). \tag{10}$$

This condition equates the marginal revenue product of labor to the local wage rate. It characterizes optimal labor demand at the firm level as a function of prices, productivity, and wages.

In equilibrium, the land rental price across different firms producing various goods within a location is expected to equalize. If wage or rent disparities exist among firms within a particular area, workers and land will be reallocated until these prices are uniform. As a result, we can simplify the wage and rent terms $w_t(n)$ and $R_t(n)$, respectively, representing the uniform wage and rent for every firm in location n at time t. Moreover, in equilibrium, entry and exit of firms drive profits to zero (net of land costs). Therefore, the firm's profits are exactly equal to the cost of land rent.

$$\pi_t^\omega(n) = p_t^\omega(n,n) z_t^\omega(n) l_t^\omega(n)^\iota - w_t(n) l_t^\omega(n) - R_t(n) = 0. \tag{11}$$

Rearranging 11 and applying FOC give an expression for the equilibrium land rent paid by firm ω in region n at time t:

$$\begin{aligned} R_t(n) &= p_t^\omega(n,n) z_t^\omega(n) l_t^\omega(n)^\iota - w_t(n) l_t^\omega(n) \\ &= \frac{w_t(n) l_t^\omega(n)}{l} - w_t(n) l_t^\omega(n) \\ &= \frac{1-l}{l} w_t(n) l_t^\omega(n). \end{aligned} \tag{12}$$

Similarly, the firm's labor hiring decision can also be expressed in terms of local rent:

$$l_t^\omega(n) = \frac{\iota R_t(n)}{(1-\iota) w_t(n)}. \tag{13}$$

These formulas imply that the firm's land rent is proportional to its labor cost and vice versa. Since the rent will be equalized across different firms in equilibrium, this tells us uniformity in firm behavior in each location n. As a result, the analysis can be simplified by omitting the superscript ω for $l_t(n)$.

3.5 Goods Trade

In this model, goods are tradable across regions subject to iceberg transportation costs. For a unit of a good shipped from origin n to destination s, only $\frac{1}{\upsilon(s,n)}$ units arrive, where the transportation cost $\upsilon(s,n) \geq 1$. The iceberg cost is symmetric, i.e. $\upsilon(s,n) = \upsilon(n,s)$, and remains constant over time to isolate the effects of other endogenous dynamics.

The price of goods produced and consumed locally can be derived from the production function. The local price of a good ω produced and consumed in region n at time t is given by:

$$p_t^\omega(n,n) = \iota^{-1} w_t(n) z_t(n)^{-1} l_t(n)^{1-l}. \tag{14}$$

When goods are shipped to locations other than where they were produced, the price of these imported goods, as consumed in the destination location s, is given by:

$$\begin{aligned} p_t^\omega(s,n) &= \upsilon(s,n)\iota^{-1} w_t(n) z_t(n)^{-1} l_t(n)^{1-l} \\ &= \frac{w_t(n)\upsilon(s,n)}{\iota Z_t(n)\epsilon_t^\omega(n) l_t(n)^{\iota-1}}. \end{aligned} \tag{15}$$

In this economy, a continuum of firms across different locations produces each variety ω. Consumers in any given location s will purchase the good from the firm offering the lowest price, whether the good is produced locally or imported. Following the framework by Eaton and Kortum (2002), the cumulative distribution function (CDF) of the prices for a good ω produced in location n and consumed in location s is:

$$\mathbb{P}\left[p_t^\omega(s,n) \le p\right] = \mathbb{P}\left[\epsilon_t^\omega(n) \ge \frac{w_t(n)\upsilon(s,n)}{\iota Z_t(n) l_t(n)^{\iota-1} p}\right]. \tag{16}$$

Plug (16) into the Fréchet distribution, we will have:

$$\begin{aligned} \mathbb{P}\left[p_t^\omega(s,n) \le p\right] &= 1 - exp\left\{-T(n)\left(\frac{w_t(n)\upsilon(s,n)}{\iota Z_t(n) l_t(n)^{\iota-1} p)}\right)^{-\delta}\right\} \\ &= 1 - exp\left\{-T(n)\left(\frac{\upsilon(s,n)\xi_t(n)}{p}\right)^{-\delta}\right\}, \end{aligned} \tag{17}$$

where $\xi_t(n) = \frac{w_t(n)}{\iota Z_t(n) l_t(n)^{\iota-1}}$.

Plug (17) into the price index, and we can extend the price index in location s at time t as:

$$\begin{aligned} P_t(s) &= \left[\int_0^1 p_t^\omega(s)^{-(\sigma-1)} d\omega\right]^{-\frac{1}{\sigma-1}} \\ &= \left[\int_0^\infty p^{-(\sigma-1)} \cdot \delta p^{\delta-1} \cdot \exp\left\{-\Phi_t(s) p^\delta\right\} dp\right]^{-\frac{1}{\sigma-1}} \\ &= \kappa\left[\sum_{n=1}^N T(n)(\xi_t(n)\upsilon(s,n))^{-\delta}\right]^{-1/\delta}, \end{aligned} \tag{18}$$

where $\Phi_t(s) = \sum_{n=1}^{N} T(n) \left(\xi_t(n)\upsilon(s,n)\right)^{-\delta}$ and $\kappa = \left[\Gamma(\frac{1-\sigma}{\delta}+1)\right]^{\frac{1}{1-\sigma}}$.

The fraction of goods produced in location n and consumed in location s, denoted as $\pi_t(s,n)$, can be derived as:

$$\begin{aligned}
\pi_t(s,n) &= \int_0^{\infty} \prod_{j\in N} exp\left\{(\frac{\upsilon(s,j)\xi_t(j)}{p})^{-\delta}\right\} exp\left\{(\frac{\upsilon(s,n)\xi_t(n)}{p})^{-\delta}\right\} \left(\upsilon(s,n)\xi_t(n)\right)^{-\delta} dp^{\delta} \\
&= \left(\upsilon(s,n)\xi_t(n)\right)^{-\delta} \int_0^{\infty} \exp\left\{-\sum_{j}^{N} \left(\upsilon(s,j)\xi_t(j)\right)^{-\delta} p^{\delta}\right\} dp^{\delta} \\
&= \frac{\left(\upsilon(s,n)\xi_t(n)\right)^{-\delta}}{\sum_{j=1}^{N} \left(\upsilon(s,j)\xi_t(j)\right)^{-\delta}}
\end{aligned} \tag{19}$$

Given the constant elasticity of substitution in consumer preference, this fraction $\pi_t(s,n)$ also represents the expenditure of consumers in location s spending on goods imported from location n at time t.

We assume no trade surplus or deficit. Firms in each location generate revenue from all the goods they sell locally and across locations. Meanwhile, consumers spend their entire income on goods, whether produced locally or imported. This trade balance assumption means that the total expenditure in any location equals its total income.

The total income in n is the sum of labor and land income:

$$Y_t(n) = w_t(n)\bar{l}_t(n)H(n) + R_t(n)H(n).$$

We impose a trade balance condition:

$$Y_t(n) = \sum_{s=1}^{N} \pi_t(s,n)Y_t(s). \tag{20}$$

By substituting the relevant expressions for labor (12) and rent (13) from previous sections, this equation simplifies to:

$$w_t(n)\bar{L}_t(n) = \sum_{s=1}^{N} \pi_t(s,n)w_t(s)\bar{L}_t(s). \tag{21}$$

3.6 Equilibrium

This section defines the conditions under which a dynamic equilibrium is achieved in the model. The economy consists of a finite set of locations where workers, firms, and land interact over time. Locations are connected through goods trade and labor migration, and productivity evolves endogenously as a function of local economic activity.

3.6.1 Definition of Equilibrium

Given:

- An initial distribution of labor $\{L_0(n)\}_{n=1}^N$, land endowments $\{H(n)\}_{n=1}^N$, and amenity $\{A_0(n)\}_{n=1}^N$;
- A set of bilateral trade cost $\{\upsilon(n,j)\}_{n=1,j=1}^{N,N}$, mobility cost $\{m_(n)\}_{n=1}^N$, and amenity cost $\{\phi(n,j)\}_{n=1,j=1}^{N,N}$;
- And model parameters governing preferences, production, productivity, amenity, and mobility $(\beta, \gamma, \sigma, \iota, \lambda_1, \delta, \alpha, \chi, \lambda)$

The sequential competitive equilibrium of this dynamic spatial model is a sequence of factor prices, goods prices, values and labor distribution $\{w_t(n), R_t(n), P_t(n), V_t(n,h), L_t(n)\}_{n=1,t=0}^{N,\infty}$ such that, at every time period t:

1) Firms maximize profits: Firms in each region choose labor inputs to maximize profits, taking wages, land rents, and their productivity draws as given. The first-order conditions for labor demand hold, and free entry ensures that profits (net of land rent) are zero in equilibrium.

2) Goods Market Clearing and Trade Balance: Each region's total income from wages and land rents equals its total expenditure on goods. Equivalently, trade is balanced in each location: the value of exports equals the value of imports.

3) Land market clearing: Land is in fixed supply and fully used by local firms. The land rental price adjusts to equate demand with the available land in each location.

4) Labor mobility: Mobile workers choose where to live based on wages, amenities, and migration costs. Their location choice follows a probabilistic rule derived from the random utility framework, incorporating idiosyncratic preferences.

5) Labor market clearing:
$$\sum_{s}^{N} H(s)\bar{l}_t(s) = \bar{L}_t \tag{22}$$

6) Productivity dynamics: Regional productivity evolves over time as a function of past productivity, labor density (reflecting agglomeration), and location-specific shocks.

7) Amenity Determination: Amenity levels in each region evolve based on exogenous baseline amenities, endogenous population density effects, and fiscal investment from local land rents.

3.6.2 Existence and Uniqueness of Equilibrium

To ensure that an equilibrium exists and is unique, we follow the analytical approaches of Allen and Arkolakis (2014) and Desmet and Rossi-Hansberg (2014). In particular, the uniqueness of equilibrium of the model depends on the relative strength of agglomeration forces (which attract labor to dense regions) and dispersion forces (which push labor away due to congestion and diminishing returns).

The model yields a unique static equilibrium at each time period if the following condition holds: The sum of agglomeration elasticities from productivity spillovers, amenity investment, and population density effects must not exceed the sum of dispersion forces from production congestion, amenity crowding, migration frictions, and trade substitutability.

Mathematically, this is expressed as:

$$\frac{\lambda_1}{\delta} + \chi < |\lambda| + (1 - \iota) + \frac{1}{\gamma},$$

where

- $\frac{\lambda_1}{\delta}$: strength of productivity agglomeration from labor density;
- χ: elasticity of amenities to rent-driven investment;
- λ: elasticity fo amenities to labor density;
- $(1 - \iota)$: dispersion force from decreasing returns to labor in production;
- $1/\gamma$: migration dispersion due to idiosyncratic preference shocks.

If this inequality is satisfied in every period, then the equilibrium allocation of labor, production, and prices is unique. If not, the model may exhibit multiple equilibria or path dependence, where small differences in initial conditions lead to persistent divergence across regions.

4 Quantitative Analysis

This section presents a quantitative analysis of the dynamic spatial model using prefecture-level data from China. The main objective is to calibrate and estimate the structural parameters of the model, discipline key mechanisms using micro and macro data, and use the resulting framework to simulate the spatial distribution of labor, wages, productivity, and migration dynamics in the presence of *hukou*-induced mobility frictions. The model features Fréchet-distribution firm productivity, land rents funding amenity investments, and free entry and exit in the production sector, generating endogenous agglomeration and dispersion forces.

The analysis proceeds in four steps. First, we describe the data sources used to construct key model variables. Second, we outline the empirical strategies used to estimate or calibrate structural parameters. Third, we derive initial conditions such as productivity and amenities for the starting year. Finally, we compute compound mobility frictions and decompose them into geographic and institutional components.

4.1 Data

The quantitative analysis of the model relies on a rich set of data sources that together capture the spatial and temporal variation across Chinese prefecture-level cities. The geographic units of analysis include 313 prefectures, harmonized across decades to account for boundary adjustments, administrative reclassifications, and missing observations. The harmonization procedure ensures consistency in geographic identifiers and comparability over time.

The starting year for model calibration is 2000. This year represents a period before major institutional reforms to the *hukou* system and thus provides a clear empirical baseline from which to evaluate the dynamic evolution of spatial outcomes.

The primary source for demographic and labor market information is the China Population Census, conducted decennially in 2000, 2010, and 2020. The census provides comprehensive data on resident population, registered *hukou* population, migration status, education, employment, and household composition. It includes information on both place of residence and place of *hukou* registration, which is critical for modeling internal migration under institutional frictions. Data at the prefecture level are aggregated from individual responses.

To supplement the census, we use microdata from the China Migrants Dynamic Survey (CMDS). This nationally representative annual survey, administered by the National Health Commission, offers detailed information on individual migrants' socioeconomic characteristics, migration histories, *hukou* status, income, employment sectors, and destination choices. It is especially useful for estimating the elasticity of migration with respect to wages and distance, and for documenting *hukou*-related penalties in mobility and welfare.

To construct initial migration flows, we also rely on the 1% micro-sample of 2000 China Population Census made available through the Integrated Public Use Microdata Series (IPUMS). This dataset contains retrospective migration data, allowing reconstruction of inter-prefecture migration flows based on the place of residence five years prior. It also enables estimation of revealed-

Prefecture-level economic data are obtained from the China City Statistical Yearbooks, published annually by the National Bureau of Statistics. These yearbooks report key variables at the prefecture level, including GDP, sectoral output, average wages, employment figures, land use, and household consumption. We use these data to calculate regional wages, labor shares, land inputs, and consumption-based utility proxies. They also provide the base values for initial conditions in the model.

Data for modeling amenities are assembled from multiple official sources. Observable amenity indicators—such as environmental quality, public infrastructure, healthcare services, educational institutions, and cultural facilities—are primarily drawn from the China City Statistical Yearbooks and Ministry of Ecology and Environment bulletins. These indicators are used later to construct a composite amenity index at the prefecture level.

Exogenous components of amenity—those not shaped by endogenous population changes—are proxied using geophysical and meteorological datasets. Climatic data, including average temperature, precipitation, humidity, and solar radiation, are obtained from the China Meteorological Data Service Center. Terrain and topography, including elevation and ruggedness, come from the Relief Degree of Land Surface Dataset of China, compiled by the National Geographic Information Center.

To compute trade frictions, we construct a matrix of bilateral trade costs across all prefecture pairs using GIS-based travel time estimates. Specifically, we apply the ArcGIS Origin-Destination (OD) Cost Matrix tool, overlaying a transportation network based on roads and railways from the year 2000. This network includes national highways, provincial roads, and major railway lines. The tool calculates the shortest travel time between each pair of prefectures, which is then transformed into the bilateral iceberg trade cost matrix across all region pairs.

These datasets together provide comprehensive coverage of spatial heterogeneity in labor, amenities, productivity, and migration decisions. They serve as the empirical foundation for the model calibration and estimation. In the subsequent sections, we describe how key structural parameters are estimated using these data.

4.2 Parameter Estimation and Calibration

4.2.1 Parameters Set to Literature Benchmarks

The discount factor β adjusted to reflect the use of decadal intervals, leading to a value of 0.78, corresponding to an annualized interest rate of 2.5%. The elasticity of substitution σ is set at 4, following Krugman (1991), capturing the degree to which goods are substitutable. This parameter critically influences the degree to which goods can be substituted for one another, with a lower elasticity indicating a higher degree of product differentiation and a greater tendency for consumers to value a diverse range of goods. This moderate elasticity assumes meaningful but not perfect substitutability.

Trade elasticity δ, which varies across industries and countries, is a key parameter. Caliendo et al. (2019) find that trade elasticity typically falls within the range of 3 to 8, which is consistent with earlier estimates by Eaton and Kortum (2002) and Head and Mayer (2014). However, intra-country trade generally faces smaller barriers compared to international trade, so the trade elasticity δ in this model is calibrated at 4.55, towards the lower end of the literature's spectrum.

The persistence of future productivity on historical productivity levels, represented by α, is set at 0.98. This parameter captures the concept of "path dependence" or "productivity persistence," where regions with a history of high productivity are more likely to maintain high productivity levels in the future due to accumulated advantages like knowledge, continuous innovation, and superior infrastructure. This value is based on studies by Allen and Donaldson (2020), Moretti (2012), Comin and Hobijn (2010), Fagerberg, Srholec, and Knell (2007) and many others, making 0.98 a reasonable value.

