Internal Migration Restrictions, Aggregate Productivity, and Spatial Growth

Yanbin Tracy Xu *

September, 2025

Abstract

China's *hukou* system links access to public services and social benefits to place of registration, creating large barriers to internal labor mobility. This paper develops a dynamic spatial general equilibrium model to quantify how relaxing institutional migration frictions affects labor reallocation, productivity growth, and welfare. The model features endogenous agglomeration economies and amenity formation driven by both population density and public investment, and is calibrated to prefecture-level data.

The results show that removing mobility barriers triggers substantial reallocation toward high-productivity regions, followed by sustained aggregate productivity gains as agglomeration forces amplify regional advantages. Long-run welfare improvements are significantly larger when migration reform is complemented by endogenous improvements in local public goods and urban amenities, highlighting complementarities between labor mobility and place-based investment. In contrast, partial or spatially uneven reforms can generate muted or even negative welfare effects when they redirect migration toward lower-growth regions.

Beyond the Chinese context, the analysis provides quantitative evidence on how internal migration frictions shape aggregate efficiency, regional inequality, and the transmission of structural reforms in large economies. The findings offer broader lessons for policymakers in emerging markets facing rapid urbanization and institutional barriers to labor mobility.

JEL Codes: R23, R11, R13, O18, R32

*Postdoctoral Researcher, Knight-Georgetown Institute, 37th and O Street NW, Washington DC 20057, USA (email:yx76@georgetown.edu)

1 Introduction

Migration is a fundamental driver of economic development, shaping labor supply, productivity, and the spatial distribution of economic activity. When workers can move freely to where their skills are most productive, reallocation raises wages, boosts aggregate productivity, and supports more efficient spatial distribution of economic activity (Eaton & Kortum, 2002; Desmet, Nagy, & Rossi-Hansberg, 2018; Monte, Redding, & Rossi-Hansberg, 2018; Moretti, 2012). In practice, however, mobility is rarely frictionless. In developing economies, migration is not only shaped by financial, social, and informational barriers (Bryan & Morten, 2019; Young, 2013), but also by formal institutional restrictions that explicitly limit internal mobility.

China's *hukou* system is one of the most prominent examples of such institutional barriers. By tying access to public services–including education, healthcare, and housing subsidies–to a person's place of registration, *hukou* raises the effective cost of migration and distorts the allocation of labor. A rich literature documents its effects on rural-to-urban migration and wage gaps (Chan & Buckingham, 2008; Song, 2014; Tombe & Zhu, 2019; Roberts, Deichmann, Fingleton, & Shi, 2012; Bosker, Deichmann, & Roberts, 2018), but most studies emphasize static outcomes. Less is known about how *hukou* frictions shape the dynamic trajectory of productivity, inequality, and welfare over time.

This paper addresses that gap by developing a dynamic spatial equilibrium model in which *hukou* acts as a persistent friction in forward-looking migration. Productivity evolves endogenously through agglomeration, and amenities respond to both congestion and rent-financed investment. Empirically, I estimate key parameters from Chinese prefecture-level data, showing that migration is highly responsive to wages, that amenities are shaped by both congestion and investment, and that local productivity growth depends on agglomeration.

Quantitative analysis reveals that full *hukou* abolition would generate large and lasting welfare and productivity gains by unlocking migration-driven agglomeration and amenity improvements. By contrast, partial reforms yield only modest benefits and can exacerbate inequality as migration flows concentrate in advantaged cities.

The paper makes three contributions. First, it extends dynamic spatial equilibrium models by explicitly embedding institutional mobility frictions. Second, it shows how *hukou* fundamentally alters long-run growth trajectories, not just static allocations. Third, it provides new policy insight: comprehensive reform delivers dynamic complementarities, whereas partial reform risks widening inequality. Relative to Tombe and Zhu (2019), who treat productivity growth as exogenous, and Cai, Caliendo, Parro, and Xiang (2022), who study the general mechanics of spatial growth, this paper shows how a specific institutional distortion reshapes those mechanics with significant policy implications.

More broadly, the findings highlight how migration frictions can distort the dynamic gains from spatial reallocation. This has implications not only for China, but also for understanding place-based policies in other developing economies.

The rest of the paper is organized as follows. Section 2 describes *hukou* policy and the migration constraints it creates. Section 3 presents the dynamic spatial equilibrium model. Section 4 discusses the data and calibration. Section 5 reports quantitative results and counterfactual analyses. Section 6 concludes with policy implications.

2 Institutional Context and Empirical Motivation

2.1 *Hukou* as an Institutional Friction

China's *hukou* (household registration) system, formally established in 1958, remains one of the most important institutional frictions shaping internal migration in a major economy. By tying access to local public services—including education, healthcare, subsidized housing, and pensions—to registration status, *hukou* effectively raises the cost of moving across prefectures. Even after reforms introduced temporary residency permits and, in some cities, point-based eligibility for *hukou* transfer, migrants without local registration remain excluded from many benefits. These restrictions discourage permanent settlement and create a persistent wedge between where workers are most productive and where they can fully access public services.

For the purpose of this paper, *hukou* is best understood as a recurring, location-specific migration cost. It is not a one-time expense, but a recurring institutional friction that shapes short-term mobility and long-run spatial development.

2.2 Migration Patterns under *Hukou*

Despite these restrictions, China has experienced the largest internal migration in human history. By 2020, nearly 380 million people–27% of the population—lived outside their registered *hukou* location, forming the so-called *floating population.* Flows have been highly asymmetric. Coastal metropolises such as Shanghai, Shenzhen, and Guangzhou have attracted massive inflows, while interior provinces such as Henan, Anhui, and Sichuan remain persistent outflow regions.

Migrants are disproportionately young, often with lower formal education than local urban residents, and concentrated in labor-intensive industries such as construction, manufacturing, and services. Their participation in the labor market has been vital to China's growth. Yet, the lack of *hukou* means they frequently occupy informal jobs and face barriers to long-term integration.

These descriptive patterns underscore *hukou*'s central role in shaping China's spatial economy. Large-scale flows occur despite high barriers, but their scale and direction are systematically conditioned by institutional rules that determine who can move there and under what conditions.

2.3 Emiprical Motivation

To illustrate the economic significance of *hukou* frictions, this subsection presents a set of stylized facts that motivate the dynamic model in Section 3.

Fact 1. Wage penalties. Migrants without local *hukou* earn substantially less than comparable *hukou* residents. Cheng, Hu, and Li (2020) estimate a 22% monthly and 32%

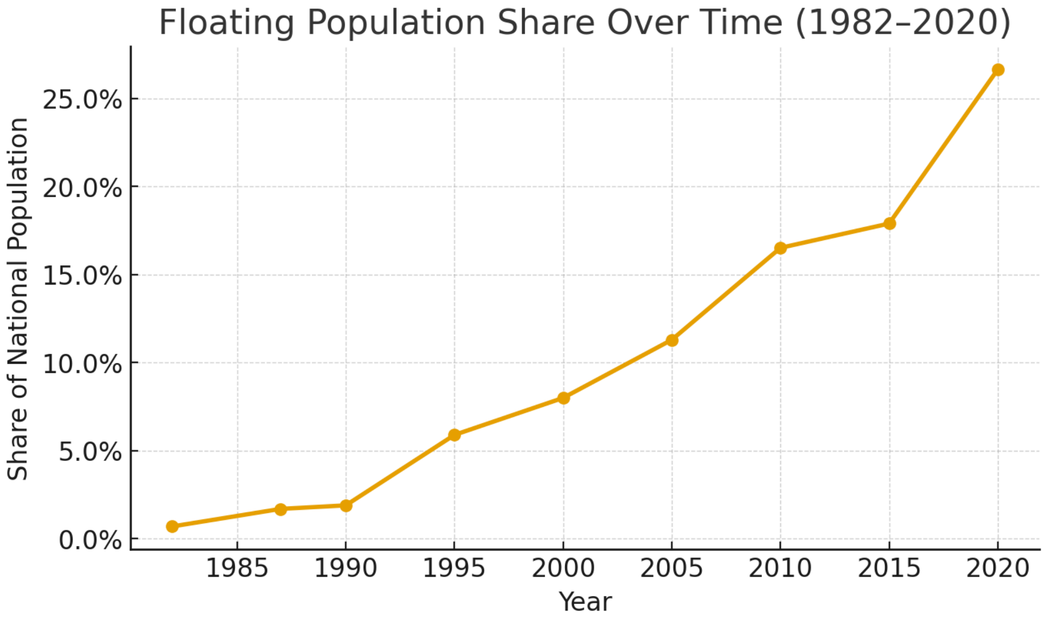

Figure 1: **Floating Population.**

hourly wage gap, even as rural migrants work 5.6% longer hours. A meta-analysis of 75 studies confirms that *hukou* status systematically depresses wages (Ma, Li, & Iwasaki, 2024), and survey evidence from HKUST (Wu & Zhang, 2015) finds that migrants earn about 17% less than locals after controlling for education, experience, and demographics.

Although much of this literature focuses on rural-to-urban migrants, *hukou* barriers extend to all individuals living outside their registration locality. The official category of floating population—nearly 27% of the national total in 2020—is defined as residents without local *hukou*, regardless of whether their origin is rural or urban. Since reforms after 2016 formally abolished the rural–urban *hukou* distinction but maintained locality-based restrictions, these wage penalties should be interpreted as a conservative measure of the broader disutilities faced by all cross-prefecture migrants. In other words, while rural–urban wage gaps are especially visible, the underlying mechanism is general: *hukou* exclusion generates a persistent utility wedge between locals and non-locals.

Fact 2. Limited access to services in destination cities. National survey data consistently show that migrants face barriers to public services—especially in Tier-1 cities[1] where inflows are largest. Migrant children are less likely to enroll in local public schools, while adults are less likely to receive employer-linked health insurance or qualify for subsidized housing. Thus, even as Tier-1 cities attract a growing share of the floating population, *hukou* restrictions prevent many of these households from fully integrating into local welfare systems.

In the model, these service gaps are captured as a persistent disutility wedge: migrants

[1] I follow the GDP Tier categorization of the prefectures by South China Morning Post. All first-tier cities have a GDP over USD300 billion; the second-tier includes cities with a GDP between USD68 billion and USD299 billion; the third-tier cities have a GDP between USD18 billion and USD67 billion; the fourth-tier cities have a GDP below USD17 billion. The detailed list can be found here: https://multimedia.scmp.com/2016/cities/

can work in Tier-1 destinations, but without local *hukou*, their effective utility remains below that of local residents, even when wages are equal.

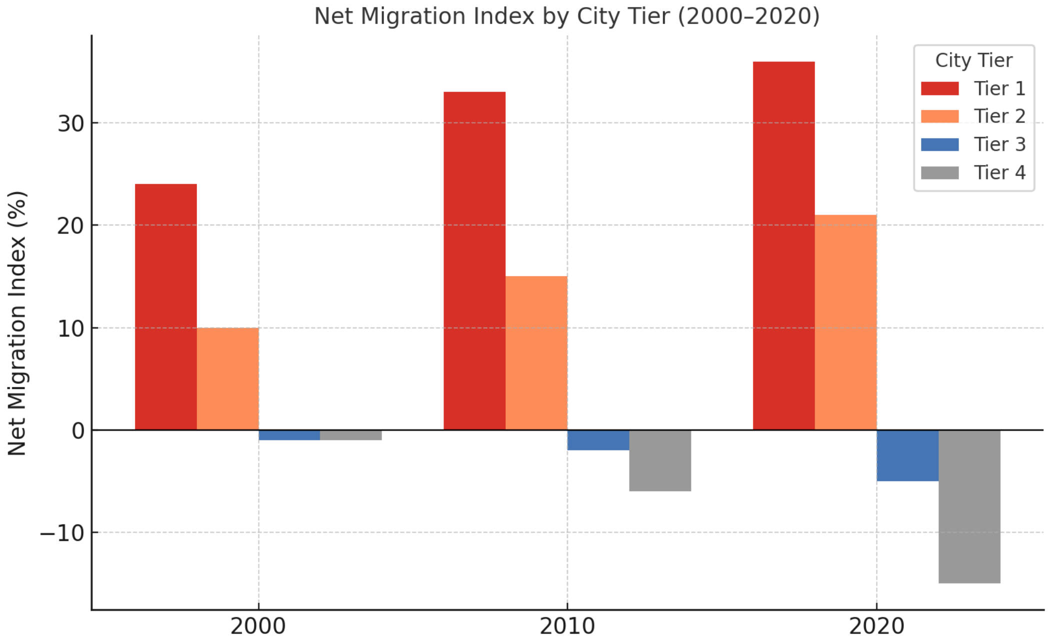

Figure 2: **Net Migration by City Tiers.**

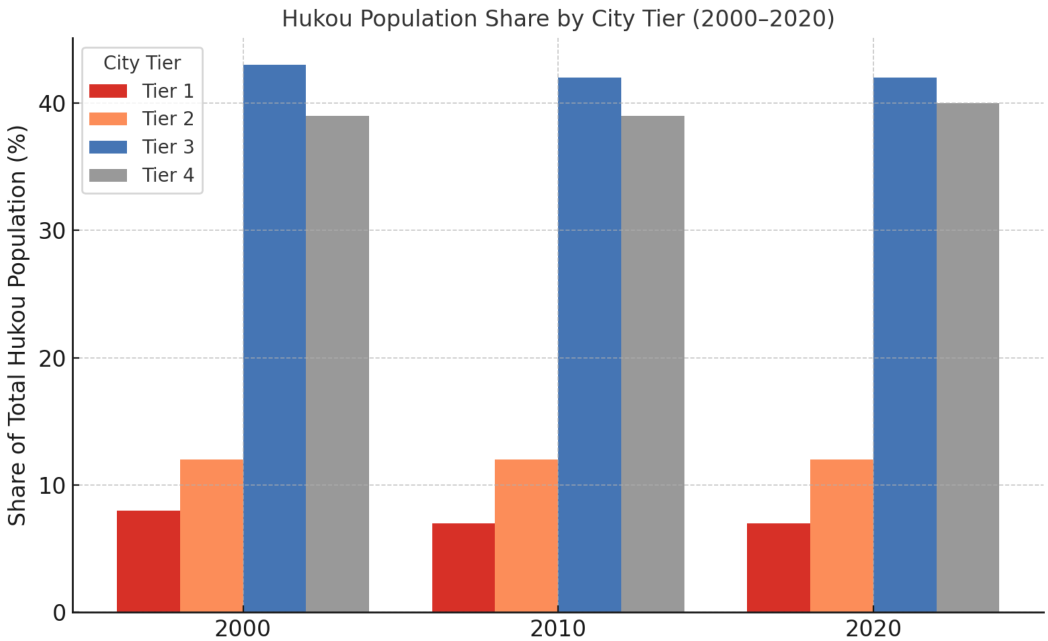

Figure 3: ***Hukou* Population by City Tiers.**

Fact 3. Persistent regional inequality. Despite massive migration flows, spatial convergence across prefectures has been modest. Census and statistical yearbook data show diverging trends between wages and GDP per capita. Wage inequality across cities has declined steadily: the cross-city Gini fell from 0.16 in 2000 to 0.11 in 2020, and the variance of log wages also decreased. This reflects migration's equalizing effect on labor market earnings, as rural and inland workers increasingly access higher-paying jobs in urban centers.

By contrast, GDP per capita inequality has remained high. The Gini coefficient of prefecture-level GDP per capita stayed around 0.40–0.42 during 2000–2020, with only modest improvement by 2020. The variance of log GDP per capita shows a similar persistence. This discrepancy highlights a critical feature of China's spatial development: while migration narrows wage gaps, *hukou*-related barriers prevent full settlement and service access, thereby sustaining wide disparities in fiscal capacity, capital deepening, and productivity across regions.

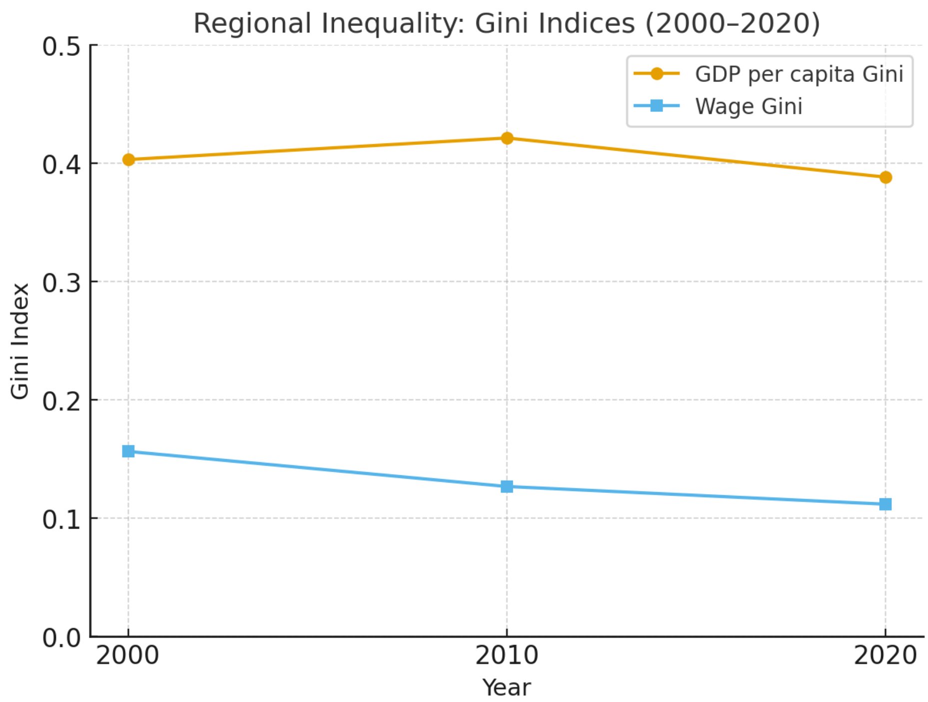

Figure 4: **Regional Inequality**

This paradox motivates the model design. The model allows wages to adjust through labor reallocation, while persistent development inequality arises from two dynamic channels. First, endogenous productivity dynamics strengthen regional advantages over time as agglomeration reinforces high-productivity locations. Second, endogenous amenity dynamics create competing forces: population inflows expand local fiscal capacity for infrastructure and services, yet congestion lowers livability, especially when migrants face limited access to benefits. Together, these dynamics explain why labor reallocation can narrow wage gaps while regional GDP per capita divergence remains entrenched.

Fact 4. Fragmented and Uneven *Hukou* Governance. Since the 1979 economic reforms, authority over *hukou* policy has progressively devolved from the central government to provincial, municipal, and prefectural levels. This decentralization has granted local governments substantial discretion to set admission criteria aligned with their distinct economic and demographic goals. The result is a patchwork of settlement regimes across China. Many small and mid-sized cities, often facing labor shortages or seeking to stimulate growth, have implemented relatively open settlement policies. In contrast, Tier-1 megacities have maintained highly restrictive systems. These are not uniform: for example, Shanghai employs a transparent points-based system, while Beijing uses an opaque employer-sponsored quota allocation model. Despite different mechanisms, the goal is the same: to tie settlement rights to factors such as educational attainment, skilled employment, and fiscal contributions.

Consequently, a migrant's ability to obtain local registration (and the accompanying access to public services like education, healthcare, and housing) is intensely place-specific. Significant variation exists even among cities of comparable size and economic profile; some have aggressively relaxed criteria to attract skilled labor and investment, while others retain barriers to manage population density and public resource strain. Over time, this decentralized approach has amplified regional divergence in migrant settlement opportunities.

This observed heterogeneity provides the empirical foundation for our counterfactual scenarios. The three modeled policy interventions—full national abolition (S0), targeted reforms exclusively in high-barrier cities (S1), and reforms focused on major economic hubs (S2)—are direct abstractions from the current institutional landscape. They allow us to assess whether incremental, locally-tailored reforms are sufficient to improve labor mobility and reduce welfare inequality, or if a more systemic, nationwide dismantling of the *hukou* system is necessary to unlock the full economic and social benefits of migration.

3 The Model

This section develops 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.

The economy consists of N prefecture-level regions indexed by $n = 1, 2, \ldots, N$, each endowed with a fixed supply of land $H(n) > 0$, constant over time. The total population is $\bar{L} = \sum_h L^h$, where each worker holds a registered *hukou* status h that denotes their registered location and remains fixed over the horizon.[2] Each worker is endowed with one unit of labor and supplies inelastically in their chosen residence. Location choices depend on wages, amenities, and migration costs, which include *hukou*-related restrictions. The resulting spatial distribution of labor shapes local market sizes, productivity, and amenity dynamics. Goods are traded between regions subject to symmetric iceberg transportation costs $\tau(n, j)$.

The baseline model is presented as a closed economy with respect to international trade, focusing on internal migration and prefecture-to-prefecture trade flows. This restriction ensures tractability and allows us to isolate the effects of *hukou* frictions. [3]

[2]This assumption reflects historically low rates of *hukou* change before 2000 and facilitates tractable large-scale simulations across multiple locations and periods.

[3]Work in progress extends the model to incorporate international linkages by introducing a "rest of the world" (ROW) node that trades with prefectures but does not participate in migration. Net exports at the prefecture level are being cleaned and incorporated as part of the regional budget constraint. This approach preserves the tractability of a closed-economy spatial equilibrium for domestic mobility

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, so wealth accumulation is not considered. However, they observe economic conditions and make location decisions to maximize their lifetime utility, subject to idiosyncratic taste shocks, mobility costs, and costs 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:

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

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; $\phi(n,h)$ captures the disutility of moving for an individual with *hukou* status h living in n;[4] $m(n,j)$ 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 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 4.2.

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).[5]

analysis, while still embedding the observed influence of international trade on local economic conditions, in line with recent work such as (Tombe & Zhu, 2019).

[4]We should think about the costs due to *hukou* as a product of all disutility costs 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 disutility rather than the whole series of disutility.

[5]This assumption ensures that total expenditures match incomes, maintaining general equilibrium

3.1.2 *Hukou* Parameter

The parameter $\phi(n,h)$ 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 ($\phi(n,h) \neq 1$), the individual may face restricted access to amenities, reflecting the barriers imposed by the *hukou* system.

This parameter $\phi(n,h)$ 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, 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 $P_t(n)$ [6] Consequently, the utility for agent i with *hukou* h in location n is:

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

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

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

consistency without introducing an additional agent.

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

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

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

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

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 beginning of each period, workers observe the economic conditions and realize idiosyncratic amenity shocks. Based on this information, 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 $P_t(\cdot)$.

Consider an individual of *hukou* type $h \in N$ currently residing in location n. The indirect utility from moving to location s in period t is given by

$$u_t^{ih}(s \mid n) \;=\; \frac{A_t(s)\, w_t(s)}{\phi(s,h)\, m(s,n)\, P_t(s)} \cdot \varepsilon_t^i(s).$$

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:

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

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:

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

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

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

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 obtains 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 $\mu_{t-1}^h(n)$, then the ratio of these individuals moving to location s at time t relative to the total labor population is $\mu_{t-1}^h(n)\Omega_t^h(s,n)$. 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:

$$H(s)\bar{l}_t(s)\mu_t^h(s) = \sum_n^N \mu_{t-1}(n)^h \Omega_t^h(s,n) H(n)\bar{l}_{t-1}(n). \tag{6}$$

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 $R_t(n)$, generates total income $R_t(n)H(n)$.

In contrast to frameworks where landlords consume goods directly, here all rental income is transferred to local governments and used exclusively for public investment. These resources finance infrastructure, public services, environmental improvements, and other components that enhance location-specific amenities. Thus, land rents provide a direct fiscal channel linking economic activity to local livability.

Formally, amenities in region n at time t combine three elements: an exogenous geographic baseline, congestion effects from higher population density, and endogenous improvements financed by land rents. This structure ensures that migration not only reallocates labor but also expands (or strains) local amenities, reinforcing feedback loops between mobility, productivity, and livability.

By abstracting from landlord consumption, the model avoids the need to specify preferences or spending patterns for an additional class of agents, while still preserving the fiscal role of land rents in shaping spatial equilibrium.

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 used by firm ω
- $L_t^\omega(n)$ 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).

Since land is fixed in supply, the land market clearing condition implies:

$$\int_0^1 H_t^\omega(n)\, d\omega = H(n). \tag{8}$$

It is therefore convenient to normalize firm variables by land. Define labor density and output per unit land as

$$q_t^\omega(n) = z_t^\omega(n) l_t^\omega(n)^\iota, \tag{9}$$

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

3.4.2 Firm-Level Productivity and Heterogeneity

We follow the approach of (Eaton & Kortum, 2002) and (Tombe & 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 Fréchet 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:

$$Z_{t+1}(n) = Z_t(n)^{\alpha} \bar{l}_t(n)^{\lambda_1} \epsilon_{t+1}(n), \tag{10}$$

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 $l_t^\omega(n)$:

$$\iota p_t^\omega(n,n) z_t^\omega(n) l_t^\omega(n)^{\iota-1} = w_t^\omega(n). \tag{11}$$

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.

$$\pi_t^\omega(n) = p_t^\omega(n,n) z_t^\omega(n) l_t^\omega(n)^\iota - w_t(n) l_t^\omega(n) - R_t(n) = 0. \tag{12}$$

Rearranging 12 and applying FOC gives an expression for the equilibrium land rent paid by firm ω in region n at time t:

$$\begin{aligned} R_t(n) &= p_t^\omega(n,n) z_t^\omega(n) l_t^\omega(n)^\iota - w_t(n) l_t^\omega(n) \\ &= \frac{w_t(n) l_t^\omega(n)}{l} - w_t(n) l_t^\omega(n) \\ &= \frac{1-l}{l} w_t(n) l_t^\omega(n). \end{aligned} \tag{13}$$

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

$$l_t^\omega(n) = \frac{\iota R_t(n)}{(1-\iota) w_t(n)}. \tag{14}$$

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:

$$p_t^{\omega}(n,n) = \iota^{-1} w_t(n) z_t(n)^{-1} l_t(n)^{1-l}. \tag{15}$$

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:

$$\begin{aligned} p_t^{\omega}(s,n) &= \upsilon(s,n)\iota^{-1} w_t(n) z_t(n)^{-1} l_t(n)^{1-l} \\ &= \frac{w_t(n)\upsilon(s,n)}{\iota Z_t(n)\epsilon_t^{\omega}(n) l_t(n)^{\iota-1}}. \end{aligned} \tag{16}$$

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 & Kortum, 2002), the cumulative distribution function (CDF) of the prices for a good ω produced in location n and consumed in location s is:

$$\mathbb{P}\left[p_t^{\omega}(s,n) \leq p\right] = \mathbb{P}\left[\epsilon_t^{\omega}(n) \geq \frac{w_t(n)\upsilon(s,n)}{\iota Z_t(n) l_t(n)^{\iota-1} p}\right]. \tag{17}$$

Plug (17) into the Fréchet distribution, we will have:

$$\begin{aligned} \mathbb{P}\left[p_t^{\omega}(s,n) \leq p\right] &= 1 - exp\left\{-T(n)\left(\frac{w_t(n)\upsilon(s,n)}{\iota Z_t(n) l_t(n)^{\iota-1} p)}\right)^{-\delta}\right\} \\ &= 1 - exp\left\{-T(n)\left(\frac{\upsilon(s,n)\xi_t(n)}{p}\right)^{-\delta}\right\}, \end{aligned} \tag{18}$$

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

Plug (18) into the price index, and we can extend the price index in location s at time t as:

$$\begin{aligned} P_t(s) &= \left[\int_0^1 p_t^{\omega}(s)^{-(\sigma-1)} d\omega\right]^{-\frac{1}{\sigma-1}} \\ &= \left[\int_0^{\infty} p^{-(\sigma-1)} \cdot \delta p^{\delta-1} \cdot \exp\left\{-\Phi_t(s) p^{\delta}\right\} dp\right]^{-\frac{1}{\sigma-1}} \\ &= \kappa \left[\sum_{n=1}^{N} T(n)(\xi_t(n)\upsilon(s,n))^{-\delta}\right]^{-1/\delta}, \end{aligned} \tag{19}$$

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:

$$\begin{aligned}
\pi_t(s,n) &= \int_0^\infty \prod_{j\in N} exp\left\{(\frac{\upsilon(s,j)\xi_t(j)}{p})^{-\delta}\right\} exp\left\{(\frac{\upsilon(s,n)\xi_t(n)}{p})^{-\delta}\right\} (\upsilon(s,n)\xi_t(n))^{-\delta}\, dp^\delta \\
&= (\upsilon(s,n)\xi_t(n))^{-\delta} \int_0^\infty \exp\left\{-\sum_j^N (\upsilon(s,j)\xi_t(j))^{-\delta} p^\delta\right\} dp^\delta \\
&= \frac{(\upsilon(s,n)\xi_t(n))^{-\delta}}{\sum_{j=1}^N (\upsilon(s,j)\xi_t(j))^{-\delta}}
\end{aligned} \tag{20}$$

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:

$$Y_t(n) = \sum_{s=1}^N \pi_t(s,n)Y_t(s). \tag{21}$$

By substituting the relevant expressions for labor (13) and rent (14) from previous sections, this equation simplifies to:

$$w_t(n)\bar{L}_t(n) = \sum_{s=1}^N \pi_t(s,n)w_t(s)\bar{L}_t(s). \tag{22}$$

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:
$$\sum_{s}^{N} H(s)\bar{l}_t(s) = \bar{L}_t \tag{23}$$

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 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 the 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 every period, then the equilibrium allocation of labor, production, and prices is unique. If not, the model may exhibit multiple equilibria, 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 re-classifications, 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 the 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.