In addition to parameters derived from existing literature, other parameter values are estimated from data, with details discussed in the following sections.

4.2.2 Parameters Estimated from Data

Labor Share in Production. Labor's contribution to production, typically ranging from 50% to 70% in the literature, is set at 0.58 in this model. While developed countries often have higher labor shares, China's rapid industrialization and capital-intensive growth have led to a declining labor share. Xu, Chen, and Li (2015) find labor shares of 71% at the sectoral level and 58% at the provincial level. Empirical analysis using prefecture-level panel data from 1993 to 2003 suggests an even lower labor share of around 34% after accounting for year and location-fixed effects, which is consistent with what Qi (2015) found using different measurements.

Amenity, Agglomeration and Dispersion. In the model, local amenity levels are endogenously determined by both population density and amenity investment funded by land rents. Specifically, the amenity level in location n at time t is specified as:

$$A_t(n) = \bar{A}(n)\bar{l}_t(n)^{\lambda} I_t(n)^{\chi},$$

where $\bar{A}(n)$ is the exogenous component of amenities that captures geographical and climate fundamentals, $\bar{l}_t(n)$ is population density, and $I_t(n)$ is amenity investment per capita, proxied by city maintenance expenditure per residence. This function form captures two key mechanisms: agglomeration or congestion effects from population density λ and the impact of fiscal investment χ on amenities.

To empirically estimate λ and χ, we adopt a two-step strategy. First, I decompose amenity levels into exogenous and endogenous components by regressing a composite amenity index on geographic and climatic fundamentals. The fitted values define $\bar{A}(n)$, while the residual captures the portion of amenities responsive to population and investment dynamics.

In the second stage, we estimate the elasticity parameters using a GMM framework. This approach accounts for potential endogeneity in both population density and rent-financed amenity investment. Specifically, we use deep lags of population and predicted rents (based on land supply and location-specific productivity) as instruments. The GMM moments are constructed to match the conditional covariance of residualized amenity outcomes with the instrument set.

The estimation yields a negative elasticity of amenities with respect to population density $\lambda = -0.39$, indicating net congestion effects in China's urban regions. The elasticity with respect to amenity investment is positive and significant $\chi = 0.22$, suggesting that increases in land rents—interpreted here as public or quasi-public investment in quality-of-life infrastructure—are associated with higher local amenity levels.

These values are used in the model to determine the strength of the feedback loop between migration, congestion, and urban investment. Importantly, the estimated negative value of λ contributes to the condition for equilibrium uniqueness (see Section 3.6), ensuring that population agglomeration does not explode in the model's spatial dynamics.

A detailed discussion of the amenity index construction, data sources, first-stage regression, variable-level descriptions, and robustness checks is provided in Appendix C1.1.

Heterogeneous Amenity Preference. In our model, agents make migration decisions based on expected utility, which includes both observed regional variables (real wage, local amenities, and migration cost) and individual-specific preference shocks drawn from a Fréchet distribution. This distribution is governed by the shape parameter γ, which characterizes the dispersion of these shocks: a higher γ implies more homogeneous preferences and greater responsiveness to differences in expected utility across locations.

However, simultaneously disentangling the behavioral responses to wages, amenity levels, and migration costs in empirical application can quickly become complex. Guided by both simplicity and precedent in spatial equilibrium literature, such as the work presented by Steven Redding and many others, I interpret γ as the elasticity of migration flows with respect to real wage differentials, holding other factors constant.

γ is estimated using a reduced-form gravity regression of bilateral migration flows on real wage differentials and distance:

$$\log M_{n\to j,t} = \gamma \log \left(\frac{w_{j,t}}{w_{n,t}} \right) - \rho \log D_{n,j} + \eta_j n + \eta_j + \eta_t + e_{njt}, \tag{23}$$

where:

- $M_{n\to j,t}$ is the observed number of migrants from origin n to destination j at time t,

- $w_{j,t}$ and $w_{n,t}$ are real wages in the destination and origin respectively,
- $D_{n,j}$ is the geodesic distance between n and j,
- η_n, η_j and η_t are destination and time fixed effects.

The estimation uses pooled data from multiple rounds of the China Migrants Dynamic Survey (CMDS). We cluster standard errors by origin-destination pair to account for unobserved bilateral heterogeneity.

The resulting estimate for γ is approximately 1.8, consistent with values reported in related spatial models (e.g., Desmet et al. (2018), Cruz and Rossi-Hansberg (2021)). Based on this evidence and to maintain model tractability, we set $\gamma = 2$ for the quantitative analysis. Robustness checks with $\pm 20\%$ variations in γ are reported in Appendix E.

Agglomeration Effects on Productivity. Local agglomeration plays a vital role in shaping regional productivity dynamics. Following the literature, we model the evolution of regional productivity $Z_t(n)$ as a function of lagged productivity and local agglomeration effects through employment density. Formally, this is expressed as:

$$logZ_t(n) = \alpha logZ_{t-1}(n) + \lambda_1 log\bar{l}_{t-1}(n) + e_t(n)$$

α captures the persistence of productivity over time, while λ_1 captures the elasticity of productivity with respect to population density, reflecting localized learning, input sharing, and knowledge spillovers.

To estimate λ_1, we use panel data on prefecture-level GDP per capita and population density from 1995 to 2005. Several estimation methods are implemented to ensure robustness and account for potential endogeneity concerns due to dynamic panel bias and omitted variable bias.

Our preferred estimate of λ_1 is 0.21, based on the difference-GMM estimator using second lags of the dependent and independent variables as instruments. This value is consistent with estimates found in the urban economics literature and supports the presence of localized productivity spillovers in China's urban areas. Estimates using other approaches range from 0.014 to 0.39, depending on instrument sets and model specifications. We provide full technical details, robustness checks, and diagnostics in Appendix E.

Bilateral Trade Costs. The estimation of bilateral trade costs follows the methodology proposed by Desmet et al. (2018), with adaptations to account for Chinese infrastructure and available data. In this model, goods are transported between cities via on-land infrastructure, incurring iceberg-type trade costs. These costs are a crucial friction affecting both prices and spatial allocation of economic activity.

To construct bilateral trade costs between all 313 prefecture-level cities in China, we use the ArcGIS Origin-Destination (OD) Cost Matrix tool. The tool computes the fastest route between each city pair using transportation networks composed of highways, roads, and railways based on geospatial data from the year 2000. For each city pair (n, s), we calculate the shortest travel time $t(n, s)$, measured in hours. We focus exclusively on land-based transport infrastructure, justified by data from China's Statistical Yearbook indicating that over 85% of freight in 2000 relied on on-land transportation.

Once the shortest travel times are computed, we translate them into iceberg trade costs using a monotonic transformation:

$$v(n, s) = 1 + t(n, s)^{0.632}$$

. This functional form captures increasing marginal trade frictions over distance and is consistent with the empirical specification, such as in Feyrer (2019). For intra-prefectural trade $(n = s)$, the travel time $t(n, n) = 0$ implies a trade cost of exactly 1 (i.e., no frictions).

The **trade elasticity parameter** governs how sensitive trade flows are to cost differences across locations. It is also the shape parameter of the Frchet distribution governing firm productivity in our model. Rather than estimating this parameter from our data, as we mentioned before, we adopt a standard value from the literature.

Model Parameters. Below is a summary of key parameters used in the model, along with their sources and justifications:

Parameters			
Parameters	Description	Value	Source/Notes
β	Discount factor	0.776	Reflecting annualized 2.5% interest rate
σ	Elasticity of substitution across varieties	4	Krugman (1991)
δ	Trade elasticity	4.55	Caliendo et al. (2019)
α	Persistence of productivity	0.98	Allen and Donaldson (2020)
λ	Elasticity of amenity w.r.t population	- 0.39	GMM-estimated from urban amenity regression
χ	Elasticity of amenity w.r.t amenity investment	0.23	GMM-estimated from urban amenity regression
γ	Elasticity of migration flows w.r.t. real income	2	Estimated from migrant flows and wage gaps
λ_1	Agglomeration elasticity in productivity dynamics	0.21	Estimated from density-productivity regressions
ι	Labor share in production	0.58	Estimated using prefectural GDP and wage data

4.3 Values for Initial Period

4.3.1 Initial Productivity and Exogenous Amenity

To solve this dynamic model, we require values for initial productivity and exogenous amenities in each location. Following the strategy in Desmet et al. (2018), we recover these unobservables using observed allocations and prices in the initial year, without assuming a balanced growth path. This approach relies on the idea that current equilibrium outcomes already encapsulate the underlying fundamentals, and these fundamentals can be inferred using the model's structure and first-order conditions.

Given initial population $\bar{L}_0(n)$, wage w_0, and land from the data, we begin by recovering initial regional productivity $Z_0(n)$ from the model's equilibrium wage equation:

$$w_0(n) = (1-\iota)^{-1}\iota^{\frac{\chi-1}{\chi}} \left[\frac{\tilde{u}_0(n)}{\bar{A}(n)}\right]^{\frac{1}{\chi}} Z_0(n)^{-\frac{1}{\chi}} \bar{l}_0(n)^{\frac{1-\iota-\lambda}{\chi}}, \tag{24}$$

where $\tilde{u}_0(n) = A_0(n)\frac{w_0(n)}{P_0(n)}$ is the deterministic component of utility excluding idiosyncratic preference, amenity loss, and migration cost.

Rearranging the equation allows us to express initial productivity:

$$Z_0(n) = (1-\iota)^{-\chi}\iota^{\chi-1}\frac{\tilde{u}_0(n)}{\bar{A}(n)}w_0(n)^{-\chi}\bar{l}_0(n)^{1-\iota-\lambda} \tag{25}$$

To implement this step, we must obtain the ratio $\frac{\bar{A}(n)}{\tilde{u}_0(n)}$ for each location. From the price index formulation derived in the trade block of the model, we obtain the following equation:

$$\left[\frac{\bar{A}(n)}{\tilde{u}_0(n)}\right]^{-\delta} w_0(n)^{-\delta(\chi+1)}\bar{l}_0(n)^{-\lambda\delta} = \kappa\left[\frac{\iota}{1-\iota}\right]^{2\chi\delta}\sum_j\left[\frac{\bar{A}(j)}{\tilde{u}_0(j)}\right]^{-\delta}T(j)w_0(j)^{-\delta(1+\chi)}\bar{l}_0(j)^{-\lambda\delta}\upsilon(n,j)^{-\delta}, \tag{26}$$

where $\kappa = \left[\Gamma(\frac{1-\sigma}{\delta}+1)\right]^{\frac{\delta}{\sigma-1}}$ is a constant.

This expression enables us to recover the ratio $\frac{\bar{A}(n)}{\tilde{u}_0(n)}$ up to a normalization, given observed wages, population, and estimated bilateral trade costs. Then, we return to the productivity equation to recover the initial productivity $Z_0(n)$ for each location.

To pin down the exogenous amenities $\bar{A}(n)$ directly, we use the consumption-based proxy for utility $\tilde{u}_0(n)$ with data on per capita consumption expenditures in 2000. This allows us to compute $\bar{A}(n)$ from the estimated ratio.

This approach ensures internal consistency between the model's equilibrium conditions and the observed spatial distribution of wages, population, and trade, while avoiding strong assumptions about future growth paths.

The full derivation of these expressions and details on the solution algorithm are provided in Appendix E.

4.3.2 Mobility Costs and Amenity Lost

To implement the model, it is crucial to estimate the mobility costs between different locations accurately. Using the exogenously computed amenity level $\bar{A}$ and the labor population data for period 0, $\bar{l}_0(n)$, along with the *hukou* population data from 1990 (denoted as period -1), we can calculate the values for $u_1(n)$ at each location. Given the assumption that the *hukou* parameter remains time-invariant—meaning individuals do not change their *hukou* status—the total population with a specific *hukou* h remains constant over time. As a result, we omit the time subscript for $\bar{L}^h$.

From (5) and (6), we have the following for period 0:

$$H(n)\bar{l}_0(n)\mu_0(n)^h = \frac{\tilde{u}_0(n)^\gamma \tilde{m}(n,s)^{-\gamma}}{\sum_{j=1}^N \tilde{u}_0(j)^\gamma \tilde{m}(j,s)^{-\gamma}} \bar{L}_{-1}^h, \tag{27}$$

where $\tilde{m}(n,s) = m(n,s)\phi(n,h)$, which incorporates the frictions due to *hukou* parameter ϕ. We refer to this as the "compound" mobility cost in the subsequent discussion.

Consider the scenario where an individual remains in their registered *hukou* location ($n = h$), implying that $\tilde{m}(n,n) = 1$, since both the *hukou* parameter and movement cost equal one:

$$H(n)\bar{l}_0(n)\mu_0(n)^{h=n} = \frac{\tilde{u}_0(n)^\gamma}{\sum_{j=1}^N \tilde{u}_0(j)^\gamma \tilde{m}(j,n)^{-\gamma}} \bar{L}_{-1}^{h=n}, \tag{28}$$

In this case, $\mu_0(n)^{h=n}$ represents the proportion of people who stay in their registered *hukou* location at time 0.

Taking the ratio of equations (27) and (28), and rearranging, we can derive a closed-form expression for the compound mobility cost from the data:

$$\tilde{m}(n,s) = \left[\frac{\mu_0(n)^{h=n} \bar{L}_{-1}^h}{\mu_0(n)^h \bar{L}_{-1}^{h=n}} \right]^{1/\gamma} \tag{29}$$

Using the locational productivity derived for period 0, $Z_0(n)$ and the labor population data $\bar{l}_0(n)$ along with the migration flows of *hukou* population, we can calculate mobility cost $\tilde{m}(n,s)$ for each location from the model. Detailed computations are provided in the appendix.

To isolate the amenity loss due to *hukou* from the compound migration cost $\tilde{m}(n,s)$, I model the migration cost based solely on distance, denoted as $m(n,s)$. This distance-based migration cost is assumed to be symmetric between two locations, reflecting the idea of an invariant baseline migration cost. It incorporates aspects such as the loss of social networks and capital at the place of origin, as well as the psychological toll of leaving family and familiar surroundings. However, it may not fully account for all tangible costs associated with moving and settling in a new area.

To discipline the elasticity parameter ρ of the distance-based migration cost, I utilize data from the CMDS (China Migration and Demographic Survey). The following gravity model is constructed to align with observed data on population, distance, and migration flows:

$$M(n,s) = G \frac{P_n^{\nu_1} P_s^{\nu_2}}{D(n,s)^\rho},$$

where G is a constant, P_n and P_s represent the populations of locations n and s, respectively, and ν_1 and ν_2 are the elasticities of migration flows with respect to the populations at the origin and destination. $D(n, s)$ denotes the distance between locations n and s, and ρ is the elasticity of migration flows with respect to distance.

From this model, I estimate $\rho = 1.32$, which aligns with the typical range of 0.7 to 2 found in the literature. This estimated migration cost allows us to isolate the "pure" frictions associated with *hukou* from the "compound" migration cost.

4.4 Counterfactuals Analysis: Removing Institutional Barriers

Having fully parameterized and calibrated the model using historical data and empirical estimates, I proceed to analyze the long-run spatial implications of institutional barriers to labor mobility—China's *hukou* system. The goal of this counterfactual exercise is to isolate the dynamic consequences of lifting migration frictions while holding other structural features of the economy constant.

Simulation Approach. To implement the counterfactual, I employ the dynamic hat algebra method. This approach facilitates tractable analysis of long-run spatial transitions by computing relative changes in endogenous variables over time, rather than solving for levels directly. The model is initialized using observed data in period $t = 0$, which reflects the status quo, where the *hukou* system is in place and continues to distort individual location choices via migration frictions embedded in the bilateral cost function $\phi(h, n) > 1$ for migrants without local registration.

I then simulate the spatial evolution of the economy under two scenarios:

- The **benchmark** scenario maintains the existing *hukou* system throughout the transition, with all bilateral migration frictions remaining unchanged over time.
- The **counterfactual** scenario abolishes the *hukou* system from period $t = 0$ onward by setting $\phi(h, n) = 1$ for all agents and destinations. This implies that all workers face equal amenity costs regardless of their origin, thereby removing institutional barriers that previously restricted mobility across regions.