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, I construct a matrix of bilateral trade costs across all prefecture pairs using GIS-based travel time estimates. Specifically, I apply the ArcGIS Origin-Destination (OD) Cost Matrix tool, overlaying a transportation network based on roads and railways from 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 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, I describe how key structural parameters are estimated using these data.

4.2 Parameter Estimation and Calibration

The quantitative implementation of the model requires assigning values to several parameters. Following standard practice in the spatial equilibrium literature, I classify parameters into three categories: (i) parameters set to values commonly used in the literature, (ii) parameters estimated directly from data, and (iii) parameters calibrated or inferred through model inversion.

4.2.1 Parameters Set to Literature Benchmarks

The discount factor β adjusted to reflect the use of decade intervals, leading to a value of 0.776, 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, Dvorkin, & Parro, 2019) find that trade elasticity typically falls within the range of 3 to 8, which is consistent with earlier estimates by (Eaton & Kortum, 2002) and (Head & 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 & Donaldson, 2020), (Moretti, 2012), (Comin & Hobijn, 2010), (Fagerberg, Srholec, & Knell, 2007) and many others, making 0.98 a reasonable value.

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, & 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, I 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, I estimate the elasticity parameters using a GMM framework. This approach accounts for potential endogeneity in both population density and rent-financed amenity investment. Specifically, I 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 E.

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 applications 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:

$$\log M_{n\to j,t} = \gamma \log\left(\frac{w_{j,t}}{w_{n,t}}\right) - \theta \log D_{n,j} + \eta_j n + \eta_j + \eta_t + e_{njt}, \tag{24}$$

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). I 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 & Rossi-Hansberg, 2021)). Based on this evidence and to maintain model tractability, I set $\gamma = 2$ for the quantitative analysis.

Agglomeration Effects on Productivity. Local agglomeration plays a vital role in shaping regional productivity dynamics. Following the literature, I 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, I 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. I provide full technical details 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 the spatial allocation of economic activity.

To construct bilateral trade costs between all 313 prefecture-level cities in China, I 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), I calculate the shortest travel time $t(n, s)$, measured in hours. I 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 land transportation.

Once the shortest travel times are computed, I 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 Fréchet distribution governing firm productivity in our model. As mentioned before, I adopt a standard value from the literature rather than estimating this parameter from our data.

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

Table 1: **Baseline Parameter Values.**

Parameter	Description	Value	Source / Notes
β	Discount factor	0.776	Implied by a 2.5% annual interest rate over a decade
σ	Elasticity of substitution across varieties	4.00	(Krugman, 1991)
δ	Trade elasticity	4.55	(Caliendo et al., 2019)
α	Persistence of regional productivity	0.98	(Allen & Donaldson, 2020)
λ	Elasticity of amenities w.r.t. population density	-0.39	GMM estimate from urban amenity regression
χ	Elasticity of amenities w.r.t. amenity investment	0.23	GMM estimate from urban amenity regression
γ	Elasticity of migration flows w.r.t. real income	2.00	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

Notes: GMM = Generalized Method of Moments. "w.r.t." denotes "with respect to." All estimates are at the prefecture level unless otherwise noted.

4.3 Values for Initial Period

4.3.1 Initial Productivity and Exogenous Amenity

Solving this dynamic model requires values for initial productivity and exogenous amenities in each location. Following the strategy in (Desmet et al., 2018), I 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:

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

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:

$$Z_0(n) = (1-\iota)^{-\chi}\iota^{\chi-1}\frac{\tilde{u}_0(n)}{\bar{A}(n)}w_0(n)^{-\chi}\bar{l}_0(n)^{1-\iota-\lambda} \tag{26}$$

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:

$$\begin{aligned} &\left[\frac{\bar{A}(n)}{\tilde{u}_0(n)}\right]^{-\delta} w_0(n)^{-\delta(\chi+1)}\bar{l}_0(n)^{-\lambda\delta} \\ = \kappa\left[\frac{\iota}{1-\iota}\right]^{2\chi\delta} \sum_j &\left[\frac{\bar{A}(j)}{\tilde{u}_0(j)}\right]^{-\delta} T(j)w_0(j)^{-\delta(1+\chi)}\bar{l}_0(j)^{-\lambda\delta}\upsilon(n,j)^{-\delta}, \end{aligned} \tag{27}$$

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 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. Full derivations and solution details are provided in Appendix A.

4.3.2 Mobility Costs and Amenity Lost

To implement the model, it is crucial to estimate mobility costs across locations. We use four sets of inputs: (i) population distributions for periods 0 and 1, (ii) *hukou* population by location for periods 0 and 1, (iii) observed migration flows between periods 0 and 1, and (iv) the exogenous amenity level $\bar{A}$ recovered in the previous subsection.

From equation (5), the bilateral migration probability of *hukou* type h moving into location n from origin s in period 0 is

$$\Omega_0^h(n,s) = \frac{\tilde{u}_0(n)^\gamma, \tilde{m}(n,s)^{-\gamma}}{\sum_{j=1}^N \tilde{u}_0(j)^\gamma, \tilde{m}(j,s)^{-\gamma}}, \tag{28}$$

where $\tilde{m}(n,s) = m(n,s)\phi(n,h)$ is the compound migration cost, consisting of distance-related cost $m(n,s)$ and the *hukou*-related penalty $\phi(n,h)$.

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. The staying probability (remaining in one's registered *hukou* location) implies

$$\Omega_0^{h=n}(n,n) = \frac{\tilde{u}_0(n)^\gamma}{\sum_{j=1}^N \tilde{u}_0(j)^\gamma, \tilde{m}(j,n)^{-\gamma}}. \tag{29}$$

Taking the ratio of equations (28) and (29), and rearranging, we can derive a closed-form expression for the compound mobility cost from the data:

$$\tilde{m}(n,s) = \left[\frac{\Omega_0^{h=n}(n,n)}{\Omega_0^h(n,s)}\right]^{1/\gamma}. \tag{30}$$

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 A.

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 θ 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:

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

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 θ is the elasticity of migration flows with respect to distance.

From this model, I estimate $\theta = 1.32$, 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 objective of this counterfactual exercise is to isolate the dynamic consequences of lifting migration frictions while holding other structural features of the economy constant.

4.4.1 Simulation Approach

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

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

1. **Baseline.** This scenario maintains the existing *hukou* system throughout the transition, with all bilateral migration frictions remaining unchanged over time.

2. **Full reform.** This scenario abolishes the *hukou* system from period $t = 0$ onward by setting $\phi(n, h) = 1$ 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.

3. **Selective/partial reform.** This scenario abolishes the *hukou* system only in designated reform prefectures. Formally, $\phi(n, h) = 1$ if n is a reform prefecture, while $\phi(n, h) > 1$ elsewhere. This captures the reality of partial reforms in which migrants gain local access rights in some destinations but not nationwide.

All scenarios are simulated for 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 $Z_t(n)$.

4.4.2 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:

1. **Labor distribution.** The reallocation of labor across regions, highlighting areas that gain or lose population once migration constraints are lifted;

2. **Productivity.** Changes in regional productivity, which evolve endogenously through agglomeration effects and innovation dynamics; and

3. **Welfare.** Welfare improvement is computed as changes in lifetime utility aggregated across individuals and weighted by the initial population distribution.

To maintain a clear empirical focus, this analysis restricts attention to the effects of institutional reforms alone. 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.

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 capture 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 5 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 less through industrial agglomeration than through environmental quality. Coastal cities in Guangdong and Zhejiang also feature high amenity levels, reflecting fertile land, trade access, and historical investment.

By 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 6 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 such as 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 sustained high levels of industrial output and labor demand.

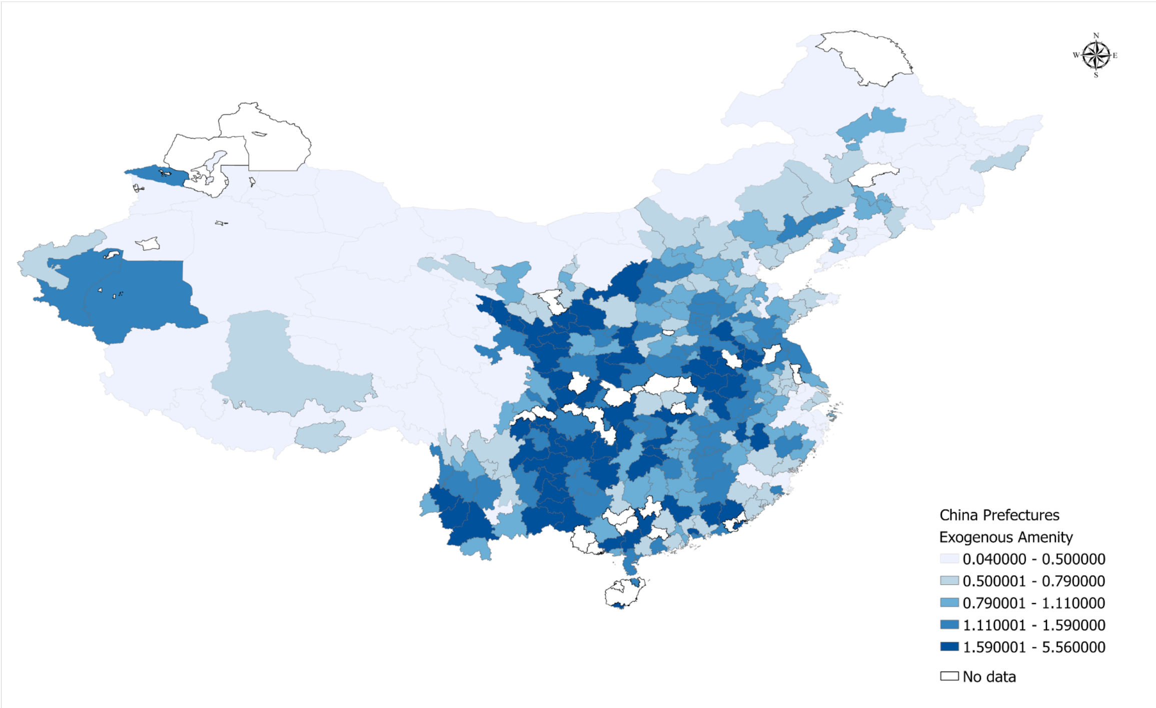

Figure 5: **Exogenous Amenities by Prefecture.**

Productivity levels in central and southwestern regions vary widely, reflecting mixed development paths. Provinces like Sichuan show moderate productivity, supported by 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 of Figure 7 displays the spatial distribution of real wages (in units of 1,000 RMB), computed by the model based on estimated productivity, exogenous 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

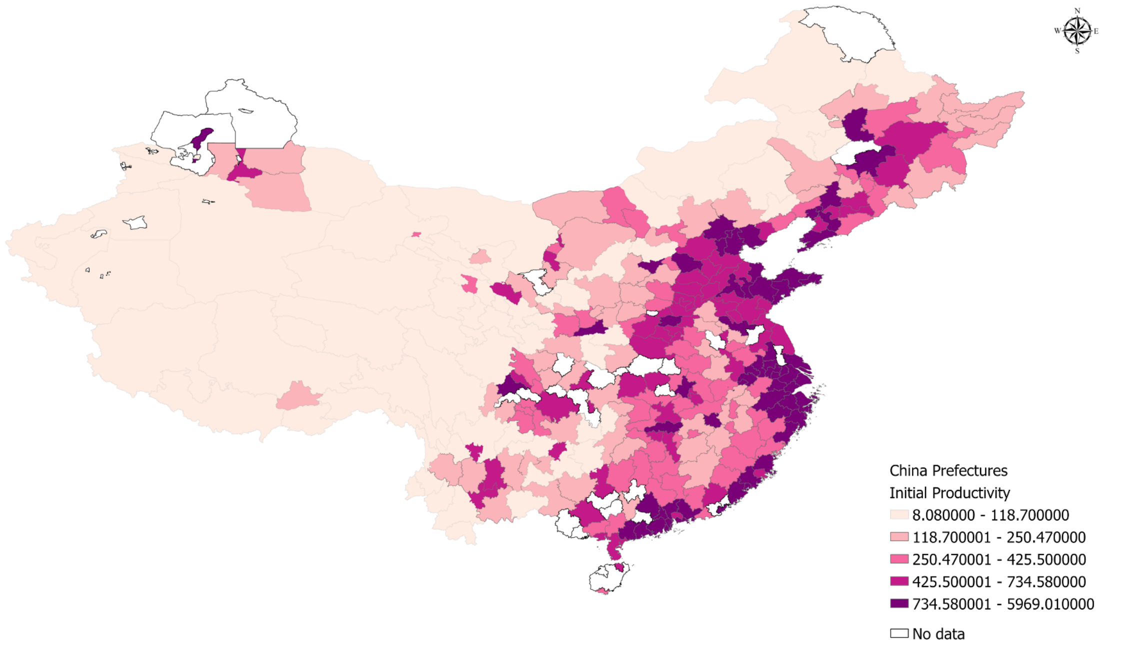

Figure 6: **Prefecture-Level Productivity in 2000.**

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.

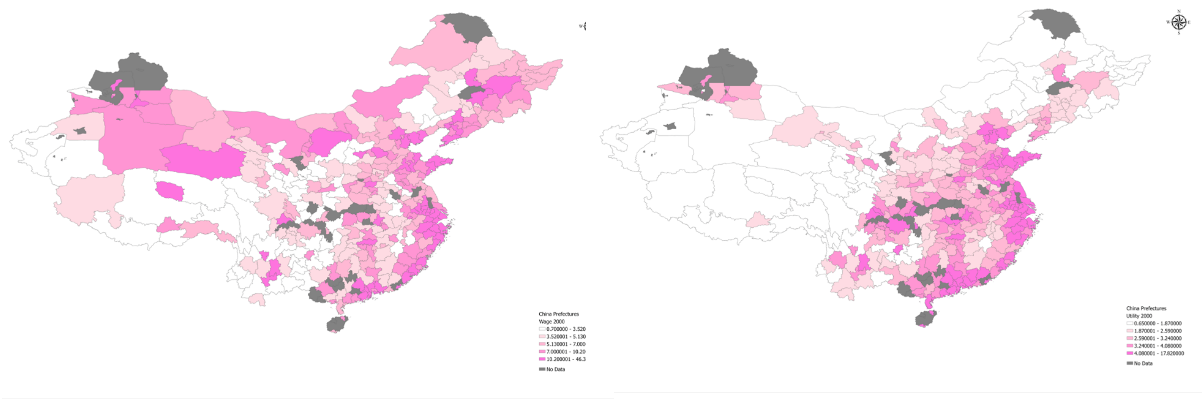

Figure 7: **Model-Implied Wage vs. Utility in 2000.**

The right panel of Figure 7 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 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.

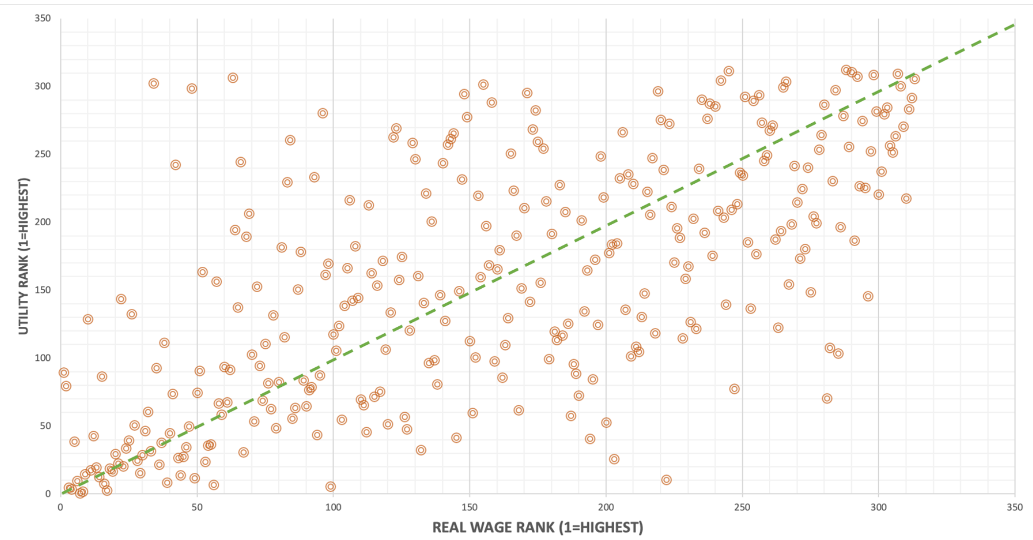

Figure 8: **Real Wage and Utility in 2000.**

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.

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. Figure 9 presents the implied total cost of moving into each prefecture. To provide a meaningful comparison, I take the average of this cost conditional on destination prefectures (migrant recipient cities). In other words, the bilateral cost is expressed as a moving-in cost specific to each location. These costs were normalized relative to the cost in Shanghai.

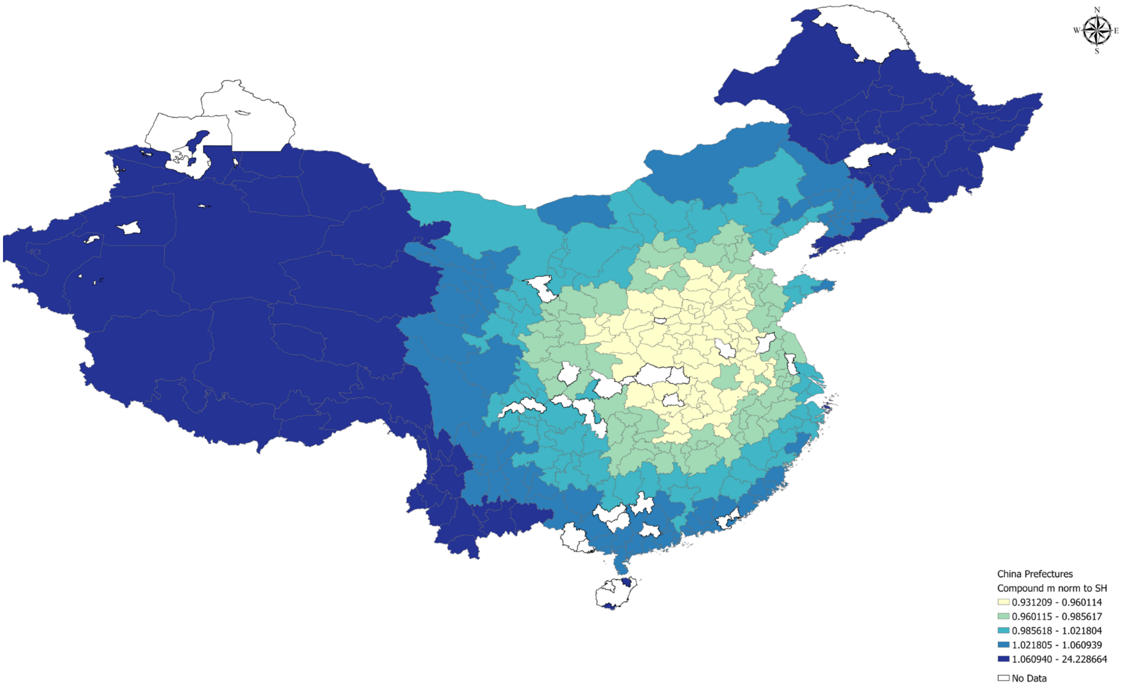

Figure 9: **Moving-in Cost at $t = 0$.**

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 (Figure 11). This component captures the relative importance of *hukou* restrictions—compared with 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, Neijiang in Sichuan province, and Longnan, Zhangye in Gansu province, the *hukou* share of total migration cost exceeds 75%, indicating that institutional restrictions—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.

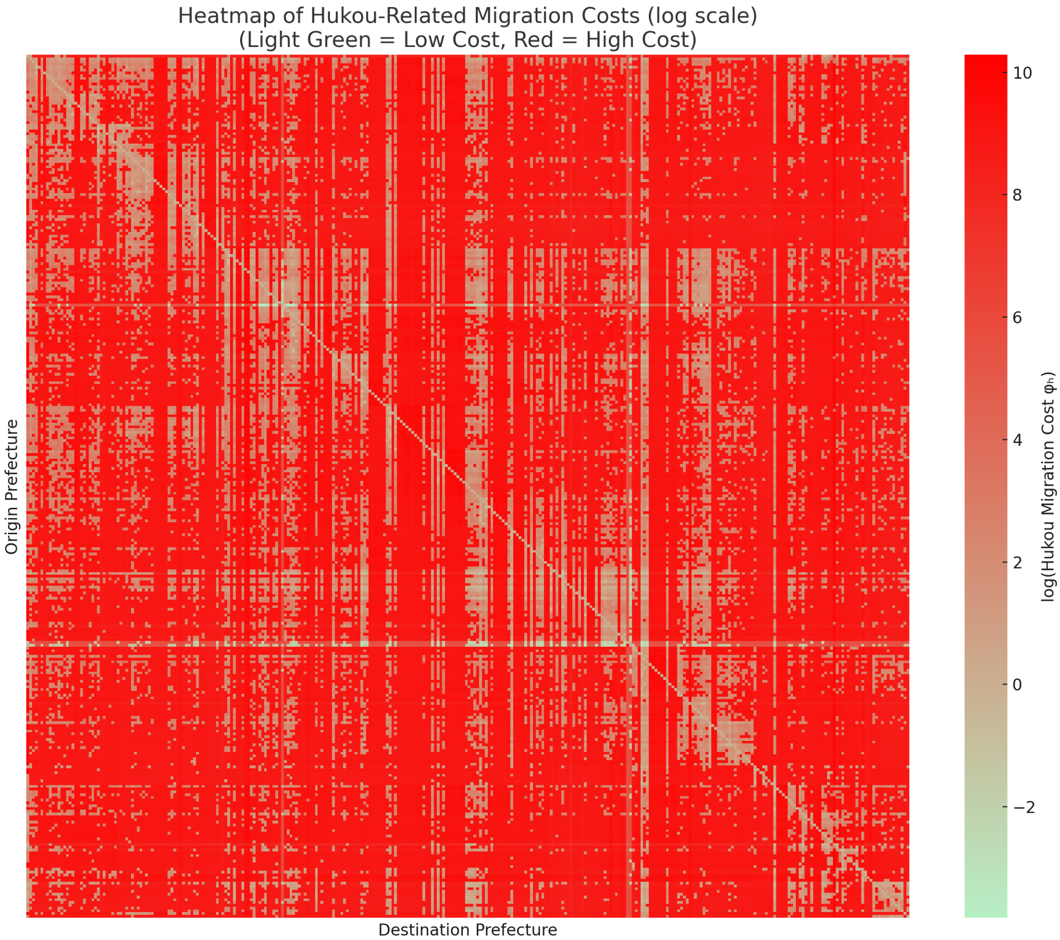

Figure 10: ***Hukou*** **Induced Migration Cost (log-scale).**

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 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

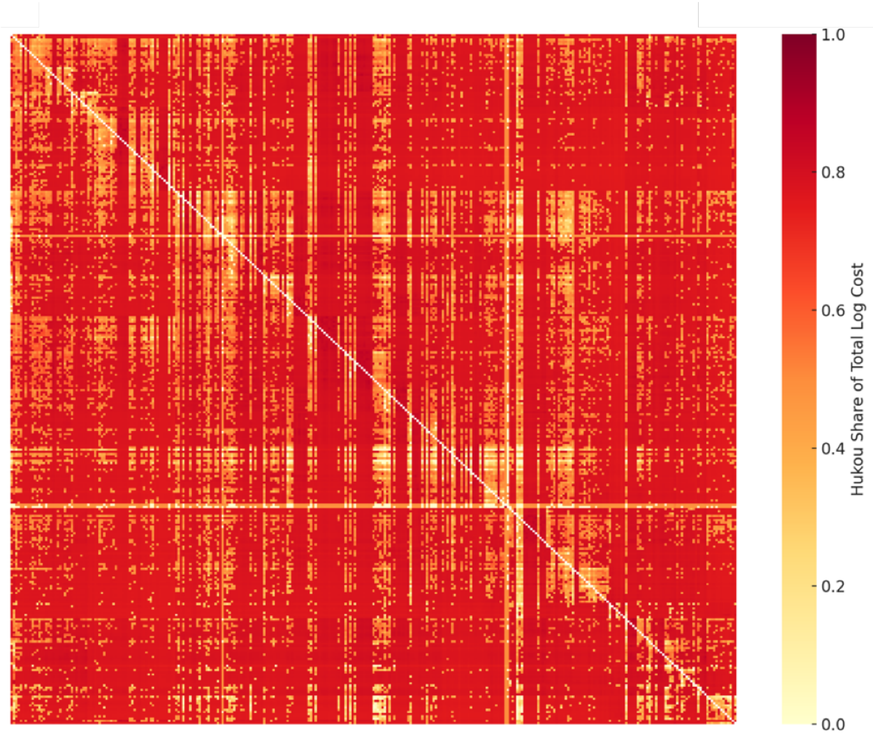

Figure 11: **Share of *Hukou* Cost in Total Migration Cost (log-scale).**

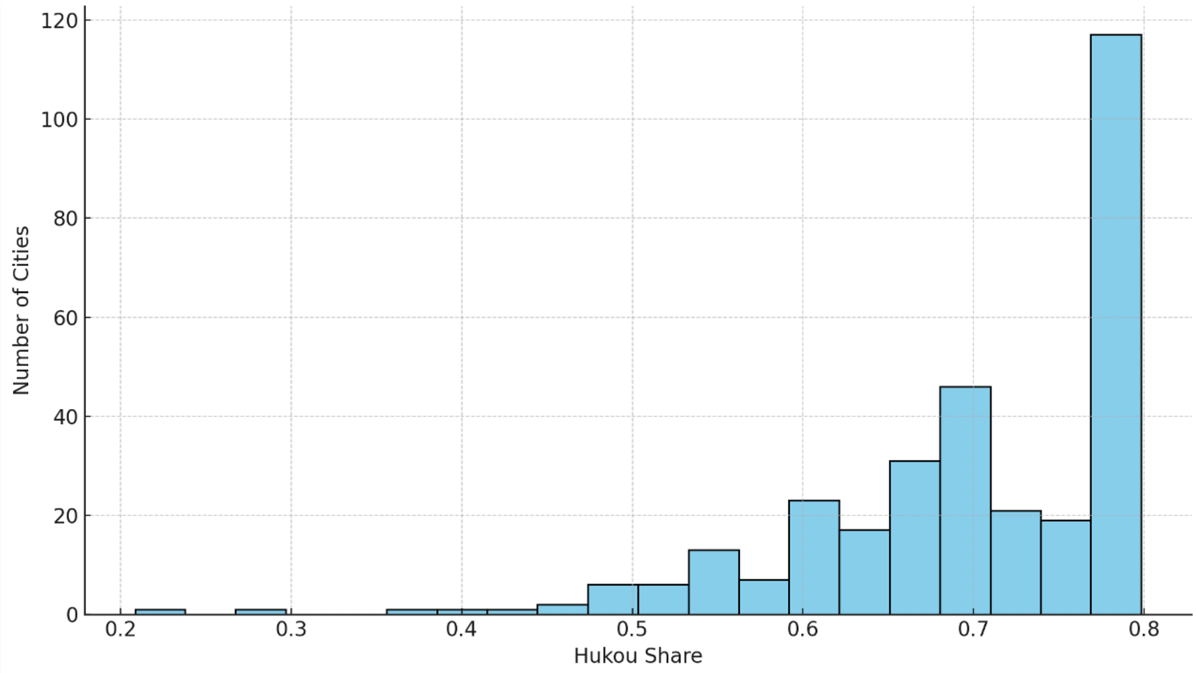

Figure 12: **Distribution of *Hukou* Cost Share in Total Migration Cost.**

structure of total migration costs. In some cases, migrants may perceive restrictive *hukou* 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 13 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 role 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: the correlation coefficient between estimated and observed GDP shares is 0.94, indicating that the model provides a reasonable approximation of cross-provincial economic structure.

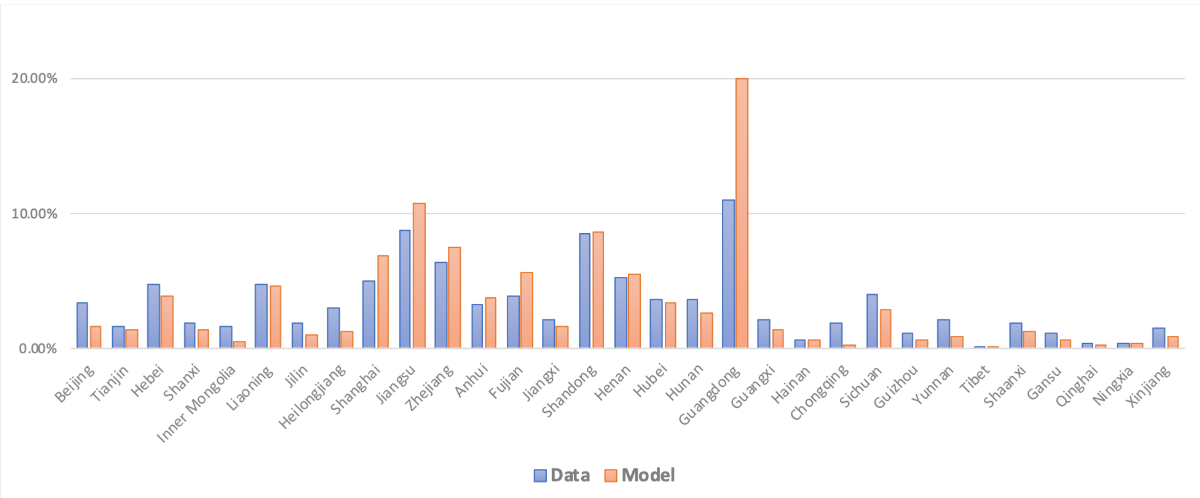

Figure 13: **Model VS. Data: Output Share in 2000.**

Next, we examine the model's performance in replicating spatial population patterns. Figure 14 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.