Both scenarios are simulated for 5-10 periods (interpreted as 50 to 100 years), allowing the economy to reach a new long-run spatial equilibrium. In each scenario, I solve for the full set of endogenous equilibrium objects over time, including regional utility levels $u_t(n)$, labor distributions $\bar{l}_t(n)$, and productivity levels $Z_t(n)$.

Interpretation and Outcomes. By comparing the paths $\hat{u}_t(n)$, $\hat{\bar{l}}_t(n)$, and $\hat{Z}_t(n)$ from the counterfactual scenario to those under the benchmark, I assess the dynamic effects of removing institutional frictions on the spatial distribution of population and productivity. The key outcomes of interest include: i) The reallocation of labor across regions, highlighting areas that gain or lose population in response to relaxed migration constraints. ii) Changes in regional productivity, which evolve endogenously as a function of local agglomeration effects and innovation dynamics. iii) Welfare improvements, computed as changes in lifetime utility, aggregated across individuals and weighted by initial population distribution.

To maintain a clear empirical focus, this analysis restricts attention to the effects of institutional reforms only. In particular, improvements in transportation infrastructure—such as the expansion of high-speed rail (HSR)—are excluded from this chapter. These infrastructure changes often co-evolve with migration and productivity, but they constitute a distinct set of policy instruments with their own modeling complexities. I therefore reserve the analysis of HSR and related transportation improvements for a separate chapter, where I will explicitly model trade costs, access, and infrastructure-induced migration incentives in interaction with institutional reforms.

Quantitative results and visualizations of the key counterfactual trajectories are presented in Section 5 and Appendix F.

5 Quantitative Results: Mobility, Productivity, and the *hukou* System

This section presents the results of the quantitative analysis introduced in Section 4, focusing on the role of the *hukou* system in shaping internal migration, productivity, and spatial development in China. We begin by examining the model's calibrated initial conditions, which reflect the economic geography at the turn of the 21st century. We then turn to the estimated migration frictions—including institutional constraints—and assess how well the model reproduces key features of the data. Finally, we conduct a counterfactual analysis to explore the long-run effects of abolishing the *hukou* system on labor reallocation, regional productivity, and aggregate welfare.

5.1 Initial Spatial Economic Conditions

The model begins in 2000 and incorporates key features of China's economic geography: exogenous amenities and initial productivity. These factors collectively shape the initial distribution of population and economic activity and help explain observed migration patterns prior to major reforms.

5.1.1 Exogenous Amenities

Figure 4 maps the estimated exogenous amenities $\bar{A}(n)$ across prefectures, reflecting geographic and environmental endowments. Contrary to the assumption that coastal areas uniformly offer the highest amenities, the map reveals a more nuanced pattern.

Some inland regions — such as Sichuan, Yunnan, and Guangxi — exhibit relatively high amenity values, driven by climate, topography, and natural beauty. These regions attract labor not through industrial agglomeration but via environmental quality. Coastal cities in Guangdong and Zhejiang also feature high amenity levels, reflecting fertile land, trade access, and historical investment.

In contrast, the western provinces (e.g., Tibet, Qinghai, Xinjiang) and parts of Inner Mongolia show lower amenity levels, consistent with harsh terrain, limited arable land, and weaker infrastructure. The northeast industrial belt (Liaoning, Jilin, Heilongjiang) shows moderate amenities, reflecting a mix of industrial legacy and declining environmental quality.

5.1.2 Initial Productivity

Figure 5 displays initial total factor productivity (TFP) across prefectures in 2000. The highest levels are concentrated in the Yangtze River Delta, Pearl River Delta, and Bohai Economic Rim, encompassing cities like Shanghai, Guangzhou, and Beijing. These areas benefited from early liberalization, foreign investment, and policy support, including Special Economic Zones. Their integration into global supply chains and proximity to export markets have sustained high levels of industrial output and labor demand.

Productivity levels in central and southwestern regions vary widely, reflecting mixed development paths. Provinces like Sichuan show moderate productivity, supported by

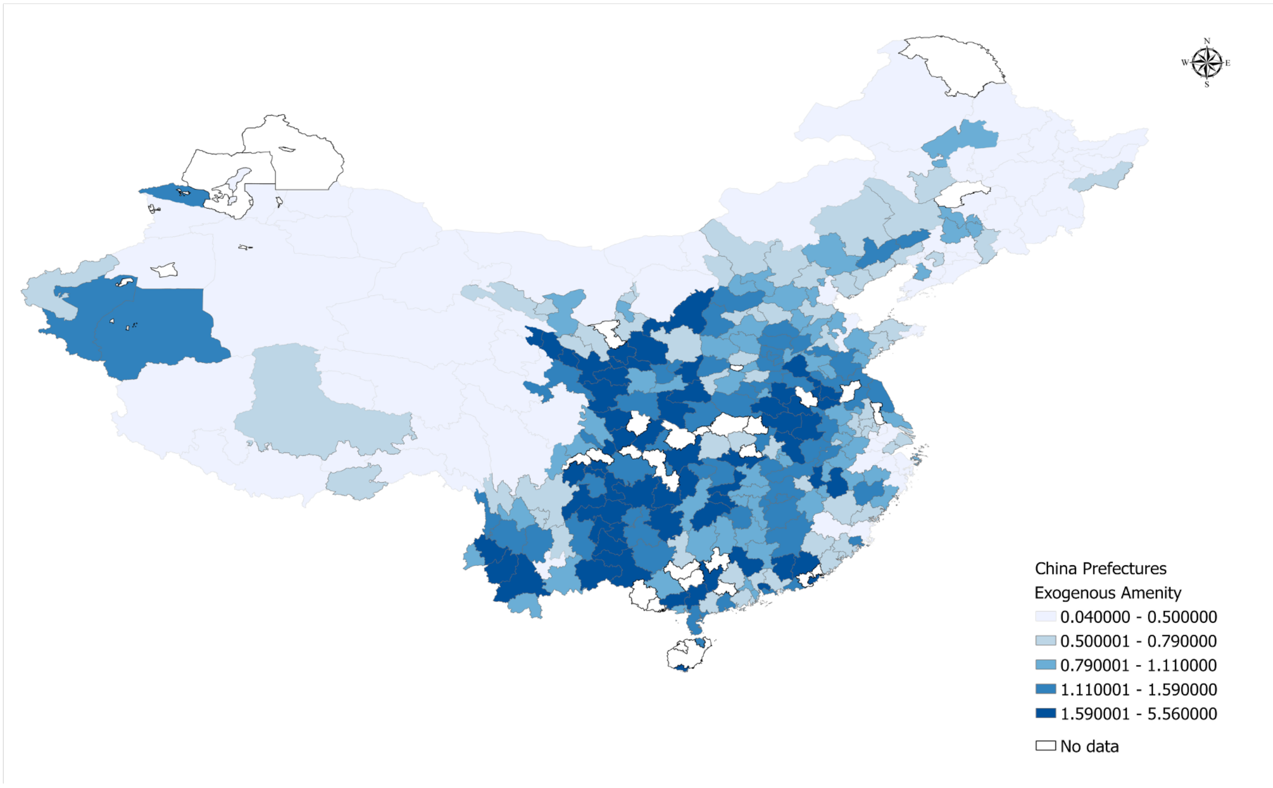


Figure 4: Exogenous Amenities by Prefecture

local manufacturing and a growing service sector. More remote inland areas remain lagging due to weaker industrial bases and transportation constraints.

The northeast, historically industrialized under the centrally planned economy, maintains moderate to high productivity in some prefectures. However, the onset of deindustrialization and state-owned enterprise reform in the 1990s led to economic stagnation and rising unemployment in parts of the region, contributing to population outflows and declining investment.

5.1.3 Initial Utility and Wage Distribution

To further characterize regional heterogeneity, the model calibrates regional deterministic utility and wage levels across locations in the initial period. These outcomes are shaped jointly by productivity, amenities, trade costs, and migration frictions. In equilibrium, they summarize the economic and social value of each location from a resident's perspective.

The left panel in figure 6 displays the spatial distribution of real wages (in units of 1,000 RMB), computed by the model based on estimated productivity, exogenous

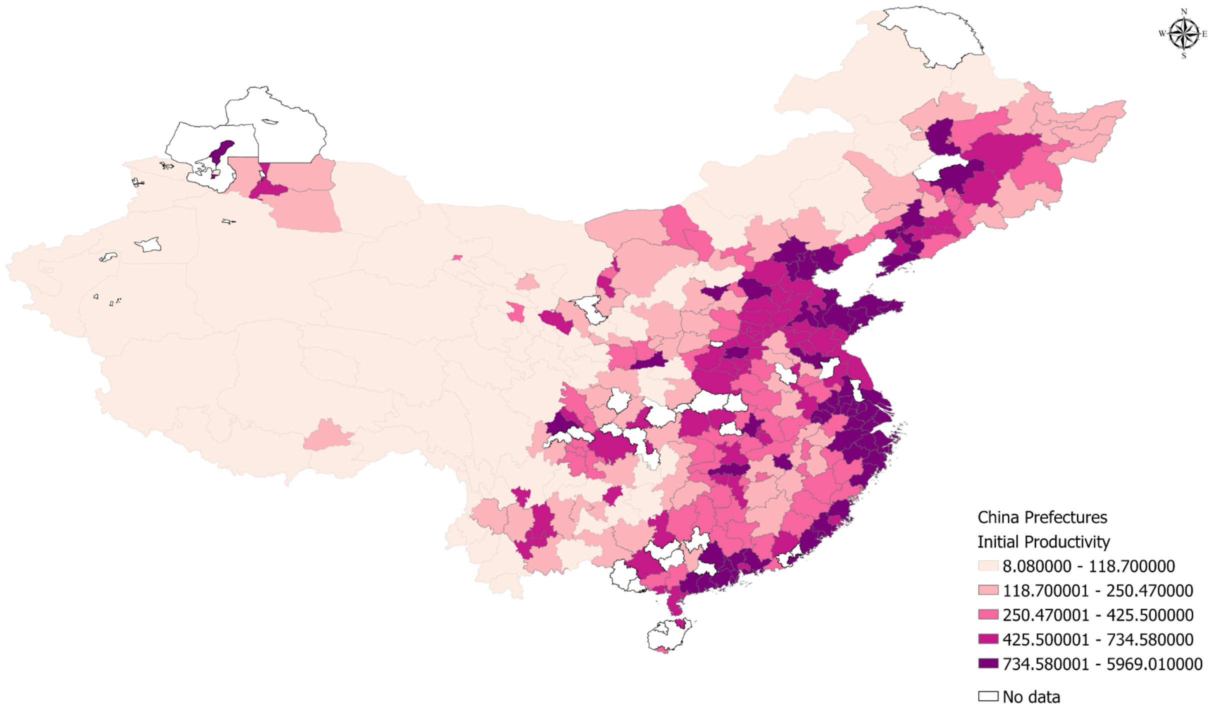


Figure 5: Prefecture-Level Productivity in 2000

amenities, and the equilibrium labor distribution. The highest wages are concentrated in coastal cities such as Shenzhen, Guangzhou, Foshan, and Zhuhai in Guangdong province; Suzhou and Wuxi in Jiangsu; and Xiamen in Fujian. Beijing also stands out, benefiting from its administrative and political centrality. Resource-intensive cities like Karamay, Dongying, Daqing, and Panzhihua show elevated wage levels, primarily due to their strong presence in extractive industries. These patterns reflect the combined influence of firm-level productivity, sectoral composition, and enhanced market access in these locations.

In contrast, western and some central inland regions show considerably lower real wages, reflecting both lower productivity and limited market access. The relative isolation of these areas — due to geography and underdeveloped infrastructure — restricts the size of local markets and firms' ability to scale, thereby constraining wage growth.

The right panel in figure 6 presents the initial utility, which incorporates both consumption and amenity components. The spatial variation in utility reflects not only trade-offs between economic opportunity and quality of life but also the reinforcing effect of high wages and high amenities in many leading cities. In particular, several top

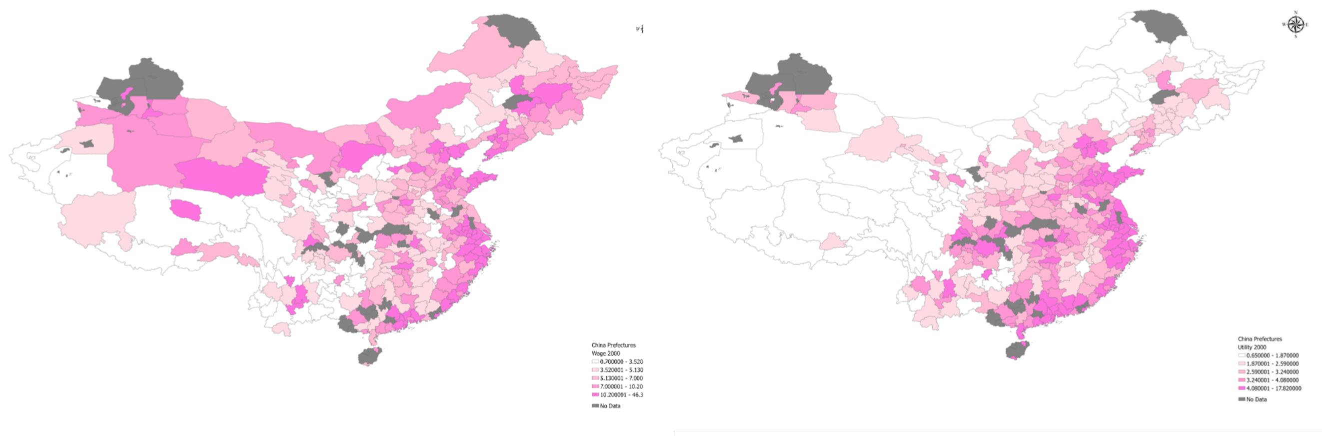


Figure 6: Model-Implied Wage vs. Utility in 2000

wage locations, such as Shenzhen, Guangzhou, and Suzhou, also rank among the highest in deterministic utility, suggesting that these places provide both strong economic returns and favorable living conditions, even after accounting for congestion and baseline frictions. This reflects a close alignment between wage and utility rankings across most regions. However, there are important exceptions. Cities like Karamay, Daqing, and Panjin exhibit high wage levels but do not rank similarly in utility. This discrepancy may be attributed to lower exogenous amenities or higher consumption goods prices due to trade frictions and geographic isolation, which reduce the effective purchasing power and desirability of these locations despite high nominal wages. Some inland cities offer relatively high utility despite modest wage levels, driven by favorable amenities, lower congestion, or lower costs of living. Meanwhile, high-income coastal cities may exhibit only moderate utility once crowding factors are accounted for.

Utility disparities across prefectures signal the presence of significant spatial frictions. While economic forces would predict greater labor reallocation toward high-productivity regions, institutional barriers — notably the *hukou* system — prevent full equalization of utility across space.

Overall, these initial conditions provide a consistent backdrop for analyzing subsequent mobility decisions and policy counterfactuals. Regional disparities in amenities, productivity, and welfare serve as key drivers of migration flows under the *hukou* regime.

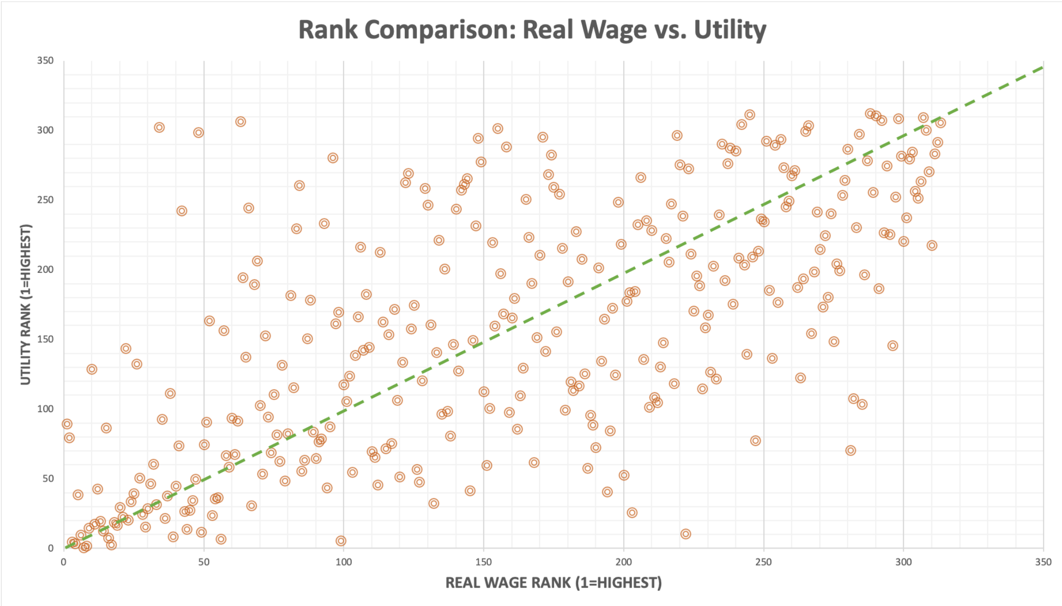


Figure 7: Real Wage and Utility in 2000

5.2 Mobility Cost and *Hukou*

The model estimates a comprehensive structure of bilateral migration frictions that includes both geographic and institutional components. These compound costs are represented as a 313 × 313 matrix of origin–destination migration barriers.