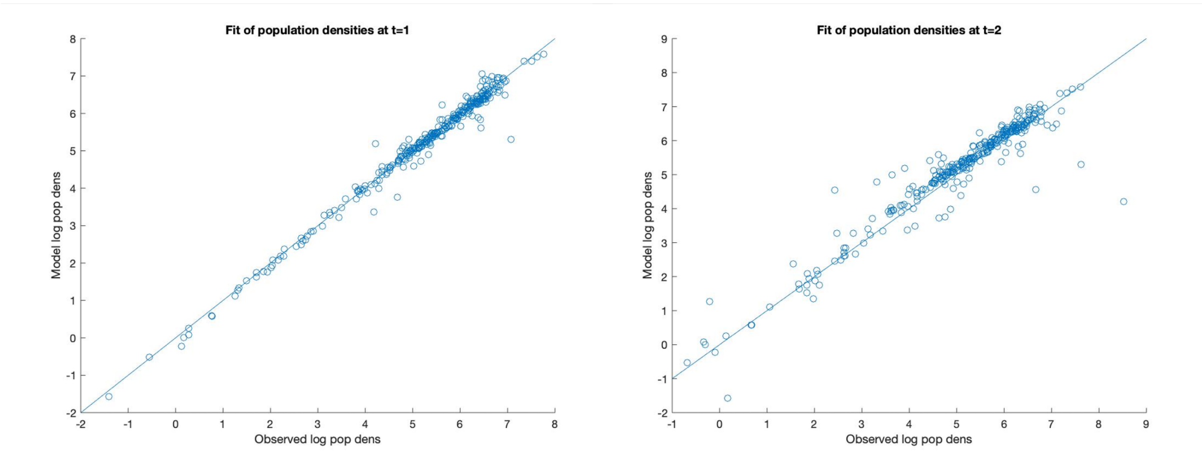

Figure 14: **Model VS. Data: Population Density in 2010 and 2020.**

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 geographic scope, then decompose the resulting welfare changes into their underlying mechanisms. We further examine the robustness of these results to alternative parameter values.

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 nationwide, 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 – High-barrier cities only.** *hukou* restrictions are abolished only in destinations where the *hukou* cost share $\Phi(n) \geq 0.75$, covering 128 prefecture-level cities and roughly 36% of the national population. These are disproportionately inland or resource-dependent (e.g., mining) cities with weaker productivity fundamentals. For all other destinations, $\phi(h, n)$ remains at its baseline level.

3. **S2 – Economic hubs only.** *hukou* restrictions are abolished only in the top 10% of cities by baseline GDP share ($\geq 1\%$ in year 2000), covering 46 cities and about 18% of the population. These hubs are predominantly coastal or highly urbanized and benefit from strong agglomeration economies.

Figure 15 shows cities affected by different reform scenarios.

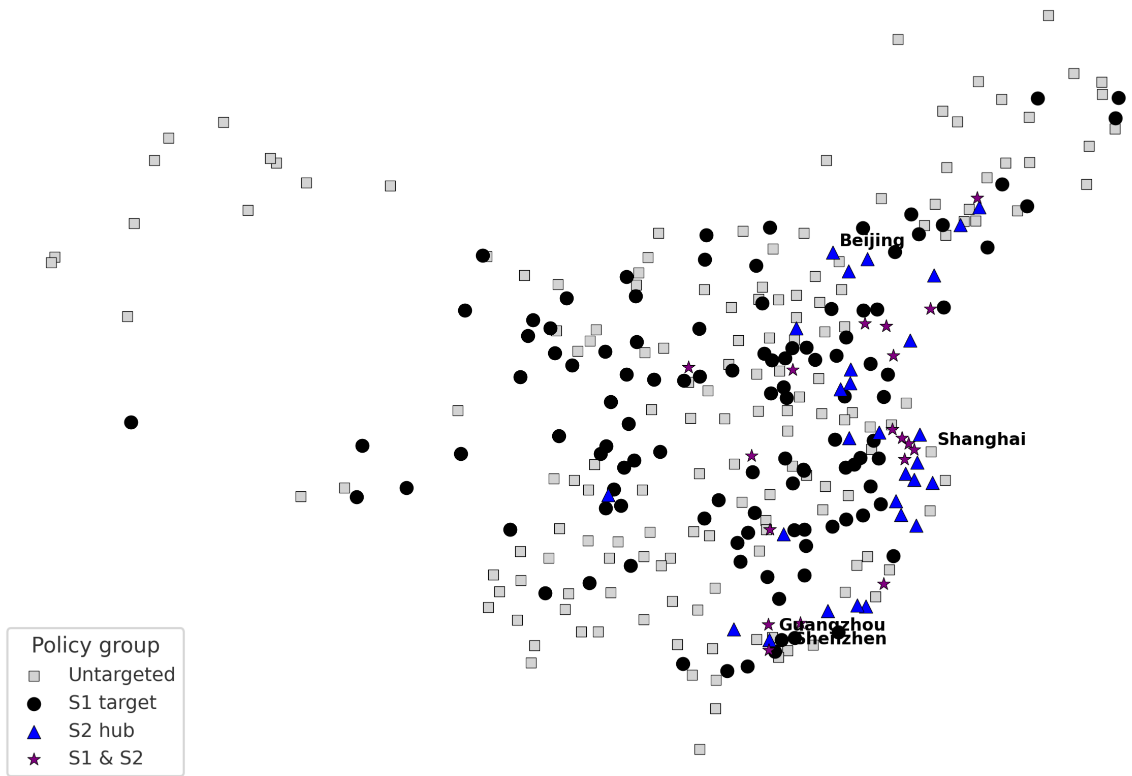

Figure 15: **Targeted Cities Under Different Reform Scenarios.**

6.2 Aggregate Results

To compute the aggregate effects reported in Table 2, we use the following definitions. All changes are expressed relative to the baseline (no reform) economy.

Aggregate productivity is the summation of regional productivities:

$$\bar{Z}_t \equiv \sum_n Z_t(n). \tag{31}$$

The cumulative percentage change is then:

$$\Delta Z_t = 100 \times \left(\frac{\bar{Z}_t^{\text{CF}}}{\bar{Z}_t^{\text{Baseline}}} - 1 \right). \tag{32}$$

National output is the sum of prefecture-level outputs:

$$Y_t = \sum_n Y_t(n). \tag{33}$$

The reported change is:

$$\Delta Y_t = 100 \times \left(\frac{Y_t^{\text{CF}}}{Y_t^{\text{Baseline}}} - 1 \right). \tag{34}$$

Aggregate welfare is based on the population-weighted utility index:

$$U_t = \sum_n L_t(n) u_t(n), \tag{35}$$

where $u_t(n)$ is deterministic utility.

Define the consumption-equivalent factor ζ that makes the baseline path indifferent to scenario s under CRRA ($\rho = 2$) and discount factor β:

$$\zeta \;=\; \left[\frac{\sum_{t=0}^{T} \beta^t \big(U_t^s\big)^{1-\rho}}{\sum_{t=0}^{T} \beta^t \big(U_t^{\text{Baseline}}\big)^{1-\rho}} \right]^{\frac{1}{1-\rho}}. \tag{36}$$

We report the level-equivalent percentage change:

$$\Delta CE(T) \;=\; 100\,(\zeta - 1), \qquad T \in \{2, 10\}. \tag{37}$$

We do not annualize CE (see table notes) since with $\rho > 1$, annualizing can be misleading.

For each scenario s and period t, we define the migration rate as the share of the total population that has relocated relative to the initial distribution:

$$M_t^s \;=\; \frac{\text{Migrants}_t^s}{\text{Population}_t}, \tag{38}$$

where Migrants_t^s denotes the total number of individuals who change location in scenario s at time t, and Population_t is the aggregate population, which is constant across scenarios.

To evaluate the effect of *hukou* reform, we compute the difference in migration rates between scenario s and the baseline:

$$\Delta M_t^s \;=\; M_t^s \;-\; M_t^{\text{Baseline}}. \tag{39}$$

Because the denominator is identical across scenarios, ΔM_t^s captures the additional share of the population that migrates as a direct consequence of *hukou* reform. This metric is reported in the final column of Table 2.

Table 2 reports short-run ($t = 2$, 20 years) and long-run ($t = 10$, 100 years) percentage changes in aggregate productivity, real output, CE welfare, and migration rates across scenarios. CE welfare is computed as a level-equivalent measure following Equations (36)–(37), which capture the uniform increase in consumption under the baseline that would make households indifferent to the counterfactual. Because $\rho > 1$ makes annualization misleading, we report CE changes only at the $t = 2$ and $t = 10$ horizons.

Table 2: **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	-7.90	698.39	-3.55	972.06	48.51	52.89	83%
S1: High-barrier only	-6.75	38.30	-8.66	37.71	-0.37	0.17	72%
S2: Economic hubs only	2.86	199.96	19.11	169.43	0.59	1.36	64%
Annualized growth (%/yr) for $t = 10$ (geometric)							
S0: Full abolition	–	2.10	–	2.89	–	–	
S1: High-barrier only	–	0.34	–	0.33	–	–	
S2: Economic hubs only	–	1.13	–	1.00	–	–	

Notes: $t = 2$ and $t = 10$ correspond to 20 and 100 years after reform, respectively. Annualized growth rates are geometric: $g_X = \left(1 + \Delta_X/100\right)^{1/100} - 1$, computed from the cumulative changes shown above. CE welfare is reported only as a cumulative level-equivalent percentage change (not annualized). Because $\rho > 1$ in the CRRA mapping, directly annualizing CE can be misleading (sign-reversing). Migration is reported as the share of the population that changes prefecture by $t = 2$.

Removing *hukou* barriers nationwide triggers a large initial migration surge, with many movers drawn to high-amenity but only moderately productive destinations. This compositional shift causes short-run declines in aggregate productivity (–7.9%) and output (–3.55%) as measured at $t = 2$. Over time, however, agglomeration effects in high-productivity hubs and endogenous amenity investments raise efficiency across the system, reversing the initial dip and producing dramatic long-run gains. After a century ($t = 10$), productivity and real output stand 699% and 972% above baseline, corresponding to annualized growth rates of 2.10%/yr and 2.89%/yr. CE welfare rises sharply—+48.5% in the short run and +52.9% in the long run—driven by both direct reallocation gains and dynamic feedbacks from productivity and amenity growth. The trajectory is consistent with the divergence patterns documented in Section 6.3: while average productivity growth initially slows due to inflows into mid-tier cities, over the long run the most productive hubs expand their lead, accounting for much of the aggregate gain.

Scenario S1 (reforms in high-barrier cities only) redirects labor toward many moderate-productivity inland cities, producing short-run declines in productivity (–6.75%) and output (–8.66%) and slightly negative CE welfare (–0.37%). By $t = 10$, productivity and output are 38.3% and 37.7% above baseline, but growth rates remain modest (0.34%/yr and 0.33%/yr). CE welfare is essentially unchanged in the long run (+0.17%), as efficiency gains are small and congestion costs limit amenity improvements.

Scenario S2 (reforms in top economic hubs only) liberalizes migration into 46 top-GDP cities, generating immediate productivity (+2.86%) and output (+19.11%) gains. By $t = 10$, productivity and output are 200% and 169% above baseline (1.13%/yr and 1.00%/yr). However, long-run CE welfare rises only 1.36%. Migration is lower than in S0 (64%), consistent with the more restricted set of destinations.

Taken together, the results show that full liberalization produces the largest and most persistent aggregate benefits, while partial reforms yield smaller or negligible welfare improvements. The CE metric highlights that efficiency gains do not translate one-for-one into welfare once congestion and cost-of-living effects are incorporated.

6.3 Growth Dynamics

Figure 16 compares the growth rates of aggregate productivity, real GDP, and utility across the Baseline, S0, S1, and S2 scenarios over the ten decades of the simulation.

6.3.1 Real GDP Growth

Nationwide removal of *hukou* restrictions (S0) produces the largest and most persistent increase in real GDP growth relative to the Baseline, reflecting substantial reallocation of labor toward more productive locations. Targeted reforms also raise GDP growth but to a lesser extent: S1 covers more prefectures (128), though many are only moderately productive, while S2 focuses on 46 top economic hubs, generating more concentrated gains despite covering fewer cities.

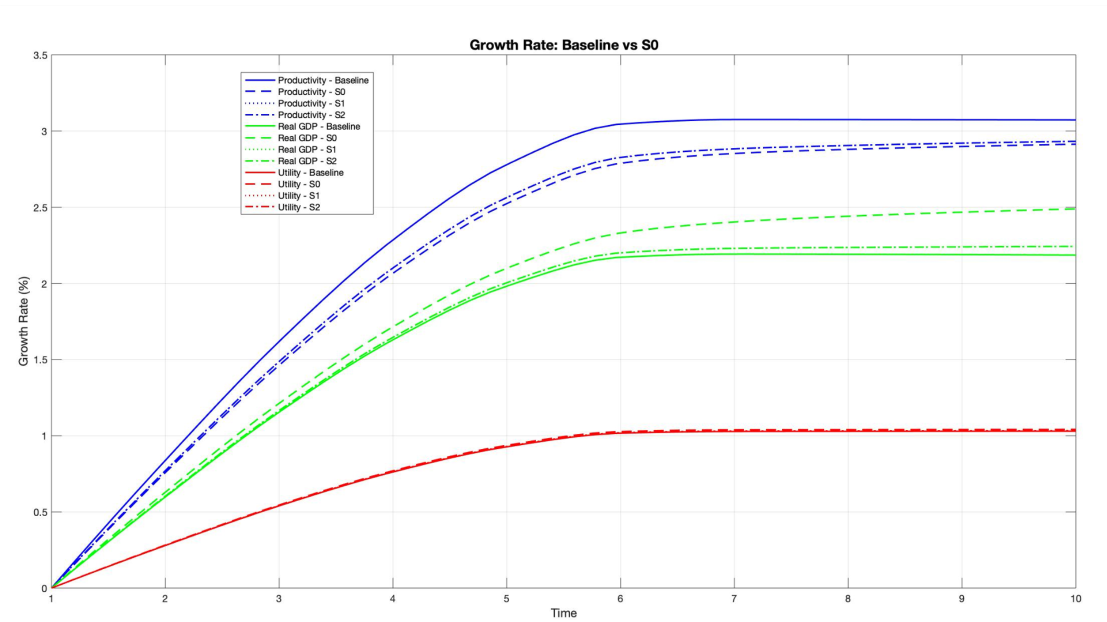

Figure 16: **Growth Dynamics.**

6.3.2 Productivity Growth

Across reform scenarios, average productivity growth rates fall slightly below the Baseline. This does not imply convergence. Instead, it reflects the concavity of agglomeration effects in the model. As major hubs absorb additional migrants, the marginal productivity gains from further concentration diminish, lowering the national average growth rate even as divergence between hubs and the periphery continues.