The resulting map reveals a complex spatial pattern in which migration costs are influenced by both physical distance and infrastructural connectivity, alongside the institutional barriers created by the *hukou* system. While geographic isolation and poor infrastructure elevate migration costs in remote areas, even well-connected urban centers can impose high institutional barriers, making it difficult for migrants to access local services and benefits without the proper *hukou* registration.

To better understand how institutional frictions contribute to internal migration barriers, we calculate the *hukou* cost share for each city as the ratio of *hukou*-induced cost to the total inward migration cost.(9) This component captures the relative importance of *hukou* restrictions—as opposed to physical distance, or geographic isolation—in shaping individuals' effective cost of relocating across prefectures.

The resulting distribution is highly uneven across space, with clear geographic and administrative patterns. In some cities, such as Suzhou, Lu'an, and Anqing in Anhui province; Luzhou, Neijing in Sichuan province; and Longnan, Zhangye in Gansu province, the *hukou* share of total migration cost exceeds 75%, indicating that institutional restric-

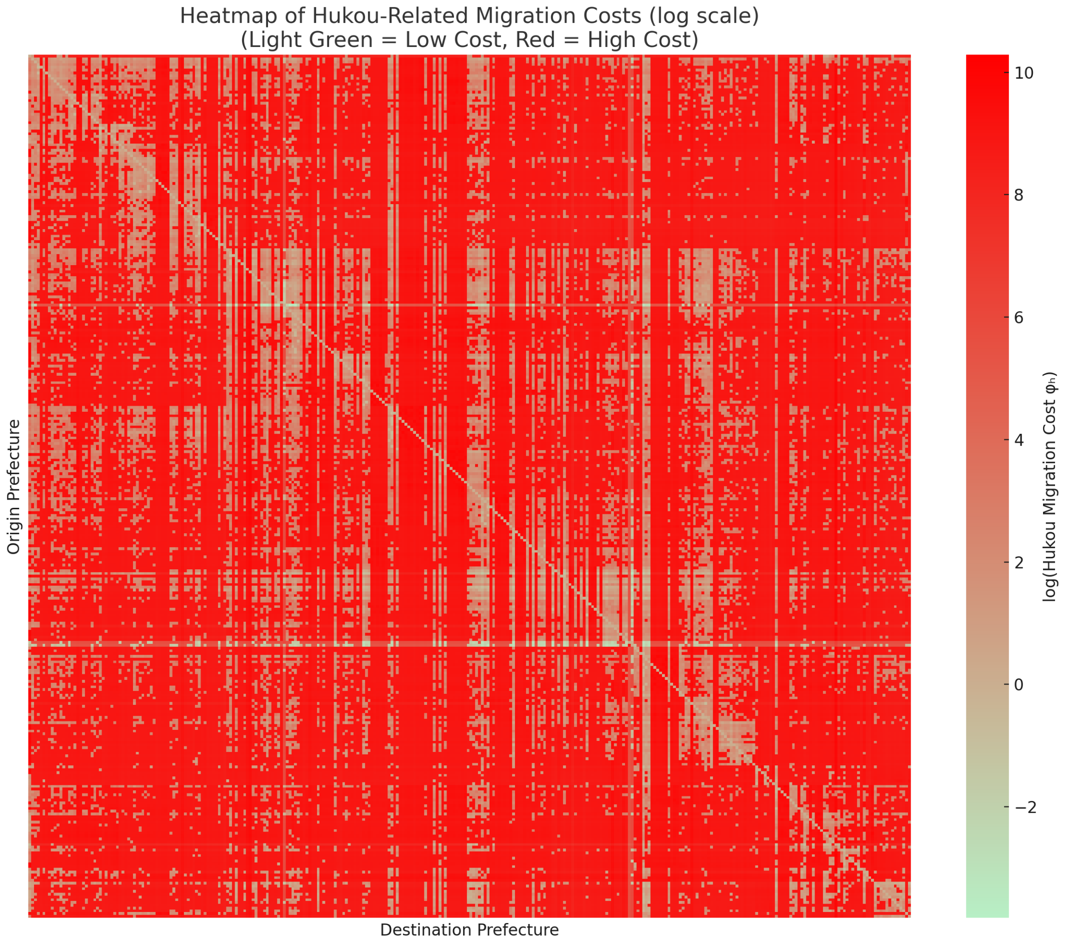


Figure 8: *hukou* Induced Migration Cost (log-scale

tions—rather than geographic or economic factors—constitute the primary deterrent to migration. These cities typically feature rigid registration systems and limited integration pathways for migrants, amplifying the disutility associated with moving in.

In contrast, other cities such as Haikou (21%), Chengdu (27%), Zhoushan (38%),Chongqing (40%), Sanya (48%), and Wenzhou (49%) show low hukou cost shares, despite varying levels of compound migration cost. These cases reflect environments where either institutional policies are more flexible or physical remoteness, trade frictions, or low connectivity dominate the migration friction landscape.

This variation across the two maps shows that some cities are dominated by institutional barriers, while others are more affected by connectivity and geographic isolation. Thus, the interplay between physical and institutional frictions is highly location-specific.

Many economically advanced cities—including Tier 1 cities such as Shenzhen (66%), Shanghai (60%), and Beijing (54%) fall into the middle range of the *hukou* cost share

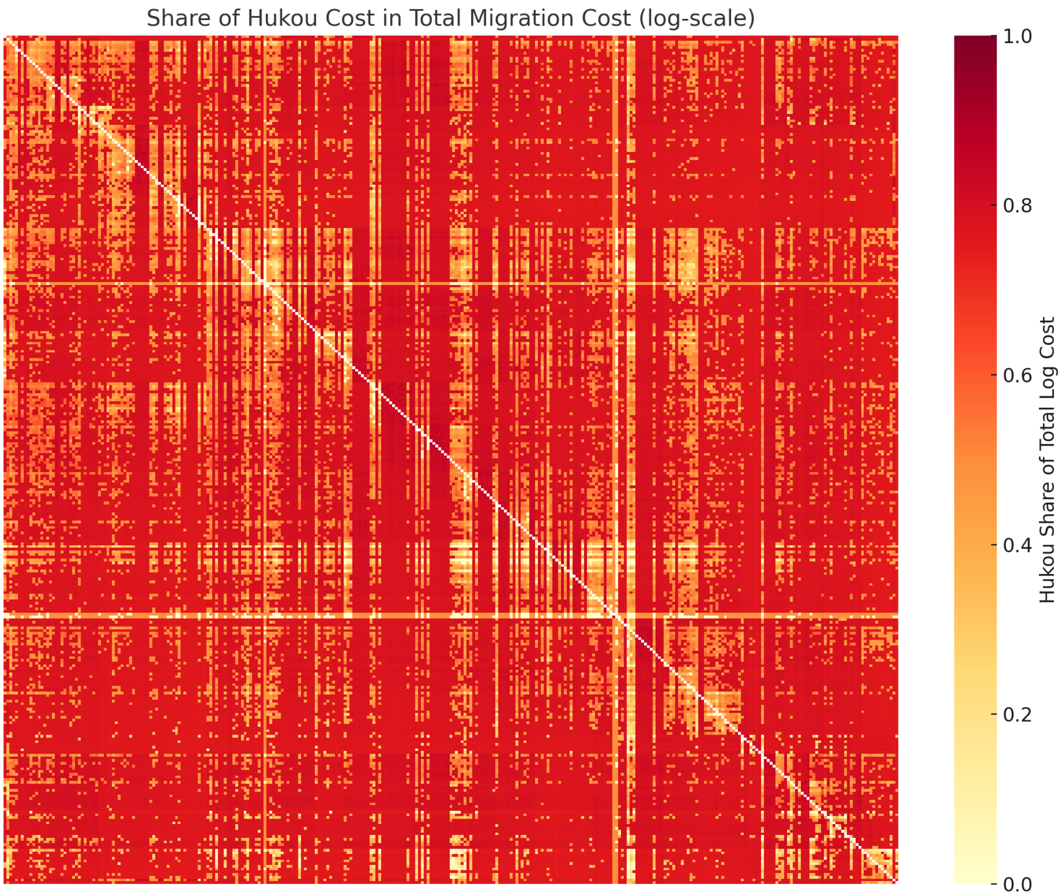


Figure 9: Share of *hukou* Cost in Total Migration Cost (log-scale)

distribution. While these cities had not yet implemented any significant *hukou* policy reforms in 2000, they remained attractive migration destinations due to their high wages and superior amenities. The hukou barrier in these cases represents a meaningful share of total migration cost, but migrants were still willing to absorb these costs to access economic opportunities. Similarly, cities like Nanjing (53%) and Changsha (53%) show moderate *hukou* frictions, suggesting that institutional barriers were present but not necessarily the dominant constraint in shaping migration decisions.

Importantly, the data challenge common narratives that associate high *hukou* barriers solely with large cities. Despite perceptions that cities like Beijing and Shanghai are among the most difficult for migrants to enter, the model suggests that their high baseline utility and economic opportunities may offset *hukou* penalties in equilibrium. The *hukou* cost share is not solely a function of city size or economic development. Instead, it reflects a combination of local administrative practices, social service availability, and the structure of total migration costs. In some cases, migrants may perceive restrictive *hukou*

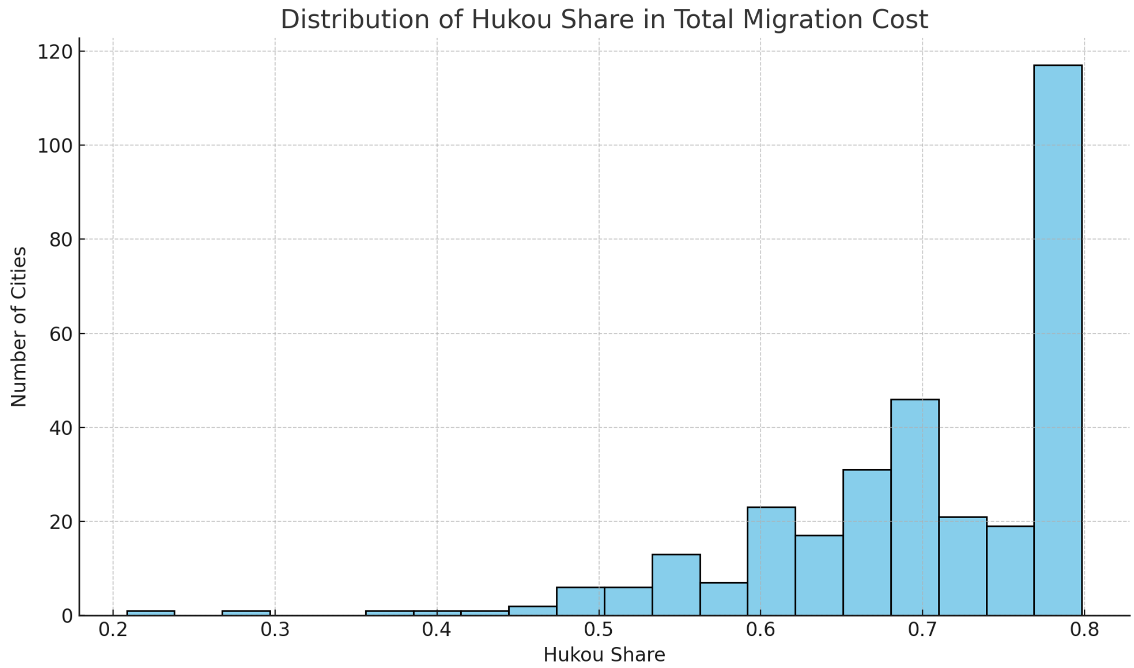


Figure 10: Distribution of *hukou* Cost Share in Total Migration Cost

policies as a larger burden than physical movement, while in others, distance and trade costs dominate the migration decisions.

This heterogeneity implies that reforming the *hukou* system could have differential impacts. In high-barrier cities, relaxing *hukou* restrictions could significantly lower total migration costs and promote labor reallocation. In contrast, in cities where *hukou* is already a minor factor, improvements in transport, housing, or wage opportunities may be more critical.

5.3 Model Validation

To assess the credibility and empirical relevance of the model, we conduct a validation exercise using untargeted moments. Specifically, we evaluate the model's ability to reproduce (i) provincial contributions to real GDP, (ii) population density across regions, and (iii) spatial patterns of labor mobility over time. These dimensions are critical for assessing whether the model reliably captures the dynamics of China's regional economy and internal migration during a period of rapid urbanization and structural transformation.

Figure 11 compares simulated and observed provincial GDP shares in the year 2000. The model successfully replicates broad spatial patterns in economic output, capturing the dominance of coastal provinces such as Guangdong, Jiangsu, and Zhejiang. These regions, historically prioritized for industrial development and foreign trade, contribute substantially to national GDP, both in the model and in the data. The model also reflects the growing economic roles of inland provinces like Henan and Hubei, which have emerged as transportation and manufacturing hubs in recent decades.

Despite this alignment, the model overestimates GDP shares for some leading provinces such as Guangdong and Shanghai, while underestimating Beijing's economic output. These discrepancies may stem from the model's assumption of static productivity in the initial period, which does not fully capture city-specific innovation or external shocks. Nonetheless, the overall fit is strong, with a correlation coefficient of 0.94 between estimated and observed GDP shares, indicating that the model provides a reasonable approximation of cross-provincial economic structure.

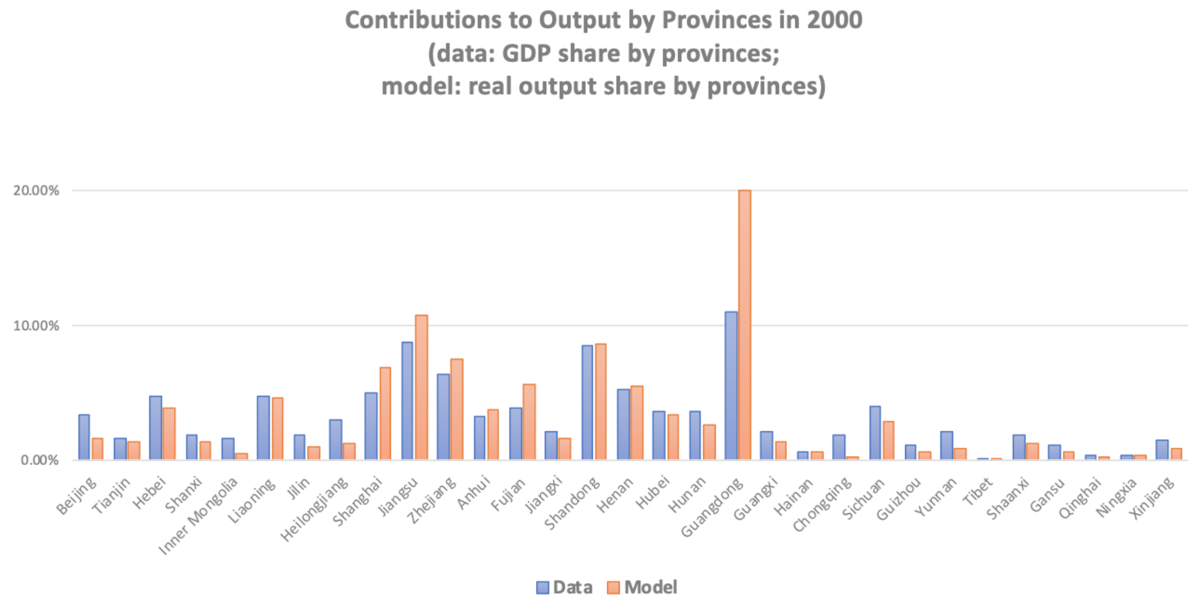


Figure 11: Output Share by Province in 2000

Next, we examine the model's performance in replicating spatial population patterns. Figure 12 shows scatter plots comparing predicted and actual population density (log-transformed) for 313 prefectures in 2010 and 2020. The model captures the broad demographic reallocation from inland to coastal regions, particularly the rapid urbanization of the Yangtze River Delta, the Pearl River Delta, and the Beijing-Tianjin-Hebei corridor. These regions attract large inflows of labor due to their superior productivity, infrastructure, and amenities—patterns that are well reflected in the simulated population distribution.

The model also reproduces labor outflows from interior provinces such as Gansu, Guizhou, and parts of Sichuan, where lower productivity and higher mobility frictions reduce local economic attractiveness. Some discrepancies remain, notably in high-density cities like Shanghai and Beijing, where the model slightly underpredicts population size. These deviations may result from not fully modeling land-use constraints, housing prices, or formal migration quotas, which could influence congestion and settlement patterns.

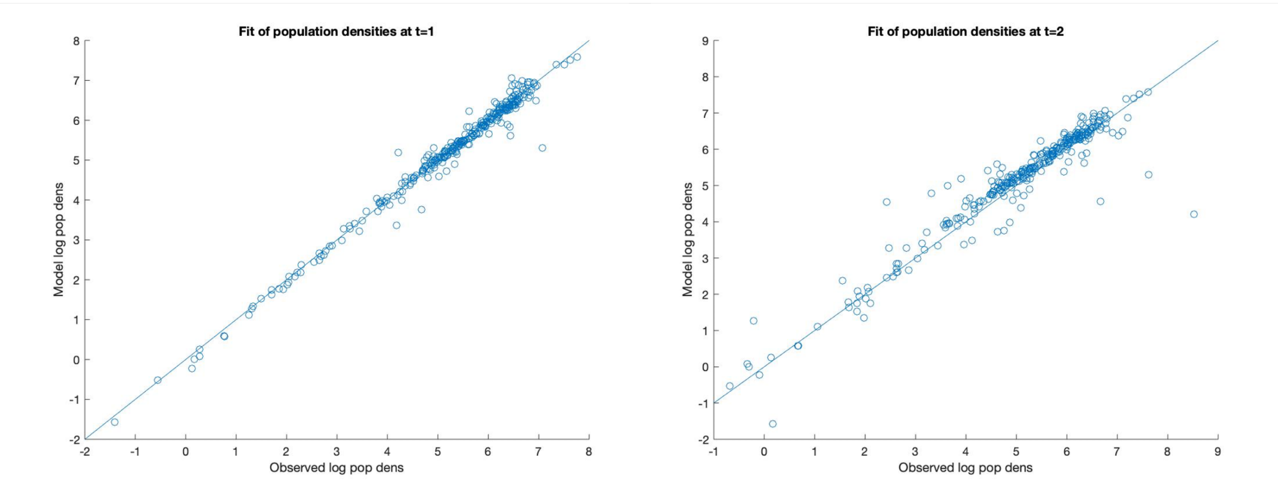


Figure 12: Population Density in 2010 and 2020

Overall, the model generates a strong empirical fit along both economic and demographic dimensions. It captures the essential drivers of regional development and labor mobility: spatial productivity gaps, endogenous amenities, trade access, and institutional barriers such as the *hukou* system. While some simplifying assumptions may understate second-order frictions, the observed alignment with untargeted moments supports the model's utility for policy simulations and counterfactual analysis in later sections.

6 Counterfactual Analysis: *hukou* Reform Scenarios

In this section, we evaluate the economic and spatial implications of abolishing or partially removing the *hukou* system under different policy designs. We first consider three reform scenarios with varying scope, then decompose the resulting welfare changes into their underlying mechanisms. We further examine the robustness of these results to alternative parameter values, and—where data permit—provide a back-of-the-envelope cost–benefit calculation.

6.1 Reform Scenarios

We analyze the following policy scenarios, all relative to the baseline calibration with observed *hukou* frictions:

1. **S0 – Full abolition:** All *hukou*-related migration costs are removed nationalwide, i.e., $\phi(h, n) = 1$ for all origin–destination pairs (h, n). Migration is subject only to physical or geographic costs $m(h, n)$.

2. **S1 – Targeted at high-barrier cities:** Abolition of *hukou* restrictions only in destinations with a *hukou* cost share $\Phi(n) \geq 0.75$. For all other destinations, $\phi(h, n)$ remains at its baseline level. This scenario targets 128 prefecture-level cities and encompasses about 36% of the total population.

3. **S2 – Targeted at economic hubs:** Abolition of *hukou* restrictions only in the top 10% of cities by baseline GDP share. Other destinations retain baseline $\phi(h, n)$. This scenario covers 46 cities and 18% of the total population.

6.2 Aggregate Results

Table 1 summarizes the short-run ($t = 2$) and long-run ($t = 10$) percentage changes in aggregate productivity, real output, and deterministic consumption-equivalent (CE) welfare for each scenario. The CE welfare measure is derived from the present discounted value (PDV) of deterministic utility following the Aiyagari-style aggregation outlined in Section E, using the CRRA-to-CE mapping in Equation(E.1).

To quantify the welfare effects of *hukou* abolition, we compute the consumption-equivalent (CE) welfare change using the deterministic utility paths from the model and a CRRA aggregator with $\rho = 2$ a standard macroeconomic benchmark. The CE welfare change, g, represents the uniform percentage increase in consumption in every location and period under the baseline that would make households indifferent between the baseline and the counterfactual.

Under the full-reform scenario, we find that the reform unleashes a large initial reallocation of labor, with short-run ($t = 2$) declines in aggregate productivity (–7.91%) and output (–3.61%) as migration surges into high-amenity but initially lower-productivity destinations. Over time, agglomeration effects and amenity investments amplify location attractiveness and raise efficiency, leading to dramatic long-run gains. After a century, productivity and real output are 710% and 976% above baseline, corresponding to annualized growth rates of 2.10%/yr and 2.89%/yr. CE welfare rises sharply in both the short run (48.0%) and long run (56.0%, or 0.46 pp/yr), with more than half of the gain attributable to endogenous amenity improvements. The results highlight the strong complementarities between migration, productivity growth, and amenity dynamics when mobility restrictions are eliminated across all regions.

we find that g equals 46.9% after the first 20 years (model period $t = 2$), and rises to 66% after 100 years (model period $t = 10$), indicating substantial long-run welfare gains. Annualizing the long-run effect implies an average welfare growth rate of approximately 0.51% per year over a century.

Table 1: Aggregate Effects of *hukou* Reform Scenarios

	Productivity		Real Output		CE Welfare		Migration
Scenario	$t = 2$	$t = 10$	$t = 2$	$t = 10$	$t = 2$	$t = 10$	$t = 2$
Cumulative change (%)							
S0: Full abolition	-7.91	710.33	-3.61	975.61	48.00	52.87	83%
S1: High-barrier only	-6.63	41.02	-8.44	38.80	-0.37	0.18	72%
S2: Economic hubs only	3.73	206.80	19.30	170.72	0.59	1.36	64%
Annualized growth (%/yr) for $t = 10$							
S0: Full abolition	–	2.10	–	2.89	–	0.46	
S1: High-barrier only	–	0.34	–	0.33	–	-0.10	
S2: Economic hubs only	–	1.13	–	1.00	–	-0.09	

Notes: $t = 2$ and $t = 10$ correspond to 20 years and 100 years after reform, respectively. Annualized growth rates are computed from the cumulative percentage change relative to baseline over the century.

Under S1—High-barrier only, this targeted reform, *hukou* abolition is confined to 128 prefecture-level cities with baseline migration cost shares above 75%, covering about 36% of the population. These cities are disproportionately inland or resource-dependent (e.g., mining) and exhibit weaker productivity fundamentals. The initial reallocation directs labor toward these less dynamic destinations, producing short-run declines in productivity (–6.63%) and output (–8.44%), with CE welfare slightly negative (–0.37%). Although productivity and output gradually recover to 41.02% and 38.80% above baseline after a century, the corresponding annualized growth rates are modest at 0.34%/yr and 0.33%/yr. CE welfare remains essentially flat in the long run (0.18%, or –0.10 pp/yr growth), indicating that the efficiency gains are insufficient to offset congestion and modest amenity improvements.

Under S2—Economic hubs only, Here, *hukou* restrictions are lifted only in cities whose baseline GDP share in 2000 exceeded 1%, covering 46 cities and about 18% of the population. These hubs are generally coastal or highly urbanized, with strong agglomeration economies. As expected, the immediate effects are positive: productivity rises 3.73% and output 19.30% after two decades, as labor reallocates toward high-productivity centers. In the long run, productivity and output gains reach 206.80% and 170.72% above baseline, corresponding to annualized growth rates of 1.13%/yr and 1.00%/yr, respectively. However, the welfare effect turns slightly negative in annualized terms (–0.09 pp/yr) despite a small positive level change (1.36%), as congestion and cost pressures in these large cities erode much of the efficiency benefit.

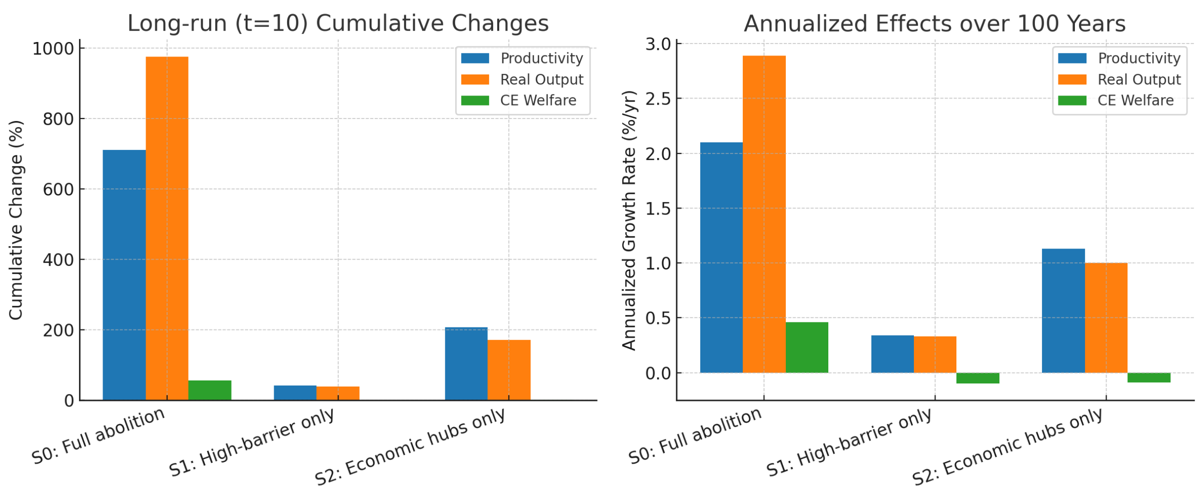


Figure 13: Net labor inflow by city: long-run effects under S0, S1, S2.

Under S0—Full abolition, hukou-related migration costs are removed nationwide. The short-run response is an initial drop in productivity (–7.9%) and output (–1.2%) as migration also flows toward lower-productivity but high-amenity areas. In the long run, however, the benefits of unrestricted labor mobility dominate: productivity grows at 2.10%/yr and output at 2.89%/yr over the century, with CE welfare improving by 0.46 pp/yr. This scenario delivers the largest and most persistent utility gains among all reforms, alongside the highest long-run growth in aggregate output and productivity.

6.3 Spatial Patterns of Adjustment

Based on the results in Table2, the abolition of the *hukou* system under Scenario S0 generates highly uneven spatial adjustments in both labor flows and welfare outcomes by $t = 10$.

On the migration side, the largest net inflow gains are concentrated in mid-sized inland and lower-tier coastal cities rather than in the megacities. Tongling, Ezhou, Ma'anshan, and Huainan—largely clustered in Anhui and surrounding provinces—each gain over 350,000 residents (measured in 10,000s). Several resource-based or industrial cities such as Hebi, Huaibei, and Wuhai also register substantial inflows. This pattern suggests that once institutional barriers are lifted, labor reallocates not only toward high-wage coastal hubs but also toward medium-size industrial centers with underutilized productive capacity.

The welfare results, however, tell a somewhat different story. The top CE welfare gains are dominated by cities starting from a relatively low baseline utility level. Jiayuguan, Huaiyin, and Anshun record increases exceeding 250% in consumption-equivalent welfare, reflecting large relative improvements from a low base. In several cases, such as Hebi and Luohe, cities appear in both the top inflow and top welfare gain lists—indicating that their gains are driven by both increased scale and improved economic conditions.

Conversely, the largest net outflows are concentrated in remote or less-connected regions, such as Ku'erle, Delingha, and Alxa, many located in Xinjiang, Qinghai, and Inner Mongolia. These regions face outmigration exceeding half a million people, reflecting both their geographic isolation and lower productivity in the counterfactual equilibrium.

In terms of welfare losses, the bottom-ranked cities are predominantly in either highly touristic coastal regions (e.g., Haikou, Sanya, Zhoushan) or large metropolitan areas such as Chengdu, Chongqing, and Wuhan. In tourist destinations, welfare declines are likely driven by congestion and increased costs without commensurate productivity gains. In large metros, the loss may be partly due to amenity dilution from rapid inflows exceeding infrastructure capacity, offsetting the benefits from scale economies.

Overall, these results highlight that removing *hukou* constraints creates a clear divergence between migration winners and welfare winners. Cities that attract the most migrants are not always those experiencing the largest proportional welfare gains, and some high-productivity urban centers may even see short-run welfare declines if congestion effects dominate. This underscores the importance of complementary urban policies to manage infrastructure, housing, and public services during large-scale migration adjustments.

Figures 14 show long-run spatial changes in net labor inflows (top panels), and deterministic welfare (bottom panels) for three counterfactual scenarios at $t = 10$.

In each map, the ten largest and ten smallest cities in each metric are highlighted, illustrating the starkly uneven spatial responses.

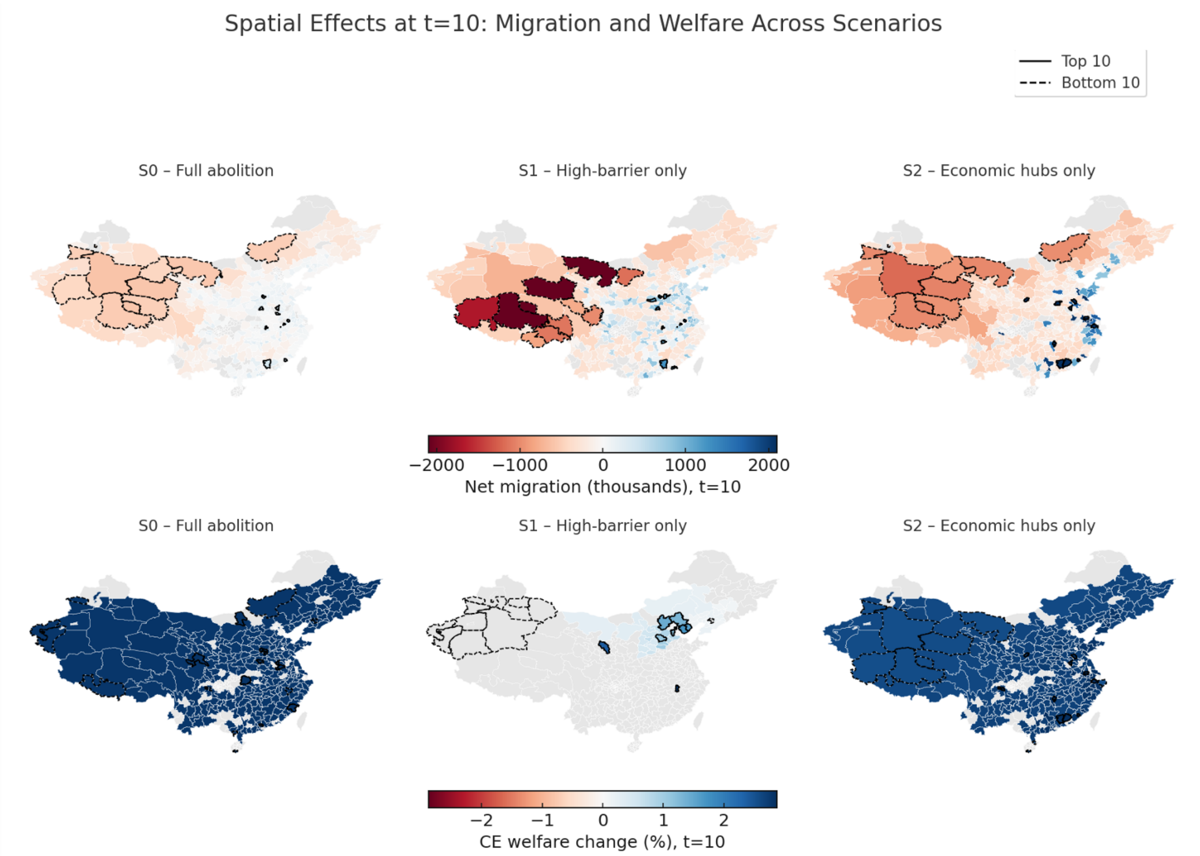


Figure 14: Net labor inflow by city and long-run effects under S0, S1, S2.