6.3.3 Utility Growth

Utility growth rates look nearly indistinguishable across scenarios. This apparent similarity is due to the offsetting effect of higher living costs and amenity congestion in the expanding hubs, which erode much of the wage gains from reallocation. However, the long-run consumption-equivalent (CE) welfare diverges. In S0, the economy experiences an immediate and substantial upward shift in utility levels, leading to a 52.89% higher CE welfare after 100 years compared to the Baseline, even though the long-run growth rate of utility is similar across scenarios. S1 and S2, by contrast, generate smaller and later level shifts in utility, resulting in much smaller CE gains.

Overall, these results highlight that the scope and targeting of migration reforms matter. Broad national liberalization produces the largest GDP and welfare gains, while targeted reforms yield more modest improvements. At the same time, GDP growth gains do not necessarily translate into higher long-run utility growth rates, underscoring the importance of considering both the level effects and general equilibrium feedbacks—such as land rents and congestion—when evaluating the welfare implications of migration policy.

6.4 Spatial Inequality Dynamics

Figures 17–19 reveal distinct patterns in the evolution of spatial inequality across wages, productivity, and utility.

6.4.1 Wage Inequality

In the Baseline, the Gini coefficient rises from 0.27 to 0.31 over ten periods, and the P90–P10 wage gap grows by roughly 15%. With full abolition (S0), the Gini climbs from 0.26 to 0.38, and the P90–P10 gap more than doubles, reflecting rapid migration toward high-wage cities. Targeting economic hubs (S2) produces the steepest divergence: the Gini exceeds 0.40 and the wage gap expands by nearly 150% relative to period 1. By contrast, the high-barrier-only reform (S1) yields an intermediate path, with inequality increases smaller than in S0 or S2.

6.4.2 Productivity Inequality

The spatial dispersion of productivity exhibits patterns similar to those for wages but with even stronger divergence forces. As shown in Figure 18, the Baseline already displays gradual widening over time, with the standard deviation of log productivity rising from

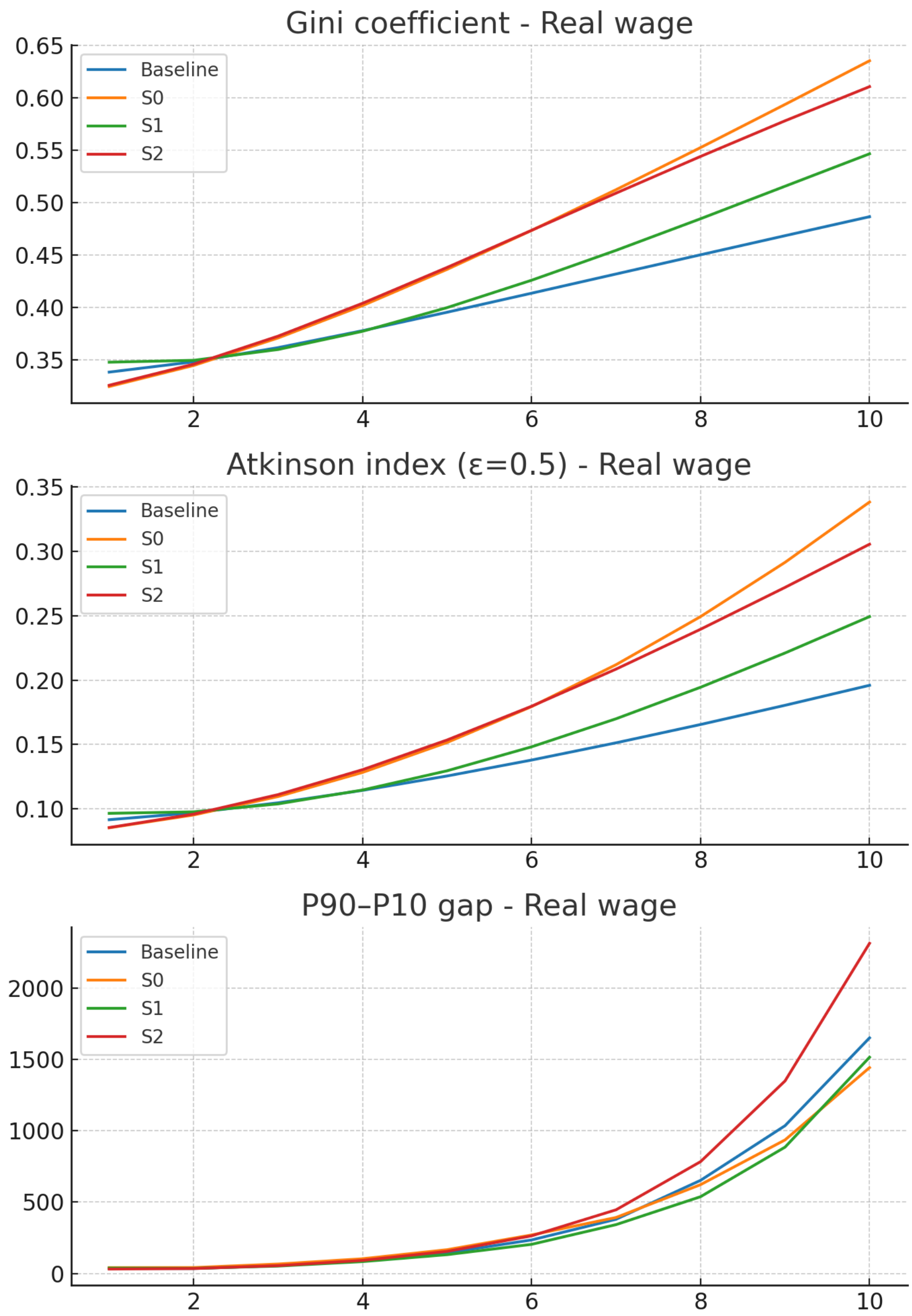

Figure 17: **Divergence or Convergence: Real Wage Inequality.**

0.61 to 1.21 over ten periods. Under complete *hukou* removal (S0), dispersion nearly doubles over the horizon, while targeted reform toward advantaged cities (S2) produces the steepest trajectory, reaching 1.45 by period 10. These results indicate that migration-induced agglomeration effects compound pre-existing productivity advantages, with the largest gains accruing to cities already well-positioned in the productivity distribution. The similarity of patterns in wage and productivity dispersion suggests that real wage divergence is driven largely by underlying productivity dynamics rather than wage-setting frictions.

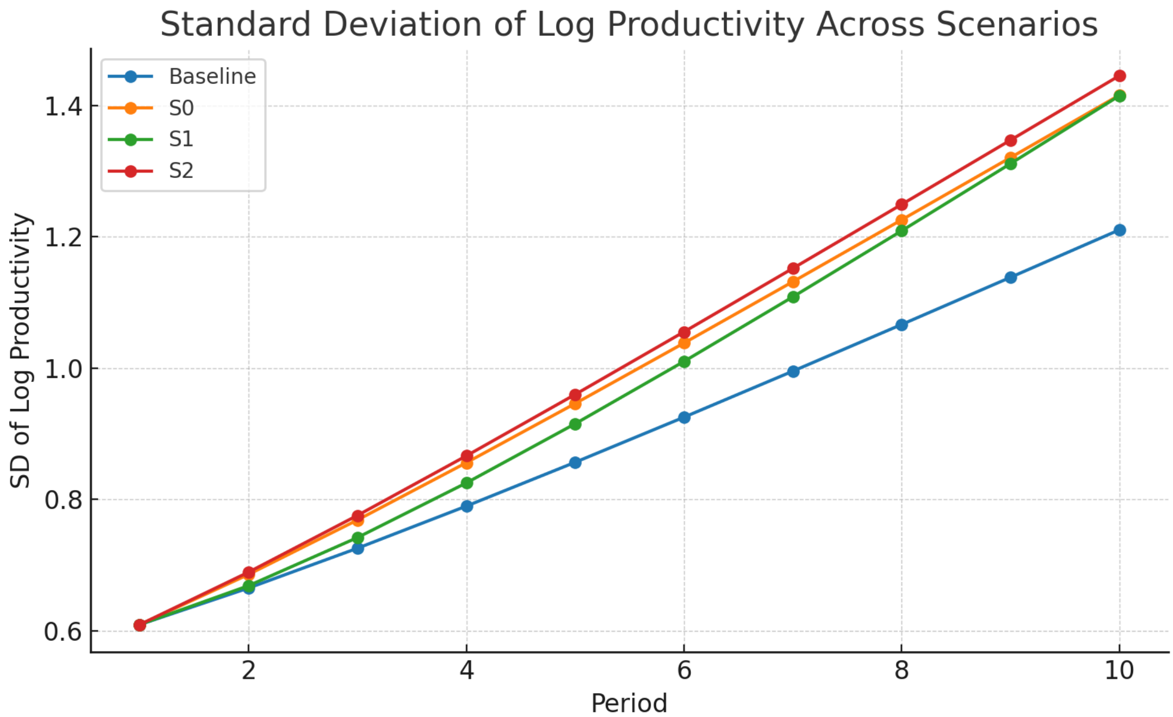

Figure 18: **Divergence or Convergence: Productivity Inequality.**

6.4.3 Utility Inequality

In contrast, the dispersion of log utility tells a different story (Figure 19). Across all scenarios, the standard deviation of log utility either declines or remains nearly flat over time. For example, in S0, dispersion falls slightly from 0.042 in period 1 to 0.033 by period 10, and in S2, the decline is of similar magnitude. This muted or negative trend, despite strong divergence in wages and productivity, can be explained by congestion and amenity effects in top cities. While migrants moving to high-wage, high-productivity cities enjoy higher incomes, these gains are partially offset by higher costs of living, reduced amenities, and crowding effects. As a result, the spatial distribution of welfare is more equal than the distributions of wages or productivity alone would suggest.

Overall, the model predicts that while *hukou* reforms—especially those focused on top-tier cities—can sharply increase income and productivity dispersion, welfare inequality remains contained due to offsetting forces. This highlights a tension: visible economic divergence coexists with muted welfare divergence, a point relevant to policy design.

6.5 Spatial Patterns of Adjustment

Based on the results in Table3, 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

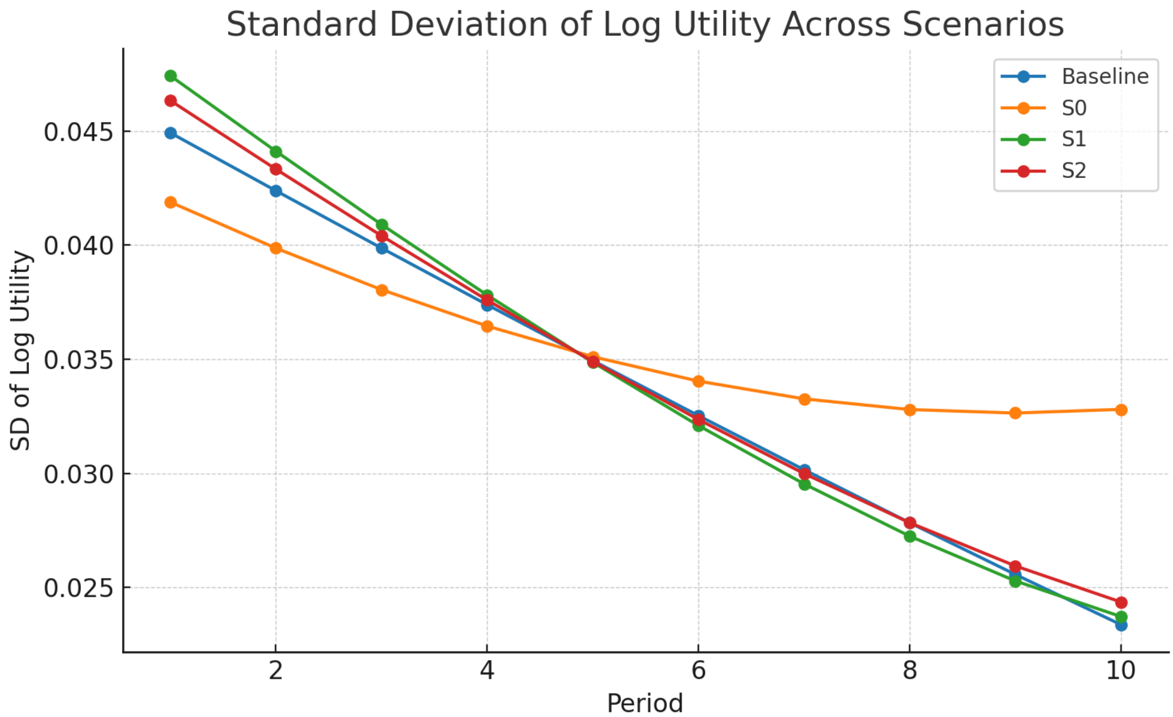

Figure 19: **Divergence or Convergence: Utility Inequality.**

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-sized 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 on infrastructure, housing, and public services during large-scale migration adjustments.

Table 3: **Largest Movers and Welfare Changes Under Full Abolition.**

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 $\rho = 2$.

Figures 20 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 highly uneven spatial responses.