In a full abolition scenario, migration inflows are concentrated in mid-sized industrial and inland cities in Anhui, Henan, and Inner Mongolia, alongside some smaller coastal manufacturing hubs. Tongling (+430 k), Ezhou (+420 k), and Ma'anshan (+380 k) are among the top destinations, reflecting underutilized productive capacity. Outflows dominate in remote western prefectures such as Ku'erle (–610 k) and Delingha (–590 k), and in certain large metropolitan areas where congestion effects offset wage advantages.

Table 2: Largest movers and welfare changes under S0

Top 10 net inflow (t=10)				Top 10 welfare gain (t=10, CE %)		
City	$\Delta\bar{l}$	Rank	hukou cost (%)	City	CE %	Rank
Tongling	+43	1	75	Jiayuguan	394	1
Ezhou	+42	2	79	Huaiyin	354	2
Ma'anshan	+38	3	79	Anshun	257	3
Huaiyin	+37	4	80	Liangyungang	218	4
Hebi	+37	4	79	Hebi	174	5
Luohe	+35	6	79	Hunjiang	168	6
Huaibei	+34	7	70	Luohe	162	7
Heyuan	+34	8	78	Jicheng	158	8
Wuhai	+32	9	74	Heze	147	9
Xiamen	+31	10	70	Tongling	146	10

Bottom 10 net inflow ($t = 10$)			Bottom 10 welfare gain (t=10, CE %)		
City	$\Delta\bar{l}$	Rank	City	CE %	Rank
Ku'erle	-61	1	Haikou	-100	1
Delingha	-59	2	Sanya	-99	2
Alxa	-56	3	Zhoushan	-99	3
Naggu	-53	4	Jingdezhen	-90	4
Hailar	-53	4	Chengdu	-89	5
Xilingguole	-52	6	Chongqing	-85	6
Yushu	-50	7	Wuhan	-78	7
Yining	-49	8	Tongchuan	-70	8
Jiuquan	-46	9	Fuzhou	-59	9
Hetian	-44	10	Jiaxing	-58	10

Notes: Net inflow measured in units of 10,000 people. CE welfare computed with $\rho = 2$.

Welfare gains are largest in lower-baseline inland cities—Jiayuguan (+394%), Huaiyin (+354%), and Anshun (+257%), while tourist-dependent cities like Sanya (–99%) and high-density metros such as Chengdu (–89%) register losses.

In high-barrier only scenario, hukou restriction is lifted in 128 high-barrier cities (hukou cost share 75%), many of which are inland or resource-dependent. Net inflows are led by Hebi (+370 k) and Luohe (+350 k), but these destinations tend to have low initial productivity and limited agglomeration economies, driving short-run declines in aggregate productivity and welfare. Gains are localized—Hunjiang (+168%) and Jicheng (+158%)—while large swathes of the coast and major hubs remain unaffected.

If hukou is lifted in economic hubs only, it means liberalizing migration into 46 top-GDP cities (each contributing 1 % of baseline GDP) channels inflows toward Beijing, Shanghai, Shenzhen, and key provincial capitals. Short-run productivity and output rise due to scale economies, but welfare gains are uneven. Some secondary hubs in the liberalized group see modest improvements—e.g., certain Pearl River Delta cities record CE welfare gains of about +2.8%—yet several megacities face welfare declines from congestion and amenity dilution.

Across all scenarios, migration winners and welfare winners do not perfectly overlap. Cities attracting the largest inflows are not always those with the highest proportional welfare gains, and some high-productivity or high-density metros experience welfare losses from congestion and amenity dilution when inflows outpace the capacity of existing amenities and infrastructure in the model. These findings highlight the importance of complementary investments in public services, housing, and transportation when reforming migration institutions.

Productivity patterns broadly mirror, but do not perfectly align with, the migration outcomes. In Scenario S0, productivity gains concentrate in a mix of coastal economic hubs and select inland industrial centers that receive significant inflows, suggesting the model's agglomeration channel is strongest where baseline productivity is already high. Many high-barrier inland cities in S1 experience a short-run productivity drop despite large inflows, reflecting the dilution of average productivity when low-productivity workers move in faster than local technology or scale economies can adjust. By contrast, S2's targeted reform in economic hubs delivers immediate productivity boosts in the short run, but over the long horizon, congestion and reduced marginal returns to agglomeration slow these gains.

6.4 Mechanism Decomposition

The aggregate and spatial patterns documented above are the net outcome of multiple offsetting forces operating through the model's structural channels. To better understand these forces, we decompose the total consumption-equivalent (CE) welfare change into three distinct components, corresponding to the main terms in the welfare aggregator:

1. **Pure reallocation:** Labor reallocates across space; productivity $\tau_t(n)$ and amenities $A_t(n)$ are fixed at baseline levels.

2. **+ Productivity feedback:** Allow $\tau_t(n)$ to update via the agglomeration elasticity λ_1, amenities fixed.

3. **+ Amenity feedback:** Allow amenities to update via λ and χ, productivity fixed.

4. **Full model:** All channels active.

We compute the contribution of each channel by sequentially shutting down the other channels in counterfactual simulations, holding all else constant. The decomposition is performed for the long run ($t = 10$), and for S0 scenario.

Table 3: Mechanism Decomposition of Welfare (CE %, long run)

Scenario	CE Gain	Incremental (pp)	Share of Total (%)
Migration only (ϕ only)	18.60	+18.60	35.1
ϕ + Productivity feedback (λ_1 on; A fixed)	28.4	**+9.80**	18.5
$\phi + Amenity feedback(\lambda$, χ on; Z fixed)	43.97	**+15.57**	29.4
Interaction ($\phi \times Z \times A$)	–	**+8.9**	16.7
Full model($\phi + \lambda_1 + \lambda, \chi$)	**52.87**	–	**100.0**

Notes: "Incremental" is the additional CE relative to the previous row. Interaction = Full − (Migration+Prod+Amenity). Shares are each component's proportion of the Full model gain.

Under the full abolition of *hukou*, long-run welfare rises by about 53% in consumption-equivalent (CE) terms. A stepwise decomposition shows that **pure spatial reallocation** accounts for **18.6 pp** (35.1% of the total gain), allowing productivity to adjust via agglomeration adds 9.8 pp (18.5%), and amenity responses contribute the largest increment, 15.57 pp (29.4%). The remaining 8.9 pp (16.7%) reflects a positive interaction between productivity and amenity channels, indicating moderate complementarities.

These magnitudes reflect two structural features of the Chinese economy. First, amenity-driven migration is a prominent channel: local governments often invest heavily in public services, infrastructure, and urban livability—financed in part by land revenues—which directly raises location attractiveness and accounts for over half of the welfare gain in the simulation. Second, migration rates are already high in many regions, particularly among young and rural-to-urban workers. As a result, removing the *hukou* barrier alone—while holding productivity and amenities fixed—delivers a sizeable welfare gain, largely through improved matching and reduced mobility costs. Nevertheless, the largest welfare improvements arise when dispersion from lower migration costs is coupled with agglomeration forces from endogenous productivity and amenity growth, producing mutually reinforcing adjustments that reshape China's spatial economy.

6.5 Sensitivity and Robustness

We examine how the long-run welfare gains respond to variation in four key parameters: (i) the amenity–density elasticity (λ), (ii) the productivity–density elasticity (λ_1), (iii) the migration elasticity (γ), and (iv) the amenity–investment elasticity (χ).

The amenity–density elasticity λ captures congestion effects from high urban density; Chinese megacities such as Beijing and Shanghai face some of the world's highest population densities, making this channel potentially strong.

The productivity–density elasticity λ_1 reflects agglomeration spillovers, which are shaped by China's industrial clustering in manufacturing and services, especially in the coastal growth poles.

The migration elasticity γ measures the responsiveness of labor flows to utility differences; despite large wage gaps across regions, the *hukou* system and social network factors keep migration costs high, so plausible values of γ may be lower than in more mobile economies.

Finally, the amenity–investment elasticity χ reflects how effectively local governments convert land-lease revenues into public services and infrastructure. In China, this channel is salient because land finance accounts for a substantial share of local fiscal revenue, and urban amenity investments have been used to attract migrants and firms.

For each parameter, we simulate a "Low" and "High" value corresponding to a one-standard-error decrease or increase from the baseline estimate (or an equivalent calibrated range where standard errors are unavailable), holding all other parameters fixed.

Given the high-congestion baseline estimate for the amenity–density elasticity ($\lambda = -0.39$), we center sensitivity around this value: Low $\lambda = -0.50$, High $\lambda = -0.20$ (stress test: -0.60 to -0.05). For the productivity–density elasticity we use Low $\lambda_1 = 0.15$, High $\lambda_1 = 0.30$ (stress: 0.10–0.40). Migration elasticity varies at $0.5\gamma_0$ and $1.5\gamma_0$ (stress: $0.25\gamma_0$–$2\gamma_0$). Amenity–investment elasticity uses Low $\chi = 0.15$, High $\chi = 0.35$ (stress: 0.00–0.50).

Table 4: Parameter Ranges for Sensitivity and Stress Tests

Parameter	Baseline	Main Range	Stress Range
Amenity–density (λ)	-0.39	$[-0.50, -0.20]$	$[-0.60, -0.05]$
Productivity–density (λ_1)	0.21	[0.15, 0.30]	[0.10, 0.40]
Migration elasticity (γ)	2	$[\gamma_0, 1.5\gamma_0]$	$[0.25\gamma_0, 2.0\gamma_0]$
Amenity–investment (χ)	0.23	[0.15, 0.35]	[0.00, 0.50]

Table 5: Sensitivity of Long-run CE Welfare Gains (%, $t = 10$)

Scenario	Baseline	Low λ	High λ	Low λ_1	High λ_1
S0: Full abolition	52.87	48.67	56.4	49.8	58.8
Scenario	**Low γ**	**High γ**	**Low χ**	**High χ**	
S0: Full abolition	61	45.3	48.2	59.99	

Notes: "Low" and "High" denote a one-standard-error decrease or increase from the baseline parameter value (or an equivalent calibrated range where standard errors are unavailable). CE values are percentage changes relative to the baseline equilibrium. $t = 10$ corresponds to the long-run equilibrium after reform.

Table 5 shows that the qualitative ranking of scenarios is robust across plausible parameter ranges: S0 (full abolition) consistently yields the largest long-run welfare gain, followed by S1 (high-barrier only) and S2 (economic hubs only). Quantitatively, welfare gains are most sensitive to changes in the migration elasticity γ and the amenity–investment elasticity χ, reflecting the central role of migration responsiveness and public-investment-driven amenity improvements in driving aggregate welfare changes. In contrast, varying the amenity–density elasticity λ or productivity–density elasticity λ_1 within empirically reasonable bounds alters the magnitudes less substantially, indicating that congestion effects and agglomeration spillovers—while important—play a more moderate role in the aggregate outcome.

Across all parameter variations, the ordering of scenarios remains unchanged, suggesting that the main policy conclusions are not driven by fine-tuned parameter assumptions.

7 Conclusion and Discussion

his paper develops and quantifies a dynamic spatial equilibrium model with endogenous productivity, amenities, and migration frictions to evaluate the aggregate and spatial consequences of abolishing China's hukou system. The model captures three key mechanisms: the dispersion force from lower migration costs, and two agglomeration forces from productivity spillovers and amenity investment. Calibration to prefecture-level data reveals that full removal of hukou barriers generates large long-run gains in productivity, output, and welfare.

Under full reform, aggregate productivity and real output grow at 2.10% and 2.89% per year over the century, respectively, with consumption-equivalent (CE) welfare rising by 0.46 percentage points per year. Decomposition analysis attributes 35.1% of total welfare gains to pure spatial reallocation, 18.5% to productivity feedbacks, and 29.4% to amenity feedbacks, with the remainder due to positive interactions. Amenity-driven migration emerges as a particularly important channel in the Chinese context, reflecting local governments' capacity to finance and deliver public service improvements through land revenues.

Partial reforms produce more modest and in some cases negative welfare effects. Targeting only high-barrier cities redirects migration toward less dynamic locations, depressing short- and long-run productivity and output. Restricting reforms to economic hubs boosts efficiency but generates congestion and cost pressures that erode long-run welfare. Sensitivity checks show that results are most responsive to amenity–density elasticity and amenity–investment elasticity, while remaining robust to plausible variation in migration elasticity and productivity–density elasticity.

The findings highlight that the welfare impact of hukou reform hinges critically on the interaction between mobility and agglomeration forces. Removing barriers in isolation yields significant but incomplete gains; the largest benefits arise when greater labor mobility is accompanied by endogenous improvements in productivity and amenities. These results have broader relevance for economies where institutional frictions limit internal migration, underscoring the need for complementary policies that enhance both the economic and livability attributes of destination cities.

References

Ahlfeldt, G. M., Bald, F., Roth, D., & Seidel, T. (2020). *The stationary spatial equilibrium with migration costs.* Unpublished.

Allen, T., & Arkolakis, C. (2014). Trade and the topography of the spatial economy. *The Quarterly Journal of Economics*, *129*(3), 1085–1140.

Allen, T., & Donaldson, D. (2020). *Persistence and path dependence in the spatial economy* (Tech. Rep.). National Bureau of Economic Research.

Au, C.-C., & Henderson, J. V. (2006). How migration restrictions limit agglomeration and productivity in china. *Journal of development economics*, *80*(2), 350–388.

Banerjee, A. V., & Duflo, E. (2007). The economic lives of the poor. *Journal of economic perspectives*, *21*(1), 141–167.

Bento, A. M., Cropper, M. L., Mobarak, A. M., & Vinha, K. (2018). The effects of urban land use regulation on urban development: Evidence from brazil. *World Bank Economic Review*, *32*(2), 371–394. Retrieved from `https://doi.org/10.1093/wber/lhw026` doi: 10.1093/wber/lhw026

Bosker, M., Deichmann, U., & Roberts, M. (2018). Hukou and highways the impact of china's spatial development policies on urbanization and regional inequality. *Regional Science and Urban Economics*, *71*, 91–109.

Bryan, G., Chowdhury, S., & Mobarak, A. M. (2014). Underinvestment in a profitable technology: The case of seasonal migration in bangladesh. *Econometrica*, *82*(5), 1671–1748.

Bryan, G., & Morten, M. (2019). The aggregate productivity effects of internal migration: Evidence from indonesia. *Journal of Political Economy*, *127*(5), 2229–2268.

Cai, S., Caliendo, L., Parro, F., & Xiang, W. (2022). *Mechanics of spatial growth* (Tech. Rep.). National Bureau of Economic Research.

Caliendo, L., Dvorkin, M., & Parro, F. (2019). Trade and labor market dynamics: General equilibrium analysis of the china trade shock. *Econometrica*, *87*(3), 741–835.

Chan, K. W., & Buckingham, W. (2008). Is china abolishing the hukou system? *The China Quarterly*, *195*, 582–606.

Ciccone, A. (2002). Agglomeration effects in europe. *European economic review*, *46*(2), 213–227.

Combes, P.-P., Duranton, G., Gobillon, L., Puga, D., & Roux, S. (2012). The productivity advantages of large cities: Distinguishing agglomeration from firm selection. *Econometrica*, *80*(6), 2543–2594.

Comin, D., & Hobijn, B. (2010). An exploration of technology diffusion. *American economic review*, *100*(5), 2031–2059.

Cruz, J.-L., & Rossi-Hansberg, E. (2021). *The economic geography of global warming* (Tech. Rep.). National Bureau of Economic Research.

De Haas, H. (2010). The internal dynamics of migration processes: A theoretical inquiry. *Journal of ethnic and migration studies*, *36*(10), 1587–1617.

Desmet, K., Nagy, D. K., & Rossi-Hansberg, E. (2018). The geography of development. *Journal of Political Economy*, *126*(3), 903–983.

Desmet, K., & Rossi-Hansberg, E. (2014). Spatial development. *American Economic Review*, *104*(4), 1211–1243.

Eaton, J., & Kortum, S. (2002). Technology, geography, and trade. *Econometrica*, *70*(5), 1741–1779.

Fagerberg, J., Srholec, M., & Knell, M. (2007). The competitiveness of nations: Why some countries prosper while others fall behind. *World development*, *35*(10), 1595–1620.

Feyrer, J. (2019). Trade and income—exploiting time series in geography. *American Economic Journal: Applied Economics*, *11*(4), 1–35.

Head, K., & Mayer, T. (2014). Gravity equations: Workhorse, toolkit, and cookbook. In *Handbook of international economics* (Vol. 4, pp. 131–195). Elsevier.

Krugman, P. (1991). Increasing returns and economic geography. *Journal of political economy*, *99*(3), 483–499.

Lagakos, D., Marshall, S., Mobarak, A. M., Vernot, C., & Waugh, M. E. (2020). Migration costs and observational returns to migration in the developing world. *Journal of Monetary Economics*, *113*, 138–154.

Liu, Z. (2005). Institution and inequality: the hukou system in china. *Journal of Comparative Economics*, *33*(1), 133-157.

Meng, X. (2012). Labor market outcomes and reforms in china. *Journal of economic perspectives*, *26*(4), 75–102.

Monte, F., Redding, S. J., & Rossi-Hansberg, E. (2018). Commuting, migration, and local employment elasticities. *American Economic Review*, *108*(12), 3855–90.

Moretti, E. (2012). *The new geography of jobs*. Houghton Mifflin Harcourt.

Morten, M., & Oliveira, J. (2018). The effects of roads on trade and migration: Evidence from a planned capital city. *NBER Working Paper*, *22158*, 1–64.

Qi, H. (2015). *The labor share question in china* (Unpublished doctoral dissertation). University of Massachusetts Libraries.

Redding, S. J., & Rossi-Hansberg, E. (2017). Quantitative spatial economics. *Annual Review of Economics*, *9*, 21–58.

Roback, J. (1982). Wages, rents, and the quality of life. *Journal of political Economy*, *90*(6), 1257–1278.

Roberts, M., Deichmann, U., Fingleton, B., & Shi, T. (2012). Evaluating china's road to prosperity: A new economic geography approach. *Regional Science and Urban Economics*, *42*(4), 580–594.

Rosen, H. S. (1979). Housing decisions and the us income tax: An econometric analysis. *Journal of Public Economics*, *11*(1), 1–23.

Song, Y. (2014). What should economists know about the current chinese hukou system? *China Economic Review*, *29*, 200–212.

Stark, O., & Bloom, D. E. (1985). The new economics of labor migration. *The american Economic review*, *75*(2), 173–178.

Tabuchi, T., & Thisse, J.-F. (2002). Taste heterogeneity, labor mobility and economic geography. *Journal of Development Economics*, *69*(1), 155–177.

Tombe, T., & Zhu, X. (2019). Trade, migration, and productivity: A quantitative analysis of china. *American Economic Review*, *109*(5), 1843–1872.

Xu, Z., Chen, Y., & Li, M. (2015). Are chinese workers paid the correct wages? measuring wage underpayment in the chinese industrial sector, 2005-2010. *Review of Radical Political Economics*, *47*(3), 446–459.

Young, J. (2013). Markets, migrants and institutional change. *China's Hukou System: Markets, Migrants and Institutional Change*, 8–26.

A Appendix A: Derivations of Demand and Supply Equilibrium Conditions

This appendix provides detailed derivations of trade shares, price index, migration shares, initial period productivity, and exogenous amenities.

A.1 Derivation of the Trade Shares

Let $p_t^\omega(n,s)$ be the price of good ω produced in location n and purchased in location s at time t.

From the FOCs, we can have a closed form for $p_t^\omega(n,s)$:

$$p_t^\omega(n,n) = \iota^{-1} w_t(n) l_t(n)^{1-\iota} Z_t(n)^{-1} \tag{A.1}$$

where $Z_t(n) = Z_{t-1}(n)^\alpha \bar{l}_{t-1}(n)^{\lambda_1} \epsilon_t^\omega(n)$.

The price of good ω produced in location n and sold in location s is its locational price multiplied by the iceberg transportation cost.

$$\begin{aligned} p_t^\omega(n,s) &= \upsilon(n,s) p_t^\omega(n,n) \\ &= \upsilon(n,s)\iota^{-1} w_t(n) l_t(n)^{1-\iota} Z_t(n)^{-1}. \end{aligned} \tag{A.2}$$

The cumulative distribution function (CDF) of the prices for a good ω produced in location n and consumed in location s is:

$$\begin{aligned} &\mathbb{P}\left[p_t^\omega(n,s) \le p\right] \\ &\qquad = \mathbb{P}\left[\upsilon(n,s)\iota^{-1} w_t(n) l_t(n)^{1-\iota} Z_t(n)^{-1} \le p\right] \\ &\qquad = \mathbb{P}\left[\epsilon_t^\omega(n) \ge \frac{\upsilon(n,s)\iota^{-1} w_t(n) l_t(n)^{1-\iota} Z_t(n)^{-1}}{p}\right] \\ &\qquad = 1 - \mathbb{P}\left[\epsilon_t^\omega(n) \le \frac{\upsilon(n,s)\iota^{-1} w_t(n) l_t(n)^{1-\iota} Z_t(n)^{-1}}{p}\right] \\ &\qquad = 1 - exp\left\{-(\frac{\upsilon(n,s)\xi_t(n)}{p})^{-\delta}\right\}, \end{aligned} \tag{A.3}$$

where $\xi_t(n) = \upsilon(n,s)\iota^{-1} w_t(n) l_t(n)^{1-\iota} Z_t(n)^{-1}$

Then, the fraction of goods produced in location n and consumed in location s can be expressed as:

$$
\begin{aligned}
\pi_t(n,s) &= \mathbb{P}\left[p_t^{\omega}(n,s) \leq \min_{j\in N} p_t^{\omega}(n,j)\right] \\
&= \int_0^{\infty} \mathbb{P}\left[\min_{j\in N} p_t^{\omega}(n,j) \geq p\right] dG_{ns,t}(p) \\
&= \int_0^{\infty} \prod_{j\in N} \mathbb{P}\left[p_t^{\omega}(n,j) \geq p\right] dG_{ns,t}(p) \\
&= \int_0^{\infty} \prod_{j\in N} \left[1 - G_{nj,t}(p)\right] dG_{ns,t}(p).
\end{aligned} \tag{A.4}
$$

If we replace $G_{ns,t}(p)$ by its expression in (19), we will have

$$
\begin{aligned}
\pi_t(n,s) &= \int_0^{\infty} \prod_{j\in N} exp\left\{(\frac{\upsilon(j,s)\xi_t(j)}{p})^{-\delta}\right\} exp\left\{(\frac{\upsilon(n,s)\xi_t(n)}{p})^{-\delta}\right\} (\upsilon(n,s)\xi_t(n))^{-\delta}\, dp^{\delta} \\
&= (\upsilon(n,s)\xi_t(n))^{-\delta} \int_0^{\infty} \exp\left\{-\sum_j^N (\upsilon(j,s)\xi_t(j))^{-\delta} p^{\delta}\right\} dp^{\delta} \\
&= \frac{(\upsilon(n,s)\xi_t(n))^{-\delta}}{\sum_j^N (\upsilon(j,s)\xi_t(j))^{-\delta}}
\end{aligned} \tag{A.5}
$$

This gives us the equation (13) in the text.

A.2 Derivation of Migration Flows

Let $\Omega_t(n,j)^h$ be the fraction of individuals, with *hukou* h, relocating from location j to location n at time t. That an individual chooses to relocate to n means she can maximize her utility in n at time t. So we have:

$$
\begin{aligned}
\Omega_t(n,j)^h &= \mathbb{P}\left[\tilde{u}_t(n)\tilde{m}_t(n,j)^{-1}\epsilon_t(n) \geq \tilde{u}_t(s)\tilde{m}_t(s,j)^{-1}\epsilon_t(s)\right] \\
&= \mathbb{P}\left[\tilde{u}_t(n)\tilde{m}_t(n,j)\epsilon_t(n) \geq \tilde{u}_t(s)\tilde{m}_t(s,j)\epsilon_t(s), s \neq n\right] \\
&= \mathbb{P}\left[\frac{\epsilon_{s,t}}{\epsilon_{n,t}} \leq \frac{\tilde{u}_t(n)^{-1}\tilde{m}_t(n,j)^{-1}}{\tilde{u}_t(s)\tilde{m}_t(s,j)}\right]
\end{aligned} \tag{A.6}
$$

Since $\epsilon_{n,t}$ follow a Fréchet distribution and i.i.d. across locations and times, the ratio of the them also follows a distribution with a CDF given by:

$$
F(\epsilon \leq z) = \exp\left(-z^{-\gamma}\right)
$$

The probability of choosing location n rather than location s then can be written as:

$$
\mathbb{P}\left[\frac{\epsilon_t(s)}{\epsilon_t(n)} \leq \frac{\tilde{u}_t(n)\tilde{m}_t(n,j)}{\tilde{u}_t(s)\tilde{m}_t(s,j)}\right] = \exp\left(-\left(\frac{\tilde{u}_t(s)\tilde{m}_t(s,j)}{\tilde{u}_t(n)\tilde{m}_t(n,j)}\right)^{\gamma}\right)
$$

The overall migration probability for location n is the probability that location n is chosen over all alternatives $s \neq n$. This the the product of the probabilities for each alternative location:

$$\begin{aligned}\mathbb{P} &= \prod_{s\neq n} \exp\left(-\left(\frac{\tilde{u}_t(s)\tilde{m}_t(s,j)}{\tilde{u}_t(n)\tilde{m}_t(n,j)}\right)^{\gamma}\right) \\ &= \exp\left(-\sum_{s\neq n}\left(\frac{\tilde{u}_t(s)\tilde{m}_t(n,s)}{\tilde{u}_t(n)\tilde{m}_t(n,j)}\right)^{\gamma}\right)\end{aligned} \tag{A.7}$$

Then the probability $\Omega_t(n,j)^h$ can be written as:

$$\Omega_t(n,j)^h = \frac{(\tilde{u}_t(n)\tilde{m}_t(n,j))^{\gamma}}{\sum_s^N (\tilde{u}_t(s)\tilde{m}_t(n,s))^{\gamma}} \tag{A.8}$$

A.3 Initial Utility and Exogenous Amenity

The part of utility that does not depend on idiosyncratic amenity shocks, $\tilde{u}_t(s)$, is affected only by the characteristics of a location and is common for all its residents:

$$\tilde{u}_t(s) = \bar{A}_t(s)\bar{l}_t(s)^{\lambda}\frac{w_t(s)}{P_t(s)}$$

If we replace $P_t(s)$ by (18), we can have:

$$\tilde{u}_t(s) = \bar{A}_t(s)\bar{l}_t(s)^{\lambda}\frac{w_t(s)}{[\sum_s^N(\xi_t(n)v(n,s))^{-\delta}]^{-1/\delta}\left[\Gamma\left(\frac{1-\sigma}{\delta}+1\right)\right]^{\frac{1}{1-\sigma}}}$$

then, by rearranging the equation, we have:

$$\left[\Gamma\left(\frac{1-\sigma}{\delta}+1\right)\right]^{\frac{-\delta}{1-\sigma}}\sum_n^N(\xi_t(n)v(n,s))^{-\delta} = (\frac{1-\iota}{\iota})^{\chi}\left[\frac{\bar{A}_t(s)}{\tilde{u}_t(s)}\right]^{-\delta}\bar{l}_t(s)^{-\lambda\delta}w_t(s)^{-\delta(1+\chi)}. \tag{A.9}$$

Combining (A.5) and (A.9), the probability of goods produced in n and consumed in s becomes:

$$\pi_t(n,s) = (v(n,s)\xi_t(n))^{-\delta}\left[\Gamma\left(\frac{1-\sigma}{\delta}+1\right)\right]^{\frac{-\delta}{1-\sigma}}\left[\frac{\bar{A}_t(s)}{\tilde{u}_t(s)}\right]^{\delta}\bar{l}_t(s)^{\lambda\delta}w_t(s)^{\delta} \tag{A.10}$$

With (A.10), the trade balance condition can be expanded as:

$$w_t(n)H(n)\bar{l}_t(n)\xi_t(n)^{\delta} = \left[\Gamma\left(\frac{1-\sigma}{\delta}+1\right)\right]^{\frac{-\delta}{1-\sigma}}\sum_s^N\left[\frac{\bar{A}_t(s)}{\tilde{u}_t(s)}\right]^{\delta}H(s)\bar{l}_t(s)^{1+\lambda\delta}w_t(s)^{1+\delta}\upsilon_t(n,s)^{-\delta} \tag{A.11}$$

Plug in $\xi_t(n)$ and rearrange (A.11):

$$\begin{aligned}&w_t(n)^{1+\delta}H(n)\bar{l}_t(n)^{1+(1-\iota)\delta}\tilde{z}_t(n)^{-\delta} \\ &= \left[\Gamma\left(\frac{1-\sigma}{\delta}+1\right)\right]^{\frac{-\delta}{1-\sigma}}\sum_s^N\left[\frac{\bar{A}_t(s)}{\tilde{u}_t(s)}\right]^{\delta}H(s)\bar{l}_t(s)^{-\lambda\delta}w_t(s)^{-\delta(1+\chi)}\upsilon_t(n,s)^{-\delta}\end{aligned} \tag{A.12}$$

B Appendix B: Data Description

B.1 Data Sources and Description

This appendix provides a detailed overview of the data sources and methodological approaches used in this study to calibrate the dynamic spatial equilibrium model and conduct the analysis of China's *hukou* system.

China Census Data: the primary data sources for this study include the China Census Data for the years 2000, 2010, and 2020, provided by the National Bureau of Statistics of China (NBS). The China Census is conducted every ten years and offers comprehensive demographic and socio-economic information across Chinese prefectures, including population size, age structure, education levels, employment status, and hukou status. These datasets were crucial for capturing changes in population distribution and labor mobility over time. The census data was accessed through the official NBS website, with some historical data obtained from academic institutions with special access to NBS archives.

City-Level Statistical Yearbooks: additional economic data was sourced from the City-Level Statistical Yearbooks, which are published annually by the NBS and various provincial and municipal statistics bureaus. These yearbooks provide detailed economic data for each prefecture-level city in China, including GDP per capita, total GDP, industrial output, disposable income, wages, and land area. This information was essential for understanding the economic context in which migration decisions are made and for calibrating the model to reflect regional economic disparities. The statistical yearbooks were accessed through the provincial statistics bureaus' online databases.

China Migrants Dynamic Survey (CMDS): conducted annually since 2010 by the National Health Commission of the People's Republic of China, provided detailed information on China's migrant population. The CMDS data includes variables related to socio-economic status, health, education, employment, and migration history, which were integral to understanding the characteristics and behavior of the migrant population under the hukou system. The CMDS data was obtained through the official website of the CMDS, which requires specific authorization for years and regions.

Integrated Public Use Microdata Series (IPUMS): Migration flow data from the initial period was derived from IPUMS International, which provides harmonized data from national censuses worldwide, including the 2000 China National Population Census. The dataset used in this study is a 1% sample of the total census population, offering a rich source of microdata on demographics, housing, employment, and migration. The IPUMS data was accessed through the IPUMS website, which provides public access to researchers upon registration.