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),

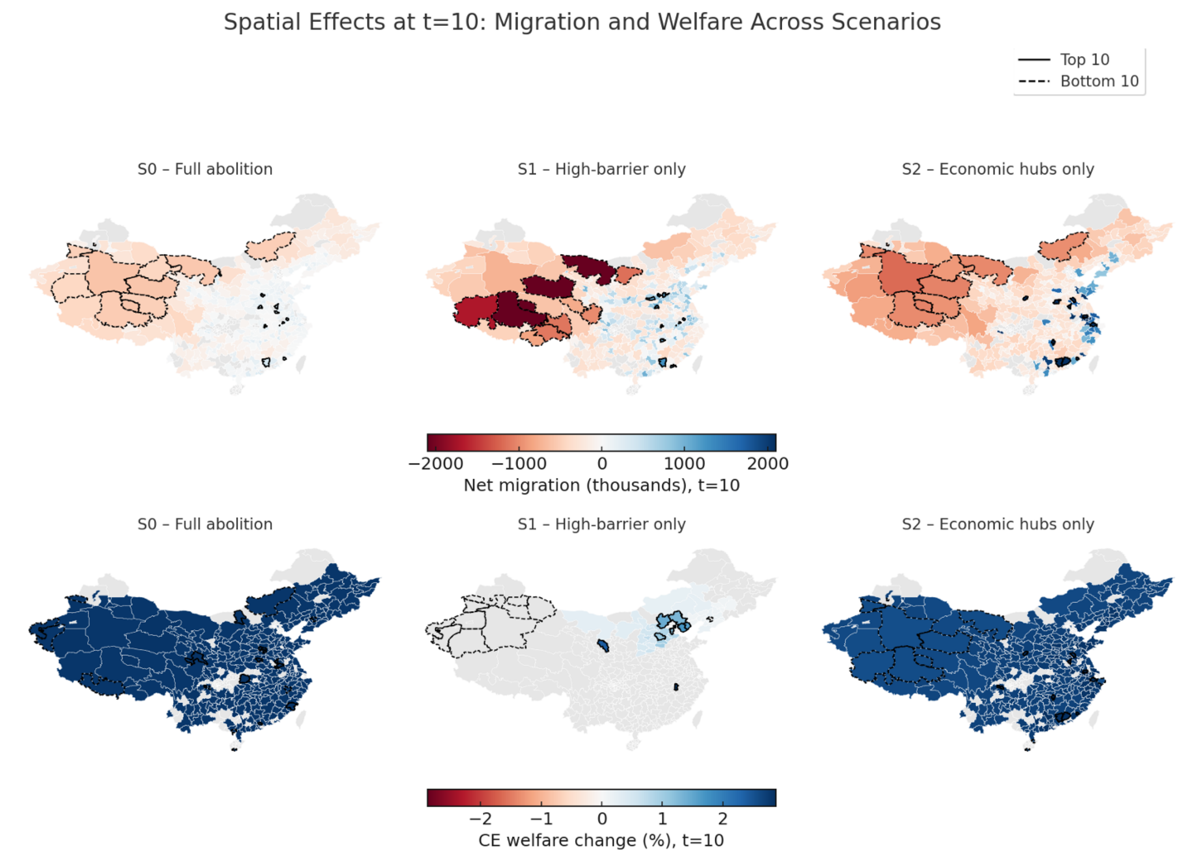

Figure 20: **Net Labor Inflow and Welfare in the Long-Run.**

and in certain large metropolitan areas where congestion effects offset wage advantages. 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 a high-barrier-only scenario, *hukou* restriction is lifted in 128 high-barrier cities (*hukou* cost share $\geq 75\%$), 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 $\geq 1\%$ 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. The central implication is that migration liberalization alone produces uneven local outcomes, shaped by initial conditions and adjustment dynamics.

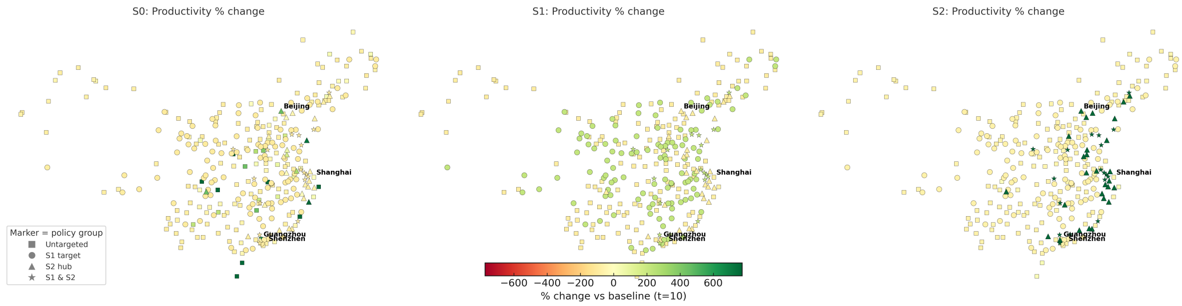

Figure 21: **Productivity Change at $t = 10$.**
Notes: Values are relative to the baseline. Colors indicate percent change. Markers denote policy groups: $\square$ Untargeted, $\circ$ S1, $\triangle$ S2, $\star$ S1&S2.

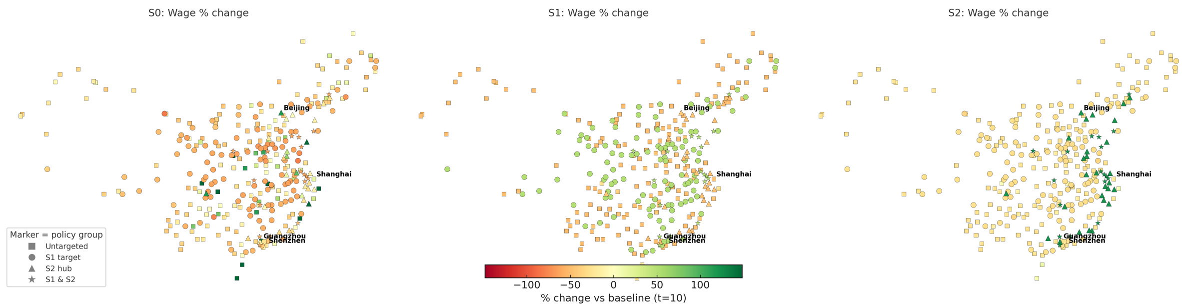

Figure 22: **Wage Change at $t = 10$.**
Notes: Values are relative to the baseline. Colors indicate percent change. Markers denote policy groups as in Fig. 21.

Productivity patterns broadly mirror, but do not perfectly align with the migration outcomes. In Scenario S0, productivity gains are broad-based, with the largest effects in coastal hubs and several inland poles. This pattern is consistent with agglomeration feedback operating on both productivity and wages once migration frictions are removed. In S1, many targeted inland cities show positive migration and wage responses but muted or even short-run declines in measured productivity. This reflects the dilution of average productivity when low-productivity workers move in faster than local technology or scale economies can adjust. Moreover, local scale/technology adjusts with a lag. By contrast, S2's targeted reform in economic hubs delivers immediate productivity boosts in the short run. But over longer horizons, marginal agglomeration returns diminish and congestion/price pressures temper the gains.

Across scenarios, productivity and wage maps are positively related but not one-for-one: in dense recipients, prices rise faster than wages (via non-tradables/land), so real income gains are smaller than productivity gains; in outflow areas, both wages and productivity fall.

6.6 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 disentangle these forces, we decompose the total consumption-equivalent (CE) welfare gain into three structural components, plus their interaction:

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, while holding amenities fixed.

3. **Amenity feedback.** Allow amenities to respond via λ and χ, while holding productivity fixed.

4. **Interaction.** The residual complementarities between productivity and amenity channels when both are active.

We compute the contribution of each channel relative to a common baseline in which only the *hukou* barrier is removed (migration-only scenario). That is, productivity and amenity feedbacks are each "added on" to the migration-only baseline rather than to one another. As a result, their magnitudes cannot be summed directly. The remaining gap between the sum of these single-channel increments and the full model outcome is reported as the interaction term, which captures complementarities when productivity and amenity channels operate simultaneously.

Under full *hukou* abolition, long-run welfare rises by about 53 pp in CE terms. Pure reallocation explains 18.6 pp (35.1% of the total), productivity feedback contributes 9.8 pp (18.5%), and amenity feedback adds 15.6 pp (29.4%). The remaining 8.9 pp (16.7%) reflects complementarities, indicating that productivity and amenity channels reinforce each other when both are active.

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.

Table 4: **Mechanism Decomposition of Long-Run Welfare Gains at $t = 10$.**

Scenario	Incremental (pp)	Share of Total (%)
Migration only (ϕ only)	+18.60	35.2
Add-on effects (relative to ϕ only):		
ϕ + Productivity feedback (λ_1 on; A fixed)	**+9.80**	18.5
ϕ + Amenity feedback (λ, χ on; Z fixed)	**+15.57**	29.5
Interaction ($\phi \times Z \times A$)	**+8.9**	16.8
Full model($\phi + \lambda_1 + \lambda, \chi$)	**52.87**	**100.0**

Notes: All increments are measured relative to the migration-only (ϕ only) situation. The productivity and amenity effects are not additive; their combined operation generates an interaction effect. Together, the baseline + add-ons + interaction sum to the full model gain.

Finally, to ensure that the results are not driven by specific modeling choices or parameter values, I conduct a series of robustness checks, including alternative migration elasticity estimates, amenity specifications, and trade elasticities. The main findings are stable across these exercises. Detailed results are reported in Appendix H.

7 Conclusion and Discussion

This paper develops a dynamic spatial equilibrium model with endogenous productivity, amenities, and migration frictions to quantify the aggregate and spatial effects of abolishing China's *hukou* system. The framework captures three key forces: a dispersion force from lower migration costs, and two agglomeration forces—productivity spillovers and amenity investment. Calibrated to prefecture-level data, the model shows that full *hukou* abolition yields large and persistent gains in productivity, output, and welfare.

Under full reform, aggregate productivity and real output increase by 2.10% and 2.89% annually over the century, respectively, while long-run consumption-equivalent (CE) welfare rises by about 53%. A decomposition of the welfare gains attributes 35.1% to pure spatial reallocation, 18.5% to productivity feedbacks, and 29.4% to amenity feedbacks, with the remainder driven by positive interactions between channels. Amenity-driven migration plays a particularly important role in the Chinese context, reflecting the ability of local governments to finance and deliver public service improvements through land revenues.

Partial reforms deliver smaller welfare changes. Limiting reform to high-barrier cities (S1) redirects migration toward less dynamic regions, dampening productivity and output. Restricting reform to economic hubs (S2) boosts efficiency in the short run but generates congestion and cost pressures that erode long-run welfare.

A new dimension of this analysis concerns the distributional consequences of reform. Even without policy change, wage and productivity inequality rise gradually over time as agglomeration benefits accrue to already-advantaged cities. Removing *hukou* restrictions accelerates this divergence, with the strongest inequality effects observed when reform is targeted to high-productivity hubs (S2). In contrast, utility inequality remains muted or declines, as higher wages in top cities are offset by congestion and amenity crowding. This divergence between income-based and welfare-based measures highlights the importance of considering non-monetary factors in spatial policy evaluation.

The welfare impact of *hukou* reform, therefore, depends critically on the policy objective. If maximizing aggregate output and growth is the priority, targeting economic hubs yields the largest efficiency gains but at the cost of greater spatial inequality. If inclusive development is the goal, reforms focused on high-barrier cities offer a more balanced path, moderating divergence while capturing a meaningful share of the growth potential. Removing migration barriers in isolation delivers substantial but incomplete gains; the largest benefits arise when greater labor mobility is complemented by productivity-enhancing investments and amenity improvements.

These findings have broader relevance for other economies with institutional barriers to internal migration, underscoring the importance of policy packages that jointly promote mobility, productivity, and livability. In the Chinese context, *hukou* reform without attention to spatial inequality could deepen regional divides, whereas coordinated reforms can foster both national growth and more balanced regional development.

7.1 Policy Implications for "Common Prosperity"

China's current "common prosperity" agenda seeks to reduce excessive regional disparities while sustaining economic dynamism. The results of this paper suggest that *hukou* reform, if implemented without complementary measures, could exacerbate spatial wage and productivity gaps despite raising aggregate output. A policy mix that combines mobility liberalization with targeted investments in lagging regions, congestion management in top cities, and mechanisms to ensure equitable access to amenities can align migration policy with the dual goals of growth and equity. In this way, *hukou* reform can serve not only as a catalyst for economic expansion, but also as a pillar of spatially inclusive development.

References

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.

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., & 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.

Cheng, H., Hu, D., & Li, H. (2020). Wage differential between rural migrant and urban workers in the people's republic of china. *Asian Development Review*, *37*(1), 43–60.

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.

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.

Ma, X., Li, Y., & Iwasaki, I. (2024). The hukou system and wage gap between urban and rural migrant workers in china: A meta-analysis. *Economics of Transition and Institutional Change*, *32*(4), 1105–1136.

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.

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

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.

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

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

Wu, X., & Zhang, Z. (2015). *Wage discrimination in urban china: How hukou status affects migrant pay* (Tech. Rep.). HKUST Institute for Emerging Market Studies.

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.

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 $p_t^\omega(n,s)$:

$$p_t^\omega(n,n) = \iota^{-1} w_t(n) l_t(n)^{1-\iota} Z_t(n)^{-1} \tag{40}$$

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.

$$\begin{aligned} p_t^\omega(n,s) &= \upsilon(n,s) p_t^\omega(n,n) \\ &= \upsilon(n,s)\iota^{-1} w_t(n) l_t(n)^{1-\iota} Z_t(n)^{-1}. \end{aligned} \tag{41}$$

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

$$\begin{aligned} &\mathbb{P}\left[p_t^\omega(n,s) \le p\right] \\ &\quad = \mathbb{P}\left[\upsilon(n,s)\iota^{-1} w_t(n) l_t(n)^{1-\iota} Z_t(n)^{-1} \le p\right] \\ &\quad = \mathbb{P}\left[\epsilon_t^\omega(n) \ge \frac{\upsilon(n,s)\iota^{-1} w_t(n) l_t(n)^{1-\iota} Z_t(n)^{-1}}{p}\right] \\ &\quad = 1 - \mathbb{P}\left[\epsilon_t^\omega(n) \le \frac{\upsilon(n,s)\iota^{-1} w_t(n) l_t(n)^{1-\iota} Z_t(n)^{-1}}{p}\right] \\ &\quad = 1 - exp\left\{-(\frac{\upsilon(n,s)\xi_t(n)}{p})^{-\delta}\right\}, \end{aligned} \tag{42}$$

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:

$$\begin{aligned} \pi_t(n,s) &= \mathbb{P}\left[p_t^\omega(n,s) \le \min_{j\in N} p_t^\omega(n,j)\right] \\ &= \int_0^\infty \mathbb{P}\left[\min_{j\in N} p_t^\omega(n,j) \ge p\right] dG_{ns,t}(p) \\ &= \int_0^\infty \prod_{j\in N} \mathbb{P}\left[p_t^\omega(n,j) \ge p\right] dG_{ns,t}(p) \\ &= \int_0^\infty \prod_{j\in N} \left[1 - G_{nj,t}(p)\right] dG_{ns,t}(p). \end{aligned} \tag{43}$$

If we replace $G_{ns,t}(p)$ by its expression in (20), we will have

$$\begin{aligned}
\pi_t(n,s) &= \int_0^\infty \prod_{j\in N} exp\left\{(\frac{\upsilon(j,s)\xi_t(j)}{p})^{-\delta}\right\} exp\left\{(\frac{\upsilon(n,s)\xi_t(n)}{p})^{-\delta}\right\} (\upsilon(n,s)\xi_t(n))^{-\delta}\, dp^\delta \\
&= (\upsilon(n,s)\xi_t(n))^{-\delta} \int_0^\infty \exp\left\{-\sum_j^N (\upsilon(j,s)\xi_t(j))^{-\delta}\, p^\delta\right\} dp^\delta \\
&= \frac{(\upsilon(n,s)\xi_t(n))^{-\delta}}{\sum_j^N (\upsilon(j,s)\xi_t(j))^{-\delta}}
\end{aligned} \tag{44}$$

This gives us the equation (14) 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:

$$\begin{aligned}
\Omega_t(n,j)^h &= \mathbb{P}\left[\tilde{u}_t(n)\tilde{m}_t(n,j)^{-1}\epsilon_t(n) \geq \tilde{u}_t(s)\tilde{m}_t(s,j)^{-1}\epsilon_t(s)\right] \\
&= \mathbb{P}\left[\tilde{u}_t(n)\tilde{m}_t(n,j)\epsilon_t(n) \geq \tilde{u}_t(s)\tilde{m}_t(s,j)\epsilon_t(s), s \neq n\right] \\
&= \mathbb{P}\left[\frac{\epsilon_{s,t}}{\epsilon_{n,t}} \leq \frac{\tilde{u}_t(n)^{-1}\tilde{m}_t(n,j)^{-1}}{\tilde{u}_t(s)\tilde{m}_t(s,j)}\right]
\end{aligned} \tag{45}$$

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:

$$\begin{aligned}
\mathbb{P} &= \prod_{s\neq n} \exp\left(-\left(\frac{\tilde{u}_t(s)\tilde{m}_t(s,j)}{\tilde{u}_t(n)\tilde{m}_t(n,j)}\right)^\gamma\right) \\
&= \exp\left(-\sum_{s\neq n}\left(\frac{\tilde{u}_t(s)\tilde{m}_t(n,s)}{\tilde{u}_t(n)\tilde{m}_t(n,j)}\right)^\gamma\right)
\end{aligned} \tag{46}$$

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

$$\Omega_t^h(n,j) = \frac{\tilde{u}_t(n)^\gamma \tilde{m}_t(n,j)^{-\gamma}}{\sum_s^N \tilde{u}_t(s)^\gamma \tilde{m}_t(n,s)^{-\gamma}} \tag{47}$$

Appendix B 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 (19), 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:

$$\left[\Gamma\left(\frac{1-\sigma}{\delta}+1\right)\right]^{\frac{-\delta}{1-\sigma}}\sum_n^N(\xi_t(n)v(n,s))^{-\delta} = (\frac{1-\iota}{\iota})^{\chi}\left[\frac{\bar{A}_t(s)}{\tilde{u}_t(s)}\right]^{-\delta}\bar{l}_t(s)^{-\lambda\delta}w_t(s)^{-\delta(1+\chi)}. \tag{48}$$

Combining (44) and (48), the probability of goods produced in n and consumed in s becomes:

$$\pi_t(n,s) = (v(n,s)\xi_t(n))^{-\delta}\left[\Gamma\left(\frac{1-\sigma}{\delta}+1\right)\right]^{\frac{-\delta}{1-\sigma}}\left[\frac{\bar{A}_t(s)}{\tilde{u}_t(s)}\right]^{\delta}\bar{l}_t(s)^{\lambda\delta}w_t(s)^{\delta} \tag{49}$$

With (49), the trade balance condition can be expanded as:

$$w_t(n)H(n)\bar{l}_t(n)\xi_t(n)^{\delta} = \left[\Gamma\left(\frac{1-\sigma}{\delta}+1\right)\right]^{\frac{-\delta}{1-\sigma}}\sum_s^N\left[\frac{\bar{A}_t(s)}{\tilde{u}_t(s)}\right]^{\delta}H(s)\bar{l}_t(s)^{1+\lambda\delta}w_t(s)^{1+\delta}v_t(n,s)^{-\delta} \tag{50}$$

Plug in $\xi_t(n)$ and rearrange (50):

$$\begin{aligned} &w_t(n)^{1+\delta}H(n)\bar{l}_t(n)^{1+(1-\iota)\delta}\tilde{z}_t(n)^{-\delta} \\ =&\left[\Gamma\left(\frac{1-\sigma}{\delta}+1\right)\right]^{\frac{-\delta}{1-\sigma}}\sum_s^N\left[\frac{\bar{A}_t(s)}{\tilde{u}_t(s)}\right]^{\delta}H(s)\bar{l}_t(s)^{-\lambda\delta}w_t(s)^{-\delta(1+\chi)}v_t(n,s)^{-\delta}\end{aligned} \tag{51}$$

Appendix C 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.

C.1 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.

C.2 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.

C.3 China Migrants Dynamic Survey (CMDS)

Conducted annually since 2010 by the National Health Commission of the People's Republic of China, it 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.

C.4 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.

The ArcGIS OD Cost Matrix Tool, part of Esri's ArcGIS software suite, was employed to estimate bilateral trade costs between cities. 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.

Appendix D 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 were 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 were 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.

Appendix E Estimation of Parameters

E.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:

$$A_t(n) = \bar{A}(n) \cdot \bar{l}_t(n)^{\lambda} \cdot I_t(n)^{\chi}, \tag{52}$$

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:

$$\log A_t(n) = \log \bar{A}(n) + \lambda \log \bar{l}_t(n) + \chi \log I_t(n). \tag{53}$$

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.

E.1.1 Constructing the Amenity Index

We measure $A_t(n)$ as a composite index summarizing multiple observable dimensions of local livability. The index is constructed using Principal Component Analysis (PCA) on standardized indicators drawn from five broad categories:

1. **Environmental Quality.** Industrial SO_2 emissions (tons/km^2), green coverage rate, arable land (thousand hectares), and noise-affected surface area.

2. **Transportation Infrastructure.** Paved road area per capita, public transit vehicles per 10,000 persons, and taxis per 10,000 persons.

3. **Healthcare.** Hospital beds per capita and physicians per capita.

4. **Education.** Public education expenditure per capita, schools per capita, and full-time teachers per capita.

5. **Cultural Environment.** Number of cinemas per capita and library books per capita.

All indicators are first normalized to remove scale effects. We then apply PCA and retain the first seven principal components, which jointly explain the majority of the variance across cities, to form the composite amenity index $A_t(n)$.

E.1.2 Measuring Amenity Investment

Amenity investment, $I_t(n)$, is proxied by per capita expenditures on city maintenance and environmental protection, sourced from the *China City Statistical Yearbooks* (base year 2000). These expenditures are closely linked to local government land-lease revenues, which finance public service and infrastructure improvements.

E.1.3 Exogenous Amenities $\bar{A}(n)$

To separate the exogenous (geography-driven) component of amenities from the endogenous component shaped by economic forces, we regress $\log A_t(n)$ on time-invariant geographical fundamentals:

1. **Topography.** Topographic relief index.

2. **Climate.** Average annual temperature and humidity.

3. **Water Resources.** Share of water bodies in total land area (National Geographic Information Center).

The predicted values from this regression, $\hat{\bar{A}}(n)$, capture the geography-driven amenity level, while the residuals represent the endogenous component responsive to density and investment.

E.1.4 Estimation of λ and χ

We estimate the relationship between endogenous amenities, population density, and amenity investment using the following specification:

$$\log A_t(n) - \log \hat{\bar{A}}(n) = \lambda \log \bar{l}_t(n) + \chi \log I_t(n) + \epsilon_t(n), \tag{54}$$

where $\bar{l}_t(n)$ denotes population density. To address endogeneity concerns (e.g., population density and amenities being jointly determined), we use deep lags of $\bar{l}_t(n)$ and $I_t(n)$ as instruments. Standard errors are clustered at the regional level.

E.1.5 Results

The GMM estimates are:

$$\hat{\lambda} = -0.39, \quad \hat{\chi} = 0.23.$$

The negative λ confirms congestion effects from higher density, while the positive χ indicates that land-rent-financed amenity investments significantly enhance local amenity levels. These parameters are subsequently used in model calibration and sensitivity analysis.

E.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:

$$\log(\text{GDPpc}_{it}) = \lambda_1 \log(\text{PopDensity}_{i,t-1}) + \alpha \log(\text{GDPpc}_{i,t-1}) + \epsilon_{it}, \tag{55}$$

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 5: **Estimation of λ_1.**

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, & 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, & 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.

Appendix F 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.

F.1 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.

F.2 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).

Appendix G 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:

$$\frac{\widehat{\text{PDV}_U}}{\text{PDV}_U} = (1+g)^{1-\sigma} \quad \Rightarrow \quad g = \left(\frac{\widehat{\text{PDV}_U}}{\text{PDV}_U}\right)^{\frac{1}{1-\sigma}} - 1, \tag{56}$$

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:

$$g = \frac{\text{PDV}^{\text{cf}}}{\text{PDV}^{\text{base}}} - 1. \tag{57}$$

Example – Full *Hukou* Abolition Scenario

Table 6: **PDV of Welfare Gains under Full *Hukou* Abolition.**

Horizon	**PDV$_{\text{base}}$**	**PDV$_{\text{cf}}$**	**CE Welfare Gain** g (%)
$t = 2$	14,000,000,000	20,900,000,000	+48.51
$t = 10$	30,800,000,000	47,000,000,000	+52.87

Notes: PDV denotes the present discounted value of aggregate consumption in the baseline (PDV$_{\text{base}}$) and counterfactual full abolition scenario (PDV$_{\text{cf}}$). g represents the percentage change in consumption-equivalent (CE) welfare relative to the baseline equilibrium.

These figures indicate that households would require nearly 49% higher permanent consumption in the baseline to match their welfare in the counterfactual by $t = 2$, and over 52% 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.

Appendix H Sensitivity and Robustness

This subsection examines how the long-run welfare gains from the full abolition of *hukou* barriers (Scenario S0) 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, shaped by China's industrial clustering in manufacturing and services, especially in coastal growth poles. The migration elasticity γ measures the responsiveness of labor flows to utility differences across locations. In the model, migration costs arise from two sources—institutional frictions from the *hukou* system and distance-based costs—so γ governs how strongly workers reallocate once those frictions change. Finally, the amenity–investment elasticity χ reflects how effectively local governments convert land-lease revenues into public services and infrastructure—an important channel in China, where land finance accounts for a substantial share of local fiscal revenue.

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. The ranges are summarized in Table 7.

Table 7: **Parameter Ranges for Sensitivity Tests.**

Parameter	Baseline	Low	High
Amenity–density (λ)	−0.39	−0.50	−0.20
Productivity–density (λ_1)	0.21	0.15	0.30
Migration elasticity (γ)	2.00	1.00	3.00
Amenity–investment (χ)	0.23	0.15	0.35

Table 8: **Sensitivity of Long-Run CE Welfare Gains at $t = 10$.**

Scenario	Baseline	Low λ	High λ	Low λ_1	High λ_1
S0: Full abolition	52.87	48.67	56.40	49.80	58.80

Scenario	Low γ	High γ	Low χ	High χ
S0: Full abolition	45.30	61.00	48.20	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.

While a full robustness check across all counterfactual scenarios is left for future work, the S0 results provide an informative benchmark. Table 8 shows that the welfare gains from complete *hukou* abolition are most sensitive to γ and χ, highlighting the importance of migration responsiveness and the role of public-investment-driven amenity improvements. By contrast, varying λ or λ_1 within empirically reasonable bounds changes the magnitudes more moderately, suggesting that congestion and agglomeration spillovers, while important, are less decisive in determining aggregate gains in this policy experiment.

Overall, these results suggest that the main policy message—nationwide *hukou* abolition yields large long-run welfare gains—remains qualitatively robust to reasonable parameter variation.

Appendix I Prefectures 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.

Appendix J Institutional Background: The *Hukou* System

This appendix provides additional institutional context on the *hukou* system in China. It traces the system's origins, major policy adjustments, and the decentralized governance structure that continues to shape migration outcomes today.

J.1 Historical Origins: 1950s–1970s

The *hukou* system was formally established in 1958 under the Regulations on Household Registration of the People's Republic of China. Its original purpose was to control rural-to-urban migration during the early stages of industrialization, when the state sought to ensure stable grain supply, prevent overcrowding in urban areas, and protect limited public services. By tying access to education, healthcare, housing, and food rations to an individual's registered *hukou* location, the system effectively bound people to their birthplace. For more than two decades, internal migration required explicit state authorization or employer sponsorship, and unauthorized relocation often led to fines or forced return. *Hukou* thus became a cornerstone of China's planned economy, deeply shaping patterns of labor allocation and urbanization.

J.2 Reforms and Decentralization: 1970s–1990s

The launch of the Reform and Opening policy in 1979 marked the first step toward loosening *hukou* restrictions. Authority over *hukou* transfers was gradually delegated from the central Ministry of Public Security to provincial and local governments, creating scope for regional variation. Temporary residence permits (zanzhuzheng) were introduced in the early 1980s, allowing rural migrants to relocate for work without gaining full local registration. Meanwhile, rapid industrial expansion in coastal provinces created unprecedented demand for labor, fueling large flows of rural migrants to cities. Although these changes facilitated mobility, access to public services remained largely restricted to registered residents, reinforcing the divide between locals and migrants.

J.3 Gradual Adjustment: 2000s

Policy adjustments in the 2000s reflected growing recognition of migrants' role in urban economies. In 2001, the State Council encouraged mid-sized cities to relax *hukou* criteria in order to absorb surplus rural labor. A landmark change came in 2003 with the abolition of the custody and repatriation system, which had allowed local authorities to detain and forcibly return migrants without *hukou*. In the mid-2000s, several provinces experimented with pilot programs that extended partial access to schooling for migrant children or included migrants in limited social insurance schemes. Nonetheless, coverage remained uneven and contingent on local fiscal capacity. By 2010, the floating population had exceeded 200 million, underscoring both the scale of migration and the persistence of institutional exclusion.

J.4 Recent Reforms and Persistent Fragmentation: 2010–present

The 2010s brought renewed calls for *hukou* reform, culminating in the National New-Type Urbanization Plan of 2014. This plan encouraged cities with fewer than three million residents to remove *hukou* restrictions entirely, while instructing larger cities to implement points-based or quota systems. In 2016, the rural–urban *hukou* distinction was formally abolished, but locality-based restrictions remained in place, meaning that migrants without local registration continued to face barriers to public services. Today, Tier-1 metropolises such as Beijing, Shanghai, Shenzhen, and Guangzhou maintain highly restrictive systems, often tying eligibility to educational attainment, housing ownership, or tax contributions. By contrast, many smaller cities have significantly relaxed criteria to attract labor and investment.

J.5 Fragmented Governance

The decentralization of *hukou* management has produced a fragmented institutional landscape. Local governments have substantial discretion to set admission rules in line with their fiscal capacity, demographic pressures, and development strategies. This discretion has resulted in starkly different regimes across cities. Shanghai employs a transparent points-based system, rewarding education, tax contributions, and employment in skilled sectors, while Beijing operates a more opaque employer-linked quota system with limited transparency. Smaller cities, facing depopulation or labor shortages, often provide relatively straightforward pathways to *hukou* acquisition. The outcome is a patchwork of local rules that determine migrants' settlement opportunities in highly uneven ways.

This fragmented governance structure directly motivates the counterfactual policy scenarios examined in the main text. The three modeled interventions—nationwide abolition (S0), selective reforms in high-barrier cities (S1), and reforms concentrated in major economic hubs (S2)—mirror the diversity of current institutional practices and allow for an evaluation of their long-run implications for labor mobility, productivity, and welfare.