In addition to demographic and economic data, climate variables such as average temperature, humidity, and sunlight hours were obtained from the China Meteorological Data Service Center, operated by the China Meteorological Administration. These variables were used as proxies for exogenous amenities in the model, influencing regional attractiveness. The climate data was accessed through the China Meteorological Data Service Center's official website, with some datasets requiring subscriptions.

Topographical and geographical data, including terrain relief and the presence of water bodies, were sourced from the National Geographic Information Center, affiliated with the Chinese Academy of Sciences. These data were used to measure exogenous amenities and their influence on population distribution. The Relief Degree of Land Surface dataset provided insights into the physical geography affecting regional development and settlement patterns. Data from the National Geographic Information Center was accessed through academic collaborations with the Chinese Academy of Sciences.

To estimate bilateral trade costs between cities, the ArcGIS OD Cost Matrix Tool, part of Esri's ArcGIS software suite, was employed. This tool calculates the fastest paths through highways, roads, and railroads, allowing for the estimation of spatial frictions in the model by assessing the time and distance-based costs associated with transportation across regions. ArcGIS software was accessed through Esri's official website, with institutional licenses provided by Georgetown.

B.2 Data Preparation

The methodology involved a rigorous calibration process, ensuring that the model accurately reflects the economic and demographic realities of Chinese prefectures from 1990 to 2020. Data processing and harmonization were critical to this effort. For census data, raw data from the NBS was processed to extract relevant variables such as population size, *hukou* status, and employment. This data was harmonized across different years to ensure consistency in definitions and measurement units. Economic variables from the city-level statistical yearbooks were adjusted for inflation and converted to real terms where necessary. In cases where data inconsistencies arose due to changes in administrative boundaries or definitions, interpolation or imputation techniques were applied.

Survey data from the CMDS was cleaned and weighted to reflect the representativeness of the sample, with missing data points addressed using multiple imputation methods. This ensured that the analysis captured the full diversity of the migrant population.

Spatial analysis was conducted using GIS tools to map the distribution of population, amenities, and economic output across Chinese prefectures. These maps provided a visual representation of the regional disparities that the model seeks to explain. The ArcGIS OD Cost Matrix Tool was instrumental in estimating trade costs, which were then integrated into the model as spatial frictions influencing migration and economic decisions.

The robustness of the model's predictions was tested through sensitivity analysis, where key parameters were varied within plausible ranges to observe their impact on the results. Alternative model specifications were also explored to ensure that the findings were not driven by specific assumptions. The counterfactual analysis, simulating the scenario of *hukou* abolition, was conducted by removing institutional constraints on labor mobility and observing the resulting changes in population distribution, productivity, and economic output. The dynamic-hat algebra method was used to solve the model under these new conditions, allowing for a direct comparison of baseline and counterfactual outcomes.

C Appendix C: Estimation of Parameters

C.1 Elasticities of Amenities

This section presents the technical estimation strategy used to identify the elasticity of amenities with respect to population density (λ) and amenity investment (χ). In the model, endogenous amenity levels are defined as:

$$A_t(n) = \bar{A}(n) \cdot \bar{l}_t(n)^{\lambda} \cdot I_t(n)^{\chi}, \tag{C.1}$$

where $A_t(n)$ is the aggregate amenity level in region n at time t, $\bar{A}(n)$ is the exogenous amenity level, $\bar{l}_t(n)$ is the total population, and $I_t(n)$ is the per capita investment in amenities. Log-linearizing yields:

$$\log A_t(n) = \log \bar{A}(n) + \lambda \log \bar{l}_t(n) + \chi \log I_t(n). \tag{C.2}$$

This specification reflects the dual nature of population density's effect on amenities. On one hand, higher population density can enhance the availability and quality of public amenities. Densely populated areas can support a broader range of infrastructure and services—such as transportation networks, educational institutions, health care facilities, and cultural venues—due to economies of scale and higher public investment returns. As emphasized by Brueckner and Largey (2008), the concentration of people in urban areas often facilitates the provision of diverse and high-quality amenities that improve residents' quality of life. In this way, agglomeration can be self-reinforcing: people are drawn to places with better amenities, which in turn can support further improvements.

On the other hand, excessive density can strain local resources and infrastructure. When urban population growth exceeds the capacity of public services and the environment, the very amenities that attracted people can degrade. Congestion, long wait times, overcrowded public spaces, and environmental degradation are frequent symptoms. Cavailhès et al. (2007) and Cohen (2006) document how such negative externalities can erode quality of life in overly dense urban areas. This nonlinearity underscores the complex interaction between density and amenity provision, where marginal increases in population can either enhance or deteriorate local amenities depending on the context.

Constructing the Amenity Index. To measure $A_t(n)$, we construct a composite amenity index using Principal Component Analysis (PCA) based on a wide range of observable indicators. These indicators cover five dimensions:

- **Environmental Quality:** industrial SO_2 emissions (tons/km^2), green coverage rate, arable land (thousand hectares), and noise surface area.
- **Transportation Infrastructure:** paved road area per capita, public transit vehicles per 10,000 persons, taxis per 10,000 persons.
- **Healthcare:** hospital beds and physicians per capita.
- **Education:** public expenditure per capita, schools per capita, and full-time teachers per capita.

- **Cultural Environment:** number of cinemas and library books per capita.

After normalizing these indicators, we apply PCA and retain the first seven principal components as the amenity index $A_t(n)$.

Measuring Amenity Investment. Amenity investment $I_t(n)$ is proxied by city maintenance and environmental protection expenditures, normalized by total population in each city. Data is obtained from China city statistical yearbooks for the year 2000.

Exogenous Amenities $\bar{A}(n)$. To isolate the endogenous component, we regress $\log A_t(n)$ on exogenous geographical fundamentals, including:

- **Topography:** topographic relief index.
- **Climate:** average temperature and humidity.
- **Water resources:** water body coverage (from National Geographic Information Center).

The residual from this regression represents the endogenous amenity component, allowing us to estimate the elasticities.

Estimation via GMM. We estimate the following equation using Generalized Method of Moments (GMM):

$$\log A_t(n) - \log \hat{\bar{A}}(n) = \lambda \log \bar{l}_t(n) + \chi \log I_t(n) + \epsilon_t(n), \tag{C.3}$$

where $\hat{\bar{A}}(n)$ denotes the predicted exogenous amenity level from the geography-based regression. To address concerns of endogeneity (e.g., population density being jointly determined with amenities), we use deep lags of population and investment as instruments. Standard errors are clustered at the regional level.

Results. The GMM estimation yields:

$$\hat{\lambda} = -0.39,$$

$$\hat{\chi} = 0.23.$$

The negative λ confirms the presence of congestion effects, while a positive χ indicates that land-rent-funded amenity investments significantly improve regional amenity levels. These values are used in the model calibration and sensitivity analysis.

C.2 Agglomeration Effects in Productivity Dynamics

This section provides technical details regarding the estimation of the agglomeration elasticity λ_1 in regional productivity dynamics, as well as the persistence parameter α. The estimation equation is:

$$\log(\text{GDPpc}_{it}) = \lambda_1 \log(\text{PopDensity}_{i,t-1}) + \alpha \log(\text{GDPpc}_{i,t-1}) + \epsilon_{it}, \tag{C.4}$$

This specification captures how past population density contributes to current productivity through agglomeration channels, while controlling for productivity persistence.

Given concerns about endogeneity of both lagged GDP per capita and population density, several estimation strategies were employed:

- **Fixed Effects (FE)** regression, controlling for unobserved time-invariant regional heterogeneity.
- **Instrumental Variables (IV)** estimation using deep lags (e.g., $L_5.\log(\text{GDPpc})$, $L_5.\log(\text{PopDensity})$) as instruments.
- **Dynamic Panel GMM** estimators with lagged levels as instruments.
- **Nonlinear GMM**

The following table summarizes the estimated values of λ_1 and α across methods:

Table 6: Estimation of Agglomeration Elasticity in Productivity Dynamics

Method	λ_1 (Agglomeration Elasticity)
Fixed Effects (Region and Year)	0.16
IV: L5.lgdppc+ Year FE	0.39
IV: L5.lgdppc and L5.popdens	0.19
IV: L5.lgdppc and L5.popdens + Year FE	0.34
`xtdpdgmm`, lag(2)	0.21
`gmm`, lag(1) instruments	0.021
`gmm`, lag(2) and (3) instruments	0.014

Across the methods, λ_1 ranges from 0.014 to 0.39. Estimates from IV using deep lags with year fixed effects yield relatively high values, suggesting stronger agglomeration externalities. Meanwhile, estimates using GMM with limited lag depth show weaker effects, potentially due to instrument weakness or multicollinearity.

The main analysis adopts $\lambda_1 = 0.21$, which is robust across multiple specifications and closely aligned with estimates found in the literature. For example, Combes, Duranton, Gobillon, Puga, and Roux (2012) report agglomeration elasticities ranging from 0.03 to 0.08 in developed countries using firm-level data, while in rapidly urbanizing developing economies, the literature often finds larger effects. For instance, Ciccone (2002) document elasticities in the range of 0.10 to 0.25 in European city data. Similarly, Bento, Cropper, Mobarak, and Vinha (2018) estimates values between 0.12 and 0.20 for Brazilian urban areas.

Given that this study focuses on Chinese prefecture-level data during a period of rapid urbanization and structural transformation, the selected value of $\lambda_1 = 0.18$ lies within the upper range of international estimates but remains consistent with existing empirical evidence on agglomeration economies in emerging markets.

D Appendix D: Approximation of Equilibrium Allocations Period 0

This section outlines the numerical procedure used to estimate spatial equilibrium allocations at the initial stage of the model (period zero), where population distributions, wages, amenities, and trade flows are jointly determined. Given the nonlinear and interdependent nature of these variables, I implement a fixed-point iterative approach to recover equilibrium values that are consistent with observed data from the year 2000 (period $t = 0$).

The estimation focuses on recovering three central objects in equilibrium: (i) the exogenous amenity component $a(n)$, (ii) the endogenous productivity level, (iii) the population density.

Initialization and Structure I initialize the ratio of exogenous amenity component $a(n)$ over the deterministic utility $\tilde{u}_t(n)$ across regions, and set the initial guess for population allocation $l_0(n)$ as a uniform distribution scaled by the total population. The key equations governing the system include: i) endogenous productivity $Z_0(n)$, which depends on wages, land area, and exogenous amenities; ii) a labor demand equation from firms, derived from equilibrium price indices and trade flows. A consistency condition requiring that the population allocation implied by these equations matches the observed population distribution.

Iterative Algorithm The solution is obtained via a two-layer fixed-point iteration:

- Inner Loop 1 (Exogenous Amenities $a(n)$): Given initial guesses of $l(n)$, I compute the implied trade-based price indices using the bilateral trade cost matrix and production parameters. These enter into an expression for local expenditure, allowing recovery of the $a(n)$ terms that rationalize trade flows and price levels. Convergence is assessed by the squared deviation between consecutive $a(n)$ vectors.

- Inner Loop 2 (Population Allocation $l(n)$): With $a(n)$ and implied $Z(n)$, I solve for labor allocations that satisfy labor demand, using the functional form derived from the model. The convergence criterion here is the squared deviation between the updated $l(n)$ and the prior guess.

- Outer Loop (Matching to Data): After both inner loops converge, I evaluate the distance between the model-implied population distribution and the observed data. This outer error is defined as the normalized deviation:

$$\text{Error}_{\text{outer}} = \frac{\|H(n) \cdot (l(n) - l_0(n))\|}{\|H(n) \cdot l_0(n)\|}.$$

  The outer loop continues until either the error falls below a specified tolerance threshold or the improvement in the outer error becomes negligible over successive iterations (i.e., convergence stalls).

E Appendix E: Consumption-Equivalent Welfare Calculation

To evaluate the welfare implications of abolishing the *hukou* system, we compute the present discounted value (PDV) of deterministic utility along the transition path, following an Aiyagari-style aggregation. In each period t, let $\tilde{u}_t(n)$ denote the deterministic component of indirect utility in location n—that is, the component implied by equilibrium prices, wages, amenities, and migration frictions, excluding the i.i.d. taste shocks that rationalize migration probabilities in the discrete-choice migration block. Aggregate deterministic utility in period t is given by:

$$U_t = \sum_{n=1}^{N} \tilde{u}_t(n) \cdot H(n) \cdot \bar{l}_t(n),$$

where $H(n)$ is the land area of location n and $\bar{l}_t(n)$ is the equilibrium labor density.

The PDV of deterministic utility is then:

$$\text{PDV}_U = \sum_{t=1}^{T} \beta^t U_t,$$

where $\beta \in (0, 1)$ is the intertemporal discount factor. We compute PDV_U for both the baseline (with *hukou*) and the counterfactual (abolishing *hukou*, i.e., $\phi(h, n) = 1$ for all (h, n)), and measure the welfare change as:

$$\Delta\text{PDV}_U = \frac{\widehat{\text{PDV}}_U - \text{PDV}_U}{\text{PDV}_U}.$$

To express this welfare change in more interpretable terms, we convert it into a consumption-equivalent (CE) welfare measure. Assuming per-period deterministic utility is homothetic in consumption with CRRA curvature ρ and separable from amenities, scaling consumption by a constant factor $(1+g)$ multiplies period utility by $(1+g)^{1-\rho}$. This implies:

$$\frac{\widehat{\mathrm{PDV}}_U}{\mathrm{PDV}_U} = (1+g)^{1-\sigma} \quad \Rightarrow \quad g = \left(\frac{\widehat{\mathrm{PDV}}_U}{\mathrm{PDV}_U}\right)^{\frac{1}{1-\sigma}} - 1, \tag{E.1}$$

where $100 \times g$ is reported as the CE welfare gain in percentage terms.

In our benchmark calibration, we set $\rho = 2$ following standard macroeconomic practice. In this case, the CE welfare gain simplifies to:

$$g = \frac{\mathrm{PDV}^{\mathrm{cf}}}{\mathrm{PDV}^{\mathrm{base}}} - 1. \tag{E.2}$$

Example – Full Hukou Abolition Scenario:

Table 7: Consumption-Equivalent Welfare Gains under Full Hukou Abolition

Horizon	$\mathrm{PDV}_{\mathrm{base}}$	$\mathrm{PDV}_{\mathrm{cf}}$	CE Welfare Gain g
$t = 2$	125,467,755.2	187,811,921.0	+49.7%
$t = 10$	88,592,675.1	147,234,635.9	+66.2%

These figures indicate that households would require nearly 50% higher permanent consumption in the baseline to match their welfare in the counterfactual by $t = 2$, and over 66% higher by $t = 10$, highlighting the long-run benefits of removing institutional migration barriers.

This approach isolates the welfare effects of *hukou* abolition that operate through deterministic changes in consumption and amenities, abstracting from transitory or idiosyncratic migration shocks. It ensures that the reported gains reflect permanent, economy-wide improvements in location-specific economic conditions rather than stochastic variation.

F Prefutures in China

China's administrative structure is hierarchical, with the country divided into several levels of government, each with its own jurisdictions and responsibilities. One of the key administrative divisions is the prefecture, an important unit that lies between the provincial and county levels. Prefectures serve as crucial links in the governance chain, managing large areas that typically encompass multiple counties, districts, and even cities. These prefectures play a vital role in implementing policies, managing resources, and serving as hubs for economic and social development.

Prefecture-level cities are particularly significant in China's urban hierarchy. They usually include a central urban area and its surrounding rural regions, with the central city acting as the administrative and economic heart. The number of prefecture-level cities in China has evolved over time due to administrative changes such as mergers, abolitions, and the establishment of new cities. As of 2000, there were 333 prefecture-level units in China, but after various adjustments, this number was streamlined to 313, which are the focus of the study in this paper. These prefectures are diverse, encompassing economically advanced coastal cities, less-developed inland regions, and a wide range of geographic, cultural, and economic environments.

Overlaying this administrative landscape is China's *hukou* system, a household registration policy that has been in place since the 1950s. The hukou system categorizes Chinese citizens based on their place of residence and birth, effectively tying individuals to a specific location and controlling their access to various social services such as education, healthcare, and housing. Initially designed to manage rural-to-urban migration and ensure social stability, the *hukou* system has become a major determinant of social mobility in China.